J Supercomput (2007) 41: 109–117
DOI 10.1007/s11227-006-0031-2

Parallel history matching and associated forecast at the center for interactive smart oilfield technologies

Ken-Ichi Nomura · Rajiv K. Kalia · Aiichiro Nakano · Priya Vashishta · Jorge L. Landa

Published online: 23 March 2007

Abstract We have developed a parallel and distributed computing framework to solve an inverse problem, which involves massive data sets and is of great importance to petroleum industry. A Monte Carlo method, combined with proxies to avoid excessive data processing, is employed to identify reservoir simulation models that best match the oilfield production history. Subsequently, the selected models are used to forecast future productions with uncertainty estimates. The parallelization framework combines: (1) message passing for tightly coupled intra-simulation decomposition; and (2) scheduler/Grid remote procedure calls for model parameter sweeps. A preliminary numerical test has included 3,159 simulations on a 256-processor Intel

K.-I. Nomura (✉)
Collaboratory for Advanced Computing and Simulations (CACS), Center for Interactive Smart Oilfield Technologies (CiSoft), University of Southern California, Los Angeles, CA 90089-0242, USA
e-mail: knomura@usc.edu

R.K. Kalia
Department of Computer Science, University of Southern California, Los Angeles, CA 90089-0242, USA
e-mail: rkalia@usc.edu

A. Nakano
Department of Physics and Astronomy, University of Southern California, Los Angeles, CA 90089-0242, USA
e-mail: anakano@usc.edu

P. Vashishta
Department of Chemical Engineering and Materials Science, University of Southern California, Los Angeles, CA 90089-0242, USA
e-mail: priyav@usc.edu

J.L. Landa
ChevronTexaco, 6001 Bollinger Canyon, San Ramon, CA 94583, USA
e-mail: Jorge.Landa@chevron.com

Xeon cluster at the USC-CACS. The results provide uncertainty estimates of unprecedented precision.

Keywords Smart oilfield · History matching and forecast · Inverse problem · Monte Carlo method · Parallel and distributed computing · Massive data sets

1 Introduction

The Center for Interactive Smart Oilfield Technologies (CiSoft) was established by Chevron and the University of Southern California (USC) in December 2003. The research goal of the CiSoft is to reduce cost and increase efficiency of oilfield operations by allowing decision making through integration of static and dynamic data, and visualization of risk, uncertainty ranges and various financial metrics, in remote collaborative environments.

In a digital oilfield governed by an intelligent supervisory management system, continuous sensor data are obtained from individual wells. In addition, information from permanently installed geophysical recorders and surface operation facilities are analyzed in real time and serve as the basis for progressively more accurate oilfield characterization, management planning and control of the wells. Furthermore, these sensor data are augmented with synthetic data from numerical simulations of subsurface oil/gas reservoirs to guide optimal decisions. Real time interpretation of such a heterogeneous data system and instant consequential analysis of alternative will open the possibilities for intelligent asset protection, recovery optimization, human safety in the oilfield and environmental protection leading to substantial economic gains over conventional operation.

One critical objective of the CiSoft is to enable a paradigm shift in reservoir management, by integrating high performance computing into novel history matching methods to reduce uncertainty and allow real-time reservoir management. History matching is an inverse problem to calibrate reservoir simulation models to the observed production history, and it is a critical and required step in optimizing decisions that are linked to the subsurface (i.e., oil/gas reservoir). The uncertainty in the resolution of the subsurface model translates into the uncertainty in predictions, which are the key input to the decision process. The development of rapid, efficient and accurate computational methods, and of associated computer infrastructure, is necessary to facilitate multiple realizations in history matching, so that one can capture the uncertainty in parameters. Fast simulation techniques utilizing high performance computing are particularly needed with the expected availability of an avalanche of real-time data, from novel sensors, and of control variables, from novel actuators.

The history matching and the assessment of the associated uncertainty in predictions constitute a complex and difficult problem. Internal work at Chevron during the last five years has resulted in a mathematical framework and associated workflows, called *history matching and associated forecast (HMAF)* [1]. The HMAF includes a probabilistic approach to the solution of the inverse problem [2], uncertainty and heterogeneity in the data, the linking of multiple forward models using the concept of "Common Earth" model, and the development of complex "proxies" (or approximate response surfaces) as a substitute for compute-expensive forward modeling (such as full field reservoir simulation).

Prior applications of the HMAF have shown that the practicality of the methods is contingent on the availability and the innovative use of high performance computers and networks. History matching of real field cases involves Monte Carlo sampling of 10^4 simulations with different model parameter variants (or realizations) to select models that best explain the past production history. This will require several processor·years of computing, and performing history matching on thousands of processors will enable overnight forecast to guide daily decision-making.

This paper describes the design of a *parallel history matching and associated forecast (P-HMAF)* framework. In the next section, we describe the P-HMAF framework on a parallel computer and its extension to a Grid of geographically distributed parallel computers—*Grid-enabled history matching and associated forecast (G-HMAF)* framework. Results of preliminary numerical tests are given in Sect. 3, and Sect. 4 contains conclusions.

2 History matching and associated forecast (HMAF) framework

The HMAF framework involves Monte Carlo (MC) sampling of subsurface reservoir simulations with different model parameter variants. The parameter variants reflect the uncertainty in geological model definitions (e.g., permeability values and the transmissibility at the faults) and initial conditions (e.g., the positions of the initial gas/oil and oil/water contacts), among others. Petroleum engineers often abstract these uncertainties into a set of $N_p = 10$–20 parameters. This introduces an exponentially large, N_p-dimensional solution space for exhaustive enumeration of combinatorial parameter sets.

The HMAF pseudorandomly generates a large number, $N_{\mathrm{sim}} \sim 10^4$, of parameter sets to sample the N_p-dimensional model parameter space. Each parameter set constitutes a distinct realization of the simulation model, and accordingly the HMAF performs N_{sim} reservoir simulations. Note that each reservoir simulation involves numerical solution of multiphase (oil, gas, water) flows discretized on a large number of mesh points, and is highly compute intensive. The HMAF then selects a subset of N_{acc} ($\ll N_{\mathrm{sim}}$) simulation models that reproduce the past oil, gas and water production history within a predetermined error tolerance, see Fig. 1. The selected simulations are used to forecast future production, and a histogram of predicted values, in turn, is used to estimate the uncertainty in the forecast. Finally, the HMAF recursively refines the population of simulation models by pseudorandomly producing a new generation of parameter sets.

2.1 Parallel history matching and associated forecast (P-HMAF) framework

In the HMAF framework, each of the N_{sim} reservoir simulations numerically integrates partial differential equations that are discretized on a large number of mesh points, to determine the time evolution of flow patterns over years for oil, gas, and water under ground. Each simulation task is a tightly coupled parallel application, which is written in a single-program multiple-data (SPMD) paradigm, with message

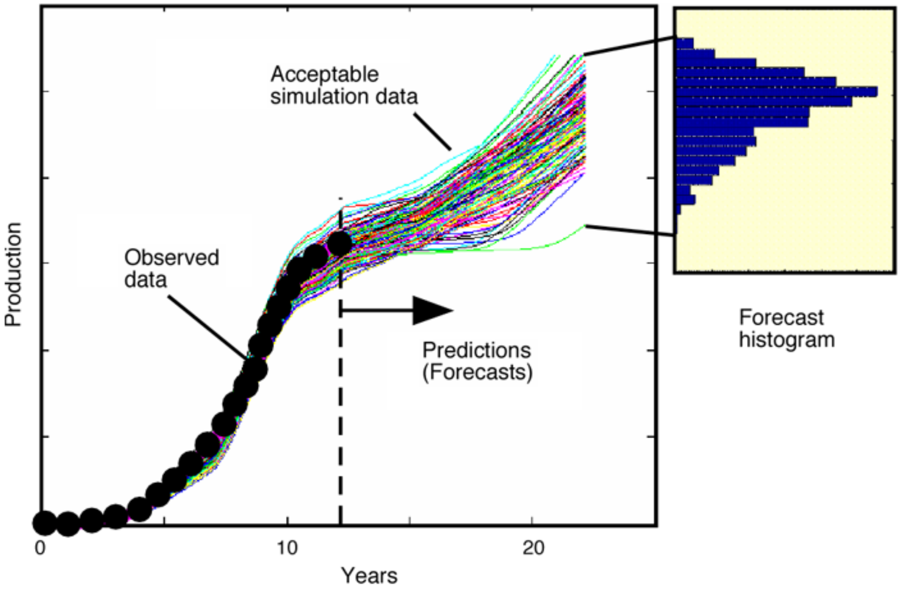

Fig. 1 Schematic of history matching and forecast. Acceptable simulations (*lines*) that match the observed production history (*circles*) are used to construct a histogram to forecast future production

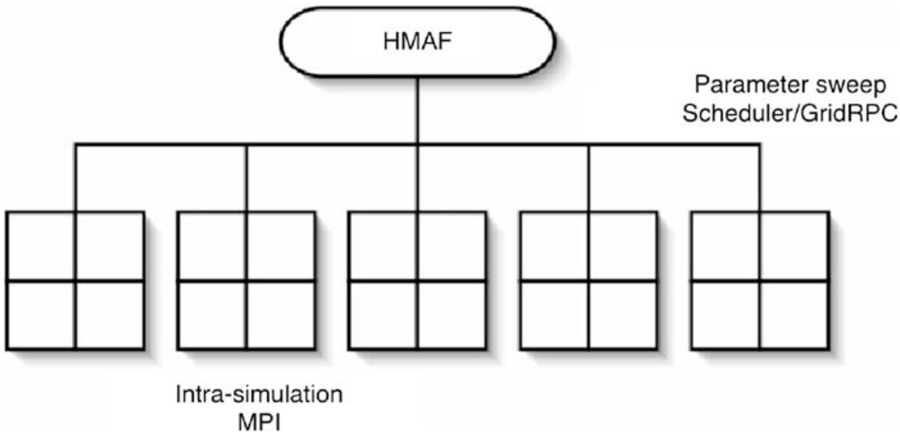

Fig. 2 Mixed parallelization framework based on intra-simulation (MPI) and parameter-sweep (scheduler/GridRPC) parallelisms

send and receive operations implemented with the message passing interface (MPI) standard, see Fig. 2.

In the HMAF, tens of thousands of independent reservoir simulations with different parameter realizations are performed, which collectively constitutes a large parameter sweep (or task farm) application. Each simulation on 1–32 processors (depending on the size of the oilfield) takes several hours of wall-clock time and produces 20 MB–1 GB of data.

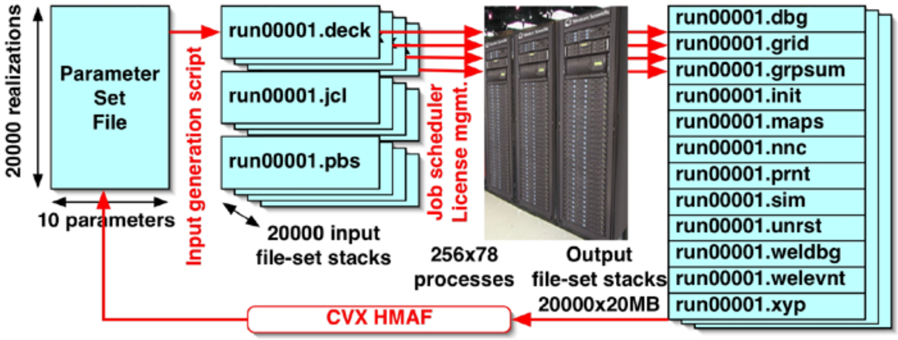

Fig. 3 Parallel history matching and associated forecast (P-HMAF) framework

We have nearly automated the workflow of parallel history matching and associated forecast (P-HMAF) framework. As a specific example, consider a case, in which model variants are summarized in terms of $N_p = 10$ unknown parameters, and an importance sampling MC technique generates $N_{\text{sim}} = 20{,}000$ simulations. In this example, each simulation is a single-processor job for ~ 3 CPU hours, and we use a 256-processor parallel computer to provide the requisite $3 \times 20{,}000 = 60{,}000$ CPU·hours of computation. The input parameter deck for this reservoir simulation includes 10 unknown parameters, among other information such as the definition of the geology. First the framework generates a file containing a batch of random realizations of the 10 parameter values. A script reads this file and generates a stack of input deck and job control files. A job scheduler then launches the batch processes on the cluster 256 at a time, see Fig. 3. Next, a new batch of realizations is generated and the process is repeated, until a total of 20,000 simulations are completed.

To avoid excessive handling of massive data and computation, the P-HMAF uses proxies to construct response surfaces (i.e., how the simulation results of interest change as a function of model parameters). The proxy smoothly interpolates the simulated response surface points, using both the function values and also when available their gradients of interest from the simulations [1]. The recursive sampling greatly reduces the uncertainty.

2.2 Grid-enabled history matching and associated forecast (G-HMAF) framework

We have also initiated an effort to implement the HMAF approach on a Grid of globally distributed parallel computers. We have developed a hybrid Grid metacomputing framework that combines: (1) Grid remote procedure call (GridRPC), which is best suited for parameter-sweep applications such as the MC tasks within the HMAF; and (2) message passing interface (MPI), which is required for tightly-coupled parallel tasks (e.g., reservoir simulations) in the HMAF. The hybrid Grid computing approach combines the flexibility and fault tolerance capabilities of GridRPC and the high performance of MPI.

We have successfully tested the hybrid GridRPC/MPI Gridification framework at the IEEE/ACM Supercomputing 2004 (SC04) Conference [3], to perform a metacomputing involving 1,792 processors on the US TeraGrid [4], which connects several Teraflop platforms via 40 Gbit/s optical fibers, and its Japanese counterpart—the

11 Tflop Super Cluster at the National Institute of Advanced Industrial Science and Technology (AIST).

More recently, we have performed metacomputing involving 153,600 processor-hours on a US-Japan Grid at the University of Southern California (USC), Pittsburgh Supercomputing Center (PSC), National Center for Supercomputing Applications (NCSA), AIST, University of Tokyo, and Tokyo Institute of Technology. Though it is difficult to exclusively access 1,000 processors for 10 consecutive days at a single supercomputer center, we have achieved it on a Grid involving 6 supercomputer centers. This opens up a possibility of overnight history matching on a Grid, with much accurate forecast than is possible today.

3 Numerical results

We have performed a proof-of-concept demonstration of the P-HMAF approach on a parallel computer. The test case involves 30-year history data from a real oilfield [5]. The P-HMAF test involves 3,159 reservoir simulation models performed on the 256-processor Linux cluster at the Collaboratory for Advanced Computing and Simulations (CACS) at USC. (The 3,159 runs have been performed in 7 batches, where the batch size depends on the number of parameters, the number of available processors, and the wall-clock time to run a single model.) Each simulation on the 2.8 GHz Intel Xeon processor takes $\sim$ 3 hours and produces 20 MB of output data (see Fig. 3), with aggregated data size 62 GB. In comparison, the best history matching and forecast to

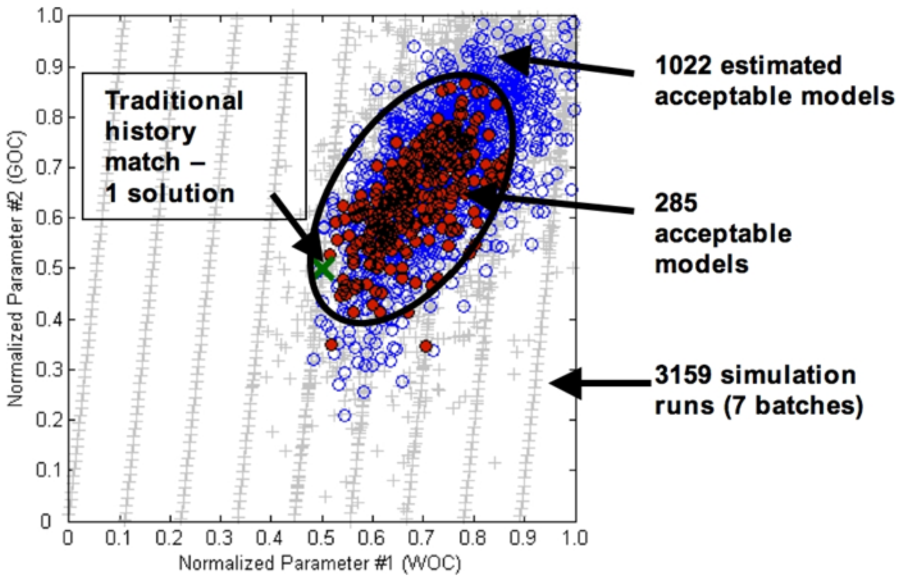

Fig. 4 Combination of two history matching parameters—normalized water-oil-contact (WOC) and gas-oil-contact (GOC) positions—for (1) all 3,159 simulation models that have been run (+ symbols), (2) 285 acceptable models that reproduce the history data within a threshold (solid circles), and (3) 1,022 estimated acceptable models. Traditional history match, in contrast, produces only one acceptable model (shown by a cross)

Table 1 Comparison of the estimated CPU times of different history match methods

Number of CPUs	Method	Number of model simulations	Wall-clock time (days)	Number of acceptable solutions
1	Gauss-Newton	7	8.8	1
1	Probabilistic history match	3159	395	n/a
256	Parallel probabilistic history match	3159	1.5	285

date has involved 600 simulation models [5]. Our parallel test increases the sampling size by orders-of-magnitude to significantly narrow the range of forecast values [6].

Probabilistic descriptions of the model parameters after incorporating the production data were estimated using proxies. Figure 4 shows multiple solutions to the history match problem projected onto the two-parameter space, where the two normalized parameters represent the position of the initial water oil contact (WOC) and that of the initial gas oil contact (GOC). In Fig. 4, the plus symbols denote all 3,159 model simulations that were run, out of which 285 models (solid circles) produced the history data within a prescribed threshold. These acceptable models were augmented by means of proxies to generate 1,022 estimated acceptable models (open circles). The large number of acceptable models results in reduced uncertainty in production forecast, which is in contrast to a traditional history match approach that produces only one acceptable model (shown as a cross).

The wall-clock time to solve the history match problem with the PHMAF approach is compared with those of other approaches in Table 1. As a comparison, we have performed a traditional history match using a gradient-based approach based on the Gauss-Newton method. Here, the gradients are estimated numerically, and with 7 model simulations, it took 8.8 days to obtain a single acceptable solution. The probabilistic history match approach in this paper produces a number of acceptable solutions, resulting in greatly reduced uncertainty in forecast. With 3 hours wall-clock time per simulation, we estimated that HMAF with 3,159 model simulations on single processors would take 395 days. On 256 processors, we have achieved the same task in 1.5 days.

Figure 5 compares the uncertainty in the predicted production value (i.e., standard deviation of the sample mean normalized by the mean) as a function of the computational cost (i.e., total number of simulation runs). The circles denote numerical experimental results for 600, 1,669 and 3,159 simulation runs. The results show that the forecast uncertainty decreases rapidly as a function of the computational cost. As a comparison, an uncertainty estimate according to a standard sample mean MC method is plotted as a dashed line. The superior uncertainty reduction in the P-HMAF results from its recursive importance sampling.

4 Conclusions

We have parallelized the history matching and associated forecast (P-HMAF) framework, which incorporates a probabilistic approach to the inverse problem solution,

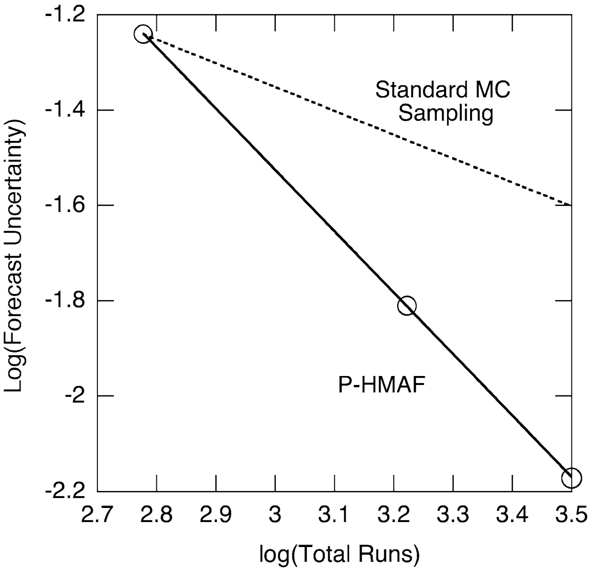

Fig. 5 Forecast uncertainty (normalized by the sample mean) as a function of the computational cost of P-HMAF (*circles*), compared with that with a standard MC sampling

data uncertainty, data heterogeneity, linking of multiple forward models using the concept of "Common Earth" model, and "proxies" as a substitute for compute-intensive forward modeling. We have also proposed to apply our hybrid Grid computing framework to Grid-enable the HMAF approach.

Currently, we are performing a 20,000-model P-HMAF test on the 256-processor Linux cluster at the USC-CACS, and planning: (1) an overnight history matching and forecast involving a larger field case on 1,024 processors of the Linux cluster at USC's high performance computing facility; and (2) a metacomputerized HMAF (G-HMAF) approach on a Grid of globally distributed parallel computers.

Though the data size (62 GB) of the preliminary test case is rather small, we anticipate a 2–3 orders-of-magnitude increase in data size for the larger field case to be studied in 2005–2006. In addition to providing future production forecast with uncertainty estimates for decision support, CiSoft research activities involve data mining from these massive simulation data sets to uncover hidden patterns and correlations to guide intelligent oilfield management, which will pose significant challenges for I/O systems.

Acknowledgements This research was partly funded by CiSoft (Center for Interactive Smart Oilfield Technologies), a Center of Research Excellence and Academic Training and a joint venture between the University of Southern California (USC) and Chevron. Computation was performed on the 266-processor Intel Xeon-based Linux cluster at the Collaboratory for Advanced Computing and Simulations (CACS) at USC. We thank Yoshio Tanaka, Hiroshi Takemiya, and Satoshi Sekiguchi at the National Institute of Advanced Industrial Science and Technology (AIST) in Japan for useful discussions on Grid computing. The authors would like to thank the management of Chevron and its partners for their permission to present this paper.

References

1. Landa JL, Guyaguler B (2003) A methodology for history matching and the assessment of uncertainties associated with flow prediction, SPE 84465. Society of Petroleum Engineers, Richardson

2. Mosegaard K, Tarantola A (2002) Probabilistic approach to inverse problems. In: International handbook of earthquake and engineering seismology, Part A. Academic Press, p 237
3. Takemiya H, Tanaka Y, Sekiguchi S, Ogata S, Kalia RK, Nakano A, Vashishta P (2006) Sustainable adaptive Grid supercomputing: multiscale simulation of semiconductor processing across the Pacific. In: Proceedings of SC06. IEEE/ACM, Tampa
4. We have used the AlphaServer at the Pittsburgh Supercomputing Center
5. Castellini A, Landa JL, Kikani J (2004) Practical methods for uncertainty assessment of flow predictions for reservoirs with significant history—a field case study. In: Proc. 9th European conf. on the mathematics of oil recovery, Cannes, France, 2004, p A-33
6. Landa JL, Kalia RK, Nakano A, Nomura K, Vashishta P (2005) History match and associated forecast uncertainty analysis—practical approaches using cluster computing. In: Proceedings of the international petroleum technology conference, IPTC 10751. Society of Petroleum Engineers, Richardson