Statistical Analysis of Seismic Phase Variability in Dense Data

A. Bakulin[1], A. Rohatgi, S. Fomel

[1] Bureau of Economic Geology

Summary

This abstract discusses phase variability in seismic data, emphasizing the novelty of treating phase as a statistical distribution rather than a deterministic value. Traditionally, phase as a statistical property has been overlooked, making this study the first to propose this approach. In real data, signal phase is often unknown and wrapped within $-\pi$ to π. While unwrapped phase is desirable for analysis, obtaining it in noisy data is challenging. Treating phase as a linear quantity is prone to misinterpretation. This study addresses such limitations by introducing circular statistics as a robust alternative for analyzing wrapped phase data.

Two methods are compared: a stack-based approach for estimating signal phase and a circular statistics method that avoids explicit unwrapping. The stack-based method derives residual phase differences by comparing traces with the stacked phase. In contrast, the circular statistics approach, based on the von Mises distribution, directly characterizes phase variability, providing unbiased estimates of the circular mean (dominant phase) and spread (concentration parameter, analogous to standard deviation).

Synthetic examples show how Gaussian and multiplicative noise affect phase distributions. The von Mises distribution reliably captures phase spread in noisy data, offering a robust framework for noise characterization and improving phase-dependent data interpretation and inversion.

Statistical Analysis of Seismic Phase Variability in Dense Data

Introduction

In modern seismic data analysis, phase plays a critical role in detecting and tracing seismic events, whether in images or pre-stack gathers. With the advent of high trace density in seismic data, we can start treating phase as a statistical distribution rather than a deterministic number. This allows us to analyze phase variability across local ensembles of traces, providing a meaningful approach to assess the quality of processing steps, data integrity, and the reliability of interpretations. The statistical perspective on phase is essential in extracting more accurate and robust insights from seismic data.

Phase Variability in Seismic Ensembles

Seismic data are inherently multi-trace, often comprising millions or billions of traces, which necessitates statistical approaches for their evaluation. For example, signal-to-noise ratio (SNR) estimations already rely on such methods to handle this complexity. Estimating the SNR of a single trace is impractical due to overlapping signal and noise. Instead, data-driven SNR methods analyze ensembles of traces from stack volumes or pre-stack gathers, assuming the signal remains constant while the noise is statistically similar. Larger ensembles provide more robust SNR estimates, especially at lower SNRs (Bakulin et al. 2022a).

Figure 1 presents a toy synthetic example with three horizontal events distorted by additive Gaussian and multiplicative noise (Bakulin et al. 2022b), demonstrating how noise introduces phase variability. The figure highlights how these noise types distort the traces and introduce variability in the signal phase. Visually, it is evident that the event phase is no longer constant. This contamination has significant implications for various seismic processing tasks, including event picking, static corrections, and the evaluation of phase-dependent attributes. Any variability in the wavelet phase directly undermines the reliability of these operations. Measuring this variability can be helpful diagnostics.

Recognizing this challenge, we propose treating phase variability as a statistical property rather than focusing on individual trace-to-trace differences or disruptions within the ensemble. We can more effectively quantify phase contamination by capturing phase variations and evaluating their statistical characteristics. This approach enables an objective assessment of phase distortion, providing a robust framework for improving seismic analysis and optimizing processing workflows.
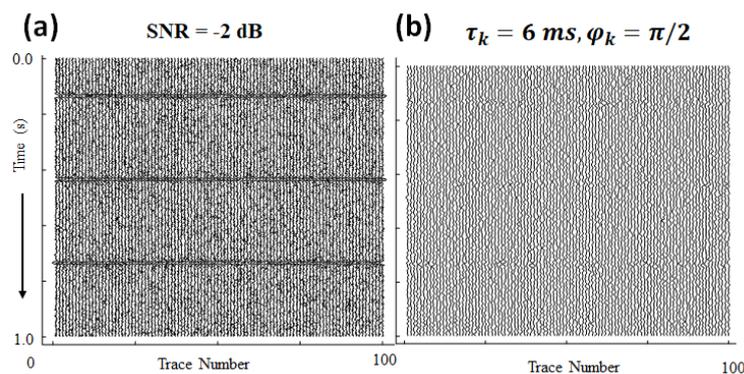

Figure 1 *Examples of different types of noise affecting an ensemble of one hundred traces:*
(a) Traces with additive Gaussian noise, SNR of -2 dB; (b) Traces with random multiplicative noise, including phase perturbations with a standard deviation of π/2 and residual static shifts with a standard deviation of 6 ms, resulting in signal variability and phase distortions. A Klauder wavelet represents the signal in all traces. This figure highlights the trace and phase variability we aim to quantify. Despite the different noise types, phase variability remains statistically consistent both spatially and vertically.

Methods of Practical Statistical Analysis for Real Data

Analyzing phase variability in real seismic data presents unique challenges compared to synthetic examples where the signal phase or wavelet is known. In real data, the signal phase is typically unknown and wrapped within the range of $-\pi$ to π. While unwrapping the phase is often desired in signal processing to analyze phase changes, this process is complex and can easily break down in areas where the phase is noisy. Wrapped phases limit the applicability of traditional linear distributions, such as the Gaussian or normal distribution. Consequently, practical phase analysis requires methods for estimating and interpreting phase variability using new tools, such as circular statistics designed for wrapped data, which can be reliably estimated without issues.

To address this, we describe two contrasting approaches: a stacking-based method to estimate the signal phase and a more general method that leverages circular statistics.

Stacking-Based Method: This two-step approach builds on data-driven SNR estimation principles (Bakulin et al. 2022b). The signal phase is derived by stacking traces along the signal moveout, producing a clean representative phase. The residual phase for each trace is calculated as the difference between its phase and the stacked signal phase and is then analyzed frequency by frequency.

In the case of multiplicative noise, the residual phase "wobbles" symmetrically around the signal phase as defined by the noise model (Bakulin et al. 2022b). A similar randomized phase perturbation pattern forms for additive Gaussian noise, following a symmetric distribution, as noted by Goodman (2020) and Rohatgi et al. (2024). While scattering noise shows frequency-dependent perturbations, additive Gaussian noise produces consistent phase randomization across frequencies.

Circular Statistics Method: This method avoids explicit signal phase estimation and stacking by directly analyzing wrapped phase values using circular statistics. Since the phase is bounded within $-\pi$ to π, the von Mises distribution—a circular analog of the Gaussian distribution—provides a natural framework for describing phase variability. The von Mises distribution captures how phase values cluster and spread around the mean value on the circular range.

The probability density function for the von Mises distribution is:
$$f(\alpha \,|\, \mu, \kappa) = \frac{\exp(\kappa \cos(\alpha - \mu))}{2\pi I_0(\kappa)} \, , \tag{1}$$
where, $I_0(\kappa)$ is the modified Bessel function of the first kind of order 0, μ represents the distribution center (analogous to the mean in a normal distribution) and κ measures concentration, with $\frac{1}{\kappa}$ analogous to variance σ^2, reflecting the spread of the distribution.

Figure 2 illustrates this method using phase distributions for additive Gaussian noise. The middle row shows histograms of phase values on a linear scale at different noise levels. In case of horizontal events without noise, the phase remains constant for all traces, collapsing the distribution to a single green bar. However, as noise increases, the phase values scatter more widely around the signal phase. At lower noise levels, the phase clusters more tightly around the signal phase (Figure 2a). As noise levels increase, the distribution tails reach the bounds $-\pi$ and π, creating an apparent bimodal pattern in linear plots (Figure 2 b, c – middle column).

This apparent bimodality arises due to the limitations of linear representation, where phase values near $-\pi$ and π appear as separate peaks rather than parts of a continuous cluster. The right column of Figure 2, displaying rose plots (histograms on circular scale), resolves this paradox by showing the phase distribution as a symmetric, petal-shaped cluster centered around the signal phase. This visualization underscores the importance of using circular coordinates to represent phase distributions accurately.

The von Mises distribution provides a robust means to quantify phase variability, even at high noise levels. The circular mean (μ) and concentration parameter (κ) describe the distribution's center and spread, enabling an accurate assessment of phase coherence and noise levels. This method is effective for typical noise types, such as multiplicative and Gaussian white noise, where phase variability behaves symmetrically.

If noise perturbs the signal phase symmetrically, the circular mean naturally provides an unbiased estimate of the signal phase. Similarly, Bakulin et al. (2022b) demonstrated that time-domain stacking also yields an unbiased estimate of the signal phase. Two practically important noise types—white additive noise and multiplicative speckle noise—generate symmetric phase distortions and conform to these principles.
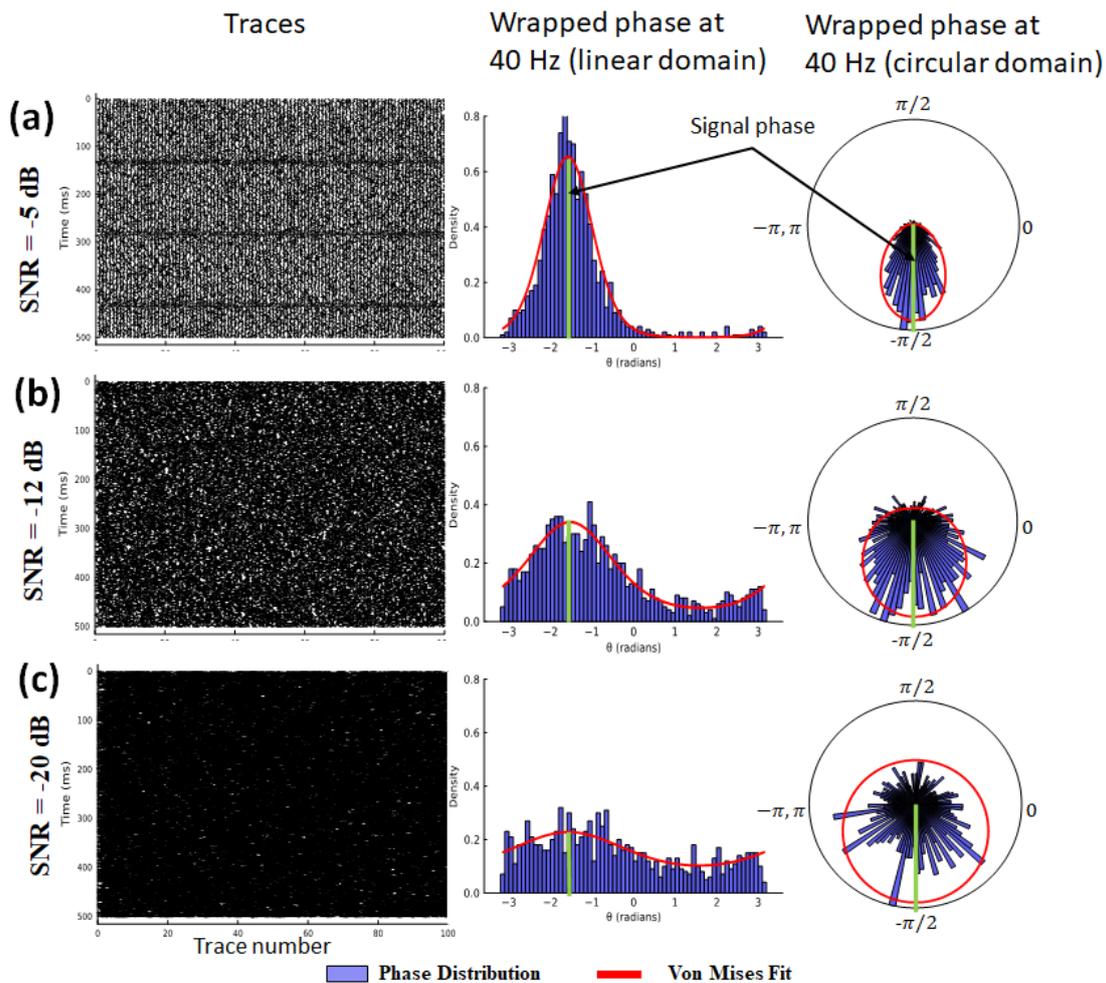

Figure 2 *Phase distributions for three horizontal reflectors with white Gaussian noise at SNR levels of -5 dB (a), -12 dB (b), and -20 dB (c). In (a), the wrapped phase is near-normal and clustered around the signal phase, as shown by the green bar. In (b), the distribution flattens with tails reaching $-\pi$ to π. In (c), the phase approaches a near-uniform distribution. The linear plots (middle column) can appear bimodal when the noisy phase clusters away from zero. At the same time, the circular plots (right column) reveal natural clustering in polar coordinates, making the appearance more intuitive and more straightforward for visual analysis and numerical assessment.*

Since phase distortion is a statistical quantity, the robustness of this estimate depends on the size of the ensemble and becomes more accurate as the ensemble size increases. Significant ensemble sizes are required for high noise levels, such as those in Figures 2b and 2c, to achieve an accurate signal phase estimate, making ensemble size selection an important consideration (Rohatgi et al., 2024b).

The spread of phase perturbations is a critical and quantifiable metric for any type of random noise. Specifically, in the case of white noise, it directly correlates with the SNR (lower κ corresponds to higher SNR). Using a von Mises fit, the circular statistics method provides a straightforward and visually intuitive way to quantify phase wobbling (as shown in the right column of Figure 2).

The concentration parameter (κ) can be translated into a standard deviation in the linear domain; however, its application is only valid for special cases where phase spreads are narrow or when the unwrapped phase can be directly estimated from the data. Both scenarios represent limiting cases (Bakulin et al. 2024).

Circular statistics offers a general and convenient approach to analyzing real data by estimating the mean and concentration parameters. This approach complements the stacking-based method while providing a more robust framework that avoids the need for explicit signal phase estimation, making it highly effective for analyzing phase variability in real seismic data. While our examples focus on a simple case of horizontal events without moveout, this approach can be extended across time, depth, or space using window-based or local attribute techniques (Fomel 2007).

Conclusions

We demonstrate that the phase in seismic data behaves as a statistical quantity with a distribution that can be leveraged for various data characterization purposes. In the presence of common noise types, such as additive Gaussian noise or multiplicative scattering noise, the phase undergoes randomization. However, this randomization follows a near-symmetric distribution that can be effectively modeled using the von Mises distribution within the framework of circular statistics.

By analyzing an ensemble of phases from a window or event, we can derive the statistical distribution of the phase within that ensemble. When the underlying signal remains constant across traces, the circular mean phase of the distribution corresponds to the signal phase. In contrast, the spread—described by the kappa parameter (κ) of the von Mises distribution—quantifies phase variability. This spread is closely correlated with the noise level, making it a helpful metric for characterizing noise magnitude.

Circular statistics provide a robust and intuitive way to capture phase variability in real seismic data, offering insights into data quality and noise levels. This method enables frequency-by-frequency analysis, allowing detailed assessment of noise levels at different frequencies, which is critical for interpretation, processing, and inversion tasks. Quantifying phase variability through statistical analysis offers a robust framework for evaluating data quality and guiding pre-processing decisions. By leveraging circular statistics and the von Mises distribution, this method enhances phase coherence assessment and helps reduce uncertainty, ensuring that seismic workflows are optimized for reliable interpretation.

References

Bakulin, A., Neklyudov, D., and Silvestrov, I. [2024] The impact of receiver arrays on suppressing seismic speckle scattering noise caused by the meter-scale near-surface heterogeneity. *Geophysics*, 89(6), 1ND-V667.

Bakulin, A., Silvestrov, I., and Protasov, M. [2022a] Signal-to-noise ratio computation for challenging land single-sensor seismic data. *Geophysical Prospecting*, 70(3), 629–638.

Bakulin, A., Neklyudov, D., and Silvestrov, I. [2022b] Multiplicative seismic noise caused by small-scale near-surface scattering and its transformation during stacking. *Geophysics*, 87(5), V419–V435.

Fomel, S. [2007] Local seismic attributes. *Geophysics*, 72(3), A29–A33.

Goodman, J.W. [2020] *Speckle phenomena in optics: Theory and applications*, 2nd ed. SPIE Press.

Mardia, K.V., and Jupp, P.E. [2000] *Directional statistics*. John Wiley & Sons.

Rohatgi, A., Bakulin, A., and Fomel, S. [2024a] Analyzing the impact of additive and multiplicative noise on seismic data analysis. *SEG Technical Program Expanded Abstracts*.

Rohatgi, A., Bakulin, A., and Fomel, S. [2024b] Phase pilot recovery: A foundation for mitigating speckle scattering noise in seismic data. *SEG Technical Program Expanded Abstracts*.