Performance Characterization of a Marine CSEM Inversion Application

Jessica Imlau Dagostini, Henrique Corrêa Pereira da Silva, Vinicius Garcia Pinto,
Eduardo Simões Lopes Gastal, Lucas Mello Schnorr
Institute of Informatics, Federal University of Rio Grande do Sul – UFRGS, Porto Alegre, Brazil
{jidagostini, hcpsilva, vgpinto, eslgastal, schnorr}@inf.ufrgs.br

***Abstract*—The prospection of areas for oil exploration is an expensive and risky process. For this reason, several computational simulations are performed to estimate the probability of finding a viable petroleum reservoir before drilling it. In this paper, we characterize the performance of a parallel application based on the marine Controlled Source Electromagnetic Method. We have identified the dominant step, and after parallelizing it, we outperform the original code with a speed-up of ≈7 times. We also detect that there is room for future gains by optimizing the workload distribution.**

***Index Terms*—oil prospection, CSEM method, MPI, OpenMP**

I. Introduction

One of the main challenges in the oil industry is to correctly map areas in the ocean where there is a higher probability of oil and gas. For this mapping, it is necessary to apply a numerical method that needs to be as precise as possible to give reliable information. Physical oil exploration has a high monetary cost, usually requiring amounts in the order of millions. Having ways to do accurately computational simulations to help with this process is essential to these industries since it can reduce the incidence of unsuccessful operations, thus saving time and monetary resources.

The usage of electromagnetic survey methods in such offshore contexts now sees an increase as another way to remove doubt from the costly exploratory process. One known method to simulate and computationally investigate ocean lands is the marine Controlled Source Electromagnetic (mCSEM) [1]. This method uses maritime receptors – fixed in the ocean floor – and uses a portable source that emits electromagnetic signals that are kept from these receptors to gather data.

This electromagnetic surveying is applied to recognize the seafloor's aspects, thus giving relevant information about the explored region's conditions. From this information, it is possible to run numerical methods and apply mathematical inversion to matching subsurface model. The data processing step takes increased attention, as it is the deciding factor that guarantees efficient employment of the collected data before the visualization phase.

This study was financed by the "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior" (CAPES) - Finance Code 001, the CNPq - Brazil scholarship for the 2º and 3º author, and the projects: FAPERGS MultiGPU (16/354-8) and GreenCloud (16/488-9), the CNPq project 447311/2014-0, the CAPES/Brafitec EcoSud 182/15, the CAPES/Cofecub 899/18, Petrobras (2016/00133-9, 2018/00263-5). The experiments presented in this paper used the PCAD infrastructure, http://gppd-hpc.inf.ufrgs.br, at INF/UFRGS.

As the exploration of specific areas in the seafloor can generate a considerable amount of data, it becomes necessary to use a high-performance application to inversion faster. Adopting a parallel approach to this high-performance application, we need to guarantee that such parallelization is efficient in using of all computing resources by stressing each resource at its maximum.

The present work aims to characterize and demonstrate improvements in the performance of an inversion mCSEM application for mapping possible oil and gas regions. This is a third-party real world application, coded in C language, using MPI and OpenMP for code parallelization. This paper is divided into four sections. Section II introduces a background about the problem, the application, and the data used to run the simulations. Section III presents the initial analysis of the application, alongside the solution of one problem in performance and the discovery of a new one. For the last, Section IV presents the conclusions and future work.

II. Basic Concepts

Electromagnetic surveying is a complex process, containing multiple steps – both physical and informational – before providing data to interpretation. Our characterized application performs a fine-grained electromagnetic data inversion to reconstruct a model of the underlying substrate based on a marine Controlled Source Electromagnetic surveying. This approach enables superior control over all the exploratory stages of offshore reservoirs and a fine-grained control in the interpretation phase of the exploratory process.

A. Offshore Exploration

In the era of rising green energy, the continuing economic development and population growth are expected to push the global energy needs in a way that outpaces the establishment of cleaner alternatives. To meet this demand, the oil and gas industries project to fill as much as 50% of the primary energy sources before the middle of this century [2]. However, easily accessible hydrocarbon reserves are either depleted or already developed, so these companies require faster, more precise acquisition and visualization techniques in order to explore more complex undeveloped options [3].

A particularly complex area of operations is offshore exploration. Deep under salt layers, these reservoirs often pose a challenge to traditional seismic surveys, as different fluids, like hydrocarbons and water, prove difficult to differentiate [4]. To provide better data before the costly physical exploratory drilling operations, electromagnetic methods can be used to provide important complementary information to aid in the decision-making process of those operations.

Figure 1 diagrams the exploratory process. The operations start with a previous formal geologic knowledge of the area and its formation, which indicates the presence of traps or other characteristic rock sequences. Then, data from one or more surveys feed visualization tools that further assist the decision-making of supplementary actions, which are either more surveys or the drilling of an exploratory well, a rather expensive operation in offshore contexts. Found the existence of oil or gas, further surveys may follow to more precisely define the extension of the reservoirs before the start of the extraction operation.

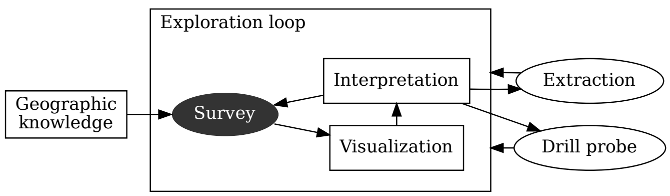

Fig. 1. The decision-making process in resevoir exploration.

B. Controlled Source Electromagnetism

Electromagnetic (EM) surveying is applied in both land and maritime contexts – though these have developed almost independently – and either actively or passively. In maritime controlled source electromagnetism (mCSEM), a towed human-made electromagnetic source emits a low-frequency continuous signal that potentially traverses water, surface, and deep into the seabed before being captured by an array of receptors either placed on the seafloor or also towed along with the source [5]. Figure 2 represents the static receivers mCSEM approach, where receptors placed at the seafloor capture reading from the boat-towed EM source. These approaches aim to collect readings of the EM signal after it crossed the underlying salt and rock sequences, altering it and, ultimately, imprinting its distinct characteristic onto the signal [6].

When crossing conductive substances of varying resistivity, the emitted wave's amplitude gets altered according to the variation on the conductive material. Such process obeys Maxwell's equations, and so could be though as the application of a direct method over an existing input: the traversal of an EM wave over the subsurface salt and rock layers [6]. In other words, the materials underground imprinted their different conductivities onto the now read EM field values. Therefore, the aim is to model back these resistivities from the read values through the inverse method, matching the readings to the most probable model of the subsurface substances that generated such readings.

However, unwanted signals contribute to the reading as well, like the wave that traveled directly to the receptor, the air-

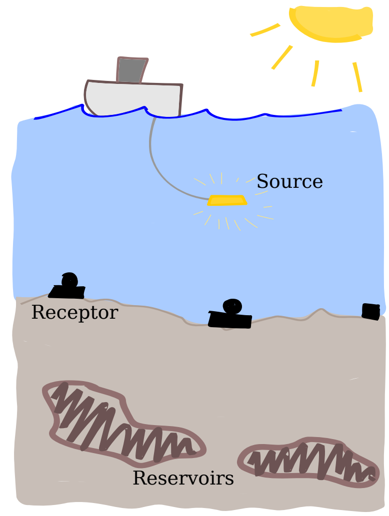

Fig. 2. Representation of the CSEM surveying process.

traversed wave (airwave), and magnetotelluric signals [4]. As to whether preprocess these readings in hopes to alleviate the influence of these signals is part of the interpretation method applied, as deepwater operations do not pick up as much airwave noise as shallow water receivers. However, noise considerably influences the sensitivity of the model to the desired ground-traversed waves.

C. CSEM Data Inversion Processing

The inversion method – as in the reconstruction of a model of resistivities that could have caused our readings – can be done with and without an initial input model. The application we characterize implements the former approach, which can prove especially useful when tied together with *a priori* geographic knowledge of the underlying substrate. This allows fine-grained control in the interpretation phase of the exploratory process, as the model-driven inversion is acutely sensitive to these initial values and so could achieve a more faithful match to the underlying rock layers' resistivities [4].

The industry's CSEM inversion methods are iterative solvers, as this is essentially an optimization problem. We generally see two phases to such algorithms: the **simulation** calculation and the **forward** step. In the former, the solver uses the current model to simulate EM signal values and then to compare the result to the read values read in the survey, checking for convergence according to a misfit delta value. The simulation requires solving of large sparse equation system and is achieved in this application using a Cholesky factorization. In the last-mentioned step, the solver forwards the model to the next iteration. This can be accomplished in several ways, like using CMP (Common Midpoint) [7], but the details of the chosen approach are part of the company's intellectual property and escape the scope of this work.

Our characterized application follows the logic outlined in Figure 3. Initially, from the highlighted input resistivity model, the application simulates synthetic data through the direct method, using the Maxwell equations that model the

EM interactions. The input survey data is then compared to the generated data and, if within a defined misfit delta, the inversion ends, as a well-matching model has been found. Otherwise, the application forwards the existing model, updating it to the next simulation iteration.

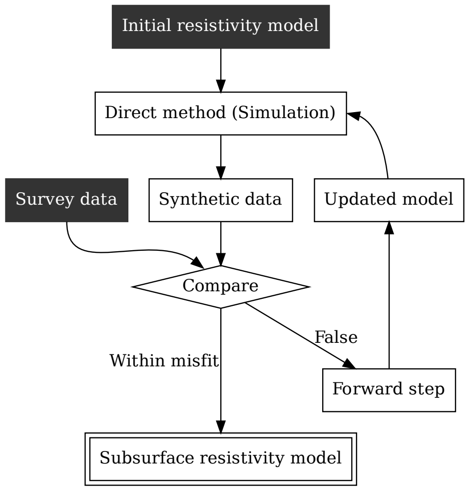

Fig. 3. Diagram representing the inversion algorithm.

III. Methodology and Results

To understand the code's behavior and provide improvements to its execution, we first characterized the application by doing a more in-depth analysis of the code. We then instrumented it by using a tracer application. By these tracing results, we were able to provide some performance gains and found other notable issues.

A. Platform Configuration

Table I shows the platform configuration of the **cei** cluster used in the executions presented at Subsections III-B, III-C, and III-D. All experiments in this work used the PCAD infrastructure, http://gppd-hpc.inf.ufrgs.br, at INF/UFRGS.

TABLE I
Configuration of Machines Used in the Experiments

	cei
Nodes	5
Processors per Node	2
Processor	Intel Xeon 4116 2.10GHz
Cores per Processor	12
Core count	24
Memory	96 GB DDR4
Interconnection	100Mbit/s
Operating System	Debian 10
Linux kernel	4.19.0-8-amd64
OpenMPI	3.1.3
GCC	8.3.0

B. Performance Characterization

The original code was already parallel, exploring both shared and distributed memory paradigms using both OpenMP and MPI. As an initial step, we have instrumented the code using ScoreP [8] to characterize each application phase's performance. The top plot of Figure 4 shows a first visualization of the obtained execution traces of this original code. We split the main region of the application in three phases: **forward**, **cholesky**, and **correction**.

One can observe that only the first node (rank 0) execute the **cholesky** step (in blue) while all other ranks remain idle waiting in a global barrier. The **forward** (red) and **correction** (green) steps are executed in parallel by all ranks, but the last ones (**correction**) are too short to be visible in the graphic.

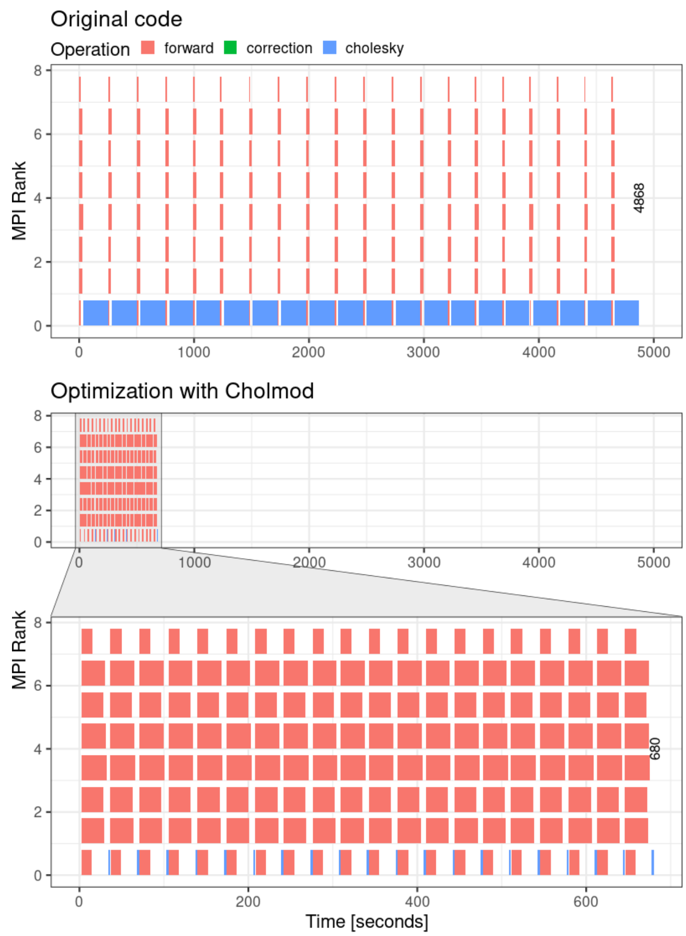

Fig. 4. Execution Trace Visualization: Original implementation (top); after Optimization with Cholmod (bottom)

C. Sparse Cholesky Solver

Standard libraries of linear algebra kernels usually follow a common API as BLAS, which makes easy switch among different implementations (e.g., Intel MKL, OpenBlas, CUBLAS). Since the original code uses a built-in implementation of the Cholesky factorization, we decided to replace it by integrating a standard one. We chose the **Cholmod** module [9] from the Suitesparse library [10], which is multithread, and proved suitable for the application and its current workload. Figure 4 shows the performance gains of this optimization, reducing the makespan from 4868 seconds to 680 seconds, i.e., a speed-up of $\approx$7 times. The **cholesky** steps are no more time dominant, with **forward** steps being now the most representatives (see zoom area at the bottom of Figure 4).

D. Load Imbalance Issues

Even with the performance gains obtained through the Cholesky parallelization, we still could notice some load problems. Taking a closer look at Figure 4, we note that the time spent in each MPI Rank during the **forward** execution is different among them. Thus, we are faced with a new load imbalance in the application execution.

The **forward** region is responsible for evolving the model by using CMPs information. Each CMP has a different amount of combinations of three data: the receptor, the source, and the frequency. These combinations define the valid values used in the model evolution and, as we could statistically prove, are directly responsible for the time spent in the **forward** phase.

The distribution algorithm used by the application considers sharing the total CMPs among MPI Ranks. However, each CMP has a different number of combinations to process, which causes the different workload to each rank. Figure 5 shows the direct relation between process time and quantity of the combinations by a rank.

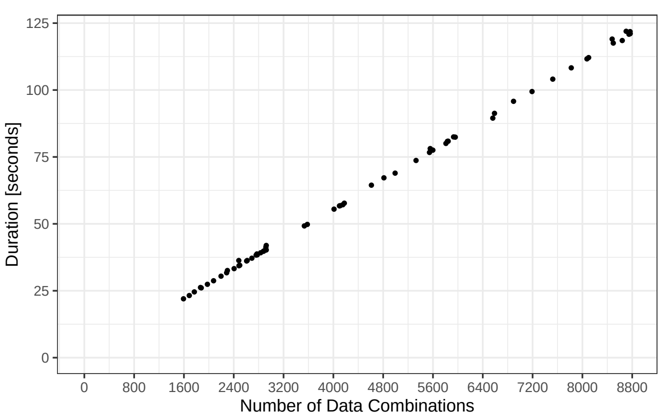

Fig. 5. Direct relation between combinations and workload by rank.

Therefore, we apply ANOVA (ANalysis Of VAriance) [11] to verify the direct association between the number of combinations allocated among MPI ranks and its execution time. The results confirm our hypothesis, with a statistical confidence of 99.7%. By them, we aim to find a better load distribution between the ranks, to balance their workload and then reduce the execution time.

As each combination of three information is directly related to a CMP, we aim to perform this distribution by allocating different CMPs to each rank without splitting its data combinations in different ranks. Thus, we need to allocate the CMPs in a combination that balances how many groups of three combinations we have in each rank. Some distributions algorithms are already in our roadmaps, such as the Round Robin and Greedy algorithms.

IV. Conclusion and Future Work

In this paper, we have characterized the performance of an application for oil prospection. This application uses the marine Controlled Source Electromagnetic Method. Even if the original code is already parallel using MPI and OpenMP, we have identified a sequential step that executes a sparse **cholesky** factorization. After replacing it by a parallel implementation provided by the **Cholmod** module, the makespan reduces by a factor of $\approx$7.

Our optimization on the **cholesky** step has highlighted another performance issue, the unbalanced workload among the MPI ranks. As future work, we plan to explore different load balancing algorithms to achieve more performance gains and to investigate different data decomposition (CMPs). We also plan to extend our analysis with new artificially created workloads in order to evaluate the scalability of the envisaged solutions.

References

[1] V. C. T. de Souza, "Modelagem Numérica de Dados mCSEM 3D Usando Computação Paralela," Ph.D. dissertation, Universidade Federal do Pará, 2007. [Online]. Available: http://repositorio.ufpa.br:8080/jspui/handle/2011/5673

[2] Z. Liu, "Chapter 4 - supply and demand of global energy and electricity," in *Global Energy Interconnection*, Z. Liu, Ed. Boston: Academic Press, 2015, pp. 101 – 182. [Online]. Available: https://doi.org/10.1016/B978-0-12-804405-6.00004-X

[3] Z. Ningning, W. Qing, W. Jianjun, H. Lianhua, L. Haowu, and L. Qian, "Characteristics of oil and gas discoveries in recent 20 years and future exploration in the world," *China Petroleum Exploration*, vol. 23, no. 1, p. 44, 2018. [Online]. Available: http://www.cped.cn/EN/10.3969/j.issn.1672-7703.2018.01.005

[4] A. Ziolkowski and D. Wright, "The potential of the controlled source electromagnetic method: A powerful tool for hydrocarbon exploration, appraisal, and reservoir characterization," *IEEE Signal Processing Magazine - IEEE SIGNAL PROCESS MAG*, vol. 29, pp. 36–52, 07 2012. [Online]. Available: https://doi.org/10.1109/MSP.2012.2192529

[5] M. Wilt and D. Alumbaugh, "Electromagnetic methods for development and production: State of the art," *The Leading Edge*, vol. 17, no. 4, pp. 487–487, 1998. [Online]. Available: https://doi.org/10.1190/1.1437997

[6] S. Constable and L. J. Srnka, "An introduction to marine controlled-source electromagnetic methods for hydrocarbon exploration," *GEOPHYSICS*, vol. 72, no. 2, pp. WA3–WA12, 2007. [Online]. Available: https://doi.org/10.1190/1.2432483

[7] J. L. Silva Crepaldi, M. P. Pereira Buonora, and I. Figueiredo, "Fast marine CSEM inversion in the CMP domain using analytical derivatives," *Geophysics*, vol. 76, no. 5, pp. F303–F313, sep 2011. [Online]. Available: https://doi.org/10.1190/geo2010-0237.1

[8] A. Knüpfer, C. Rössel, D. a. Mey, S. Biersdorff, K. Diethelm, D. Eschweiler, M. Geimer, M. Gerndt, D. Lorenz, A. Malony, and et al., "Score-p: A joint performance measurement run-time infrastructure for periscope, scalasca, tau, and vampir," *Tools for High Performance Computing 2011*, p. 79–91, 2012. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-31476-6_7

[9] T. A. Davis, "User guide for cholmod: a sparse cholesky factorization and modification package," *Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, USA*, 2008.

[10] S. Kolodziej, M. Aznaveh, M. Bullock, J. David, T. Davis, M. Henderson, Y. Hu, and R. Sandstrom, "The suitesparse matrix collection website interface," *Journal of Open Source Software*, vol. 4, no. 35, p. 1244, Mar 2019. [Online]. Available: http://dx.doi.org/10.21105/joss.01244

[11] E. George, J. S. Hunter, W. G. Hunter, R. Bins, K. Kirlin IV, and D. Carroll, *Statistics for experimenters: design, innovation, and discovery*. Wiley New York, 2005.