https://doi.org/10.2352/ISSN.2470-1173.2020.14.COIMG-307

Reconstruction of 2D Seismic Wavefields from Nonuniformly Sampled Sources

Laura Galvis [1], Juan M. Ramírez [1], Edwin Vargas [1], Ofelia Villarreal [2], William Agudelo [3], Henry Arguello [1]
[1] **Universidad Industrial de Santander, Santander, Colombia**
[2] **Cooperativa de Tecnólogos e Ingenieros de la Industria del Petróleo y Afines, TIP**
[3] **Instituto Colombiano del Petróleo, ICP, Ecopetrol S.A., Santander, Colombia**

Abstract
In a 3D seismic survey, the source sampling in a regular grid is commonly limited by economic costs, geological constraints, and environmental challenges. This non-uniform sampling cannot be ignored since the lack of regularity leads to incomplete seismic data with missing 2D wavefields. Notice that the post-processing tasks have been developed under the assumption that 3D seismic data are obtained from a regular sampling. Therefore, signal recovery from incomplete data becomes a crucial step in the seismic imaging processing flow. In this work, we propose a pre-processing step that includes the nonuniformly acquired wavefields in a finer regular grid, such that shot gathers are stacked considering the actual spatial location of the sources. Then, based on the 3D curvelet transform, a sparse signal recovery algorithm that considers an interpolation operator is employed in order to reconstruct the missing wavefields in a regular grid. The performance of the proposed seismic reconstruction approach is evaluated on a real data set.

Introduction
A seismic data set is the collection of units known as traces. Each source and receiver pair generates one of these traces, and the set of all the traces together provides a spatio-temporal sampling of the reflected wavefield, which contains different arrivals corresponding to the various interactions of the incident wavefield with the subsurface [1]. The follow-up processing and interpretation, such as the velocity analysis, normal-moveout, statics corrections, and the migration imaging process, demands integrity of the seismic data, which requires the reconstruction of incomplete data. Even more, in terms of compression, the requirement of higher resolutions, which, in turn, demands the increasing amount of data collection, as well as the expansion of explored areas, makes data compression an important and challenging issue [2].

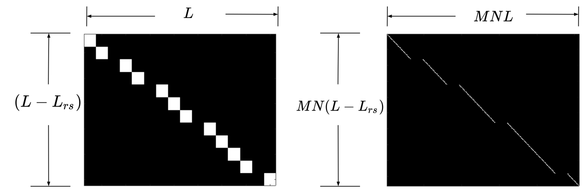
The reconstruction of seismic data in the state-of-the-art can be achieved through different schemes. First, methods based on the propagation characteristics of the seismic wave, which solve the inverse problem by Dip Moveout (DMO) or Azimuth Moveout (AMO) [3]. Second, methods based on prediction filtering, where high-frequency information is predicted from low-frequency information using a frequency estimation method, such as the F-X domain prediction error reconstruction [4]. Third, emerging computational imaging approaches consisting of reconstruction methods based on the compressive sensing theory [5]. These methods use the transformation of the seismic data into a sparse domain and exploit the incoherence to establish bounds for the exact recovery of the signal [6] [7].

The compressive sensing (CS) theory states that it is possible to recover compressible signals from fewer samples than those required by the Nyquist-Shannon sampling theorem. Since the seismic data is not sparse per se, the signal should exhibit sparsity in a transform domain. Some transforms have been used into the sparse representation of seismic data in the literature, Fourier, Seislet, Curvelet, among others. Nonetheless, some of them achieve an optimally sparse representation given its properties, such as the Curvelets, which exploit smoothness along arriving wavefronts and differentiates among different signal components on the basis of location, angle, and frequency content. The implementations of the fast discrete curvelet transform assume a regular sampling along all axes [1]. However, in real acquisitions, there are some limitations, mainly due to the positions of sources and receivers. To address this issue, we propose a re-organization of the data cube such that the shots are stacked, considering the actual spatial locations of the sources. After this step referred to as a binning process, the reconstruction of a number of sources or complete shots can be performed using the 3D curvelet transform. Moreover, to improve the reconstructions, an initialization with an interpolating operator is also proposed in the recovery process. The feasibility and effectiveness of the proposed method are verified by reconstructing actual 2D seismic data.

Compressive sensing in seismic reconstruction
Considering a 2D seismic acquisition, the ideal discrete seismic data of the wavefield sampling on a regular grid can be represented as a discrete data volume $\mathbf{F} \in \mathbb{R}^{N\times M\times L}$, where N is the number of discrete-time samples captured by each receiver at a seismic shot, M is the number of receivers, and L is the number of shots required for building the desired data cube $\mathbf{F}$. However, due to different reasons, such as economic limitations, environmental constraints, and elimination of low SNR acquired traces, the observed seismic field data is irregular and incomplete. Therefore, seismic data recovery is a crucial step in the seismic imaging processing flow. In this work, we consider a regular sampling along the receiver line, and we assume there are missing shots along the source line. More formally, the observed seismic data can be modeled as

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\eta}, \tag{1}$$

where $\mathbf{g} \in \mathbb{R}^{NM(L-L_{rs})\times 1}$ represents the acquired data in vector form with L_{rs} as the number of suppressed sources, $\mathbf{f} \in$



(a) Matrix **Z** (b) Sensing Matrix **H**

Figure 1. *(a) Matrix* $\mathbf{Z}$ *for* $L=16$*, and* $L_{rs}=4$ *(b) sensing matrix* $\mathbf{H}$ *for* $M=4$*,* $N=4$*,* $L=16$*, and* $L_{rs}=4$*.*

$\mathbb{R}^{NML\times 1}$ is the desired 3D data reorganized in vector form, $\mathbf{H}\in\mathbb{R}^{NM(L-L_{rs})\times NML}$ is the sensing matrix that describes the acquisition process, and $\boldsymbol{\eta}\in\mathbb{R}^{NM(L-L_{rs})\times 1}$ denotes the additive noise vector whose components are considered as independent and identically distributed (iid) random samples following a zero-mean Gaussian distribution.

To describe the structure of the sensing matrix $\mathbf{H}$, consider $z'=\{z_1,z_2,\ldots,z_{L_{rs}}\}$ as the index set of removed sources on a fine regular grid. Therefore, $\mathbf{H}$ is given by

$$\mathbf{H}=\mathbf{Z}\otimes\mathbf{I}_{MN}, \tag{2}$$

where $\mathbf{Z}\in\mathbb{R}^{L_{rs}\times L}$ is a rectangular matrix containing the location information of the removed sources, $\otimes$ denotes the Kronecker product operator, and $\mathbf{I}_{MN}$ is the identity matrix with dimensions $MN\times MN$. Figure 1 (a) and (b) pictorially depict the matrix $\mathbf{Z}$ and the sensing matrix $\mathbf{H}$, respectively, for $M=4$, $N=4$, $L=16$, and $L_{rs}=4$.

In this work, we aim at finding the missing 2D wavefield slices, corresponding to the removed sources in the seismic data acquisition process, from available measurements. To this end, the recovery of the missing slices is formulated as a reconstruction problem of the 3D desired seismic data $\mathbf{f}$ from undersampled measurements $\mathbf{g}$ in the context of compressive sensing theory, this reconstruction is called sparse recovery. In this context, a target signal can be recovered from compressive measurements with high probability when it exhibits a sparse representation in a given basis $\boldsymbol{\Psi}$, i.e., the desired signal can be succinctly described as $\mathbf{f}=\boldsymbol{\Psi}\boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is the sparse vector that represents the target signal in the basis domain. Therefore, the compressive measurements, defined in Eq. (1), can also be expressed as

$$\mathbf{g}=\mathbf{H}\boldsymbol{\Psi}\boldsymbol{\Theta}+\boldsymbol{\eta}. \tag{3}$$

Reconstruction problem

Based on the assumption that the components of the acquisition noise vector $\boldsymbol{\eta}$ are iid random samples following a zero-mean Gaussian distribution, the reconstruction problem reduces to the minimization of the ℓ_2-norm of the error vector between the measurements and an undersampled version of the desired 2D seismic data. However, the structure of the sensing matrix $\mathbf{H}$ leads to an ill-posed optimization problem. To overcome this drawback, the fact that the desired signal can be described as a sparse vector in a given representation basis is exploited by including a regularization term in the optimization problem. Therefore, the reconstruction of the 3D seismic data is formulated as

$$\hat{\boldsymbol{\Theta}}=\arg\min_{\boldsymbol{\Theta}}\left\{\frac{1}{2}\|\mathbf{g}-\mathbf{H}\boldsymbol{\Psi}\boldsymbol{\Theta}\|_2^2+\lambda\|\boldsymbol{\Theta}\|_1\right\}, \tag{4}$$

where λ is the regularization parameter that controls the trade-off between the error term and the sparsity inducing term. There are many algorithms to solve the sparse regularization problem in Eq. (4). To name a few, the gradient projection for sparse reconstruction (GPSR) [8], the iterative shrinkage thresholding algorithm (ISTA) [9], and the alternating direction method of multipliers (ADMM) [10], are state-of-the-art algorithms. In this work, an ADMM-based algorithm is used to solve the problem in (4).

Reconstruction with an interpolating operator

Iterative algorithms that solve the optimization problem in Eq. (4), require to compute the adjoint operator of $\mathbf{H}$. The form of the sampling operator $\mathbf{H}$ leads to an adjoint operator $\mathbf{H}^\dagger$ that maps the irregular data along the source line onto a regular grid with zeros in the positions corresponding to the L_{rs} removed sources. Thus, the reconstruction of the seismic data using sparse regularization problems as described in Eq. (4) is very poor. To overcome this problem, we propose an alternative model to the sparse recovery that additionally includes an initialization in the form of an interpolating operator. Specifically, first, consider the interpolation along the source line of the observed seismic data $\mathbf{g}$ given by

$$\bar{\mathbf{g}}=\mathbf{R}\mathbf{g}, \tag{5}$$

where $\mathbf{R}$ is an interpolating operator and $\bar{\mathbf{g}}\in\mathbb{R}^{NML\times 1}$ is the interpolated data. Then, the optimization problem is re-formulated including the initialization as

$$\hat{\boldsymbol{\Theta}}=\arg\min_{\boldsymbol{\Theta}}\left\{\frac{1}{2}\|\bar{\mathbf{g}}-\boldsymbol{\Psi}\boldsymbol{\Theta}\|_2^2+\lambda\|\boldsymbol{\Theta}\|_1\right\}. \tag{6}$$

Binning pre-processing step

When seismic data is acquired in land, it is common to find several obstacles limiting the uniform localization of receivers, sources, or both. A synthetic representation of a non-uniform sensing in the source line is shown in Fig. 2, the common offset gather view clearly illustrates how continuity along wavefronts is lost when casting non-uniform sensed data to a regular grid. With the aim to use the 3D Curvelet transform to sparsify the seismic data as in Eqs. (4, 6), a binning pre-processing step is proposed to cast the irregularly sampled data into a regular grid maintaining the continuity required on this basis.

The non-uniformly sampled data $\mathbf{F}$ has M samples over time, N uniform located receivers in the receiver line, and L non-uniform sensed sources, where the $\{\ell_1,\ldots,\ell_L\}$ sources can not be uniform located in a grid. Therefore, a new distribution of sources is required. The source interval (SI) defines the distance between sources, and it is variable for the data cube $\mathbf{F}$. The transformation of the non-uniform data cube $\mathbf{F}$ into a new data cube $\hat{\mathbf{F}}$ is achieved defining a uniform SI parameter S, and a margin error of this interval ε. The x and y spatial coordinates of the sources are used to calculate the actual distance between the sources as $d=\{d_1,\ldots,d_{L-1}\}$, where d_1 is the distance between the first and

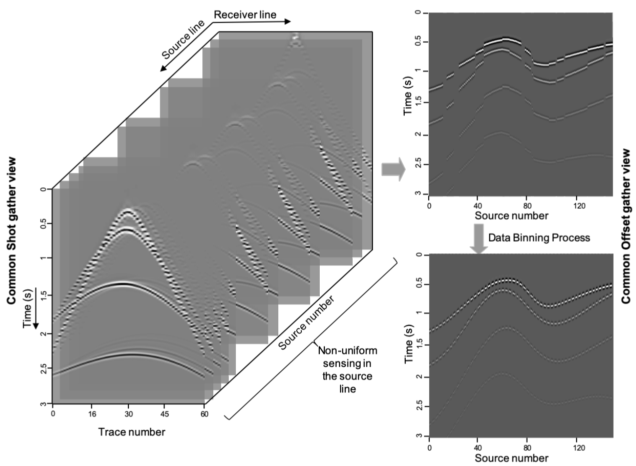


Figure 2. Synthetic seismic data. The common shot gather view shows the nonuniformly sampled data in the source line. The irregular data recasting onto a regular grid destroys the continuity along the arriving wavefronts in the top common offset gather. In the bottom common shot gather, the binning process cast the irregular sampled data into a regular grid to maintain the continuity.

the second source. The uniform grid is constructed dividing the cumulative distance $c = \sum d_i$ by the S defined previously, and the resulting value gives the number of intervals in the uniform grid, and therefore the new number of sources $\hat{L}$ in the data cube $\hat{\mathbf{F}}$. Then, the distances d are evaluated to find the nearest ℓ source to the new grid positions, and this operation is defined as,

$$\delta(d_i) = \begin{cases} \hat{\ell}_i = \ell_i, & \text{if } d_i \leq (S+\varepsilon) \text{ or } d_i \geq (S-\varepsilon) \\ \hat{\ell}_i = \mathbf{0}, & \text{otherwise ,} \end{cases} \tag{7}$$

where $\mathbf{0} \in \mathbb{R}^{M \times N}$ is a zero matrix, representing a non-sensed shot. The new data cube is then represented as $\hat{\mathbf{F}} \in \mathbb{R}^{M \times N \times \hat{L}}$, where the three axis are uniformly sensed.

Simulations and Results

A post-stack 2D real split-spread seismic data of a certain Colombian area is used to verify the effectiveness of the seismic data reconstruction. To evaluate the performance, the signal-to-noise ratio (SNR) quality metric is calculated.

The real seismic data consists of 150 sources and 295 receivers; the receiver interval is 20 m, the source interval $S = 25$ was established, as well as a margin error $\varepsilon = 5$. Each receiver recorded 2001 time samples within a sampling period of 4 ms. The application of the binning pre-processing step resulted in 58 shots assigned to source positions in the uniform grid, and 184 additional shots were added; these additional shots are zero-valued as presented in Eq. (7), this, to maintain the continuity in the source line. The complete uniform/binning data cube has a total of 242 shots. Then, to simulate the incomplete data cube, 6 and 11 randomly selected shots were removed, this is around 10 % and 20 % of compression from the original shots included in the binning data cube $\hat{\mathbf{f}}$.

The reconstruction of the removed sources was obtain solving the two formulations, the sparse recovery in Eq. (4), and the sparse recovery with the initialization using the interpolating operator in Eq. (6). The optimization is numerically solved with the ADMM algorithm, and the representation basis used is the 3D Curvelet. Figures 3 and 4 show the reconstruction of two removed shots for the two different scenarios of compression, 10% and 20%. The original shot is presented in the top, and the reconstructions using sparse recovery and sparse recovery with an interpolating operator are respectively shown in the bottom. The quality achieved when using the initialization and the binning data cube is noticeable. The SNR values for the reconstructed shots are included in the figures to facilitate the comparison. In the 20% compression scenario, Fig. 4, the proposed sparse recovery with the interpolating operator approach outperforms the results attained with the sparse recovery model. Table 1 reports the overall performance in the reconstruction of 6 and 11 removed shots, obtained for the two recovery methods. Notice the improvement when using the proposed method, which achieved around 10 and 7 dB of gain.

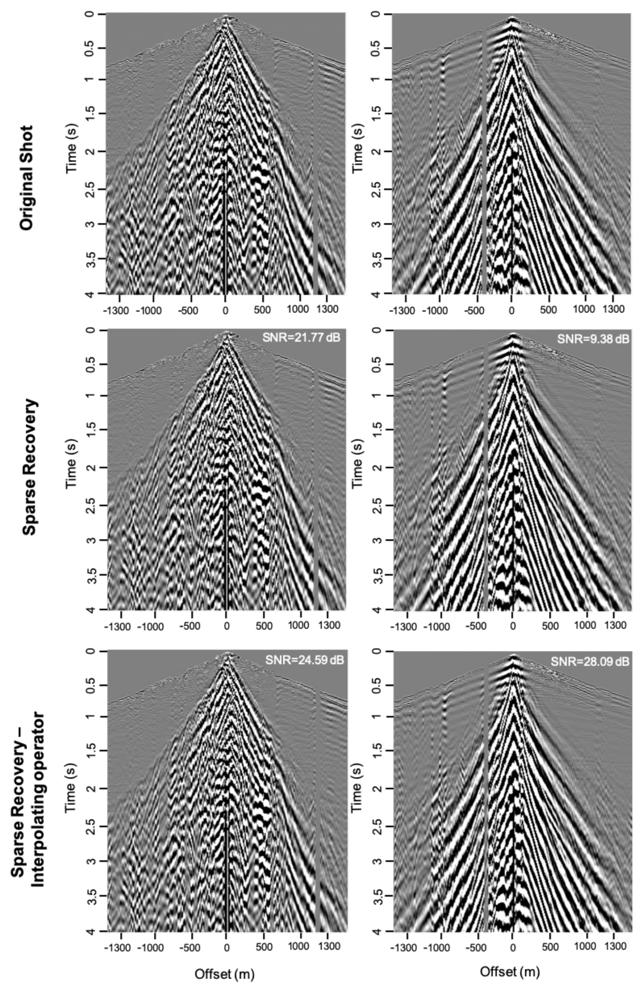


Figure 3. Reconstruction of 2 missing shots of a 2D wavefield in a 10% compression scenario for the two recovery methods.

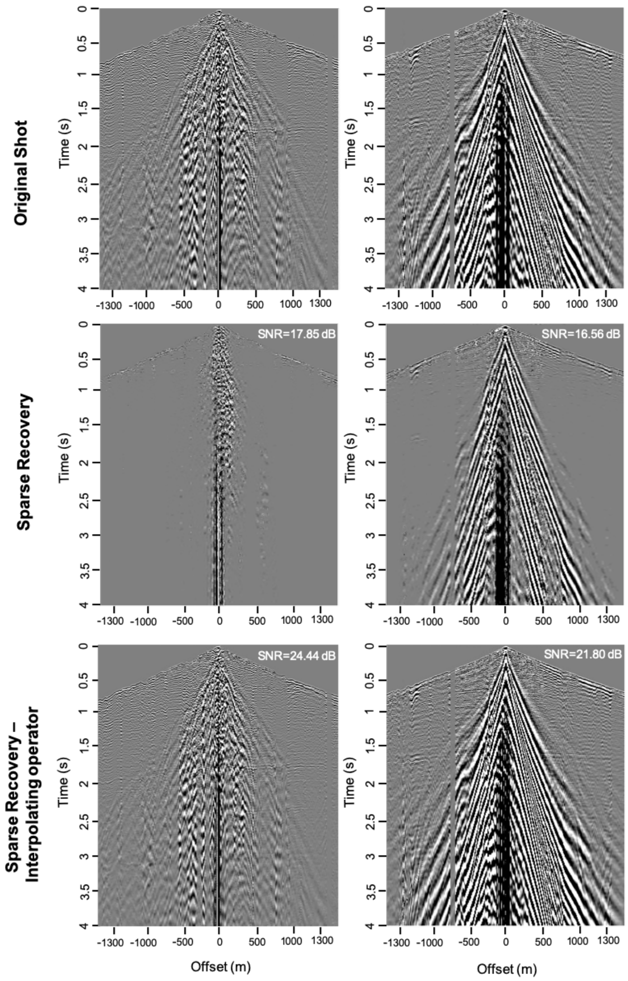


Figure 4. *Reconstruction of 2 missing shots in a* 20% *compression scenario for the two sparse recovery methods.*

Conclusions

A binning pre-processing of nonuniformly sampled seismic acquisition and its wavefield reconstruction have been proposed. The binning pre-processing allows to stack shot gathers accounting the actual spatial locations of the sources, to maintain the continuity along the source line. The reconstruction of missing shots is performed using a sparse recovery algorithm, based on the 3D Curvelet transform considering an initialization with an interpolation operator. Simulations using real 2D seismic line demonstrate the performance of the proposed approach.

Mean SNR achieved with the sparse recovery methods, for 2 compression scenarios

Number of removed sources (Compression)	mean SNR [dB]	
	Sparse Recovery	Sparse Recovery - initialization with interpolation
6 sources (10%)	14.6841	**24.8245**
11 sources (20%)	13.1754	**20.933**

Acknowledgments

This work was supported by UIS-ECOPETROL (AC 27), Vicerrectoría de Investigación y Extensión VIE-UIS (2486), and Colciencias-Programa de estancias postdoctorales (811-2018).

References

[1] G. Hennenfent and F. J. Herrmann, "Seismic denoising with nonuniformly sampled curvelets," *Computing in Science Engineering*, vol. 8, no. 3, pp. 16–25, May 2006.

[2] R. G. Baraniuk and P. Steeghs, "Compressive sensing: A new approach to seismic data acquisition," *The Leading Edge*, vol. 36, no. 8, pp. 642–645, 2017. [Online]. Available: https://doi.org/10.1190/tle36080642.1

[3] A. Canning and G. H. F. Gardner, "Regularizing 3-D data sets with DMO," *Geophysics*, vol. 61, no. 4, pp. 1103–1114, 08 1996. [Online]. Available: https://doi.org/10.1190/1.1444031

[4] S. Spitz, "Seismic trace interpolation in the f-x domain," *GEOPHYSICS*, vol. 56, no. 6, pp. 785–794, 1991. [Online]. Available: https://doi.org/10.1190/1.1443096

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

[6] O. P. Villarreal, K. León, D. Espinosa, W. Agudelo, and H. Arguello, "Compressive sensing seismic acquisition by using regular sampling in an orthogonal grid," in *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec 2017, pp. 1–5.

[7] K. L. López, J. M. Ramírez, W. Agudelo, and H. Arguello, "Regular multi-shot subsampling and reconstruction on 3d orthogonal symmetric seismic grids via compressive sensing," in *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*, April 2019, pp. 1–5.

[8] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of selected topics in signal processing*, vol. 1, no. 4, pp. 586–597, 2007.

[9] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 57, no. 11, pp. 1413–1457, 2004.

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

Author Biography

Laura Galvis Received the B.Sc. and M.Sc. degrees in computer science from the Universidad Industrial de Santander (UIS), Colombia, in 2009 and 2012, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Delaware (UDEL), Newark, DE, USA, in 2018. She is currently a postdoctoral researcher at UIS, sponsored by Colciencias and UIS. Her main research areas are computational optical imaging, compressive sensing, and high-dimensional signal processing.

Author Biography

Juan Marcos Ramírez received the B.S. in electrical engineering, M.S. in biomedical engineering and the Doctor degree in applied sciences from the Universidad de Los Andes, Mérida, Venezuela, in 2002, 2007 and

2017, respectively. His research interest includes sparse image representation, seismic signal processing and spectral image classification.

Author Biography

Edwin Vargas received the B.S. degree in electronic engineering from the Universidad Industrial de Santander, Bucaramanga, Colombia, in 2016, where he is currently pursuing the master's degree with the Department of Electrical and Computer Engineering. His research interests focus on the areas of high dimensional signal processing, compressive sensing, and dictionary learning.

Author Biography

Ofelia P. Villarreal received the B.E.E. degree from the University of Valle in Cali, Colombia, in 2006 and the Master degree from the Electrical and Computer Engineering Department, University of Delaware, Newark, DE, USA in 2014. Her current research interests include seismic signal and image processing.

Author Biography

William Agudelo received his Master's degree and the Ph.D. degree in Earth Sciences - Geophysics from the Université Pierre et Marie Curie (Paris 6) in 2001 and 2005, respectively. He is currently Senior Research Geophysicist at the R&D Center (Centro de Innovación y Tecnología - ICP) of the Colombian National Oil Company, Ecopetrol S.A.. His research interests include the implementation and development of new technologies in multicomponent seismic, Seismic Imaging, Seismic Acquisition Design and Surface Waves.

Author Biography

Henry Arguello received the Master's degree in electrical engineering from the Universidad Industrial de Santander, Bucaramanga, Colombia, and the Ph.D. degree from the Electrical and Computer Engineering Department, University of Delaware, Newark, DE, USA in 2003 and 2013, respectively. He is currently a Titular Professor at the Computer Science Department, Universidad Industrial de Santander. His research interests include statistical signal processing, high-dimensional signal coding and processing, optical imaging, optical code design, and computational imaging.