

Article

Particle Swarm Optimization and Uncertainty Assessment in Inverse Problems

José L. G. Pallero ¹ , María Zulima Fernández-Muñiz ² , Ana Cernea ² ,
Óscar Álvarez-Machancoses ² , Luis Mariano Pedruelo-González ² , Sylvain Bonvalot ^{3,4} 
and Juan Luis Fernández-Martínez ^{2,*} 

¹ ETSI en Topografía, Geodesia y Cartografía, Universidad Politécnica de Madrid, 28031 Madrid, Spain; jlg.pallero@upm.es

² Grupo de Problemas Inversos, Optimización y Aprendizaje Automático, Departamento de Matemáticas, Universidad de Oviedo, 33007 Oviedo, Spain; zulima@uniovi.es (M.Z.F.-M.); cerneadoina@uniovi.es (A.C.); oscalmac@gmail.com (Ó.Á.-M.); pedrueloluis@uniovi.es (L.M.P.-G.)

³ Laboratoire GET, Université de Toulouse, CNRS, IRD, CNES, 31400 Toulouse, France; sylvain.bonvalot@ird.fr

⁴ Bureau Gravimétrique International (BGI), 31401 Toulouse, France

* Correspondence: jlfm@uniovi.es; Tel.: +34-985-103-199

Received: 29 November 2017; Accepted: 25 January 2018; Published: 30 January 2018

Abstract: Most inverse problems in the industry (and particularly in geophysical exploration) are highly underdetermined because the number of model parameters too high to achieve accurate data predictions and because the sampling of the data space is scarce and incomplete; it is always affected by different kinds of noise. Additionally, the physics of the forward problem is a simplification of the reality. All these facts result in that the inverse problem solution is not unique; that is, there are different inverse solutions (called equivalent), compatible with the prior information that fits the observed data within similar error bounds. In the case of nonlinear inverse problems, these equivalent models are located in disconnected flat curvilinear valleys of the cost-function topography. The uncertainty analysis consists of obtaining a representation of this complex topography via different sampling methodologies. In this paper, we focus on the use of a particle swarm optimization (PSO) algorithm to sample the region of equivalence in nonlinear inverse problems. Although this methodology has a general purpose, we show its application for the uncertainty assessment of the solution of a geophysical problem concerning gravity inversion in sedimentary basins, showing that it is possible to efficiently perform this task in a sampling-while-optimizing mode. Particularly, we explain how to use and analyze the geophysical models sampled by exploratory PSO family members to infer different descriptors of nonlinear uncertainty.

Keywords: inverse problems; nonlinear inversion; noise and regularization; model reduction; uncertainty analysis; particle swarm optimization (PSO)

1. Introduction

Inverse problems belong to the class of ill-posed problems (see [1]); that is, either the solution does not exist, it is not unique, or it is unstable, because it does not depend continuously on the observed data. Uncertainty in discrete inverse problems is due to noise in data, incomplete data coverage, modeling assumptions and numerical approximations, among other causes. Traditionally, the ill-posed character of nonlinear inverse problems has been approached via local optimization methods combined with regularization techniques, but the uncertainty appraisal that is done using linearization techniques is only accurate in the neighborhood of the inverse solution (see [2]).

Typical inverse problems in geophysics are, for example, the determination of the mass distribution in the earth using gravity measurements taken at the terrain surface (see for instance [3]), finding the slowness distribution of a geological medium (see for instance [4]), the determination of the reservoir permeability field from historical data (see for instance [5]), and the characterization of the geoelectrical structure of the subsurface via resistivity imaging (see for instance [6]), to cite some of the examples our research group has investigated.

Discrete inverse problems consist of solving the following set of equations:

$$\mathbf{F}(\mathbf{m}) = \mathbf{d}^{\text{obs}} \quad (1)$$

where $\mathbf{m} \in \mathcal{M}$ is the discrete inverse model; $\mathbf{d}^{\text{obs}} \in \mathcal{D}$ is the observed data array, which belongs to the data space \mathcal{D} ; and $\mathbf{F} : \mathcal{M} \rightarrow \mathcal{D}$ represents the forward problem functional. A particular model \mathbf{m} is expressed by its coordinates in a given basis set of the model space as $\mathbf{m} = \{m_1, m_2, \dots, m_n\}$. Analogously, the observed data can be written as $\mathbf{d}^{\text{obs}} = \{d_1, d_2, \dots, d_s\}$. The model and data spaces are Euclidean spaces, to be able to measure distances between predictions and observations as well as distances between models with respect to a reference. If the forward operator \mathbf{F} corresponds to a linear application $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^s$, the inverse problem is linear. Otherwise, the inverse problem is nonlinear; that is, there is a nonlinear dependence of \mathbf{F} on \mathbf{m} .

The idea of solving an inverse problem by only determining one *best* model requires revision (see [7]). Additionally, inverse problems are a special kind of optimization problem, because the observed data enters into the cost function. Therefore, as a result of the effect of noise in data, the global optimum of the cost function is never the model that has generated the observed data, in case this model exists (see [8] for details).

In mathematical terms, the uncertainty analysis consists of finding the family M_{tol} of plausible models \mathbf{m} that fit the observed data \mathbf{d}^{obs} within a given error tolerance «tol»:

$$\mathbf{m} \in M_{\text{tol}} : \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{\text{obs}}\|_p \leq \text{tol} \quad (2)$$

where p is the norm used in the data space, usually the L_2 norm ($p = 2$). Therefore, the uncertainty analysis is related to the cost-function topography (see [2]). In the linear case, the misfit function is valley-shaped, elongated in the directions corresponding to the smaller singular values of \mathbf{F} . In the case of nonlinear problems, this topography persists (up to a first-order approximation) in the directions corresponding to the smaller singular values of the Jacobian of \mathbf{F} . These valleys adopt a curvilinear shape, which is frequently named a banana or croissant shape (see [2]), because of the dependence of the Jacobian on the discrete models \mathbf{m} . Additionally, several local minima might also coexist in the misfit landscape and are related to the nonlinear effects ignored by the Gauss–Newton approximation to the Hessian. This fact also explains why linearization techniques fail in addressing the uncertainty analysis of nonlinear inverse problems.

The structure of this paper is as follows: Section 2 briefly describes the effect of noise in data and that of the regularization for linear and nonlinear inverse problems (see [2,8–10] for a deep understanding of the topic). This analysis is important for understanding that uncertainty assessment via linearization techniques always leads to incorrect results, as these techniques are not able to analyze the nonlinear uncertainty. Section 3 is a brief presentation of the different particle swarm optimization (PSO) family members. Section 4 shows the application of this methodology to gravity inversion in sedimentary basins. Section 5 briefly talks about the use of model reduction techniques and their combination with PSO to address the solution and model appraisal in high-dimensional inverse problems. Finally, Section 6 outlines the main conclusions of this paper, aiming to convince practitioners that uncertainty analysis is a compulsory step in inversion that can be efficiently done via exploratory global optimization methods.

2. The Effect of Noise and Tikhonov's Regularization in Inverse Problems

The concept of noise is intimately related to the quality of the solution found via inversion, as a poor understanding of its effect may cause the signal, the noise and the modeling errors to be confused, introducing very important biases. The difference between noise and modeling errors is also in some situations very difficult to find. Typically, modeling errors are confounded with colored noise, while measurement noise has a random structure. In the case of nonlinear inverse problems, we denote by $c_p(\mathbf{m})$ the perturbed cost function due to the influence of noise $\delta\mathbf{d}$ in data; then we have

$$\begin{aligned} c_p(\mathbf{m}) &= \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{\text{obs}}\|_2^2 = \|\mathbf{F}(\mathbf{m}) - (\mathbf{d}^{\text{true}} - \delta\mathbf{d})\|_2^2 \\ &= \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{\text{true}}\|_2^2 + \|\delta\mathbf{d}\|_2^2 - 2\delta\mathbf{d}^T [\mathbf{F}(\mathbf{m}) - \mathbf{d}^{\text{true}}] \\ &= c(\mathbf{m}) - 2\delta\mathbf{d}^T [\mathbf{F}(\mathbf{m}) - \mathbf{d}^{\text{true}}] + \|\delta\mathbf{d}\|_2^2 \end{aligned} \quad (3)$$

where $\mathbf{d}^{\text{obs}} = \mathbf{d}^{\text{true}} + \delta\mathbf{d}$ is the sum of the true observations plus the noise. The minimum of $c_p(\mathbf{m})$ is achieved in the model \mathbf{m}_0 , where the gradient of the cost function is null: $\nabla c_p(\mathbf{m}_0) = \mathbf{0}$. $\mathbf{JF}_{\mathbf{m}_0}$ being the Jacobian of \mathbf{F} in \mathbf{m}_0 , the stationary condition implies the following (see [10] for details):

$$\nabla c(\mathbf{m}_0) = 2\mathbf{JF}_{\mathbf{m}_0}^T \delta\mathbf{d} \quad (4)$$

instead of $\nabla c(\mathbf{m}_0) = \mathbf{0}$, as would happen in the noise-free case. In the presence of data noise, the minimum of the perturbed cost function will never correspond to the minimum in the noise-free case. Therefore, uncertainty analysis is always needed to assess the presence of other plausible scenarios, which could be located in different low-misfit basins.

Additionally, using mathematical analysis, it has been proved (see [2,9,10]) that noise deforms the cost-function topography (and particularly the region of equivalence) in linear and nonlinear inverse problems. The method consists of establishing the condition for which the equivalence region for a given tolerance value has the same spatial extent under the presence of data noise with respect to the noise-free case. In the case of linear inverse problems, this deformation is homogeneous, whereas in the case of nonlinear inverse problems, it depends on the misfit region, decreasing the size of the regions with lower misfits and eventually increasing the regions of medium–high misfits. In this way, finding a model with a lower misfit becomes a harder task, but locating the region of medium misfits using search methods might become simpler. This fact has been proved via numerical experimentation in synthetic cases (see [5]).

The following two-dimensional (2D) synthetic example (presented in [10]) serves to show the influence of noise in nonlinear inverse problems. The data predictions in the different x_k points are given by

$$y_k = \alpha \left(1 - e^{-\beta x_k}\right) + \delta_k \quad (5)$$

where δ_k represents the effect of the observational noise. The inverse problem consists of identifying the parameters (α, β) from a set of data points $\{(x_1, y_1), (x_2, y_2), \dots, (x_s, y_s)\}$ that have been generated using $(\alpha_t, \beta_t) = (20, 0.1)$ as the true model. We have added two different levels of white noise to the observations ($\delta_1 \rightarrow N(0, 0.05)$ and $\delta_2 \rightarrow N(0, 0.075)$), and the nonlinear parameter identification problem has been solved using the Gauss–Newton technique with $\mathbf{m}_0 = (\alpha_0, \beta_0) = (15, 0.06)$ as the initial guess.

Figure 1 shows the results obtained for both inversions compared to the true model. The nonlinear (green line) and the linearized (black line) equivalent regions of 9% relative error are also shown. Some interesting observations can be made:

1. The typical banana shape of the region of equivalence in nonlinear inverse problems can be observed (see [2]). Additionally, the hyperquadric corresponding to the linearized region of equivalence (computed in the Gauss–Newton solution) represents the nonlinear uncertainty region only locally. The uncertainty assessment performed by taking into account the linearized

region of equivalence will produce wrong results, particularly in highly nonlinear problems with complex cost-function topographies, as the linearized region contains models that do not belong to the nonlinear region of equivalence and ignores others that do belong to this region.

2. The nonlinear and linearized equivalence regions of 9% relative misfit decrease in size as the noise level increases.
3. The solution in the noisy case might move further away from the true solution as the noise level increases, particularly in the absence of the regularization term. The procedure might converge to a solution located in another disconnected basin of the cost-function topography, depending on the reference model that is used. This also depends on the statistical distribution of the noise that is added to perturb the observed data.

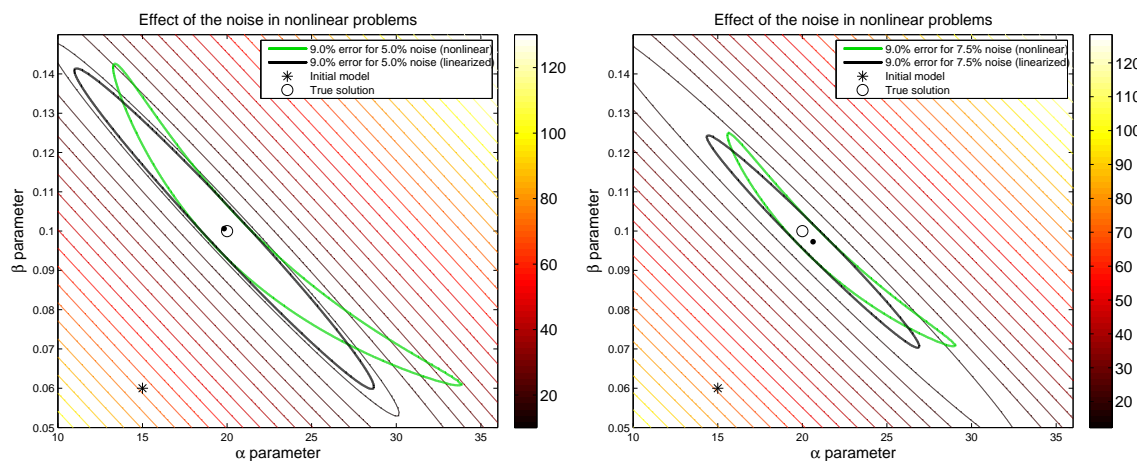


Figure 1. Synthetic numerical example. The left figure shows the effect of noise for a nonlinear problem without and with white noise, $N(0, 0.05)$. Right figure shows the case with white noise, $N(0, 0.075)$.

A similar analysis can be done to understand the effect of the regularization. As has already been explained, inverse problems are unstable, and this fact provokes dramatic perturbations of their solution when they are solved via local optimization techniques that build a numerical approximation of the inverse operator. To stabilize their solution, different types of regularization have been introduced (see [11–13]). Tikhonov and Arsenin introduced the concept of a *correctness set*, for which the inverse operator is continuous and the inverse problem becomes well-posed. They also proved that it is possible to achieve a stable pseudo-solution (also called quasi-solution) by performing minimization in the correctness set (see [14]). A way of constructing this set is to approximate the original ill-posed problem by a class of well-posed regularized problems of the following type: $\mathbf{m}_\epsilon = \mathbf{F}_\epsilon^{-1}$, where $\epsilon > 0$ is the regularization parameter, designed to approach zero. The simplest way to define \mathbf{F}_ϵ is to introduce in the regularized misfit function $c_\epsilon(\mathbf{m})$ a penalization term that takes into account the distance to a reference model \mathbf{m}^{ref} :

$$c_p(\mathbf{m}) = \|\mathbf{F}(\mathbf{m}) - \mathbf{d}^{\text{obs}}\|_p + \epsilon \|\mathbf{m} - \mathbf{m}^{\text{ref}}\|_q \tag{6}$$

where p and q represent respectively the norms adopted in the data and model spaces (usually the Euclidean norm).

The regularization induces a deformation of the linearized region of equivalence for the regularized inverse problem. This deformation is non-homogeneous and anisotropic; that is, it affects differently each of the principal directions of the region of equivalence, bounding this region more intensely along the axes with higher uncertainty. Tikhonov’s regularization might also be responsible for the introduction of local minima in the cost-function topography (see [2]).

Figure 2 presents, for the above-mentioned synthetic numerical example, the linearized equivalent region around two different models, which belongs to the nonlinear equivalent region of 9% relative

misfit (green line): $\mathbf{m}_0 = (20, 0.1)$ (black line) and $\mathbf{m}_0 = (25, 0.08)$ (red line). The parameters used in this numerical experiment were the following: the initial guess to compute the inversion coincided with the reference model adopted in the regularization, $\mathbf{m}^{\text{ref}} = (15, 0.06)$; the damping parameter was $\varepsilon^2 = 0.1$; and a noise term $\delta \rightarrow N(0, 0.05)$ was added to the synthetic data. As was previously outlined, the linearized region of equivalence only represents the nonlinear region of equivalence locally. Additionally, this analysis depends strongly on the model that it is used to calculate the Jacobian.

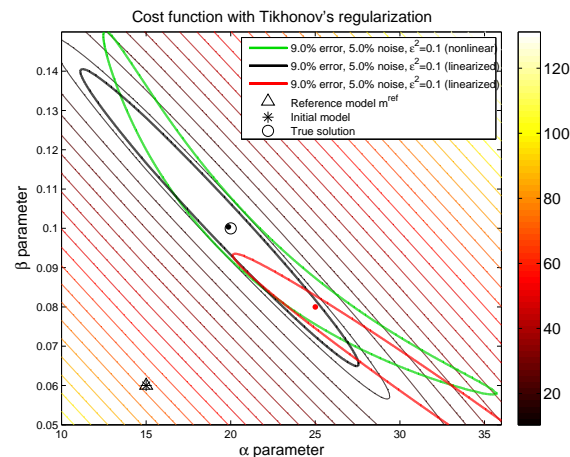


Figure 2. Synthetic numerical example. Comparison of the nonlinear and linearized regions of equivalence with Tikhonov's regularization for two different models located on the nonlinear region of equivalence; noise: $\delta \rightarrow N(0, 0.05)$, regularization parameter: $\varepsilon = 0.1$; $\mathbf{m}^{\text{ref}} = (15, 0.6)$.

3. Sampling Uncertainty via the Particle Swarm Optimization Algorithm

The solution of nonlinear problems via local optimization techniques provides geophysical models that strongly depend on the initial model that is adopted and on the kind of regularization that it is used. Particularly, the reference model used in the regularization term has a crucial effect in the inversion. These techniques are used in combination with linear uncertainty analysis to assess the quality of the solution that has been obtained. Nevertheless, as has been shown in Figures 1 and 2, the linearized uncertainty region (or linearized region of equivalence) only partially accounts for the true nonlinear uncertainty. Other possibilities concern the use of Monte Carlo methods that randomly sample the nonlinear equivalence region following a Bayesian approach (see [15]). However, these methods are not feasible for high dimensions because of the curse of dimensionality (see [16]).

The point of view followed by Tarantola [7] does not consist of the solution display supplemented by a technical description of the corresponding uncertainties, but rather by a display of as many solutions as possible consistent with the prior information and the observed data. In fact, the uncertainty analysis in inverse problems consists of sampling low-misfit elongated curvilinear valleys of the data-prediction cost function to obtain a representative set of samples from the nonlinear uncertainty region (see [2,8] for details).

Global optimization methods are a good option to sample the set of equivalent models when they are used in exploratory form. This is the case of binary genetic algorithms with a high mutation rate (see [17]) and also for different members of the PSO family (see [18]). These methods do not suffer from instability as they do not solve the inverse problem as an optimization problem, but as a search or sampling problem, using only the forward model to calculate the misfit of the different sampled models. These methods do not construct an approximation of the inverse operator. Therefore, the regularization term and the initial model are not needed. Additionally, in the case of costly forward solutions, it is crucial for these algorithms to be able to sample the region of uncertainty with a limited number of forward solves. PSO shows a good trade-off between exploration and exploitation.

3.1. The PSO Family

PSO (see [19]) is a global optimization algorithm that is inspired by the behavior of bird flocks and fish schools searching for food, for which a swarm of particles examines the space of solutions in order to optimize a given cost function. The PSO flowchart for optimization problems is as follows: the PSO particles (geophysical model parameters in our case) are represented by vectors of length equal to the optimization problem unknowns. The initial set of particles is initialized using random positions \mathbf{x}^0 and null velocities \mathbf{v}^0 , and the cost function (Equation (3)) is evaluated to compute the misfit (prediction error) of each particle in the swarm. As iterations progress, the position and velocity of each is updated as a function of its fitness and the corresponding values of its neighbors. At time-step $k + 1$, the PSO algorithm updates the individuals' positions \mathbf{x}^{k+1} and velocities \mathbf{v}^{k+1} as follows:

$$\begin{cases} \mathbf{v}_i^{k+1} = \omega \mathbf{v}_i^k + \phi_1 (\mathbf{g}^k - \mathbf{x}_i^k) + \phi_2 (\mathbf{l}_i^k - \mathbf{x}_i^k) \\ \mathbf{x}_i^{k+1} = \mathbf{x}_i^k + \mathbf{v}_i^{k+1} \end{cases} \quad (7)$$

In Equation (7), \mathbf{g}^k is the global best position in the whole swarm, \mathbf{l}_i^k is the i th particle's best position, $\phi_1 = r_1 a_g$ and $\phi_2 = r_2 a_l$ are random accelerations (global and local, respectively), and ω is the inertia weight (a real constant); r_1 and r_2 are random variables uniformly distributed in $(0, 1)$ to weight the global and local acceleration constants a_g and a_l , but other statistical distributions could also be used. The PSO parameter tuning consists of providing proper values for the inertia constant ω and for the local and global accelerations a_g and a_l , to achieve the exploration of the low-misfit regions of the cost function.

PSO can be interpreted as a double stochastic gradient in the model space. It is a special case of the generalized PSO (GPSO) algorithm (see [20]), which was developed following the analogy of a damped mass–spring scheme with unit mass, damping factor $1 - \omega$, and total stiffness constant $\phi = \phi_1 + \phi_2$. A full family of particle swarm optimizers (CC-PSO, CP-PSO, PP-PSO and RR-PSO) was also developed on the basis of this analogy using different discretizations for the velocity and the accelerations (see [21–23] for details). The stochastic stability analysis of the PSO trajectories was performed in [20,21,24] and served to establish the relationship between PSO convergence and the first- and second-order stability of the trajectories of the particles considered as stochastic processes. Suitable PSO parameters (ω , a_g and a_l) are located in the neighborhood of the upper border of the second-order stability region for each member of the PSO family. This result is also true for any statistical distribution of the PSO parameters (see [25]). The cloud versions of these algorithms are based on this mathematical analysis (see [18]). In the cloud versions, no parameter tuning (inertia, local and global accelerations) is needed, as each particle has its own parameters that are randomly selected from a set of PSO parameters that are located in the neighborhood of the upper limit of their second-order stability regions. Additionally, in the cloud design, each particle has its corresponding time-step Δt value. The exploration of the search space increases when Δt is greater than 1.0. Conversely, the algorithm freezes the solution found when Δt is decreased to values lower than 1.0.

The case of the RR-PSO algorithm is slightly different, as it has been numerically shown that good parameter sets are located along the line $\bar{\phi} = 3(\omega - 3/2)$, mainly for inertia values $\omega > 2$ (see [22]). This straight line does not depend on the cost function and remains invariant when the number of parameters increases. Additionally, this line is located in a region of medium attenuation and very high frequency for the particle trajectories. This feature confers to RR-PSO a good equilibrium between exploration and exploitation, allowing for an efficient and exploratory search.

The numerical experiments using different well-known benchmark functions have shown that the best-performing algorithm of the PSO family is RR-PSO (see [22]). CP-PSO is the most exploratory family member. Therefore, it is an interesting choice for performing nonlinear uncertainty analysis and exploring the cost-function topography efficiently. PP-PSO has the same velocity update as GPSO, but the positions of the particles are in time t instead of $t + 1$. PP-PSO has a more exploratory character than GPSO but a lower convergence rate. Finally, CC-PSO has showed the fastest convergence rate.

In any case, it is always convenient to check their results with a new problem to select the more appropriate version. In general terms, all the PSO family members provide excellent results as long as the parameter tuning is correct, and this is related to the second-order stability of the trajectories. This way, no fancy or strange mechanisms are needed to avoid the misunderstood phenomenon of premature convergence. All these mathematical results make PSO a very unique algorithm, different in any case from other heuristic approaches (see [26,27]). Additionally, the PSO family members are able to provide a set of representative samples of the nonlinear region of equivalence, which can be used to infer an approximate posterior of the model parameters in nonlinear inverse problems much faster than Monte Carlo methods and much more realistically than linear analysis techniques combined with local optimization methods.

3.2. Application of the PSO Family to Geophysical Inversion

As an example of this methodology, we show the application to gravity inversion in geophysical exploration. The use of PSO in applied geophysics was introduced in [18,28–30]. Particularly, Fernández-Martínez et al. presented in [18] the application of GPSO, CC-PSO and CP-PSO to the solution and appraisal of a 1D-DC (Vertical Electric Sounding, VES) resistivity inverse problem, successfully comparing the results with Metropolis–Hastings. In gravity inversion, PSO has been used in synthetic examples to compare with other global and local search methods without performing an uncertainty assessment (see [31–33]). This way of proceeding is not correct, as we have already conveniently explained. PSO was used for the first time in gravity inversion applied to basement depth determination in sedimentary basins in [3,34].

Prior to the PSO sampling, we need to specify the lower and upper bounds of the parameter search space, the number of particles and iterations, and the basis set in which the sampling will be performed. The search space should be designed according to the geological prior knowledge that we have at disposal for the model parameters; that is, the algorithm will explore the portion of the cost function bounded by the search limits. In some problems, the observations themselves can be used to estimate such limits. For example, the upper and lower depth limits of the search space in basement depth gravity inversion can be computed using Bouguer's plate formulation (see [3,34]). The search space design can be used also to impose absolute constraints on some parameters and reduce their uncertainty and the trade-offs with others. This information can come from boreholes or from other geophysical techniques. The use of all possible sources of prior information is a key concept in inverse problems. Nevertheless, all these pieces of information must be carefully examined to avoid inconsistencies (see [7]).

The number of particles depends on the computational cost needed to solve any individual forward problem (see [5]). Nevertheless, when a maximum number of forward solves can be achieved, it is more convenient to increase the number of iterations than the number of particles in each iteration. Despite this, only a fast numerical experimentation can help to establish the optimum trade-off between these two parameters.

Regarding the stopping criteria in this sampling procedure, the PSO algorithm is designed to finish by iterations. Nevertheless, the PSO design has to avoid the collapse of the swarm towards a unique particle that would introduce a bias in the posterior analysis. For that reason, it is very important to monitor the dispersion of the swarm with iterations. In each iteration, the dispersion can be calculated through the median distance between the particles and their center of gravity with respect to the median distance in the first iteration, where the dispersion is considered to be 100%. In the case that this collapse occurs, there are several alternatives to avoid the oversampling of the misfit region: (1) to stop the sampling and finish the PSO execution; (2) to increase the exploration by increasing the time-step or introducing repulsive forces (see [18]); (3) to consider in the posterior analysis all the collapsed particles as a unique posterior model (see [5]). It is important to understand that the PSO sampling-while-optimizing procedure is just a numerical approximation of the Bayesian posterior. Finally, the selection of the basis set to achieve the sampling is explained at the end of the paper.

4. Uncertainty Analysis Using the PSO Sampling; Application in Gravimetry

Uncertainty analysis consists of obtaining a representative sample of the region M_{tol} to perform posterior statistics by simple counting, so as to adopt risk decisions. PSO obtains a collection of models from the nonlinear uncertainty region. All these models explain the observations with an error level lower than the prescribed tolerance (see Equation (2)). The rest (those that do not fulfill the misfit criterion) come from different regions of higher misfit. Therefore, the use of a global search algorithm in their exploratory form provides a more general overview about the equivalent solutions of the inverse problem than local optimization methods. According to Tarantola (see [7]) and Scales and Snieder (see [35]), this collection of models can be considered itself the inverse problem solution, as all of the models explain the observations at the prescribed error level and are prior compatible. The model with minimum misfit is another model in this set, and, as has been explained in the section concerning the effect of noise, it is the only model that could not have generated the observed data. Additionally, it can be shown via synthetic experiments that the true model will belong to the nonlinear region of uncertainty and will have a higher misfit. This fact suggests the need of performing a correct uncertainty analysis.

4.1. Gravity Inversion in Sedimentary Basins

Estimation, using surface gravity measurements, of the vertical separation interface between two media that have different densities (where the upper, embedded in the lower, has a density $\rho_u < \rho_l$, as is common in sedimentary basins) is a nonlinear inverse problem. Gravity inversion is a classical tool, widely used in geophysics, oil and gas exploration, hydrogeology, glaciology, and so forth (see [36–40]). The gravity inverse problem, as a potential field method, has an inherent non-uniqueness in the solution, which leads to an infinite number of mass distributions generating the same gravitational signal (see, e.g., [14,41–43]). It is therefore mandatory to introduce some kind of constraints and take into account external information (as borehole data, seismic profile information, contrasted prior models, etc.) in order to bound the set of plausible solutions.

Once the gravity is observed and corrected (specific gravimeter corrections, time-dependent signals, atmospheric effects, etc.), the measurements are transformed into anomalies Δg through the classic and well-known free-air, Bouguer's plate and terrain reductions, and the subtraction of the normal gravity field signal (see [38,40], for example). These values, called complete Bouguer's anomalies, contain information about the crust–mantle interface and the mass distribution in a wide range of depths for the earth's crust; thus their influence is considered as regional. Although some sedimentary basins whose area covers thousands of square kilometers exist, their depths are commonly small compared to the earth's crust thickness; thus for a correct problem interpretation, it is necessary to isolate the part of Δg generated only by the sediments. This part is commonly called the local or residual anomaly, while the part of the signal due to the regional and deep mass distributions is known as the regional trend. These components can be isolated as a previous step or during the inversion process itself. A variety of methods exist to perform this task (see [44–46] for some examples).

Residual anomalies Δg on sedimentary basins usually have negative values, as the anomalous mass derived from a contrast density between the sediments and the basement has a negative sign: $\Delta g = \rho_s - \rho_b < 0$. These anomalies are related to the basin depth (sediments layer thickness), and they constitute the fundamental observable for gravity inversion. This signal might have several contributions:

1. The anomaly due to the sediments themselves.
2. Observational noise, which is always present.
3. Modeling errors introduced by Bouguer's anomaly correction, among others.
4. Modeling errors introduced by the regional trend suppression.

Noise and modeling errors introduced in points 2, 3 and 4 contribute to the lack of uniqueness in the solution and therefore to the uncertainty problem. These contributions should be properly taken into account in the interpretation of the results.

The most commonly used techniques for the solution of this inverse problem are based on local optimization methods (see [47], for example) or even the sequential solution of the forward problem (see [48]). Global search methods have also been proposed for this particular problem in [49], although the uncertainty assessment of the solution was not performed. For basin modeling, several alternatives have been suggested (for both 2D and three-dimensional (3D) models), such as, for example, its approximation as regular geometric bodies (see [50]), irregular polygons (see [51]), or the discretization in rectangles and prisms for 2D and 3D models, respectively (see [47,52]). Regarding the density contrast $\Delta\rho$ between the sediments and the basement, the nonlinear approach for the problem, in which the interface depths are the unknowns, imply that $\Delta\rho$ must be postulated as a constant value, or being variable with depth.

4.2. Argelès-Gazost Basin Gravity Inversion

As an example, we refer in this section to the gravity inverse problem in sedimentary basins in three dimensions [34]. In this work, the basement depth of the Argelès-Gazost basin, an ancient glacial valley located in the French Pyrenees, was determined via PSO using gravity anomalies measured on the terrain's surface. Figure 3 (left) shows the approximate limits of the basin (blue line) and the points at which the gravity anomaly was measured (a total of 117 points, 69 of them located inside the basin limits). The basin was divided into 300 prisms of $225\text{ m} \times 255\text{ m}$ in horizontal coordinates. The model parameters of the inverse problem were the depths of the basement in this grid. A density contrast between the sediments and the basement of $\Delta\rho = -600\text{ kg m}^{-3}$ was used, and for the PSO execution (PP-PSO algorithm produced the best results) a swarm of 200 models and 200 iterations was employed; thus 40,000 forward models were solved, and the execution took less than 5 min on a personal computer (Intel Core i7-4800MQ 2.70 GHz running Matlab R2017a on a Debian GNU/Linux operating system). Figure 3 (right) shows the computed gravity anomaly (upper image) corresponding to the best model and the best model itself, that is, the basement depths (lower image). The best model had a relative error of 9.72%. The first step in the uncertainty analysis is the selection of the sampled models whose misfits are below a given error tolerance (working error bound); that is, the uncertainty analysis is done only with the PSO sampled models located in the low-misfit regions of the cost function. In this case, the error tolerance used was 15%, and PSO was able to sample 25,429 models within this region.

The first approach consisted of establishing a minimum–maximum range for each parameter, taking into account the geophysical models that have been sampled within the 15% nonlinear uncertainty region. Figure 4 (left) shows a West–East profile (that which includes the deepest point, where the profile number corresponds to the I prism index in Figure 3, right lower) of the Argelès-Gazost basin. It can be observed that the best PSO model had a misfit of 9.72%; the deepest point reached a depth from the surface of about 332.7 m, and it belonged to a range of 288–379 m.

The second approach consisted of calculating the mean (or median) and standard deviation (or interquartile range) for each geophysical parameter, by averaging the above-mentioned geophysical models. This way, a “most-probable mean model” can be determined, together with some confidence intervals. Clearly, these results are approximate in nonlinear inversion, as the posterior distribution of the model parameters is typically non-Gaussian and/or multimodal. Figure 4 shows the best model found by PSO with the 15% nonlinear equivalence region (left) and the mean model with its 95% confidence interval (right). The misfit of the mean model (11.5%) was higher than that of the best model, but it was located inside the 15% tolerance region. The deepest point was located at 317 m below the surface, and its 95% confidence interval was 300–334 m. These are two different ways of visualizing the uncertainty.

Additionally, different percentiles for the deepest point can be determined from the set of sampled models. Figure 5 (left) shows the empirical cumulative density function corresponding to the deepest prism in the Argelès-Gazost basin. The inferred parameter distribution is multimodal, as can be seen clearly in Figure 5 (right). This fact is an example of the complex cost-function landscape of this kind

of inverse problem and the limited validity of a unique geophysical model found by local optimization. This information is crucial to assess risk.

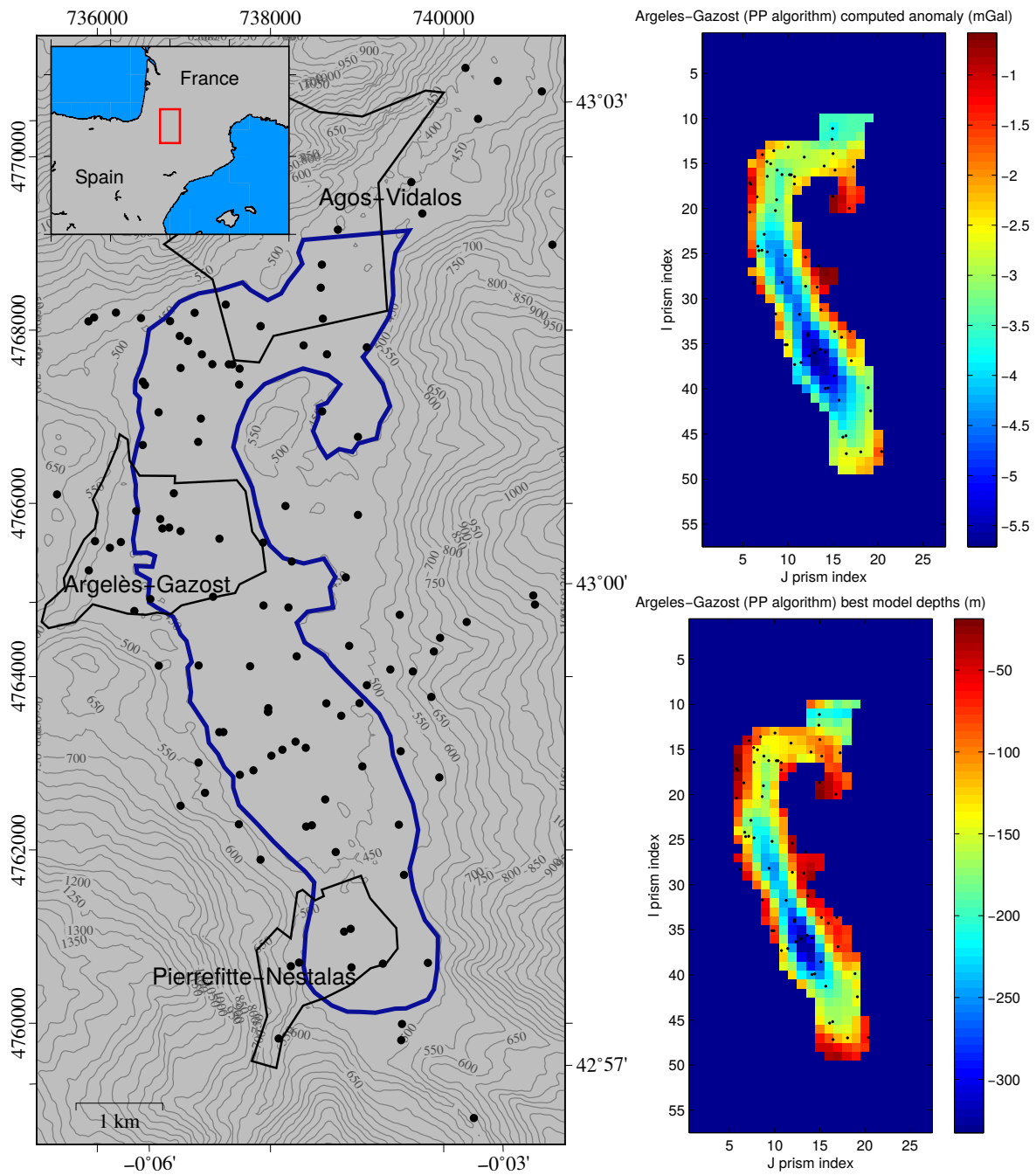


Figure 3. Left: Contour of the Argeles-Gazost basin (blue line) and observational points (black dots). Upper right: Computed gravity anomaly of the best model after the particle swarm optimization (PSO) execution. Lower right: Best depth model estimated via PSO (PP algorithm). Figure from [34].

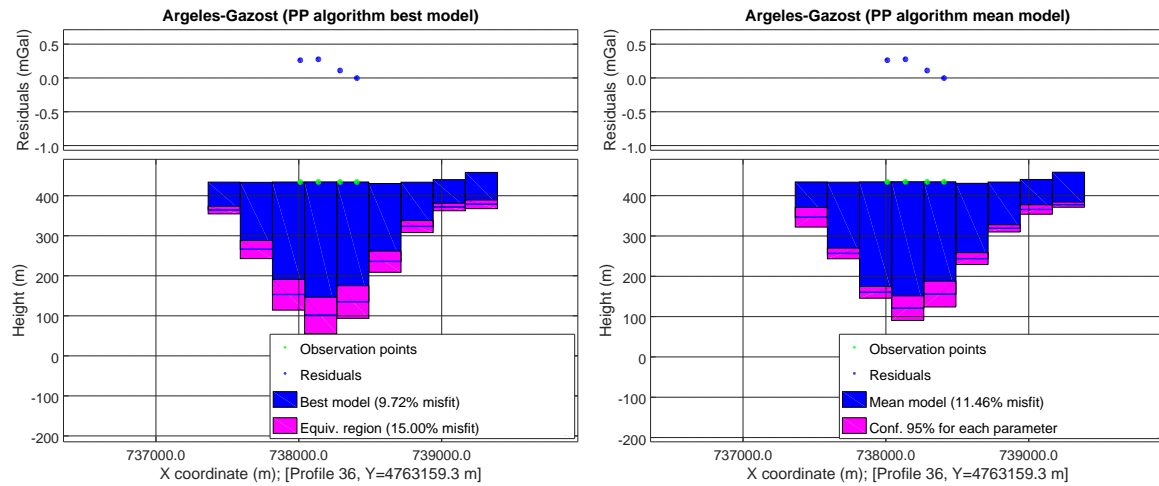


Figure 4. Left: West–East profile of the Argelès–Gazost basin containing the deepest point, belonging to the best particle swarm optimization (PSO) model (PP algorithm) and equivalent region of 15%. Right: Mean model corresponding to the same profile and confidence intervals of 95% for each parameter.

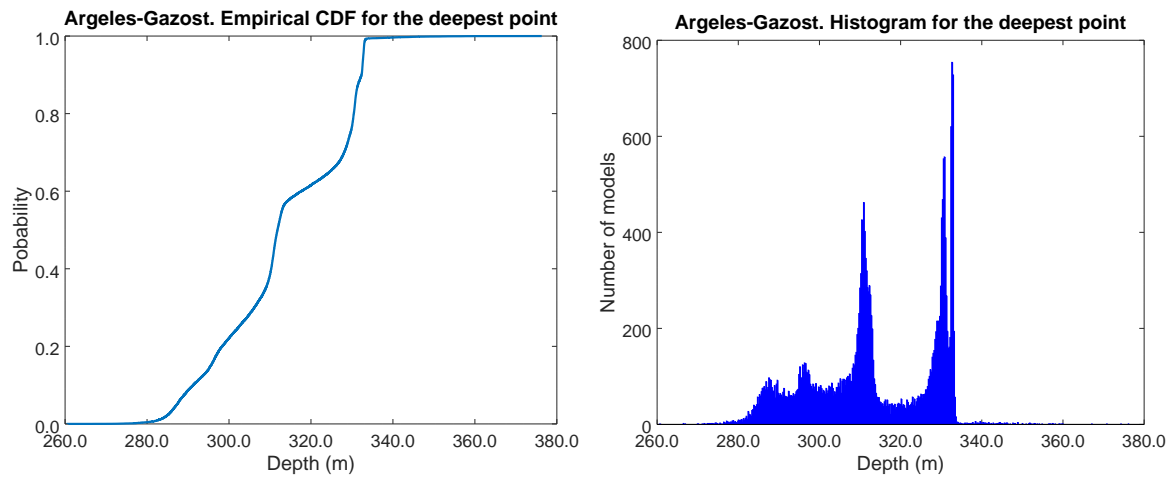


Figure 5. Argelès–Gazost basin. Left: Empirical cumulative density function corresponding to the deepest prism. Right: Absolute histogram for the deepest prism.

5. PSO Combined with Model Reduction Techniques

PSO can be used for sampling using different basis sets. This fact depends on the number of model parameters of the inverse problem. It has to be understood that the number of parameters of the inverse and forward problems does not have to be the same, as model reduction techniques can be used for sampling purposes. In general terms, when the number of parameters is lower than 100, the sampling can be performed in the original basis set, although the curse of dimensionality indicates that it is impossible to efficiently sample a model space with more than 10 independent dimensions. Clearly, this is done by informing the search space. This strategy was used for instance in [3,30] using approximate solution methods. For very high dimensional inverse problems, it is compulsory to reduce the dimension by changing the basis set in which the sampling is performed. The reason is as follows: a highly underdetermined linear system transforms to an overdetermined linear problem if the number of basis terms in which the solution is found is smaller than the number of observed data points. Model reduction also serves to simplify the topography of the cost function and can be interpreted as a way of reducing uncertainty to the set of plausible models (see [53]).

The basis set can be established on the basis of a set of prior scenarios that are generated via conditional simulation. Then, their spectral basis can be established using principal component

analysis (PCA) to avoid redundancy and reduce the dimension. This strategy was used in reservoir history matching (see [5]). The model reduction can also be based on the spectral decomposition of a model that is already located in the low-misfit region via local optimization methods. This strategy was followed in the electrical tomography problem in two dimensions (see [6]). In this case, the spectral basis set was established via the singular value decomposition (SVD), and the PSO sampling was performed in this reduced space. A high-frequency term can be added to the expansion to improve the quality of the sampling. In this case, it is recommended to have a reduced space whose dimension is lower than 10, which is the limit established by the curse of dimensionality in an isotropic space.

6. Conclusions

Inverse problems in the industry usually have a very underdetermined character. Particularly, in geophysical inversion, it is common to deal with data that is insufficient in quantity and is affected by noise, which together with the modeling and numerical errors cause the inverse problem solution to have an intrinsic uncertainty. Uncertainty always exists, and it is a synonymous of ambiguity; that is, the observed data never contain enough information to uniquely determine a true unique inverse solution. Nevertheless, practitioners tend to minimize the importance of uncertainty in inverse solutions, relying on the facts that (1) uncertainty always has a random structure that is impossible to be known; (2) regularization techniques provoke the vanishing of the equivalent solutions; and (3) linear analysis techniques provide a fairly good approximation of the real uncertainty. These three assumptions are false and might motivate that the decisions taken on the basis of this kind of analysis are erroneous, as a consequence inducing large economic losses.

Therefore, a proper nonlinear uncertainty analysis is always needed and should be understood in a very pragmatic way, as its final end is to be able to unravel plausible scenarios that were not and cannot be unravelled by the inversion. Global search methods offer a good alternative to Bayesian sampling methods when they are correctly designed. The most important feature is to be able to explore the nonlinear uncertainty region. We have shown some examples in gravimetry inversion, and we have referred to some of our previous work that used these methodologies. The final aim of this paper is to convince practitioners that no further inversions without their corresponding uncertainty analysis should be accepted.

Author Contributions: Juan Luis Fernández-Martínez and José L. G. Pallero conceived the paper; Juan Luis Fernández-Martínez, José L. G. Pallero, María Zulima Fernández-Muñiz and Luis Mariano Pedruelo-González worked on the analysis of the effect of the noise and Tikhonov's regularization in inverse problems; Juan Luis Fernández-Martínez, María Zulima Fernández-Muñiz and Ana Cernea worked on the development of the PSO algorithms; José L. G. Pallero and Juan Luis Fernández-Martínez perform the gravity inverse problem experiment; Sylvain Bonvalot provided the gravity data and interpreted the results under the geophysical point of view; José L. G. Pallero, Juan Luis Fernández-Martínez and Óscar Álvarez-Machancoses performed the statistical analysis, and Juan Luis Fernández-Martínez developed the section of model reduction techniques. Juan Luis Fernández-Martínez and José L. G. Pallero wrote the paper. All authors have read and approved the final manuscript. José L. G. Pallero completed his PhD at the Group of Inverse Problems, Optimization and Machine Learning under the supervision of Juan Luis Fernández-Martínez

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

References

1. Hadamard, J. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton Univ. Bull.* **1902**, *13*, 49–52. (In French)
2. Fernández-Martínez, J.L.; Fernández-Muñiz, M.Z.; Tompkins, M.J. On the topography of the cost functional in linear and nonlinear inverse problems. *Geophysics* **2012**, *77*, W1–W15.
3. Pallero, J.L.G.; Fernández-Martínez, J.; Bonvalot, S.; Fudym, O. Gravity inversion and uncertainty assessment of basement relief via Particle Swarm Optimization. *J. Appl. Geophys.* **2015**, *116*, 180–191.
4. Fernández-Martínez, J.L.; Pedruelo-González, L.M. Anisotropic Mean Traveltime Curves: A Method to Estimate Anisotropic Parameters from 2D Transmission Tomographic Data. *Math. Geosci.* **2009**, *41*, 163–192.

5. Fernández-Martínez, J.L.; Mukerji, T.; García-Gonzalo, E.; Suman, A. Reservoir characterization and inversion uncertainty via a family of particle swarm optimizers. *Geophysics* **2012**, *77*, M1–M16.
6. Fernández-Martínez, J.L.; Shu, S.; Sirieix, C.; Fernández-Muñiz, Z.; Riss, J. Uncertainty analysis and probabilistic segmentation of electrical resistivity images: The 2D inverse problem. *Geophys. Prospect.* **2017**, *65*, 112–130.
7. Tarantola, A. Popper, Bayes and the inverse problem. *Nat. Phys.* **2006**, *2*, 492–494.
8. Fernández-Martínez, J.L.; Fernández-Muñiz, Z.; Pallero, J.L.G.; Pedruelo-González, L.M. From Bayes to Tarantola: New insights to understand uncertainty in inverse problems. *J. Appl. Geophys.* **2013**, *98*, 62–72.
9. Fernández-Martínez, J.L.; Pallero, J.L.G.; Fernández-Muñiz, Z.; Pedruelo-González, L.M. The effect of noise and Tikhonov's regularization in inverse problems. Part I: The linear case. *J. Appl. Geophys.* **2014**, *108*, 176–185.
10. Fernández-Martínez, J.L.; Pallero, J.L.G.; Fernández-Muñiz, Z.; Pedruelo-González, L.M. The effect of noise and Tikhonov's regularization in inverse problems. Part II: The nonlinear case. *J. Appl. Geophys.* **2014**, *108*, 186–193.
11. Tikhonov, A.N.; Arsenin, V.Y. *Solutions of Ill-Posed Problems*; V. H. Winston & Sons: Washington, DC, USA, 1977.
12. Engl, H.W.; Hanke, M.; Neubauer, A. *Regularization of Inverse Problems*; Springer: Berlin, Germany, 1996.
13. Kaltenbacher, B.; Neubauer, A.; Scherzer, O. *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*; Walter de Gruyter: Berlin, Germany, 2008.
14. Zhdanov, M.S. *Geophysical Inverse Theory and Regularization Problems*, 1st ed.; Elsevier: Amsterdam, The Netherlands, 2002.
15. Scales, J.A.; Tenorio, L. Prior information and uncertainty in inverse problems. *Geophysics* **2001**, *66*, 389–397.
16. Curtis, A.; Lomax, A. Prior information, sampling distributions, and the curse of dimensionality. *Geophysics* **2001**, *66*, 372–378.
17. Fernández-Álvarez, J.P.; Fernández-Martínez, J.L.; Menéndez-Pérez, C.O. Feasibility Analysis of the Use of Binary Genetic Algorithms as Importance Samplers Application to a 1-D DC Resistivity Inverse Problem. *Math. Geosci.* **2008**, *40*, 375–408.
18. Fernández-Martínez, J.L.; García-Gonzalo, E.; Fernández Álvarez, J.P.; Kuzma, H.A.; Menéndez Pérez, C.O. PSO: A powerful algorithm to solve geophysical inverse problems. Application to a 1D-DC resistivity case. *J. Appl. Geophys.* **2010**, *71*, 13–25.
19. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks, Perth, Australia, 27 November–1 December 1995; pp. 1942–1948.
20. Fernández-Martínez, J.L.; García-Gonzalo, E. The Generalized PSO: A New Door for PSO Evolution. *J. Artif. Evol. Appl.* **2008**, *2008*, doi:10.1155/2008/861275.
21. Fernández-Martínez, J.L.; García-Gonzalo, E. The PSO family: Deduction, stochastic analysis and comparison. *Swarm Intell.* **2009**, *3*, 245–273.
22. Fernández-Martínez, J.L.; García-Gonzalo, E. Stochastic Stability and Numerical Analysis of Two Novel Algorithms of PSO Family: PP-PSO and RR-PSO. *Int. J. Artif. Intell. Tools* **2012**, *21*, 1240011.
23. García-Gonzalo, E.; Fernández-Martínez, J.L.; Cernea, A. Four-Points Particle Swarm Optimization Algorithms. *J. Mult.-Valued Log. Soft Comput.* **2014**, *22*, 239–266.
24. Fernández-Martínez, J.L.; García-Gonzalo, E. Stochastic Stability Analysis of the Linear Continuous and Discrete PSO Models. *IEEE Trans. Evol. Comput.* **2011**, *15*, 405–423.
25. García-Gonzalo, E.; Fernández-Martínez, J.L. Convergence and stochastic stability analysis of particle swarm optimization variants with generic parameter distributions. *Appl. Math. Comput.* **2014**, *249*, 286–302.
26. García-Gonzalo, E.; Fernández-Martínez, J.L. A Brief Historical Review of Particle Swarm Optimization (PSO). *J. Bioinform. Intell. Control* **2012**, *1*, 3–16.
27. Fernández-Martínez, J.L.; García-Gonzalo, E. Particle swarm optimisation: Time for uniformisation. *Int. J. Comput. Sci. Math.* **2013**, *4*, 16–33.
28. Fernández-Álvarez, J.P.; Fernández-Martínez, J.L.; García-Gonzalo, E.; Menéndez-Pérez, C.O. Application of the particle swarm optimization algorithm to the solution and appraisal of the vertical electrical sounding inverse problem. In Proceedings of the 11th Annual Conference of the International Association of Mathematical Geology (IAMG '06), Liège, Belgium, 3–8 September 2006.
29. Shaw, R.; Srivastava, S. Particle swarm optimization: A new tool to invert geophysical data. *Geophysics* **2007**, *72*, F75–F83.

30. Fernández-Martínez, J.L.; García-Gonzalo, E.; Naudet, V. Particle swarm optimization applied to solving and appraising the streaming-potential inverse problem. *Geophysics* **2010**, *75*, WA3–WA15.
31. Yuan, S.; Wang, S.; Tian, N. Swarm intelligence optimization and its application in geophysical data inversion. *Appl. Geophys.* **2009**, *6*, 166–174.
32. Toushmalani, R. Gravity inversion of a fault by Particle Swarm Optimization (PSO). *SpringerPlus* **2013**, *2*, 315.
33. Toushmalani, R. Comparison result of inversion of gravity data of a fault by particle swarm optimization and Levenberg-Marquardt methods. *SpringerPlus* **2013**, *2*, 462.
34. Pallero, J.L.G.; Fernández-Martínez, J.L.; Bonvalot, S.; Fudym, O. 3D gravity inversion and uncertainty assessment of basement relief via Particle Swarm Optimization. *J. Appl. Geophys.* **2017**, *139*, 338–350.
35. Scales, J.A.; Snieder, R. The Anatomy of Inverse Problems. *Geophysics* **2000**, *65*, 1708–1710.
36. Dobrin, M.B. *Introduction to Geophysical Prospecting*, 2nd ed.; McGraw-Hill: New York, NY, USA, 1960.
37. Nettleton, L.L. *Gravity and Magnetism in Oil Prospecting*; McGraw-Hill: New York, NY, USA, 1976.
38. Telford, W.M.; Geldart, L.P.; Sheriff, R.E.; Keys, D.A. *Applied Geophysics*; Cambridge University Press: Cambridge, UK, 1976.
39. Parker, R.L. *Geophysical Inverse Theory*; Princeton University Press: Cambridge, UK, 1994.
40. Blakely, R.J. *Potential Theory in Gravity and Magnetic Applications*; Cambridge University Press: Cambridge, UK, 1995.
41. Skeels, D.C. Ambiguity in gravity interpretation. *Geophysics* **1947**, *12*, 43–56.
42. Al-Chalabi, M. Some studies relating to non uniqueness in gravity and magnetic inverse problems. *Geophysics* **1971**, *36*, 835–855.
43. Hinze, W.J.; von Frese, R.R.B.; Saad, A.H. *Gravity and Magnetic Exploration. Principles, Practices and Applications*, 1st ed.; Cambridge University Press: Cambridge, UK, 2013.
44. Beltrão, J.F.; Silva, J.B.C.; Costa, J.C. Robust polynomial fitting method for regional gravity estimation. *Geophysics* **1991**, *56*, 80–89.
45. Mickus, K.L.; Aiken, C.L.V.; Kennedy, W.D. Regional-residual gravity anomaly separation using the minimum-curvature technique. *Geophysics* **1991**, *56*, 279–283.
46. Xu, Y.; Hao, T.; Li, Z.; Duan, Q.; Zhang, L. Regional gravity anomaly separation using wavelet transform and spectrum analysis. *J. Geophys. Eng.* **2009**, *6*, 279–287.
47. Barbosa, V.C.F.; Silva, J.B.C.; Medeiros, W.E. Gravity inversion of basement relief using approximate equality constraints on depths. *Geophysics* **1997**, *62*, 1745–1757.
48. Chen, Z.; Zhang, X.M.S. 3D gravity interface inversion constrained by a few points and its GPU acceleration. *Comput. Geosci.* **2015**, *84*, 20–28.
49. Parker, P.B. Genetic Algorithms and Their Use in Geophysical Problems. Ph.D. Thesis, University of California at Berkeley, Berkeley, CA, USA, 1999.
50. Rao, D.B. Analysis of gravity anomalies of sedimentary basins by an asymmetrical trapezoidal model with quadratic density function. *Geophysics* **1990**, *55*, 226–231.
51. Al-Chalabi, M. Interpretation of gravity anomalies by non-linear optimisation. *Geophys. Prospect.* **1972**, *20*, 1–16.
52. Silva, J.B.C.; Costa, D.C.L.; Barbosa, V.C.F. Gravity inversion of basement relief and estimation of density contrast variation with depth. *Geophysics* **2006**, *71*, J51–J58.
53. Fernández-Martínez, J.L. Model reduction and uncertainty analysis in inverse problems. *Lead. Edge* **2015**, *34*, 1006–1016.

