Relevance Vector Machines with Uncertainty Measure for Seismic Bayesian Compressive Sensing and Survey Design

Georgios Pilikos
Maxwell Centre
Department of Physics
University of Cambridge
Cambridge, United Kingdom
ggp29@cam.ac.uk

A.C. Faul
Maxwell Centre
Department of Physics
University of Cambridge
Cambridge, United Kingdom
acf22@cam.ac.uk

***Abstract*—Seismic data acquisition in remote locations involves sampling using regular grids of receivers in a field. Extracting the maximum possible information from fewer measurements is cost-effective and often necessary due to malfunctions or terrain limitations. Compressive Sensing (CS) is an emerging framework that allows reconstruction of sparse signals from fewer measurements than conventional sampling rates. In seismic CS, the utilization of sparse solvers has proven to be successful, however, algorithms lack predictive uncertainties. We apply the Relevance Vector Machine (RVM) to seismic CS and propose a novel utilization of multi-scale dictionaries of basis functions that capture different variations in the data. Furthermore, we propose the use of a new predictive uncertainty measure using the information from the neighbours of each estimation to produce accurate uncertainty maps. We apply the RVM to different seismic signals and obtain state-of-the-art reconstruction accuracy. Using the RVM and its predictive uncertainty map, it is possible to quantify risk associated with seismic data acquisition and at the same time guide future survey design.**

***Keywords*-Compressive Sensing; Bayesian Inference; Seismic Data Acquisition;**

I. Introduction

Seismic surveys are used in order to visualize the interior structure of our planet. The process uses an artificial source (shot) of body waves at the surface and reflections from deep impedance changes at rock layer boundaries are recorded by a grid of surface receivers. Efficient data acquisition would utilize as few shots and/or receivers as possible commensurate with adequate signal quality. Frequently, surface infrastructure can lead to gaps in the placement of shots and/or receivers leading to undesirable gaps in subsurface coverage. CS allows perfect reconstruction from fewer measurements than those conventionally used by the Shannon/Nyquist rate. The assumption for CS to work is that the signal has to be sparse either in the acquisition domain or sparse in a specified transform. Modelling a signal as a linear combination of basis functions [1] is a popular choice. Examples of dictionaries in the seismic CS literature include the Fourier Transform [2], the Radon Transform [3] or the Curvelet Transform [4] used in conjunction with a sparse solver. Projection Onto Convex Sets (POCS) [5] utilizes the Fourier Transform to sparsify the available data and uses hard or soft thresholding [6] to choose the essential basis functions. Iterative-Reweighted Least Squares was proposed [7] to suppress the artifacts in the Fourier domain. The Iterative Soft Thresholding (IST) and the Curvelet Transform were used successfully for seismic CS [4]. A faster version of IST was proposed [8], namely the Fast Iterative Soft Thresholding Algorithm (FISTA), and then applied to seismic data [9]. Spectral Projected Gradient for L1 (SPGL1) [10] was proposed with great reconstruction accuracy and is still considered state-of-the-art in seismic CS [11]. Nevertheless, the speed of convergence is important and thus a faster gradient projection method based on curvelets was proposed recently [11] with comparative reconstruction accuracy but faster computational time. Tensor Completion algorithms [12] were also proposed to solve this issue and to scale to larger dimensions. A recent comparison of 5D solvers has shown that POCS preserves better the amplitudes in the signals compared to other algorithms in the literature [13].

A solver in this context should be able to use arbitrary dictionaries of basis functions, either analytic or learned from seismic data [14]. The utilization of basis functions should ideally be multi-scale [14] in order to capture different variations in the data and it should provide a predictive uncertainty in its estimations. With this, it would be possible to trust areas or even re-shoot in uncertain regions if necessary. In this paper, we apply the RVM [15] to seismic CS. We extend its capabilities by exploiting the probabilistic nature of the algorithm through the use of a recent predictive uncertainty measure. We create a network of RVMs that refine their estimations and at each stage different dictionaries of basis functions could be used. We apply the RVM on various seismic signals and compare it with POCS and SPGL1. We show that the RVM outperforms POCS in accuracy and the RVM with appropriate basis functions matches the performance of the SPGL1 providing state-of-the-art reconstruction. We also provide an exhaustive analysis of the trade-off between accuracy and

computational time and show how the proposed predictive uncertainty measure can be used for seismic survey design.

The paper is organized as follows. Firstly in section II, an overview of the CS theory is given in a Bayesian framework. Then, in section III, the predictive uncertainty measure is described. In section IV, the typical seismic survey setup is discussed along with experiments on different seismic signals. A trade-off analysis of computational time and reconstruction accuracy is provided, comparisons with POCS and SPGL1 are shown and illustrations of uncertainty maps are included. Conclusions are made in section V.

II. Bayesian CS Theory

CS provides perfect reconstruction of sparse signals using unconventional sampling schemes under certain assumptions. Let $\boldsymbol{w} \in \mathbb{R}^N$ be the signal in the sparse domain. The matrix $\boldsymbol{\Psi} \in \mathbb{R}^{N\times N}$ transforms the signal from the sparse domain to the acquisition domain and the goal is reconstruction using M measurements where $M \ll N$. These measurements are described by:

$$\mathbf{y} = \boldsymbol{\Omega}\boldsymbol{\Psi}\mathbf{w}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^M$ are the measured data and $\boldsymbol{\Omega} \in \mathbb{R}^{M\times N}$ is the sensing matrix. The i-th column of $\boldsymbol{\Psi} \in \mathbb{R}^{N\times N}$ is the i-th basis element $\boldsymbol{\psi}_i \in \mathbb{R}^N$ evaluated at all N points of interest. There is a question as to what $\boldsymbol{\Omega}$ should be, with the Gaussian matrix being popular. However, this completely disregards limitations of the real world where the choice of sampling and thus sensing method might be severely restricted. To this end, $\boldsymbol{\Omega}$ is the zero matrix, apart from one non-zero entry equal to 1 per row. Thus, a measurement is either taken or not. With this in mind, we simplify:

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{w}, \tag{2}$$

where $\boldsymbol{\Phi} = \boldsymbol{\Omega}\boldsymbol{\Psi}$. The i-th column of $\boldsymbol{\Phi}$ is the i-th basis element evaluated at only M points, denoted by $\boldsymbol{\phi}_i \in \mathbb{R}^M$. One solution for the under-determined system of equations in (2) is to set a sparsity constraint. This can be achieved by minimizing the l_0 norm of $\mathbf{w}$ which represents the number of non-zero elements of the signal. However, this is intractable in general [16]. The breakthrough was made by a series of papers ([17], [18]) arriving at an approximation by minimizing the l_1 norm.

Sparse Bayesian learning was introduced in CS [19] which treats the problem as regression. The training data are the measurements and the model's coefficients are included in the sparse signal:

$$f(x_i;\mathbf{w}) = \mathbf{w}^T\boldsymbol{\psi}(x_i) + \epsilon_i, \tag{3}$$

where ϵ_i represents noise and $\boldsymbol{\psi}(x_i) \in \mathbb{R}^N$ contains the values of all N basis functions at point x_i. After training the model, the values of an unknown data point, x_*, can be inferred by:

$$f(x_*;\mathbf{w}) = \mathbf{w}^T\boldsymbol{\psi}(x_*). \tag{4}$$

Therefore, the objective is to find $\mathbf{w}$ that fits the training data best. In order to obtain predictive uncertainties, a probabilistic framework through the RVM [20] [21] is used. A zero mean Gaussian prior distribution is chosen to satisfy the sparsity condition of the coefficients, defined by:

$$p(\mathbf{w}|\boldsymbol{\alpha}) = \prod_{i=1}^{N}\mathcal{N}(w_i|0,\alpha_i^{-1}), \tag{5}$$

where $\boldsymbol{\alpha} = (\alpha_1,...,\alpha_N)$ represent the inverse variance for each coefficient's distribution. The measurements are assumed to have additive Gaussian noise and therefore the likelihood model is a Gaussian distribution defined by:

$$p(\mathbf{y}|\mathbf{w},\sigma^2) = \prod_{i=1}^{M}\frac{1}{\sqrt{2\pi\sigma^2}}e^{\left(-\frac{(y_i-f(x_i;\mathbf{w}))^2}{2\sigma^2}\right)}. \tag{6}$$

The posterior distribution of $\mathbf{w}$ is given by Bayes' rule:

$$p(\mathbf{w}|\mathbf{y},\boldsymbol{\alpha},\sigma^2) = \frac{p(\mathbf{y}|\mathbf{w},\sigma^2)p(\mathbf{w}|\boldsymbol{\alpha})}{p(\mathbf{y}|\boldsymbol{\alpha},\sigma^2)}. \tag{7}$$

This is a multivariate Gaussian distribution with posterior covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{N\times N}$ and mean $\boldsymbol{\mu} \in \mathbb{R}^N$:

$$\boldsymbol{\Sigma} = (\sigma^{-2}\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \mathbf{A})^{-1}, \tag{8}$$

$$\boldsymbol{\mu} = \sigma^{-2}\boldsymbol{\Sigma}\boldsymbol{\Phi}^T\boldsymbol{y}, \tag{9}$$

where $\mathbf{A} \in \mathbb{R}^{N\times N}$ is a diagonal matrix with the hyper-parameters $\boldsymbol{\alpha}$ on the diagonal. The RVM obtains the optimum $\boldsymbol{\alpha}$ and σ^2 is either fixed at an assumed level or inferred from the data. This is done by maximizing the marginal log-likelihood for $\boldsymbol{\alpha}$ and σ^2 [20]. The predictive mean at an unknown data point, x_*, can be inferred by:

$$f(x_*) = \boldsymbol{\psi}(x_*)^T\boldsymbol{\mu}. \tag{10}$$

The predictive variance is given by:

$$\sigma_*^2 = \sigma^2 + \boldsymbol{\psi}(x_*)^T\boldsymbol{\Sigma}\boldsymbol{\psi}(x_*), \tag{11}$$

where σ^2 controls the assumed noise level.

III. Predictive Uncertainty Measure

It has been noted [22] that the predictive variance depends on the choice of basis functions. It is common to choose basis functions which decay quickly or basis functions with finite support. Therefore, all elements of $\psi(x_*)$ could be small or zero. In this case, $f(x_*)$ and σ_*^2 are close to zero, making the predictive distribution meaningless. To resolve this issue, we utilize a different measure that uses the change in the logarithm of the marginal likelihood when including a data point, x_*, in the model. The change is given by [1]:

$$\Delta\mathcal{L} = \log\left[\frac{1}{\sqrt{2\pi}\sigma_*}\exp\left(-\frac{(f(x_*)-y_*)^2}{2\sigma_*^2}\right)\right], \tag{12}$$

where y_* is the true value at the estimated data point. Thus, the log likelihood is changed by the logarithm of

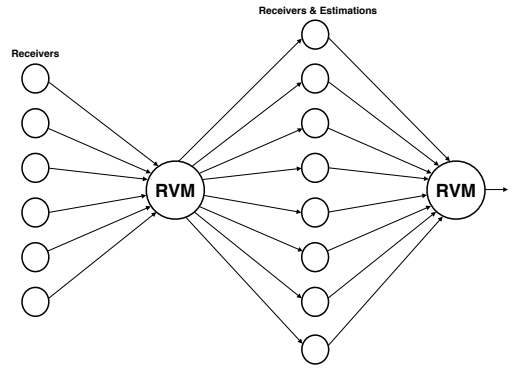


Figure 1. Layers of Relevance Vector Machines that propagate the first estimations and the original receivers' values to the next stage using the predictive uncertainty measure. Each instance of the RVM is different from the other by the choice of the dictionary of basis functions.

the likelihood of the new data value, y_*, at x_* given the predictive probability distribution $\mathcal{N}(f(x_*), \sigma_*^2)$. Since $\sigma_* \geq \sigma$, the change lies between $-\infty$ and $log\ \frac{1}{\sqrt{2\pi}\sigma}$. If the likelihood of the data is small, the log likelihood is reduced largely indicating that the model's prediction is poor at x_*. We want a measure of how good the predictive probability distribution $\mathcal{N}(f(x_*), \sigma_*^2)$ is, however, y_* is unavailable. Thus, we estimate the probability distribution of y_* to be normal with mean and variance given by:

$$\bar{\mu} = \underset{x_i \in \mathcal{S}}{\text{mean}}\{y_i\}, \tag{13}$$

$$\bar{\sigma}^2 = \underset{x_i \in \mathcal{S}}{\text{var}}\{y_i\}, \tag{14}$$

where $\mathcal{S}$ is a subset of the training data points (neighbours). The expected change in log likelihood is given by [1]:

$$E\left[\Delta\mathcal{L}\right] = log\frac{1}{\sqrt{2\pi}\sigma_*} - \frac{\bar{\sigma}^2 + (\bar{\mu} - f(x_*))^2}{2\sigma_*^2}. \tag{15}$$

The second term is the important one. If the predictive probability distribution does not match well with the probability distribution estimated from the data in the neighbourhood, the expected change in log likelihood is a negative number. Using the measure in equation (15), it is possible to create an uncertainty map with the largest negative values being the most uncertain regions. Furthermore, using the uncertainty map, it is possible to create a deep model, using more layers of RVMs. Each RVM could utilize a different dictionary of basis functions and as input, it could use the output of the previous RVM by propagating only the estimations that are trustworthy. Figure 1 illustrates this network architecture.

IV. Seismic Survey Experiments

In a typical seismic survey, body waves created by a surface source pass down into the earth with a proportion being reflected back by impedance changes caused by changes in rock properties at depth. The reflected waves arrive back at the surface and are recorded by receivers, where the output of a receiver is a digitized time series recorded at a constant sample rate. The seismic traces can be organized in different ways: a line of receivers can be viewed over all time stamps, called the shot domain. Another way of looking at the data would be to take the output of all receivers at a single instant of time, known as a time slice.

In order to evaluate the performance of our proposed system, seismic data were extracted from a synthetic data set, called SEAM-2 [23], that was kindly provided by BP. If a receiver is missing, a column of data would be missing in the shot domain and a single point would be missing from the time slice. Exhaustive experiments were made in the time slice domain. For the RVM with many layers, estimated data points are propagated through the network when the uncertainty is large. We provide examples of predictive uncertainty maps and also illustrate the CS capabilities in the shot domain. To ensure that the results are consistent over different instances of time slices with different structures, we have extracted two hundred 128×128 patches. These patches were then used to randomly remove receivers and then reconstruct them. In all experiments, the SPGL1[1], the POCS[2] and RVM[3] packages were used.

A. Trade-off between accuracy and time

The main variables that can change in the network architecture are the number of layers, the patch size and the basis functions. Due to the sequential nature of the model, the more layers we use the longer the computational time. To test the depth of the architecture, we used the two-dimensional multi-scale Haar Wavelet Transform. By using the smallest scale with support of 2×2, the finest details of the signal were captured. Then, the basis functions of the second scale with support of 4×4 were used to capture larger regions and so on. In this experiment, we used networks with one, two and three layers and for each we also varied the patch size between 8×8, 16×16 and 32×32. In addition, we used the Discrete Cosine Transform (DCT) with one layer. Two hundred patches of seismic time slices were decimated in different percentages and then reconstructed by each of these configurations. The Feature Similarity (FSIM) index [24] was used, which captures the degradation of a signal as perceived in features (FSIM $= 1$ is perfect reconstruction). Figure 2 shows the mean FSIM against the computational time in seconds. By using 80% of the receivers, the accuracy is high for all configurations. Smaller patches take less time to execute, along with architectures with fewer layers. By using 40%, there is a difference in accuracy between the first layer of the Haar wavelets and the rest. This is due to the fact that there are regions that are uncovered. The RVM with

[1] http://www.cs.ubc.ca/labs/scl/spgl1

[2] http://www.freeusp.org/synthetics/POCS_example/index.html

[3] http://www.miketipping.com/downloads.htm

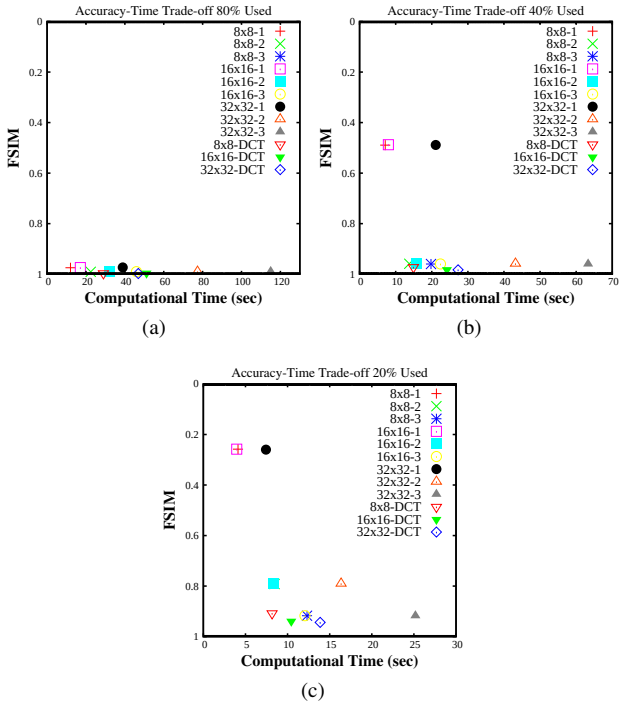


Figure 2. Accuracy and computational time trade-off for various configurations of the network. Each configuration is labelled first with its patch size and then with the number of layers used. (a) shows the trade-off when 80% of receivers are used, (b) uses 40% of receivers and (c) uses 20%.

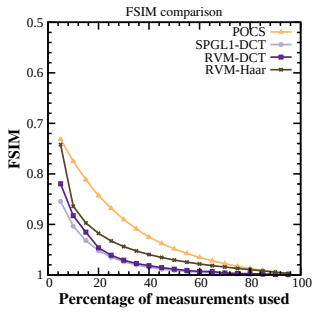


Figure 3. Seismic FSIM comparison

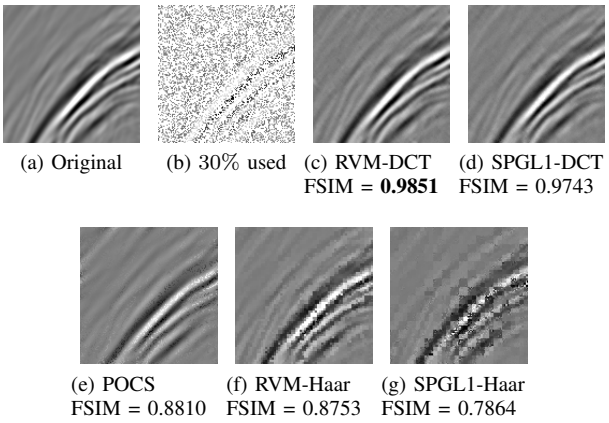


Figure 4. Example of a time slice from the SEAM 2 dataset (a) is decimated to only 30% of the receivers (inverted for visualization) (b). Reconstruction by the RVM with DCT (c) outperforms both the SPGL1 with DCT (d) and POCS (e). The RVM with Haar (f) performs better than the SPGL1 with Haar (g).

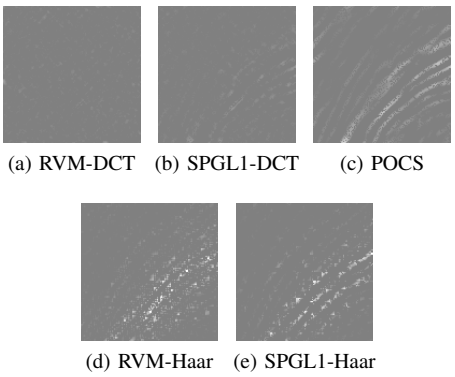
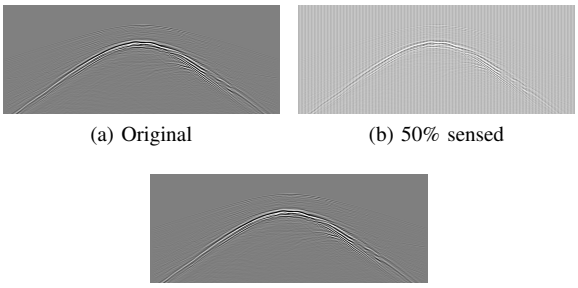


Figure 5. Difference maps of the original signal from Figure 4 compared to each respective reconstruction. All difference maps show the absolute difference but shifting all values by the median value of the domain for illustration purposes. (a) shows the error in the RVM using DCT basis functions which is minimal. (b) shows the error of the SPGL1 with DCT being more visible but still small. (c), (d) and (e) show errors in regions of large changes in the original signal.

the DCT provides slightly better accuracy. By using 20% of the receivers, the accuracy is now clustered in three groups. The RVM with three layers can achieve similar results in accuracy to that of the RVM with DCT illustrating that the multi-scale usage of simple basis functions can match the performance of the domain specific basis functions.

B. Comparisons

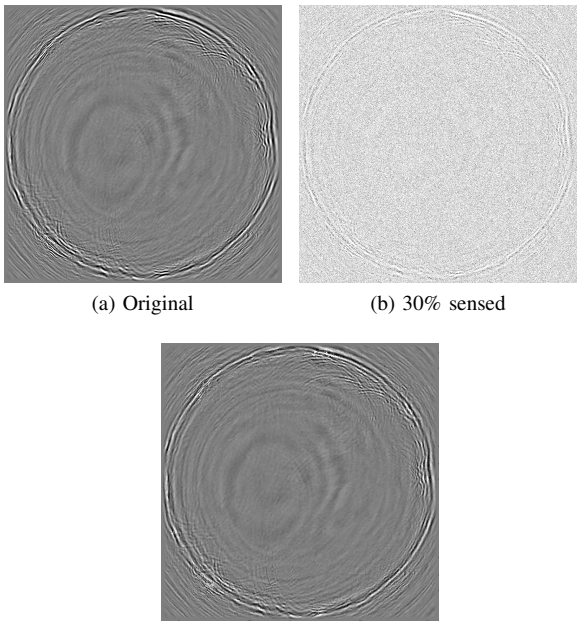
We mentioned earlier that the Fourier Transform is used in seismic data processing as a sparsifying transform and thus we compare the solvers using the DCT. The goal is to make a comparison on reconstruction accuracy between POCS, SPGL1 with DCT, the RVM with DCT and the RVM with Haar wavelets (for completeness) using three layers on 32×32 patches. We decimated two hundred seismic time slices in different percentages and reconstructed them using all solvers. The mean FSIM for all solvers is plotted against the percentage of measurements used in Figure 3. Examples of different reconstructions can be seen in Figure 4 along with the respective difference maps for each solver in Figure 5. We have also provided a reconstruction of SPGL1 with Haar wavelets of scale three in Figure 4. This shows that in this setting, the RVM does better at reconstructing than SPGL1 and it shows the great importance of choosing the correct basis functions for the solver. The RVM with DCT performs as good as the SPGL1 with DCT with 20% of receivers and more and they both outperform POCS. The RVM with Haar is better than POCS but is not as good as the RVM with DCT and SPGL1 with DCT. For seismic shot



(a) Original (b) 50% sensed

(c) Reconstruction

Figure 6. 500×1281 shot record (a) reconstruction using the RVM (c) with three layers. Each line is the entire signal of one receiver and decimation is done every other line (b) (inverted for visualization).



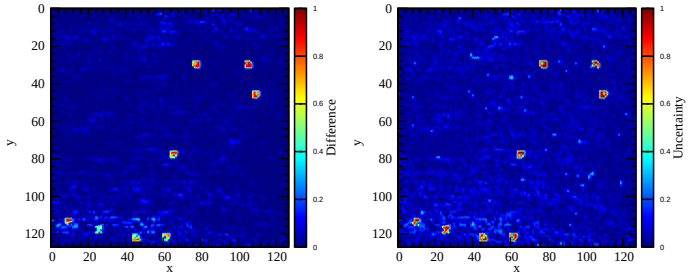
(a) Original (b) 30% sensed

(c) Reconstruction

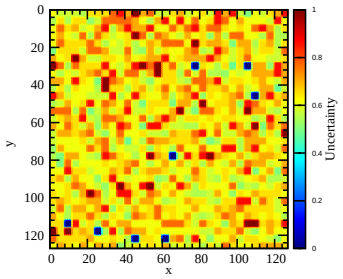
Figure 7. 1281×1281 time slice (a) reconstruction using the RVM (c) with three layers from 30% (b) of the original (inverted for visualization).

(a) Original (b) 50% used (c) Early stop

Figure 8. The algorithm was stopped in the second layer using Haar Wavelets to show an example of poor reconstruction. This is done in order to compare the uncertainty maps. (b) is inverted for visualization purposes.



(a) Differences (b) Proposed Measure



(c) Predictive Variance

Figure 9. Figure (a) illustrates the differences between the original time slice and the reconstruction. It shows that the majority of the area is reconstructed accurately (in blue) with some areas being reconstructed poorly. Figure (b) shows the uncertainty map produced by the proposed predictive uncertainty measure. Figure (c) illustrates the predictive variance which fails to capture the error. All predictive uncertainties were shifted and normalized in the $0-1$ range. The proposed measure produces large uncertainty in regions of large reconstruction error and vice versa.

records, lines in the signal are removed. An example of 50% of missing lines is shown in Figure 6b where every other line is discarded and the reconstruction using the RVM with Haar and three layers is shown in Figure 6c. Furthermore, a complete time slice can be seen in Figure 7.

C. Uncertainty maps

As discussed, the predictive variances from equation (11) do not behave as expected and thus we have proposed the use of a new predictive uncertainty measure in equation (15). In order to illustrate this, an example of an uncertainty map was produced using the proposed measure and compared to the predictive variances. Figure 8a shows the original signal, Figure 8b shows the efficiently sensed seismic signal from 50% of receivers and Figure 8c shows the failed reconstruction (stopped early). Figure 9a shows the error differences between the original and the reconstructed, Figure 9b shows the uncertainty map produced by the new measure and Figure 9c shows the uncertainty map produced by the RVM with the original predictive variances. The model with the proposed measure is more certain about regions of good reconstruction and less certain in regions that the model did not perform well. The proposed uncertainty measure can be used for seismic survey design, by using the available data, it can suggest regions to place receivers to minimize uncertainty.

V. Conclusion

In this work, we applied the RVM to the seismic CS framework and used a new predictive uncertainty measure.

Using this measure, we were able to create networks of RVMs with many layers. Due to the different possibilities of the network, we performed experiments to examine the trade-off between reconstruction accuracy and computational time. We compared the RVM with state-of-the-art seismic reconstruction algorithms and demonstrated that it has comparative reconstruction accuracy with SPGL1 and outperforms POCS. In addition, we examined the uncertainty map produced by the proposed measure and illustrated that it compares well with the errors in the reconstruction as opposed to the predictive variances of the original RVM.

The proposed framework offers great advantages such as state-of-the-art reconstruction accuracy and practical computational time through the utilization of smaller patches which can be processed in parallel. It also provides a multi-scale dictionary framework that is able to capture different variations in the data using appropriate basis functions, analytic or learned from data. It is possible to improve the reconstruction accuracy by passing only the relevant information from each layer. In addition, the predictive uncertainty measure can be very useful when determining the level of risk associated with the reconstruction on particular regions. This can provide an on-line data acquisition process where the algorithm guides the seismic survey design in the future.

ACKNOWLEDGMENT

We would like to thank BP for kindly allowing the publication of this research. In addition, we would like to thank Ray Abma, Chris Brookes, Mark Foster, James Keggin and Neil Philip from BP for discussions on seismic data acquisition. Also, we would like to thank Nikolaos Nikiforakis, Alan O'Neill, Carola Schoenlieb and Michael Tipping for various aspects of the paper. We also thank SEAM-II for the model and Carl Regone of BP for modelling the seismic data. We acknowledge funding from the EPSRC through a CASE studentship.

REFERENCES

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

[2] M. Sacchi, T. Ulrych, and C. Walker, "Interpolation and extrapolation using a high-resolution discrete fourier transform," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 31–38, Jan 1998.

[3] D. O. Trad, T. J. Ulrych, and M. D. Sacchi, "Accurate interpolation with high-resolution time-variant radon transforms," *Geophysics*, vol. 67, no. 2, pp. 644–656, 2002.

[4] F. J. Herrmann and G. Hennenfent, "Non-parametric seismic data recovery with curvelet frames," *Geophysical Journal International*, vol. 173, no. 1, pp. 233–248, 2008.

[5] R. Abma and N. Kabir, "3d interpolation of irregular data with a pocs algorithm," *Geophysics*, vol. 71, no. 6, pp. E91–E97, 2006.

[6] A. Stanton, M. D. Sacchi, R. Abma, and J. A. Stein, "Mitigating artifacts in projection onto convex sets interpolation," *SEG Technical Program Expanded Abstracts 2015*.

[7] P. M. Zwartjes and M. D. Sacchi, "Fourier reconstruction of nonuniformly sampled, aliased seismic data," *Geophysics*, vol. 72, no. 1, pp. V21–V32, 2007.

[8] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Img. Sci.*, vol. 2, no. 1, pp. 183–202, Mar. 2009.

[9] D. O. Prez, D. R. Velis, and M. D. Sacchi, "High-resolution prestack seismic inversion using a hybrid fista least-squares strategy," *Geophysics*, vol. 78, no. 5, pp. R185–R195, 2013.

[10] E. van den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2009.

[11] J. Cao, Y. Wang, and B. Wang, "Accelerating seismic interpolation with a gradient projection method based on tight frame property of curvelet," *Exploration Geophysics*, vol. 46, pp. 253–269, 2015.

[12] N. Kreimer and M. D. Sacchi, "A tensor higher order singular value decomposition (hosvd) for prestack simultaneous noise reduction and interpolation," *SEG Technical Program Expanded Abstracts 2011*.

[13] A. Stanton, N. Kreimer, D. Bonar, M. Naghizadeh, and M. Sacchi, "A comparison of 5d reconstruction methods," *SEG Technical Program Expanded Abstracts 2012*.

[14] L. Zhu, E. Liu, and J. H. McClellan, "Seismic data denoising through multiscale and sparsity-promoting dictionary learning," *Geophysics*, vol. 80, no. 6, pp. WD45–WD57, 2015.

[15] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.

[16] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, Apr. 1995.

[17] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb 2006.

[18] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[19] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, June 2008.

[20] A. C. Faul and M. E. Tipping, "Analysis of sparse bayesian learning," in *Advances in Neural Information Processing Systems 14*. MIT Press, 2001, pp. 383–389.

[21] M. E. Tipping and A. Faul, "Fast marginal likelihood maximisation for sparse bayesian models," in *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, 2003.

[22] C. E. Rasmussen and J. Quiñonero Candela, "Healing the relevance vector machine through augmentation," in *Proceedings of the 22nd International Conference on Machine Learning*, 2005.

[23] M. Oristaglio, "Seam phase ii land seismic challenges," *The Leading Edge*, vol. 31, no. 3, pp. 264–266, 2012. [Online]. Available: http://dx.doi.org/10.1190/1.3694893

[24] L. Zhang, D. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, Aug 2011.