Computer Science • Vol. 10 • 2009

Anna Pięta*, Justyna Bała*

COMPARING PARALLEL PROGRAMMING ENVIRONMENTS FOR THE JOINT INVERSION OF GEOELECTRICAL DATA

The article presents the comparison of the implementation of the inverse problem in geoelectrical methods in two different parallel computational environments. Combination of Monte Carlo method and Multistart algorithm was applied in the inversion process. Parallelization was done by fine grain decomposition. Execution time, speed-up and efficiency received for parallel algorithms in both computational environments were presented and analyzed.

Keywords: *parallel computing, inverse problem, geoelectrical methods*

PORÓWNANIE RÓWNOLEGŁYCH ŚRODOWISK OBLICZENIOWYCH NA PRZYKŁADZIE INWERSJI POŁĄCZONEJ DANYCH GEOELEKTRYCZNYCH

W artykule przedstawiono porównanie równoległej implementacji zagadnienia odwrotnego dla metod geoelektrycznych w dwóch różnych środowiskach obliczeniowych. Do rozwiązania zadania odwrotnego użyto algorytmu Monte Carlo – Multistart. W przypadku równoległej realizacji zastosowano drobnoziarnistą dekompozycję inwersji danych geoelektrycznych. Analizowano czas, przyspieszenie i efektywność algorytmu równoległego w dwóch różnych środowiskach obliczeniowych.

Sowa kluczowe: *obliczenia równoległe, zagadnienie odwrotne, metody geoelektryczne*

1. Introduction

Computation plays significant role in analyzing and solving scientific and engineering problems. Due to increases in computational power many kinds of problems can be solved and more and more realistic solutions can be obtained nowadays. However, there are still various scientific disciplines where greater computational power and better modeling techniques are required. Inversion of geophysical data of more than one measurement sets done simultaneously can be an example of such problems.

* Department of Geoinformatics and Applied Computer Science, Faculty of Geology, Geophysics and Environmental Protection, AGH University of Science and Technology, Krakow, Poland, `apieta@geol.agh.edu.pl`, `jbala@geol.agh.edu.pl`

The main limitation of the inverse problem is the great number of computations which have to be preformed to find the values of the model parameters characterizing the system under investigation. Procedure of the joint inversion of geophysics data is an important method which can overcome the ambiguity of interpretation of these data. Applying joint inversion of two data sets increases computational time almost twice with respect of the computational time of single data set inversion.

One of the ways to overcome problem of time-consuming computationally intensive numerical algorithm is the application of the parallel computing environment. This technique is commonly used in solving geophysical problems [1]. In this paper parallel algorithm based on fine grain domain decomposition is presented. The goal of the paper is comparison of the performance of parallel inverse algorithm in two different computational environments. Algorithm is tested in cluster of personal computers and in supercomputer with cache-coherent non-uniform access (ccNUMA) architecture. Common used metrics of the parallel implementation for a given parallel computer has been also presented.

In the following sections briefly description of problem of the joint inversion of geoelectrical data and its results for a synthetic data (Section 2.1 and 2.2 respectively) has been presented. In the next sections (Section 3, 3.1 and 3.2 respectively) implementation of parallel joint inversion problem, technical details of the two test parallel environment and results obtaining during the test have been described.

2. Joint inversion in geoelectrical methods

In the inverse problem, the aim is to reconstruct the model on the basis a set of measurements. The process of solving the inverse problem is called in geophysics inversion. In the ideal case, inversion describes how the data should be transformed in order to reproduce the model.

In general there are two reasons why the estimated model differs from the true model. The first reason is the non-uniqueness of the inverse problem that causes several (usually infinite) models to fit the data. The second reason is that real data (and physical theories more often than we would like) are always contaminated with errors and the estimated model is therefore affected by these errors as well. Therefore a model appraisal has two aspects, non-uniqueness and error propagation [2]. An efficient way to overcome internal ambiguities is the use of the joint inversion, which means the integration of various groups of data records (arising from physically or geometrically different methods and surveys) into a single inversion algorithm. In this paper application of joint inversion of Vertical Electrical Soundings (VES) and Electromagnetic (EM) soundings has been presented.

Surface geophysical techniques like geoelectrical methods are non-invasive and cost-effective alternatives for obtaining information on groundwater (both quantity and quality). Such surveys usually are combined with selected test borehole observations and used in geological interpretation. Electrical and electromagnetic geophysical methods have been widely used in engineering, mining and groundwater investigations because of good correlation between electrical properties of rocks and fluid content of geological formations [3]. Among various geophysical methods, the direct current method is probably the most popular in groundwater studies due to the simplicity of the technique and easy interpretation of the data. The EM sounding method is relatively new. In Poland it has been developed more intensively since the mid-1980s and has been commonly used in environmental and hydrogeological investigation [4]. Both geoelectrical methods are often used together in geophysical measurments. For the reasons mentioned above these two methods were chosen to present results of the joint inversion process.

Similar to other geophysical methods, electrical (VES) and electromagnetic (EM) methods also suffer from ambiguity in interpretation due to the phenomena of the principle of equivalence [5]. Electromagnetic and electrical resistivity data sets are generated by the same physical property (electrical conductivity) of the Earth by different physical principles and hence these data sets can be regarded as genetically related. Therefore, when a conductivity structure of the subsurface is investigated, joint inversion of these data sets should yield better results in comparison with the joint inversion of EM or VES with other data sets which are not genetically related, for example seismic, gravity, magnetic, etc. [6]. Results presented in this work show that using joint inversion of electrical and electromagnetic sounding can decrease the ambiguity inherited to each method.

2.1. Synthetic Example

Geophysical data are often interpreted assuming horizontally layered geologic models. One dimensional (1D) inversion of Electromagnetic (EM) or Electrical (VES) resistivity data sets is a very simple and fast tool for mapping the vertical variation in the electrical conductivity of the Earth. In addition, 1D inversion results are very useful in constructing initial models for multidimensional interpretations. In this work synthetic curves (Fig.1b) for the three-layer geological model (Fig.1a) were generated to present the joint inversion results. In order to simulate measurement error random noise was added to generated data. The influence of joint inversion on ambiguity in the inverse problem in geoelectrical methods has been presented.

In this paper combination of Pure Random Search method and Multistart algorithm was applied [7]. Pure random search (PRS) is the simplest Monte Carlo algorithm. The domain of possible solution is surveyed at random with uniform distribution. The point where the function's value is minimal is returned as a solution. After a given number of samplings, the local method is lunched at the solution point. Multistart is similar to PRS, but the local method is launched after each sampling.

In presented algorithm computations were carried out in two stages. In the first step of calculations starting models were randomly chosen from parameter space. From all randomly selected starting models only 1000 models with the smallest matching errors (errors between field and theoretical curves – calculated as ε (1)) were chosen. From these points in the second step of calculation, minimization procedure was carried out. Powell's conjugate directions algorithm [8] was used as the local minimization method.

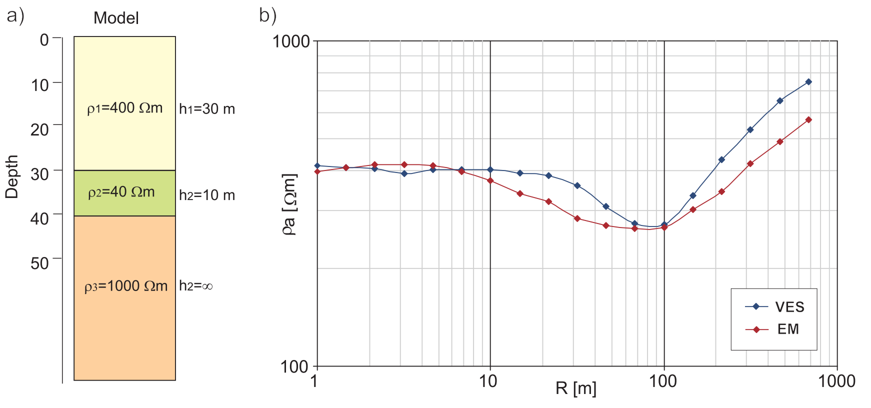

Fig. 1. The model (a) used for computations and genereted curves (b) for electrical (VES) and electromagnetic (EM) method

The main aim of the inverse problems is minimization of the chosen objective function also called error function. To combine EM and VES apparent resistivity data the following objective function was applied:

$$\varepsilon = \frac{1}{2N}\left[\sum_{i=1}^{N}\left(\frac{ln(\rho_i{}^o) - ln(\rho_i{}^c)}{ln(\rho_i{}^o)}\right)^2_{VES} + \sum_{i=1}^{N}\left(\frac{ln(\rho_i{}^o) - ln(\rho_i{}^c)}{ln(\rho_i{}^o)}\right)^2_{EM}\right] \tag{1}$$

where $\rho_i{}^o$ i $\rho_i{}^c$ are observed and computed apparent resistivity for Vertical Electrical Sounding (VES) and Electromagnetic (EM) sounding, and N is a number of measurements.

2.2. Results

In order to analyse the ambiguity of solutions, models computed by separated and joint inversions are presented (Fig.2). Grayscale shows measured (generated synthetic curve) and computed curves matching (error function value (1)). The red line in this pictures presents assumed model (Fig.1a).

For the separated inversion a lot of models with small value of error function (dark lines) were obtained. But most of them significantly differ from the assumed one (red line). In case of application of joint inversion most obtained models have smaller value of error function and most often they are close to the true one (Fig.1a).

The correlation matrix (Table 1) helps to understand the relation between different model parameters. There is a strong positive correlation between ρ_2 and h_2. This means that if ρ_2 and h_2 are both increased or decreased, the apparent resistivity curve will not change. It is connected with the equivalence phenomenon. Analysis of the obtained outcomes demonstrate that the use of joint inversion gives significantly better results in solving values of parameters with equivalence in comparison with separate inversions of VES and EM soundings.

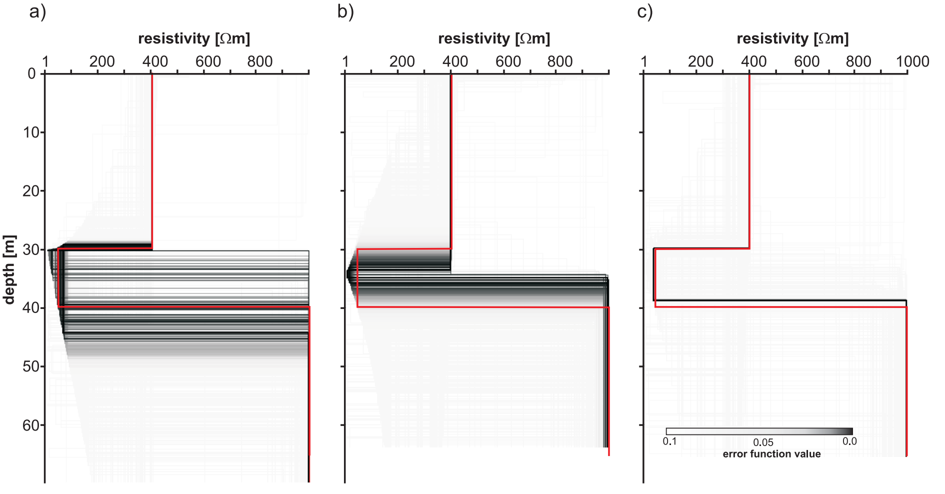

Fig. 2. The obtained final models compared to the true one (red line) for individual inversions of VES (a) and EM (b) and joint inversion (c)

Table 1
The correlation matrix for model parameters

	ρ_1	h_1	ρ_2	h_2	ρ_3
ρ_1	1	−0.44	−0.24	0.23	−0.11
h_1	−0.44	1	−0.19	0.17	0.13
ρ_2	−0.24	0.19	1	**1.00**	−0.77
h_2	0.23	0.17	**1.00**	1	0.78
ρ_3	−0.24	0.13	−0.77	0.78	1

3. Implementation of the joint inversion algorithm

Parallelism was introduced into the inverse problem by decomposition of the optimization process [9]. Parallel algorithm based on the master -– slaves paradigm and domain decomposition was used. In this algorithm either master or slave computational nodes have complete information about two analyzed geological data structure. In the first step of computation master node randomly search parameter space in order to select the best starting points for local minimization procedure. All points with values of error function (see Equation 1) smaller than assumed are sent to slave nods. In this step master-node send starting points to all slave-nodes successively. In the second step of computation master and slave nodes repeat sending – receiving sequences. Master send starting points to idle slave node whenever it send him back the results of previous minimization procedure, then slave launch local minimization procedure and so on. When slave-nodes are searching for local minimum of objective function, master node randomly searches parameter space for next suitable starting point. Parallel optimization algorithm is finished when satisfying amount of data is collected. This algorithm is shown in Figure 3.

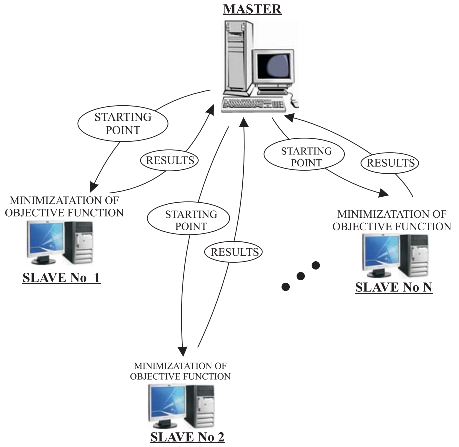

Fig. 3. Illustration of parallel joint inversion algorithm

The algorithm for the parallel joint inversion of geoelectrical data can be then written as follows:

Master node algorithm:

```
BEGIN
 Startup data structure, variables and constraints
 Read two analyzed geoelectrical data sets
  REPEAT
    Randomly choose starting point from parameter space
    Evaluate error function
    IF error value is smaller then assumed value THEN
       send the starting point to idle slave node
       save the received results
    ELSE
       Continue
    ENDIF
  UNTIL satisfying amount of data is collect (1000 here)
END
```

Slave nodes algorithm:

```
BEGIN
   Startup data structure, variables and constraints
   Read two analyzed geoelectrical data sets
        REPEAT
         launch local minimization procedure from point sent by
         master node sent the result to master node
        UNTIL master node finish collecting all results
END
```

The computational time of parallel joint inversion algorithm for the same model and the same computational environment can be different. This is due to the fact that time of single minimization procedure can be different even if nearest parameter space points are used as starting points for local minimization method.

3.1. Test environment

Parallel joint inversion algorithm was performed in two different environments. An example of shared and distributed memory computational machine was used to test parallelization of joint inversion of geoelectrical data. Whole code was written in C language. MPICH, freely available, portable implementation of Message Passing Interface was used to implement parallel algorithm. The first environment was PC cluster of several nodes. Each node has following parameters: processor with hyper-threading technology (Intel Pentium 4, 2.8 GHz), 1024 MB RAM, Gigabit Ethernet network adapter All nodes in cluster were connected by Gigabit Ethernet switch. Computation was done in Fedora Core 3, (kernel: 2.6.12-1.1381 [smp]) operating system.

The second environment was 128-processor SGI Altix 3700 machine in ACK CYFRONET AGH, the cache-coherent non-uniform access multiprocessor architecture with memory physically distributed among the nodes but globally addressable to all processors through the interconnection network. Each processor has following parameters: processor Intel Itanium 2, 1.5 GHz and 512 GB RAM. Computation was done in SUSE Linux Enterprise Server 10 operating system. In the following sections the performance and metrics of the parallel join inversion optimization algorithm in both parallel environmental are shown.

3.2. Results

The commonly used measurements of the performance of the parallel programs like speed up and efficiency has been presented in Figure 4 and 5.

There are relatively small differences between measurements of the performance in both parallel computational environments. For both PC cluster and shared memory environment decreasing of computational time with increasing of number of computational nodes can be seen (Fig. 4 and 5). Difference between computational time for a given size of computational environment are due to differences in parameters of computational nodes (PC cluster has better parameters of a single processor).

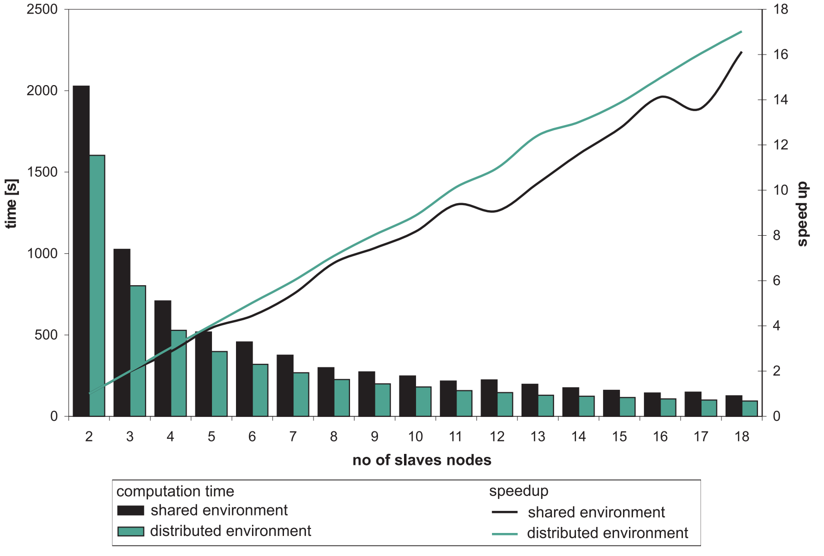

Fig. 4. Relation between speed up and number of computers for shared and distributed parallel environments

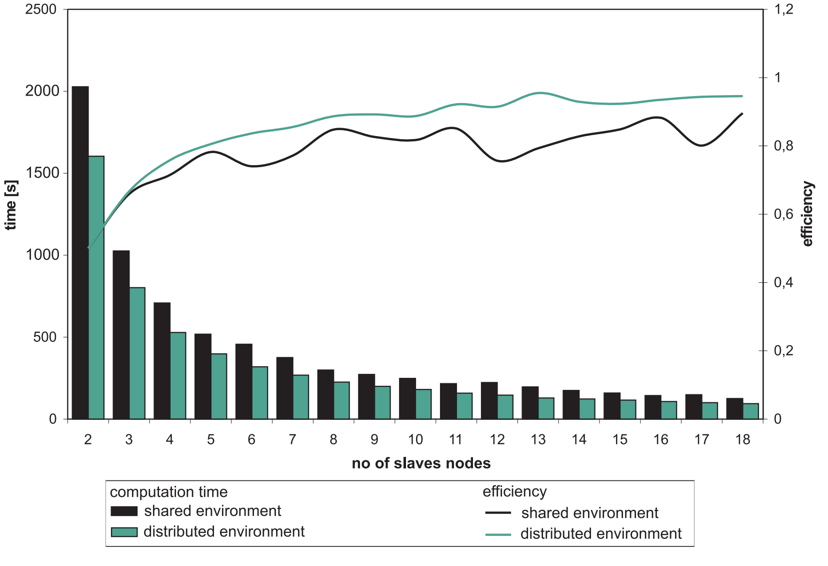

Fig. 5. Relation between efficiency and number of computers for shared and distributed parallel environments

For both computational environment speedup and efficiency relationship show that more computational nodes we use, better speedup we receive and more efficient we utilize the computational environment (Fig.4 and 5 respectively). Speedup and efficiency relationship received for PC cluster computational environment increase monotonously whereas the same curves for shared memory environment are non-uniform.

There are two possible reasons for such behavior: efficiency of communication layer or differences in execution time of single minimization procedure due to Monte Carlo type of computation. In Figure 6 histograms of execution time of the of single minimization procedure measured for inverse algorithm launched twice times (series A and B) are presented.

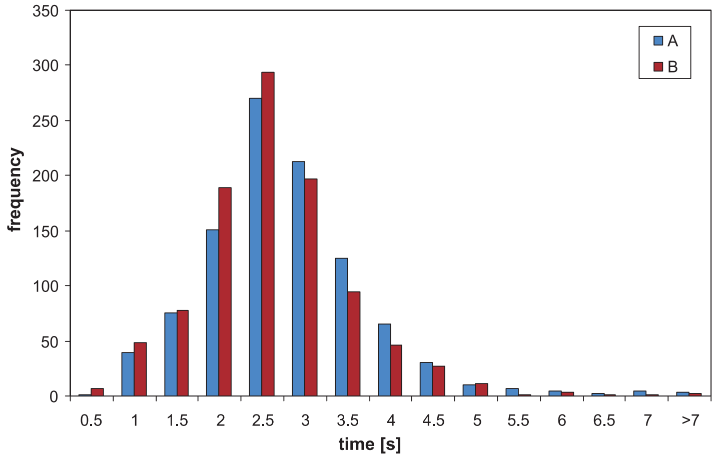

Fig. 6. Histogram of the computational time for single minimization procedure for twice joint inversion of geoelectrical data

Chi-square test of homogeneity was applied to determine whether times of single minimization procedure for both realization of joint inversion algorithm come from a specific distribution. The null hypothesis states that both series come from the same distribution cannot be rejected at the 0.01 level of significance. This proves that the Monte-Carlo type of computation has insignificant influence on differences between the computational times of the joint inversion parallel algorithm. Differences between speed up and efficiency curves are caused mainly by the different architecture of both computational environment and implementation of Message Passing Interface with suits better for distributed memory computational environment.

4. Summary

In this paper parallel joint inversion code suitable for both distributed memory environment (PC cluster of Intel Pentium 4) and shared memory environment (SGI Altix

3700) has been presented. Commonly used parallel execution measurements like execution time, speed up and efficiency has been presented for both computational environments. Speed up and efficiency relations received for parallel implementation of joint inversion algorithm have generally the same character for both computational platforms. The differences between speed up and efficiency curves that can be noticed are caused mainly by different architecture of both computational environments and implementation of Message Passing Interface. The other feature that can increase differences among speed up and efficiency relation for both computational platforms is different execution time of a single local minimization event. Differences among time of local minimalization events are always observed and are independent from computational environment. They depend only on randomly chosen starting point and, as was shown above, it has insignificant influence on difference between execution time, speedup and efficiency of joint inversion parallel algorithm observed in two computational environments.

Result presented in this paper shows that the fine grain decomposition proposed for joint inversion of geoelectrical data suits well for both shared and distributed memory environment.

Acknowledgements

This work was supported in part by grant no. MEiN/SGI3700/AGH/033/2006 and project no.11.11.140.561 AGH University of Science and Technology (Faculty of Geology Geophysics and Environmental Protection, Department of Geoinformatics and Applied Computer Science).

References

[1] Danek T., Franczyk A.: *Parallel and distributed seismic wave-field modeling.* TASK Quarterly: scientific bulletin of Academic Computer Centre in Gdansk, vol. 8, 2004, 573–582

[2] Snieder R., Trampert J.: *Inverse Problems in Geophysics.* New York, Springer 1999

[3] McNeill J. D.: *Use of Electromagnetic Methods for Groundwater Studies. Geotechnical and Environmental Geophysicsv.* vol. 1, 1990, Society of Exploration Geophysicists, Tulsa, OK, 191–218

[4] Moscicki W., Antoniuk J.: *Application of geoelectric methods into studying of geological environment influenced by human activity.* Publications of the Institute of Geophysics Polish Academy of Sciences, 2002, 179–193

[5] Zhdanov M. S., Keller G. V.: *The geoelectrical methods in geophysical exploration.* Amsterdam, Elsevier 1994

[6] Sharma S. P., Kaikkonen P.: *Appraisal of equivalence and suppression problems in 1D EM and DC measurements using global optimization and joint inversion.* Geophysical Prospecting, vol. 47, 1999, 219–249

[7] Pszczoła G., Leśniak A.: *Non-linear optimization methods for small earthquake locations*. TASK Quarterly: scientific bulletin of Academic Computer Centre in Gdańsk, vol. 8, 2004, 583–590

[8] Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P.: *Numerical Recipes in C: the art of scientific computing*. Second Edition, New York, Cambridge University Press 1992

[9] Schnabel R. B.: *A Wiew of the Limitations, Opportunities and Challenges in Parallel Nonlinear Optimization*. Parallel Computing, vol. 21, 1995, 875–905