Z-noise attenuation using Gaussian Mixtures Models and unsupervised learning

Francesco Perrone and Sergio Grion, Shearwater GeoServices

SUMMARY

In Ocean Bottom Multicomponent acquisitions, the particle motion recordings can be heavily contaminated with shear-wave noise. The presence of this shear-wave noise, especially on the *Z*-component, hinders the pre-processing steps required to achieve an accurate calibration of the pressure and *Z*-component for up/down wave separation, also referred to as *PZ* summation. Failure to effectively attenuate this noise may reduce the quality of the separation and consequently the quality of the processing products that depend on it (up/down deconvolution and imaging). We implement a simple and flexible approach that operates in the Radon domain and uses clustering techniques borrowed from unsupervised machine learning to identify the coefficients that are likely noise and those that are likely signal. We use a set of local structural and similarity attributes to define the space of features for clustering and test the methodology on a real data example.

Introduction

Multicomponent ocean bottom acquisition systems allow the recording of the full vector particle motion and pressure components of the seismic wavefield. Because of the physics of wave propagation, multicomponent data make possible to naturally separate up-going and down-going propagating waves, that in turn allow the identification and separation of primary reflections from surface related multiples (Amundsen 2001). The separated components can then be used in imaging schemes like mirror-migration (Dash et al. 2009), which produce a wider image of the subsurface because of the wider aperture sampled by multiples with respect to primary reflections. However, not all components are created equal and they are affected by different types of noise in different ways. The particle motion recordings are usually (depending on the medium properties at the water bottom) heavily contaminated with shear-wave noise (Paffenholz et al. 2006), which appears mostly random in the common-shot domain, but strong and coherent in common-receiver gathers. The presence of this shear-wave noise on the *Z*-component hinders the pre-processing steps required to achieve an accurate calibration of the pressure and *Z*-component up/down wave separation, also known as *PZ* summation. Failure to effectively attenuate this noise may reduce the quality of the separation and consequently the quality of the processing products that depend on it (up/down deconvolution and imaging).

Because of the high variability of the *Z*-component noise, several strategies have been developed to tackle this denoising problem. Practically all of them rely on multidimensional transformations of the input data into dimensionally redundant spaces (e.g., Craft and Paffenholz 2007), such as the curvelet domain, and on measures of similarity between the *Z* component and the pressure data, which are usually free from shear wave noise or at least less contaminated. In this work, we propose an approach that operates in the linear Radon domain and uses clustering techniques borrowed from unsupervised machine learning (Pedregosa et al. 2011) to identify the coefficients that are likely noise and those that are likely signal. The clustering step produces a noise model that we then subtract from the input *Z*-component data.

Signal-Noise separation using Gaussian Mixture Models

The basic idea shared by most *Z*-noise attenuation approaches is to transform the data to a domain where the signal and noise components are clearly separated and thus denoising reduces to muting out the noise components. The most popular domains for signal-noise separation are the linear Radon domain (Craft and Paffenholz 2007; Naeini et al. 2011; Poole et al. 2012), the (complex) wavelet and curvelet domains (Peng et al. 2014; Yu et al. 2011). The advantage of wavelets and curvelets comes from their locality and the ability to automatically extract the orientation of the features in the data at every spatial location. The Radon transform can be implemented in a local fashion to achieve the same objective. The drawback is their computational cost and, in the case of the local Radon transform, extra care is required to avoid numerical artefacts caused by local windowing. We avoid computing expensive transforms in redundant domains like the curvelet-domain and use instead a strategy based on the estimation of local attributes to identify the noise in the Radon domain. We focus our attention on structural attributes (anisotropy and orientation) extracted from gradient square tensors (van Vliet and Verbeek 1995). Fomel (2007) proposed an inverse problem formulation for the estimation of local attributes such as similarity and instantaneous frequency. We use the approach proposed by Hale (2006) for computing local correlations.

Several denoising solutions based on techniques borrowed from the machine learning community have been recently proposed. Stork (2017) uses a pattern recognition strategy based on estimated covariance matrices to identify the signal and reject the noise components. Turquais et al. (2017) develop a coherent-noise attenuation approach based on dictionary learning. Once the signal and noise dictionaries are estimated, the latter can be used to identify and subtract the noise from the data. Our approach is also based on unsupervised machine learning techniques but does not involve the estimation of a dictionary; instead, we estimate a set of structural attributes in the Radon domain from the *P* and *Z* component (structural anisotropy, local correlations, similarity of orientation) and we estimate a probability density function in the multidimensional space spanned by these attributes using a Gaussian

Mixture Model (GMM). A GMM approximates the probability density function (PDF) of a random variable as linear superposition of Gaussian kernels, whose parameters (weights, means and covariances) are estimated using an Expectation-Maximization approach. (Press et al. 2007).

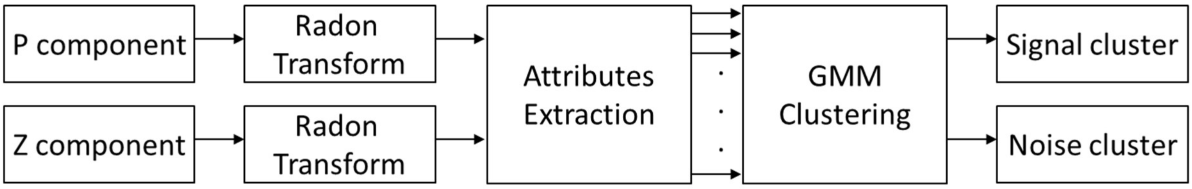

Figure 1 *Unsupervised classification flow using Gaussian Mixture Models.*

Noise and signal cluster in different areas of the multidimensional space spanned by the attributes: the signal displays high anisotropy, high correlation and similar orientation on the *P* and *Z* component; the noise clusters closer to the origin of the attribute space. The centre, covariance matrix and weights of the two clusters allow a classification of each component in the Radon domain (Figure 1). It is important to scale the attributes and have them span similar ranges of values. In this work, the attributes are all defined on the unit interval [0,1]. The noise estimate is not used "as is"; the envelope of the noise is instead compared with the envelope of the noisy Z-component and if the ratio between the two is beyond a certain threshold that Radon component is subtracted from the data. The algorithm is the following:

- Transform *P* and *Z* to the Radon domain
- Compute the attributes for classification
- GMM clustering and extraction of the noise component
- Soft-thresholding of the input data based on the envelope of the input and estimated noise
- Inverse Radon transform the denoised *Z*-component

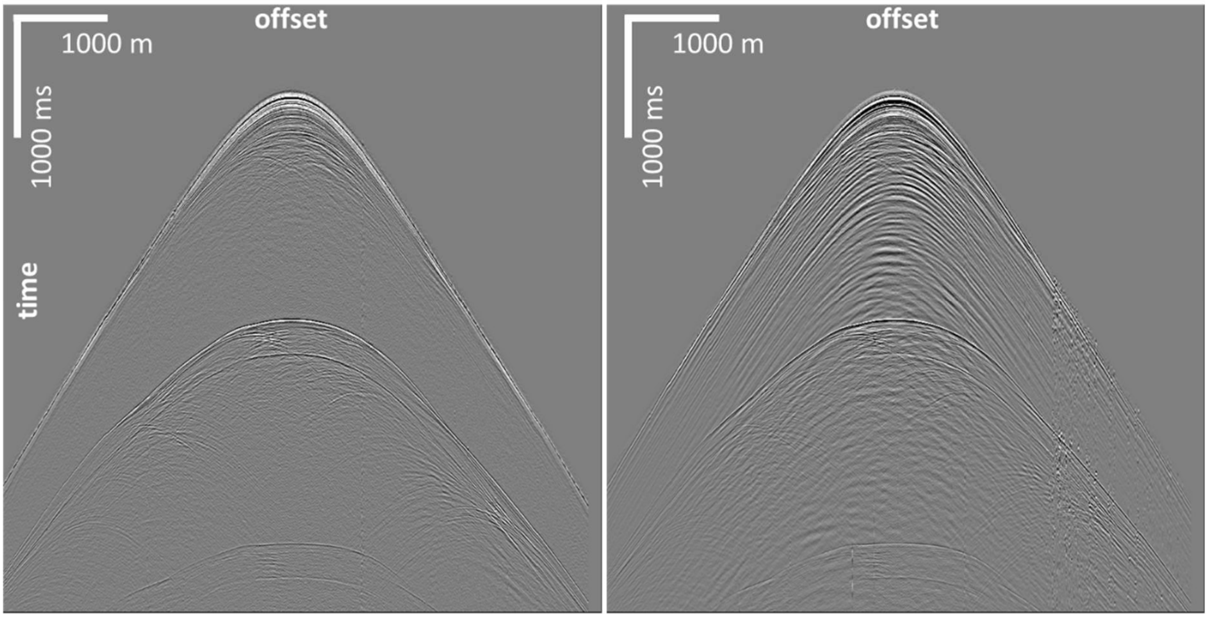

Figure 2 *Input pressure (left) and Z component (right). Notice the different "structural" character of the shear-wave with respect to the signal.*

Additionally, the noise estimate can be obtained and used in an adaptive subtraction step to further optimize the denoising of the input *Z*-component. Here, we output directly the denoised *Z*-component. The domain in which the local attributes are estimated plays an important role. In the space-time domain, signal and noise have a higher level of overlap than in the Radon domain. The Radon domain further localizes the analysis without implementing more computationally heavy transformations (e.g., curvelets). However, the proposed procedure is only partly effective if signal and noise are overlapping in the transformed Radon domain. A careful choice of the thresholding prevents signal damage at the cost of reduced noise attenuation.

Examples

We test the clustering-based Z noise attenuation approach on the Cascadia dataset: an academic Ocean Bottom Node dataset acquired offshore western Canada. The water bottom is around 1300m depth, which leads to a clear temporal separation between the direct arrival and primary reflections package and their surface-related multiples. The Z-component is strongly contaminated by shear-wave noise. Figure 2 shows the pressure and Z-component recorded at the first node location: notice the very consistent moveout of the shear-wave noise along the record and the similarity with the reflection signal, which makes the denoising problem particularly challenging. Figure 3 shows the comparison between the input Z-component and the denoising result. The clustering-based denoising is effective in attenuating the strong coherent noise between the direct arrival and the first water bottom multiples. However, some leakage of diffracted multiples can be observed in the subtracted noise (Figure 3, right).

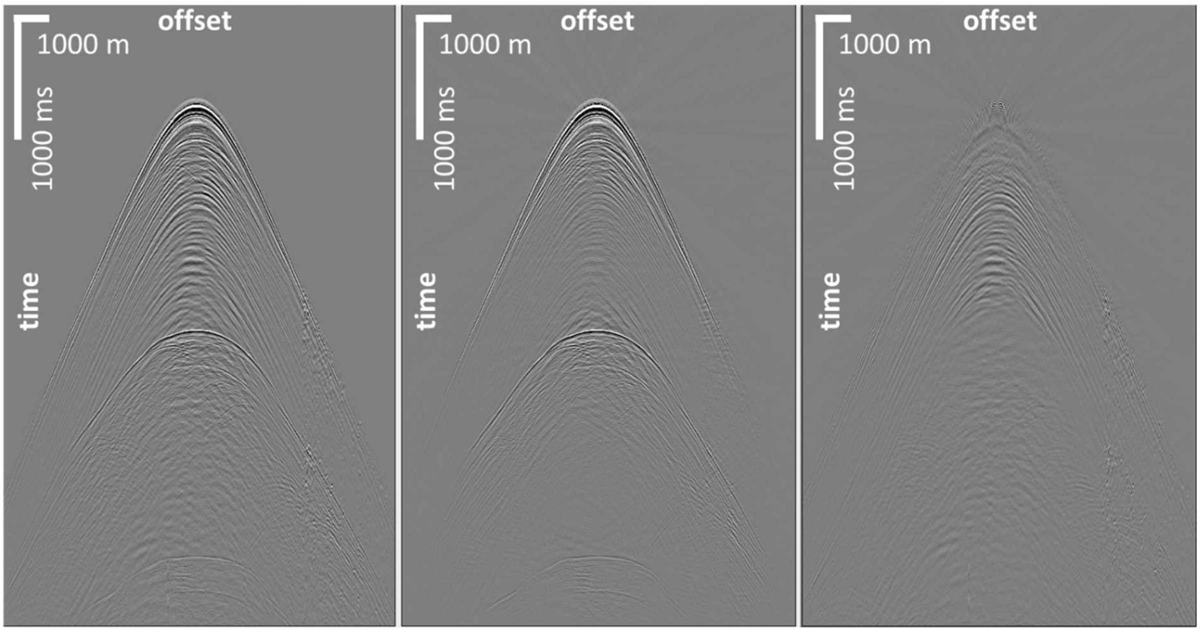

Figure 3 *Input (left), denoised Z component (centre) and noise (right). The strong coherent noise that contaminates the gather between the direct arrival and its surface related multiple is largely attenuated.*

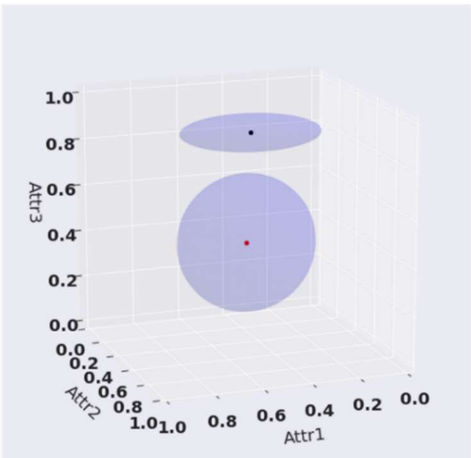

Figure 4 *Centers of the clusters estimated via GMM for the example in Figure 3. The widths of the ellipsoids are the square root of the diagonal elements of the covariance matrices. Notice the clear separation of the clusters. The red point is closer to the origin and represent the noise cluster.*

Figure 4 refers to the denoising example in Figure 3 and shows the centres of the clusters with the 1-standard-deviation-wide ellipsoid superimposed to indicate the size of the clusters. The axes represent the estimated local attributes. Notice the separation between the two clusters and how, at least for this example, the noise cluster (centred around the red dot) is more circularly symmetrical than the signal cluster (centred around the black dot).

Conclusions

We discuss a noise identification and attenuation strategy based on local attributes estimated in the linear Radon domain. We estimate local attributes in Radon domain and then perform clustering to identify the points that are more likely to contain noise components. The estimated Gaussian Mixture Model (GMM) naturally supplies the probability that each Radon component represents signal rather than noise. We extract a noise estimate using the GMM clusters and then we attenuate the noise in the Radon domain using a soft thresholding strategy. The number of attributes can be increased without changing the algorithm at the cost of a more computationally expensive clustering step. However, operating in the linear Radon domain limits the amount of separation that can be achieved with respect to sparser domains, like complex curvelets; once again the trade-off is between computational cost and accuracy. Real data examples show the effectiveness of the methodology.

Acknowledgements

The authors thank Ken Smith for the useful discussion about clustering and Shearwater GeoServices for permission to publish this work. The Cascadia data were collected with support from the Natural Science and Engineering Research Council and the Geological Survey of Canada.

References

Amundsen, L. [2001] Elimination of free-surface related multiples without the need of the source wavelet. *Geophysics*, **66**(1), 327-341.
Craft, K., L. and Paffenholz, J. [2007] Geophone noise attenuation and wave-field separation using multi-dimensional decomposition technique. *77th SEG Annual Meeting*, Expanded Abstracts, 2630-2634.
Dash, R., Spence, G., Hyndman, R., Grion, S. Wang, Y. and Ronen, S. [2009] Wide-area imaging from OBS multiples. *Geophysics*, **74**(6), Q41-Q47.
Fomel, S. [2007] Local seismic attributes. *Geophysics*, **72**(3), A29-A33.
Hale, D. [2006] Fast local cross-correlations of images. *76th SEG Annual Meeting*, Expanded Abstracts, 3160-3164.
Naeini, E. Z., L., Baboulaz, L. and Grion, S. [2011] Enhanced wavefield separation of OBS data. *73rd EAGE Conference and Exhibition*, Extended Abstracts, B003.
Paffenholz, J., Shurtleff, R., Hays, D. and Doherty, P. [2006] Shear Wave Noise on OBS Vz Data – Part I Evidence from Field Data. *68th EAGE Conference and Exhibition*, Extended Abstracts, B046.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. [2011] Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, **12**, 2825-2830.
Peng, C., Jin, H. and Wang, P. [2014] Noise attenuation for multi-sensor streamer data via cooperative de-noising. *84th SEG Annual Meeting*, Expanded Abstracts, 1878-1882.
Press, W. H., Teukolsky, S. A., Vetterling, W.T., Flannery, B. P. [2007] *Numerical Recipes: The art of scientific computing*. Cambridge University Press, New York.
Poole, G., Casasanta, L. and Grion, S. [2012] Sparse τ-p Z-noise attenuation for ocean-bottom data. *82nd SEG Annual Meeting*, Expanded Abstracts, 1-5.
Stork, C. [2017] Removing complex land noise with modern pattern recognition using Machine Learning. *87th SEG Annual Meeting*, Expanded Abstracts, 6089-6093.
Turquais, P., Asgedom, E. G. and Söllner W. [2017] Coherent noise suppression by learning and analyzing the morphology of the data. *Geophysics*, **82**(6), V397-V441.
van Vliet, L. J. and Verbeek, P.W. [1995] Estimators for orientation and anisotropy in digitized images. *ASCI'95*, Proceedings of the first Annual Conference of the Advanced School for Computing and Imaging, 442-450.
Yu, Z., Kumar, C. and Ahmed [2011] Ocean bottom seismic noise attenuation using local attribute matching filter. *81st SEG Annual Meeting*, Expanded Abstracts, 3586-3590.