Geophysical Prospecting EAGE EUROPEAN ASSOCIATION OF GEOSCIENTISTS & ENGINEERS
Geophysical Prospecting, 2022, 70, 629–638 doi: 10.1111/1365-2478.13178

Research Note: Signal-to-noise ratio computation for challenging land single-sensor seismic data

Andrey Bakulin[1], Ilya Silvestrov[1] and Maxim Protasov[2]*

[1] *EXPEC Advanced Research Center, Saudi Aramco, Dhahran, Saudi Arabia, and* [2] *Institute of Petroleum Geology and Geophysics, Siberian Branch of The Russian Academy of Sciences, Novosibirsk, Russia*

Received July 2021, revision accepted December 2021

ABSTRACT

This study explores practical methods of estimating the signal-to-noise ratio of challenging seismic data containing a signal too small to be identified visually, as often occurring in a desert environment with single-sensor surveys. As a result, the estimation can only be performed using an ensemble of traces. We compare several data-driven estimation approaches and reveal their practical limits using a controlled experiment. We identify a stacking-based method as the most robust with the broadest range of applicability. We relate the number of traces in the ensemble to the lowest reliably estimated absolute signal-to-noise ratio possible with this method using such an ensemble. We support the findings using synthetic and single-sensor field data with low signal-to-noise ratios down to -60 dB. The proposed methodology allows reliable data-driven estimation of very low absolute signal-to-noise ratios directly from pre-stack seismic data.

Key words: Acquisition, noise, signal processing.

INTRODUCTION

The signal-to-noise ratio (SNR) is an essential characteristic of seismic data. It is useful during data acquisition to measure the quality of recordings, relate them to the acquisition parameters, as well as compare different geological areas. SNR is also helpful for evaluating a processing flow during data analysis, tuning its variable parameters and diagnosing problematic areas that require additional improvement and refinement. This is especially important for modern dense single-sensors acquisitions recording petabytes of data with low SNR, which quality cannot be assessed and controlled manually. Achieving high SNR post-stack and pre-stack is a requirement for successful seismic interpretation and inversion, respectively. SNR is typically defined as a ratio of signal energy to noise energy. A definition of the signal and the noise can vary depending on the application. During the acquisition phase, a window before first arrivals may be used as an ambient noise representation, whereas data after first arrivals are often considered as the signal. This allows control of the reliability of the excitation and recording systems. However, it does not provide any information about the quality of the recorded target reflected events. During the processing phase, the processed data with reflections revealed from background noise can be considered the signal, while the removed energy is the noise. The overall objective of the processing is to increase SNR by successively removing various noises and restoring the signals. However, processing progress is often assessed using subjective human judgment instead of definitive data-driven metrics. Quantitative SNR estimates can be helpful to compare different noise attenuation and signal enhancement algorithms and adjust their parameters. However, we demonstrate that most approaches do not usually provide an actual absolute SNR of the data but rather a relative metric showing improvement in the SNR. In contrast, seismic interpretation and inversion rely on reaching a certain absolute SNR for them to succeed. Indeed, the actual SNR of some target reflected wave is an intrinsic and definitive property of the data itself. Knowing this

*E-mail: protasovmi@ipgg.sbras.ru

true SNR value provides valuable insights into the actual data quality. These insights can lead to the development of new acquisition designs and directly identify its impact on the effectiveness of the current state-of-the-art data processing flow and the development of new novel seismic processing algorithms.

In the past, different algorithms for SNR computation were proposed based on statistical properties of the data and some common assumptions on signal and noise (Hatton *et al.*, 1986; Liu & Li, 1997; Belousov, 2011). The classical approach is based on a normalized cross-correlation of two traces known as a coefficient of coherency (Foster & Guinzy, 1967; Bosworth *et al.*, 2008). Although this coefficient can theoretically give an exact SNR value, its direct usage based on only a pair of traces is inaccurate. Instead, more statistically reliable estimates utilizing an averaging over an ensemble of traces are used in practice (Belousov, 2011). Another approach for SNR computation is based on stacking the traces themselves rather than their normalized cross-correlations. A well-known semblance formula also employs a similar summation. Although semblance is usually used in many seismic applications as a coherency measure, it is equal to the ratio of signal energy to total energy over a selected window, as Neidell and Taner (1971) show. Thus, the SNR value can be easily derived from semblance and vice versa. The third approach, often used in seismic data processing, is based on singular value decomposition (Key *et al.*, 1987; Chen & Fu, 1993; Zhao *et al.*, 2019). In this method, the SNR estimate relies on the approximation that coherent signal energy is concentrated in the first singular value of a data matrix, whereas incoherent noise energy is uniformly distributed within all singular values.

The approaches mentioned above for SNR estimation use the same assumptions such as constant signal and noise levels over an ensemble of traces, zero mean of noise, and its statistical independence from trace to trace and from the signal. For moderate and high SNR, all these methods tend to produce similar results. However, their performance and the estimated levels differ significantly for more challenging data with extremely weak pre-stack signals and low SNR. Examples of such data are modern high-channel-count and single-sensors datasets acquired in desert environments without large traditional source and receiver arrays where SNR can go down to −40 dB and less (Bakulin *et al.*, 2020; Cordery, 2020). In addition, the estimated SNR often exhibits a dependency on the number of traces used in the analysis even when all the assumptions about statistical properties of noise and signal are met and hence does not provide an actual SNR level. In this study, we illustrate these effects in more detail. We briefly compare different algorithms for SNR computation and show the advantages of the stacking-based approach at low SNR. Also, we demonstrate how the SNR estimate of the stacking method may approach actual SNR when an appropriate number of traces is selected for the estimation ensemble. Based on these relationships, we propose a simple, practical recipe that can estimate the actual absolute SNR values of noisy seismic data acquired without large source and receiver arrays in the field.

THEORY

Let us consider a seismic data window after moveout corrections and assume that it can be represented as a superposition of signal and noise:

$$d_{ij} = s_i + n_{ij},\ i = 1, \ldots, N,\ j = 1, \ldots, M, \tag{1}$$

where $d_{ij} = d(t_i, x_j)$ is a noisy seismic trace with time index i and trace index j, $s_i = s(t_i)$ is a signal, which is common for all traces and $n_{ij} = n(t_i, x_j)$ is a noise that varies from trace to trace. In this study, the signal means a flattened reflection event. In contrast, all the remaining interfering seismic energy and the signal distortions are treated as noise. We use common assumptions that the noise has zero mean, it is statistically independent from trace to trace and from the signal and its energy is constant over an ensemble of traces. According to a standard definition, the signal-to-noise ratio (SNR) in the given window is calculated as the ratio of signal energy to noise energy and can be estimated as follows:

$$\mathrm{SNR} = \frac{M \sum_{i=1}^{N} s_i^2}{\sum_{j=1}^{M} \sum_{i=1}^{N} n_{ij}^2}, \tag{2}$$

or in decibel scale:

$$\mathrm{SNR}^{dB} = 10\log_{10}(\mathrm{SNR}). \tag{3}$$

Under the assumption that the noise energy is constant from trace to trace, Equation (2) describes the SNR for every trace in the seismic data (1).

In practice, signal and noise are unknown, and rarely can they be completely separated. An ensemble of traces is invoked to perform a data-driven SNR estimation in practice to overcome this limitation. In this study, several data-driven methods for SNR calculation are considered and contrasted. A brief description is provided for each method, and their applicability to low SNR environments is assessed.

When we introduce a zero-lag cross-correlation of two traces:

$$R_{d_k d_l} = \sum_{i=1}^{N} d_{ik} d_{il}, \tag{4}$$

an averaged normalized cross-correlation can be written as follows (Neidell & Taner, 1971; Hatton, Worthington & Makin, 1986):

$$\gamma = \frac{2}{(M-1)\,M} \sum_{k=1}^{M} \sum_{l=k+1}^{M} \frac{R_{d_k d_l}}{\sqrt{R_{d_k d_k}}\sqrt{R_{d_l d_l}}}. \tag{5}$$

If we assume that the noise has the same energy on all traces, and also that its cross-correlation between different traces and between signals is zero, SNR can be calculated as

$$\mathrm{SNR}_{\mathrm{cor}} = \frac{\gamma}{1-\gamma}. \tag{6}$$

This equation is similar to the one used by Belousov *et al.* (2011). However, the SNR calculation is done after the averaging of the cross-correlations rather than before. This gives more stable results when applied to very noisy data.

Analogously, SNR based on stacking method can be estimated as (Liu & Li, 1997):

$$\mathrm{SNR}_{\mathrm{stack}} = \frac{S}{1-S}, \tag{7}$$

where semblance S is defined as the normalized ratio of stacked output and raw input signals or equivalently expressed in terms of cross-correlations as (Neidell & Taner, 1971):

$$S = \frac{\sum_{i=1}^{N} \left(\sum_{j=1}^{M} d_{ij}\right)^2}{M \sum_{i=1}^{N} \sum_{j=1}^{M} d_{ij}^2} = \frac{\sum_{k=1}^{M} \sum_{l=1}^{M} R_{d_k d_l}}{M \sum_{l=1}^{M} R_{d_l d_l}}. \tag{8}$$

Semblance is a popular coherency measure for stacking velocity analysis and other seismic applications. Neidell and Taner (1971) analysed its properties for coherency analysis. They showed that semblance is superior to the averaged normalized cross-correlation coefficient (5) for this application. They also demonstrated that the semblance coefficient is equal to a ratio of signal energy to total energy under the assumption that the noise sum over all channels at any time is zero, thus leading to Equation (7) (see also the Appendix).

The SNR estimation method based on singular value decomposition (Key *et al.*, 1987; Chen & Fu, 1993) uses similar statistical assumptions for signal and noise as before and is based on the assumption that all the signal energy is concentrated in the major singular value of the data matrix d_{ij}, whereas noise is uniformly distributed and represented by all singular values. Approximating noise energy by the difference between data energy and signal energy as in Equations (6) and (7), we arrive at SNR estimate for the SVD-based method (Liu & Li, 1997):

$$\mathrm{SNR}_{svd} = \frac{\sigma_1^2 - \sum_{j=2}^{M} \sigma_j^2 / (M-1)}{M \cdot \sum_{j=2}^{M} \sigma_j^2 / (M-1)}. \tag{9}$$

It is worth stressing that all three Equations (6), (7), and (9) for estimation of the SNR value are purely data-driven. Each method uses only noisy seismic data as an input and estimates the actual SNR in the considered data window.

NUMERICAL RESULTS

Controlled experiment

Let us compare three methods using controlled experiments in which the data's signal-to-noise ratio (SNR) is known. To illustrate the difficulties one can face when applying the described methods to noisy data, we start with a trivial and reproducible example. Figure 1(a) shows a window of synthetic data after moveout correction containing pure signal according to Equation (1). An example of white Gaussian noise is shown in Fig. 1(b). Figures 1(c) and (d) show examples of data after the noise has been added to the signal with SNR values of 10 dB and 0 dB, respectively. We generate a number of datasets with different SNR values varying from −60 dB to 20 dB and apply Equations (6), (7), and (9) for data-driven SNR estimation (Fig. 2). When the signal level in the data is high, all the methods provide reliable estimations of the actual SNR values. However, the results deviate from the true values when the signal falls below the noise floor. For example, when 100 traces are used in the analysis (Fig. 2a), the lowest reliably estimated SNR values vary from −12 dB to −18 dB, depending on the method. This can be improved by using more traces, as shown in Fig. 2(b), where their total number is increased to 5000. In this case, we were able to estimate even lower SNR values than before reliably. However, a limit of applicability of all methods still exists. With 5000 traces in use, the lowest correctly estimated SNR is −38 dB. This shows that even in an ideal situation when all assumptions considered in the methods are fulfilled, some limitations of the above methods do not allow the estimation of the actual absolute SNR in cases of very noisy data. In all cases, the stacking-based method (7) provides the best results, which allows for getting true SNR values for more noisy data.

Real noise in pre-stack data is different from the white Gaussian noise used in the first example. To test the SNR estimation methods in more realistic settings, we use as

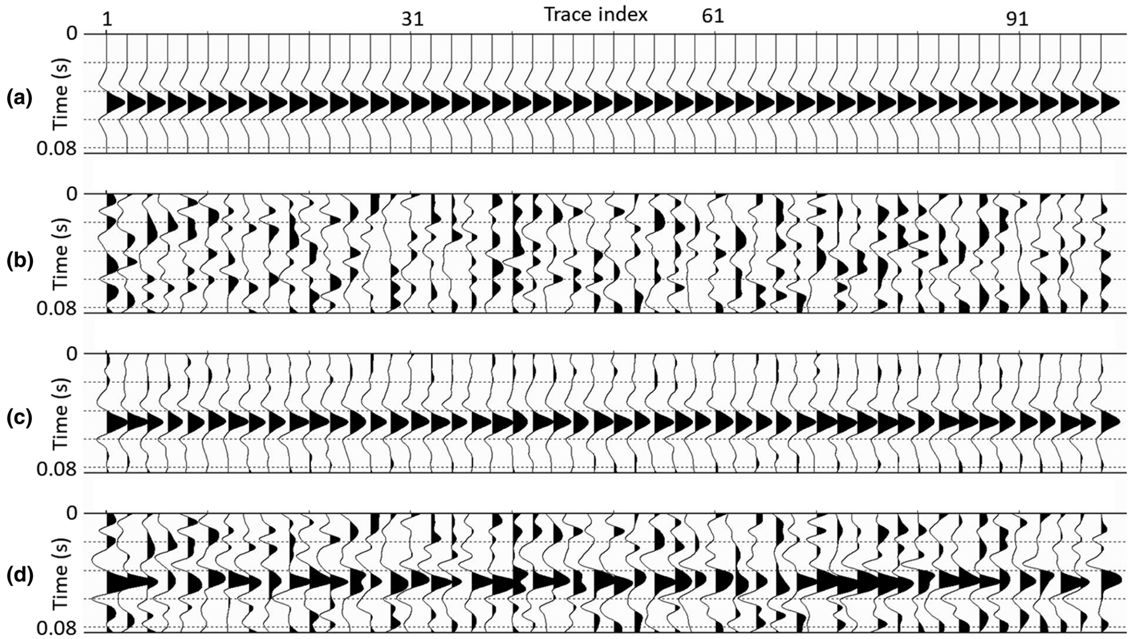

Figure 1 (a) Ideal coherent synthetic signal. (b) Random noise with Gaussian distribution. Noisy data is simulated as a superposition of synthetic signal and random noise with actual SNR of (c) 10 dB and (d) 0 dB.

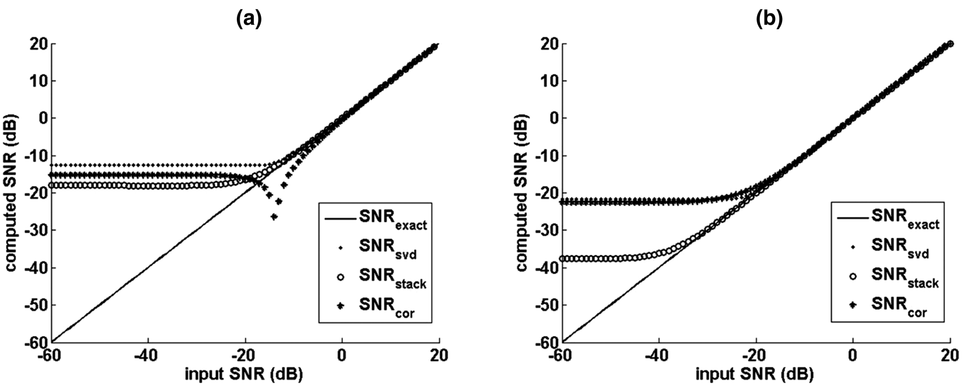

Figure 2 Estimated versus true signal-to-noise ratios for simulated synthetic data shown in Fig. 1(a) with a controlled level of synthetic random noise: (a) data consist of 100 traces; (b) data consist of 5000 traces. Each estimation method is shown with a different line: solid – ground-truth value computed with the exact formula (2); stars – correlation-based formula (6); circles – semblance-based formula (7); dots – SVD-based formula (9).

the noise a late-time portion of seismic data (starting from 5 seconds) taken from field single-sensor seismic records acquired in an arid environment (Pecholcs *et al*., 2012; Cordery, 2020) after one of the intermediate processing stages (Fig. 3a). Figure 4 shows the estimated versus true SNR results when such noise was added to the synthetic signal from Fig. 1(a) with different energy levels. As before, we observe that each method has its own limits of applicability, which improves when more traces are used in the analysis. The singular value decomposition-based method showed the worst result in this case and correctly identified the SNR values only above −5 dB. The other two methods show more reliable results. However, the stacking-based Equation (7) still gives a more accurate and more stable estimation for wider ranges of low-level SNR values.

Let us define the lowest reliably estimated SNR as a value where the difference between an estimated and a true SNR value starts to exceed 3 dB, which can be considered a practical tolerance for noisy data. It can be shown that under certain assumptions about statistical properties of signal

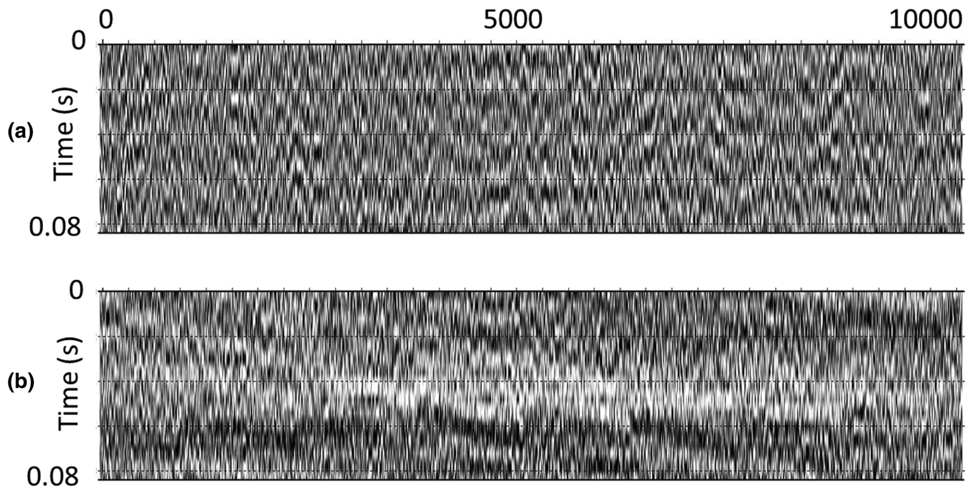

Figure 3 Windows extracted from real data and used for controlled SNR experiments: (a) noise-dominated data window extracted from a late-time portion (starting from 5 seconds); (b) signal-dominated data window (400–1000 m offsets) after NMO corrections containing a relatively strong reflection event.

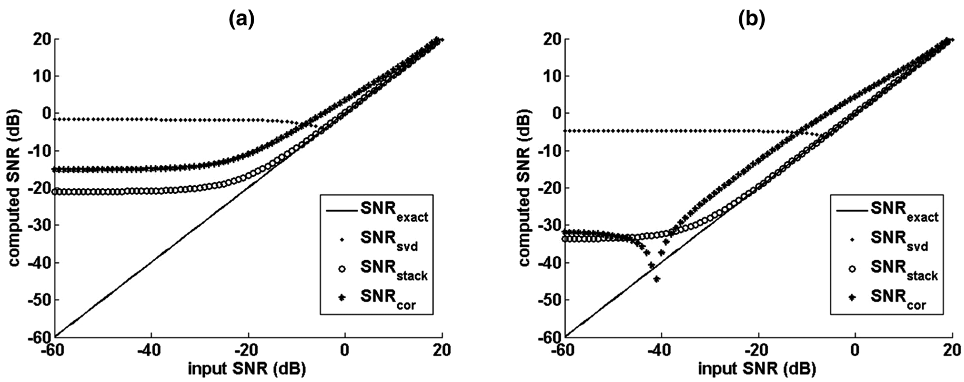

Figure 4 Estimated versus true signal-to-noise ratios for simulated synthetic data shown in Fig. 1(a) with a controlled level of realistic random noise from Fig. 3(a): (a) data consist of 100 traces; (b) data consist of 5000 traces. Each estimation method is shown with a different line: solid – ground-truth value computed with the exact formula (2); stars – correlation-based formula (65); circles – stacking-based method with formula (7); dots – SVD-based formula (9).

and noise, this lowest reliably estimated SNR value decreases when the number of data traces used for analysis increases. For the stacking-based method, the minimum required number of traces to reliably estimate the given $\mathrm{SNR}_{\mathrm{LRE}}$ is equal to (see the Appendix):

$$M_{LRE} = \left\lceil 1 + \frac{1}{\mathrm{SNR}_{LRE}} \right\rceil = \left\lceil 1 + 10^{-\frac{\mathrm{SNR}_{LRE}^{dB}}{10}} \right\rceil, \qquad (10)$$

where $\lceil\cdot\rceil$ means rounding toward positive infinity. However, for low SNR values below about −10 dB, the first term can often be neglected in practice leading to further simplification similar to a common square-root law:

$$M_{LRE} \approx \frac{1}{\mathrm{SNR}_{LRE}} = 10^{-\frac{\mathrm{SNR}_{LRE}^{dB}}{10}}. \qquad (11)$$

For example, for accurate computation of SNR values around −30 dB, one needs to have at least ~1000 traces. However, for reliable computation of SNR values down to −40 dB, at least ~10,000 traces are required. One simple physical interpretation of Equations (10) and (11) can be given if we recall that increase in SNR due to stacking of *M* traces is described by a factor of *M*.Therefore, Equations (10) and (11) suggest that ensemble size should be selected so that SNR reaches ~0 dB after stacking to achieve reliable estimation. Figure 1(c) provides a graphical illustration of the 0 dB case, suggesting that if stacking reveals a visually identifiable signal comparable to noise, then SNR can be robustly estimated (with the error of ~3dB). However, below this limit, signal estimate with stacking remains too inaccurate. As a result,

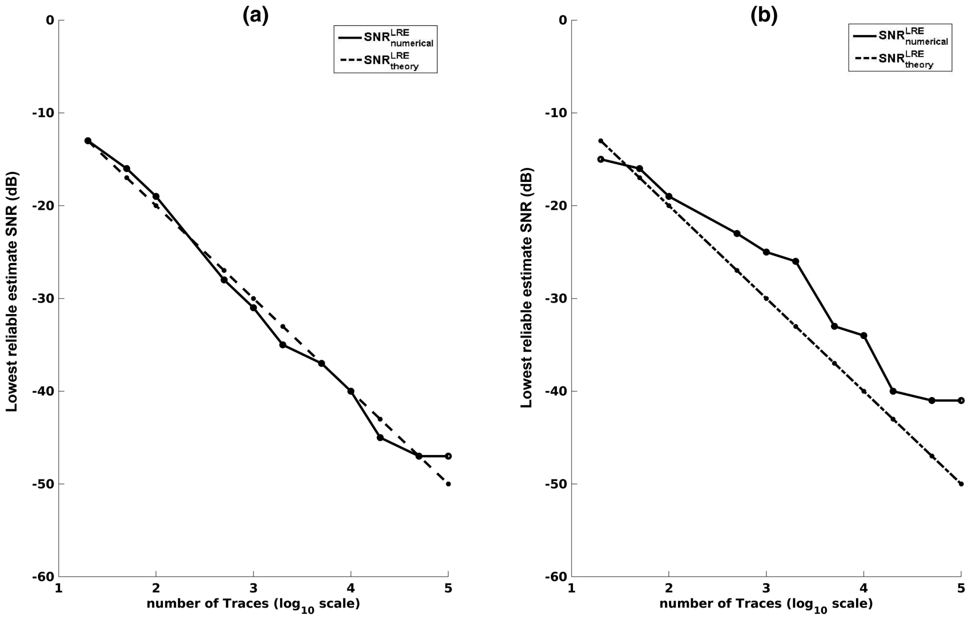

Figure 5 Lowest reliable theoretical SNR values (numerically estimated and predicted) obtained using the stacking-based method are shown as a function of the number of traces used in the estimation ensemble: (a) random noise; (b) real data noise.

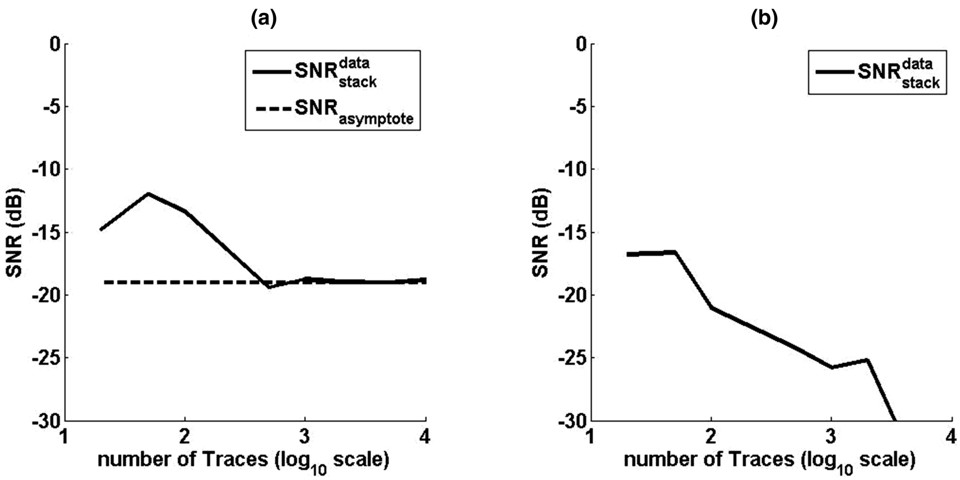

Figure 6 (a) Estimated SNR values for a field land single-sensor data from Fig. 3(b); (b) estimated SNR values for noise window shown in Fig. 3(a). On (a), SNR reaches a plateau of −19 dB considered as an actual absolute SNR. The presence of such a plateau is theoretically predicted and confirmed by a numerical example in Fig. A1. On (b), SNR estimate continues to decline without reaching any plateau (e.g. no signal or $\mathrm{SNR}^{dB} = -\infty$).

absolute SNR is overestimated by more than 3 dB, as shown in the Appendix.

Figure 5(a) and (b) shows the relationship's numerical verification for the lowest reliable SNR values for random Gaussian and realistic data noises. Both figures were obtained using controlled numerical experiments similar to those shown in Figs 2 and 4. We observe the good correspondence between the theoretical Equation (10) and actual results for synthetic random noise. For realistic noise, more traces are required to achieve reliable estimation than predicted by the theoretical curve. For example, 5000 traces are needed for real noise to reliably estimate SNR of −30 dB instead of 1000 for random noise. Likewise, ∼100,000 traces are required to reliably estimate SNR of −40 dB instead of 10,000 for random Gaussian noise. This controlled numerical experiment suggests that the theoretical relation (10) underestimates the

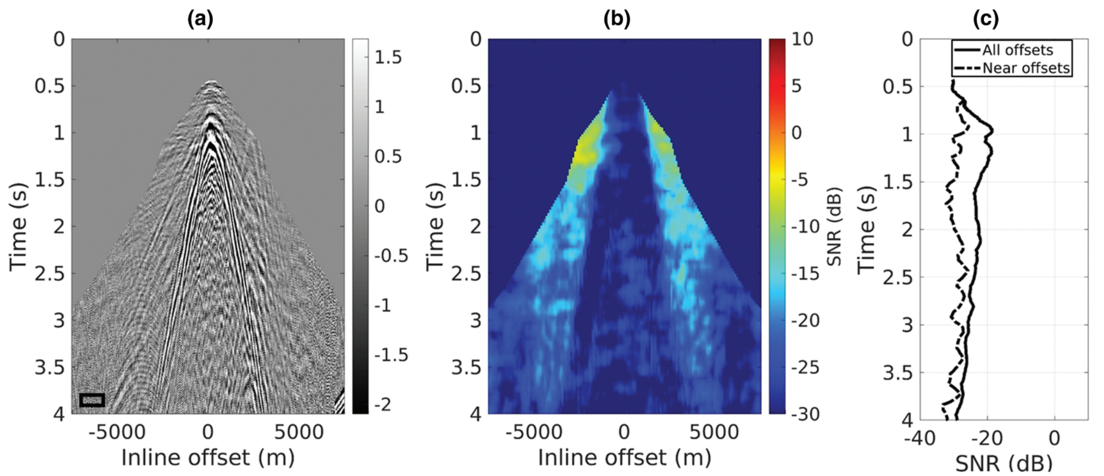

Figure 7 An example of SNR computation using the stacking-based method on a 3D cross-spread gather extracted from raw field single-sensor land dataset: (a) 2D slice (corresponding to a fixed source) from a 3D cross-spread gather after normal moveout corrections; (b) corresponding 2D slice of pre-stack SNR calculated with Equation (7); (c) vertical profiles of an average SNR computed for the whole offset range as well as near offsets only (0–1 km). The black rectangle shows the 2D cross-section of the time-space window (1250 m × 1250 m × 0.1 s) used in SNR computation.

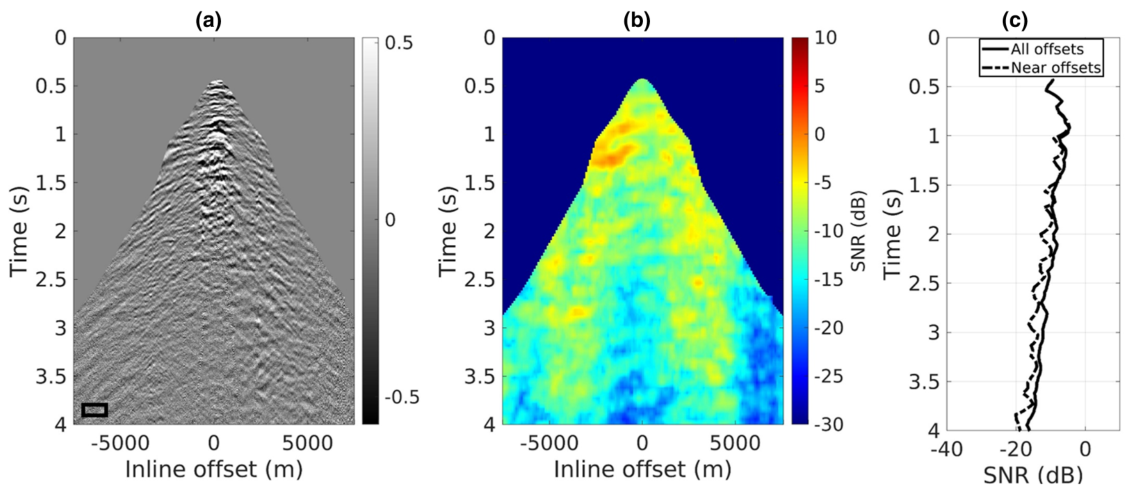

Figure 8 Same as Fig. 7 but for a gather after typical pre-processing. Compared to Fig. 7, observe consistently higher SNR values with the average uplift of ∼15 dB.

required number of traces to estimate the actual SNR value reliably and that more traces may be required for real data.

Real-data example: single reflector

As the following example, we perform SNR estimation using the stacking method on a portion of real data with a target reflection event. Figure 3(b) shows the time window extracted from the same land single-sensor seismic dataset. This data window contains a relatively strong reflected event, which could be confirmed by examining the corresponding stack section. Although processing, including noise attenuation, has already been applied, the signal on the pre-stack data remains very weak and hardly visible behind the noise. Figure 6(a) shows estimated SNR values using the stacking method for a gradually increasing number of traces extracted from the considered window. In the beginning, when the number of traces is small, the output SNR decreases with increasing the window's size. After around 1000 traces, the estimated SNR value reaches a plateau at around −19 dB, which can be

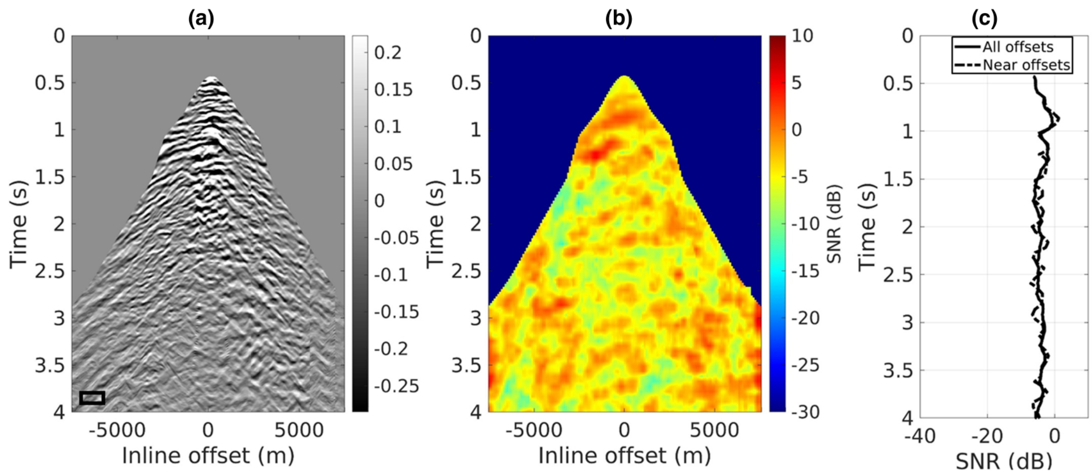

Figure 9 Same as Fig. 8 but for a gather after pre-processing with an additional enhancement using non-linear beamforming. Observe higher visibility of pre-stack events on (a). This is further substantiated by consistently higher SNR on (b) and the additional average SNR uplift of ~10 dB due to non-linear beamforming (compare Figs 8c and 9c).

considered the actual SNR of the target reflected event. Such asymptotic behaviour is theoretically expected for SNR estimate obtained with the stacking method as shown in the Appendix and demonstrated in Fig. A1. Once we reach the minimum required number of traces, adding more traces to the ensemble does not change the estimate. Finally, Fig. 6(b) shows the application of the stacking-based SNR estimation method to the noisy data window from Fig. 3(a). As no signal is expected in this case ($\mathrm{SNR}^{dB} = -\infty$), the SNR estimate decreases with the increasing number of traces, and a plateau is not reached. These examples, along with the theoretical dependence provided in the Appendix, lead us to a practical recipe for a processor on choosing the required number of traces for reliable SNR estimation in real data when the absolute value of the signal remains unknown. Trace number corresponding to the start of the plateau (e.g. Fig. 6a) can be considered a required size of the ensemble needed to obtain a reliable SNR estimate with formula (7). Otherwise, when an insufficient number of traces is used, then SNR estimates remain biased towards overly optimistic values.

Real-data example: full gather

The same signal-to-noise ratio (SNR) estimation technique can be extended to a three-dimensional (3D) pre-stack gathers instead of a single target reflector. Figure 7(a) shows a two-dimensional slice through a 3D cross-spread gather formed by fixing one receiver line and one source line orthogonal to each other. With a maximum inline offset of 6250 m, a cross-line offset of 4200 m and a sampling of 12.5 m in both directions, the total number of traces in this 3D gather is 672,000. A conventional global moveout correction should be applied to the data to make the reflection events flat, as assumed by Equation (1). Additional local moveout correction based on automatic coherency search as done in non-linear beamforming (NLBF) (Bakulin *et al.*, 2020), or other similar methods, can help to account for any residual moveout, making the estimation less sensitive to velocity errors. SNR computed from a single ensemble is output to the geometric centre of the ensemble. Figures 7–9 show the SNR estimates after three different processing stages: raw field gathers, data after typical pre-processing and pre-processing with added NLBF data enhancement. NLBF enhancement is selected as a mere example of a processing step designed to enhance the signal and suppress noise via local stacking. For the acquisition geometry at hand and selected cross-spread domain, the estimation ensemble of 10,000 traces translates into a local window of 1200 m by 1200 m, formed by 97 adjacent sources and 97 receivers in lateral directions and 0.1 seconds in the time direction. Theoretically, this allows getting a reliable SNR down to −40 dB, as indicated by Fig. 4(a). Three-dimensional space-time window defined above is moved along the gather (in time and space) to generate the SNR attribute at every point. When the required minimum window is much larger than the trace sampling, this results in a smooth behaviour of the SNR attribute. One can observe that typical SNR values for raw data are between −40 dB and −20 dB. The ground-roll noise cone at the near offsets (Fig. 7a) is clearly identified as an inner triangle with low SNR because surface waves do not coherently stack inside the ensembles due to their low apparent velocity. After a typical data processing, the SNR consistently increases across the gathers, falling into a range of −20 dB to −5 dB. After an additional step of data enhancement with NLBF, SNR values further improve and become within a range of −5 dB to 5 dB. In this case, the average increase in pre-stack SNR value after the typical pre-processing is approximately estimated as 15 dB. An additional data enhancement leads to an average improvement of another 10 dB, as shown in Figs 7(c), 8(c) and 9(c). We note that according to the theoretical curve from the controlled experiment (Fig. 4a), all these absolute SNR values can be trusted and used to assess data quality during acquisition and processing.

CONCLUSIONS

This study considers practical aspects of signal-to-noise ratio (SNR) calculation for challenging seismic data from desert environments obtained with modern high-density acquisitions using small source/receiver arrays or single sensors. The main goal is to identify an automatic data-driven method that allows obtaining a robust absolute value of SNR for target reflected waves, often hidden by coherent propagating modes of noise generated in the near-surface. We show that estimates from different commonly used SNR calculation algorithms such as correlation-based, stacking-based and singular value decomposition-based approaches can significantly vary in the presence of strong noise and deviate from the correct values. We reveal the practical limits of each algorithm by using a controlled SNR experiment providing verifiable metrics for the comparison. The stacking-based method, closely related to the semblance formula, shows the most stable and reliable results for very noisy data and allows accurate SNR estimations down to −40 dB and lower. Such a low signal level is not unusual for single-sensor data from a desert environment. As shown by the single-sensor example, conventional pre-processing improves the SNR by around 15 dB. Additional signal enhancement may be needed to further improve SNR to the level required by signal-demanding applications such as pre-stack inversion or reservoir characterization. Establishing and tracking reliable data-driven SNR metrics from acquisition to processing to inversion may allow quantitative assessment long sought by practicing geophysicists.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Van Do (Saudi Aramco) for valuable and encouraging discussions related to the topic of this study.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Andrey Bakulin https://orcid.org/0000-0002-6638-7821
Ilya Silvestrov https://orcid.org/0000-0002-2394-3576
Maxim Protasov https://orcid.org/0000-0003-0449-1639

REFERENCES

Bakulin, A., Silvestrov, I., Dmitriev, M., Neklyudov, D., Protasov, M., Gadylshin, K. and Dolgov, V. (2020) Nonlinear beamforming for enhancement of 3D prestack land seismic data. *Geophysics*, 85(3), V283–V296.

Belousov, A. (2011) Standard evaluation of the quality of field seismic material. *Devices and Systems of Exploration Geophysics*, 03(37), 31–36. (In Russian).

Bosworth, B.T., Bernecky, W.R., Nickila, J.D., Adal, B. and Carter, G.C. (2008) Estimating signal-to-noise ratio (SNR). *IEEE Journal of Oceanic Engineering*, 33(4), 414–418.

Chen, Z.D. and Fu, D.D. (1993) *Estimating signal noise ratio of multitrace seismic data by singular value decomposition*. Jianghan Petroleum Institute. (In Chinese).

Cordery, S. (2020) An effective data processing workflow for broadband single-sensor single-source land seismic data. *The Leading Edge*, 39, 401–410.

Foster, M.R. and Guinzy, N.J. (1967) The coefficient of coherence: its estimation and use in geophysical data processing. *Geophysics*, 32, 602–616.

Hatton, L., Worthington, M.H. and Makin, J. (1986) *Seismic data processing: theory and practice*. Hoboken, NJ: Blackwell Scientific Publications.

Key, S.C., Kirlin, R.L. and Smithson, S.B. (1987) Seismic velocity analysis using maximum-likelihood weighted eigenvalue ratios. *SEG Technical Program Expanded Abstracts*, pp. 461–464.

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

Neidell, N. and Taner, M. (1971) Semblance and other coherency measures for multichannel data. *Geophysics*, 36, 482–497.

Pecholcs, P., Al-Saad, R., Al-Sannaa, M., Quigley, J., Bagaini, C., Zarkhidze, A., *et al*. (2012) A broadband full azimuth land seismic case study from Saudi Arabia using a 100,000 channel recording system at 6 terabytes per day: acquisition and processing lessons learned. *SEG Technical Program Expanded Abstracts*, pp. 1–5.

Zhao, Y., Mao, N.B. and Chen, X. (2019) Calculation method of the signal-to-noise ratio attribute of seismic data based on structural orientation. *Applied Geophysics*, 16, 455–462.

APPENDIX

This Appendix derives a theoretical relation between the accuracy of signal-to-noise ratio (SNR) calculation according to stacking method (7) and the number of traces used for estimation. Equation (8) can be rewritten as

$$S = \frac{\sum_{i=1}^{N}\left[s_i^2 + 2s_i\overline{n_i} + (\overline{n_i})^2\right]}{\sum_{i=1}^{N}\left[s_i^2 + 2s_i\overline{n_i} + \frac{1}{M}\sum_{j=1}^{M}\left(n_{ij}\right)^2\right]} = \frac{R_{ss} + 2R_{s\bar{n}} + R_{\bar{n}\bar{n}}}{R_{ss} + 2R_{s\bar{n}} + R_{nn}}, \quad \text{(A1)}$$

where $\overline{n_i} = \frac{1}{M}\sum_{j=1}^{M} n_{ij}$ is an averaged noise over an ensemble of traces, R_{xy} is a zero-lag cross-correlation of pair of traces according to Equation (4). In the second part of Equation (A1), we assume that noise zero-lag auto-correlations R_{nn} are equal for all traces. Assuming that signal and noise are uncorrelated and taking into account that zero-lag auto-correlation is equal to energy, we can rewrite (A1) as

$$S = \frac{e_s + e_{\bar{n}}}{e_s + e_n} = \frac{e_s + \frac{1}{M}e_n}{e_s + e_n}, \quad \text{(A2)}$$

where e_s and e_n are signal and noise energy corresponding to a single trace. In the second part of Equation (4), we assume that an ensemble averaging of M traces reduces the noise energy as $\frac{1}{M}$, which is true for uncorrelated noise with zero mean. As a result, we arrive at the equation describing an estimate obtained with the stacking method ($\mathrm{SNR}_{\mathrm{stack}}$) as

$$\mathrm{SNR}_{\mathrm{stack}} = \frac{S}{1-S} = \frac{e_s + \frac{1}{M}e_n}{e_n - \frac{1}{M}e_n}. \quad \text{(A3)}$$

Assuming $\frac{1}{M} \to 0$ and retaining only linear terms with respect to $\frac{1}{M}$, we arrive at a simpler approximation:

$$\mathrm{SNR}_{\mathrm{stack}} \approx \mathrm{SNR} + \frac{1}{M}\frac{e_s + e_n}{e_n}. \quad \text{(A4)}$$

We can conclude that the estimated SNR value using the stacking method always exceeds the actual SNR and approaches it as $\frac{1}{M}$ when the number of traces involved in the calculation increases. The accuracy of the estimation can be written as

$$\frac{\mathrm{SNR}_{\mathrm{stack}} - \mathrm{SNR}}{\mathrm{SNR}} = \frac{1}{M}\frac{e_s + e_n}{e_s}. \quad \text{(A5)}$$

To achieve a given estimation accuracy ε for a given SNR, the minimum required number of traces should be

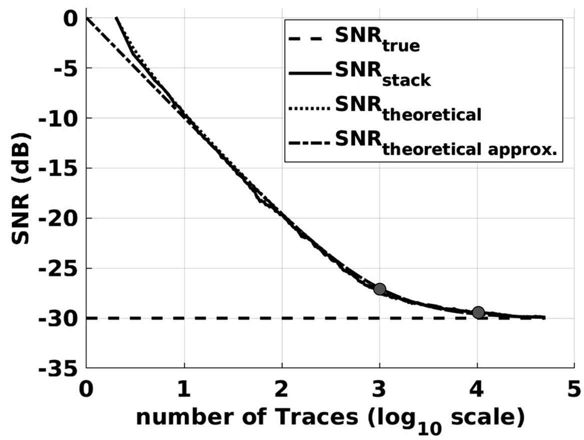

Figure A1 Comparison of numerical and theoretical dependence of SNR estimate using stacking method on the number of traces for the controlled synthetic experiment. Observe good agreement with all curves asymptotically reaching a plateau of true SNR value (−30 dB). As predicted, the accuracy of $\varepsilon = 1$ (∼3 dB error) is reached at 1000 traces, whereas the accuracy of $\varepsilon = 0.1$ requires 10,000 traces as marked by grey dots. $\mathrm{SNR}_{\mathrm{theoretical}}$ and $\mathrm{SNR}_{\mathrm{theoretical\ approx.}}$ are computed using Equations (A.3) and (A.4), respectively.

$$M_{LRE} = \frac{1}{\varepsilon}\frac{e_s + e_n}{e_s} = \frac{1}{\varepsilon}\left(1 + \frac{1}{\mathrm{SNR}}\right). \quad \text{(A6)}$$

A minimum number of traces increases with decreasing SNR. If we set $\varepsilon = 1$, we arrive at the following ratio between the estimated and actual SNR:

$$\frac{\mathrm{SNR}_{\mathrm{stack}}}{\mathrm{SNR}} = 2, \quad \text{(A7)}$$

or in decibel scale:

$$\mathrm{SNR}_{\mathrm{stack}}^{dB} - \mathrm{SNR}^{dB} \approx 3, \quad \text{(A8)}$$

which can be considered as a reasonable tolerance for practical applications with noisy seismic data. Figure A1 validates the derived equations via comparison with a controlled numerical experiment using Gaussian random noise of −30 dB. Theoretical estimates (A3) and (A4), and the actual estimate obtained using Equations (7) and (8) asymptotically approach plateau of correct values. We confirm that 1000 and 10,000 traces provide estimates of accuracy $\varepsilon = 1$ and $\varepsilon = 0.1$, respectively.

