

Article

A Graph-Space Optimal Transport Approach Based on Kaniadakis κ -Gaussian Distribution for Inverse Problems Related to Wave Propagation

Sérgio Luiz E. F. da Silva ^{1,2,*} , João M. de Araújo ^{3,†} , Erick de la Barra ^{4,†}  and Gilberto Corso ^{3,†} 

¹ Department of Applied Science and Technology, Politecnico di Torino, 10129 Torino, Italy

² Geoscience Institute, Fluminense Federal University, Niterói 24210-346, RJ, Brazil

³ Department of Theoretical and Experimental Physics, Federal University of Rio Grande do Norte, Natal 59072-970, RN, Brazil; gfcorso@gmail.com (G.C.)

⁴ School of Business, Universidad Católica del Norte, Coquimbo 1780000, CO, Chile

* Correspondence: sergio.dasilva@polito.it

† These authors contributed equally to this work.

Abstract: Data-centric inverse problems are a process of inferring physical attributes from indirect measurements. Full-waveform inversion (FWI) is a non-linear inverse problem that attempts to obtain a quantitative physical model by comparing the wave equation solution with observed data, optimizing an objective function. However, the FWI is strenuously dependent on a robust objective function, especially for dealing with cycle-skipping issues and non-Gaussian noises in the dataset. In this work, we present an objective function based on the Kaniadakis κ -Gaussian distribution and the optimal transport (OT) theory to mitigate non-Gaussian noise effects and phase ambiguity concerns that cause cycle skipping. We construct the κ -objective function using the probabilistic maximum likelihood procedure and include it within a well-posed version of the original OT formulation, known as the Kantorovich–Rubinstein metric. We represent the data in the graph space to satisfy the probability axioms required by the Kantorovich–Rubinstein framework. We call our proposal the κ -Graph-Space Optimal Transport FWI (κ -GSOT-FWI). The results suggest that the κ -GSOT-FWI is an effective procedure to circumvent the effects of non-Gaussian noise and cycle-skipping problems. They also show that the Kaniadakis κ -statistics significantly improve the FWI objective function convergence, resulting in higher-resolution models than classical techniques, especially when $\kappa = 0.6$.

Keywords: κ -Gaussian distribution; optimal transport; seismic imaging; cycle skipping; non-linear optimization; Wasserstein metric; inverse problems; wave propagation



Citation: da Silva, S.L.E.F.; de Araújo, J.M.; de la Barra, E.; Corso, G. A Graph-Space Optimal Transport Approach Based on Kaniadakis κ -Gaussian Distribution for Inverse Problems Related to Wave Propagation. *Entropy* **2023**, *25*, 990. <https://doi.org/10.3390/e25070990>

Academic Editor: Ali Mohammad-Djafari

Received: 3 May 2023

Revised: 15 June 2023

Accepted: 25 June 2023

Published: 28 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The task of inferencing physical parameters from indirect observations arises in various practical problems. Determining parameters that cannot be directly observed remains a complex issue and involves a robust set of tools that compose the theoretical basis of the inverse problem theory [1]. The goal of an inverse problem consists of obtaining a quantitative model m that explains the observations (or observed data) by matching modeled data $d^{mod} = G(m)$ to observed data d^{obs} , in which G denotes the so-called forward operator. The forward operator maps the variables from the model space to the data space through a physical law [2]. For instance, we may want to determine the thermal diffusivity of a material (physical system) by analyzing the observed data: the temporal distribution of the diffusing material density at a determined location. In this regard, the model consists of the collective diffusion coefficient, and a diffusion equation represents the forward operator G . So, the diffusion coefficients are determined by optimizing an objective function, which measures the distance between modeled and observed data.

In this work, we consider a non-linear inverse problem that has attracted increasing interest in several fields, such as astrophysics [3], biomedicine [4], machine learning [5],

and geophysics [6], named the Full-Waveform Inversion (FWI) [7]. FWI is a powerful imaging technique to obtain high-resolution quantitative physical models by analyzing the complete information of a collection of waveforms [8]. In particular, we consider the FWI technique in a geophysical context, in which the forward problem consists of simulating the propagation of acoustic waves by solving a wave equation. In this regard, the acoustic wave equation represents the forward operator G . At the same time, the data d and the model m are the pressure waveforms and the distribution of acoustic wave velocities (coefficients of the wave equation) of the subsurface medium. The inverse problem involves inferring the coefficients of the wave equation (model parameters) by comparing the modeled data (wave equation solution) with the observed data by employing an objective function [9].

The objective function based on the least-squares method (or squared l_2 -norm) is the most employed for handling FWI issues [7]. The least-squares objective function (from now on, classical objective function) computes the square root of the sum of the absolute squares of the residual data (or errors), the difference between the modeled and the observed data. Each objective function is closely connected to a statistical interpretation of the errors [2]. The classical objective function bears a relationship to Gaussian statistics. Indeed, in this classical framework, the errors are assumed to be independent and identically distributed according to a Gaussian probability distribution [10]. However, this assumption is sometimes adequate since the errors seldom are Gaussian in non-linear problems [11,12]. Let us remind the reader that errors arise from different natures, comprising the noise in the observations and uncertainties related to the physical rule employed in the forward problem. In fact, non-Gaussian noises are present in geophysical datasets and are caused by several elements, such as weather-related mechanisms [13] and instrument noise [14]. Several objective functions based on non-Gaussian statistics have been presented in the literature as alternative criteria. Non-Gaussian distributions exhibit much longer tails than the Gaussian ones, a crucial feature for dealing with erratic data (outliers) [15]. Several works have shown the effectiveness of non-Gaussian criteria in geophysical data inversions, such as objective functions based on Laplace distribution [16], Student's t distribution [17], generalized approaches [18,19], and hybrid criteria [20,21].

Recently, ref. [22] introduced a new non-Gaussian criterion, namely, the κ -objective function, based on the Kaniadakis statistics (or κ -statistics) [23–27], which is robust to erratic data. The κ -objective function assumes that the errors are independent and identically distributed according to the κ -deformation of a Gaussian distribution (or κ -Gaussian distribution), in which the classical approach is a particular case [28]. The κ -Gaussian distribution arises from optimizing the Kaniadakis κ -entropy as a generalization of the well-known Gaussian distribution [29]. The κ -criterion exhibits robust characteristics thanks to the much longer tail of the κ -Gaussian distribution than the classical Gaussian probability function, which is crucial to mitigate the effects of non-Gaussian errors in FWI problems [30].

Due to the high computational efforts to solve the wave equation several times during the FWI process, the minimization process of the objective function is usually solved by local optimization methods [7]. Thus, FWI is prone to trapping into a non-informative local minimum if the initial background velocity model is not kinematically accurate [31]. Such an intrinsic limitation of FWI is associated with the absence of low-frequency contents, causing cycle-skipping issues [32,33]. Cycle-skipping is a phase ambiguity problem when the phase correspondence between two waveforms is greater than half a wavelength [34]. Although the approaches mentioned above are robust to non-Gaussian errors, they measure sample-by-sample the data misfit, making them sensitive to cycle skipping. Hence, a vast body of objective functions has been introduced for mitigating the cycle-skipping effects, such as those based on the waveform envelopes [35], convolutional filters [36], non-parametric techniques [37], and optimal transport metrics [38]; this is the methodology employed in the present study.

The theory of optimal transport (OT) was formally introduced by Gaspard Monge [39], who sought to understand the most effective allocation of resources by redistributing

materials (mass) from sources to sinks. In recent years, OT theory has received much attention in broad literature (e.g., refs. [40–43]), such as geophysics issues [44–46]. However, the OT-based objective function is suitable for comparing probability distributions, that is, positive and normalized quantities, two requirements that seismic signals do not greet due to their oscillatory nature. In this way, waveforms are commonly distorted through transformations to satisfy the probability axioms, which may manufacture unwanted information. Indeed, several applications have demonstrated the effectiveness of OT-based objective functions to mitigate the effects of phase ambiguity; however, all assume that the errors obey Gaussian statistics.

In this work, we explore the κ -objective function in the context of OT theory to introduce an objective function resistant to non-Gaussian noise and less sensitive to cycle-skipping issues. In this regard, we propose a robust framework for matching seismic waveforms using the Wasserstein distance, a well-posed relaxation of the OT formulation. Furthermore, inspired by ref. [47], we consider the representation of the waveforms in the graph space suitable for real large-scale problems [33]. In this approach, the waveforms are represented by Dirac delta functions in a two-dimensional space (amplitude *versus* time).

We organize the present work as follows. In Section 2, we briefly introduce the theoretical basis of inverse problems in the context of κ -Gaussian statistics and their robustness properties. In Section 3, we present a well-posed relaxation of the original optimal transport formulation using the Kaniadakis κ -objective function. Then, in Section 4 we present FWI based on optimal transport and κ -Gaussian statistics in the context of the adjoint state method. In Section 5, we demonstrate how the proposed objective function deals with cycle-skipping issues and non-Gaussian noise by considering a Brazilian pre-salt case study. Finally, we devote Section 6 to the final remarks and future applications.

2. Inverse Problems in the Context of Kaniadakis κ -Statistics

In science issues, several practical problems are data-centric. Indeed, determining a quantitative physical model that explains the observations is crucial to more accurately model and describe a wide variety of existing physical systems. In this context, the inverse problem theory is an excellent tool.

From a practical point of view, an inverse problem is formulated as an optimization task for obtaining a quantitative model by comparing modeled data to observed data. Modeled data are calculated using an appropriate physical law. The comparison between modeled and observed data is performed through an objective function. In the classical approach, the objective function is constructed from the assumption that the errors (the difference between modeled and observed data) obey Gaussian statistics. Let $\vec{\varepsilon} = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N\}$ be the errors. From the assumption that the errors are independent and identically distributed according to a standard Gaussian distribution,

$$p_0(\varepsilon_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\varepsilon_i^2\right), \quad (1)$$

we can determine the associated likelihood function as follows [11]:

$$\mathcal{L}_0 = \prod_{i=1}^N p_0(\varepsilon_i) = \left(\frac{1}{\sqrt{2\pi}}\right)^N \exp\left(-\frac{1}{2}\sum_{i=1}^N \varepsilon_i^2\right), \quad (2)$$

where $\mathcal{L}_0$ is the Gaussian likelihood. The use of index 0 will become clear later on. It is worth remembering that the standard Gaussian distribution can be determined from the maximization of the Boltzmann–Gibbs–Shannon entropy subject to the normalization condition

$$\int_{-\infty}^{+\infty} p(\varepsilon) d\varepsilon = 1 \quad (3)$$

and the unit variance constraint,

$$\int_{-\infty}^{+\infty} \varepsilon^2 p(\varepsilon) d\varepsilon = 1. \quad (4)$$

In inverse problems, the errors $\vec{\varepsilon}$ depend on the model parameter m and are computed through the difference between modeled ($d^{mod} = G(m)$) and observed (d^{obs}) data, i.e., $\varepsilon_i(m) = d_i^{mod}(m) - d_i^{obs}$, where G represents the forward operator. In this way, obtaining the model parameter can be performed by employing the maximum likelihood estimation (MLE) method, which is achieved by maximizing the likelihood function as follows:

$$\hat{m} = \max_m \mathcal{L}_0(m|d^{obs}) \quad (5)$$

where $\hat{m}$ represents the estimated model. The MLE estimates an unknown model parameter by considering that its optimal value maximizes the probability that the observed data are measured. Since the minimum of the negative log-likelihood coincides with the maximum of the likelihood function, maximizing $\mathcal{L}_0$ (5) is equivalent to minimizing the negative log-likelihood, i.e.,

$$\max_m \mathcal{L}_0(m|d^{obs}) \equiv \min_m -\ln(\mathcal{L}_0(m|d^{obs})). \quad (6)$$

From the principle of maximum likelihood, an objective function ϕ_0 can be obtained from [11]:

$$\phi_0(m) \propto -\ln(\mathcal{L}_0(m|d^{obs})), \quad (7)$$

which can be rewritten as:

$$\phi_0(m) \propto \frac{N}{2} \ln(2\pi) + \frac{1}{2} \sum_{i=1}^N \varepsilon_i^2 \quad (8)$$

$$\phi_0(m) = \frac{1}{2} \sum_{i=1}^N \varepsilon_i^2. \quad (9)$$

We notice that minimizing Equation (8) or (9) is the same since the term $\frac{N}{2} \ln(2\pi)$ is constant. The latter equation is well-known and used in solving problems via the least squares method. Please see Section 2 of ref. [48] for more detail.

However, due to the non-Gaussianity of the errors, it is reasonable to assume that the errors are non-Gaussian. In this study, we consider that the errors are distributed according to a Kaniadakis κ -Gaussian distribution of the form [22]:

$$p_\kappa(\varepsilon_i) = \frac{1}{Z_\kappa} \exp_\kappa(-\beta_\kappa \varepsilon_i^2), \quad (10)$$

where Z_κ is a normalizing constant, β_κ is a scale parameter, and

$$\exp_\kappa(y) = \exp\left(\frac{1}{\kappa} \operatorname{arcsinh}(\kappa y)\right) = \left(\sqrt{1 + \kappa^2 y^2} + \kappa y\right)^{\frac{1}{\kappa}} \quad (11)$$

with $0 \leq |\kappa| < 1$, is the κ -exponential function [26], a generalization of the exponential function. The κ -exponential becomes the ordinary exponential function in the limit $\kappa \rightarrow 0$: $\exp_0(y) = \exp(y)$.

Considering the normalization (3) and unitary variance (4) conditions, we obtain

$$Z_\kappa = \sqrt{\frac{\pi}{\beta_\kappa}} \frac{|2\kappa|^{-1/2} \Gamma\left(\frac{1}{|2\kappa|} - \frac{1}{4}\right)}{1 + \frac{1}{2}|\kappa| \Gamma\left(\frac{1}{|2\kappa|} + \frac{1}{4}\right)} \quad (12)$$

and

$$\beta_\kappa = \frac{|2\kappa|^{-1} \frac{1 + \frac{1}{2}|\kappa|}{1 + \frac{3}{2}|\kappa|} \Gamma\left(\frac{1}{|2\kappa|} - \frac{3}{4}\right) \Gamma\left(\frac{1}{|2\kappa|} + \frac{1}{4}\right)}{\Gamma\left(\frac{1}{|2\kappa|} + \frac{3}{4}\right) \Gamma\left(\frac{1}{|2\kappa|} - \frac{1}{4}\right)} \quad (13)$$

holding for $|\kappa| < 2/3$. The standard Gaussian distribution (1) is a particular case of the κ -Gaussian distribution (10) in the classical limit $\kappa \rightarrow 0$ since $\lim_{\kappa \rightarrow 0} \beta_\kappa = \frac{1}{2}$ and $\lim_{\kappa \rightarrow 0} Z_\kappa = \sqrt{2\pi}$. Figure 1 depicts the plots of the κ -Gaussian distribution (10) for typical κ -values, with the solid black curve referring to the standard Gaussian distribution ($\kappa \rightarrow 0$).

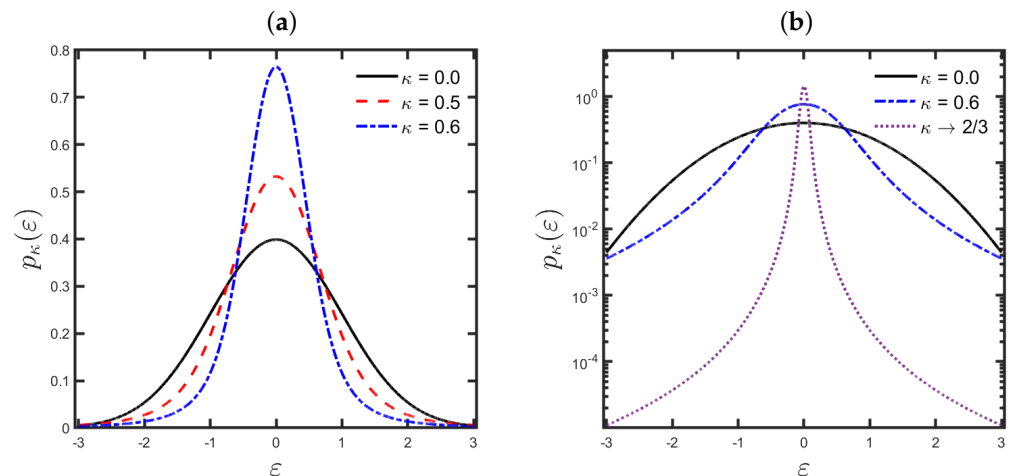


Figure 1. Probability plots of the κ -Gaussian distribution (10) for some κ -values using (a) a linear scale and (b) a linear scale on the axis of ordinates, and a logarithmic scale on the axis of abscissas. The solid black line represents the standard Gaussian distribution ($\kappa \rightarrow 0$).

Because we assume that the errors are independent and identically distributed by the power law distribution represented in (10), we can calculate the corresponding objective function by estimating the most likely state using the probabilistic maximum likelihood method:

$$\min_m \phi_\kappa(m) \equiv \max_m \mathcal{L}_\kappa(m|d^{obs}), \quad (14)$$

where $\mathcal{L}_\kappa(m|d^{obs}) := \prod_{i=1}^N p_\kappa(\varepsilon_i(m))$ represents the likelihood function. It is crucial to remember that minimizing the negative log-likelihood is the same as maximizing the likelihood function. In this way, the objective function ϕ_κ can be obtained from (14):

$$\phi_\kappa(m) \propto N \ln(Z_\kappa) - \sum_{i=1}^N \ln \left[\exp_\kappa \left(-\beta_\kappa \varepsilon_i^2(m) \right) \right] \quad (15)$$

$$\phi_\kappa(m) = - \sum_{i=1}^N \ln \left[\exp_\kappa \left(-\beta_\kappa \varepsilon_i^2(m) \right) \right], \quad (16)$$

where ϕ_κ is the κ -objective function, which converges to the classical objective function (9) in the limit $\kappa \rightarrow 0$.

The κ -objective function is not easily influenced by aberrant measurements (outliers), as it is based on κ -Gaussian criteria [49]. To demonstrate this, we compute the influence function Y related to the objective function. According to ref. [50], a statistical criterion is not robust if $Y \rightarrow \pm\infty$ under $|\varepsilon| \rightarrow \infty$, and robust (outlier-resistant) if $Y \rightarrow 0$ under $\varepsilon \rightarrow \pm\infty$. Given a model m_i , the influence function is defined by [50]:

$$Y_\kappa(m_i) := \frac{\partial \phi_\kappa(\varepsilon|m_i)}{\partial \varepsilon}, \quad (17)$$

where $\phi_\kappa(\varepsilon|m_i)$ is the κ -objective function computed from the errors ε given the model m_i . Thus, the κ -objective function generates the following influence function:

$$Y_\kappa = \frac{2\beta_\kappa \varepsilon}{\sqrt{1 + \kappa^2 \beta_\kappa^2 \varepsilon^4}} \quad (18)$$

for $0 < \kappa < 2/3$, with $Y_\kappa = Y_\kappa(m_i)$ and $\varepsilon = \varepsilon(m_i)$. We notice that, as ε tends to $\pm\infty$, the influence function associated with the κ -criterion (Y_κ) approaches to 0; the κ -objective function is then robust (outlier-resistant). Indeed, the influence function in its valid domain ($0 < \kappa < 2/3$) is proportional to $1/\varepsilon$ for large errors (suppressing these) and to ε for small errors (magnifying these). Figure 2 depicts the behavior of the κ -objective function and the associated influence function.

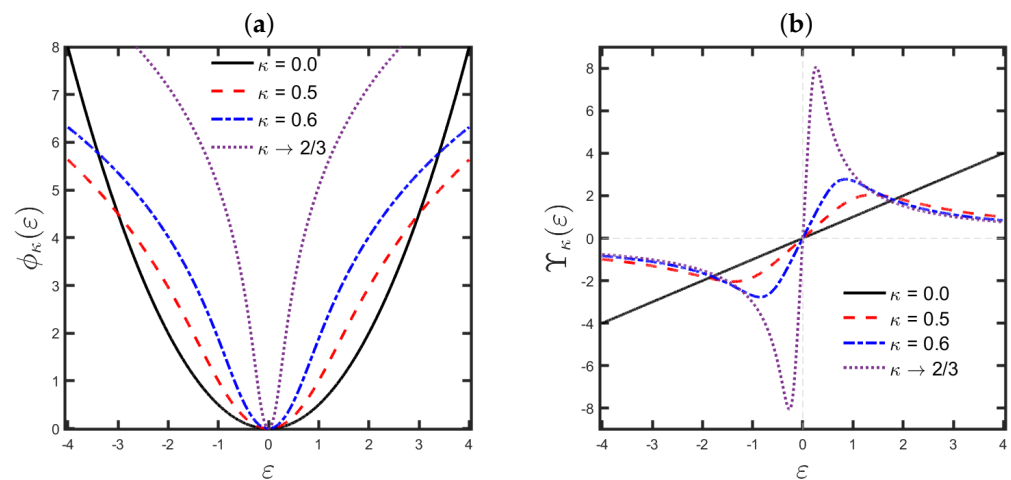


Figure 2. (a) Graphical representation of the κ -objective function (16), and (b) the associated influence function (18) for some κ -values. The solid black line represents the classical criterion ($\kappa \rightarrow 0$).

3. Optimal Transport Metric Based on Kaniadakis κ -Statistics

In 1781, Gaspard Monge first raised a challenger transportation problem [39], which consisted of moving a pile of sand from one place to another optimally and without losing mass. As formulated by Monge, the optimal transport (OT) issue is an ill-posed problem; hence the solution, if it exists, is not unique. Nearly 200 years later, Leonid Kantorovitch proposed a well-posed relaxation of Monge's OT problem in the context of optimal economic resource allocation [51]. In this regard, Kantorovich proposed what is now known as the Kantorovich–Rubinstein metric (also referred to as Wasserstein distance), which earned him the 1975 Nobel Memorial Prize in Economic Science.

The Wasserstein criterion is a metric that defines a distance between two probability distributions. Let us consider two sets of points $\Omega_1 = \{x_i; i = 1, 2, \dots, N_1\}$ and $\Omega_2 = \{y_j; j = 1, 2, \dots, N_2\}$, in which each point x_i and y_j are represented by “mass” functions, namely $\mu(x_i)$ and $v(y_j)$, respectively. Considering the mass conservation constraint ($\sum_i \mu(x_i) = \sum_j v(y_j) = 1$), we can define the κ -optimal total transport cost $\mathcal{W}_\kappa$ as [30,40]:

$$\mathcal{W}_\kappa(\mu, v) = \min_{\mathcal{T} \in \Lambda(\mu, v)} \sum_{i,j} \mathcal{T}_{i,j} \phi_{\kappa,i,j}, \quad (19)$$

where $\Lambda(\mu, v)$ denotes the set of transport maps $\mathcal{T}$ defined in

$$\Lambda(\mu, v) = \left\{ \mathcal{T}_{i,j} \geq 0, \forall (i, j); \sum_{j=1}^{N_2} \mathcal{T}_{i,j} = \mu(x_i), \forall i; \sum_{i=1}^{N_1} \mathcal{T}_{i,j} = v(y_j), \forall j \right\}. \quad (20)$$

The transport map $\mathcal{T}$ assigns how many “sand particles” from $\mu(x_i)$ should be transported to $v(y_j)$ for each pair (x_i, y_j) , while the κ -objective function maps each pair (x_i, y_j)

to $[0; +\infty]$. The Monge–Kantorovich transportation relaxed problem (19), using the Kaniadakis κ -statistics, can therefore be solved by determining an optimal transport plan $\mathcal{T}$ that minimizes the κ -optimal total transport cost $\mathcal{W}_\kappa$ from μ to ν , given an κ -objective function ϕ_κ .

Let us consider a metric space $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ formed by a set of probability measures, in which $\mathcal{X}$ and $\mathcal{Y}$ are two separable and complete metric spaces with $\mu \in \mathcal{X}$ and $\nu \in \mathcal{Y}$. From this point forward, for practical reasons, we assume that $\mathcal{X} = \mathcal{Y} \subset \mathbb{R}^N$ (with $N_1 = N_2 = N \in \mathbb{N}$). In addition, let us consider the mass distributions μ and ν represented in terms of the Dirac delta function as follows: $\mu(x) = \frac{1}{N} \sum_{l=1}^N \delta(x - u_l)$ and $\nu(y) = \frac{1}{N} \sum_{l=1}^N \delta(y - w_l)$, in which $u_l \in \Omega_1$ and $w_l \in \Omega_2$ point out the data points describing $\mu(x)$ and $\nu(x)$. In this context, we can reformulate the optimization problem in Equation (19) as follows:

$$\mathcal{W}_\kappa(\mu, \nu) = \min_{\mathcal{T}_{i,j}} - \frac{1}{N} \sum_{i,j=1}^N \mathcal{T}_{i,j} \ln \left\{ \exp_\kappa \left[-\beta_\kappa \left(\mu(x_i) - \nu(y_j) \right)^2 \right] \right\} \quad (21)$$

subject to

$$\mathcal{T}_{i,j} \geq 0, \quad \sum_{j=1}^N \mathcal{T}_{i,j} = 1, \quad \sum_{i=1}^N \mathcal{T}_{i,j} = 1. \quad (22)$$

From a practical viewpoint, we notice that solving (21) consists of obtaining an optimal transport plan that links data points from $\mathcal{P}(\mathcal{X})$ to the corresponding data points in $\mathcal{P}(\mathcal{Y})$ that minimizes the κ -optimal total transport cost $\mathcal{W}_\kappa$. Although each element of the optimal transport plan $\mathcal{T}_{i,j}$ can assume fractional values, a classic result states that the optimal solution values are integer values, specifically 0 or 1 when the constraints described in Equation (22) are considered [52,53]. Indeed, obtaining the minimum of $\mathcal{W}_\kappa$ implicates solving a combinatorial optimization issue, which can be defined as:

$$\mathcal{W}_\kappa(\mu, \nu) = \min_{\sigma \in \mathcal{S}} - \frac{1}{N} \sum_{i=1}^N \ln \left\{ \exp_\kappa \left[-\beta_\kappa \left(\mu(x_{\sigma(i)}) - \nu(y_i) \right)^2 \right] \right\}, \quad (23)$$

where σ represents a permutation solution for the linear sum assignment problem in (21) related with $\mathcal{T}$, and $\mathcal{S}(N) = \{1, 2, \dots, N\}$ is a set of permutations. Equation (23) represents the Wasserstein metric in the context of κ -Gaussian statistics.

Naturally, the Wasserstein metric based on Kaniadakis κ -statistics appreciates the advantages provided by κ -Gaussian statistics. However, this approach in this format is only valid for comparing probability distributions, which is not interesting for geophysical applications like the FWI case. This incompatibility is because seismic signals are not normalized and positive-definite quantities like probability functions.

4. Kaniadakis κ -Graph-Space Optimal Transport FWI

4.1. FWI Based on Kaniadakis κ -Gaussian Distribution

In this section, we present the main elements of FWI based on the Kaniadakis κ -Gaussian distribution, the metric explained in Section 2. The FWI is a non-linear inverse problem whose main goal consists of inferring a quantitative physical model by comparing modeled waveforms (modeled data) with measured waveforms (observed data) [7]. FWI is often formulated as a gradient-based minimization due to the computational costs, in which the model parameters are iteratively updated, from an initial model m_0 , as follows [8]:

$$m_{i+1} = m_i - \alpha_i h_\kappa(m_i) \quad \text{for } i = 0, 1, 2, \dots, N_{iter}, \quad (24)$$

where m represents the model parameter, $\alpha_i > 0$ is the so-called step length [54], N_{iter} represents the number of FWI iterations, and h_κ denotes the descent direction at the i -th iteration.

In this work, we employ a non-linear conjugate gradient optimization method based on the so-called Polak–Ribière–Polyak algorithm. In this regard, the descent direction is defined by [55,56]:

$$h_\kappa(m_i) = \begin{cases} \nabla_m \phi_\kappa(m_0) & , \text{ if } i = 0 \\ \nabla_m \phi_\kappa(m_i) + \zeta_\kappa(m_i) h_\kappa(m_{i-1}), & \text{ for } i = 1, 2, \dots, N_{iter} \end{cases} \quad (25)$$

with

$$\zeta_\kappa(m_i) = \frac{\nabla_m \phi_\kappa(m_i) (\nabla_m \phi_\kappa(m_i) - \nabla_m \phi_\kappa(m_{i-1}))}{\nabla_m \phi_\kappa(m_{i-1}) \nabla_m \phi_\kappa(m_{i-1})}, \quad (26)$$

where $\nabla_m \phi_\kappa(m)$ is the gradient of the κ -objective function.

Thus, it is remarkable that the objective function plays a crucial role in obtaining models via FWI, which is defined for our problem as (10):

$$\min_m \phi_\kappa(m) := - \sum_{s,r} \int_0^T \ln \left\{ \exp_\kappa \left[-\beta_\kappa \left(\Gamma_{s,r} \psi_s(\vec{x}, m, t) - d_{s,r}(\vec{x}_{s,r}, t) \right)^2 \right] \right\} dt, \quad (27)$$

where $\Gamma_{s,r} \psi_s = d_{s,r}^{mod}$ and $d_{s,r} = d_{s,r}^{obs}$ represent the modeled and observed data generated by the seismic source s and recorded in the receiver r , while $\vec{x} \in \mathbb{R}^2$ and $t \in [0, T]$ denote the spatial coordinates and the seismic acquisition time.

It is worth mentioning that the observed data $d_{s,r}$ are registered only in the receiver positions $\vec{x} = \vec{x}_{s,r}$, the available and chosen positions during a seismic survey. The seismic wavefield ψ_s is computed in the entire physical domain for each seismic source s by solving a wave equation. Thus, $\Gamma_{s,r}$ represents a sampling operator that acts as a measurement processor onto the receiver r from the source s . In this work, we consider the acoustic case; therefore, ψ_s are the pressure wavefields that satisfy the following model:

$$\frac{1}{c^2(\vec{x})} \frac{\partial^2 \psi_s(\vec{x}, t)}{\partial t^2} - \nabla^2 \psi_s(\vec{x}, t) = g_s(t) \delta(\vec{x} - \vec{x}_s) \quad (28)$$

where g_s represents a seismic source signature at the fixed position $\vec{x} = \vec{x}_s$, c is the P -wave velocity model of the medium, and ∇^2 denotes de Laplacian operator.

Thus, the gradient of the κ -objective function (27) with respect to the model parameters is given by:

$$\nabla_m \phi_\kappa(m) = \frac{\partial \phi_\kappa(m)}{\partial m_l} = 2\beta_\kappa \sum_{s,r} \int_0^T \frac{\mathcal{J}_{s,r}(m, t) \Delta d_{s,r}(m, t)}{\sqrt{1 + \kappa^2 \beta_\kappa^2 \Delta d_{s,r}^4(m, t)}} dt, \quad (29)$$

where $\Delta d_{s,r}(m, t) = \Gamma_{s,r} \psi_s(\vec{x}, m, t) - d_{s,r}(\vec{x}_{s,r}, t)$ represents the error (or residual data) and

$$\mathcal{J}_{s,r}(m, t) = \frac{\partial}{\partial m_l} \left(\Gamma_{s,r} \psi_s(\vec{x}, m, t) \right) \quad (30)$$

is known as the Fréchet derivative. It is worth emphasizing that FWI problems involve many elements from the model parameters that typically comprise 10^6 to 10^{12} variables (coefficients of the wave equation). In this context, we need to solve the wave equation once in the forward modeling process plus at least 10^6 times in calculating the gradient of the κ -objective function through Fréchet derivatives, being unfeasible in industrial problems.

4.2. Adjoint-State Method

Since calculating Fréchet derivatives can be computationally prohibitive, we compute the gradient of the κ -objective function using the adjoint-state method, which was developed in the 1970s [57]. There are several ways to formulate the state-adjoint approach, such as in techniques based on the augmented Lagrangian method or Green's functions. How-

ever, in this work we consider the perturbation theory to calculate the gradient efficiently. We notice that the κ -objective function can be written as:

$$\phi_\kappa(m) = f(\psi(m), m), \quad (31)$$

where ψ is a state variable that belongs to the complex space $\mathcal{Q}$; ψ satisfies the following equation of state:

$$F(\psi(m), m) = \mathcal{A}(m, t)\psi(m, t) - g(t) = 0, \quad (32)$$

in which we suppress the subscript s for the sake of a simplified notation. In the latter equation, $\mathcal{A}(m, t)\psi(t) = q(t)$ represents the wave equation written in a compact form, where $\mathcal{A}(m, t) = m \frac{\partial^2}{\partial t^2} - \nabla^2$ is the d'Alembert wave operator with $m = \frac{1}{c^2(\vec{x})}$ belonging to the real space $\mathcal{M}$, whilst $g(t) = g_s(t)\delta(\vec{x} - \vec{x}_s)$.

Suppose we consider an arbitrary variation δm concerning the model parameter m . In that case, the state variable ψ will be disturbed by a variation $\delta\psi$; consequently, the κ -objective function in Equation (31) will also be disturbed. In this way, we have to:

$$\delta\phi_\kappa = \frac{\partial f(\psi, m)}{\partial m} \delta m + \left\langle \frac{\partial f(\psi, m)}{\partial \psi_j}, \delta\psi \right\rangle_{\mathcal{Q}}, \quad (33)$$

where we only consider the first-order terms in δm and $\delta\psi$. Furthermore, ψ_j is any element of the space $\mathcal{Q}$, and $\langle \cdot, \cdot \rangle_{\mathcal{Q}}$ is the inner product in $\mathcal{Q}$.

It is worth emphasizing that the perturbations δm and $\delta\psi$ also induce variations in the equation of state (32). Moreover, assuming that there is a unique solution ψ for any model parameter m , we can state that $\psi + \delta\psi$ is the unique solution of $F(\psi + \delta\psi, m + \delta m) = 0$. In other words, for a physical realization ψ (that is, $F(\psi, m) = 0$), we have the following first-order development in δm and $\delta\psi$:

$$F(\psi + \delta\psi, m + \delta m) = F(\psi, m) + \left(\frac{\partial F(\psi, m)}{\partial m} \right) \delta m + \left(\frac{\partial F(\psi, m)}{\partial \psi_j} \right) \delta\psi = 0. \quad (34)$$

From the latter equation, we have that the perturbation in the state variable ψ is given by:

$$\delta\psi = - \left(\frac{\partial F(\psi, m)}{\partial \psi_j} \right)^{-1} \left(\frac{\partial F(\psi, m)}{\partial m} \right) \delta m, \quad (35)$$

where a^{-1} denotes the inverse of a . So, replacing the resulting from Equation (35) in Equation (33), we have an efficient way to compute the gradient of the κ -objective function without the Fréchet derivatives:

$$\delta\phi_\kappa = \frac{\partial f(\psi, m)}{\partial m} \delta m - \left\langle \frac{\partial f(\psi, m)}{\partial \psi_j}, \left(\frac{\partial F(\psi, m)}{\partial \psi_j} \right)^{-1} \left(\frac{\partial F(\psi, m)}{\partial m} \right) \delta m \right\rangle_{\mathcal{Q}}. \quad (36)$$

On the other hand, to obtain an intuitive way to calculate the gradient, Equation (36) can be rewritten so that in the inner product in $\mathcal{Q}$, one of the terms varies only with ψ and the other with m . For this, we consider the following adjoint operator property for any x and y variables:

$$\langle x, \mathcal{R}y \rangle = \langle \mathcal{R}^\dagger x, y \rangle \quad (37)$$

where $\mathcal{R}^\dagger$ is the adjoint operator of $\mathcal{R}$, while the superscript $\dagger$ represents the adjoint operation (complex-conjugate transpose). Applying the property (37) to the second term of Equation (36), we obtain:

$$\delta\phi_\kappa = \frac{\partial f(\psi, m)}{\partial m} \delta m - \left\langle \left[\left(\frac{\partial F(\psi, m)}{\partial \psi_j} \right)^{-1} \right]^\dagger \frac{\partial f(\psi, m)}{\partial \psi_j}, \left(\frac{\partial F(\psi, m)}{\partial m} \right) \delta m \right\rangle_{\mathcal{Q}}. \quad (38)$$

Furthermore, if we consider a new state variable v belonging to the complex space $\mathcal{V}$, given by:

$$v = \left[\left(\frac{\partial F(\psi, m)}{\partial \psi_j} \right)^{-1} \right]^\dagger \frac{\partial f(\psi, m)}{\partial \psi_j}, \quad (39)$$

where v is the first term of the inner product in Equation (38), we have the following equation of state:

$$\left(\frac{\partial F(\psi, m)}{\partial \psi_j} \right)^\dagger v = \frac{\partial f(\psi, m)}{\partial \psi_j}, \quad (40)$$

which is known as the adjoint-state equation [58,59], and therefore v is called the adjoint-state variable.

In summary, the calculation of the gradient of the κ -objective function through the state-adjoint method is given by:

$$\nabla_m \phi_\kappa(m) = \frac{\partial \phi_\kappa(m)}{\partial m} = \frac{\partial f(\psi, m)}{\partial m} - \left\langle v, \frac{\partial F(\psi, m)}{\partial m} \right\rangle_{\mathcal{Q}}, \quad (41)$$

where the state-adjoint variable v is calculated from the state-adjoint equation in (40). In this way, for our problem we have:

$$f(\psi_s(m, t), m) = - \sum_r \ln \left\{ \exp_\kappa \left[-\beta_\kappa \left(\Gamma_{s,r} \psi_s(m, t) - d_{s,r}(t) \right)^2 \right] \right\}, \quad (42)$$

where $\phi_\kappa(m) = \sum_s f(\psi_s(m, t), m)$. In addition, for any model parameter m , let ψ_s be a solution of the equation of state given in (32), that is, a physical realization. We obtain:

$$F(\psi_s(t), m) = \mathcal{A}(m, t) \psi_s(t) - g_s(t) = 0 \quad (43)$$

and

$$f(\psi_s(t), m) = - \sum_r \ln \left\{ \exp_\kappa \left[-\beta_\kappa \left(\Gamma_{s,r} \psi_s(t) - d_{s,r}(t) \right)^2 \right] \right\} = f(\psi_s(t)). \quad (44)$$

Therefore, we obtain the following derivatives of the equation of state:

$$\frac{\partial F(\psi_s(t), m)}{\partial m} = \frac{\partial \mathcal{A}(m, t)}{\partial m} \psi_s(t) \quad \text{and} \quad \frac{\partial F(\psi_s(t), m)}{\partial \psi_{s_j}} = \mathcal{A}(m, t), \quad (45)$$

and the following for the κ -objective function:

$$\frac{\partial f(\psi_s(t), m)}{\partial m} = 0 \quad \text{and} \quad \frac{\partial f(\psi_s(t), m)}{\partial \psi_{s_j}} = \sum_r \frac{2\beta_\kappa \Gamma_{s,r}^\dagger (\Gamma_{s,r} \psi_s(t) - d_{s,r}(t))}{\sqrt{1 + \kappa^2 \beta_\kappa^2 (\Gamma_{s,r} \psi_s(t) - d_{s,r}(t))^4}}. \quad (46)$$

Thus, the gradient of the κ -objective function via the state-adjoint method is given by substituting the derivatives calculated in Equations (45) and (46) in Equations (40) and (41), as follows:

$$\nabla_m \phi_\kappa(m) = - \sum_s \int_0^T \left\langle v_s(\vec{x}, t; \kappa), \frac{\partial^2 \psi_s(\vec{x}, t)}{\partial t^2} \right\rangle_{\vec{x}} dt \quad (47)$$

with v_s being the solution of the adjoint-wave equation given by

$$m(\vec{x}) \frac{\partial^2 v_s(\vec{x}, t)}{\partial t^2} - \nabla^2 v_s(\vec{x}, t) = \sum_r \frac{2\beta_\kappa \Gamma_{s,r}^+ (\Gamma_{s,r} \psi_s(t) - d_{s,r}(t))}{\sqrt{1 + \kappa^2 \beta_\kappa^2 (\Gamma_{s,r} \psi_s(t) - d_{s,r}(t))^4}} \quad (48)$$

where $m(\vec{x}) = \frac{1}{c^2(\vec{x})}$.

In search of a physical meaning for Equations (47) and (48), let us consider a new state variable given by $\lambda_s(\vec{x}, t) = v_s(\vec{x}, T - t)$. So, the latter equation becomes:

$$m(\vec{x}) \frac{\partial^2 \lambda_s(\vec{x}, t)}{\partial t^2} - \nabla^2 \lambda_s(\vec{x}, t) = \sum_r \frac{2\beta_\kappa \Gamma_{s,r}^+ (\Gamma_{s,r} \psi_s(T - t) - d_{s,r}(T - t))}{\sqrt{1 + \kappa^2 \beta_\kappa^2 (\Gamma_{s,r} \psi_s(T - t) - d_{s,r}(T - t))^4}}. \quad (49)$$

We notice that the adjoint-state variable λ_s is calculated in reverse time from Equation (49), i.e., starting the wave propagation from the final time T to the initial time 0. For this reason, this state-adjoint variable is commonly called the backpropagated wavefield, while Equation (49) is called the adjoint-wave equation, in which the right-hand term is named the adjoint source [59]. In this way, the κ -objective function gradient is calculated efficiently from the cross-correlation of the forward wavefield with the backpropagated wavefield.

In this context, computing the gradient via the state-adjoint method for each seismic source requires solving the wave equation only twice, first in the forward modeling and second in backpropagation modeling. We also point out that the robustness properties of the objective function discussed in Section 2 are indispensable in calculating the gradient. Indeed, the influence function (18) gives the adjoint source used in the inversion process. The particular classical case $\kappa \rightarrow 0$ provides the residual data as the adjoint source.

4.3. κ -Graph-Space Optimal Transport FWI

From a statistical point of view, non-Gaussian criteria are critical to handle noisy datasets in FWI analysis [9]. In this sense, we also consider the κ -Gaussian-based metric to deal with a challenging issue in FWI called cycle skipping [7]. Cycle skipping occurs when the initial model used in the FWI process is not kinematically accurate or lacks low-frequency contents in the analyzed dataset [34]. So, we consider the criterion based on κ -Gaussian statistics in the context of OT metric to mitigate the effects of non-Gaussian errors and cycle-skipping issues. However, let us remind the reader that the OT metric measures the distance between probability distributions, incompatible with a comparison between seismic signals, as discussed earlier. Thus, to work around this incompatibility, in this work we represent the non-normalized and oscillatory waveforms in the graph space [47]. Graphs are mathematical structures formed by ordered pairs of disjoint sets (V, E) , where V denotes the so-called vertices and E represents an edge that connects paired vertices [60].

Hence, we discretize the waveforms $d(t)$ as an ensemble of ordered pairs of the form $\{(t_i, d_i) \in \mathbb{R}^2; i = 1, 2, \dots, N\}$ with $d_i = d(t_i)$. So, the graph-transformed representation of a discretized waveform $\mathbf{d} = \{d_i; i = 1, 2, \dots, N\}$ is defined as:

$$\mathcal{G} : \mathbf{d} \rightarrow \mathcal{G}(\mathbf{d}) = d^{\mathcal{G}}(y, t) \quad \mathbb{R}^N \rightarrow \mathcal{D}(\mathbb{R}^2), \quad (50)$$

where $\mathcal{G}$ denotes the graph transformation, $d^{\mathcal{G}}(y, t)$ is the graph-transformed waveform, and $\mathcal{D}(\mathbb{R}^2)$ is a probability space on $\mathbb{R}^2$. The graph-transformed waveform is defined as:

$$d^{\mathcal{G}}(y, t) = \frac{1}{N} \sum_{i=1}^N \delta(t - t_i) \delta(y - d_i), \quad (51)$$

where y is associated with the waveform amplitude. In this way, waveforms are represented by normalized and positive quantities.

However, in many contexts like FWI, one needs to calculate the derivative of waveforms on some occasions (as explained in the previous sections); the Dirac delta function is not differentiable. Due to this, we consider a smoothed graph transformation by representing Dirac functions using κ -Gaussian distributions (10). Thus, the κ -graph-transformed representation of a discretized waveform is given by [61]:

$$\begin{aligned} \mathcal{G}_{\kappa} : \mathbf{d} &\rightarrow \mathcal{G}_{\kappa}(\mathbf{d}) = d^{\mathcal{G}_{\kappa}}(y, t) \\ \mathbb{R}^N &\rightarrow \mathbb{C}^{\infty}(\mathbb{R}, \mathbb{R}_*^+) \end{aligned} \quad (52)$$

with

$$d^{\mathcal{G}_{\kappa}}(y, t) = \frac{1}{Z_{\kappa}} \sum_{i=1}^N \exp_{\kappa} \left(-\beta_{\kappa}(t - t_i)^2 \right) \exp_{\kappa} \left(-\beta_{\kappa}(y - d_i)^2 \right), \quad (53)$$

where $\mathbb{C}^{\infty}(\mathbb{R}, \mathbb{R}_*^+)$ represents a set of strictly positive and infinitely differentiable functions. In this context, the graph-space κ -OT objective function is defined as:

$$\phi_{\mathcal{W}_{\kappa}^{\mathcal{G}_{\kappa}}}(m) := \sum_{s,r} \mathcal{C}_{\kappa} \left(d_{s,r}^{mod}, d_{s,r}^{obs} \right), \quad (54)$$

where $d_{mod,i}^{\mathcal{G}_{\kappa}} = (t_i, d_{mod,i})$ and $d_{obs,i}^{\mathcal{G}_{\kappa}} = (t_i, d_{obs,i})$, and $\mathcal{C}_{\kappa}(d^{mod}, d^{obs}) = \mathcal{W}_{\kappa}^{\mathcal{G}_{\kappa}}(d_{mod}^{\mathcal{G}_{\kappa}}, d_{obs}^{\mathcal{G}_{\kappa}})$ represents the κ -Wasserstein criterion applied to the graph-transformed seismic data. The κ -Wasserstein distance $\mathcal{W}_{\kappa}^{\mathcal{G}_{\kappa}}(d_{mod}^{\mathcal{G}_{\kappa}}, d_{obs}^{\mathcal{G}_{\kappa}}) = \mathcal{W}_{\kappa}^{\mathcal{G}_{\kappa}}$ is then computed via the following minimization task:

$$\mathcal{W}_{\kappa}^{\mathcal{G}_{\kappa}} = \min_{\sigma \in \mathcal{S}(N)} - \sum_{i=1}^{N_t} \ln \left\{ \exp_{\kappa} \left[-\beta_{\kappa} \left(t_{\sigma(i)} - t_i \right)^2 \right] \exp_{\kappa} \left[-\beta_{\kappa} \left(d_{\sigma(i)}^{mod} - d_i^{obs} \right)^2 \right] \right\}. \quad (55)$$

where N_t denotes the number of time samples for each waveform, while σ is the permutation solution for the linear sum assignment problem in (21) related with a transport map $\mathcal{T}$, and $\mathcal{S}(N) = \{1, 2, \dots, N\}$ is an ensemble of permutations. For simplicity, we multiply the κ -Wasserstein distance (23) by the scalar N in the latter equation. Indeed, optimizing $\mathcal{W}_{\kappa}$ is equivalent to optimizing the product $N \times \mathcal{W}_{\kappa}$. Equation (55) represents the FWI objective function based on κ -OT, namely, the κ -GSOT-FWI, for short, in reference to κ -Graph-Space Optimal Transport FWI.

The gradient of the κ -GSOT-objective function (55), that is, the derivative of $\mathcal{W}_{\kappa}^{\mathcal{G}_{\kappa}}$ with respect to the model parameters, is given by:

$$\nabla_m \mathcal{W}_{\kappa}^{\mathcal{G}_{\kappa}}(m) = \frac{\partial \mathcal{W}_{\kappa}^{\mathcal{G}_{\kappa}}(m)}{\partial m} = \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \sum_{i=1}^{N_t} \mathcal{J}_{s,r,i}(m) \mathcal{U}_{s,r,i}(m; \kappa), \quad (56)$$

in which

$$\mathcal{U}_{s,r,i}(m; \kappa) = \frac{2\beta_{\kappa} \left(d_{s,r,\sigma(i)}^{mod}(m) - d_{s,r,i}^{obs} \right)}{\sqrt{1 + \kappa^2 \beta_{\kappa}^2 \left(d_{s,r,\sigma(i)}^{mod}(m) - d_{s,r,i}^{obs} \right)^4}}. \quad (57)$$

is the adjoint-source related with the κ -GSOT-FWI framework, while N_s , N_r and N_t represent the number of seismic sources, receivers, and time samples used in the acquisition of seismic data.

The statistical interpretation of the residual data (error) associated with the κ -Gaussian statistics is preserved in the κ -GSOT-FWI case. The critical difference is that in the approach without OT, the waveforms are compared sample by sample. In contrast, in the κ -GSOT-FWI approach, the waveforms are analyzed more completely, comparing each time sample of the observed data with all the time samples of the modeled data in according to an optimal assignment using the permutation solution σ .

Figure 3 shows a flow chart of the FWI algorithm, which is an iterative process, which means that model updates are computed concerning the previous model as described by Equation (24). The first step, called Initial Setup, consists of introducing the input variables, i.e., the initial model, the parameters of the seismic acquisition (the positions of the sources and receivers, the seismic source signature, acquisition time). After configuring and organizing all the input variables of the FWI algorithm, modeled wavefields are obtained in the forward problem through the numerical solution of the wave Equation (28) by employing the finite difference method [62]. Then, a sampling operation ($\Gamma_{s,r}$) is carried out from the modeled wavefields ψ_s to obtain the modeled data ($d_{s,r}^{mod} = \Gamma_{s,r}\psi_s$), extracting the wavefields in the positions of the seismic acquisition receivers. After, the objective function gradient is obtained through the adjoint-state method described in Section 4.2 and used to update the model following Equation (24). Finally, the FWI algorithm checks whether the optimization process reached the pre-defined stopping criteria (which, in our case, was the maximum number of iterations equal to 50). As long as the criteria are not met, the cycle is repeated. If so, the iterative process is interrupted, and the resulting model is the one that minimizes the difference between modeled and observed data.

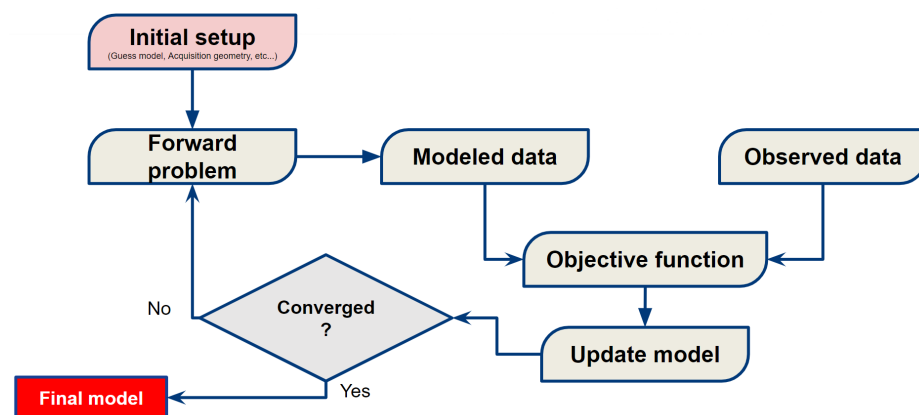


Figure 3. Flow chart of the full-waveform inversion (FWI) process.

5. Numerical Experiments

To demonstrate how the κ -GSOT-FWI deals with non-Gaussian noise and cycle-skipping issues, we carried out numerical examples involving a 2D acoustic time-domain FWI to estimate a P-wave velocity model in a typical Brazilian pre-salt oil region. Such an Earth model, namely, Chalda, represents a region with approximate dimensions 16 by 7 km in lateral distance and depth, respectively, as depicted in Figure 4a. Our problem has 720,484 unknown variables because we discretize the Chalda model in a regular grid with 12.5 m spacing, generating 562 and 1282 grid cells in the vertical and horizontal directions, respectively.

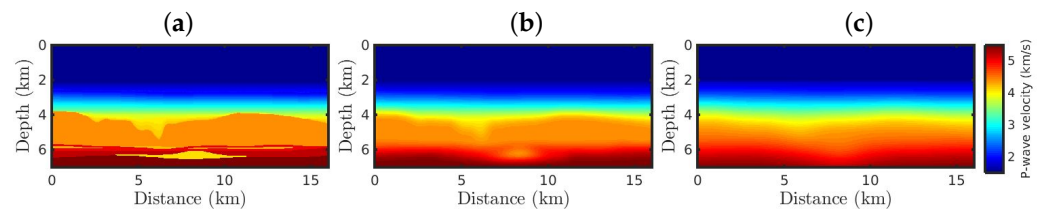


Figure 4. (a) Chalda model representing the Brazilian pre-salt oil region, used as the true model. Initial models used in the (b) first and (c) second scenarios.

In all numerical experiments, we consider a seismic survey comprising 161 seismic sources equally spaced every 75 m at 12.5 m in-depth. We employ a Ricker wavelet as a seismic source, which is mathematically described by: $f(t) = (1 - 2\pi^2\mu_p^2 t^2) \exp(-\pi^2\mu_p^2 t^2)$, in which μ_p represents the peak frequency (maximum energy in the spectrum of frequencies). Moreover, to simulate a sparse node acquisition, named the ocean bottom nodes survey, we take into account 21 receivers implanted on the ocean floor at 400 m intervals. We consider the Chalda model depicted in Figure 4a as a benchmark (or true model). Thus, we generate a seismic dataset by considering the true model, the acquisition geometry, and the finite difference method with second and eighth order approximations for time and space. In order to simulate an infinite medium, we implement the perfectly matched layer [63] absorbing boundaries for spatial discretization. We consider 7 s as the seismic acquisition time at a sampling rate of 2 ms. In addition, to simulate a realistic case, we also employ a high-pass filter on the seismic dataset to remove energy less than 2.5 Hz.

In the FWI experiments, we consider two scenarios involving different initial models to confirm the significance of our proposal. In the first one, we consider an initial model similar to the true model, which is depicted in Figure 4b. We produce such a velocity model by weakly smoothing the true model by applying a Gaussian filter with a standard deviation of 250 m. This scenario's idea is to simulate a seismic imaging process starting from a kinetically accurate model. We call this model the Good Model. In contrast, we produce the second initial model, referring to the second scenario, by applying a more severe Gaussian filter with a standard deviation of 750 m. We call this model the Bad Model. We notice that the Bad Model lacks the main structures of the true model, particularly in the pre-salt oil region, as depicted in Figure 4c. Since the Bad Model is kinematically inaccurate, it generates cycle-skipped data [34].

For each initial-model scenario, we conduct time-domain FWI by applying the classical FWI based on Gaussian statistics, and the κ -GSOT objective function (55) in the classical limit $\kappa \rightarrow 0$ and for $\kappa = 0.1, 0.3, 0.5$ and 0.6 . We consider 50 FWI iterations in all numerical experiments. To evade the so-called inversion crime, we perform the forward modeling using a different algorithm than the one used to generate the observed dataset. In this regard, our algorithm solves the forward problem using a finite difference scheme with second and fourth order approximations for time and space in a regular grid with 25 m spacing. In addition, we consider two different circumstances concerning the type of noise in the seismic dataset. First, we consider a dataset contaminated by Gaussian noise with a signal-to-noise ratio (SNR) of 20. In contrast, in the second circumstance, we consider a non-Gaussian noise from which the dataset is polluted by Gaussian noise with an SNR of 20 and a collection of spikes (erratic data or outliers) with different amplitudes. In this regard, 10% of the time samples were contaminated by outliers, where locations were randomly drawn. The spikes have intensities that range from $5P$ to $15P$ multiplied by the original waveform amplitude, where P is a standard normal random variable.

We perform our numerical simulations using a computer hosting a quad-core processor at 3.50 GHz and 256 GB RAM. Each FWI iteration takes approximately 6 min; 71.4% is associated with calculating the gradient of the objective function using the state-adjoint method described in Section 4.2, 26.7% is related to the forward modeling process (i.e., in generating the modeled data by numerically solving Equation (28)), 1.3% of this time is

dedicated to solving the combinatorial optimization problem in (55), and 0.6% is spent on the rest of the algorithm in I/O initialization and initial set-ups loading.

Figures 5 and 6 show the FWI resulting P-wave models starting from the Good Model for the Gaussian and non-Gaussian noise cases, respectively. From a visual inspection, when only Gaussian noise is considered, all resulting models are satisfactory (Figure 5) since they are very similar to the true model (Figure 4a), regardless of the κ -value. Such successful results are due to the weak Gaussian noise in the observed data simultaneously with a kinetically accurate initial model (Figure 4b).

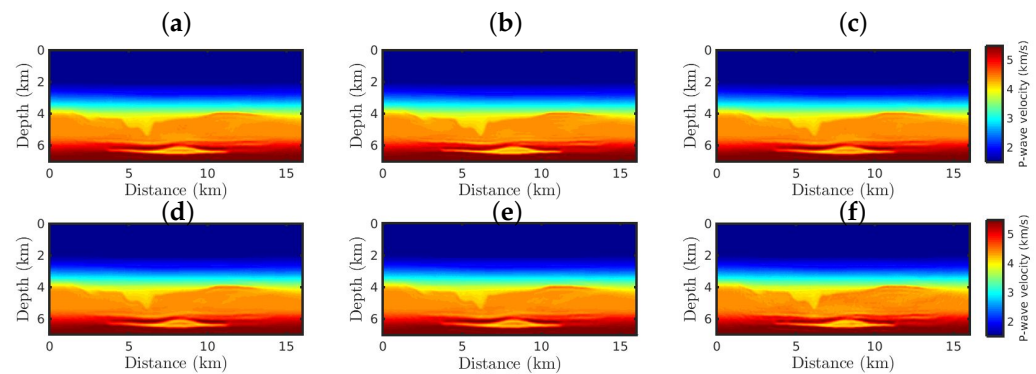


Figure 5. The resulting models starting from the Good Model for the Gaussian noise case, by employing the (a) classical FWI approach, and the κ -GSOT-FWI framework with (b) $\kappa \rightarrow 0$, (c) $\kappa = 0.1$, (d) $\kappa = 0.3$, (e) $\kappa = 0.5$, and (f) $\kappa = 0.6$.

Furthermore, we quantitatively compare our FWI resulting models with the true model by employing Pearson's correlation coefficient (R) and the normalized root-mean-square (NRMS), defined as

$$R = \frac{\text{cov}(c^{true}, c^{inv})}{\text{std}(c^{true}) \text{std}(c^{inv})} \quad \text{and} \quad \text{NRMS} = \left[\frac{\sum_i (c_i^{true} - c_i^{inv})^2}{\sum_i (c_i^{true})^2} \right]^{1/2}, \quad (58)$$

where c^{true} and c^{inv} are the true and the resulting models, while $\text{cov}(\cdot)$ and $\text{std}(\cdot)$ denote covariance and standard deviation, respectively. The R-value ranges from -1 to 1 , with -1 representing, in this context, a wrong resulting model, while 1 represents a perfect resulting model. The NRMS-value range from 0 (perfect resulting model) to ∞ (wrong resulting model).

Table 1 summarizes the comparative metrics between the true model and the P-wave velocity models resulting from the first scenario by analyzing data contaminated only by Gaussian noise. In this table, we can see that all resulting models have a low error and are strongly correlated with the true model ($R \geq 0.8$, following the strength-scale suggested by ref. [64]).

Table 1. The comparative metrics between the true model and the resulting models, depicted in Figure 1, from the first scenario in the Gaussian noise case. R represents the Pearson's correlation coefficient, while NRMS represents the normalized root-mean-square.

Strategy	κ	NRMS	R
Classical FWI	-	0.0256	0.9982
κ -GSOT-FWI	0.0	0.0272	0.9979
	0.1	0.0278	0.9979
	0.3	0.0288	0.9977
	0.5	0.0293	0.9976
	0.6	0.0277	0.9979

However, when non-Gaussian noise is considered, the classical approach fails as expected (Figure 6a). Such a wrong model is due to the classical approach being based on Gaussian statistics and sensitive to cycle-skipping issues. Figure 6b shows the resulting model from the classical GSOT-FWI, which is also based on Gaussian statistics. However, the Wasserstein metric was able to mitigate the effects of the outliers, building a satisfactory model. Nevertheless, as the κ -value increases (which means a more significant deviation from Gaussian behaviors), the κ -GSOT-FWI models present a better resolution (Figure 6c–f), especially in the deeper regions of the analyzed area. Although all κ -GSOT-FWI models are strongly correlated with the true model, the case $\kappa = 0.6$ has a higher Pearson's coefficient and a smaller NRMS error, as summarized in Table 2.

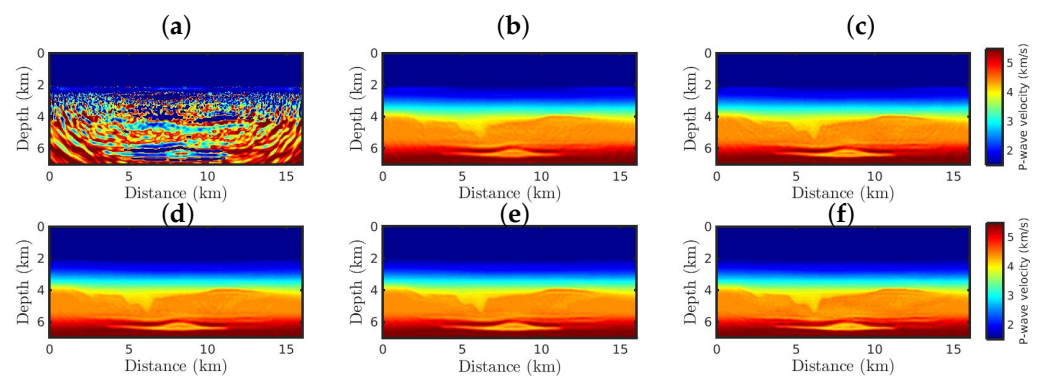


Figure 6. The resulting models starting from the Good Model for the non-Gaussian noise case, by employing the (a) classical FWI approach, and the κ -GSOT-FWI framework with (b) $\kappa \rightarrow 0$, (c) $\kappa = 0.1$, (d) $\kappa = 0.3$, (e) $\kappa = 0.5$, and (f) $\kappa = 0.6$.

Table 2. The comparative metrics between the true model and the resulting models, depicted in Figure 6, from the first scenario in the non-Gaussian noise case. R represents the Pearson's correlation coefficient, while NRMS represents the normalized root-mean-square.

Strategy	κ	NRMS	R
Classical FWI	-	0.3009	0.7627
κ -GSOT-FWI	0.0	0.0306	0.9974
	0.1	0.0296	0.9976
	0.3	0.0293	0.9976
	0.5	0.0293	0.9976
	0.6	0.0277	0.9979

Figures 7 and 8 show the FWI resulting P-wave models starting from the Bad Model for the Gaussian and non-Gaussian noise cases, respectively. From a visual inspection, it is noticeable that the classical FWI approach fails when the initial model is kinetically inaccurate, regardless of whether the data are polluted by Gaussian or non-Gaussian noise, as depicted in Figures 7a and 8a. In contrast, the FWI based on the κ -GSOT approach generates satisfactory models when Gaussian noise is considered, regardless of the κ -value (Figure 7b–f). Again, as the κ -value increases, the resulting models (Figure 7) are closer to the true model (Figure 4a), as endorsed by the statistical metrics summarized in Table 3.

Finally, in the second scenario with non-Gaussian noise, the resulting models are drastically affected by the outliers and poverty from the initial model, as depicted in Figure 8. However, when the κ -GSOT-based objective function is applied, the large geological structures of the true model are reconstructed regardless of the κ -value. However, the case $\kappa = 0.6$ reveals a P-wave velocity model (Figure 8f) that is quite accurate and comparable to the true model (Figure 4a). Likewise, the case $\kappa = 0.6$ generated a model closer to the true model, as summarized in Table 4. Indeed, in all numerical tests, the κ -GSOT-FWI for $\kappa = 0.6$ generated accurate velocity models, leading to accurate parameter estimations.

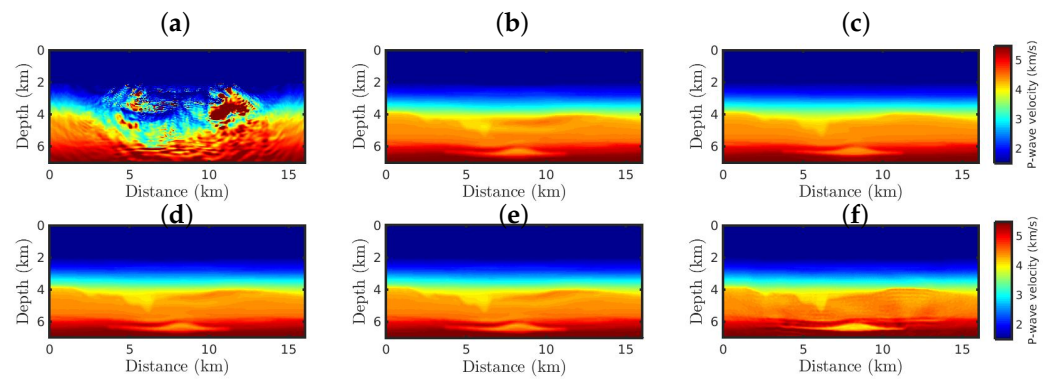


Figure 7. The resulting models starting from the Bad Model for the Gaussian noise case, by employing the (a) classical FWI approach, and the κ -GSOT-FWI framework with (b) $\kappa \rightarrow 0$, (c) $\kappa = 0.1$, (d) $\kappa = 0.3$, (e) $\kappa = 0.5$, and (f) $\kappa = 0.6$.

Table 3. The comparative metrics between the true model and the resulting models, depicted in Figure 7, from the second scenario in the Gaussian noise case. R represents the Pearson's correlation coefficient, while NRMS represents the normalized root-mean-square.

Strategy	κ	NRMS	R
Classical FWI	-	0.2947	0.7982
κ -GSOT-FWI	0.0	0.0362	0.9964
	0.1	0.0333	0.9970
	0.3	0.0363	0.9964
	0.5	0.0372	0.9962
	0.6	0.0341	0.9968

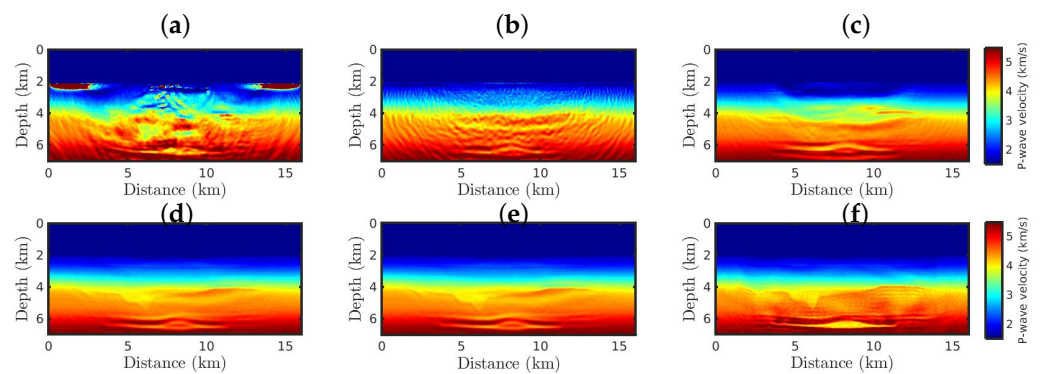


Figure 8. The resulting models starting from the Bad Model for the non-Gaussian noise case, by employing the (a) classical FWI approach, and the κ -GSOT-FWI framework with (b) $\kappa \rightarrow 0$, (c) $\kappa = 0.1$, (d) $\kappa = 0.3$, (e) $\kappa = 0.5$, and (f) $\kappa = 0.6$.

Table 4. The comparative metrics between the true model and the resulting models, depicted in Figure 8, from the second scenario in the non-Gaussian noise case. R represents the Pearson's correlation coefficient, while NRMS represents the normalized root-mean-square.

Strategy	κ	NRMS	R
Classical FWI	-	0.2715	0.7192
κ -GSOT-FWI	0.0	0.0610	0.9899
	0.1	0.0673	0.9874
	0.3	0.0564	0.9913
	0.5	0.0578	0.9908
	0.6	0.0509	0.9928

Figure 9 shows the normalized κ -GSOT-objective function decay for all numerical tests, in which panels (a) and (b) refer to the first scenario, while panels (c) and (d) correspond to the second scenario. In this regard, the left column refers to the case in which Gaussian noise is considered, and the right column is the non-Gaussian noise case. The convergence curve of the classical objective is represented by the solid black line in Figure 9. We notice that the classical objective function monotonically decays only in the most straightforward situation, where the initial model is the Good Model, and the data is contaminated by Gaussian noise (Figure 9a). In this case, the classical approach is the most indicated because the convergence rate is higher than our proposal, in addition to generating more accurate models (as summarized in Table 1). In cases where the noise is non-Gaussian or when the inversion process starts from the Bad Model, our proposal with $\kappa = 0.6$ exhibits a higher objective function decay rate (see red curves in Figure 9b–d), reconstructing P-wave velocity models closer to the true model, as summarized in Tables 2–4.

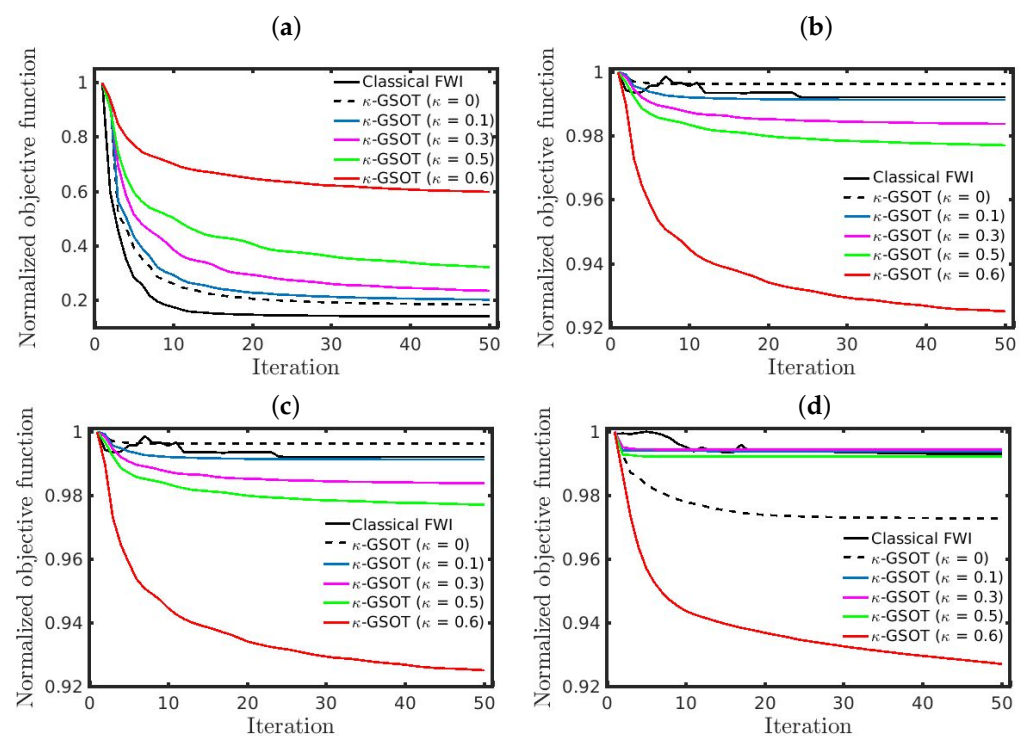


Figure 9. Convergence curves for the first scenario with (a) Gaussian noise, (b) non-Gaussian noise, and for the second scenario with (c) Gaussian noise, (d) non-Gaussian noise.

6. Final Remarks

In this work, we have examined the portability of the objective function based on the graph-space optimal transport and Kaniadakis κ -Gaussian statistics in the FWI context. In particular, we have analyzed the robustness of our proposal in mitigating two critical problems in seismic imaging via FWI, which are associated with cycle-skipping issues and the non-Gaussian nature of the errors. We have set up an objective function by employing the probabilistic maximum likelihood method for computing the most probable state using a κ -Gaussian distribution. Furthermore, we have formulated the FWI in a relaxed version of the optimal transport problem, known as the Kantorovich–Rubinstein metric or Wasserstein distance. So, we have considered the graph of the seismic data rather than the original data because the optimal transport framework is predicated on the idea that the compared entities adhere to the probability axioms. We named our proposal the κ -Graph-Space Optimal Transport FWI (or κ -GSOT-FWI, for short).

The Brazilian pre-salt case study disclosed how the κ -GSOT-FWI could be employed to deal with flawed initial models and non-Gaussian noise. The findings have demonstrated

that the classical approach is ineffective in producing accurate physical models when the initial model is crude or if the observed waveforms are contaminated by non-Gaussian errors. However, when the initial model is kinetically precise and the data well-behaved, the classical approach is the best alternative in terms of computational cost. The results also revealed that the κ -GSOT-FWI lessens the impact of phase ambiguity and non-Gaussian errors on the waveform inversion, demonstrating that our proposal is a powerful way to deal with non-linear inverse problems related to wave propagation. Moreover, we notice that the κ -GSOT-FWI produces more accurate models than those produced by classical approaches, leading to a notable improvement in objective function convergence. Additionally, our numerical experiments demonstrated that a more significant deviation from a Gaussian behavior (which in our applications was typified by the $\kappa = 0.6$ case) results in a more authentic P-wave velocity model. However, our proposal depends on the choice of a hyperparameter, which demands special investigations on how to obtain it in a real setting application. This issue should be examined in future applications.

From a practical point of view, extensive and arduous data processing is required to engineer a good initial model to alleviate phase-ambiguity issues and eliminate erratic data points. In this context, the κ -GSOT-FWI decreases the requirement of human subjectivity, which is appealing for automated techniques to analyze, for instance, recent big datasets. Thus, the κ -OT-based approach has enormous potential for dealing with modern data-centric problems. As a perspective, we intend to test our proposed methodology to analyze field data and evaluate its robustness from several initial conditions. Finally, we underline how readily our concept may be applied to a wide variety of inverse problems, ranging from estimating critical exponents of power-law distributions to modern artificial intelligence applications.

Author Contributions: Conceptualization, S.L.E.F.d.S., J.M.d.A., E.d.l.B. and G.C.; Methodology, S.L.E.F.d.S., J.M.d.A., E.d.l.B. and G.C.; Software, S.L.E.F.d.S.; Validation, S.L.E.F.d.S., J.M.d.A., E.d.l.B. and G.C.; Formal analysis, S.L.E.F.d.S., J.M.d.A., E.d.l.B. and G.C.; Investigation, S.L.E.F.d.S.; Resources, J.M.d.A. and G.C.; Data curation, S.L.E.F.d.S.; Writing—original draft, S.L.E.F.d.S., J.M.d.A., E.d.l.B. and G.C.; Supervision, G.C.; Project administration, J.M.d.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

FWI	Full-Waveform Inversion
GSOT	Graph-Space Optimal Transport
OT	Optimal Transport
MLE	Maximum Likelihood Estimation
NRMS	Normalized Root-Mean-Square

References

1. Razavy, M. *An Introduction to Inverse Problems in Physics*, 1st ed.; World Scientific: Hackensack, NJ, USA, 2020; pp. 1–388.
2. Menke, W. *Geophysical Data Analysis: Discrete Inverse Theory*, 4th ed.; Academic Press: London, UK, 2018; pp. 1–352.
3. Hanasoge, S.M.; Tromp, J. Full waveform inversion for time-distance helioseismology. *APJ* **2014**, *784*, 69. [[CrossRef](#)]
4. Guasch, L.; Calderón Agudo, O.; Tang, M.X.; Nachev, P.; Warner, M. Full-waveform inversion imaging of the human brain. *NPJ Digit. Med.* **2020**, *3*, 28. [[CrossRef](#)] [[PubMed](#)]
5. Robins, T.; Camacho, J.; Calderón Agudo, O.; Herraiz, J.L.; Guasch, L. Deep-Learning-Driven Full-Waveform Inversion for Ultrasound Breast Imaging. *Sensors* **2021**, *21*, 4570. [[CrossRef](#)] [[PubMed](#)]
6. Cao, J.; Brossier, R.; Górszczyk, A.; Métivier, L.; Virieux, J. 3-D multiparameter full-waveform inversion for ocean-bottom seismic data using an efficient fluid–solid coupled spectral-element solver. *Geophys. J. Int.* **2022**, *229*, 671–703. [[CrossRef](#)]

7. Virieux, J.; Operto, S. An overview of full-waveform inversion in exploration geophysics. *Geophysics* **2009**, *74*, WCC1–WCC26. [\[CrossRef\]](#)
8. Fichtner, A. *Full Seismic Waveform Modelling and Inversion*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 1–343.
9. Brossier, R.; Operto, S.; Virieux, J. Which data residual norm for robust elastic frequency-domain full waveform inversion? *Geophysics* **2010**, *75*, R37–R46. [\[CrossRef\]](#)
10. de Lima, I.P.; da Silva, S.L.E.F.; Corso, G.; de Araújo, J.M. Tsallis Entropy, Likelihood, and the Robust Seismic Inversion. *Entropy* **2020**, *22*, 464. [\[CrossRef\]](#)
11. Tarantola, A. *Inverse Problem Theory and Methods for Model Parameter Estimation*, 1st ed.; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2005; pp. 1–352.
12. Khokhlov, A.; Hulot, G. On the cause of the non-Gaussian distribution of residuals in geomagnetism. *Geophys. J. Int.* **2017**, *209*, 1036–1047. [\[CrossRef\]](#)
13. Elboth, T.; Reif, B.A.; Andreassen, Ø. Flow and swell noise in marine seismic data. *Geophysics* **2009**, *74*, Q17–Q25. [\[CrossRef\]](#)
14. Hlebnikov, V.; Elboth, T.; Vinje, V.; Gelius, L.-J. Noise types and their attenuation in towed marine seismic: A tutorial. *Geophysics* **2021**, *86*, W1–W19. [\[CrossRef\]](#)
15. Claerbout, J.F.; Muir, F. Robust modeling with erratic data. *Geophysics* **1973**, *38*, 826–844. [\[CrossRef\]](#)
16. Crase, E.; Pica, A.; Noble, M.; McDonald, J.; Tarantola, A. Robust elastic nonlinear waveform inversion: Application to real data. *Geophysics* **1990**, *55*, 1942–2156. [\[CrossRef\]](#)
17. Aravkin, A.Y.; Friedlander, M.P.; Herrmann, F.J.; van Leeuwen, T. Robust inversion, dimensionality reduction and randomized sampling. *Math. Program.* **2012**, *135*, 101–125. [\[CrossRef\]](#)
18. Da Silva, S.L.E.F.; da Costa, C.A.N.; Carvalho, P.T.C.; de Araújo, J.M.; dos Santos Lucena, L.; Corso, G. Robust full-waveform inversion using Q-Statistics. *Phys. A* **2020**, *548*, 124473. [\[CrossRef\]](#)
19. Silva, S.A.; da Silva, S.L.E.F.; de Souza, R.F.; Marinho, A.A.; de Araújo, J.M.; Bezerra, C.G. Improving Seismic Inversion Robustness via Deformed Jackson Gaussian. *Entropy* **2021**, *23*, 1081. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Bube, K.P.; Langan, R.T. Hybrid $L1/L2$ Minimization Appl. Tomography. *Geophysics* **1997**, *62*, 1045–1346. [\[CrossRef\]](#)
21. Guitton, A.; Symes, W.W. Robust inversion of seismic data using the Huber norm. *Geophysics* **2003**, *68*, 1126–1422. [\[CrossRef\]](#)
22. Da Silva, S.L.E.F.; Silva, R.; dos Santos Lima, G.Z.; de Araújo, J.M.; Corso, G. An outlier-resistant κ -generalized approach for robust physical parameter estimation. *Physica A* **2022**, *600*, 127554. [\[CrossRef\]](#)
23. Kaniadakis, G. Non-linear kinetics underlying generalized statistics. *Phys. A* **2001**, *296*, 405–425. [\[CrossRef\]](#)
24. Kaniadakis, G. Statistical mechanics in the context of special relativity. *Phys. Rev. E* **2002**, *66*, 056125. [\[CrossRef\]](#)
25. Kaniadakis, G.; Scarfone, A.M. A new one-parameter deformation of the exponential function. *Phys. A* **2002**, *305*, 69–75. [\[CrossRef\]](#)
26. Kaniadakis, G. Statistical mechanics in the context of special relativity. II. *Phys. Rev. E* **2005**, *72*, 036108. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Kaniadakis, G. Theoretical Foundations and Mathematical Formalism of the Power-Law Tailed Statistical Distributions. *Entropy* **2013**, *15*, 3983–4010. [\[CrossRef\]](#)
28. Da Silva, S.L.E.F.; dos Santos Lima, G.Z.; de Araújo, J.M.; Corso, G. Extensive and nonextensive statistics in seismic inversion. *Phys. A* **2021**, *563*, 125496. [\[CrossRef\]](#)
29. Wada, T.; Suyari, H. κ -generalization of Gauss' law of error. *Phys. Lett. A* **2006**, *348*, 89–93. [\[CrossRef\]](#)
30. Da Silva, S.L.E.F.; Carvalho, P.T.C.; de Araújo, J.M.; Corso, G. Full-waveform inversion based on Kaniadakis statistics. *Phys. Rev. E* **2020**, *101*, 053311. [\[CrossRef\]](#)
31. Bunks, C.; Saleck, F.M.; Zaleski, S.; Chavent, G. Multiscale seismic waveform inversion. *Geophysics* **1995**, *60*, 1457–1473. [\[CrossRef\]](#)
32. Liu, X.; Zhu, T.; Hayes, J. Critical zone structure by elastic full waveform inversion of seismic refractions in a sandstone catchment, central Pennsylvania, USA. *J. Geophys. Res. Solid Earth* **2022**, *127*, e2021JB023321. [\[CrossRef\]](#)
33. Górszczyk, A.; Brossier, R.; Métivier, L. Graph-space optimal transport concept for time-domain full-waveform inversion of ocean-bottom seismometer data: Nankai Trough velocity structure reconstructed from a 1D model. *J. Geophys. Res. Solid Earth* **2021**, *126*, e2020JB021504. [\[CrossRef\]](#)
34. Hu, W.; Chen, J.; Liu, J.; Abubakar, A. Retrieving Low Wavenumber Information in FWI: An Overview of the Cycle-Skipping Phenomenon and Solutions. *IEEE Signal Process. Mag.* **2018**, *35*, 132–141. [\[CrossRef\]](#)
35. Bozdağ, E.; Trampert, J.; Tromp, J. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurement. *Geophys. J. Int.* **2011**, *185*, 845–870. [\[CrossRef\]](#)
36. Warner, M.; Guasch, L. Adaptive waveform inversion: Theory. *Geophysics* **2016**, *81*, R429–R445. [\[CrossRef\]](#)
37. Carvalho, P.T.C.; da Silva, S.E.F.; Duarte, E.F.; Brossier, R.; Corso, G.; de Araújo, J.M. Full waveform inversion based on the non-parametric estimate of the probability distribution of the residuals. *Geophys. J. Int.* **2022**, *229*, 35–55. [\[CrossRef\]](#)
38. Métivier, L.; Brossier, R.; Méridot, Q.; Oudet, E.; Virieux, J. Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion. *Geophys. J. Int.* **2016**, *205*, 345–377. [\[CrossRef\]](#)
39. Monge, G. *Mémoire sur la Théorie des Déblais et des Remblais*, 1st ed.; Histoire de l'Académie Royale des Sciences de Paris: Paris, France, 1781; pp. 666–704.
40. Villani, C. *Optimal Transport: Old and New*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1–976.
41. Figalli, A. The Optimal Partial Transport Problem. *Arch. Ration. Mech. Anal.* **2010**, *195*, 533–560. [\[CrossRef\]](#)
42. Ambrosio, L.; Gigli, N. User's Guide to Optimal Transport. In *Modelling and Optimisation of Flows on Networks. Lecture Notes in Mathematics*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 2062, pp. 1–155.

43. Wang, S.; Stavrou, P.A.; Skoglund, M. Generalizations of Talagrand Inequality for Sinkhorn Distance Using Entropy Power Inequality. *Entropy* **2022**, *24*, 306. [\[CrossRef\]](#)
44. Messud, J.; Poncet, R.; Lambaré, G. Optimal transport in full-waveform inversion: Analysis and practice of the multidimensional Kantorovich–Rubinstein norm. *Inverse Probl.* **2021**, *36*, 065012. [\[CrossRef\]](#)
45. Sambridge, M.; Jackson, A.; Valentine, A.P. Geophysical inversion and optimal transport. *Geophys. J. Int.* **2022**, *231*, 172–198. [\[CrossRef\]](#)
46. Da Silva, S.L.E.F.; Karsou, A.; de Souza, A.; Capuzzo, F.; Costa, F.; Moreira, R.; Cetale, M. A graph-space optimal transport objective function based on q-statistics to mitigate cycle-skipping issues in FWI. *Geophys. J. Int.* **2022**, *231*, 1363–1385. [\[CrossRef\]](#)
47. Métivier, L.; Allain, A.; Brossier, R.; Mériçot, Q.; Oudet, E.; Virieux, J. Optimal transport for mitigating cycle skipping in full-waveform inversion: A graph-space transform approach. *Geophysics* **2018**, *83*, R515–R540. [\[CrossRef\]](#)
48. De Lima, J.V.T.; da Silva, S.L.E.F.; de Araújo, J.M.; Corso, G.; dos Santos Lima, G.Z. Nonextensive statistical mechanics for robust physical parameter estimation: The role of entropic index. *Eur. Phys. J. Plus* **2021**, *136*, 269. [\[CrossRef\]](#)
49. Dos Santos Lima, G.Z.; de Lima, J.V.T.; de Araújo, J.M.; Corso, G.; Da Silva, S.L.E.F. Generalized statistics: Applications to data inverse problems with outlier-resistance. *PLoS ONE* **2023**, *18*, e0282578. [\[CrossRef\]](#)
50. Hampel, F.R.; Ronchetti, E.M.; Rousseeuw, P.J.; Stahel, W.A. *Robust Statistics: The Approach Based on Influence Functions*, 1st ed.; Wiley-Interscience: Hoboken, NJ, USA, 2005; pp. 1–502.
51. Kantorovich, L. On the Translocation of Masses. *J. Manag. Sci.* **1958**, *5*, 1381–1382. [\[CrossRef\]](#)
52. Birkhoff, G. Three observations on linear algebra. *Univ. Nac. Tucumán. Rev. A* **1946**, *5*, 147–151.
53. Burkard, R.; Dell’Amico, M.; Martello, S. *Assignment Problems*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2012; pp. 1–393.
54. Nocedal, J.; Wright, S.J. *Numerical Optimization*; Springer: New York, NY, USA, 2006; pp. 1–597.
55. Polak, B.; Ribière, G. Note sur la convergence des méthodes de directions conjuguées. *Rev. Fr. Inform. Rech. Oper.* **1969**, *3*, 35–43.
56. Polyak, B.T. The conjugate gradient method in extreme problems. *USSR Comput. Math. Math. Phys.* **1969**, *9*, 94–112. [\[CrossRef\]](#)
57. Chavent, G. Identification of function parameters in partial differential equations. In *Identification of Parameter Distributed Systems*; R. E. Goodson and M. Polis: New York, NY, USA, 1974; pp. 155–156.
58. Haber, E.; Ascher, U.M. On optimization techniques for solving nonlinear inverse problems. *Inverse Probl.* **2000**, *16*, 1263–1280. [\[CrossRef\]](#)
59. Plessix, R.-E. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophys. J. Int.* **2006**, *167*, 495–503. [\[CrossRef\]](#)
60. Bollobás, B. *Modern Graph Theory, Graduate Texts in Mathematics*, 1st ed.; Springer: New York, NY, USA, 1998; pp. 1–408.
61. da Silva, S.L.E.F.; Kaniadakis, G. κ -statistics approach to optimal transport waveform inversion. *Phys. Rev. E* **2022**, *106*, 034113. [\[CrossRef\]](#)
62. Grossmann, C.; Roos, H.-G.; Stynes, M. *Numerical Treatment of Partial Differential Equations*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2007; p. 23.
63. Berenger, J.-P. A perfectly matched layer for the absorption of electromagnetic waves. *J. Comput. Phys.* **1994**, *114*, 185–200. [\[CrossRef\]](#)
64. Evans, J.D. *Straightforward Statistics for the Behavioral Sciences*, 1st ed.; Brooks/Cole Publishing Company: Monterey, CA, USA, 1996; pp. 1–600.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.