

# Posterior Sampling using Particle Swarm Optimizers and Model Reduction Techniques

J. L. Fernández Martínez, Stanford University, University of California-Berkeley, USA and University of Oviedo, Spain

E. García Gonzalo, University of Oviedo, Spain

Z. Fernández Muñiz, University of Oviedo, Spain

G. Mariethoz, Stanford University, USA

T. Mukerji, Stanford University, USA

---

## ABSTRACT

Inverse problems are ill-posed and posterior sampling is a way of providing an estimate of the uncertainty based on a finite set of the family of models that fit the observed data within the same tolerance. Monte Carlo methods are used for this purpose but are highly inefficient. Global optimization methods address the inverse problem as a sampling problem, particularly Particle Swarm, which is a very interesting algorithm that is typically used in an exploitative form. Although PSO has not been designed originally to perform importance sampling, the authors show practical applications in the domain of environmental geophysics, where it provides a proxy for the posterior distribution when it is used in its explorative form. Finally, this paper presents a hydrogeological example how to perform a similar task for inverse problems in high dimensional spaces through the combined use with model reduction techniques.

**Keywords:** High Dimensional Spaces, Inverse Problems, Model Reduction Techniques, Particle Swarm, Posterior Sampling

---

## PARTICLE SWARM OPTIMIZATION (PSO) APPLIED TO INVERSE PROBLEMS

Particle swarm optimization is a stochastic evolutionary computation technique inspired by the social behavior of individuals (called

particles) in nature, such as bird flocking and fish schooling (Kennedy & Eberhart, 1995).

Let us consider an inverse problem of the form  $\mathbf{F}(\mathbf{m}) = \mathbf{d}$ , where  $\mathbf{m} \in \mathbf{M} \subset \mathbb{R}^n$  are the model parameters,  $\mathbf{d} \in \mathbb{R}^s$  the discrete observed data, and

$$\mathbf{F}(\mathbf{m}) = (f_1(\mathbf{m}), f_2(\mathbf{m}), \dots, f_s(\mathbf{m}))$$

DOI: 10.4018/jaec.2010070102

---

is the vector field representing the forward operator and  $f_j(\mathbf{m})$  is the scalar field that accounts for the  $j$ -th data. Inverse problems are very important in science and technology and sometimes referred to as, parameter identification, reverse modeling, etc. The “classical” goal of inversion given a particular data set (often affected by noise), is to find a unique set of parameters  $\mathbf{m}$ , such the data prediction error  $\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_p$  in a certain norm  $p$ , is minimized.

The PSO algorithm to approach this inverse problem is at first glance very easy to understand and implement:

1. A prismatic space of admissible models,  $\mathbf{M}$ , is defined:

$$l_j \leq m_{ij} \leq u_j, \quad 1 \leq j \leq n, \quad 1 \leq i \leq N_{\text{size}}$$

where  $l_j$ ,  $u_j$  are the lower and upper limits for the  $j$ -th coordinate of each particle in the swarm,  $n$  is the number of parameters in the optimization problem and  $N_{\text{size}}$  is the swarm size.

2. The misfit for each particle of the swarm is calculated,  $\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_p$  and for each particle its local best position found so far (called  $\mathbf{l}_i(k)$ ) is determined as well as the minimum of all of them, called the global best ( $\mathbf{g}(k)$ ).
3. The algorithm updates at each iteration the positions  $\mathbf{x}_i(k)$  and velocities  $\mathbf{v}_i(k)$  of each model in the swarm. The velocity of each particle  $i$  at each iteration  $k$  is a function of three major components:
  - a) The inertia term, which consists of the old velocity of the particle,  $\mathbf{v}_i(k)$  weighted by a real constant,  $\omega$ , called inertia.
  - b) The social learning term, which is the difference between the global best position found so far (called  $\mathbf{g}(k)$ )

and the particle’s current position ( $\mathbf{x}_i(k)$ ).

- c) The cognitive learning term, which is the difference between the particle’s best position (called  $\mathbf{l}_i(k)$ ) and the particle’s current position ( $\mathbf{x}_i(k)$ ):

$$\mathbf{v}_i(k+1) = \omega \mathbf{v}_i(k) + \phi_1(\mathbf{g}(k) -$$

$$\mathbf{x}_i(k)) + \phi_2(\mathbf{l}_i(k) - \mathbf{x}_i(k))$$

$$\mathbf{x}_i(k+1) = \mathbf{x}_i(k) + \mathbf{v}_i(k+1)$$

$\omega$ ,  $a_g$ ,  $a_l$  are the PSO parameters: inertia and local and global acceleration constants;  $\phi_1 = r_1 a_g$ ,  $\phi_2 = r_2 a_l$  are the stochastic global and local accelerations, and  $r_1$ ,  $r_2$  are vectors of random numbers uniformly distributed in  $(0, 1)$  to weight the global and local acceleration constants. In the classical PSO algorithm these parameters are the same for all the particles in the swarm.

In an inverse problem, the positions are the coordinates of the model  $\mathbf{m}$  in the search space, and the velocities represent the perturbations needed to find the low misfit models.

The PSO algorithm can be physically interpreted as a particular discretization of a stochastic damped mass-spring system (Martínez et al., 2008; Martínez & Gonzalo, 2008):

$$\begin{aligned} \mathbf{x}_i''(t) + (1 - \omega) \mathbf{x}_i'(t) + (\phi_1 + \phi_2) \mathbf{x}_i(t) = \\ \phi_1 \mathbf{g}(t - t_0) + \phi_2 \mathbf{l}_i(t - t_0) \end{aligned} \quad (1)$$

This model has been addressed as the PSO continuous model since it describes (together with the initial conditions) the continuous movement of any particle coordinate in the swarm  $\mathbf{x}_i(t)$ , where  $i$  stands for the particle index, and  $\mathbf{g}(t)$  and  $\mathbf{l}_i(t)$  are its local and global attractors. In model (1) the trajectories are allowed to be delayed a time  $t_0$  with respect

to the trajectories. Using this physical analogy we were able to analyze the PSO particle's trajectories (Martínez et al., 2008) and to explain the success in achieving convergence of some popular parameters sets found in the literature (Carlisle & Dozier, 2001; Clerc & Kennedy, 2002; Trelea, 2003). Also we derived a whole family of PSO algorithms (Martínez & Gonzalo, 2009; Gonzalo & Martínez, 2009) considering different differences schemes for  $\mathbf{x}_i''(t)$

and  $\mathbf{x}_i'(t)$ :

1. GPSO or centered-regressive PSO  $t_0 = 0$

$$\begin{aligned} v(t + \Delta t) &= (1 - (1 - \omega)\Delta t)v(t) + \\ &\quad \phi_1\Delta t(g(t) - x(t)) + \phi_2\Delta t(l(t) - x(t)) \\ x(t + \Delta t) &= x(t) + v(t + \Delta t)\Delta t. \end{aligned}$$

The GPSO algorithm is the generalization of the PSO algorithm for any time step  $\Delta t$ , (PSO is the particular case for  $\Delta t = 1$ ). These expressions for the velocity and position are obtained by employing a regressive scheme in velocity and a centered scheme in acceleration.

2. CC-PSO or centered - centered PSO ( $t_0 = 0$ ).

$$\begin{aligned} x(t + \Delta t) &= x(t) + \\ &\quad \left[ \frac{2 + (\omega - 1)\Delta t}{2}v(t) + \phi_1 \frac{\Delta t}{2}(g(t) - x(t)) + \phi_2 \frac{\Delta t}{2}(l(t) - x(t)) \right] \Delta t, \\ v(t + \Delta t) &= \frac{2 + (\omega - 1)\Delta t}{2 + (1 - \omega)\Delta t}v(t) + \\ &\quad \frac{\Delta t}{2 + (1 - \omega)\Delta t} \sum_{k=0}^1 \left[ \phi_1(l(t + k\Delta t) - x(t + k\Delta t)) \right. \\ &\quad \left. + \phi_2(g(t + k\Delta t) - x(t + k\Delta t)) \right] \end{aligned}$$

3. CP-PSO or centered-progressive PSO ( $t_0 = \Delta t$ ).

$$\begin{aligned} v(t + \Delta t) &= \\ &\quad \frac{(1 - \phi\Delta t^2)v(t) + \phi_1\Delta t(g(t) - x(t)) + \phi_2\Delta t(l(t) - x(t))}{1 + (1 - \omega)\Delta t}, \\ x(t + \Delta t) &= x(t) + v(t)\Delta t \end{aligned}$$

4. PP-PSO or progressive-progressive PSO ( $t_0 = 0$ ).

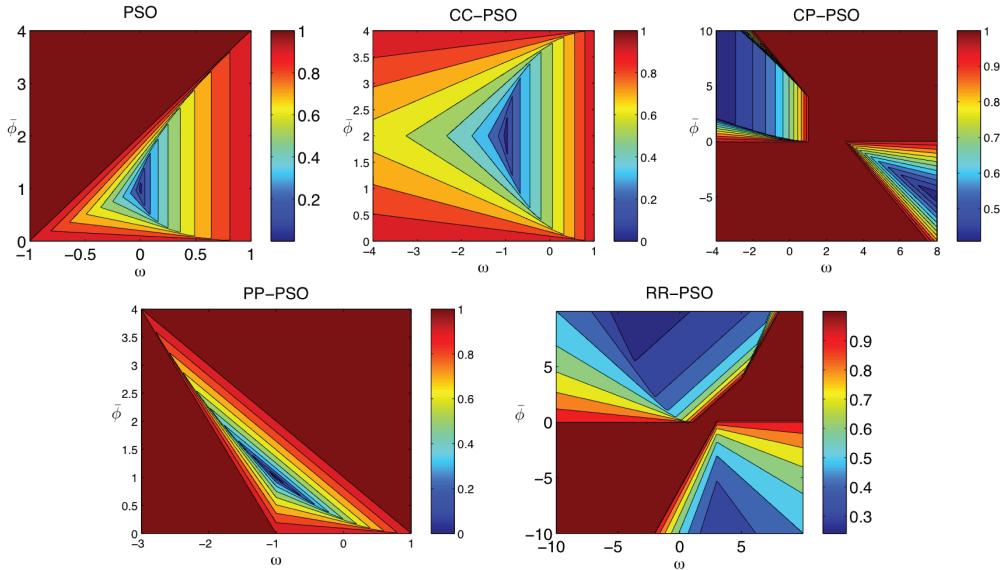
$$\begin{aligned} v(t + \Delta t) &= (1 - (1 - \omega)\Delta t)v(t) + \\ &\quad \phi_1\Delta t(g(t) - x(t)) + \phi_2\Delta t(l(t) - x(t)), \\ x(t + \Delta t) &= x(t) + v(t)\Delta t \end{aligned}$$

5. RR-PSO or regressive-regressive PSO ( $t_0 = \Delta t$ )-

$$\begin{aligned} v(t + \Delta t) &= \frac{v(t) + \phi_1\Delta t(g(t) - x(t)) + \phi_2\Delta t(l(t) - x(t))}{1 + (1 - \omega)\Delta t + \phi\Delta t^2}, \\ x(t + \Delta t) &= x(t) + v(t + \Delta t)\Delta t. \end{aligned}$$

Comparing these versions it is also possible to observe that GPSO and RR-PSO act as integrators since first they update the velocity of the particle and thereafter its position. CC-PSO is a differentiator and uses two consecutive centers of attraction. Finally CP-PSO and PP-PSO update at the same time the position and velocity. Also in the case of CP-PSO and RR-PSO the centre of attraction and the trajectories are delayed. Thus, it is logical to conclude that these versions have the greater exploratory capabilities.

Also, we performed the full stochastic analysis of the PSO continuous and discrete models (GPSO) (Martínez & Gonzalo, 2009b; Martínez et al., 2010). This analysis served to investigate the GPSO second order trajectories, to show the convergence of the discrete versions (GPSO) to the continuous PSO model as the discretization time step goes to zero, and to explain the role of the cost function on the first and second order continuous and discrete dynamical systems.

*Figure 1. First order stability regions for different family members and corresponding spectral radii*

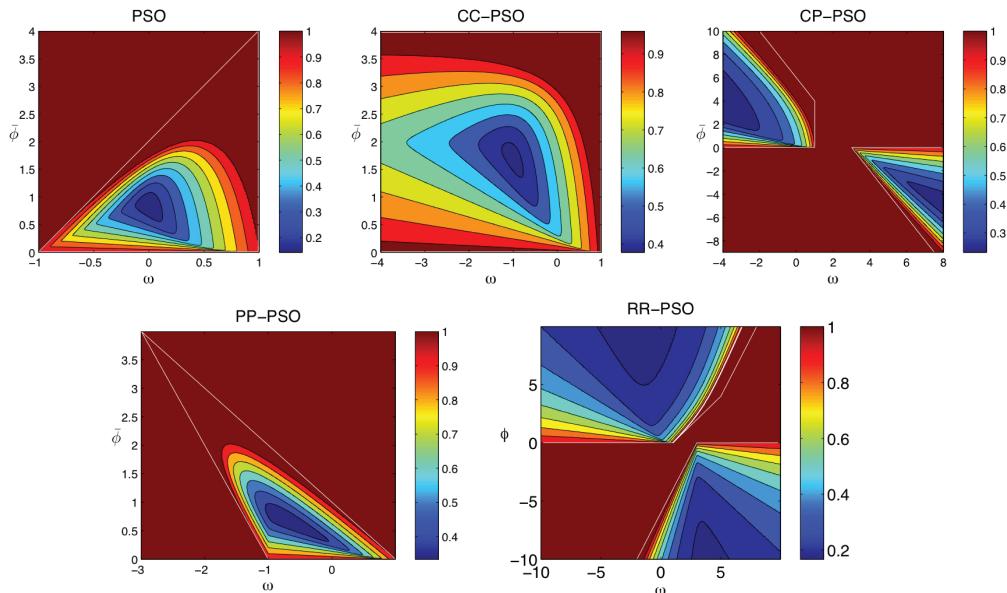
The consistency of the different PSO family members has been related to the stability of their first and second order trajectories (Martínez et al., 2008; Martínez & Gonzalo, 2009). Figure 1 shows the first order stability regions for all family members and the corresponding spectral radii that control the attenuation of the mean trajectories. The type of mean trajectories depends on the character of the eigenvalues of the first order difference equation. Basically there are four kinds of trajectories: damped oscillatory in the complex eigenvalue region, symmetrically and asymmetrically zigzagging in the regions of negative real eigenvalues and almost monotonous decreasing character in the region of positive real eigenvalues. Maximum exploration is reached in the complex region.

The second order trajectories (Martínez & Gonzalo, 2009) show a similar kind of behavior. Similar to Figure 1, Figure 2 shows the second order stability regions for all family members and the corresponding spectral radii. The second order spectral radius controls the rate of attenuation of the second order moments of the particle trajectories (variance and temporal covariance between  $x(t)$  and  $x(t + \Delta t)$ )

). The second order moment (variance) becomes unbounded for points located outside the boundaries of the second order stability zone. This feature provides a more exploratory capability but can also affect negatively the convergence rate. Easy functions to optimize such as de Jong-f4 have the convergence zone inside the second order stability region, but difficult functions as Rosenbrock and/or Griewank benefit need points with higher exploration capabilities to succeed in locating the low misfit region. Other factor that has been proved important is the frequency of oscillation of the second order trajectories. In the RR-PSO algorithm the points with the best convergence have a high frequency oscillatory character for the second order trajectories..

These results have been confirmed by numerical experiments with different benchmark functions in several dimensions. Figure 3 shows for each family member the contour plots of the misfit error (in logarithmic scale) after a certain number of iterations (300) for the Rosenbrock function. Figure 4 shows similar results for the Griewank function.

*Figure 2. Second order stability regions for different family members and corresponding spectral radii*



Other benchmark functions can be used but basically these two functions show the numerical complexities that we are interested in for real inverse problems: minima located in very flat valleys and/or surrounded by many other local minima. This numerical analysis is

done for a lattice of  $(\omega, \bar{\phi})$  points located in the corresponding first order stability regions over 50 different simulations. For GPSO, CC-PSO and CP-PSO, better parameter sets  $\omega, a_g, a_l$  are located on the first order complex region, close to the upper border of the second order stability region (where the variance of the trajectories becomes unbounded) and around the intersection to the median lines of the first stability regions (where the temporal covariance between trajectories is close to zero) (Martínez & Gonzalo, 2009). The PP-PSO does not converge for  $\omega < 0$ , and the good parameter sets are in the complex region close to the limit of second order stability and to  $\bar{\phi} = 0$ . This result can be partially altered when the velocities are

clamped or the time step is decreased. The good parameters sets for the RR-PSO are concen-

trated around the line  $\bar{\phi} = 3(\omega - 3/2)$ , mainly for inertia values greater than two. This line is located in a zone of medium attenuation and high frequency of trajectories.

## WHY EXPLORATION IS NEEDED TO EVALUATE UNCERTAINTY

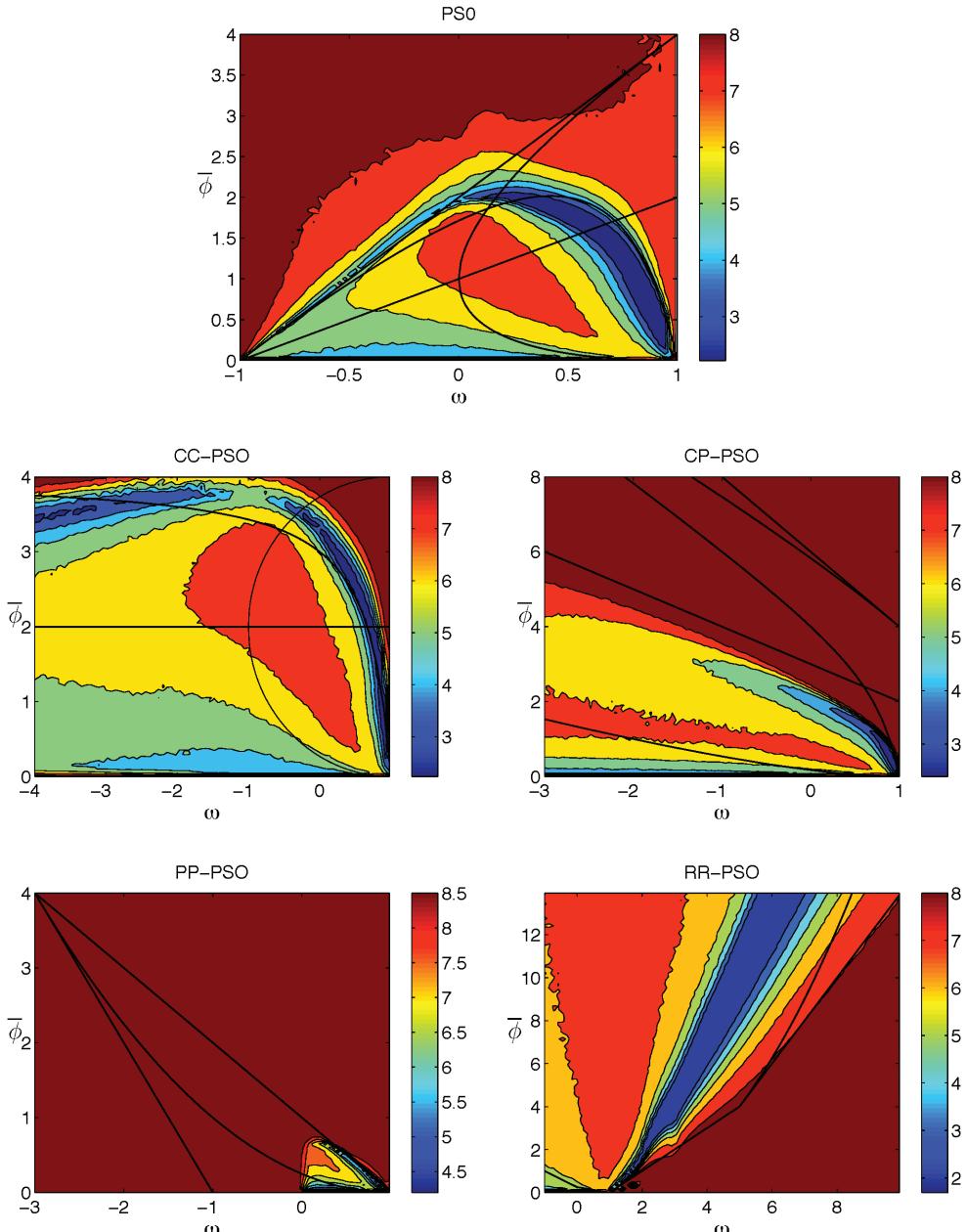
Most inverse problems can be written in discrete form as:

$$\mathbf{d} = \mathbf{F}(\mathbf{m}), \mathbf{d} \in R^s, \mathbf{m} \in R^n, \quad (2)$$

where  $\mathbf{d}$  is the observed data,  $\mathbf{m}$  is the vector containing the model parameters, and  $\mathbf{F}$  is the physical model.

Given a particular observed data set,  $\mathbf{d}_0$  the goal of inversion is to find a set of reasonable parameters,  $\mathbf{m}_0$  such that  $\mathbf{F}(\mathbf{m}_0) = \mathbf{d}_0$ .

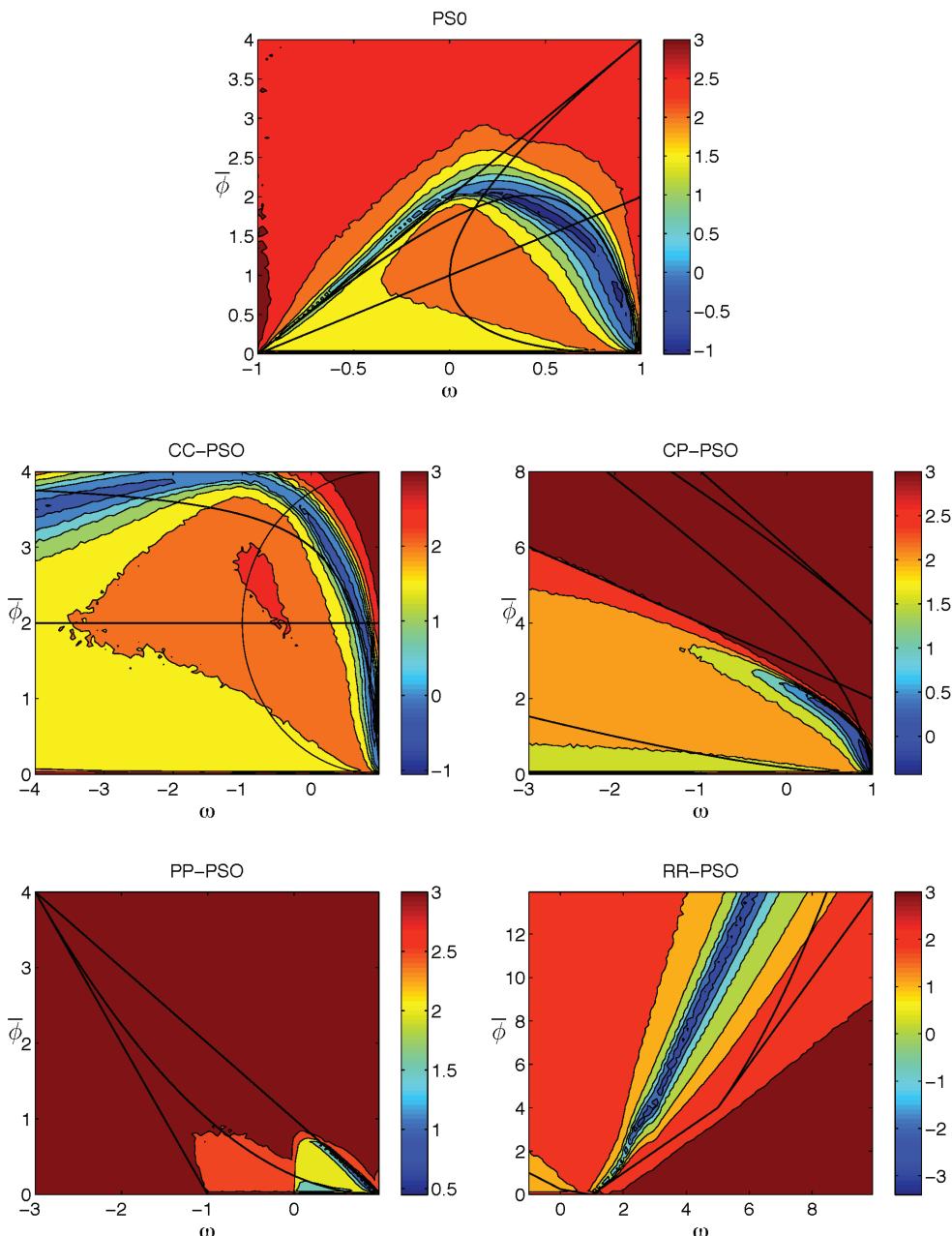
*Figure 3. Logarithmic median misfit errors for the Rosenbrock function in 50 simulations (after 300 iterations) for different family members*



Usually this problem either has no solution or many, and thus, the solution of the inverse

problem has to be constructed (Tarantola, 2004; Parker, 1994; Aster et al., 2005). The inverse

Figure 4. Logarithmic median misfit errors for the Griewank function in 50 simulations (after 300 iterations) for different family members



problem is then solved as an optimization problem: finding  $\mathbf{m}_0$  that minimizes the data

prediction error, expressed in a certain norm,  $\|\mathbf{d} - \mathbf{F}(\mathbf{m})\|_p$ .

The above optimization problem turns out to be ill-posed and the main reasons are:

1. The forward model  $\mathbf{F}$  is a simplification of reality, hypothesis and numerical approximations included.
2. Data are noisy and only partially sample the domain of interest.
3. Most applications of inverse modeling do not include enough (in quantity and quality) *a priori* knowledge to constraint the space of possible solutions.

These three points reveal inverse problems to be very different from other kind of optimization problems since physics and data are involved on the cost function,  $\|\mathbf{d} - \mathbf{F}(\mathbf{m})\|_p$ .

Although a wide variety of optimization algorithms exist in the scientific literature, how to select the appropriate algorithm to solve an inverse problem is still an active subject of research. These methods involve different well known local and global optimization methods such as Bayesian approaches, neural networks, Kalman filters, kernel methods and support vector machines, etc. In addition, finding the global minima of the error function is often complicated by the presence of many local minima and/or the error space topography surrounding the global minima being very flat and oblong. Primarily, the nature of the error space topography is dependent upon the forward model, or physics, but the overall picture is a flat valley shape. To understand this fact, let us consider a model  $\mathbf{m}_0$  located in the low misfit

region, that is  $\|\mathbf{F}(\mathbf{m}_0) - \mathbf{d}\|_2 < tol$ . If we take a Taylor expansion of  $\mathbf{F}(\mathbf{m})$  centered in  $\mathbf{m}_0$ , then:

$$\mathbf{F}(\mathbf{m}) - \mathbf{F}(\mathbf{m}_0) \approx \mathbf{JF}_{\mathbf{m}_0} (\mathbf{m} - \mathbf{m}_0),$$

$$\begin{aligned} \|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2 &= \|\Delta\mathbf{d}\|_2^2 + (\mathbf{m} - \mathbf{m}_0)^T \\ &\quad \mathbf{JF}_{\mathbf{m}_0}^T \mathbf{JF}_{\mathbf{m}_0} (\mathbf{m} - \mathbf{m}_0) + 2\Delta\mathbf{d}^T \mathbf{JF}_{\mathbf{m}_0} (\mathbf{m} - \mathbf{m}_0), \end{aligned}$$

where  $\Delta\mathbf{d} = \mathbf{F}(\mathbf{m}_0) - \mathbf{d}$ . Therefore  $\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2 = tol^2$  is the hyper quadric of equation:

$$\begin{aligned} &(\mathbf{m} - \mathbf{m}_0)^T \mathbf{JF}_{\mathbf{m}_0}^T \mathbf{JF}_{\mathbf{m}_0} (\mathbf{m} - \mathbf{m}_0) + \\ &2\Delta\mathbf{d}^T \mathbf{JF}_{\mathbf{m}_0} (\mathbf{m} - \mathbf{m}_0) + \|\Delta\mathbf{d}\|_2^2 = tol^2. \end{aligned}$$

Introducing the singular value decomposition of the Jacobian matrix  $\mathbf{JF}_{\mathbf{m}_0} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ , this equation transforms into:

$$\begin{aligned} &\Delta\mathbf{m}_{B_V}^T \mathbf{\Sigma}^T \mathbf{\Sigma} \Delta\mathbf{m}_{B_V} + \\ &2\Delta\mathbf{d}_{B_V}^T \mathbf{\Sigma} \Delta\mathbf{m}_{B_V} + \|\Delta\mathbf{d}\|_2^2 = tol^2, \end{aligned}$$

where " $\mathbf{m}_{B_V}$ " stands for the vector  $\mathbf{m} - \mathbf{m}_0$  referred to the  $\mathbf{V}$  base and  $\Delta\mathbf{d}_{B_V}$  is the error data prediction  $\mathbf{F}(\mathbf{m}_0) - \mathbf{d}$  referred to the base  $\mathbf{U}$ . The equivalent models locally in  $\mathbf{m}_0$  will have the direction of the vectors of the  $\mathbf{V}$  base and the axes are proportional to the inverse of the singular values  $(1/\lambda_k)$  in each  $\mathbf{v}_k$  direction. In the directions of the components that belong to the null-space of  $\mathbf{JF}_{\mathbf{m}_0}$ , the valley has infinite length. This means that along these directions the data are not sensible to model variations if the inverse problem was linear. Thus the equivalent region has locally a valley shape around model  $\mathbf{m}_0$ .

Additionally, data noise can lead to an increase in the size of the valley topography and/or in certain cases to an increased presence of local minima (see for instance, Alvarez et al., 2008; Martínez & González, 2008). Local optimization methods are not able to discriminate among the multiple choices consistent with the end criteria and may land quite unpredictably at any point on the valley region. These pathologies are treated through regularization techniques

and the use of “good” prior information and/or initial guesses.

Once the solution of an inverse problem is found, an important issue is the model appraisal to quantify the uncertainty of the estimation. When using local optimization algorithms the appraisal is usually done through linearization of the inverse problem functional. F. Bayesian approaches and Monte Carlo methods (Scales & Tenorio, 2001; Omre & Tjelmland, 1996; Mosegaard & Tarantola, 1995) can be used to solve the inverse problem in a different way accomplishing implicitly the model appraisal task. In this case, the inverse problem is formulated such that the posterior probability distribution is sampled many times in a random walk, with a bias towards increased sampling of areas of higher probability to accurately approximate the whole posterior probability (the so-called importance sampling). The drawback of such approaches is that they are computationally expensive, and they might not even be feasible at all in case of slow forward problems.

Posterior sampling techniques are closely related to global optimization algorithms, which can be used to provide a proxy for the true posterior distribution. In many practical situations, prior information is not available, and global optimization methods are a good alternative for avoiding the strong dependence of the solution upon noisy data. Most of these techniques are stochastic and heuristic and try to sample the low misfit region of model parameters, i.e., the region **E** of the model space **M** containing the models that fit the observed data within a given tolerance. As mentioned above, this sampling procedure acquires full sense in a Bayesian framework. Nevertheless, independent of the name that practitioners use to classify the inversion procedure, the main mechanism involved is sampling.

One of the peculiarities of global algorithms is that they do not try to solve the inverse problem as an optimization problem, but as a sampling problem, and thus, they do not need a prior model to stabilize the inversion. In fact the only prior information that these algorithms need is to specify the search space, which can be

very large if no prior information is available. In some cases, the space of solutions is so large that some form of regularization is necessary to constrain the solution sampling to a range of acceptable limits.

Global optimization algorithms include, among others, well known techniques such as Genetic algorithms (Holland, 1992), Simulated Annealing (Kirkpatrick et al., 1983), Particle Swarm Optimization (Kennedy & Eberhart, 1995), Differential evolution (Storn & Price, 1997), the Neighborhood algorithm (Sambridge, 1999a). Previous work has shown that, under certain conditions, global algorithms can provide accurate measures for model uncertainty. Mosegaard and Tarantola (1995) used Simulated Tempering to perform importance sampling of the model parameters through the Metropolis-Hastings algorithm. More recently, Álvarez et al. (2008) have shown numerically the ability of the binary Genetic Algorithms to perform an acceptable sampling of the posterior distribution when they are used in an exploration capacity. The Neighborhood Algorithm can also be used to accomplish this task (Sambridge, 1999b).

From this analysis we have shown that exploration is a very important property when we want to achieve a theoretically correct posterior sampling. In the case of genetic algorithms the exploration depends basically on the mutation probability. For simulated annealing the exploration is achieved by fixing the cooling parameter (temperature) to one.

Although PSO has not been designed to perform importance sampling, it provides a proxy for the posterior distribution for the model parameters if it is used in its explorative form (see for instance Martínez et al., 2010a). In the next section we show how to achieve exploration.

## HOW TO ACHIEVE EXPLORATION USING PSO

PSO can be viewed as a set of algorithms that can be used for exploitation (looking for a

unique global minimum) and/or exploration (sampling the low misfit region in the model space) purposes. The character of the PSO version (explorative or exploitative) depends on the PSO numerical parameters  $(\omega, a_g, a_l, \Delta t)$ .

The  $(\omega, a_g, a_l)$  point that has been selected influences the algorithm explorative behavior. As we have already commented, the greatest explorative behavior is achieved when the  $(\omega, a_g, a_l)$  point is close or even below the second order stability upper limit. The reason is that the variance increases its value with the mean acceleration  $\bar{\phi}$ . Above the second order stability limit, the variance does not attenuate with time (Martínez & Gonzalo, 2008, 2009). Also, the median of the first order stability region is the line where no temporal correlation exists at stagnation between trajectories,  $x(t)$  and  $x(t + \Delta t)$  (Martínez & Gonzalo, 2008). In the transient state, this temporal correlation goes to zero within a very small number of iterations (Martínez & Gonzalo, 2009b).

This median line has also the following special properties:

1. If for a given  $\omega$  value, the acceleration parameters  $a_g, a_l$  are chosen such that the

mean value  $\bar{\phi} = \frac{a_g + a_l}{2}$  falls on the median line. Then, every drawn value of  $\phi = \phi_1 + \phi_2 = r_1 a_g + r_2 a_l$ , will lie inside the first order stability region (Martínez et al., 2008).

2. The different set of parameters proposed in the literature for the PSO case (Carlisle & Dozier, 2001; Clerc & Kennedy, 2002; Trelea, 2003) lie close to the intersection of the median line and the upper limit hyperbola of second order stability (Martínez & Gonzalo, 2008). This fact has been also confirmed for all the PSO versions by means of numerical experiments (Martínez et al., 2009a).

Also, the total acceleration  $\phi = \phi_1 + \phi_2 = r_1 a_g + r_2 a_l$ , is a random variable with a trapezoidal density function centered on the mean value,  $\bar{\phi} = \frac{a_g + a_l}{2}$ , and with variance  $\sigma^2 = \frac{a_g^2 + a_l^2}{12}$ . Furthermore, if  $a_g = a_l$ , the density function becomes triangular. For the same total mean acceleration,  $\bar{\phi} = \frac{a_g + a_l}{2}$ , the use of dissimilar values of local and global accelerations terms,  $a_g, a_l$ , has the following consequences:

1. An increase in the dispersion of the total acceleration values at each PSO stage, since  $\phi$  exhibits a trapezoidal distribution function with a greater variance than the corresponding of the triangular distribution function (case of  $a_g = a_l$ ).
2. An increase of exploration when  $a_l > a_g$ , since particles are allowed to cover more freely the search space, making entrapment in local minima less likely.
3. Conversely, an increase of the exploitative task when  $a_g > a_l$ , since particles are strongly following the global leader direction. Entrapment around local minima becomes then more likely.
4. In the case  $a_l = a_g$ , the local and global terms have the same importance and the random draws for the total acceleration,  $\phi = r_1 a_g + r_2 a_l$ , are concentrated around its mean value  $\bar{\phi} = a_g = a_l$ .

The time step ( $\Delta t$ ) parameter is a numerical constriction factor used to achieve stability. It is possible to show analytically that the first and second order stability regions increase their size and tend to  $[\omega < 1, \bar{\phi} > 0]$  when  $\Delta t$  goes to zero. In this case, the exploration is increased around the global best solution. Conversely, when  $\Delta t$  is greater than one

the first and second order stability regions shrink in size and the exploration is increased in the whole search space. This feature might help to avoid entrapment in local minima. This feature has inspired us to create the lime and sand algorithm that alternates values of  $\Delta t$  greater and lower than one depending on the iterations (Martínez & Gonzalo, 2008).

When the inertia value approaches to one, the damping coefficient  $(1 - \omega)$  approaches zero, making the swarm explore a broader area of the search space. Stability is lost when  $\omega = 1$ . The role of the inertia constant  $\omega$  is also to avoid elitism, since the  $\mathbf{g}$  particle, which is the global best in iteration  $k$  ( $\mathbf{x}_g(k) = \mathbf{g}(k)$ ), will be on iteration  $k+1$ :

$$\mathbf{x}_g(k+1) = \mathbf{g}(k) + w\mathbf{v}_g(k), \quad (3)$$

i.e., its position will be changed by its velocity in iteration  $k$ ,  $\mathbf{v}_g(k)$ , weighted by the inertia constant,  $w$ . Relationship (3) also shows that if the velocity  $\mathbf{v}_g(k)$  is close to zero the swarm will collapse towards the global best, that is, the amount of exploration performed by PSO at the later stages of inversion is very limited and constrained to the global best neighborhood. This has the effect of oversampling the low misfit region.

Exploration can also be increased by:

1. Introducing repulsive forces into the dynamic of the swarm by switching to negative the sign of the acceleration constants. This repulsive effect is introduced when the mean distance between the particles of the swarm and the global best is less than a certain percentage (for instance 10%) of the initial mean dispersion. This percentage might depend on the inverse problem and the number of parameters used to describe the forward model. Also, in the posterior analysis we can take into account this effect by considering all the particles to be the same in each of these iterations.

2. Optimizing using the cloud versions (Gonzalo & Martínez, 2009; Martínez & Gonzalo, 2010) where each particle in the swarm has different inertia (damping) and acceleration (rigidity) constants. The results obtained for very hard benchmark functions in several dimensions using the PSO-cloud algorithm are comparable to the reference misfits published in the literature. This design avoids two main drawbacks of the PSO algorithm: the tuning of the PSO parameters and the clamping of the velocities. The criteria for choosing the cloud points it is not very rigid since points close to the second order convergence border achieve good results.

These strategies have been used to solve inverse problems in environmental geophysics (Martínez et al., 2009a, 2010a, b). Typically in these cases the number of parameters is low (tens to hundreds). These techniques can be also applied to inverse problems having a higher number of dimensions (thousands) by a combined use with model reduction techniques (Martínez et al., 2009c; Mukerji et al., 2009). In the next sections we show several examples in environmental geophysics and reservoir engineering.

## APPLICATION TO ENVIRONMENTAL GEOPHYSICS

The VES method is a low-cost DC geophysical method to characterize the depth-variation of the subsurface resistivity distribution. This technique exhibits a wide range of environmental applications.

The VES data acquisition method deployed with the Schlumberger configuration is as follows:

1. On the surface at each station, two current and two potential electrodes are symmetrically laid on both sides of a fixed central point. In successive stations, the

two external current electrodes are moved apart from each other increasing their mutual distance - $s$ - while holding the two inner potential electrodes fixed in place at a much shorter distance,  $b$ .

2. At each position, the voltage difference  $\Delta V(s)$  is measured and the apparent resistivity is calculated:

$$\rho_a^o(s; b) = K(s, b) \cdot \frac{\Delta V(s)}{I},$$

where  $K(s, b)$  is a real constant with a known dependency on parameters  $s, b$ , and  $I$  is the injected current.

In the VES model, the terrain is assumed to be horizontally stratified, characterized by the resistivities -  $\rho_k$  - and the thicknesses -  $t_k$  - of the electrical layers. The representation of the subsurface is accomplished by a vector  $\mathbf{m} = (\rho_1, t_1, \rho_2, t_2, \dots, \rho_{n-1}, t_{n-1}, \rho_n)$  belonging to a  $2n-1$  dimensional vector space  $\mathbf{M}$ , where  $n$  is the number of layers. The analysis of the VES mathematical model can be found for instance in Koefoed (1979) and involves the solution of the Laplace equation under cylindrical symmetry.

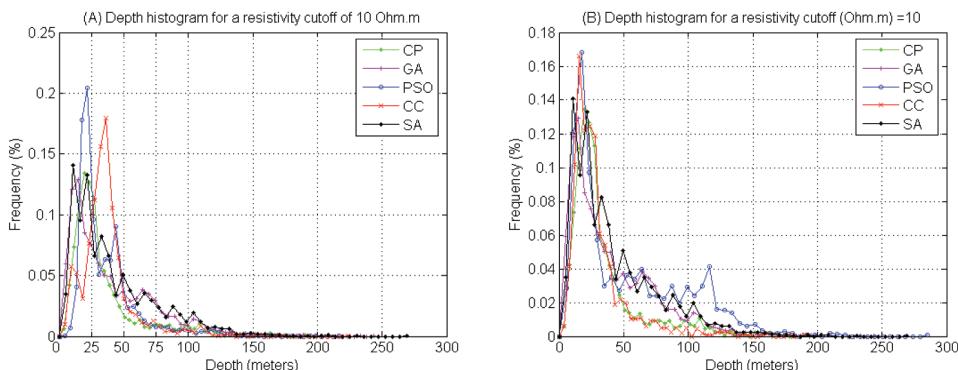
The inverse problem consists of estimating the model parameters of  $\mathbf{m}$  given the measured data,  $\mathbf{A}_a^o(\mathbf{s}) = (\rho_a^o(s_1), \rho_a^o(s_2), \dots, \rho_a^o(s_{n_d}))$ , for an array  $\mathbf{s} = (s_1, s_2, \dots, s_{n_d})$  of distances between the current electrodes (the so-called apparent resistivity curve), taking into account the physics of the problem. The VES inverse problem is nonlinear and over-determined since the number of available apparent resistivities,  $n_d$ , is often much higher than the number of geo-electrical parameters,  $2n-1$ . The number of geo-electrical layers,  $n$ , remains an additional unknown of the VES inverse problem. In the absence of prior information, the strategy adopted is to consider the lower number of geo-electrical layers that correctly fit the observed apparent resistivity curve such that the under-

determined part of the inverse problem is not increased. Although it has a low number of parameters, the VES inverse problem is a very challenging one due to its ill-posed character. Apart from the effect of noise, the main reason for ill-posedness of the VES inverse problem is that the spatial support of the resistivity field is unknown.

In this section we show the application of this geophysical method to monitor salt-water intrusion in a coastal aquifer in southern Spain, in a region where the use of groundwater for agriculture is very intensive.

The aim is to establish the depth of the intrusion probabilistically. For comparison purposes we have used Simulated Annealing with a temperature fixed to the value of one (no cooling) to correctly estimate the posterior distribution of the intrusion depth (Mosegaard & Tarantola, 1995). In all the cases, the swarm size was 200 models and the number of iterations 100. Initial seed was distributed uniformly on the model search space in  $R^{11}$ . For the SA case, the algorithm was programmed in matrix form to run in parallel a swarm of 200 models, nevertheless each model execution was autonomous. We have also used binary GA with a high mutation probability (0.3) and the algorithm was not elitist (Álvarez et al., 2008). For the PSO family members (PSO, CC-PSO and CP-PSO) the  $(w, a_l, a_g)$  values adopted are located on the intersection of the median line of the first order stability region and the second order stability region with  $a_l = 2a_g$ . Figure 5 A shows the histogram of the depth of the intrusion for a resistivity cut-off of 10 Ohms-meter for different global algorithms. The histograms provided by the PSO family members have approximately the same mode, but PSO and CC-PSO overestimate the mode probability, considering SA to be the correct importance sampler. This means that PSO and CC-PSO oversample the low misfit error region. CP-PSO provides the more accurate histogram. The results obtained for PSO and CC-PSO can be improved working in points with a higher exploration rate (for example those located on the

*Figure 5. Posterior analysis on the region of 20% relative error. Comparison between the different depths of intrusion histograms. A) Results using standard versions. B) Results obtained using clouds and introducing repulsive forces.*



median line of temporal uncorrelation above the second order stability region) and/or introducing repulsive forces to disperse the swarm in the last iterations. Results are shown in Figure 5 B. The use of the cloud versions also improved the quality of the posterior histogram. In the absence of more concluding theoretical proofs, the experimental results shown in this section explain that exploration is the key to perform a correct posterior sampling.

## HOW TO EXPLORE HIGH DIMENSIONAL SPACES? APPLICATION TO RESERVOIR ENGINEERING

The use of global optimization algorithms in inverse modeling is hampered mainly by two facts:

1. The computation time needed to solve each individual forward run.
2. The number of parameters needed to describe the inverse solution. In some cases the number of parameters used to solve the forward problem is very high (several thousands) due to the fine discretization used in the model space to achieve accurate data predictions. This also causes the inverse problem to be highly ill-posed.

Monte Carlo techniques and global optimization methods become unfeasible for high dimensional problems, although there are some attempts to deal with a high number of variables and fast forward evaluations (Cui et al., 2006, 2008). The main reason is that the base used to solve the inverse problem is the same than the one that is used to perform the forward predictions. We propose to use global optimization algorithms (PSO in this case) in a reduced model space, that is, to adopt a “more-informed” base in which we solve the inverse problem. Model reduction techniques can be used for this purpose. Its use is based on the fact that the inverse model parameters are not independent. Conversely, there exist correlations between model parameters introduced by the physics of the forward problem in order to fit the observed data. We propose to take advantage of this fact to reduce the number of parameters that are used to solve the identification problem. To illustrate this idea let us consider an underdetermined linear inverse problem of the form:

$$\mathbf{G}\mathbf{m} = \mathbf{d}, \quad \mathbf{G} \in M(s, n), \quad n \gg s,$$

where  $\mathbf{G}$  is the forward linear operator and  $s$ ,  $n$  stand respectively for the dimensions of the

data and model spaces. The solution to this linear inverse problem is expanded as a linear combination of a set of independent models  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_q\}$ :

$$\mathbf{m} \in \langle \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_q \rangle = \sum_{k=1}^q \alpha_k \mathbf{v}_k, \quad q << n$$

Now the inverse problem consists in finding a model  $\mathbf{m}$  in a subspace of  $\mathbb{R}^n$  of dimension  $q$ , fulfilling:

$$\mathbf{G}\mathbf{V}\pm = \mathbf{d} \rightarrow \mathbf{G}\mathbf{v}_i = \mathbf{b}_i \rightarrow \mathbf{B}\pm = \mathbf{d}, \quad \mathbf{B} \in M(s, q).$$

This amounts to finding the weights  $\pm$  of the linear combination. Although this linear system might still be ill-posed, the effect of this methodology is to reduce the space of equivalent solutions. Additionally, depending on the values of  $s$  and  $q$ , the new linear system might be over-determined. This methodology can be easily generalized to nonlinear inverse problems, because once the base  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_q\}$  is determined, the search is performed on the  $\pm$ -space.

The use of a reduced set of basis vectors that are consistent with our prior knowledge allows to regularize the inverse problem and to reduce the space of possible solutions. Several techniques can be used to construct these bases such as the Principal Component Analysis (PCA) of the model covariance, the Singular Value Decomposition, the Discrete Cosine Transform (DCT) and the Discrete Wavelet Transform (Martínez et al., 2010c).

Between all these techniques, the Principal component analysis (Pearson, 1901) is very useful in geosciences because it allows inputting into the base different geological scenarios. This method has been extensively used in several fields, such as weather prediction and operational oceanography, fluid dynamics, turbulence, statistics, reservoir engineering, etc. Sometimes is also known under other terminologies such as Proper Orthogonal

Decomposition or Orthogonal Empirical bases. The PCA transforms a number of correlated variables into a smaller number of uncorrelated variables called principal components. The resulting transformation is such that the first principal component accounts for as much of the variability and each succeeding component accounts for as much of the remaining variability as possible (Jolliffe, 2002). Usually PCA is performed in the data space, but in this case it is used to reduce the dimensionality of the model space based on a priori models obtained from conditional geostatistical realizations that have been constrained to static data. Applied to our context, PCA consists in finding an orthogonal base of the experimental covariance matrix estimated with these prior geological models, and then selecting a subset of the most important eigenvalues and associated eigenvectors that are used as a reduced model space base.

The method works as follows:

1. First let us imagine that we are able to generate a n ensemble  $\mathbf{X} = [\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_q]$  of plausible scenarios that are constrained using the available prior information. None of these scenarios obviously fit the observed data with the prescribed tolerance  $tol$ . Also these models might not constitute an independent set of vectors, that is, they might contain redundant information. Random field simulations techniques can be used for this purpose. For instance, let us suppose that all the scenarios will be modeled as stochastic processes of the type  $\mathbf{m}_i = \mathbf{t} + \mathbf{r}_i$ , where  $\mathbf{t}$  is a fixed deterministic trend and the residual  $\mathbf{r}_i$  having a known prior covariance  $\mathbf{C}$ . It is then possible to generate different unconditional simulations by  $\mathbf{m}_i = \mathbf{t} + \mathbf{L}\mathbf{u}_i$ , where  $\mathbf{u}_i$  is a standard Gaussian process and  $\mathbf{L}$  is the Cholesky decomposition of the covariance matrix  $\mathbf{C} = \mathbf{LL}^T$ . The different scenarios  $\mathbf{m}_i$  can also correspond to different trends ( $\mathbf{t}$ ) or spatial (time) covariance  $\mathbf{C}$ . These techniques can be implemented with two and

three dimensional fields (Deutsch & Journel, 1992; Le Ravalec-Dupin et al., 2000; Remy et al., 2009).

3. The problem consists in finding a set of patterns  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_q\}$  that provide an accurate low dimensional representation of the original set with  $q$  being much lower than the dimension of the model space. PCA does it by diagonalizing the prior experimental covariance matrix

$$\mathbf{C}_{\text{prior}} = \frac{1}{N} \sum_{k=1}^N (\mathbf{m}_k - \bar{\mathbf{m}})(\mathbf{m}_k - \bar{\mathbf{m}})^T$$

where  $\bar{\mathbf{m}} = \frac{1}{N} \sum_{k=1}^N \mathbf{m}_k$  is the experimental ensemble mean. This ensemble covariance matrix is symmetric and semi-definite positive, hence, diagonalizable with orthogonal eigenvectors  $\mathbf{v}_k$ , and real semi-definite positive eigenvalues. Eigenvectors  $\mathbf{v}_k$  are called principal components. Eigenvalues can be ranged in decreasing order, and we can select a certain number of them to match most of the variability of the models. That is, the  $d$  first eigenvectors represent most of the variability in the model ensemble. The centered character of the experimental covariance is crucial to maintain consistency after reconstruction.

4. Then, any model in the reduced space is represented as a unique linear combination of the eigenmodels:

$$\mathbf{m} = \bar{\mathbf{m}} + \sum_{k=1}^q a_k \mathbf{v}_k.$$

where  $\mu$  is the model experimental mean. The orthonormal character of the vectors provides to this base a telescopic (nested) character; that is, if we add the next eigenvector to the base, the vector will be expressed in these two bases as follows:

$$\begin{aligned} \mathbf{m} - \mu &= (a_1, a_2, \dots, a_q)_{\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_q\}} \\ &= (a_1, a_2, \dots, a_q, 0)_{\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_q, \mathbf{v}_{q+1}\}} \end{aligned}$$

This property allows an easy implementation of a multi-scale inversion approach adding more eigenvalues to match higher frequencies to the model  $\mathbf{m}$  as needed. Determining which level of detail we have to consider is an important question since all finer scales might not be informed by the observed data, that is, they might belong to the null space of our local linear forward operator. By truncating the number of PCA terms that we use in the expansion we are setting these finer scales of heterogeneity (high frequencies of the model) to zero, avoiding also the risk of over fitting the data. In other words, the use of a truncated PCA base provides a kind of natural smoothing of the solution.

The model covariance used in the PCA analysis, instead of being the prior covariance from a set of stochastic simulations, can also be a posterior covariance estimated by linearizing the forward operator about the last iteration, called  $\mathbf{m}_f$ . Based on linear inverse theory this posterior covariance can be computed using:

$$\mathbf{C}_{\text{pos}} \propto (\mathbf{JF}_{\mathbf{m}_f}^T \mathbf{JF}_{\mathbf{m}_f})^{-1}$$

$\mathbf{JF}_{\mathbf{m}_f}$  is the model Jacobian matrix computed at  $\mathbf{m}_f$ . This linear covariance matrix constitutes the prior information for our uncertainty estimation method and must be computed as the first step. The Jacobian matrix is rank deficient. The dimension of the null space of the Jacobian serves to account locally for the linear uncertainty analysis around the base model. To invert the matrix  $\mathbf{JF}_{\mathbf{m}_f}^T \mathbf{JF}_{\mathbf{m}_f}$ , truncation (Moore-Penrose pseudo-inverse) and/or damping techniques can be used. In this last case this inverse also gathers the influence of the left singular vectors that lie on the null space of  $\mathbf{JF}_{\mathbf{m}_f}$  and that are typically related to the

high frequencies in the model. Finally another different estimate for the posterior covariance can be computed as an experimental covariance from the ensemble of sampled posterior models in a certain region of error tolerance (for instance 10%). In this case the change to the reduced base is performed when this sampled posterior covariance is calculated.

### **Application to Hydrogeological Inverse Problems**

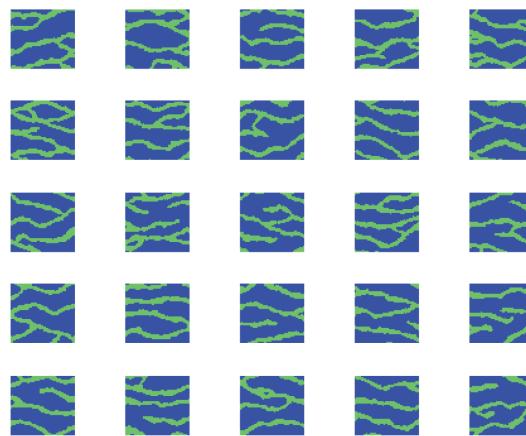
To illustrate this approach we use a synthetic problem where the transmissivity of an underground water reservoir has to be imaged using only the head measurements in nine wells. The problem was designed such that it has a large number of equivalent solutions, since the dimension of the model space is 10.000 and the number of data is only nine. To address this problem we include the prior information on a geologically consistent base. Figure 6 shows a set of scenarios that have been generated using multiple point geostatistics (Mariethoz et al., submitted; Mariethoz & Renard, 2010), with known prior information about this reservoir that consists in a training image displaying channels structures (Strebelle, 2002). Figure 7 shows the first 20 terms of the PCA base calculated using 2000 such scenarios. It can be observed that the terms of the PCA base reflect different features that are present on the original scenarios. The features' frequency becomes higher as their index increases. In this case we are using a PCA base reflecting the prior covariance. We solve this hydrogeological inverse problem using different numbers of PCA terms. Figure 8 shows the convergence rate curves deduced from this analysis. It can be observed that with 30, 50 and 100 PCA, PSO reaches the region of low misfits within 30 iterations. Also, when the number of PCA terms is increased to 200, the search becomes harder since the reservoir model has increased its high frequency content. Figure 9 shows the comparison between the posterior mean and standard deviation deduced from the samples having a relative misfit lower than 0.010 using an explorative version (CP-PSO), and those provided by a rejection

algorithm that used 100.000 forward problem evaluations (Mariethoz et al., submitted). It can be observed that PSO gives a good approximation of the posterior distribution with only 1500 evaluations, sampling very different kinds of reservoir models.

### **Application to Reservoir Engineering: The History Matching Problem**

The second application we would like to show is a petroleum engineering history matching problem. Numerical models and inverse problems are very much used in reservoir characterization to improve oil production. Solving the history matching problem provides to the reservoir engineers an update of the spatial distribution of physical reservoir properties that can be used in later stages for reservoir management. This problem has a very ill-posed character that increases with the noise level in the production data and with the reservoir complexity. In this case  $\mathbf{F}(\mathbf{m})$  it is composed of multiple components: a reservoir flow simulator to predict the production data; a wave propagation model and inversion (diffraction tomography) to reconstruct the seismic velocities from the seismic traces measured at the boreholes; a geostatistical model to constrain the spatial structure of the reservoir, and finally a rock physics model that takes into account the facies-specific relations between porosity, permeability, saturations and elastic velocities. The reservoir model is composed of 4000 cells organized in ten layers of 20x20 pixels extracted from the Stanford VI sand and shale synthetic reservoir (Castro et al., 2005). The ensemble of plausible reservoirs (one thousand) was generated by geostatistical techniques (Strebelle, 2002). These realizations of the reservoir span what we think could be the variability of our model space. All these models are conditioned to the borehole data (facies measured at the wells). To perform the inversion we used the cloud versions (Gonzalo & Martínez, 2009; Martínez et al., 2010) of the different PSO optimizers (Martínez & Gonzalo, 2009). As we have previously discussed, the

*Figure 6. Different scenarios for a shale and sand reservoir generated by multiple-point geostatistics*

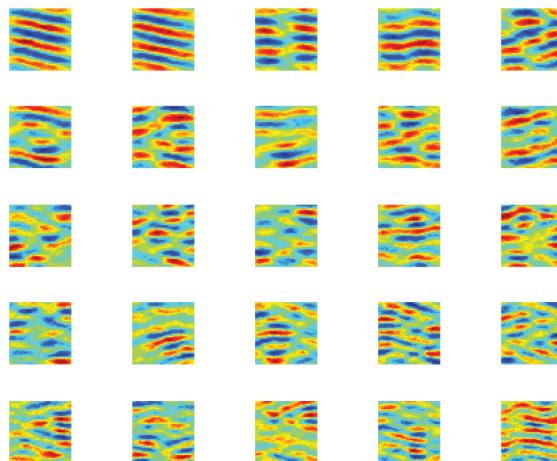


cost function in real inverse problems involves observed data that are always noisy. This makes the optimization problem to be highly non-convex. Then, it is critical to explore the space of possible solutions and be able to sample the posterior distribution in the model space, rather than to achieve a very low misfit with the danger of over fitting the data. A good balance between exploitation and exploration is the key in this case. This feature is achieved for the particle swarm optimizers through the cloud design.

Figure 10 shows the optimum facies model found by the CP-PSO algorithm (the most explorative version) in the presence of 5% of Gaussian noise compared to the true model. We also show the uncertainty analysis deduced from the samples that can be associated to this “optimum” facies model and belong to the low misfit region. Although the true model is binary (sand and shale) the optimum facies model and the interquartile range show a continuous color gradation due to the truncation adopted on the PCA base.

It can be observed that the inverted model approaches the true synthetic model, and although they are different, the uncertainty measures in each pixel serve to account for the

difference between these models. In practice the uncertainty analysis is performed by taking into account the particles that have been collected on a certain region of the low misfit area, with a chosen misfit cut-off. We also keep track of the evolution of the median distance between the global best and the particles of the swarm. When this distance is smaller than a certain percentage of the initial value (5% for instance) this means that the swarm has collapsed towards the global best. Once this happens we can either stop the algorithm, or continue iterating, but in the posterior analysis we count all the particles in this collapsed swarm as one. Taking them into account individually has the effect of overestimating the probability in the low misfit area due to oversampling. The algorithm also has the possibility to increase the exploration by adopting a new center of attraction based on previous statistics or by dispersing the swarm introducing repulsive forces by switching to negative the sign of the acceleration constants. Finally, based on the selected samples, it is possible to produce averages (E-types) over the samples and interquartile range maps to establish facies probabilities. The posterior marginal distributions indicate how any individual PCA term is resolved, and the posterior covari-

*Figure 7. First twenty-five terms of the PCA base*

ance helps us understanding the existing linear tradeoffs between model parameters. These last measures are related to the topography of the misfit surface.

In absence of more concluding theoretical proofs, the experimental results show that

exploration is the key to perform a correct posterior sampling. By using these algorithms in explorative form, we can use the samples that are gathered in low misfit regions to deduce a proxy of uncertainty for the model parameters.

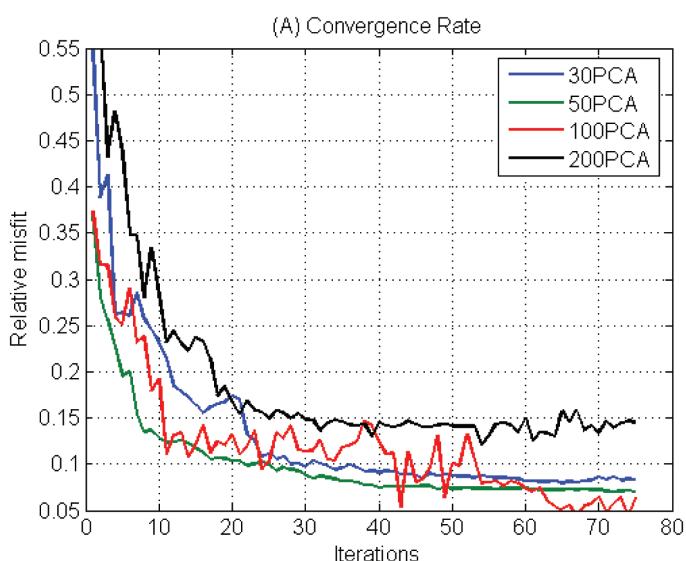
*Figure 8. Convergence curves for PSO as a function of the number of PCA terms*

Figure 9. Comparison between the posterior mean and standard deviation deduced using CP-PSO with 30 PCA members, and the rejection algorithm

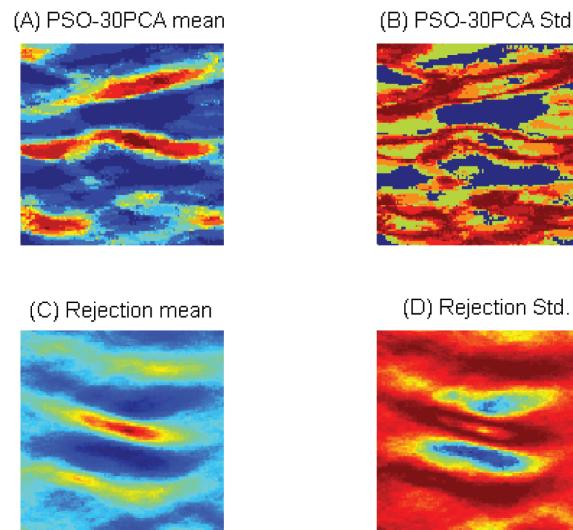
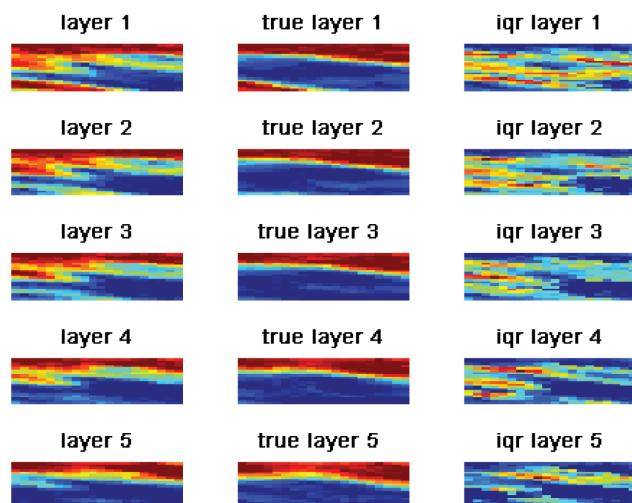


Figure 10. Five first horizontal slices from the synthetic reservoir showing the inverted solution, true model, and uncertainty on the model parameters (inter-quartile range)



## CONCLUSION

Inverse problems are ill posed. Uncertainty analysis around the best solution is always needed to perform risk analysis. Although the particle swarm optimizers have not been designed to perform posterior sampling, we show numerically that these algorithms can provide a proxy for the posterior distribution if they are used in their explorative form. Based on stochastic stability analysis, we show different ways to achieve exploration in particle swarm. Finally we show that the combined use of Particle Swarm Optimizers and model reduction techniques allows addressing real-world applications with thousands of parameters. The use of model reduction techniques is based on the fact that the inverse model parameters are not independent. Conversely, there exist correlations between model parameters introduced by the physics of the forward problem in order to fit the observed data.

## ACKNOWLEDGMENT

This work benefited from one-year sabbatical grant (2008-2009) at the University of California Berkeley (Department of Civil and Environmental Engineering) given by the University of Oviedo (Spain), and by the “Secretaría de Estado de Universidades y de Investigación” of the Spanish Ministry of Science and Innovation. We also acknowledge the financial support for the present academic year coming from the University of California Berkeley, the Lawrence Berkeley National Laboratory (Earth Science Division), and the Energy Resources Engineering Department of Stanford University (Stanford Center for Reservoir Forecasting and Smart Fields Consortia). We also acknowledge David Echeverría and Eduardo Santos (Stanford University) for lending us the forward programs to model the oil reservoir example presented in this paper.

## REFERENCES

- Aster, R. C., Thurber, C. H., & Borchers, B. (2005). *Parameter estimation and Inverse problems*. New York: Elsevier Academic Press.
- Carlisle, A., & Dozier, G. (2001). An off-the-shelf PSO. In *Proceedings of the Workshop on Particle Swarm Optimization 2001*, Indianapolis, IN.
- Castro, S., Caers, J., & Mukerji, T. (2005). *The Stanford VI Reservoir*. Palo Alto, CA: Stanford Center for Reservoir Forecasting (SCRF).
- Clerc, M., & Kennedy, J. (2002). The particle swarm-explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation*, 6(1), 58–73. doi:10.1109/4235.985692
- Cui, Z., Cai, X., Zeng, J., & Sun, G. (2008). Particle swarm optimization with FUSS and RWS for high dimensional functions. *Applied Mathematics and Computation*, 205, 98–108. doi:10.1016/j.amc.2008.05.147
- Cui, Z., Zeng, J., & Sun, G. (2006). A Fast Particle Swarm Optimization International . *Journal of Innovative Computing*, 2, 1365–1380.
- Deutsch, C., & Journel, A. (1992). *GSLIB: Geostatistical Software Library* (p. 340). New York: Oxford University Press.
- Echeverría, D., & Mukerji, T. (2009 July). A robust scheme for spatio-temporal inverse modelling of oil reservoirs. In R. S. Anderssen, R. D. Braddock, & L. T. H. Newham (Eds.), *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation*, Cairns, Australia (pp. 4206–421).
- Gonzalo, G. E., & Martínez, F. J. L. (2009a). Design of a simple and powerful Particle Swarm optimizer. In *Proceedings of the International Conference on Computational and Mathematical Methods in Science and Engineering (CMMSE2009)*, Gijón, Spain.
- Gonzalo, G. E., & Martínez, F. J. L. (2009b). *The PP-GPSO and RR-GPSO (Tech. Rep.)*. Oviedo, Spain: University of Oviedo, Department of Mathematics.
- Holland, J. (1992). Genetic algorithms. *Scientific American*, 66–72. doi:10.1038/scientificamerican0792-66
- Jolliffe, I. T. (2002). *Principal Component Analysis* (p. 487). New York: Springer-Verlag.

- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of IEEE International Conference on Neural Networks (ICNN '95)*, Perth, WA, Australia (pp. 1942-1948).
- Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220, 671-680. doi:10.1126/science.220.4598.671
- Koefoed, O. (1979). *Geosounding principles*. Amsterdam: Elsevier.
- Le Ravalec-Dupin, L., Noetinger, B., & Hu, L. (2000). The FFT moving average (FFT-MA) generator: An efficient numerical method for generating and conditioning Gaussian simulations. *Mathematical Geology*, 32(6), 701-723. doi:10.1023/A:1007542406333
- Mariethoz, G., & Renard, P. (2010). Reconstruction of incomplete data sets or images using Direct Sampling. *Mathematical Geosciences*, 42(3), 245-268. doi:10.1007/s11004-010-9270-0
- Mariethoz, G., Renard, P., & Caers, J. (n.d.). Bayesian inverse problem and optimization with Iterative Spatial Resampling. *Water Resources Research*.
- Mariethoz, G., Renard, P., & Straubhaar, J. (n.d.). The direct sampling method to perform multiple-points simulations. *Water Resources Research*.
- Martinez, F. J. L., Alvarez, F. J. P., Gonzalo, G. E., Pérez, M. C. O., & Kuzma, H. A. (2008). Particle Swarm Optimization (PSO): A simple and powerful algorithm family for geophysical inversion. In *Proceedings of the SEG Annual Meeting, SEG Expanded Abstracts* (Vol. 27, p. 3568).
- Martinez, F. J. L., & González, P. L. M. (2008). Anisotropic Mean traveltimes curves: a method to estimate anisotropic parameters from 2D transmission tomographic data. *Mathematical Geosciences*, 41(2), 163-192. doi:10.1007/s11004-008-9202-4
- Martinez, F. J. L., & Gonzalo, G. E. (2008). The generalized PSO: A new door for PSO evolution. *Journal of Artificial Evolution and Applications*, 15.
- Martinez, F. J. L., & Gonzalo, G. E. (2009). The PSO family: deduction, stochastic analysis and comparison. *Swarm Intelligence*, 3, 245-273. doi:10.1007/s11721-009-0034-8
- Martinez, F. J. L., & Gonzalo, G. E. (2009b). *Stochastic stability analysis of the linear continuous and discrete PSO models (Tech. Rep.)*. Oviedo, Spain: University of Oviedo, Department of Mathematics.
- Martínez, F. J. L., & Gonzalo, G. E. (2010). *What makes Particle Swarm Optimization a very interesting and powerful algorithm? Handbook of Swarm Intelligence-- Concepts, Principles and Applications Series on Adaptation, Learning, and Optimization*. New York: Springer.
- Martínez, F. J. L., & Gonzalo, G. E. (2010b). Advances in Particle Swarm Optimization. In *Proceedings of the Seventh Conference on Swarm Intelligence (Ants 2010)*.
- Martínez, F. J. L., Gonzalo, G. E., & Alvarez, F. J. P. (2008). Theoretical analysis of particle swarm trajectories through a mechanical analogy. *International Journal of Computational Intelligence Research*, 4, 93-104. Retrieved from <http://www.ijcir.com/>.
- Martínez, F. J. L., Gonzalo, G. E., Álvarez, F. J. P., Kuzma, H. A., & Pérez, M. C. O. (2010a). *PSO: A Powerful Algorithm to Solve Geophysical Inverse Problems. Application to a 1D-DC Resistivity Case*. Journal of Applied Geophysics.
- Martínez, F. J. L., Gonzalo, G. E., Muñiz, F. Z., & Mukerji, T. (2010d). *How to design a powerful family of Particle Swarm Optimizers for inverse modeling. New Trends on Bio-inspired Computation*. Transactions of the Institute of Measurement and Control.
- Martínez, F. J. L., Gonzalo, G. E., & Naudet, V. (2010b). *Particle Swarm Optimization applied to the solving and appraisal of the Streaming Potential inverse problem*. *Geophysics. Hydrogeophysics Special Issue*.
- Martínez, F. J. L., Kuzma, H. A., García-Gonzalo, M. E., Díaz, F. J. M., Álvarez, F. J. P., & Pérez, M. C. O. (2009). Application of Global Optimization Algorithms to a Salt Water Intrusion Problem. In *Proceedings of the Symposium on the Application of Geophysics to Engineering and Environmental Problems*, 22, 252-260.
- Martínez, F. J. L., Mukerji, T., Muñiz, F. Z., Gonzalo, G. E., Tompkins, M. J., & Alumbaugh, D. L. (2009b). *Complexity reduction in inverse problems: wavelet transforms, DCT, PCA and geological bases*. *Eos. Trans.*
- Martínez, F. J. L., Tompkins, M. J., Muñiz, F. Z., & Mukerji, T. (2010c, September). How to construct geological bases for inversion and uncertainty appraisal? In *Proceedings of the International Groundwater Symposium*, Valencia, Spain.
- Mosegaard, K., & Tarantola, A. (1995). Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research*, 100(B7), 12431-12447. doi:10.1029/94JB03097

- Mukerji, T., Martínez, F. J. L., Ciaurri, E. D., & Gonzalo, G. E. (2009). Application of particle swarm optimization for enhanced reservoir characterization and inversion. *Eos, Transactions, American Geophysical Union*, 90(52).
- Omre, H., & Tjelmland, H. (1996). *Petroleum geostatistics (Tech. Rep.)*. Trondheim, Norway: Norwegian University of Science and Technology, Department of Mathematical Sciences.
- Parker, R. L. (1994). *Geophysical Inverse Theory*. Princeton, NJ: Princeton University Press.
- Pearson, K. (1901). Principal components analysis. *The London . Edinburgh and Dublin Philosophical Magazine and Journal*, 6, 566.
- Remy, N., Boucher, A., & Wu, J. (2009). *Applied Geostatistics with SGeMS: A User's Guide* (p. 284). Cambridge, UK: Cambridge University Press.
- Sambridge, M. (1999). Geophysical inversion with a neighborhood algorithm. I. Searching a parameter space. *Geophysical Journal International*, 138, 479–494. doi:10.1046/j.1365-246X.1999.00876.x
- Scales, J.A., & Tenorio, L. (2001). Prior information and uncertainty in inverse problems. *Geophysics*, 66, 389–397. doi:10.1190/1.1444930
- Storn, R., & Price, K. (1997). Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11, 341–359. doi:10.1023/A:1008202821328
- Strebelle, S. (2002). Conditional simulation of complex geological structures using multipoint statistics. *Mathematical Geology*, 34, 1–21. doi:10.1023/A:1014009426274
- Tarantola, A. (2004). *Inverse Problem Theory and Model Parameter Estimation*. Philadelphia: SIAM.
- Trelea, I. C. (2003). The particle swarm optimization algorithm: convergence analysis and parameter selection. *Information Processing Letters*, 85(6), 317–325. doi:10.1016/S0020-0190(02)00447-7