GEOPHYSICS, VOL. 84, NO. 2 (MARCH-APRIL 2019); P. P15–P25, 13 FIGS., 5 TABLES.
10.1190/GEO2018-0145.1

Bayesian modeling for uncertainty quantification in seismic compressive sensing

Georgios Pilikos[1] and A. C. Faul[1]

ABSTRACT

Compressive sensing is used to improve the efficiency of seismic data acquisition and survey design. Nevertheless, most methods are ad hoc, and their only aim is to fill in the gaps in the data. Algorithms might be able to predict missing receivers' values, however, it is also desirable to be able to associate each prediction with a degree of uncertainty. We used beta process factor analysis (BPFA) and its variance. With this, we achieved high correlation between uncertainty and respective reconstruction error. Comparisons with other algorithms in the literature and results on synthetic and field data illustrate the advantages of using BPFA for uncertainty quantification. This could be useful when modeling the degree of uncertainty for different source/receiver configurations to guide future seismic survey design.

INTRODUCTION

Seismic data acquisition involves sampling the seismic wavefield at or near the earth's surface. A source at the surface creates a wavefield that is reflected and refracted by changes in impedance. Surface receivers record the reflected wavefield generally on a regular grid. But some of those receivers may be missing, caused either by malfunction or because they could not be placed in the survey's required location (e.g., because of a surface obstruction). To overcome this, signal reconstruction algorithms are used to replace or restore the output of the missing receivers (traces). Most of the modern algorithms use the principle of compressive sensing (CS), which uses the assumption that the signal of interest is either sparse (a few nonzero elements) in nature or in some other basis. In the seismic CS literature, sparsity is assumed using the Fourier (Sacchi et al., 1998), the Radon (Trad et al., 2002), the curvelet (Herrmann and Hennenfent, 2008), or the focal transform (Kutscha and Verschuur, 2016) to name a few.

A popular method that uses the Fourier transform is the projection onto convex sets (POCS) (Abma and Kabir, 2006), which transforms the available data and uses hard or soft thresholding (Stanton et al., 2015) to reconstruct the desired signals. Iteratively reweighted least squares were also proposed (Zwartjes and Sacchi, 2007) in the Fourier domain. Spectral projected gradient for L1 (SPGL1) (van den Berg and Friedlander, 2009) is another method that solves the problem by using a predefined dictionary of basis functions that provide a sparse representation to solve the l_1-norm minimization problem. Tensor completion (Kreimer and Sacchi, 2011, 2012) algorithms were also proposed to scale for larger dimensions. Other techniques with prediction filters use nonaliased low frequencies of seismic data to reconstruct the aliased parts (Spitz, 1991; Porsani, 1999; Naghizadeh and Sacchi, 2007). Alternative methods have been proposed that do not require a predefined dictionary of basis functions to be used such as Beckouche and Ma (2014), Zhu et al. (2015), and Turquais et al. (2015). These techniques train on available seismic data and learn a dictionary that can be used for sparse representation.

Recently, new approaches in seismic CS have been proposed that use principles from the Bayesian statistics and machine learning literature. The relevance vector machine (RVM) (Tipping, 2001; Tipping and Faul, 2003) has been applied to seismic interpolation (Pilikos and Faul, 2016) with success, matching the performance of the SPGL1 for time slices. In addition, Pilikos and Faul (2016) use an uncertainty measure improvement and illustrate an uncertainty map for the predictions. Another method that has been used is the beta process factor analysis (BPFA) to learn a dictionary of basis from the available seismic data for interpolation and denoising (Pilikos and Faul, 2017). It has been shown that BPFA outperforms the SPGL1 and POCS when processing time slices. Furthermore, when the data are reordered in the shot record domain, there is no sign of aliasing in the frequency wavenumber (f-k) domain (Pilikos et al., 2017).

Manuscript received by the Editor 27 February 2018; revised manuscript received 17 August 2018; published ahead of production 21 November 2018; published online 11 February 2019.
[1]University of Cambridge, Laboratory for Scientific Computing, Maxwell Centre, Department of Physics, J. J. Thomson Avenue, Cambridge CB3 0HE, UK. E-mail: ggp29@cam.ac.uk; acf22@cam.ac.uk.

The core idea behind Bayesian machine learning is the construction of models using probability distributions over random variables. This provides flexibility in modeling because it is possible to incorporate prior knowledge and guide the model. Furthermore, it allows the model to provide uncertainty information about its predictions. Bayesian statistics has a long history in solving inverse problems in geophysics. Duijndam (1988a, 1988b) gives a comprehensive introduction to the field, and later Ulrych et al. (2001) write a tutorial with seismic applications. Malinverno and Briggs (2004) expand this using empirical Bayes for uncertainty quantification. Other applications of Bayesian estimation can be found in Wang et al. (2008) for seismic wavefield separation and Fjeldstad and Grana (2018) for petrophysics-seismic inversion to name a few. In this paper, we will use Bayesian machine learning, as mentioned in the previous paragraph, for seismic CS to create probabilistic data-driven models and at the same time create uncertainty maps. To avoid confusion, this type of models do not refer to velocity models but rather to constructions with general assumptions that are adaptable to the available data.

The first model that we will use is the RVM. This model uses a sparsity promoting prior distribution in the form of a hyper-prior over the coefficients of a linear combination of basis functions. By learning the appropriate parameters, the model can provide a predictive mean and predictive variance of the desired model. These can then be used for prediction and uncertainty quantification. Nevertheless, the predictive variance was found to behave counter-intuitively when using basis functions with finite support (Rasmussen and Quiñonero Candela, 2005), providing small uncertainty when data points are far from the model and vice versa. To overcome the problematic predictive variance, Faul and Pilikos (2016) propose a new uncertainty measure for the RVM. This calculates the expected change in the likelihood that the predicted data point would have. Using this proposed measure, Pilikos and Faul (2016) apply it to seismic data with some preliminary results.

The second model that we will use is the BPFA (Zhou et al., 2012). This model uses a different approach to enforce sparsity on the coefficients of the linear combination of the desired model. This is achieved using a Bernoulli distribution to control whether a coefficient is zero or not. The parameter that controls the Bernoulli distribution is governed by a Beta distribution to allow flexibility in the level of sparsity. In addition, this is then element-wise multiplied with a normal distribution to produce the value of a desired coefficient. This method of modeling provides exact zero coefficients as opposed to the RVM. Another advantage is that it also learns a dictionary of basis from the available data. This provides another level of flexibility that makes it more accurate in reconstructions. BPFA was compared with other algorithms in the literature and obtained state-of-the-art reconstructions without any signs of aliasing in the f-k domain (Pilikos and Faul, 2017; Pilikos et al., 2017). In this paper, we propose to use BPFA to create uncertainty maps. We calculate the variance for each prediction obtained by the inference process that uses Gibbs sampling. Using the variance, we show that it is possible to obtain much better uncertainty maps for the reconstructed signals compared with others in the literature.

The structure of the paper is as follows: first, an introduction to Bayesian machine learning is given, providing basic definitions and explanation of the RVM. Various modifications to the predictive variance of the RVM are also discussed. In addition, BPFA is described along with how the variance of its predictions can be used to create uncertainty maps for seismic CS. Experiments on sections of time slices are provided along with representative uncertainty maps and reconstructions for all algorithms. Furthermore, uncertainty maps for shot records are provided. A thorough comparison and analysis on thousands of uncertainty maps using different methods illustrates their performance in detail using the Spearman's correlation coefficient. Stacking of uncertainty maps is also provided that improves the correlation with the reconstruction error. Finally, we include an example on field data along with conclusions.

BAYESIAN MACHINE LEARNING

Using data-driven models to describe real-world observations has increased in popularity. Uncertainty is an integral part of the model and the measurements, and models that are able to capture it are very desirable. Bayesian machine learning is a framework that tackles this by allowing the construction of models using probability distributions over random variables. The Bayes rule is defined by

$$p(\boldsymbol{\Theta}|\mathbf{K}) = \frac{p(\mathbf{K}|\boldsymbol{\Theta})p(\boldsymbol{\Theta})}{p(\mathbf{K})}, \tag{1}$$

where $\boldsymbol{\Theta}$ is the collection of all unknown variables and $\mathbf{K}$ are the available observations. The term $p(\boldsymbol{\Theta})$ is the prior distribution of the variables that capture our prior belief of how they are distributed, $p(\mathbf{K}|\boldsymbol{\Theta})$ is the likelihood function that gives the probability of the observations being generated using a respective configuration of $\boldsymbol{\Theta}$, and $p(\mathbf{K})$ is the distribution of the observations given in the current model. Finally, $p(\boldsymbol{\Theta}|\mathbf{K})$ is the posterior distribution of the variables given in the observations. This is the quantity of interest to learn the model and make predictions.

Research in Bayesian machine learning involves the construction of models by defining appropriate prior distributions and likelihood functions. Then, an inference algorithm is used that is able to infer the variables of interest. In the next sections, we will discuss two models that incorporate sparsity for CS using the Bayesian framework.

The RVM

Consider a collection of training data $\mathbf{K} = \{\mathbf{k}^{(i)}\}_{i=1}^{N}$, with each $\mathbf{k}^{(i)} \in \mathbb{R}^c$ being the coordinates of a receiver (stored in a vector). Each receiver has a scalar corresponding target magnitude, $t^{(i)} \in \mathbb{R}$. A Bayesian regression problem is the prediction of $t^{(*)}$ from an unseen $\mathbf{k}^{(*)}$. To be able to make predictions, a model has to be constructed that effectively describes the training data. A dictionary of basis functions $\phi_l, l = 1, \ldots, L$ is used as part of the model to transform the input space into a desired domain. For CS, this domain has to be sparse. The basis functions are assumed predefined, fixed, and nonlinear. The problem is still linear in the coefficients but in a transformed input space. The model becomes

$$t^{(i)} = \sum_{l=1}^{L} w_l \phi_l(\mathbf{k}^{(i)}) + \epsilon^{(i)} = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{k}^{(i)}) + \epsilon^{(i)}, \tag{2}$$

where $\mathbf{w} \in \mathbb{R}^L$ are the coefficients of the linear combination of the transformed input data, $\epsilon^{(i)} \sim \mathcal{N}(0, \sigma^2)$ is the independent and identically distributed (i.i.d.) additive Gaussian noise, and $\boldsymbol{\phi}(\mathbf{k}^{(i)}) = [\phi_1(\mathbf{k}^{(i)}), \phi_2(\mathbf{k}^{(i)}), \phi_l(\mathbf{k}^{(i)}), \ldots, \phi_L(\mathbf{k}^{(i)})]^T$ with each $\phi_l(\mathbf{k}^{(i)})$ being a certain basis function applied to a particular data point. The corresponding model likelihood is given by

$$p(t^{(i)}|\mathbf{w},\mathbf{K},\sigma^2)=\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(t^{(i)}-\mathbf{w}^T\phi(\mathbf{k}^{(i)}))^2\right\}. \tag{3}$$

A required assumption is that the acquired signal is sparse (or could be transformed to a sparse domain). Thus, a prior probability distribution on the coefficients $\mathbf{w}$ that promote sparsity is used. A normal prior distribution is preferred for ease of calculations. Each coefficient w_l is associated with a different variance controlled by the precision (inverse of variance) α_l. The prior distribution is given by

$$p(\mathbf{w}|\alpha)=\prod_{l=1}^{L}\mathcal{N}(w_l;0,\alpha_l^{-1})=\prod_{l=1}^{L}\sqrt{\frac{\alpha_l}{2\pi}}e^{-\frac{\alpha_l}{2}w_l^2}, \tag{4}$$

which is a product of zero mean normal distributions with a separate precision α_l. This allows some probability mass to potentially concentrate on a few w_l (with small α_l) and others to be zero or close to zero (with infinite α_l) resulting in the desired sparsity. A suitable prior for α is the Gamma distribution (Tipping, 2001). During the inference stage, some coefficients $\mathbf{w}$ are deemed relevant, and others tend to zero hence the name of the RVM.

Inference involves finding the optimum configuration of α that maximizes the model likelihood. We use a fast version of the RVM as proposed by Tipping and Faul (2003). After obtaining suitable maximizing estimates for α and thus $\mathbf{w}$, these can be used for predictions. A predictive distribution is constructed (Tipping, 2001) as a normal distribution $\mathcal{N}(m_*,\sigma_*^2)$, where $m_*=\boldsymbol{\mu}^T\phi(\mathbf{k}^{(*)})$ and $\sigma_*^2=\sigma^2+\phi(\mathbf{k}^{(*)})^T\Sigma\phi(\mathbf{k}^{(*)})$. The terms $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ are given by

$$\boldsymbol{\Sigma}=(\sigma^{-2}\boldsymbol{\Phi}^T\boldsymbol{\Phi}+\mathbf{A})^{-1}\quad\text{and}\quad\boldsymbol{\mu}=\sigma^{-2}\boldsymbol{\Sigma}\boldsymbol{\Phi}^T\mathbf{t}, \tag{5}$$

where $\mathbf{A}=\text{diag}(\alpha_1,\alpha_2,\ldots,\alpha_L)$ and $\boldsymbol{\Phi}\in\mathbb{R}^{N\times L}$ is the basis matrix containing all basis functions at all measured data points. The function $\phi(\mathbf{k}^{(*)})$ contains all basis functions evaluated at $\mathbf{k}^{(*)}$, m_* is called the predictive mean, and σ_*^2 is called the predictive variance where the former gives the prediction and the latter gives the uncertainty.

This predictive distribution is heavily dependent on the model because it depends on $\phi(\mathbf{k}^{(*)})$. It is customary to choose basis functions for the dictionary, which decay quickly when moving away from their center, or basis functions with finite, compact support. Therefore, the degenerate case is possible, that is, $\phi(\mathbf{k}^{(*)})$ is close to, or even equal to zero, and thus, the predictive probability distribution becomes $\mathcal{N}(0,\sigma^2)$, which is meaningless. This was noted by Rasmussen and Quiñonero Candela (2005) with examples of predictive variances opposite to what would be desirable (small predictive variance should be close to the training data and large away from it). Thus, we use the discrete cosine transform (DCT) as dictionary of basis functions with the RVM to minimize the problem of degeneration and at the same time obtain higher reconstruction accuracy (Pilikos and Faul, 2016).

Rasmussen and Quiñonero Candela (2005) propose to augment the RVM by adding a basis function centered at the test point that might potentially be far from the support of all the previously added basis functions. By doing that, the training of the model does not change before test time. Thus, for every missing/test data point, one new basis function is added. We will use this to illustrate a possible uncertainty map obtained by the augmentation of the RVM. However, to achieve the best possible results with the RVMs augmentation, tuning is necessary such as the choice of the new basis function and its corresponding parameters. In addition, a corresponding weight and precision of that weight α_* need to be set. Tuning the RVM with augmentation is not the purpose of this paper, and thus, we only include one example during comparisons.

Expected change in model likelihood

We will assume that we have trained a model of the RVM with the available data and then treat the missing receivers as new data points. Following the approach in Faul and Pilikos (2016), we calculate the expected change in model likelihood (ExpCML) for the current model, when a data sample $(\mathbf{k}^{(*)},t^*)$ is added. Let $\mathcal{S}$ be a subset of the samples, this could be all available or neighboring receivers of $\mathbf{k}^{(*)}$. We estimate the probability distribution of t^* to be normal with mean and variance

$$\bar{m}=\underset{\mathbf{k}^{(i)}\in\mathcal{S}}{\text{mean}}\{t^{(i)}\},\qquad\bar{\sigma}^2=\underset{\mathbf{k}^{(i)}\in\mathcal{S}}{\text{var}}\{t^{(i)}\}. \tag{6}$$

With this estimate, the expected change when considering $\mathbf{k}^{(*)}$ in the logarithm of the marginal likelihood (Faul and Pilikos, 2016) is given by

$$E[\Delta\mathcal{L}]=\log\frac{1}{\sqrt{2\pi}\sigma_*}-\frac{\bar{\sigma}^2+(\bar{m}-m_*)^2}{2\sigma_*^2}, \tag{7}$$

where $\Delta\mathcal{L}$ is the change in the logarithm of the marginal likelihood. This expected change creates an uncertainty map with the largest negative values being the most uncertain regions. For further details, refer to Faul and Pilikos (2016).

Beta process factor analysis

For the previous methods, the discussion was given with a predefined dictionary of basis functions that limit the modeling capabilities. BPFA (Paisley and Carin, 2009; Zhou et al., 2012) is a method that overcomes this limitation. We will use $\mathbf{D}\in\mathbb{R}^{d\times L}$ to denote the dictionary of basis to distinguish it from the predefined dictionary of basis functions where L is the number of basis. We assume that the data matrix $\mathbf{X}\in\mathbb{R}^{d\times T}$ is generated by an underlying process with its columns $\mathbf{x}^{(i)}$, $i=1,\ldots,T$ each corresponding to a patch of seismic data. Note that T is the number of patches that the algorithm uses. In all our experiments, we will use a patch size of 8×8 in sections of 128×128 of time slices (prestack).

Each $\mathbf{x}^{(i)}$ contains 64 entries of $\mathbf{k}^{(i)}$ because each $\mathbf{k}^{(i)}$ corresponds to a receiver and each patch contains 8×8 receivers. Each patch is generated by

$$\mathbf{x}^{(i)}=\mathbf{D}\mathbf{w}^{(i)}+\boldsymbol{\epsilon}^{(i)}, \tag{8}$$

where $\mathbf{w}^{(i)}$ contains the patch's coefficients. In the case of training the BPFA model with missing data, a sampling matrix is required. We will use $\boldsymbol{\Delta}^{(i)}\in\{0,1\}^{m^{(i)}\times d}$ constructed by removing rows from the identity matrix of the corresponding missing locations in $\mathbf{x}^{(i)}$ and $m^{(i)}$ being the number of available measurements in that specific patch. We will denote $\mathbf{y}^{(i)}\in\mathbb{R}^{m^{(i)}}$ and define it by

$$\mathbf{y}^{(i)} = \mathbf{\Delta}^{(i)}\mathbf{x}^{(i)}. \tag{9}$$

Note that in practice, we do not collapse the signal as in the RVM but rather insert zeros at the missing locations.

Continuing with the definition of the model, the coefficients $\mathbf{w}^{(i)}$ depend on two variables given by $\mathbf{w}^{(i)} = \mathbf{z}^{(i)} \odot \mathbf{s}^{(i)}$, where $\mathbf{z}^{(i)} \in \{0,1\}^L$ and $\mathbf{s}^{(i)} \in \mathbb{R}^L$ correspond to the sparsity and the actual value, respectively. This means that if the coefficient is nonzero, the corresponding basis is used when generating $\mathbf{x}^{(i)}$. The term $\epsilon^{(i)}$ is modeled as zero mean Gaussian noise. Using the likelihood for the variables and appropriate prior distributions (Pilikos and Faul, 2017), the Bayes rule as defined in equation 1 gives the posterior distributions for $\mathbf{D}$ and $\mathbf{w}^{(i)} = \mathbf{z}^{(i)} \odot \mathbf{s}^{(i)}$, where $\odot$; is the element-wise multiplication. It can be shown that $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_l, \mathbf{d}_L]$ is drawn from

$$p(\mathbf{d}_l|-) \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{d}_l}, \boldsymbol{\Sigma}_{\mathbf{d}_l}), \tag{10}$$

where

$$\boldsymbol{\Sigma}_{\mathbf{d}_l} = \left(d\mathbf{I} + \gamma_\epsilon \sum_{i=1}^{T} (z_l^{(i)} s_l^{(i)})^2 (\mathbf{\Delta}^{(i)})^T \mathbf{\Delta}^{(i)} \right)^{-1} \tag{11}$$

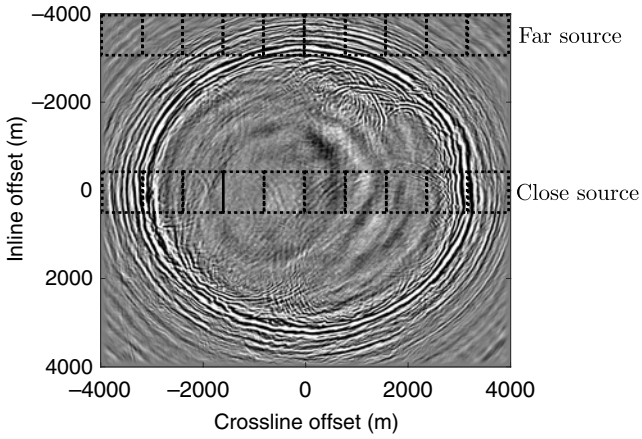

Figure 1. An example of a time slice extracted from the SEAM II data set. We partition time slices into sections of 128×128.

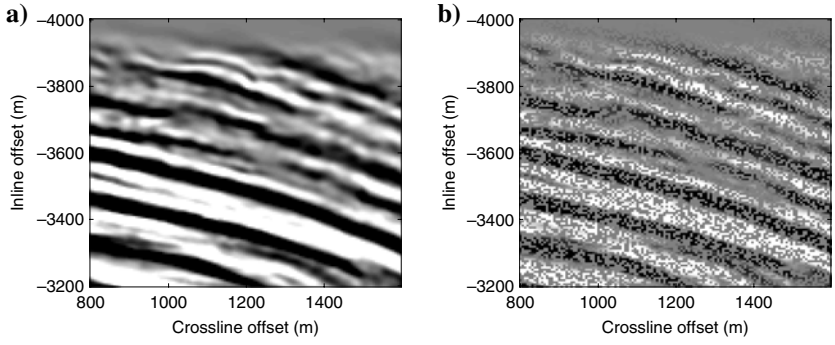

Figure 2. (a) Original signal used for evaluating uncertainty maps of dimension 128×128 receivers. This was extracted from the SEAM II data set. (b) The same signal using only 50% of receivers kept randomly.

and

$$\boldsymbol{\mu}_{\mathbf{d}_l} = \gamma_\epsilon \boldsymbol{\Sigma}_{\mathbf{d}_l} \sum_{i=1}^{T} z_l^{(i)} s_l^{(i)} \tilde{\mathbf{x}}_{-l}^{(i)} \tag{12}$$

with

$$\begin{aligned} \tilde{\mathbf{x}}_{-l}^{(i)} = {} & (\mathbf{\Delta}^{(i)})^T \mathbf{y}^{(i)} - (\mathbf{\Delta}^{(i)})^T \mathbf{\Delta}^{(i)} \mathbf{D} (\mathbf{s}^{(i)} \odot \mathbf{z}^{(i)}) \\ & + (\mathbf{\Delta}^{(i)})^T \mathbf{\Delta}^{(i)} \mathbf{d}_l (s_l^{(i)} z_l^{(i)}), \end{aligned} \tag{13}$$

and γ_ϵ being the noise precision. The posterior distribution of $z_l^{(i)}\ \forall\, i$ can be shown to be

$$p(z_l^{(i)}|-) \sim \text{Bernoulli}\left(\frac{p_1}{p_0 + p_1}\right), \tag{14}$$

where p_1 is the posterior probability that $z_l^{(i)} = 1$ and is given by

$$p_1 = \pi_l \exp\left\{ -\frac{\gamma_\epsilon}{2} ((s_l^{(i)})^2 \mathbf{d}_l^T (\mathbf{\Delta}^{(i)})^T \mathbf{\Delta}^{(i)} \mathbf{d}_l - 2 s_l^{(i)} \mathbf{d}_l^T \tilde{\mathbf{x}}_{-l}^{(i)}) \right\}. \tag{15}$$

On the other hand, the posterior probability that $z_l^{(i)} = 0$ is given by $p_0 = 1 - \pi_l$. Then, the posterior distribution of $s_l^{(i)}\ \forall\, i$ can be shown to be

$$p(s_l^{(i)}|-) \sim \mathcal{N}(\mu_{s_l^{(i)}}, \Sigma_{s_l^{(i)}}), \tag{16}$$

where

$$\boldsymbol{\Sigma}_{s_l^{(i)}} = \begin{cases} (\gamma_s + \gamma_\epsilon \mathbf{d}_l^T (\mathbf{\Delta}^{(i)})^T \mathbf{\Delta}^{(i)} \mathbf{d}_l)^{-1} & \text{if } z_l^{(i)} = 1 \\ \gamma_s^{-1} & \text{if } z_l^{(i)} = 0 \end{cases} \tag{17}$$

and

$$\mu_{s_l^{(i)}} = \begin{cases} \gamma_\epsilon \boldsymbol{\Sigma}_{s_l^{(i)}} \mathbf{d}_l^T (\mathbf{\Delta}^{(i)})^T \mathbf{\Delta}^{(i)} \tilde{\mathbf{x}}_{-l}^{(i)} & \text{if } z_l^{(i)} = 1 \\ 0 & \text{if } z_l^{(i)} = 0 \end{cases} \tag{18}$$

with γ_s the precision for the weights. By drawing from these distributions and using equation 8, one estimate of $\mathbf{x}^{(i)}$ is obtained. We will see in the next section how we can use the collection of the draws and consequently the estimations of $\mathbf{x}^{(i)}$ to create uncertainty maps. Further information regarding the prior distributions, implementation and parameter details can be found in Pilikos and Faul (2017) where this model was used for seismic CS and denoising.

Variance of BPFAs estimations

BPFA estimates the model's variables using expressions depending on other variables that are considered fixed for a given iteration. This is called a Gibbs iteration, and the entire procedure of obtaining an estimate for one variable (using the posterior conditional distribution)

given the others is called Gibbs sampling. We included only three in the previous section to facilitate the discussion. The complete set of expressions can be found in the Appendix of Zhou et al. (2012).

A feature of the algorithm is that it splits the signal into smaller patches and uses them in a sequential manner. To have enough training data to learn the dictionary of basis, overlapping patches are used. We chose a patch size of 8×8 in all our BPFA experiments in sections of time slices of 128×128 as discussed in the next section. In the first iteration, the algorithm extracts all 256 patches (there are $128/8 = 16$ patches along the horizontal axis and $128/8 = 16$ patches along the vertical axis). This is essentially what an algorithm would extract without overlaps. Then, in the second iteration, the algorithm shifts the starting point of extraction by one receiver down. This results in the extraction of 240 patches (there are again 16 patches along the horizontal axis and $16 - 1 = 15$ patches along the vertical axis because the last one is not a complete 8×8 patch). This procedure continues for all 64 receivers in a given patch, resulting in 64 such sets of patches in which each set contains hundreds of patches.

For the first Gibbs iteration of the algorithm, the inference starts using the first set of patches. Then, in the second iteration, the second set is extracted and then first and second sets are used for estimation. After all 64 iterations, all sets are used at the same time to infer the variables of interest using $(128 - 7)^2 = 14,641$ patches in total. In the final iteration, each model variable is drawn from its corresponding posterior distribution, and it is used to calculate the receivers' value for all patches using equation 8. Therefore, each receiver's value is inferred various times because it is contained in various patches (at most 64 in our example). To obtain the final values, the mean of each receiver's value is obtained by averaging over all its estimated values. In the same manner, the uncertainty of the prediction at that receiver location is obtained by calculating the variance of all its estimated values. Zhou et al. (2009) provide further information regarding the patch processing procedure.

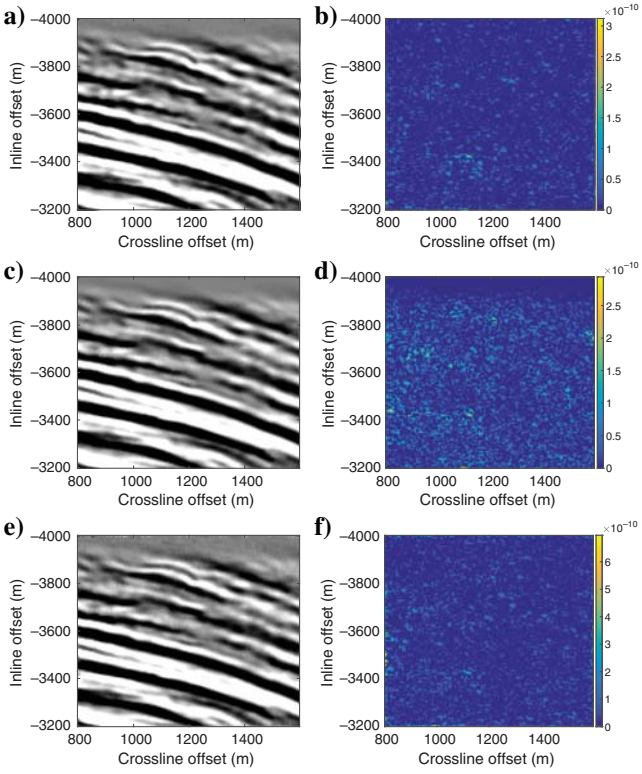

Figure 3. (a) The RVM and DCT with $Q = 36.7270$ db, (b) the respective reconstruction error, (c) the BPFA reconstruction with $Q = 28.9016$ db, (d) the BPFA error, (e) the POCS reconstruction with $Q = 24.0410$ db, and (f) its respective error.

UNCERTAINTY MAPS FOR SYNTHETIC DATA

To test the uncertainty maps, we used a 3D synthetic data set called SEG Advanced Modeling II (SEAM II) (SEG, 2018a). It contains an artificial source (shot) and a grid of receivers. We will work in the common-shot domain in which a gather contains the output of all receivers obtained from the same source. The data set has a 6.25 m spatial sampling. There are 1281 receivers along each line, and there are 1281 lines spanning 8000 m covering vertical and horizontal directions. The temporal sampling is 6 ms, and each trace has 500 samples resulting in 3 s of recordings.

We will focus on the reconstruction of sections of time slices similar to Pilikos and Faul (2016, 2017) and Pilikos et al. (2017). We extract 128×128 sections for receiver lines that pass far from and close to the source as shown in Figure 1. To interpolate a shot record (x-t domain), we process time slices independently and store all the data into a 3D cube. Then, we extract the receiver line of interest by using all time samples and all receivers at that location. Pilikos et al. (2017) follow this procedure and illustrate that it is possible to reconstruct under-sampled data without

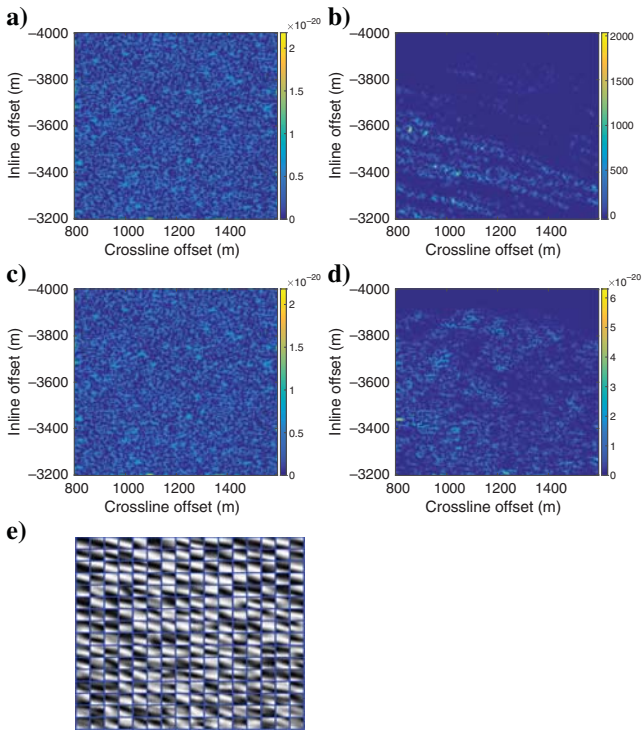

Figure 4. (a) The uncertainty map using the RVMs original predictive variance and DCT, (b) the ExpCML, (c) the uncertainty map with augmentation, (d) the BPFA uncertainty map, and (e) the respective learned dictionary of basis by BPFA. All figures are produced while reconstructing the signals of Figure 3.

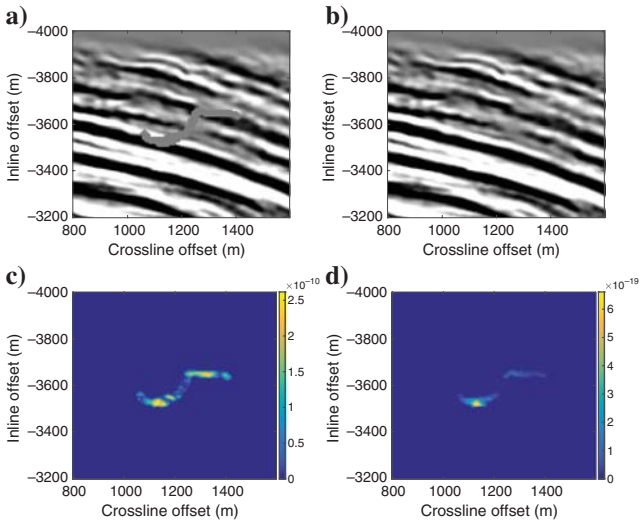

Figure 5. (a) An artificial surface obstruction resulting in 390 receivers missing, (b) the BPFA reconstruction with $Q = 28.6081$db, (c) the reconstruction error, and (d) the uncertainty map.

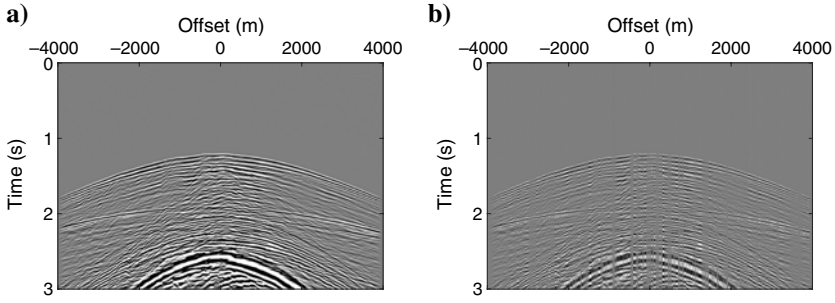

Figure 6. (a) Original signal and (b) using only 50% of the receivers randomly.

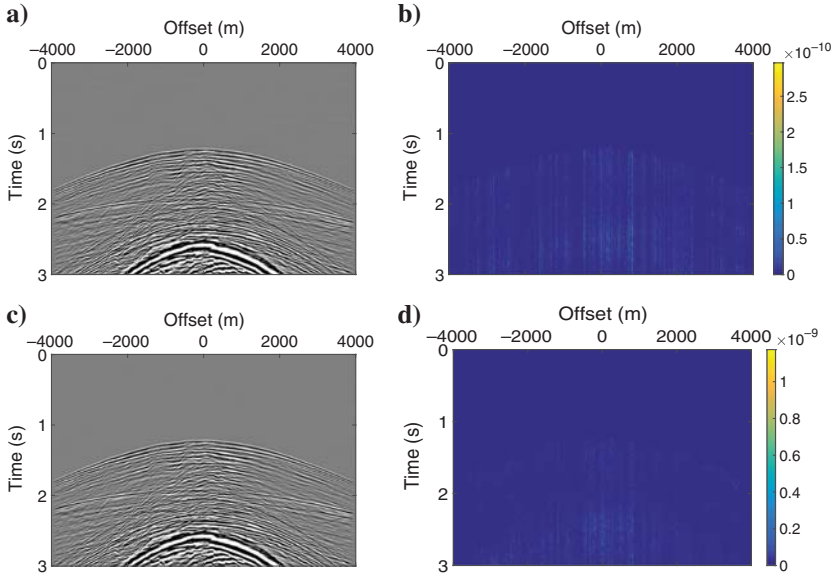

Figure 7. (a) The RVMs recovery, (b) error, (c) BPFAs recovery, and (d) error.

any aliasing. For the RVM results, we used the package from Tipping (2018), and for the BPFA results, we used the package from Zhou (2018). All experiments were performed as single-core jobs on machines with Intel(R) Xeon(R) CPU E5-2650 with 2.00 GHz.

Uncertainty visualization using different methods

One way to illustrate the effectiveness of the methods is to visualize the uncertainty maps produced along with the respective reconstruction error. Figure 2a shows the original section of a time slice. Figure 2b shows the signal using 50% of receivers. The receivers were muted randomly by going through the signal and drawing a random number between 1 and 100. If that number was less than 50, it was kept otherwise not considered. Note that random subsampling reduces aliasing but introduces incoherent noise (Kumar et al., 2015). This type of random sampling is not realistic in seismic surveys, nevertheless, it is used in this work to illustrate uncertainty.

As discussed, the RVM uses a predefined dictionary of basis functions, therefore, we chose the DCT that was used successfully by Pilikos and Faul (2016). Figure 3a shows the RVM reconstruction with the reconstruction quality Q defined by,

$$Q = 10 \log \frac{\|\mathbf{x}\|_2^2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2}, \qquad (19)$$

where $\mathbf{x}$ is the original signal and $\hat{\mathbf{x}}$ is the reconstruction. Figure 3b shows the reconstruction error. Figure 3c illustrates the reconstruction using the BPFA with lower Q, and Figure 3d shows its respective reconstruction error. For completeness, we include a reconstruction using the POCS algorithm (Abma and Kabir, 2006) mentioned earlier. Figure 3e shows the reconstruction using POCS with reconstruction quality lower than the RVM and the BPFA methods. The respective reconstruction error is included in Figure 3f. We can see that both Bayesian algorithms perform better than POCS. Nevertheless, this is not the purpose of the paper. Pilikos and Faul (2016, 2017) provide a comprehensive comparison of reconstruction accuracy of these algorithms. We will continue the discussion for uncertainty quantification.

Figure 4a shows the predictive variance produced by the RVM, which is spread around the signal without any particular pattern. Figure 4b shows the expected change in the model shifted so that the higher the values the more uncertain the model. This is achieved by negating all the values and then adding the minimum value to each location. This shifts the range from zero to positive values. As it can be seen, it captures the variance of the signal with resemblance to the original. Nevertheless, it does not show any similarities with the reconstruction error produced by the RVM. Figure 4c shows the predictive variance with augmentation using Gaussian basis functions at test points. The uncertainty map is similar to the original predictive variance.

On the other hand, the uncertainty map produced by BPFA can be seen in Figure 4d. The learned dictionary of basis is illustrated in Figure 4e, which captures the direction of the largest variations in the signal. The uncertainty map produced by BPFA shows good correlation with the reconstruction error. It looks more informative than the expected change in Figure 4b, but it is not easy to compare with the rest of the methods. Thus, a quantitative comparison is required and will be given later in this paper. Note that we will not follow the comparison with the RVMs augmentation and the ExpCML because, for a fair comparison, they require extensive parameter tuning and is not the purpose of the paper.

Before providing a quantitative comparison of uncertainty maps, we illustrate another example of reconstruction using the BPFA. In this case, we remove 390 receivers and create a gap similar to that caused by a surface obstruction at the time of acquisition. In this case, it could be a river where receivers cannot be placed. Figure 5a shows the signal with the gap, and Figure 5b shows the BPFA reconstruction. The reconstruction error obtained by the BPFA is shown in Figure 5c, and the uncertainty map is shown in Figure 5d. We can see that the reconstruction is of high quality at $Q = 28.6081$ db albeit, with some error. Nevertheless, this error is captured by the uncertainty map, illustrating the regions where the BPFA is uncertain matching with the location of error.

Sorting the uncertainty maps in the shot record domain

Following the time slice processing approach that was used by Pilikos et al. (2017), we process one section of a time slice at a time, and later combine all reconstructed sections into a shot record. This way, it is possible to visualize the uncertainty over the entire shot record and also visualize any aliasing in the f-k domain. Entire receivers are missing, as shown in Figure 6b, which means that for each time slice, the same samples are missing. To obtain an uncertainty map using time slice processing, we randomly removed receivers from 5000 sections of time slices (500 time samples for 10 sections) of 128×128 dimensions (last section is 128×129) and then reconstructed them. This was done for the RVM and the BPFA for different percentages of receivers.

Figure 6a shows the original signal, and Figure 6b shows the signal with only 50% of the receivers used. Figure 7a shows the RVMs reconstruction and the respective reconstruction error can be seen in Figure 7b. The same signals for BPFA can be seen in Figure 7c and 7d, respectively. We continue the discussion on the uncertainty, and we plot the uncertainty maps in Figure 8 for the signals in Figure 7. Figure 8a shows the variance of the BPFA normalized to one. That is, within each section of a time slice, we divide each receiver with the maximum value in that section. Different sections have varying levels of variance with a wide range of values. If we normalized over the entire shot record, many details would be lost. Figure 8b shows the respective uncertainty map produced by normalizing the RVMs predictive variance using the same procedure. The RVMs predictive variance is equal to one when the receivers have not yet recorded reflections (i.e., the variance is one prior to the first arrival). This is due to the fact that the variance in this region is very low, and the RVMs

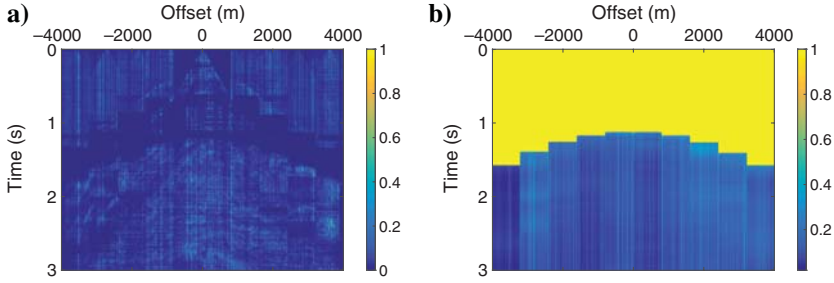

Figure 8. (a) The BPFAs uncertainty map and (b) RVMs uncertainty map.

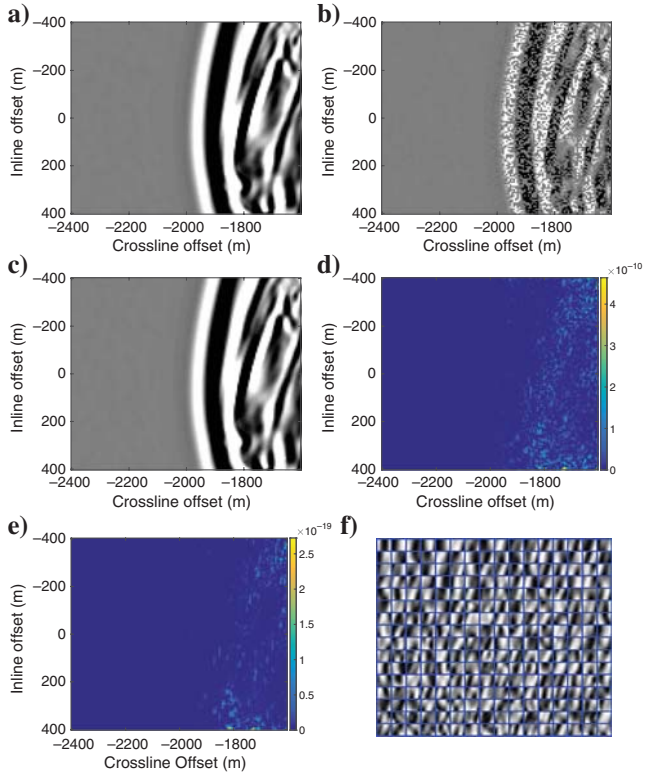

Figure 9. (a) Original, (b) 50% of receivers, (c) the BPFAs reconstruction with $Q = 33.8258$ db, (d) the BPFAs reconstruction error, (e) the BPFAs variance, and (f) the learned dictionary.

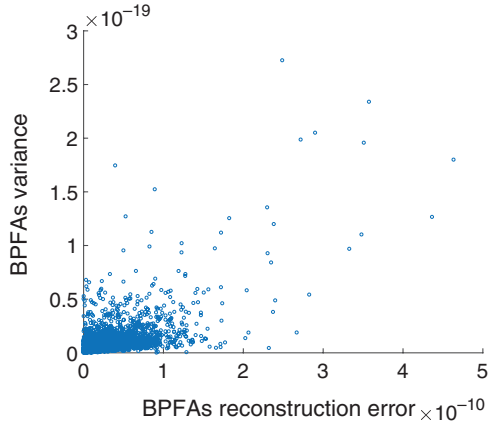

Figure 10. Direct scatter plot between the BPFAs variance and the respective reconstruction error.

reconstruction is the same everywhere using one large basis function for reconstruction. This in turn produces the worst possible uncertainty of the same value. We will quantify the overall performance in the next section.

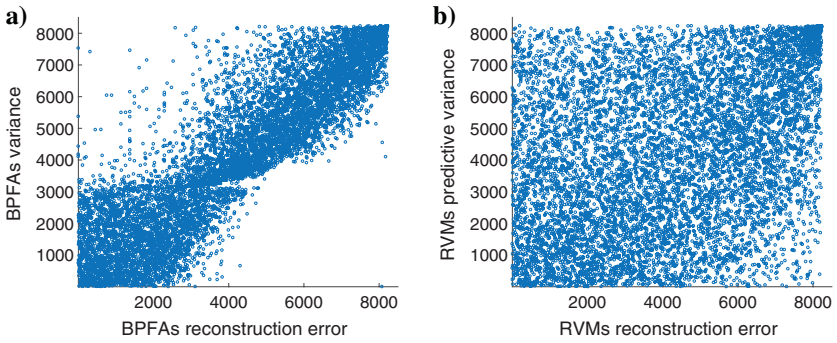

Figure 11. Ranked scatter plots for (a) BPFAs variance (0.7064) and (b) RVMs predictive variance (0.3514) against their respective reconstruction errors. In the previous sentence, the values in parenthesis are the Spearman's correlation coefficients for each relationship as defined in equation 20.

Table 1. Mean uncertainty quantification for 2000 sections (1–200 time samples) of far source signals.

Spearman's correlation coefficient between variance and reconstruction error			
Percentage used	30%	50%	70%
RVMs predictive variance	0.0027	0.0021	0.0012
BPFAs variance	0.4092	0.4056	0.3724

Table 2. Mean uncertainty quantification for 3000 sections (201–500 time samples) of far source signals.

Spearman's correlation coefficient between variance and reconstruction error			
Percentage used	30%	50%	70%
RVMs predictive variance	0.2779	0.2962	0.1967
BPFAs variance	0.5254	0.5221	0.5337

Table 3. Mean uncertainty quantification for 1000 sections (1–100 time samples) of close to source signals.

Spearman's correlation coefficient between variance and reconstruction error			
Percentage used	30%	50%	70%
RVMs predictive variance	0.0718	0.1477	0.1174
BPFAs variance	0.5804	0.5169	0.5083

Uncertainty quantification using the Spearman's correlation coefficient

From the visualizations in the previous section, it can be seen that the BPFAs variance creates informative uncertainty maps. To evaluate this over various signals and scenarios, we need a quantitative metric. Then, we can compare it against the RVM and also get a better understanding of the methods. To do this, we propose to check how much the reconstruction error correlates with the uncertainty map in a corresponding receiver location. That is, ideally, the larger the reconstruction error in a data point the larger the uncertainty and vice versa.

Scatter plots

Consider the signals in Figure 9 with a direct scatter plot between the BPFAs variance and the respective reconstruction error in Figure 10. We can see that many data points are concentrated near the origin, and it is not clear how the two variables are correlated. Therefore, we transform them to their ranked version, which means that we sort the data points by their magnitude, with data points having the same magnitude being assigned the same rank. The ranked version is now plotted for all variables of interest. Figure 11a shows the same data points as in Figure 10 but now ranked according to their value. This allows the data points to spread out instead of being near the origin. Comparing the scatter plots in Figure 11, we can see that the BPFAs variance correlates much better with the reconstruction error.

Analysis using the Spearman's correlation coefficient

To evaluate the performance over numerous sections of time slices, we will use the Spearman's correlation coefficient. This coefficient evaluates how two variables are related to each other, i.e., as one grows what happens to the other. The Spearman's correlation coefficient is defined as

$$\text{Spearman's correlation coefficient} = \frac{\sum_{i=1}^{n}(x_i-\mu_x)(y_i-\mu_y)}{\sqrt{\sum_{i=1}^{n}(x_i-\mu_x)^2}\sqrt{\sum_{i=1}^{n}(y_i-\mu_y)^2}}, \quad (20)$$

where n is the number of receiver locations used in the calculation (at predicted locations), $\{x_i\}_{i=1}^{n}$ and $\{y_i\}_{i=1}^{n}$ are the variables of interest, and μ_x and μ_y are their respective means. Note that each variable is ranked as we have seen in Figure 11. The value ranges from -1 to $+1$. Positive correlation means that as one grows so does the other. On the other hand, if the other decreases in value, it has negative correlation. In our case, we would like to check the correlation between the variables that we use to create uncertainty maps (BPFAs variance and RVMs predictive variance) and their respective reconstruction errors.

We will use 5000 sections of time slices far from and near the source as illustrated in Figure 1 for three percentages (30%, 50%, and 70%) of receivers used. We split the evaluation for far from

source signals and close to source signals and also for different time samples.

Table 1 shows the mean Spearman's correlation coefficient, calculated for both uncertainty methods along with three different percentages for the first 2000 sections of time slices. For these results, we used the sections of time slices when t, the index of the time sample, is less than or equal to 200. There are many sections that include almost zero signal, where the receivers have not yet recorded the signal. It is clear that the BPFAs variance is positively correlated with the reconstruction error with a higher correlation coefficient than the RVM. In addition, it is only slightly affected by the percentage of the available receivers used provided that there are enough training data to learn a dictionary of basis. On the other hand, the RVMs predictive variance correlation is very close to zero. Table 2 shows the average Spearman's correlation coefficient for the last 3000 sections of time slices when $t > 201$. In this region, there are sections with larger variance. The correlation is higher for both algorithms in general due to the presence of signals with higher variance. The BPFAs variance is more consistent and not significantly affected by the percentage as opposed to the RVMs predictive variance.

Moving on to closer to source sections of time slices, Table 3 shows the average Spearman's correlation coefficient for 1000 sections when $t < 101$. This region is characterized with very high and low variance. We can see that the BPFAs variance is still more correlated than the RVMs predictive variance. Table 4 shows the average correlation for the remaining 4000 sections when $t > 100$. The BPFAs variance is again better with the RVMs predictive variance obtaining improved results. From the results of quantifying uncertainty for far from the source and closer to the source, we can see that the BPFAs performance is better in all cases. However, its performance varies depending on the region of the signal that it operates.

STACKING OF UNCERTAINTY MAPS

From the previous section, we have seen that the correlation of the uncertainty maps with the respective reconstruction errors vary at different instances. Viewing the uncertainty map at one time sample is useful. Nevertheless, it does not provide the complete uncertainty because each receiver has 500 time steps of varying correlation levels. To get a complete understanding of uncertainty for a receiver, it is useful to take into account all time samples associated with it. One option would be to sort the uncertainty into the x-t domain as it was done in Figure 8. However, this does not provide a quantitative metric for the complete uncertainty. Therefore, we decided to stack all 500 uncertainty maps for each receiver together and take the average value per receiver location.

Figure 12 shows various stacked uncertainty maps for a section of receivers from a time slice in the same location as the section in Figure 9 but using all 500 time samples. We can see the BPFAs stacked variance in Figure 12a and the respective stacked reconstruction error in Figure 12b. The uncertainty map picks up regions of large reconstruction error with a Spearman's correlation coefficient equal to 0.8306. This is higher than the individual uncertainty maps because the averaging of uncertainties helps.

The RVMs stacked predictive variance is included in Figure 12c with the respective stacked reconstruction error in Figure 12d. We can see that the RVMs uncertainty map is also significantly improved providing higher correlation with the error and a Spearman's correlation coefficient equal to 0.7526. The averaging helps to amplify uncertainties that are correctly calculated and diminishes the randomness in uncertainties.

To get a better understanding of the performance of the BPFAs stacked variance and the RVMs stacked predictive variance, we repeat the experiment for all 20 sections of time slices (10 far from source and 10 close to source) averaging their 500 uncertainty maps. Table 5 shows the average Spearman's correlation coefficient for these 20 stacked sections. The BPFAs variance provides very high correlation with the stacked reconstruction error showing

Table 4. Mean uncertainty quantification for 4000 sections (101–500 time samples) of close to source signals.

Spearman's correlation coefficient between variance and reconstruction error			
Percentage used	30%	50%	70%
RVMs predictive variance	0.2904	0.3003	0.2275
BPFAs variance	0.3804	0.3632	0.3685

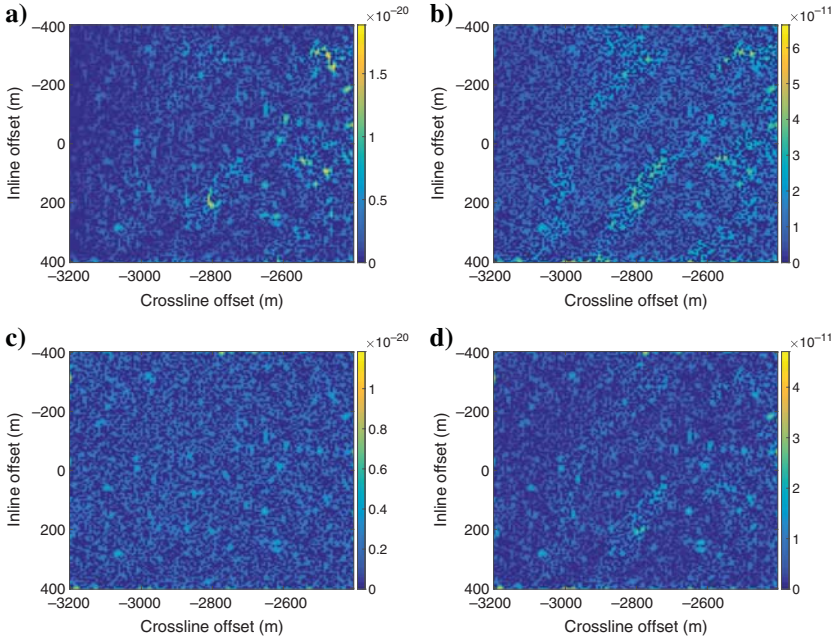

Figure 12. Average uncertainty for different methods using all 500 uncertainty maps produced per time samples for each receiver. (a) BPFAs stacked variance (0.8306), (b) BPFAs stacked reconstruction error, (c) RVMs stacked predictive variance (0.7526), and (d) the RVMs stacked reconstruction error. In parenthesis, the Spearman's correlation coefficient for each relationship as defined in equation 20.

the improvements obtaining with averaging. The same is true for the RVMs predictive variance that also provides very high correlation. Overall, the BPFAs variance still produces better uncertainty maps compared with the RVMs predictive variance.

Table 5. Mean uncertainty quantification of 20 sections stacked with 500 uncertainty maps (1–500 time samples) per percentage.

Spearman's correlation coefficient between variance and reconstruction error			
Percentage used	30%	50%	70%
RVMs predictive variance	0.8503	0.7164	0.6185
BPFAs variance	0.8533	0.9082	0.9168

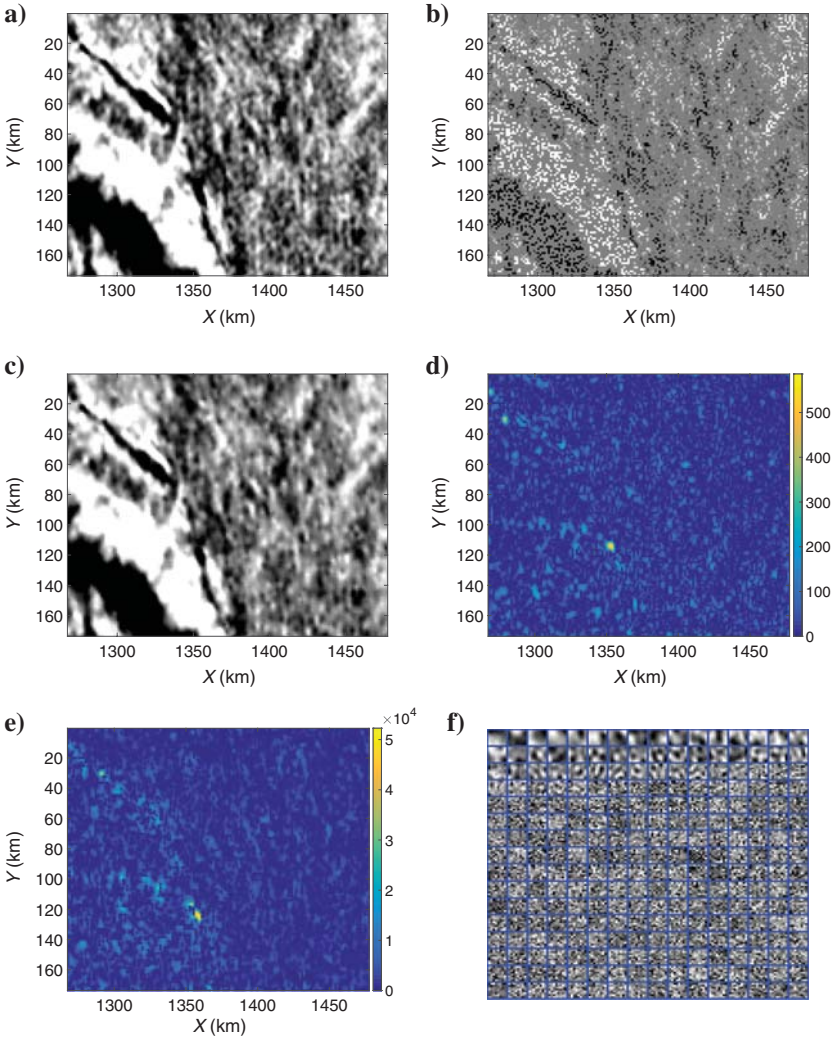

Figure 13. (a) Original, (b) using 50% of receivers, (c) the BPFAs reconstruction, (d) the BPFAs reconstruction error, (e) the BPFAs variance with Spearman's correlation coefficient equal to 0.8114, and (f) the learned basis for a section from the Parihaka data set.

UNCERTAINTY MAPS FOR FIELD DATA

We will now illustrate the BPFAs performance on field data. We will use the Parihaka data set that is a 3D seismic survey provided by New Zealand Petroleum and Minerals (NZPM) and obtained from the SEG wiki website (SEG, 2018b). Further information about this data set can be found in Cohen et al. (2006). A time slice is processed, using only 50% of receivers. Figure 13a shows an original section from the Parihaka data set. Figure 13b shows the same signal with 50% of the receivers used, and Figure 13c shows the BPFAs reconstruction. Figure 13d shows the reconstruction error, and Figure 13e shows the variance of BPFAs samples. The latter is highly correlated with the reconstruction error with a Spearman's correlation coefficient equal to 0.8114. Figure 13f shows the learned basis from the available data. We can see that the BPFA learns a dictionary of basis that captures the important features in the data, reconstructs the signal well, and the uncertainty map is highly correlated with the reconstruction error. This illustrates its effectiveness to process complex signals found in field data.

CONCLUSION

The challenge to associate each estimation with a degree of certainty is becoming more and more relevant with the growth of seismic CS. Ideally, the uncertainty map produced by an algorithm should correlate well with the reconstruction error when we evaluate them on a known signal. We used BPFA to obtain reconstructions and at the same time create uncertainty maps. This was achieved by exploiting its probabilistic nature during inference, which involves random draws. By drawing different variables and consequently predicting different receivers' values, it is possible to obtain a collection of predictions and then estimate a mean and a variance for each receiver. When the variance is small, BPFA trusts the value better and vice versa.

We compared BPFAs uncertainty maps on thousands of sections of time slices against other algorithms in the literature. The comparison was done using the Spearman's correlation coefficient that evaluates the correlation between the uncertainty map and the respective reconstruction error. We showed that BPFA produces uncertainty maps that are well-correlated with the error, better than other methods. Scatter plots also illustrate this along with visualizations of various time slices. We used all time slices and visualized the uncertainty maps for shot records and their reconstructions. Comparison with a more traditional algorithm, namely POCS, was also provided showing that the BPFA obtains higher reconstruction accuracy.

We used signals with different structures during evaluations, and we investigated how the BPFAs performance changes with varying levels of variance. The level of uncertainty quantification was found not to be heavily affected by the variance. Nevertheless, when the variance is very

small or close to zero, the uncertainty also behaves badly because BPFA is not able to learn a dictionary of basis. Therefore, care is required when splitting the signal into sections so as to ensure that there is enough variance. The acceptable level of signal's variance is subject for future research.

From this research, Bayesian estimation and analysis can help in the construction of accurate uncertainty maps for seismic data acquisition systems. BPFA is a great example of how this can be achieved without compromising the reconstruction accuracy. This could be useful for seismic survey design where it is often desirable to save costs by making source/receiver spacings larger. If we can model the degree of uncertainty that this will cost, it could have a huge impact in making the right compromise for a certain survey.

ACKNOWLEDGMENTS

We would like to thank BP for the support and for allowing the publication of this research. Furthermore, we would like to thank R. Abma and N. Philip from BP for very useful discussions on seismic data acquisition. Also, we would like to thank N. Nikiforakis for technical input. We thank SEAM-II for the model and C. Regone of BP for modeling the seismic data as well as thank NZPM for providing the Parihaka field data set. We acknowledge funding from the Engineering and Physical Sciences Research Council (EPSRC) through an Industrial CASE studentship (1502944).

DATA AND MATERIALS AVAILABILITY

Please contact the first author for details.

REFERENCES

Abma, R., and N. Kabir, 2006, 3D interpolation of irregular data with a POCS algorithm: Geophysics, **71**, no. 6, E91–E97, doi: 10.1190/1.2356088.

Beckouche, S., and J. Ma, 2014, Simultaneous dictionary learning and denoising for seismic data: Geophysics, **79**, no. 3, A27–A31, doi: 10.1190/geo2013-0382.1.

Cohen, C. R., L. J. Christianson, C. Bates, R. P. Laney, and G. A. Morton, 2006, Pogo New Zealand's 3D seismic: New standards and structural/stratigraphic insights in the Taranaki Basin: Proceedings of the New Zealand Crown Minerals.

Duijndam, A. J. W., 1988a, Bayesian estimation in seismic inversion. Part I: Principles: Geophysical Prospecting, **36**, 878–898, doi: 10.1111/j.1365-2478.1988.tb02198.x.

Duijndam, A. J. W., 1988b, Bayesian estimation in seismic inversion. Part II: Uncertainty analysis: Geophysical Prospecting, **36**, 899–918, doi: 10.1111/j.1365-2478.1988.tb02199.x.

Faul, A. C., and G. Pilikos, 2016, The model is simple, until proven otherwise: How to cope in an ever-changing world: Data for Policy, Frontiers of Data Science for Government: Ideas, Practices and Projections.

Fjeldstad, T., and D. Grana, 2018, Joint probabilistic petrophysics-seismic inversion based on Gaussian mixture and Markov chain prior models: Geophysics, **83**, no. 1, R31–R42, doi: 10.1190/geo2017-0239.1.

Herrmann, F. J., and G. Hennenfent, 2008, Non-parametric seismic data recovery with curvelet frames: Geophysical Journal International, **173**, 233–248, doi: 10.1111/j.1365-246X.2007.03698.x.

Kreimer, N., and M. D. Sacchi, 2011, A tensor higher order singular value decomposition (HOSVD) for prestack simultaneous noise reduction and interpolation: 81st Annual International Meeting, SEG, Expanded Abstracts, 3069–3074, doi: 10.1190/1.3627833.

Kreimer, N., and M. D. Sacchi, 2012, A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation: Geophysics, **77**, no. 3, V113–V122, doi: 10.1190/geo2011-0399.1.

Kumar, R., C. D. Silva, O. Akalin, A. Y. Aravkin, H. Mansour, B. Recht, and F. J. Herrmann, 2015, Efficient matrix completion for seismic data reconstruction: Geophysics, **80**, no. 5, V97–V114, doi: 10.1190/geo2014-0369.1.

Kutscha, H., and D. Verschuur, 2016, The utilization of the double focal transformation for sparse data representation and data reconstruction: Geophysical Prospecting, **64**, 1498–1515, doi: 10.1111/1365-2478.12362.

Malinverno, A., and V. A. Briggs, 2004, Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes: Geophysics, **69**, 1005–1016, doi: 10.1190/1.1778243.

Naghizadeh, M., and M. D. Sacchi, 2007, Multistep autoregressive reconstruction of seismic records: Geophysics, **72**, no. 6, V111–V118, doi: 10.1190/1.2771685.

Paisley, J., and L. Carin, 2009, Nonparametric factor analysis with beta process priors: Proceedings of the 26th Annual International Conference on Machine Learning, 777–784.

Pilikos, G., and A. C. Faul, 2016, Relevance vector machines with uncertainty measure for seismic Bayesian compressive sensing and survey design: Proceedings of the 15th IEEE International Conference on Machine Learning and Applications, 925–930, doi: 10.1109/ICMLA.2016.0166.

Pilikos, G., and A. C. Faul, 2017, Bayesian feature learning for seismic compressive sensing and denoising: Geophysics, **82**, no. 6, O91–O104, doi: 10.1190/geo2016-0373.1.

Pilikos, G., A. C. Faul, and N. Philip, 2017, Seismic compressive sensing beyond aliasing using Bayesian feature learning: 87th Annual International Meeting, SEG, Expanded Abstracts, 4328–4332, doi: 10.1190/segam2017-17558742.1.

Porsani, M. J., 1999, Seismic trace interpolation using halfstep prediction filters: Geophysics, **64**, 1461–1467, doi: 10.1190/1.1444650.

Rasmussen, C. E., and J. Quiñonero Candela, 2005, Healing the relevance vector machine through augmentation: Proceedings of the 22nd International Conference on Machine Learning, 689–696.

Sacchi, M., T. Ulrych, and C. Walker, 1998, Interpolation and extrapolation using a high-resolution discrete Fourier transform: IEEE Transactions on Signal Processing, **46**, 31–38, doi: 10.1109/78.651165.

SEG, 2018a, Parihaka data set, http://wiki.seg.org/wiki/parihaka-3D, accessed 2 February 2018.

SEG, 2018b, SEAM data set, http://seg.org/news-resources/research-and-data/seam, accessed 8 January 2018.

Spitz, S., 1991, Seismic trace interpolation in the f-x domain: Geophysics, **56**, 785–794, doi: 10.1190/1.1443096.

Stanton, A., M. D. Sacchi, R. Abma, and J. A. Stein, 2015, Mitigating artifacts in projection onto convex sets interpolation: 85th Annual International Meeting, SEG, Expanded Abstracts, 3779–3783, doi: 10.1190/segam2015-5754691.1.

Tipping, M. E., 2001, Sparse Bayesian learning and the relevance vector machine: Journal of Machine Learning Research, **1**, 211–244.

Tipping, M. E., 2018, RVM software, http://www.miketipping.com/downloads.html, accessed 8 January 2018.

Tipping, M. E., and A. Faul, 2003, Fast marginal likelihood maximisation for sparse Bayesian models: Proceedings of the 9th International Workshop on Artificial Intelligence and Statistics, 3–6.

Trad, D. O., T. J. Ulrych, and M. D. Sacchi, 2002, Accurate interpolation with high-resolution time-variant Radon transforms: Geophysics, **67**, 644–656, doi: 10.1190/1.1468626.

Turquais, P., E. G. Asgedom, W. Sollner, and E. Otnes, 2015, Dictionary learning for signal-to-noise ratio enhancement: 85th Annual International Meeting, SEG, Expanded Abstracts, 4698–4702, doi: 10.1190/segam2015-5846080.1.

Ulrych, T. J., M. D. Sacchi, and A. Woodbury, 2001, A Bayes tour of inversion: A tutorial: Geophysics, **66**, 55–69, doi: 10.1190/1.1444923.

van den Berg, E., and M. P. Friedlander, 2009, Probing the pareto frontier for basis pursuit solutions: SIAM Journal on Scientific Computing, **31**, 890–912, doi: 10.1137/080714488.

Wang, D., R. Saab, O. Yilmaz, and F. J. Herrmann, 2008, Bayesian wavefield separation by transform-domain sparsity promotion: Geophysics, **73**, no. 5, A33–A38, doi: 10.1190/1.2952571.

Zhou, M., 2018, BPFA software, http://mingyuanzhou.github.io/Code.html, accessed 8 January 2018.

Zhou, M., H. Chen, J. Paisley, L. Ren, L. Li, Z. Xing, D. Dunson, G. Sapiro, and L. Carin, 2012, Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images: IEEE Transactions on Image Processing, **21**, 130–144, doi: 10.1109/TIP.2011.2160072.

Zhou, M., H. Chen, L. Ren, G. Sapiro, L. Carin, and J. W. Paisley, 2009, Nonparametric Bayesian dictionary learning for sparse image representations, Advances in neural information processing systems 22: Cusson Associates, 2295–2303.

Zhu, L., E. Liu, and J. H. McClellan, 2015, Seismic data denoising through multiscale and sparsity-promoting dictionary learning: Geophysics, **80**, no. 6, WD45–WD57, doi: 10.1190/geo2015-0047.1.

Zwartjes, P. M., and M. D. Sacchi, 2007, Fourier reconstruction of nonuniformly sampled, aliased seismic data: Geophysics, **72**, no. 1, V21–V32, doi: 10.1190/1.2399442.