

Full-waveform inversion based on Kaniadakis statisticsSérgio Luiz E. F. da Silva^{ⓧ,*}, Pedro Tiago C. Carvalho^{ⓧ,†} and João M. de Araújo[‡]*Departamento de Física Teórica e Experimental, Universidade Federal do Rio Grande do Norte, Natal, Rio Grande do Norte, Brazil*Gilberto Corso[§]*Departamento de Biofísica e Farmacologia, Universidade Federal do Rio Grande do Norte, Natal, Rio Grande do Norte, Brazil*

(Received 1 April 2020; accepted 12 May 2020; published 27 May 2020)

Full-waveform inversion (FWI) is a wave-equation-based methodology to estimate the subsurface physical parameters that honor the geologic structures. Classically, FWI is formulated as a local optimization problem, in which the misfit function, to be minimized, is based on the least-squares distance between the observed data and the modeled data (residuals or errors). From a probabilistic maximum-likelihood viewpoint, the minimization of the least-squares distance assumes a Gaussian distribution for the residuals, which obeys Gauss's error law. However, in real situations, the error is seldom Gaussian and therefore it is necessary to explore alternative misfit functions based on non-Gaussian error laws. In this way, starting from the κ -generalized exponential function, we propose a misfit function based on the κ -generalized Gaussian probability distribution, associated with the Kaniadakis statistics (or κ -statistics), which we call κ -FWI. In this study, we perform numerical simulations on a realistic acoustic velocity model, considering two noisy data scenarios. In the first one, we considered Gaussian noisy data, while in the second one, we considered realistic noisy data with outliers. The results show that the κ -FWI outperforms the least-squares FWI, providing better parameter estimation of the subsurface, especially in situations where the seismic data are very noisy and with outliers, independently of the κ -parameter. Although the κ -parameter does not affect the quality of the results, it is important for the fast convergence of FWI.

DOI: [10.1103/PhysRevE.101.053311](https://doi.org/10.1103/PhysRevE.101.053311)**I. INTRODUCTION**

Full-waveform inversion (FWI) is a wave-equation-based methodology to estimate the physical parameters of the subsurface structures by exploiting the full information of the waveforms recorded in a seismic survey [1–3]. The FWI is a nonlinear inverse problem, in which the forward problem consists of simulating the wave propagation from a seismic source to the receivers by using the wave equation solution. The inverse problem consists of iteratively determining the coefficients of the wave equation (model parameters) data-driven by information extracted from the recorded waveforms (observed seismic data). In this process, the FWI employs an optimization method, usually a gradient-based algorithm [4]. However, the FWI is inherently an ill-posed problem in the sense of Hadamard [5], and therefore it has no unique solution [4]. Also, the noise decreases the accuracy of the measurements and makes the FWI unstable, increasing the ill-posedness of the problem.

Usually, this technique is formulated as a least-squares constrained optimization problem [3] to find an earth model that minimizes the difference between the observed and the modeled seismic data (residuals or errors) [6]. In this

approach, the least-squares solution of the inversion problem is associated with maximum likelihood in the well-known Gauss's error law. However, we are challenged because the distribution of residuals is seldom Gaussian in typical nonlinear problems [7,8]. Several studies have shown that criteria based on non-Gaussian approaches have been successful in FWI applications to real data, for instance, using the Cauchy distribution and sech criterion [9]. Reference [10] compare several criteria showing that the FWI based on non-Gaussian statistics has better performance, with more reliable results. Recently, FWI based on a non-Gaussian distribution has been successfully used in noise environments [11].

In this study, we use the κ -generalized exponential function of the Kaniadakis framework proposed in Refs. [12–14] to introduce an alternative misfit function which is based on the κ -generalized Gauss's law of error [15]. This probability distribution is obtained from the maximization of Kaniadakis entropy and it is known as the κ -Gaussian probability distribution [13,15]. The κ -Gaussian distribution is an ingenious extension of the Gaussian distribution derived from fundamental statistical mechanics assumptions [16]. A wide class of systems obeys κ -Gaussian statistics, such as in the derivation of the equipartition law of energy to obtain a modified gravitational constant [17], a generalization of Gauss's law of error [15], stellar systems [18,19], as well as detection of extreme changes in stock prices [20].

This paper is organized as follows: In Sec. II we briefly review the usual formulation of the FWI. In Sec. III we use the κ -generalized exponential to generalize the Gaussian

*sergioluizufn@ufrn.edu.br

†pedrotcc@gmail.com

‡joaomedeiros@fisica.ufrn.br

§gfcorso@gmail.com

distribution, and in Sec. IV, we apply the probabilistic maximum-likelihood method [8,21] to obtain the κ -misfit function. In addition, we explicitly derive the gradient function that corresponds to this misfit function. In Sec. V, we describe in detail the parameters used in the numerical experiments to test the accuracy of our methodology on a classical model used in exploration geophysics. We compare the standard Gaussian least-squares FWI results with the FWI based on κ -statistics in Sec. VI. Finally, in Sec. VII, we discuss the advantages of our approach and show the superiority of the methodology to overcome large errors in the empirical data.

II. FULL-WAVEFORM INVERSION

The conventional FWI approach is formulated as a least-squares optimization problem [3] given by

$$\min_{\mathbf{m}} \phi_G(\mathbf{m}) := \frac{1}{2} \sum_{s,r} \int_0^T [\Gamma_{s,r} u_s(\mathbf{m}, t) - d_{s,r}(t)]^2 dt, \quad (1)$$

where t denotes the time, \mathbf{m} represents the model parameters (coefficients of the wave equation), and T is the maximum value of the recording time. In addition, $\Gamma_{s,r} u_s$ and $d_{s,r}$ are the modeled and observed seismic data with $\Gamma_{s,r}$ representing the sampling operator (onto the receiver r of the source s). Finally, the variable u_s is the wave field which satisfies the acoustic wave equation:

$$\nabla^2 u_s(\mathbf{x}, t) - m(\mathbf{x}) \frac{\partial^2 u_s(\mathbf{x}, t)}{\partial t^2} = f(t) \delta(\mathbf{x} - \mathbf{x}_s), \quad (2)$$

where \mathbf{x} is the position (spatial coordinates) and ∇^2 the Laplacian operator. The term denoted by $f(t) \delta(\mathbf{x} - \mathbf{x}_s)$ represents the seismic source at the position \mathbf{x}_s and δ is the Dirac Delta function. In this framework, the model parameters are the inverse of the squared velocity: $m(\mathbf{x}) = \frac{1}{c^2(\mathbf{x})}$, for a medium with acoustic velocity c . Equation (2) can be compactly written as

$$A(m) u_s(t) = f_s(t), \quad (3)$$

where $A(m) = \nabla^2 - m(\mathbf{x}) \frac{\partial^2}{\partial t^2}$ is the space-time differential operator and $f_s(t) = f(t) \delta(\mathbf{x} - \mathbf{x}_s)$. It is important to point out that the spatial coordinate (\mathbf{x}) is implicit in Eq. (3) and henceforth.

In this usual approach, the statistical interpretation of the misfit, $\varepsilon = \Gamma_{s,r} u_s - d_{s,r}$, is that all errors (ε) are randomly distributed by a standard Gaussian probability distribution. In this way, note that the minimization of the expression (1) is equivalent to maximizing the Gaussian likelihood:

$$\mathcal{L}_G(\mathbf{m}) \propto \prod_{s,r} \exp \left\{ -\frac{1}{2} \int_0^T [\Gamma_{s,r} u_s(\mathbf{m}, t) - d_{s,r}(t)]^2 dt \right\}, \quad (4)$$

which is equivalent to minimizing the negative logarithm of $\mathcal{L}_G(\mathbf{m})$.

III. KANIADAKIS PROBABILITY DISTRIBUTION

Based on the kinetic interaction principle, Kaniadakis proposed a generalized statistics [12–14] by introducing the κ -generalized exponential function which is defined as

$$\exp_{\kappa}(y) = (\sqrt{1 + \kappa^2 y^2} + \kappa y)^{\frac{1}{\kappa}}, \quad (5)$$

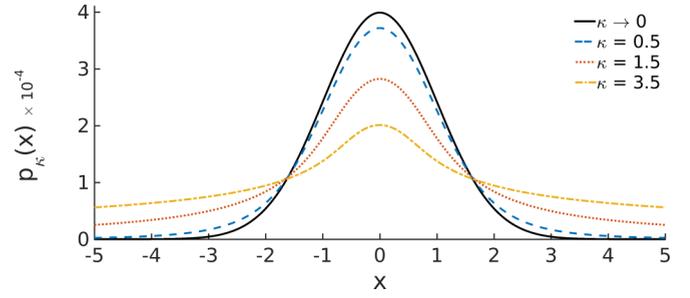


FIG. 1. Probability plots of κ -Gaussian distributions for typical values of κ . The κ -Gaussian resembles the Gaussian distribution, but with a long tail. The solid black curve represents the limit case corresponding to the Gaussian distribution ($\kappa \rightarrow 0$).

where κ is a real parameter that characterizes the deviation of the system from the standard Gaussian behavior. In the $\kappa \rightarrow 0$ case, Eq. (5) recovers the usual exponential function [12].

We use the Kaniadakis exponential to generalize the Gaussian distribution. In this way, the κ -Gaussian probability distribution associated with the κ -parameter can be defined as

$$p_{\kappa}(y) = C_{\kappa} \exp_{\kappa} \left(-\frac{1}{2} y^2 \right), \quad (6)$$

where C_{κ} is the normalizing factor to satisfy Kolmogorov probability axioms. So, Eq. (6) can be rewritten as

$$p_{\kappa}(y) = C_{\kappa} \left[\sqrt{1 + \frac{\kappa^2}{4} y^4} - \frac{\kappa}{2} y^2 \right]^{1/\kappa}. \quad (7)$$

We notice that in the limit $\kappa \rightarrow 0$ the standard Gaussian distribution is recovered.

Figure 1 shows the probability plots for typical values of κ , in which the solid black curve represents the Gaussian distribution ($\kappa \rightarrow 0$). In the $\kappa > 0$ cases, κ -Gaussian resembles the Gaussian distribution being symmetrical and bell-shaped, but with long tails that decrease as a power law: $p_{\kappa}(y) \propto y^{2/k}$. In the $\kappa \rightarrow \infty$ particular case, κ -Gaussian tends to a flat distribution.

IV. KANIADAKIS MISFIT FUNCTION

One important motivation of our study comes from the error (ε) distribution being, usually, non-Gaussian in nonlinear problems [7,8]. In this section, we will follow the opposite way from Sec. II. Given the probability function, what is the associated misfit function? Thus, assuming that all residual data ($\mathbf{y} = y_1, y_2, \dots, y_n$) are independent and randomly distributed by a κ -Gaussian distribution, the misfit function is given by the negative log-likelihood κ -function:

$$-\ln[\mathcal{L}_{\kappa}(\mathbf{y})] = -\ln \left[\prod_{i=1}^n p_{\kappa}(y_i) \right], \quad (8)$$

which is

$$-\ln[\mathcal{L}_{\kappa}(\mathbf{y})] = -\ln \left[\prod_{i=1}^n C_{\kappa} \left(\sqrt{1 + \frac{\kappa^2}{4} y_i^4} - \frac{\kappa}{2} y_i^2 \right)^{\frac{1}{\kappa}} \right] \quad (9)$$

and also can be written as

$$-\ln[\mathcal{L}_\kappa(\mathbf{y})] = -\ln(C_\kappa) - \frac{1}{\kappa} \sum_{i=1}^n \ln \left(\sqrt{1 + \frac{\kappa^2}{4} y_i^4} - \frac{\kappa}{2} y_i^2 \right). \quad (10)$$

However, in optimization problems, the main goal is to obtain the position of the optimal minimum instead of the amplitude of the misfit function. Therefore, we can ignore the term from the normalization constant in Eq. (10) [21], since any two functions $f(x)$ and $g(x) = f(x) + \text{constant}$ share the same minimum for a given x value. In this way, minimizing Eq. (10) is equivalent to minimizing the function

$$\phi_\kappa(\mathbf{y}) = -\frac{1}{\kappa} \sum_{i=1}^n \ln \left(\sqrt{1 + \frac{\kappa^2}{4} y_i^4} - \frac{\kappa}{2} y_i^2 \right), \quad (11)$$

which can also be written as

$$\phi_\kappa(\mathbf{y}) = -\sum_{i=1}^n \ln \left[\exp_\kappa \left(-\frac{y_i^2}{2} \right) \right]. \quad (12)$$

We assume that $\phi_\kappa(\mathbf{y})$ is the κ -misfit function and therefore the FWI problem becomes the following optimization problem:

$$\min_{\mathbf{m}} \phi_\kappa(\mathbf{m}) := -\sum_{s,r} \int_0^T \ln \left\{ \exp_\kappa \left[-\frac{1}{2} \Delta d_{s,r}^2(\mathbf{m}, t) \right] \right\} dt, \quad (13)$$

in which $\Delta d_{s,r}(\mathbf{m}, t) = \Gamma_{s,r} u_s(\mathbf{m}, t) - d_{s,r}(t)$ is the residual data.

Note that at the limit $\kappa \rightarrow 0$, the κ -misfit function becomes the standard FWI misfit function:

$$\lim_{\kappa \rightarrow 0} \phi_\kappa(\mathbf{m}) = -\sum_{s,r} \int_0^T \ln \left\{ \exp_0 \left[-\frac{1}{2} \Delta d_{s,r}^2(\mathbf{m}, t) \right] \right\} dt, \quad (14)$$

in which $\exp_0(y) = \exp(y)$ and $\ln[\exp(y)] = y$; thus,

$$\lim_{\kappa \rightarrow 0} \phi_\kappa(\mathbf{m}) = -\sum_{s,r} \int_0^T \left[-\frac{1}{2} \Delta d_{s,r}^2(\mathbf{m}, t) \right] dt, \quad (15)$$

which is equivalent to Eq. (1), which means

$$\lim_{\kappa \rightarrow 0} \phi_\kappa(\mathbf{m}) = \phi_G(\mathbf{m}). \quad (16)$$

Equation (13) is the κ -misfit function, and we call κ -FWI the formulation of the FWI that employs the Kaniadakis statistics.

Gradient of κ -FWI misfit function

FWI is typically performed by an iterative gradient-based method, in which the model parameters are updated along the gradient-descent direction of a misfit function ϕ [3]. Starting from a model \mathbf{m}_0 (initial model), the minimization problems formulated in Eqs. (1) and (13) are solved according to

$$\mathbf{m}_{j+1} = \mathbf{m}_j - \alpha_j \mathbf{D} \nabla_{\mathbf{m}} \phi(\mathbf{m}_j), \quad (17)$$

where α_j is a step length computed through a search strategy [22] in the j th iteration, and \mathbf{D} is a positive-definite matrix [22]. The gradient of the misfit function $\phi(\mathbf{m}_j)$ with respect to model parameters is denoted by $\nabla_{\mathbf{m}} \phi(\mathbf{m}_j) = \frac{\partial \phi(\mathbf{m}_j)}{\partial \mathbf{m}}$.

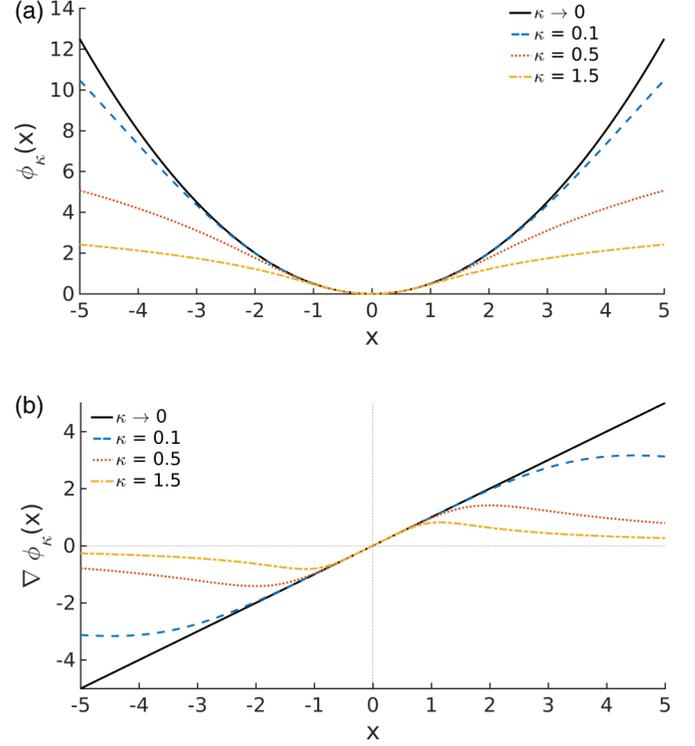


FIG. 2. Behavior of κ -misfit function and its derivative. The black curve represents the conventional misfit function ($\kappa \rightarrow 0$).

To obtain the gradient of the κ -misfit function $\phi_\kappa(\mathbf{m})$, one can compute the partial derivatives related to each model parameter m_l by

$$\frac{\partial \phi_\kappa(\mathbf{m})}{\partial m_l} = -\sum_{s,r} \int_0^T \frac{\frac{\partial}{\partial m_l} \left\{ \exp_\kappa \left[-\frac{1}{2} \Delta d_{s,r}^2(\mathbf{m}, t) \right] \right\}}{\exp_\kappa \left[-\frac{1}{2} \Delta d_{s,r}^2(\mathbf{m}, t) \right]} dt \quad (18)$$

or

$$\frac{\partial \phi_\kappa(\mathbf{m})}{\partial m_l} = \sum_{s,r} \int_T^0 J_{s,r}(\mathbf{m}, t) B_{s,r}^\kappa(\mathbf{m}, t) dt \quad (19)$$

with

$$J_{s,r}(\mathbf{m}, t) = \frac{\partial}{\partial m_l} [\Gamma_{s,r} u_s(\mathbf{m}, t)] \quad (20)$$

and

$$B_{s,r}^\kappa(\mathbf{m}, t) = \frac{\kappa \Delta d_{s,r}(\mathbf{m}, t) [1 - \kappa \Delta d_{s,r}^2(\mathbf{m}, t)]}{\exp_\kappa \left[-\frac{1}{2} \Delta d_{s,r}^2(\mathbf{m}, t) \right]}, \quad (21)$$

where $J_{s,r}$ is known as the Fréchet derivative [3].

For comparison, the gradient of the conventional misfit function is given by

$$\frac{\partial \phi_G(\mathbf{m})}{\partial m_l} = \sum_{s,r} \int_T^0 J_{s,r}(\mathbf{m}, t) \Delta d_{s,r}(\mathbf{m}, t) dt. \quad (22)$$

Figure 2 shows the behavior of the κ -misfit function, and its derivative, for the conventional FWI ($\kappa \rightarrow 0$) and the κ -FWI with $\kappa = 0.1, 0.5, 1.5$, and 3.5 . The conventional misfit function high-weights large errors [solid black curve in Fig. 2(b)], while in our proposal, the errors are down-weighted according to their residual amplitudes and the κ -parameter.

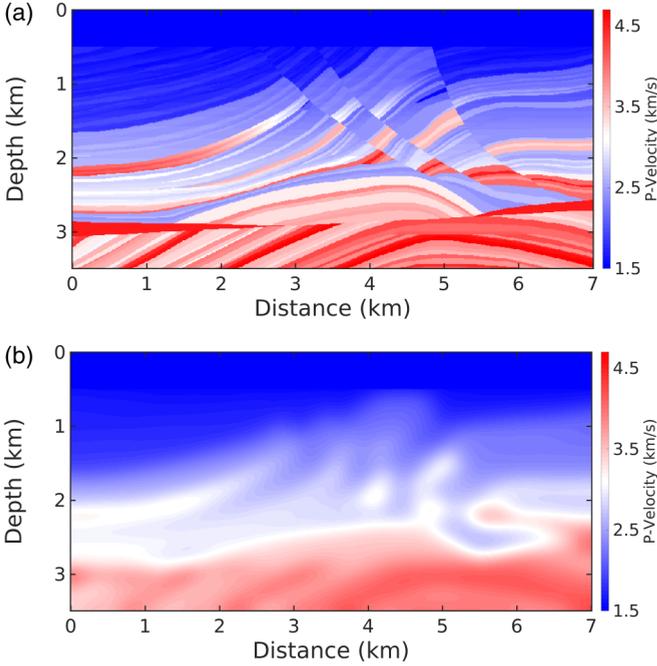


FIG. 3. Acoustic wave velocity models used in this study: (a) Marmousi model (true model) and (b) starting model (initial model), which was obtained by Gaussian smoothing of the true model.

This characteristic is the main reason κ -FWI is expected to be less sensitive to large errors.

V. NUMERICAL EXPERIMENTS

To illustrate how the κ -FWI improves the robustness of FWI to noise, we used the Marmousi acoustic velocity model as a benchmark (true model). This model is based on the geology of the Kwanza basin region (Angola) [23] and it is widely employed in seismic imaging studies because of its complexity [24]. The model contains a geometry with many reflectors and abrupt velocity variations: from 1.5 km/s to 4.7 km/s; the Marmousi model is depicted in Fig. 3(a). This model consists of 701 and 351 grid cells in the horizontal and vertical directions, respectively (246 051 total model parameters). In addition, the maximum depth is 3.5 km, and the maximum lateral distance is 7.0 km.

A Ricker wavelet [25] is used as the seismic source for all numerical experiments. In the time domain, it is defined as

$$f(t) = (1 - 2\pi^2\nu_p^2 t^2) \exp(-\pi^2\nu_p^2 t^2), \quad (23)$$

where ν_p is the peak frequency (most energetic frequency). In this study, we considered the peak frequency $\nu_p = 10$ Hz, as depicted in Fig. 4 together with its spectrum.

By using the true model, Fig. 3(a), and the seismic source defined in Eq. (23), the data set was generated with 82 equally spaced sources located every 80 m, from 250 m to 6730 m, at 10 m depth. The seismic wave field modeling is calculated with Eq. (2) using a classical 9-point finite-difference stencil with the perfectly matched layer absorbing boundaries [26] to simulate an infinite medium. We considered an acquisition

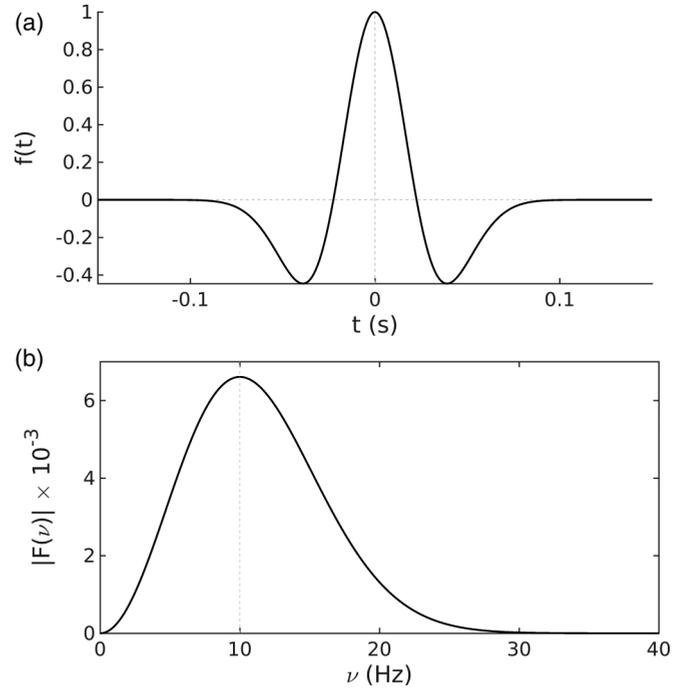


FIG. 4. Seismic source: (a) Ricker wavelet with a peak frequency of 10 Hz and amplitude given by Eq. (23), and its (b) frequency amplitude spectrum $|F(\nu)|$, where ν is the frequency.

with 171 receivers located every 40 m, from 40 m to 6860 m, deployed at 500 m depth, and the recording time was $T = 5$ s.

We performed simulations taking into account two different noise scenarios: (i) a data set contaminated by Gaussian noise with a signal-to-noise ratio (SNR) of 20 dB, and (ii) the same data set with Gaussian noise plus spiky noise (outliers). The SNR is computed by the ratio between the signal power and the noise power. The spikes were added over 15% of the recorded seismic traces (chosen randomly using a uniform distribution) by rescaling the signal amplitudes by a factor of 15α , in which α follows a standard Gaussian distribution that simulates a realistic noise [27]. Tests have been carried out for the noiseless data case, but the results from conventional FWI and κ -FWI methods were similar, and therefore we will not report this ideal case in this article. Figures 5(a) and 5(b) show examples of the observed data (seismograms) for these two scenarios, respectively, for the 41st seismic source.

In the minimizing problem, we used an optimization algorithm of quasi-Newton methods named *limited-memory* Broyden-Fletcher-Goldfarb-Shanno (*l*-BFGS) [22]. In this approach, the matrix \mathbf{D} in Eq. (17) is an approximation of the inverse of the Hessian matrix computed from previous gradients [22,28]. The step length is computed through a line search procedure that satisfies the Wolfe conditions [22]. The initial model was constructed from the true model using a Gaussian smoother with a standard deviation of 325 m, as depicted in Fig. 3(b). For each numerical experiment, we computed 50 *l*-BFGS iterations.

The gradient of the misfit function $\phi(\mathbf{m})$ is computed efficiently using the adjoint-state method [29], in which it is performed by cross-correlating the forward wave field u_s

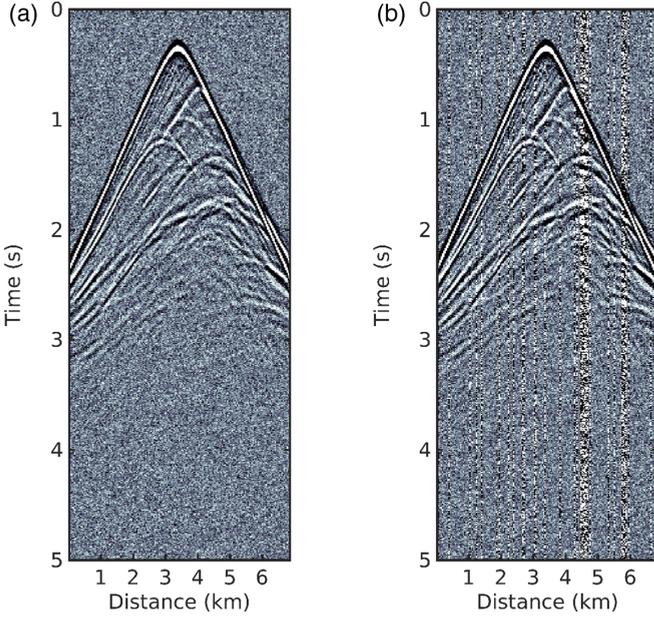


FIG. 5. Example of the observed data. Seismograms of the 41st source for scenarios with (a) Gaussian noise (first scenario), and (b) Gaussian noise plus spiky noise (second scenario).

[solution of Eq. (2)] with the adjoint wave field λ_s [2,30],

$$\nabla_m \phi(m) = \sum_s \left\langle \frac{\partial A(m)}{\partial m} u_s(t), \lambda_s(t) \right\rangle, \quad (24)$$

where $\langle \cdot \rangle$ is the inner product in the wave field space. In this approach, the wave field λ_s is the solution of the adjoint-wave equation [29]:

$$A^\dagger(m) \lambda_s(t) = \sum_r J_{s,r}^\dagger \mu_{s,r}(m, t), \quad (25)$$

where the superscript symbol \dagger refers to the adjoint operator (transpose conjugate) and $\mu_{s,r}$ is the adjoint source.

The adjoint source is equivalent to the multiplicative term of the Fréchet matrix [29]. In this way, the adjoint source for the conventional FWI is given by the residual data [3], while for our proposal it is given by Eq. (21). Comparing the gradients of the conventional FWI and the κ -FWI, Eqs. (22) and (19), one notices that the gradient of the proposed misfit function is the conventional FWI gradient weighted by the factor $\kappa [1 - \kappa \Delta d_{s,r}^2(\mathbf{m}, t)] / \exp_\kappa[-\frac{1}{2} \Delta d_{s,r}^2(\mathbf{u}, t)]$. In other words, the i th term of Eq. (19) is down-weighted by the magnitude of the i th residual.

Figure 6 shows the adjoint sources, at the first iteration, for the 41st seismic source. The adjoint wave field computed with conventional FWI is sensitive to large errors [see Fig. 6(a)], which means that the gradient shows a strong noise footprint. In contrast, the adjoint wave field computed from the κ -FWI becomes less sensitive to outliers in the data set when compared to the gradient of the conventional misfit function, in which large errors are dampened in the data inversion. This phenomenon is depicted in Figs. 6(b)–6(d) for $\kappa = 1.0, 5.0$, and 10.0 . We emphasize that the adjoint sources, for the first noise scenario, are quite similar; therefore they are not illustrated here.

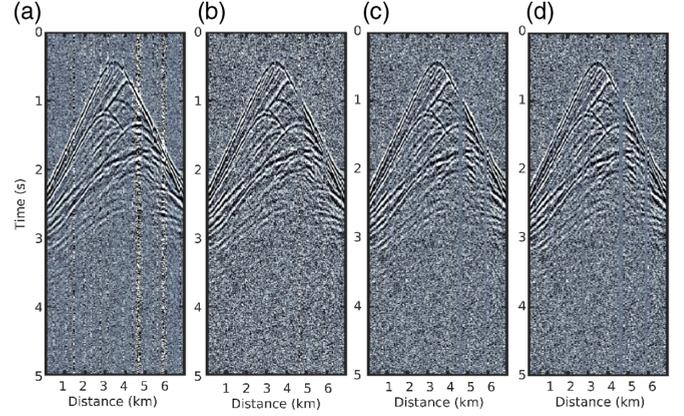


FIG. 6. Adjoint sources for the second scenario, at the first iteration, for (a) conventional FWI, and κ -FWI with (b) $\kappa = 1.0$, (c) $\kappa = 5.0$, and (d) $\kappa = 10.0$, for the 41st source.

Finally, for each noise scenario, we perform four inversions, in which the first one is based on the conventional approach, and the last three are based on our proposal with $\kappa = 1.0, 5.0$, and 10.0 . We emphasize that the position of the seismic source and receivers and the initial model are the same in all numerical experiments.

VI. RESULTS

Herein we compare the inversion results obtained from the κ -FWI and the conventional FWI, for the two noise scenarios described in Sec. V. The inversion results for the first scenario are depicted in Figs. 7(a)–7(d), for the conventional FWI and κ -FWI with $\kappa = 1.0, 5.0$, and 10 , respectively. Both the conventional FWI and κ -FWI yield estimations close to the true model as expected due to weak Gaussian noise. In fact, the first scenario considered was just a test of the stability of these methods in order to show the good functioning of the algorithms developed by the authors.

Since the true model is known we compare it quantitatively with the FWI reconstructions to test the quality of the inversion results. In this perspective, we compute three statistical measures: the normalized root-mean-square (NRMS), Pearson's correlation coefficient (R) [31], and the structural similarity (SSIM) index [32].

We define the NRMS as

$$\text{NRMS} = \left[\frac{\sum_i (c_i^{\text{true}} - c_i^{\text{inv}})^2}{\sum_i (c_i^{\text{true}})^2} \right]^{1/2}, \quad (26)$$

where c^{true} is true model and c^{inv} corresponds to the inversion result. The NRMS varies from 0 (perfect fit) to ∞ (bad fit). Pearson's correlation coefficient measures the statistical relationship between the true model and the inverted model based on the method of covariance [31]. Finally, the SSIM is a metric of similarity between images. Both R and SSIM metrics will vary between -1 (bad similarity) to 1 (perfect similarity). The statistical measures for the first noise scenario are summarized in Table I.

Our simulations show that both methods are comparable for the weak Gaussian noise scenario (see Fig. 7). However, the measures presented in Table I show that the velocity model

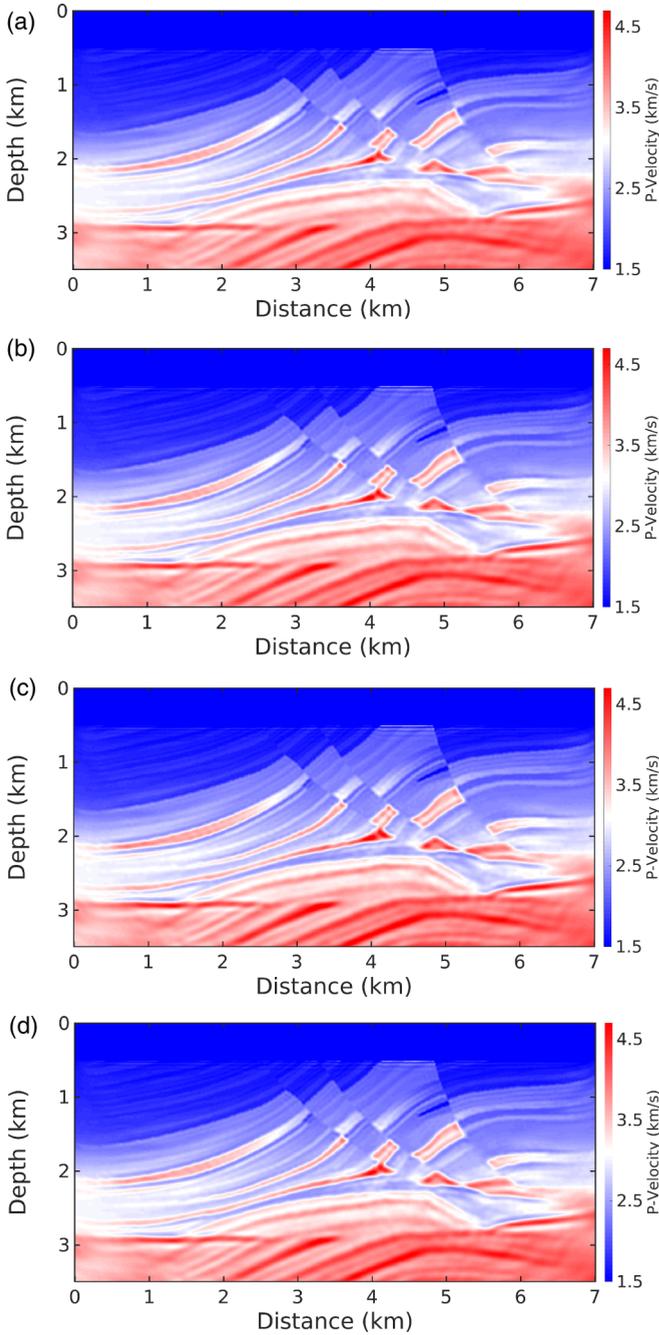


FIG. 7. Inversion results for the first scenario with (a) conventional FWI, and κ -FWI for (b) $\kappa = 1.0$, (c) $\kappa = 5.0$, and (d) $\kappa = 10.0$.

TABLE I. Main statistics at the first scenario: The normalized root-mean-square (NRMS) is based on the misfit between the true model and the inverted model. Pearson’s correlation coefficient (R) and the structural similarity index (SSIM) measure similarities between models.

Strategy	NRMS	R	SSIM
Conventional FWI	0.089	0.963	0.579
Our proposal ($\kappa = 1.0$)	0.084	0.967	0.601
Our proposal ($\kappa = 5.0$)	0.086	0.966	0.599
Our proposal ($\kappa = 10.0$)	0.087	0.965	0.588

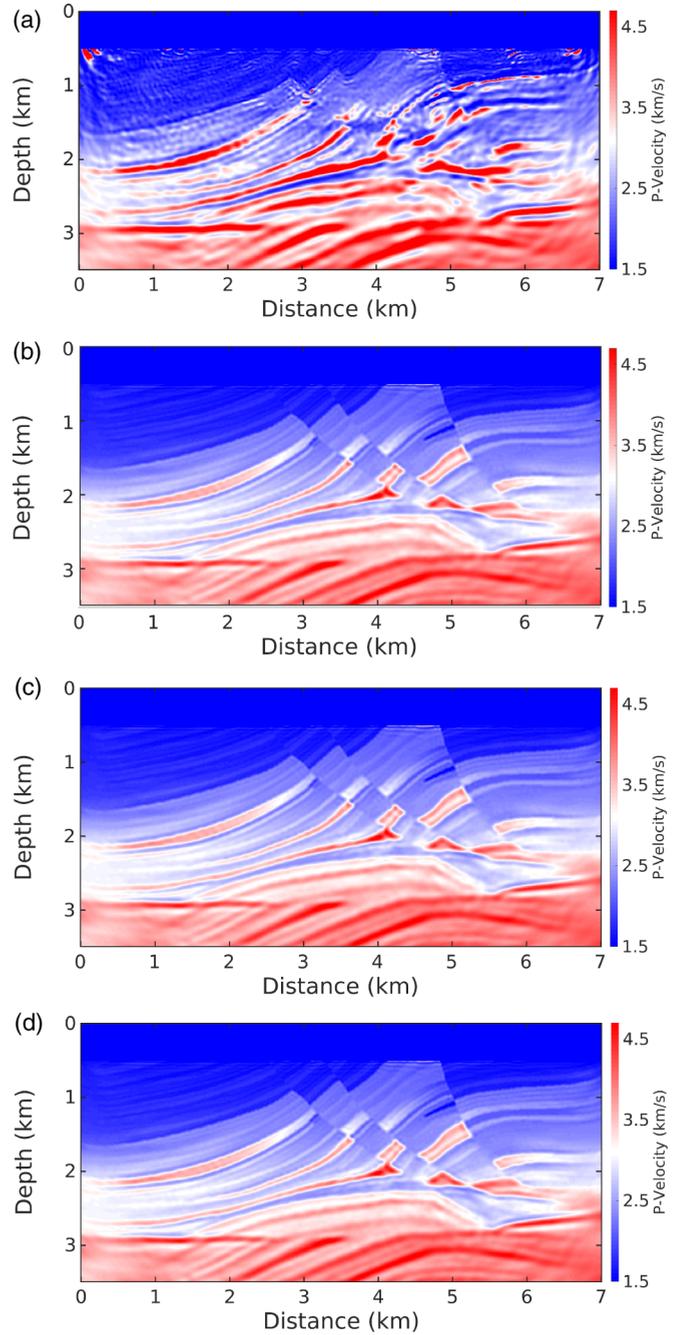


FIG. 8. Inversion results for the second scenario with (a) conventional FWI, and κ -FWI for (b) $\kappa = 1.0$, (c) $\kappa = 5.0$, and (d) $\kappa = 10.0$.

reconstructed with the κ -FWI has greater similarity (R and SSIM) and less error (NRMS) compared to the conventional FWI, especially in the $\kappa = 1.0$ case.

The second scenario consists of a more realistic application. The results are presented in Fig. 8. Note that the inversion with the conventional FWI fails to retrieve the velocity model satisfactorily [see Fig. 8(a)]. The standard inversion shows many artifacts, while our proposal provides images close to the true model, independently of the value of κ .

As expected, when non-Gaussian noisy data are used, the conventional FWI fails to recover most of the velocity model. Otherwise, the performance of the κ -FWI is very satisfactory

TABLE II. Main statistics at the second scenario: The normalized root-mean-square (NRMS) is based on the misfit between the true model and the inverted model. Pearson’s correlation coefficient (R) and the structural similarity index (SSIM) measure similarities between models.

Strategy	NRMS	R	SSIM
Conventional FWI	0.144	0.907	0.363
Our proposal ($\kappa = 1.0$)	0.097	0.955	0.529
Our proposal ($\kappa = 5.0$)	0.086	0.966	0.587
Our proposal ($\kappa = 10.0$)	0.084	0.967	0.597

and it is comparable to the conventional FWI result in the first scenario. Finally, Table II lists the NRMS, R , and SSIM for the inversion results obtained in the second scenario, in which it indicates that our proposal with $\kappa = 10$ produces the velocity model with lowest NRMS and the highest similarity (R and SSIM).

However, in real applications, we do not know the true model, and therefore the results are judged in the domain of seismic images. In the present work, we used the reverse time migration (RTM), which provides a subsurface reflectivity using the observed data and the estimated velocity model [1,33]. Figure 9 shows the RTM images from the velocity models shown in Fig. 8; it is notable that the κ -FWI produces a more accurate reflector map [see Figs. 9(b)–9(d)] than the conventional FWI [see Fig. 9(a)], especially for reflectors of deep structures. In addition, for shallow structures (approximately up to 2 km), the RTM image for the conventional FWI result, Fig. 9(a), has many artifacts that do not match the true model.

Figure 10 shows the evolution of the misfit functions, per iteration, for these two scenarios. In the first scenario, the conventional FWI showed a faster decay of the misfit function [solid black line in Fig. 10(a)] followed by the κ -FWI with $\kappa = 1.0, 5.0$, and 10.0 . In contrast, the κ -FWI with $\kappa = 10.0$ showed a faster misfit function decay [dash-dotted purple line in Fig. 10(b)] followed by the κ -FWI with $\kappa = 5.0$ and 1.0 , and the conventional FWI, in the second scenario.

VII. CONCLUSION

We have presented a misfit function to mitigate the influence of noise in the reconstruction of the velocity models by using the FWI method. Based on the κ -generalized exponential function, we proposed a misfit function based on the κ -Gaussian probability distribution, which is linked with the κ -generalization of Gauss’s law of error [15]. We call our proposal by the abbreviation κ -FWI, in reference to the Kaniadakis statistics (or κ -statistics). A numerical study with a complex acoustic velocity model demonstrates the effectiveness and robustness of our proposal for two noisy circumstances: first, a simple case with a weak Gaussian noise, and in the second one, a realistic noise circumstance. The first scenario was just a test to show the stability of these methods.

The results show that the κ -FWI is a powerful methodology in noisy environments. The proposed misfit function is less sensitive to the presence of noise, especially to non-Gaussian

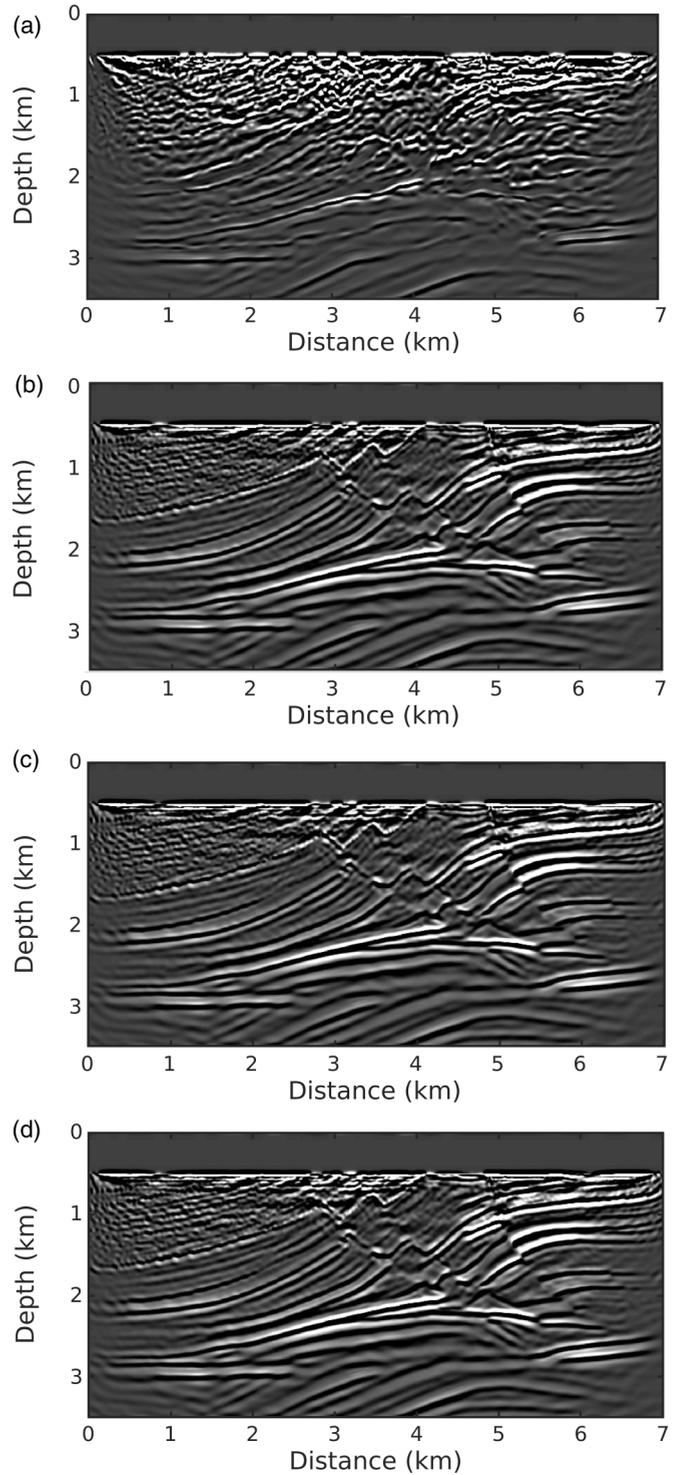


FIG. 9. RTM images for (a) conventional FWI resulting model, and κ -FWI resulting model with (b) $\kappa = 1.0$, (c) $\kappa = 5.0$, and (d) $\kappa = 10.0$.

noise with outliers. In addition, the κ -FWI provides a better model parameter estimation than the traditional approach, for no additional computational cost.

We notice that the quality of the results was not affected by the value of κ . The long tail of the κ -Gaussian distribution incorporated the weighted residuals, in which the residual

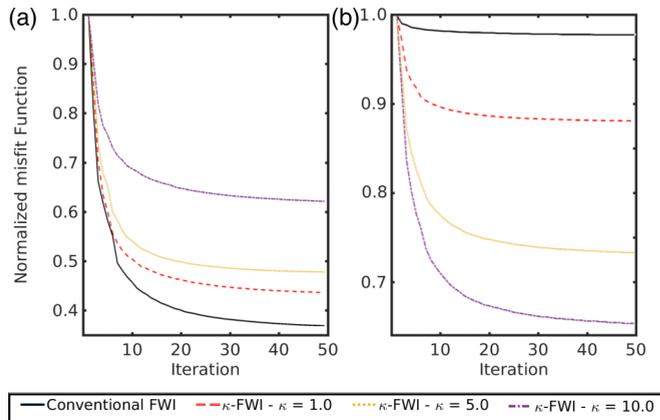


FIG. 10. Convergence. Evolution of misfit functions by iteration for (a) first and (b) second scenarios.

values that are much greater than the minimal error tend toward zero, as explicit in Eq. (19). However, the κ -parameter influences the FWI convergence. In the first scenario, we already expected that the conventional FWI convergence to be faster than our proposal, since the conventional FWI starts from the assumption that the errors are Gaussian, while our proposal deforms the Gaussian distribution. As the value of κ increases, the convergence is more slow. In contrast, in the second and more realistic scenario, the κ -FWI with $\kappa = 10.0$ shows the fastest convergence, followed by the κ -FWI cases

with $\kappa = 5.0$ and 1.0 , and the conventional FWI. As the value of κ increases, the convergence is more fast because the assumed distribution for the errors will be more distant from the usual Gaussian.

As a perspective, we intend to study the effect of the κ -parameter in the inversion considering other types of noise. In fact, the choice of the optimal κ is a challenging task to be addressed in this methodology. Our experience indicated that the best κ should be related to the quality of the data. To conclude, the κ -FWI is a valuable tool in exploration geophysics. We believe that this methodology is especially suited for robust inversion used in automated approaches, which means, without heavy data preprocessing work.

ACKNOWLEDGMENTS

The authors gratefully acknowledge support from Shell Brazil through the New Methods for Full-Waveform Inversion project at Federal University of Rio Grande do Norte (UFRN) and the strategic importance of support given by ANP through the R&D levy regulation. The authors also gratefully acknowledge support from the Brazilian National Council for Scientific and Technological Development (CNPq). The authors would like to thank Jorge L. Lopez from Shell for careful reading, comments, and suggestions on the manuscript, which greatly improved the clarity of this article. We also gratefully acknowledge valuable discussions with the researcher Carlos A. N. da Costa at UFRN; this ongoing exchange is of importance to our thinking.

- [1] P. Lailly, *Conference on Inverse Scattering: Theory and Application* (SIAM, Philadelphia, PA, 1983).
- [2] A. Tarantola, *Geophys.* **49**, 1259 (1984).
- [3] J. Virieux and S. Operto, *Geophys.* **74**, WCC1 (2009).
- [4] A. Fichtner, *Full Seismic Waveform Modelling and Inversion* (Springer, Berlin, 2010).
- [5] J. Hadamard, *Princeton Univ. Bull.* **13**, 49 (1902).
- [6] L. Sirgue and R. G. Pratt, *Geophys.* **69**, 231 (2004).
- [7] L. Amundsen, *Geophys.* **56**, 2027 (1991).
- [8] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (SIAM, Philadelphia, PA, 2005).
- [9] E. Crase, A. Pica, M. Noble, J. McDonald, and A. Tarantola, *Geophys.* **55**, 527 (1990).
- [10] R. Brossier, S. Operto, and J. Virieux, *Geophys.* **75**, R37 (2010).
- [11] S. L. E. F. da Silva, C. A. N. da Costa, P. T. C. Carvalho, J. M. de Araújo, L. S. Lucena, and G. Corso, *Phys. A: Stat. Mech. Appl.* **548**, 124473 (2020).
- [12] G. Kaniadakis, *Phys. A: Stat. Mech. Appl.* **296**, 405 (2001).
- [13] G. Kaniadakis, *Phys. Rev. E* **66**, 056125 (2002).
- [14] G. Kaniadakis, *Phys. A: Stat. Mech. Appl.* **365**, 17 (2006).
- [15] T. Wada and H. Suyari, *Phys. Lett. A* **348**, 89 (2006).
- [16] S. R. Kawa and S.-J. Lin, *J. Geophys. Res.* **108**, 4201 (2003).
- [17] E. M. Abreu, J. A. Neto, E. M. Barboza, Jr., and R. C. Nunes, *Phys. A: Stat. Mech. Appl.* **441**, 141 (2016).
- [18] J. C. Carvalho, R. Silva, J. D. do Nascimento, Jr., B. B. Soares, and J. R. de Medeiros, *Europhys. Lett.* **91**, 69002 (2010).
- [19] E. Bento, J. Silva, and R. Silva, *Phys. A: Stat. Mech. Appl.* **392**, 666 (2013).
- [20] E. Moretto, S. Pasquali, and B. A. Trivellato, *Eur. Phys. J. B* **90**, 1434 (2017).
- [21] W. Menke, *Geophysical Data Analysis: Discrete Inverse Theory* (Academic Press, London, 1984).
- [22] J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer, New York, 2006).
- [23] G. S. Martin, R. Wiley, and K. J. Marfurt, *Leading Edge* **25**, 156 (2006).
- [24] R. Versteeg, *Leading Edge* **13**, 927 (1994).
- [25] N. Ricker, *Bull. Seism. Soc. Am.* **33**, 197 (1943).
- [26] J. P. Berenger, *J. Comput. Phys.* **114**, 185 (1994).
- [27] T. Elboth, B. A. Reif, and O. Andreassen, *Geophys.* **74**, Q17 (2009).
- [28] L. Métivier and R. Brossier, *Geophys.* **81**, F1 (2016).
- [29] R.-E. Plessix, *Geophys. J. Int.* **167**, 495 (2006).
- [30] G. R. Pratt, C. Shin, and G. J. Hick, *Geophys. J. Int.* **133**, 341 (1998).
- [31] K. Pearson, *Philos. Trans. Royal Soc. London* **187**, 253 (1896).
- [32] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, *IEEE Trans. Image Process.* **13**, 600 (2004).
- [33] H.-W. Zhou, H. Hu, Z. Zou, Y. Wo, and O. Youn, *Earth-Sci. Rev.* **179**, 207 (2018).