Quality control of 3D prestack land seismic data with a focus on data enhancement

Andrey Bakulin, Ilya Silvestrov, EXPEC Advanced Research Center, Saudi Aramco; Maxim Protasov, Institute of Petroleum Geology and Geophysics, Novosibirsk, Russia*

Summary

We present an approach for quality control (QC) of massive prestack land seismic datasets, focusing on evaluating data enhancement procedures critical for processing success. The QC scheme contains such measures as coherency, signal-to-noise ratio, frequency content, and phase characteristics. We apply this to single-sensor data from the desert environment. We show that proposed metrics can quantitatively measure challenging data quality even when the signal remains mostly invisible. We further demonstrate similar analysis after data enhancement emphasizing their role in the data processing. The proposed QC method allows efficient evaluation of seismic acquisition and processing.

Introduction

Field acquisition requires robust quality control and assessment of the recorded data. Likewise, each step of the seismic processing should be assessed using similar quality control diagnostics of some sort. Typical quality control may include displays of such data attributes as amplitude spectra and autocorrelograms (Yilmaz, 2001). QC gives an understanding of how a particular processing step affects signal and noise behavior. In general, visualization of seismograms and amplitude spectrum (during acquisition or before and after each processing step) is necessary for standard QC. A row of geophysical papers presents workflows for data QC during acquisition and processing. Nateganov et al. (2018) introduce several complex quality criteria for analyzing data processing results. In the sense of data quality, the most interesting metric is "signal quality." It consists of three parameters: resolution, bandwidth index (BWI), and signal-to-noise ratio (SNR).

Many standard data QC approaches become ineffective when dealing with land seismic data acquired with small arrays and single sensors, especially from the desert environment. Often, reflected waves are invisible on prestack gathers and may remain heavily disturbed even throughout the processing stages. Therefore, we need a robust approach and metrics that can identify signal quality despite these challenges. Also, such data require powerful enhancement approaches such as supergrouping and nonlinear beamforming (Bakulin et al., 2018, 2020). Quantitative and automated comparison of the data quality before and after data enhancement could greatly assist the processing and lessen the human bias. We propose a robust, practical approach to QC any prestack data, including single-sensor data from the most challenging desert environment. We demonstrate application on progressively larger subsets from an ensemble of traces to prestack gathers to 2D or 3D volumes.

We use semblance (Neidell and Tuner, 1971) as a coherency measure. The most critical metric for seismic data quality control is the SNR. An SNR is typically defined as the ratio of signal power to noise power. Several different SNR computation algorithms were proposed (Liu and Li, 1997; Belousov, 2011). Among these algorithms, the most practically interesting SNR computation is based on estimating signal by stacking coherent events after moveout corrections (Liu and Li, 1997). After analyzing several algorithms using controlled SNR experiments, we concluded that the semblance-based approach provides the most robust SNR measurement for challenging land seismic data with a relatively low SNR. Also, to better characterize signal presence and evolution during processing, we propose to use standard amplitude spectral metrics (Belousov, 2011) along with new phase metrics introduced in this study.

QC measures

Let us assume that seismic data is the superposition of signal and noise:

$$d_{ij} = s_{ij} + n_{ij}, \;\; i = 1, \dots, N, \;\; j = 1, \dots, M, \qquad (1)$$

where $d_{ij} = d(t_i, x_j)$ is recorded seismic data, $s_{ij} = s(t_i, x_j)$ is signal, and $n_{ij} = n(t_i, x_j)$ is noise, N is the number of time samples, and M is the number of traces in the data.

Coherency

Coherency is the most straightforward indicator of the presence of signal presence. The evolution of this metric during processing can be very insightful. A well-known coherency measure is a semblance (Neidell and Tuner, 1971)

$$Semblance = \frac{\sum_{i=1}^{N}\left(\sum_{j=1}^{M} d_{ij}\right)^2}{M\sum_{i=1}^{N}\sum_{j=1}^{M} d_{ij}^2}. \qquad (2)$$

It only requires raw data as an input and can be measured in varying space-time windows or subsets of the data.

Signal-to-noise ratio (SNR)

Definition of SNR is the ratio of signal power to noise power. SNR can be formulated in the logarithmic scale as follows:

$$SNR = 10\, log_{10}\left(\sum_{i=1}^{N}\sum_{j=1}^{M} s_{ij}^2 \Big/ \sum_{i=1}^{N}\sum_{j=1}^{M} n_{ij}^2\right). \qquad (3)$$

Assuming that signals contained in a set of windowed seismic records are linearly correlated, SNR can be computed only using raw data as the only input (Liu and Li, 1997):

$$SNR_{stack} = 10\, log_{10} \frac{\frac{1}{M}\sum_{i=1}^{N}\left(\sum_{j=1}^{M} d_{ij}\right)^2}{\left(\sum_{i=1}^{N}\sum_{j=1}^{M} d_{ij}^2 - \frac{1}{M}\sum_{i=1}^{N}\left(\sum_{j=1}^{M} d_{ij}\right)^2\right)}. \qquad (4)$$

This formula relies on signal estimate coming from stacking coherent events. Taking into account the definition of the semblance, we arrive at SNR representation via semblance, as can be seen from the original paper by Neidell and Taner (1971):

$$SNR_{stack} = 10\, log_{10} \frac{Semblance}{1-Semblance}. \qquad (5)$$

Spectral measures
One of the most critical measures of the data is its spectral content. We propose to track energy partitioning between low $E_{low}(f_1, f_2)$, medium $E_{mid}(f_2, f_3)$, and high-frequency $E_{high}(f_3, f_4)$ ranges, where frequencies f_i defines the bounds of each range. Each energy is expressed as a percentage of overall amplitude energy ($E_{low} + E_{mid} + E_{high} = 1$) occupied by a predefined frequency range.

Finally, severe phase distortions observed in the data from the desert environment leads to a phase stability metric STD_{phase} that describes the standard deviation of phase angle at a fixed frequency. Here, phase angle is defined at frequency f as:

$$phase(f) = angle(u(f)) \cdot {}^{180}/_{\pi}. \qquad (6)$$

If $STD_{phase} = 0$, then the phase is constant across the selected window. In contrast, for the varying phase, we obtain a non-zero value from 0 to 360 degrees. Since STD_{phase} is frequency-dependent, in the QC process, we propose to select some essential frequencies, such as central and maximum frequencies (Belousov, 2011).

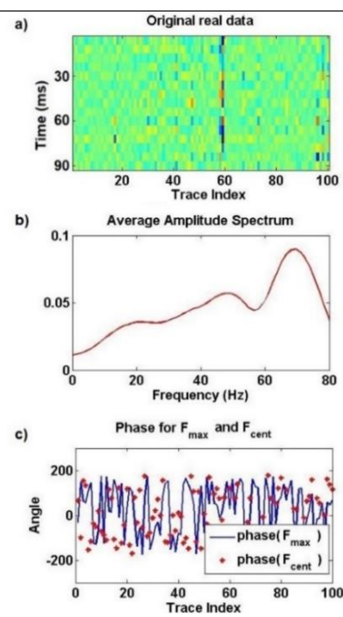

Figure 1: Original real data: (a) ensemble of traces after NMO; (b) average amplitude spectrum; (c) phase at maximum and central frequencies computed for every trace. Here $STD_{phase}(F_{cent}) = 100^0$.

QC of a single ensemble of traces
Let us demonstrate these metrics' usefulness on an extremely challenging single-sensor field dataset from the desert environment, which already passed through a standard processing sequence including conventional land noise attenuation techniques. Initially, let us select an ensemble of 100 traces after NMO. We further subselect a 90 ms time that is known to contain the reflected event. We compare the results for the original seismic data (Figure 1) and the data after enhancement (Figure 2) with nonlinear beamforming or NLBF (Bakulin et al., 2020) performed in cross-spread domain. NLBF is a data-driven approach for enhancing prestack data that utilizes a form of local stacking. The reflected event remains mostly invisible behind the remained noise carpet in the original data (Figure 1a). The amplitude spectrum shows a suspicious peak at 70 Hz (Figure 1b). At central and maximum frequency, phase values show a vast spread within the ensemble reaching around ±100 degrees. After applying NLBF, the reflected event becomes visible although suffering from residual static variation (Figure 2a). Amplitude spectra become skewed towards lower frequencies, and higher frequencies are greatly reduced (Figure 2c). Phase angle variations are significantly diminished (Figure 2c), resulting in a smaller standard deviation STD_{phase}. This suggests that enhanced records display signs of the signal as opposed to the original traces overwhelmed by noise.

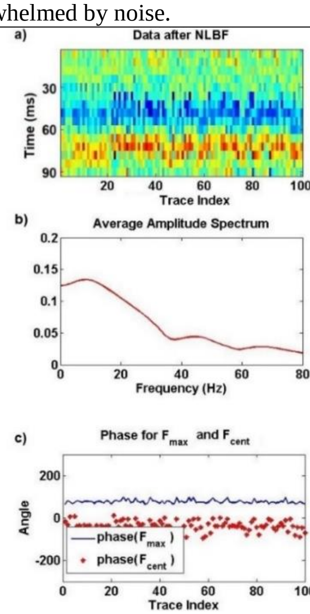

Figure 2: Real data after nonlinear beamforming: (a) ensemble of traces after NMO; (b) average amplitude spectrum; (c) phase at maximum and central frequencies computed for every trace. Here $STD_{phase}(F_{cent}) = 30^0$.

10.1190/segam2021-3576612.1

QC of prestack gathers

In prestack gather, all data QC attributes vary with time and offset. So we compute time and offset-variant semblance/SNR along the gather using local ensembles centered around output point and applying moveout corrections to flatten the events. Figures 3 and 4 show such computations for data before and after NLBF enhancement. The crucial parameter for reliable SNR computation is the number of traces. Our investigations show that 100 traces provide reliable SNR estimation only down to -17 dB. Therefore, we use larger local ensembles of around 10,000 traces allowing reliable SNR measurement down to -40 dB. We call this a minimum required ensemble size for reliable SNR estimation. One can observe that typical SNR values for data before NLBF are between -20 dB and -30 dB. In contrast, after NLBF, SNR values are between 0 dB and -10 dB. We conclude that the semblance-based approach provides the most robust SNR estimate that remains accurate down to -40 dB and even lower during our investigation. These displays show typical behavior of QC metrics versus offset and time, allowing to evaluate data quality for acquisition and processing. Despite pre-stack events being invisible on the data before NLBF — these metrics computed for properly selected ensembles can be trusted in absolute values for the quantitative analysis of the data.

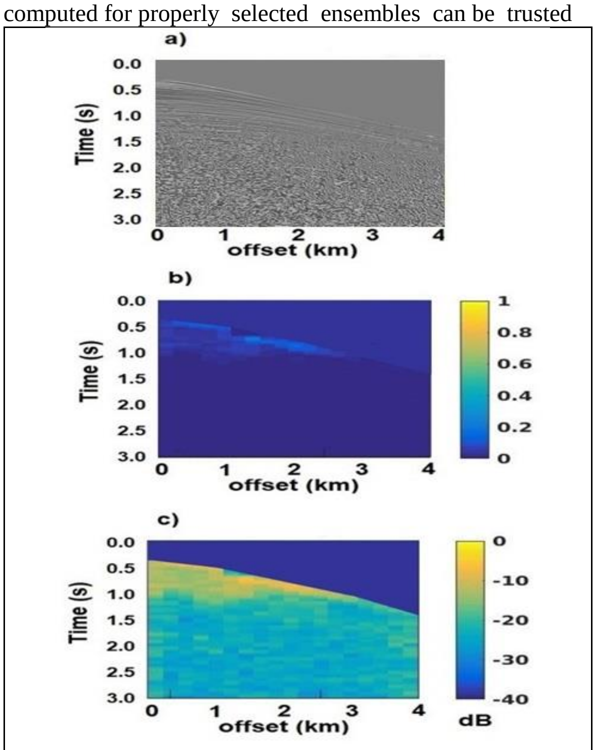

Figure 3: (a) Part of a cross-spread gather (traces for a single source sorted by offset) for original single-sensor land data after typical pre-processing; (b) semblance computed in a moving window fashion for prestack gather from (a); (c) prestack SNR derived from semblance using equation (4).

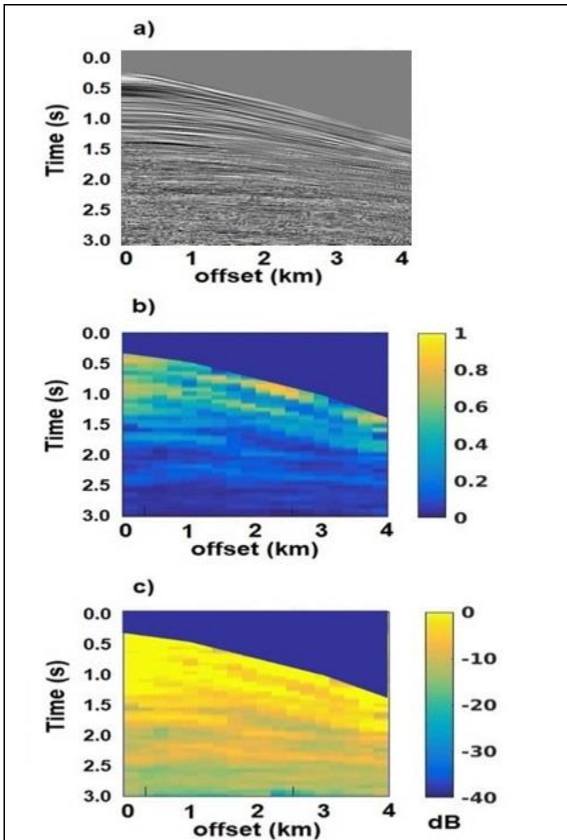

Figure 4: Same as Figure 3 but for gather after enhancement with nonlinear beamforming. Observe higher visibility of prestack events on (a). This is further confirmed by higher values of semblance (b) and SNR (c), with the average improvement of ~0.2 for semblance and ~15 dB for SNR.

QC of 2D and 3D volumes

When we go to the next scale of analyzing a 2D or 3D volume, it is convenient to compute one averaged QC metric per single gather. While this attribute still characterizes prestack data, it is now an averaged quantity describing all offsets. It is further convenient to select a target time window for analysis so resulting metrics can be displayed as a map (for 3D) or graphs (for 2D). Such maps can be visualized for different target windows or as time-dependent 3D volumes similar to gather QC. Let us demonstrate this for a single 2D subline out of 3D data. In this example, the 2D line consists of 456 cross-spread gathers. Every gather contains from 150,000 up to 300,000 traces. Likewise, we analyze all proposed QC measures for the original real data and the data after nonlinear beamforming. Looking at the QC metric for original real data (Figure 5), we see that noise is dominating.

10.1190/segam2021-3576612.1

SNR has low values varying between -20 dB and -40 dB. Coherency is hovering just above zero. We stress that when ensemble size increases above the minimum required size, SNR and coherency values should not depend on the ensemble size provided the noise level remains similar. High-frequency content is prevalent due to intense high-frequency noise. The standard deviation of the phase suggests an extreme level of distortions. Despite severe challenges, NLBF substantially increases coherency values and SNR (Figure 6a,b). Also, medium frequencies dominate the energy spectrum after NLBF (Figure 6c). The standard deviation of the phase is reduced, indicating that recovery of the signal phase has started. A bigger summation aperture inside NLBF leads to all QC metrics improving: coherency and SNR rise, medium frequencies dominate even more, while phase standard deviation becomes lower. Such behavior suggests that higher frequencies are dominated by the noise that should be suppressed. QC metrics inform us that NLBF overall is doing a relatively good job. Before NLBF, the metrics' distribution along the line remains highly variable, reflecting actual acquisition conditions with rapidly changing surface noises of various origins. NLBF effectively suppresses the strongest noise raising the overall SNR level and leading to less lateral variability of the metrics along the 2D line. This makes good sense for flat-lying geology, where a signal is expected laterally consistent while the distribution of surface noise is highly irregular.

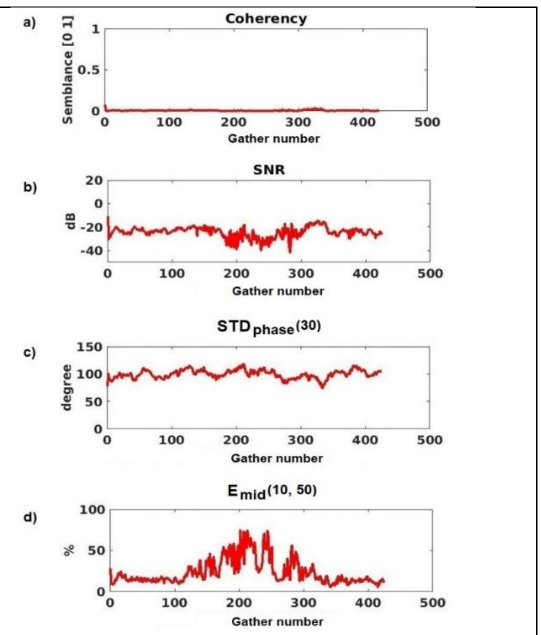

Figure 5: Computed QC measures for the original real data: (a) coherency, (b) SNR, (c) standard deviation of phase at phase deviation at the middle frequency 30 Hz, (d) percentage of overall energy occupied by medium frequencies (low: 2-10, mid:10-50 Hz; high: 50-80 Hz).

Conclusions

We present a robust, practical scheme for QC of large amounts of prestack land seismic data. The scheme uses the following QC metrics: coherency, SNR, energy partitioning between low, medium, and higher frequencies content, and phase stability metrics (also for low, medium, and high frequencies). We apply the proposed scheme to challenging single-sensor seismic data from the desert environment. We show that the number of analyzed traces is an essential parameter for reliable SNR estimation. We quantify prestack SNR after linear noise removal to be between -20 to -30 dB. Such a ratio is too low for either reliable conventional QC or further processing. We demonstrate that nonlinear beamforming can effectively enhance the data raising coherency and SNR, suppressing energetic high-frequency noise. This leads to the recovery of less contaminated data, especially at low and medium frequencies. Finally, the proposed QC tool quantitatively shows that NLBF enables efficient processing of challenging modern 3D high-channel-count and single-sensor data. These improvements lead to better land seismic imaging and reservoir characterization.

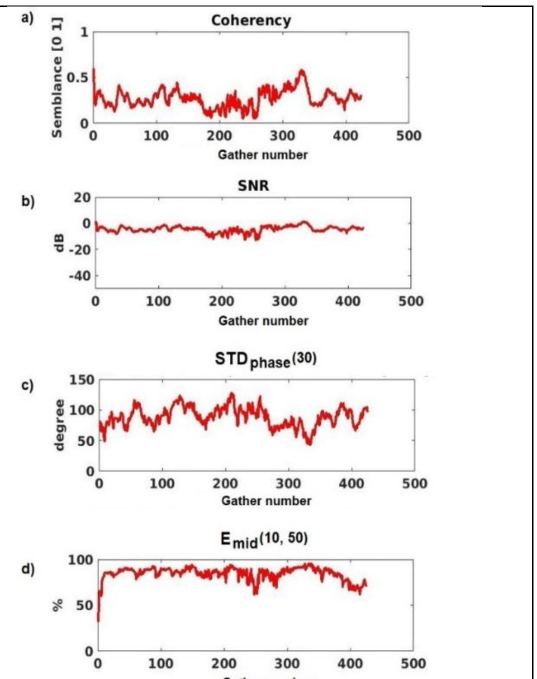

Figure 6: Same as Figure 6 but for data after enhancement with NLBF.

REFERENCES

Bakulin, A., I. Silvestrov, M. Dmitriev, D. Neklyudov, M. Protasov, K. Gadylshin, and V. Dolgov, 2020, Nonlinear beamforming for enhancement of 3D prestack land seismic data: Geophysics, **85**, no. 3, V283–V296, doi: https://doi.org/10.1190/geo2019-0341.1.

Bakulin, A., P. Golikov, M. Dmitriev, D. Neklyudov, P. Leger, and V. Dolgov, 2018, Application of supergrouping to enhance 3D prestack seismic data from a desert environment: The Leading Edge, **37**, 200–207, doi: https://doi.org/10.1190/tle37030200.1.

Belousov, A., 2011, Standard evaluation of the quality of field seismic material: Devices and Systems of Exploration Geophysics, **03**, 31–36 (In Russian).

Liu, Y., and C. Li, 1997, Some methods for estimating the signal/noise ratio of seismic data: Oil Geophysical Prospecting, **32**, 257–262 (In Chinese with English abstract).

Nateganov, A., A. Popov, and B. Pagliccia, Quantitative quality control: a tool for seismic data processing monitoring and comparison: 80th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: https://doi.org/10.3997/2214-4609.201800239.

Neidell, N., and M. Taner, 1971, Semblance and other coherency measures for multichannel data: Geophysics, **36**, 482–497, doi: https://doi.org/10.1190/1.1440186.

Yilmaz, O., 2001, Seismic data analysis: Society of Exploration Geophysicists.