Anais IX Simpósio Brasileiro de Sensoriamento Remoto, Santos, Brasil, 11-18 setembro 1998, INPE, p. 1133-1143.

Genetic-Annealing Parameter Estimation for Intensity SAR Data

PAULO RICARDO DOS SANTOS MENDONÇA[1]
ALEJANDRO C. FRERY[1]

[1]FALLSIDE LAB
CAMBRIDGE UNIVERSITY ENGINEERING DEPARTMENT
TRUMPINGTON STREET, CAMBRIDGE
CB2 1PZ UNITED KINGDOM

[2]UNIVERSIDADE FEDERAL DE PERNAMBUCO
DEPARTAMENTO DE INFORMÁTICA
CP 7851
50732-970 RECIFE, PE BRAZIL

Abstract. Finding the maximum likelihood estimators for some distributional parameters of intensity data in Synthetic Aperture Radar (SAR) images is a very difficult optimisation problem due to, among other reason, the presence of several local maxima in the objective function, the analytical intractability of the expressions involved and numerical instabilities. A possible approach to this problem is the use of stochastic optimisation techniques, such as simulated annealing and genetic algorithms, that do not get trapped into local maxima hills and, thus, make it possible to deal with very general distributions. This work shows the results of such approach in real situations, with images obtained from urban areas.

Keywords: Remote Sensing, SAR images, multiplicative model, parameter estimation, stochastic optimisation.

1 Introduction

Statistical tools have long been used to tackle some problems related to images. The stochastic nature of these objects, and the excellent results frequently obtained with this statistical approach, stimulated the development of a vast bulk of methods and techniques.

Most of these tools are based either on quite mild hypothesis (for instance, histogram equalization that assumes no distribution at all) or on the Gaussian distribution (Wiener filter, usual maximum likelihood classification, etc.).

Statistical modelling and parameter estimation are very important issues in the problem of processing and analysis of SAR images. These parameters can be associated to types of targets and, therefore, they can be used as important features for image classification.

The Gaussian distribution is so frequently used because, among other reasons, there are many techniques associated to this hypothesis. This distribution has been used for two centuries, and its properties are well known and many computational methods are available to deal with it. Synthetic Aperture Radar (SAR) images can be successfully modelled by a class of distributions that do not belong to the Gaussian distribution.

When SAR images are used, instead of optical data, the exception becomes the rule: the Gaussian hypothesis is seldom confirmed. This is mainly due to the coherent nature of the illumination, and the consequences of this departure range from poor results, when classical tools are applied, to the need of studying and proposing new methods for SAR image processing and analysis. Therefore, parameter estimation is a problem to be carefully addressed.

Among the several estimators available, the maximum likelihood ones are those with the "best" theoretical characteristics, like asymptotic convergence to normality, asymptotic consistency and efficiency, etc. Therefore, it would be desirable to estimate the parameters of distributions of SAR images data with those estimators. Nevertheless, the complexity of the models for such images usually yields to intractable problems, both from an analytical or numerical view point (Frery et al, 1997). To deal with that situation, the use of a stochastic optimisation technique, that combines genetic algorithms and simulated annealing into a single evolutionary strategy, is proposed in this work.

In Section 2 the general problem of maximum likelihood estimation is posed. The data models for distributions of SAR images are recalled in Section 3. This Chapter also introduces the problem of maximum likelihood estimation for the particular distributions that model SAR images data. An overview of stochastic optimisation is given in Section 4, with a short description of simulated annealing, genetic algorithms and the hybrid genetic annealing algorithm. The results of the last section are applied in Section 5, to the design of an algorithm aiming at solving the problem presented in Section 3. Experimental results are shown in Section 0, and the conclusions are discussed in Section 6.

2 Maximum likelihood estimation problem

Given the data set $\mathbf{x} = \{x_i\}_{i=1}^{N}$ of N independent samples of a random variable and the probability distribution $F_{\mathbf{p}}$ with density $f_{\mathbf{p}}$ under the vector of real parameters $\mathbf{p} = (p_1, \ldots, p_m)$, the *maximum likelihood estimator* for the vector of parameters $\mathbf{p}$ under the data set $\mathbf{x}$ and the distribution $F_{\mathbf{p}}$ is (if it exists) the vector $\hat{\mathbf{p}}$ such that $V(\hat{\mathbf{p}}) = \prod_{i=1}^{N} f_{\mathbf{p}}(x_i)$ is maximum. Since the logarithmic function is a strictly increasing function, the maximisation of this function is equivalent to the problem of finding $\max_{\hat{\mathbf{p}}} \sum_{i=1}^{N} f_{\mathbf{p}}(x_i)$.

For some particular distributions, e. g. Gaussian, the maximum likelihood estimation problem has simple analytical and well known solutions. But if the modelling distribution is more complex, analytical solutions may not be feasible. When faced to this situation, the analyst has to choose between using other estimation technique (the substitution method, for instance) and looking for optimisation tools. This is the case for the models presented in this work, and the use of a numerical method to find the estimator vector $\hat{\mathbf{p}}$ is presented.

3 Models and inference for SAR images

Data from SAR images can be modelled as the product of two independent random components: one due to the terrain backscatter and one due to the speckle noise. The usual model for n looks intensity speckle is a Gamma distribution, denoted as $Y_I \sim \Gamma(n, n)$, with density given by

$$f_{Y_I}(y) = \frac{n^n}{\Gamma(n)} y^{n-1} \exp(-ny), \quad y,n > 0 .$$

The number of looks, namely n, can be assumed known beforehand. A possible model for the distribution of the intensity backscatter return is the inverse generalised Gaussian, denoted as $X_I \sim N(a, g, l)$ with density

$$f_{X_I}(x) = \frac{(l/g)^{a/2}}{2K_a(2\sqrt{lg})} x^a \exp\left(-\frac{g}{x} - lx\right), \quad x > 0$$

where K_a is the modified Bessel Function of the third kind, and parameters space

$$\begin{cases} g > 0, & l \geq 0, \quad \text{if } a < 0, \\ g > 0, & l > 0, \quad \text{if } a = 0, \\ g \geq 0, & l > 0, \quad \text{if } a > 0. \end{cases}$$

The intensity return corresponds to the product of the aforementioned backscatter and speckle $Z_I = X_I \cdot Y_I$ results in a distribution denoted as $G_I(a, g, l, n)$ with density

$$f_{Z_I}(z) = \frac{n^n (l/g)^{a/2}}{\Gamma(n) K_a(2\sqrt{lg})} z^{n-1} \left(\frac{g + nz}{l}\right)^{\frac{a-n}{2}} K_a\left(2\sqrt{l(g + nz)}\right), \quad z > 0$$

The above distribution has several limit cases of interest, depending on the parameters, which are presented and discussed in Frery et al. (1997). The particular case that will be here discussed in that corresponding to the situation $-a, g > 0$, $l = 0$, that yields to a distribution known as $G_I^0(a, g, n)$, whose density is given by

$$f_{Z_I}(z) = \frac{n^n \Gamma(n-a) z^{n-1}}{g^a \Gamma(n)\Gamma(-a)(g + nz)^{n-a}}$$

Solving the maximum likelihood problem for the data set $(z_1, \ldots, z_N)$ and the distribution characterised by the density given above consists of maximising the following equation with respect to both variables a and l

$$\ell(a, l) = N \log \frac{n^n \Gamma(n-a)}{g^a \Gamma(n)\Gamma(-a)} + (n-1) \sum_{1 \leq i \leq N} \log z_i + (a - n) \sum_{1 \leq i \leq N} (g + nz_i)$$

This maximisation is the problem addressed in the next sections.

4 Stochastic optimisation

Optimisation is the search for parameters or solutions for a given problem that minimise or maximise a certain function or functional, conditioned or not to a set of restrictions.

There is a restricted class of optimisation problems with a simple analytical solution, that can be found by usual methods of calculus (Apostol, 1967a; Apostol, 1967b). Nevertheless such approach is not feasible in most of the optimisation problems of real interest due to, among other reasons, the computational complexity aspects as well as the lacking of closed analytical expressions for the involved functions.

On the other hand, Nature is every time dealing (quite successfully) with optimisation problems of extreme complexity. Two examples are the evolution of species and the annealing of molten metals. In the former example living beings, mankind for instance, are exposed to natural competition and to the influence of the environment. Those best adapted survive to keep on fighting, wile the weaker are eliminated of the evolutive competition.

The annealing of molten metals is strongly connected to an amazing optimisation problem, stated as the Second Law of Thermodynamics: *closed systems evolve so that its entropy raises* (Hill, 1960). Each atom of a system behaves in order that this condition is satisfied. In a molten metal, numbers of the order of 10^{23} atoms, in random motion, build into a single crystal, among the almost uncountable possible arrangements, in obedience to the Second Law.

Besides being natural phenomena, the two above examples have in common the presence of random elements. The crossover and mutation of species are affected by so many different and fuzzy parameters that a single deterministic analysis of this problems becomes impossible and useless. In the state transition of metals, the success of Statistical Thermodynamics over the Laplacian deterministic approach clearly shows that such problem cannot be properly coped without the aid of an stochastic approach. Thus, in the solution of these two formidable optimisation problems, Nature deals with or uses randomness. That is the basis for the study of stochastic optimisation. Without loss of generality, will we consider in the next sections only minimisation problems.

4.1 Simulated annealing algorithms

The fundamental ideas of simulated annealing algorithms were first introduced by Metropolis et al (1953), as a method to determine physical and chemical properties of a set of atoms in transition to thermal equilibrium. In this seminal work this technique did not used any cooling schedule, a central issue in annealing algorithms, since this cooling idea had no connection with the problem under analysis. The use of the Metropolis method in the search for the solution for more generic optimisation problems is due to Kirkpatrick et al (1983), where the concept of cooling schedule was also introduced. Geman and Geman (1984) were the first to state an lower bound to the temperature decay rate that guarantees the convergence of the algorithm to a global minimum or maximum point. A basic simulated annealing algorithm is described in the next scheme.

1	Generate initial state $\mathbf{x}_t$
2	Generate noise with temperature T : $\mathbf{r}_t$
3	Generate candidate to new state $\mathbf{c} = \mathbf{x}_t + \mathbf{t}_t$
4	Test the acceptance of candidate:
	4.1 If candidate is accepted then $\mathbf{x}_{t+1} = \mathbf{c}$
	4.2 Else $\mathbf{x}_{t+1} = \mathbf{x}_t$
5	Decrease T
6	Goto step 2

4.2 The Metropolis algorithm

Let S be a closed system, and X be the set of possible states in S. Consider also $\mathbf{x}_t, \mathbf{x}_{t+1} \in X$, with associated internal energies, or cost functions, $E(\mathbf{x}_t)$ and $E(\mathbf{x}_{t+1})$. Finally, define $\Delta E(\mathbf{x}_t) = E(\mathbf{x}_t) - E(\mathbf{x}_{t+1})$.

The Transition Acceptance Probability (TAP) between the states $\mathbf{x}_t$ and $\mathbf{x}_{t+1}$ is defined as

$$\Pr(\mathbf{x}_t, \mathbf{x}_{t+1}) = \begin{cases} \exp(-\Delta E(\mathbf{x}_t)/(kT(t))) & \text{if } \Delta E(\mathbf{x}_t) < 0 \\ 1 & \text{else} \end{cases}$$

where $T(t)$ is the current temperature of the process and k is a positive constant. This probability benefits transitions from states with larger costs to states with smaller associated costs. Next result shows another (less intuitive) advantage of using this TAP.

Theorem 1: if all states in a system with transitions ruled by the TAP can visited at any temperature $T > 0$, the set of states distribution converges to a Boltzmann-Gibbs distribution, i.e.,

$$\lim_{t\to\infty} \Pr(\mathbf{x}_t = \mathbf{x}) = \frac{\exp(-E(\mathbf{x})/(kT))}{\sum_X \exp(-E(\hat{\mathbf{x}})/(kT))}$$

4.3 The state visiting process

Nothing was said about the rules of visiting of states so far. This issue and the concept of cooling schedule are the fundamental differences between the Metropolis Algorithm and simulated annealing. In the former, a state $\mathbf{x}_{t+1}$ is randomly selected "around" the previous state $\mathbf{x}_t$ with uniform distribution among its neighbours. In a simulated annealing algorithm the selection of a new state is done by a temperature dependent distribution. The time variation of the temperature is the called "cooling schedule". Expressing the temperature as a monotonic decreasing function of time, i.e. the probability density of the distribution probability that selects states in an annealing algorithm denoted $g_t(\cdot)$, must satisfy the relation $g_t(\cdot) \to d(\cdot)$ as $t \to \infty$.

A formal approach to realisations of g_t can be found in Mendonça (1997). For the purposes of this work, it is enough to consider that g_t depends only on the distance $\Delta\mathbf{x}_t = \|\mathbf{x}_{t+1} - \mathbf{x}_t\|$ between the current and visited states, i.e., $g_t(\mathbf{x}_{t+1}, \mathbf{x}_t) = g_t(\Delta\mathbf{x}_t)$, and on the result stated in the following Theorem.

Theorem 2: A sufficient and necessary condition for the convergence of an annealing algorithm is that $\sum_{t=t_0}^{\infty} g_t(\Delta\mathbf{x}_t) = \infty$.

This theorem is a landmark in the theory of simulated annealing, since if establishes an analytical tool in the design of algorithms with assured convergence.

4.4 The n-fast simulated annealing algorithm

The algorithm introduced in Geman and Geman (1984) could not be directly used due to its slow convergence speed. This problem was overcame in Szu and Hartley (1987), with the development of the *Fast Simulated Annealing Algorithm* (FSA), and further improvement was obtained in Mendonça (1997) and Mendonça and Caloba (1997), with the *n-Fast Simulated Annealing Algorithm* (n-FSA), whose main results are highlighted in the following. Consider

1	$P \leftarrow$ generate initial population
2	while the stopping criteria is not satisfied
	2.1 $(a'', a''') \leftarrow \text{crossover}(a, a'', P)$
	2.2 $P \leftarrow \text{substitution\_criterion\_1}(a'', a''', P)$
	2.3 $b' \leftarrow \text{mutation}(b)$
	2.4 $P \leftarrow \text{substitution\_criterion\_2}(b', P)$
3	endwhile

- A random variable, corresponding to the state visiting distribution probability $nR = (1+R)^n - 1$, where R is the norm of a D-dimensional Cauchy distributed random vector in polar co-ordinates;
- The cooling schedule for convergence $T_{n,R}(t) = T(0)/(1+t)^n$

Theorem 3: Let a be the desired probability of a transition between states at a distance L apart at $t = 0$. Then the initial temperature for the cooling schedule must be $T(0) = L/((1+\tan(p_D^{-1}(1-a)))^n - 1)$, where $p_D = \int_0^j D\Gamma((1+D)/2)/(\sqrt{p}\,(1+D)/2)\,\text{sen}^{D-1}\,dq$, in order to guarantee convergence.

4.5 The basic genetic algorithm

The basic structure of a genetic algorithms is presented in the next scheme. The first step of the algorithm is the generation of a population of feasible solutions to the optimisation problem. This population must have great diversity and must be as uniformly distributed over the space of search as possible. Then, while the stopping criteria is not satisfied, i.e., while a "good" solution is not found among the individuals of the population, the crossover and mutation operators are executed.

The crossover corresponds to a binary operation between two selected individuals, yielding to an offspring or pair of offspring. This operator is usually designed so that the offspring has as many common good characteristics to its parents as possible. This results in an heuristic local search procedure, since the solution found by crossover may not be too apart from its parents.

Then the offspring (or offspring) is added to the original population and from this new set of individuals, one (or two) are removed, keeping the size of the initial population. The substitution criteria must be coherent with the evolution of the population.

The mutation is an unary operation that corresponds to the adding of a random perturbation in a selected individual, producing a mutant. Again, the mutant individual is added to the

population, and a consistent substitution criteria is used to keep its size. The substitution criteria for mutation does not have to be the same as the one for the crossover. This loop goes on until the stopping criteria is fulfilled.

There are some doubts related to the validity of the genetic operator of mutation. Some serious researchers believe that the use of this operator may degenerate the algorithm to a random exhaustive (Tanomaru, 1995). Let us now consider this question from a different point of view.

4.6 Genetic or annealing?

Several works in simulated annealing stress the advantages of this algorithm over genetic algorithms, and vice-versa. It is evident that certain algorithms are better suited to particular problems than others, and this contributes to this discussions (Wolpert and Macready, 1995).

Some claimed advantages of genetic algorithms over simulated annealing are their flexibility and ease of parallel implementations. Annealing, on the other hand, has stronger analytical fundaments, resulting in a better understanding and control of the algorithm; it also has a convergence proof. In general, genetic algorithms have a greater adaptability to combinatorial problems, while continuous or piecewise continuous cost functions are better handled by annealing algorithms.

4.7 Metropolis in genetic algorithms

Due to the difficulty in modelling the evolution of species, there are few analytical tools involved in the study and development of genetic algorithms. In opposition, the simulated annealing algorithms are based in Statistical Thermodynamics and, therefore, supported by strong mathematical arguments.

If the state transition probability, i.e. the probability of substitution of an old individual by a new one, is given by the TAP, and if the visited state, or individual generation, is such that all states can always be visited, the state distribution will converge to a Boltzmann-Gibbs distribution (Theorem 1).

The point now is how to ensure that all states can always be visited. The crossover genetic operator is designed in such way that the common characteristics of its parents are preserved and, thus, it is not suitable to yield to populational diversity. However, the mutation operator can be easily projected to guarantee the variety in the generation of individuals. So applying TAP as a substitution criteria *between the original and the mutant individuals* for the mutation operator results in a Boltzmann-Gibbs distribution for the population, and in the convergence of the Genetic-Metropolis Algorithm.

4.8 Annealing in genetic algorithms

Now it is possible to introduce a cooling schedule into the Genetic-Metropolis Algorithm, resulting in a full Genetic-Annealing Algorithm. To obtain that it is enough to generate the mutant by the addition of a perturbation under a probability distribution with the distribution

presented in Section 4.4, and an appropriate cooling schedule. Details can be found in Mendonça (1997) and Mendonça and in Mendonça and Caloba (1997).

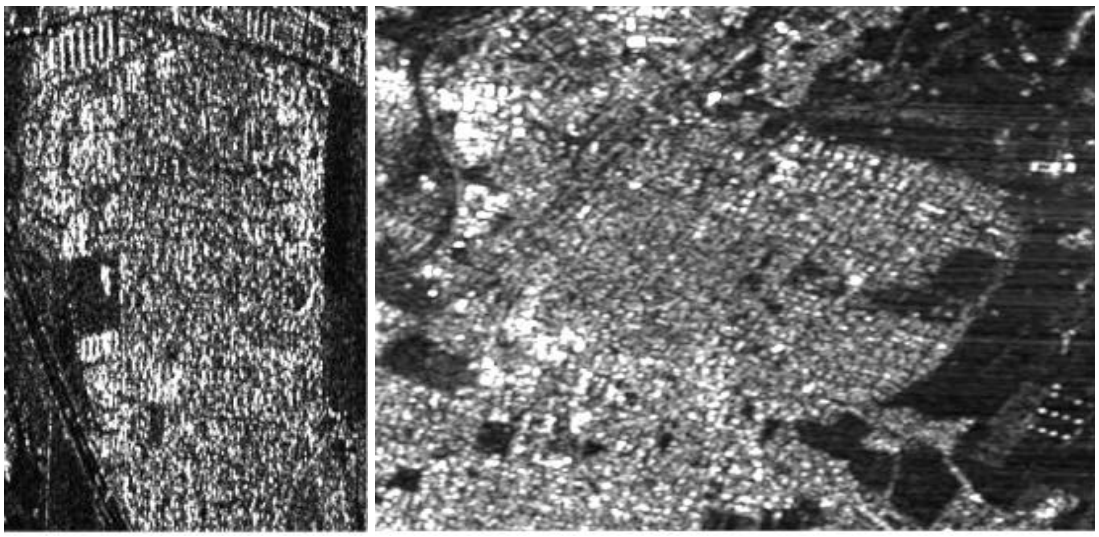
5 Parameter estimation for the G_I^0 distribution

The flexibility and power of the Genetic Annealing Algorithm motivated its use in the problem of parameters estimation of the G_I^0 distribution. Forthcoming sections will present particular aspects of the implementation of the algorithms, and the simulation results for real data.

5.1 Implementation of the algorithm

The Genetic Annealing Algorithm starts with the generation of a population of individuals. In Frery et al (1997) parameter estimation is performed by the moments method. In this manner, the population of estimators *around* this previous estimator can be generated, obtaining with a good chance that the solution found for the maximisation of the aforementioned likelihood equation has a greater likelihood than that of the estimator found by the moments method.

In all the tested situations, a initial population of 100 individuals (or estimators) was large enough to find a good solution with less than 1000 generations, and all images were sub sampled to a matrix of size 20×20. For technical details about the images, shown in the next Figure, the reader is referred to Frery et al (1997).



5.2 Representation of the population

The actual representation of parameters was (Mendonça, 1997), due to its simplicity and to the continuous nature of the problem. The population was the generated as a 100×4 matrix, following the scheme bellow:

	$\hat{a}$	$\hat{g}$	Prob. Selection	Likelihood
Entry for ind. 0	•	•	•	•
Entry for ind. 1	•	•	•	•
…	…	…	…	…

Entry for ind. 98	•	•	•	•
Entry for ind. 99	•	•	•	•

The likelihood column corresponds to the value of the log-likelihood function evaluated at the parameters in the corresponding line. The matrix is then sorted from the best to the worst individual. It is important to observe that, due to the limited machine precision, the final values of this column may equal infinity, and the correspondent degenerated individuals must be discarded from the population. This results in an effective population of size N, with a few less individuals than previously designed. Nevertheless, the number of degenerated individuals must be small, since the estimator found by the moments method is an excellent starting point.

The column **Probability of Selection** is filled with a positive parameter proportional to the difference between the correspondent value in the likelihood column plus a threshold, chosen here as 1, and the likelihood of the worst (or last, since they are already sorted) individual. The column is then normalised so that it sums to 1. This procedure, shown in detail in the following equation, generates a number that will be used as a probability of selection for individuals in the crossover and mutation operators, selecting with greater probability those individuals with a better adaptation. Denoting the likelihood as ℓ and a certain threshold as d:

$$\text{Prob. Selection}(i) = \frac{\ell(i) - \ell(N-1) + d}{\sum_{j=1}^{N-1} \ell(j) - N(\ell(N-1) - d)}$$

5.3 Crossover operator

The first step in the application of the crossover operator is the selection of a pair of individuals, henceforth named r_1 and r_2 (for *relatives*). The column with the assigned probabilities is used in this task. First, a random number x in the interval [0,1] is generated. Given the set of probabilities $\{p_j\}_{j=0}^{N-1}$, the index i of the selected individual is $i = \sum_{j=0}^{N-1} u(x - p_j)$, where u is the unitary step function. This procedure is repeated until a pair of different individuals is selected.

The crossover itself is done by convex combination of the parents, and the offspring is generated as shown

$$(\hat{a}_{\text{offspring}}, \hat{g}_{\text{offspring}}) = \frac{p_1(\hat{a}_{r_1}, \hat{g}_{r_1}) + p_2(\hat{a}_{r_2}, \hat{g}_{r_2})}{p_1 + p_2}$$

where $\hat{a}_{r_i}, \hat{g}_{r_i}$ and p_i are the parameters a, g and the probability of selection of the relative i.

The log-likelihood of the offspring is evaluated, and compared to the log-likelihood of its relatives. If it is greater than any one of them, the offspring replaces the relative with smaller adaptation. Then, the probabilities of selection are re-evaluated, and the mutation operator is again applied.

5.4 Mutation operator

Analogously to the crossover operator, the probabilities of selection are used to choose an individual m that will mutate. The mutation itself is processed in a quite similar way as in the crossover, from an individual i and a perturbation j: $(\hat{a}_m, \hat{g}_m) = (\hat{a}_i, \hat{g}_i) + (\hat{a}_j, \hat{g}_j)$.

The log-likelihood of the mutant is evaluated, and a selection between the original individual and the mutant is done, under the rule given in Theorem 1.

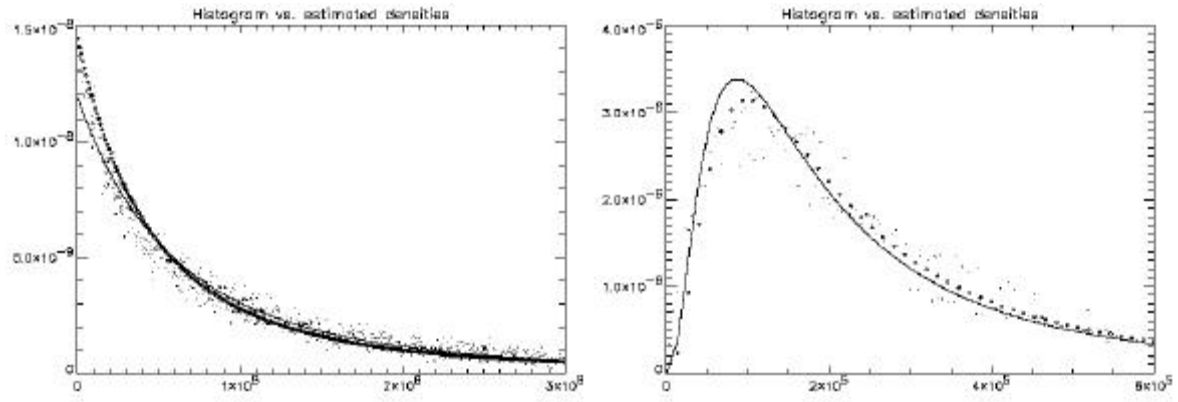
An important detail is that this operation is **not** used when the selected candidate to mutation is the best individual. Such procedure is necessary since, on the contrary to a simple annealing algorithm, there are no guarantees that the same individual will be obtained in the next generation, so it must ensured that the best individual is not lost by mutation.

5.5 Experimental results

This section presents comparisons between the genetic annealing algorithm and the moments method (MO for short), presented in Frery et al (1997). All images were sub-sampled to 20×20, which is an immediate advantage of the proposed method over the moments, since the last need samples of sizes as large as 100000.

The next Table summarises the results of the first and second experiments. Next figure shows the fit of the estimated densities over the image histogram.

	$\hat{a}$		$\hat{g}$		Log-likelihood	
MO	-2.018	-1.807	401329	124839000	-1234378.2	$-2.146947.6$
ML	-1.770	-2.416	317960	202774000	-1234334.1	-2145896.4



It is evident that in both cases the log-likelihood obtained by the optimisation procedure is higher than that attained by the moments method. This yields to a better representation of the data, a quite desirable feature for processing and analysis algorithms.

It is important to note that the log-likelihood in the tables is evaluated for the whole images data set, although the ML estimator is evaluated only for a small sub-sampled set of the image data, while the moments method uses all the available data.

6 Conclusions

This work presented a novel approach to the problem of maximum likelihood estimation for some SAR image data distributions, based on a fusion of genetic algorithms and simulated annealing. The results obtained are superior, in the log-likelihood sense, to those obtained by the moments method. The method itself is more general and robust, and allows the use of smaller samples than the moments technique.

The next step of this work will be the use of this method in the estimation of parameter of the G_I distribution. Parallel implementations will also be considered.

7 References

Apostol, T.M. *Calculus*, vol. 1. John Wiley & Sons, Singapore, 2. Ed., 1967.

Apostol, T.M. *Calculus*, vol. 2. John Wiley & Sons, Singapore, 2. Ed., 1967.

Frery, A.C.; Müller, H.J.; Yanasse, C.C.F.; Sant'Anna, S.J.S. A model for extremely heterogeneous clutter. *IEEE Trans. Geosc. Rem. Sens.*, 35(3):648–659, 1997.

Geman, S; Geman, D. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Patt. An. Mach. Int.*, 6:721–741, 1984.

Hill, T.L. *Introduction to statistical thermodynamics*. Addison-Wesley, MA, 1960.

Kirkpatrick, S.; Gelatt Jr., C.D.; Vecchi, M.P. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

Mendonça, P.R.S. *Novos algoritmos de simulated annealing*. MSc Thesis, COPPE/PEE, Universidade Federal de Rio de Janeiro, 1997.

Mendonça, P.R.S.; Caloba, L.P. New simulated annealing algorithms. In: *Proceedings 1997 IEEE International Symposium on Circuits and Systems*, June 1997, to appear.

Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N. Equation of state calculation by fast computing machines. *Jr. Chem. Phys.*, 6:1087–1092, 1953.

Szu, L; Hartley, H. Fast simulated annealing. *Phys. Lett. A*, 122(3,4):157–162, 1987.

Tanomaru, J. Motivação, fundamentos e aplicações de algoritmos genéticos. In: *II Congresso Brasileiro de Redes Neurais*, Curitiba, 1995.

Wolpert, D.H.; Macready. W.G. *No free lunch theorems for search*. Technical report, Santa Fe Institute, 1995.