# TunaOil: A Tuning Algorithm Strategy for Reservoir Simulation Workloads

Felipe Portella*†‡, Josep Ll. Berral*†
*Petróleo Brasileiro S.A. (PETROBRAS), Rio de Janeiro, Brazil
†Barcelona Supercomputing Center, Barcelona, Spain
‡Universitat Politècnica de Catalunya, Barcelona, Spain
E-mail: felipe@portella.com.br, josep.berral@bsc.es

***Keywords—Reservoir Simulation, Algorithm Configuration, Machine Learning, Parameter Tuning***

## I. Extended Abstract

The state-of-the-art techniques to tune the numerical parameters of reservoir simulators are based on running numerous simulations, specific for that purpose, to find good candidates. As the simulations for real petroleum fields require a considerable amount of time, optimizing parameters using this approach is costly in terms of time and computing resources. The main objective of this work, therefore, is to present a new methodology to optimize the numerical parameters of the reservoir simulations. It is common in the oil and gas (O&G) industry to use ensembles of models in different workflows to reduce the uncertainty associated with the forecasting of O&G production. We can leverage the runs needed to create such ensembles, to extract the information we can use to optimize the numerical parameters in future runs.

To achieve this, we mine past execution logs from many simulations with different numerical configurations and build a performance model that is based on features extracted from the data. This performance model takes general information about petroleum fields and the simulation parameters as inputs, allowing it to generalize to different unseen reservoir models. Experiments show that the presented system can correctly produce good configurations in a much-reduced time, within a history matching workflow that generates hundreds of simulations.

### A. Reservoir Simulations and History Matching

Essentially, reservoir simulation allows engineers to replicate the history of the production of oil, gas, and water from the reservoir over a time frame to forecast the future; this provides answers to a series of questions that are critical to different business strategies, for exploiting the oilfield.

Engineers are constantly looking for efficient tools to tune the simulation process to make it faster and achieve better decisions. The reservoir simulators available on the market allow users to tune numerical parameters, which can affect the performance and quality of the simulation significantly. However, these numerical parameters vary among simulators, making the selection of them difficult. Moreover, the parameter space is big and co-relations can exist between parameters, making this a non-trivial manual task.

The utility of the reservoir model, however, results from its ability to predict the behavior of the reservoir field in terms of the production of water, oil, gas, pressure, etc. To calibrate the model, engineers use a method called *History Matching* (HM). The explicit purpose of HM is to assign values to the parameters of the model to be optimized, such that it replicates the behavior observed during a past production period, leading to better forecasting. One tool used for automatic HM is the family of Kalman filters.

A Kalman filter (KF) is a mathematical method – robust to noise in the data – that uses all of the observed measurements, to produce estimates of unknown variables in linear systems. The ensemble Kalman filter (EnKF) is a Monte Carlo extension of the Kalman filter, capable of working on non-linear systems, which is applicable to the problem of HM in reservoir simulation. Various other extensions have been proposed for the petroleum industry. Nowadays, the standard HM method used by energy companies is the ensemble smoother with multiple data assimilation (ES-MDA) [1].

### B. Proposal

The objective is to develop a performance model that can achieve a faster overall ES-MDA runtime by dynamically tuning the simulations being executed. To improve the optimization process, we extract a feature vector from the output logfile of the simulation execution. This feature vector gathers important data about the underlying execution that is used to refit a performance model trained with more than 20,000 different reservoir simulations, leading to a better oracle.

### C. Experimental Environment

TunaOil was evaluated with three black-oil reservoir models – listed in Table I – that have multiple geological realizations. A realization is an uncertain representation of the rock-fluid properties, such as porosity, horizontal and vertical permeability, net-to-gross, and initial water saturation. Table I shows the number of realizations available, the mean elapsed time of the realizations in seconds (simulated with the default parameters or the engineer manually-tuned parameters) and the number of simulations with different configurations performed in the ES-MDA workflow for each reservoir model. The times reported are for the reservoir simulator using a 48-core (without Hyper-Threading) node.

The number of simulations represents the total number of executions required by the ES-MDA algorithm to perform the HM process, as it simulates each realization five times. Therefore, the total time to run the OLYMPUS case, considering the default numerical parameters of the reservoir simulation, was slightly more than 8 hours (119 seconds multiplied by 250 executions). Using the manual configuration selected by the reservoir engineer, the total time was over 28 hours. All

these timings are cumulative, as in a computational cluster with multiple nodes, we can run some cases at the same time on different nodes, reducing the wall time (the perceived total time for the end-user). As the ES-MDA needs the results of all the realizations to apply the Kalman Gain calculation and prepare the next batch of realizations to be simulated, the maximum number of nodes to be used in parallel is defined by the number of realizations. Considering the same example of the OLYMPUS case, by using 50 nodes, we reduce the wall time to the end-user to roughly 10 minutes. The same applies to all the workloads listed in the table.

**OLYMPUS** [2] is a synthetic reservoir model developed by TNO in 2017 for a benchmark study on field development optimization. The model has a grid of 341K cells. The **UNISIM-I** [3] is a synthetic model based on a real data sample from the Namorado Field in the Campos Basin, Brazil, while the **UNISIM-II** [4] is a synthetic model based on a carbonate offshore reservoir that represents the Brazilian pre-salt. The UNISIM-II has 190K cells, while the original UNISIM-I has 93K cells. However, the UNISIM-I directly used in our work was the fine geological model, which has more than 11M cells.

### *D. Results*

Our experiments show that TunaOil can improve the execution time of the base case by up to 40%, increasing the material balance error, on average, by less than 1% and the gas, oil, and water production error by less than 2%. Figure 1 shows the best result achieved by our methodology among the OLYMPUS workload – detailed in Section I-C – when compared to the engineer configuration. This figure shows the overall system performance impact on the ES-MDA workflow. The speed-up of the simulations was evaluated together with the impact on the quality of the outputs produced by the simulations.

### *E. Prior Unsuccessful Work*

The original proposal was to develop a general oracle that can present good numerical parameters for any unseen reservoir model. The first attempts to reach that goal were unsuccessful, leading to performance models that, in many cases, were even worse than the default parameters used by the reservoir simulator. When the suggestions led to good execution times, the results of the simulations were outside the acceptable engineer-error margin. It appears that "similar" models can perform differently due to factors that are difficult to characterize, that is, the degree of heterogeneity and non-linearities resulting from the characteristics of the problem, such as flow rates, mass transfer between phases, etc. The solution to overcome these issues was to include the oracle inside a workflow that simulates the "same" reservoir model multiple times, such as in an HM process. That way, we can use the first iterations to refit the performance model, providing the extra missing knowledge to characterize it.

### *F. Conclusion*

This work introduces the use of a performance model to dynamically tune the numeric parameters of petroleum reservoir models, reducing the overall application runtime without the need for additional simulations or a separate optimization study. Our experiments demonstrated that the oracle built was able to predict the proper effect of the changes in the solver options, in terms of simulation time and quality. The experiments have shown that our oracle makes accurate predictions in a broadly used workflow in the petroleum engineering area with black-oil models. The idea can be easily extended for other types of workflows, such as optimization processes or other types of models, such as compositional models. Ultimately, it would be feasible to couple the oracle developed in the central scheduler system of an energy company, such as Petrobras, to perform live optimization of any reservoir simulation being submitted to their HPC infrastructure, reducing time and associated costs.

| Reservoir Model | Number of Realizations | Mean Default Time | Mean Engineer Time | Number of Simulations |
|---|---|---|---|---|
| OLYMPUS | 50 | 119 | 411 | 250 |
| UNISIM-II | 500 | 669 | 645 | 2500 |
| UNISIM-I Fine | 48 | 72.071 | 56.360 | 240 |

TABLE I. WORKLOADS USED IN THE EVALUATION (TIMES IN SEC).

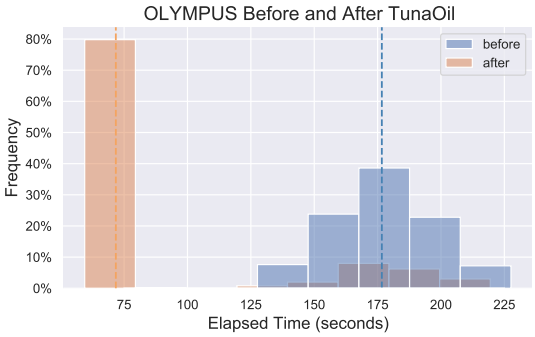


Fig. 1. Histogram of the OYMPUS simulations executed in an ES-MDA with and without TunaOil. The dashed lines represent the medium of the values.

## II. ACKNOWLEDGMENT

The author would like to thank PETROBRAS for funding this work, CMG for providing the simulator used in this research, and LNCC for providing their HPC infrastructure for the experiments. This work is under submission to the Computers and Geosciences Journal.

## REFERENCES

[1] A. A. Emerick and A. C. Reynolds, "Ensemble smoother with multiple data assimilation," *Computers Geosciences*, vol. 55, pp. 3–15, jun 2013.

[2] R. Fonseca, E. Della Rossa, A. Emerick, R. Hanea, and J. Jansen, "Overview Of The Olympus Field Development Optimization Challenge," in *16th European Conference on the Mathematics of Oil Recovery (ECMOR XVI)*, Barcelona, Spain, sep 2018.

[3] G. D. Avansi and D. J. Schiozer, "UNISIM-I: Synthetic Model for Reservoir Development and Management Applications," *International Journal of Modeling and Simulation for the Petroleum Industry*, 2015.

[4] M. Correia, J. Hohendorff, A. T. F. S. Gaspar, and D. Schiozer, "UNISIM-II-D : Benchmark Case Proposal Based on a Carbonate Reservoir," in *SPE Latin American and Caribbean Petroleum Engineering Conference*, Quito, Ecuador, 2015.

**Felipe Portella** received his degree in Informatics in 2003 and an M.Sc in Computer Science in 2008 from PUC-Rio. He is an IT Consultant at the Brazilian energy company Petróleo Brasileiro S.A. (PETROBRAS), working with petroleum reservoir simulation workloads in HPC environments. He is currently a Ph.D. student at the Universitat Politècnica da Catalunya (BarcelonaTech-UPC) in partnership with the "Data-Centric Computing" group of the Barcelona Supercomputing Center (BCN-CNS).