XIX International Conference on Water Resources
CMWR 2012
University of Illinois at Urbana-Champaign
June 17-22, 2012

MODEL ANALYSIS AND DECISION SUPPORT (MADS) FOR COMPLEX PHYSICS MODELS

Velimir V. Vesselinov* and Dylan R. Harp*

* Los Alamos National Laboratory (LANL)
Computational Earth Science Group, Earth and Environmental Sciences Division
EES-16, MS T003, Los Alamos, NM 87545, USA
e-mail: vvv@lanl.gov, dharp@lanl.gov, web page: http://mads.lanl.gov

Key words: model predictions, global optimization, global uncertainty analysis, decision support

Summary. MADS (Model Analysis & Decision Support; http://mads.lanl.gov) is an open-source object-oriented code that provides an integrated computational framework for a wide range of model-based analyses, and supports scientifically defensible decision making. The code targets the implementation of computationally efficient and robust algorithms that can perform model-based analyses requiring a relatively small number of model evaluations, and can execute simulations in parallel on multiprocessor clusters. Two new algorithms have been recently developed and implemented in MADS for global uncertainty analysis (*ABAGUS*: Agent-Based Analysis of Global Uncertainty and Sensitivity) and global optimization (*Squads*). These algorithms have been applied in a series of investigations related to groundwater environmental management at the Los Alamos National Laboratory Site including the Sandia Canyon chromium investigation.

1 INTRODUCTION

Complex physics models are frequently applied to perform various types of model-based analyses such as parameter estimation, uncertainty quantification, risk assessment and decision support. However, these models are typically computationally intensive requiring relatively long execution times (usually from several minutes to several hours). This limits our ability to perform detailed model-based analyses of model estimates and predictions because these analyses typically require a substantial number of model runs (usually from about 100 to more than 1,000,000 depending on the analysis approach). Therefore, it is important to develop computationally efficient and robust algorithms that can perform model-based analyses requiring a relatively small number of model evaluations. The code MADS[1] (Model Analysis & Decision Support) is designed to provide an integrated framework for a wide range of computationally efficient and robust model-based analyses to supports scientifically defensible decision making. Here we will discuss recent developments in two algorithms implemented in the MADS toolkit: (1) Agent-Based Analysis of Global Uncertainty and Sensitivity (*ABAGUS*) and (2) *Squads*.

2 MADS

MADS (Model Analyses & Decision Support) is an open-source, object-oriented code that provides an integrated computational framework for a wide range of model-based analyses, and supports scientifically-defensible decision making. The code can be executed under different computational modes, which include (1) sensitivity analysis, (2) model calibration (parameter estimation), (3) uncertainty quantification, (4) model selection, (5) model averaging, and (6) decision analysis. MADS can be externally coupled with any existing model simulator through integrated modules that read/write input and output files using a set of template and instruction files. MADS can also work with existing control, template and instruction files developed for the code PEST[2]. The code is internally coupled with a series of built-in analytical simulators (currently the analytical solutions are for contaminant transport in aquifers[3] only). In addition, MADS can be used as a library (toolbox) for internal coupling with any existing object-oriented simulator using object-oriented programming.

MADS provides (1) efficient parallelization, (2) runtime control, restart, and preemptive termination, (3) several Latin-Hypercube sampling techniques (including Improved Distributed Sampling[4]), (4) several derivative-based local Levenberg-Marquardt optimization methods (including geodesic acceleration[5]), (5) several single- and multi-objective global optimization methods (including Particle Swarm Optimization[6], PSO[7], TRIBES[8]), and (6) local and global sensitivity and uncertainty analyses, and (7) model analysis and decision support techniques such as GLUE[9] and information gap[10]. MADS is characterized by several unique features:

- Adaptive execution with minimum input from the user; for most analyses, all the parameters controlling the performance of the algorithms for model-based analyses are estimated internally in the code; if needed, the user has the flexibility to specify a wide range of options.
- The same problem input file is sufficient and can be applied to perform all the possible model analyses supported in the code.
- Most of the model analyses (e.g. calibration, uncertainty quantification, and sensitivity analysis, etc.) can be performed using a discretized parameter space; this can substantially reduce the computational effort to perform model analyses of complex physics models with long execution time.
- By default, all the model parameters are internally normalized and transformed in a manner that substantially improves the derivative-based optimization algorithms.
- Highly-parameterized inversion, where the number of model parameters is substantially greater than the number of model constraints (calibration targets or model observations); a similar approach is called SVD assist in the code PEST[2].
- 'Acceptable' calibration ranges for each optimization target can be implemented; in this way, the optimization can be directed to search for models producing estimates only within user-defined acceptable calibration ranges. For example, once the model estimates are within acceptable calibration ranges the optimization is terminated.
- Allows the use of an acceptable calibration range for the objective function; in this way, acceptable model solutions can be identified as those producing objective functions below a predefined cutoff value; once the objective function is decreased below the cutoff value, the

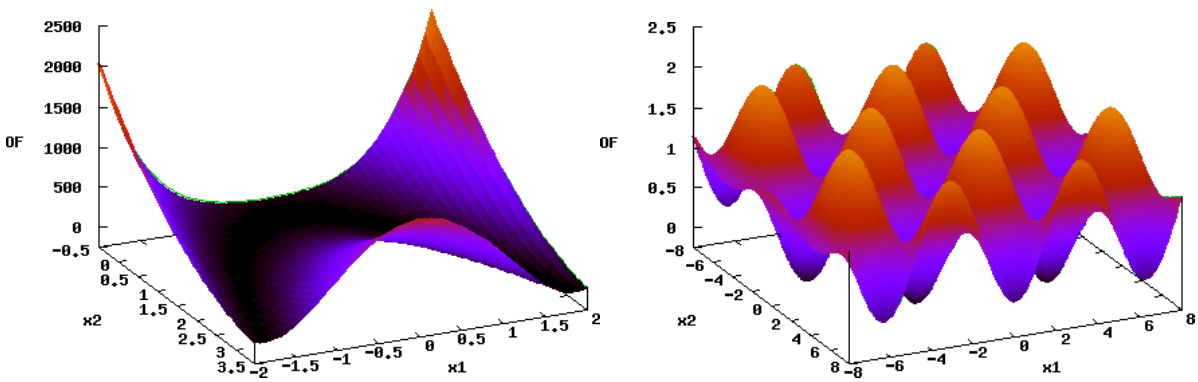

Figure 1: Two-dimensional Rosenbrock (left) and Griewank (right) test functions with global minima at (1,1) and (0,0), respectively.

optimization is terminated.

- Implements a series of alternative objective functions (OF).
- Provides the option to perform a series of optimizations with random initial guesses for optimization parameters; the code can also automatically retry the optimization process using a series of random initial guesses until an acceptable calibration is achieved.
- Automatically detects and utilizes the available parallel resources; automatically analyzes the runtime performance of the available parallel hosts (processors); hosts not capable of performing the requested parallel jobs are ignored; automatically tracks the multiple model files during parallel execution automatically; for the user, there is no difference between serial (using single processor) and parallel mode of execution.
- Performs automatic bookkeeping of all the model results for efficient restart and rerun of MADS jobs (e.g. if the previous job was not completed) and additional posterior analyses.
- Allows the user to perform different types of analyses based on model results stored during previous MADS runs; for example, model runs obtained during model calibration can be utilized in posterior Monte Carlo analyses.
- Object-oriented design of MADS allows for relatively easy integration with other object-oriented optimization or sampling techniques.

The code is written in C/C++ and tested on various Unix platforms (Linux, Mac OS X, Cygwin MS Windows). MADS source code and other files needed to execute the synthetic problems presented here are available at http://mads.lanl.gov.

3 ABAGUS

ABAGUS[11] is novel approach to global uncertainty and sensitivity analyses of modeling results utilizing concepts from agent-based modeling implemented in MADS. The explored

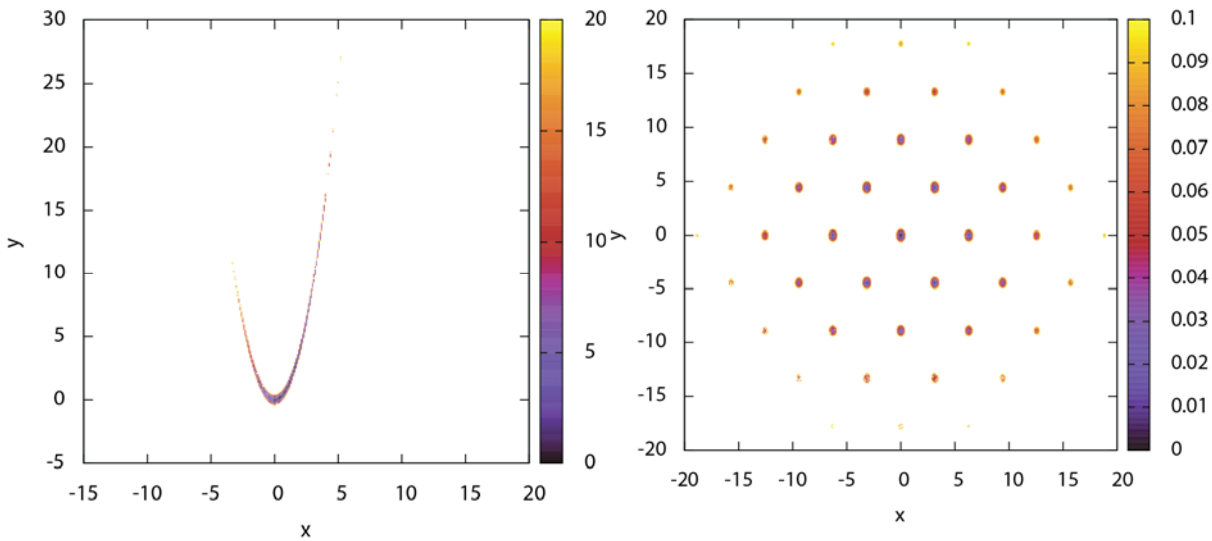

Figure 2: ABAGUS uncertainty estimates for Rosenbrock (left) and Griewank (right) functions.

model parameter space is discretized and sampled by a particle swarm where the particle locations represent unique model parameter sets. Particle locations are optimized based on a model performance metric using a standard particle swarm optimization[7] (PSO) algorithm. Model evaluations are stored in KD-Trees[12], eliminating the need for model reruns for already evaluated parameter sets. *ABAGUS* sculpts the response surface to discourage reinvestigation of "collected" regions of the parameter space and encourage global exploration. *ABAGUS* utilizes a hierarchical discretization scheme that provides automatic refinement of the parameter-space exploration based on run-time performance.

The performance of *ABAGUS* is tested on two-dimensional Rosenbrock and Griewank test functions (Fig. 1). The Griewank function has numerous local areas of attraction, but a single global minimum of zero at (0,0). In the two-dimensional case, the function has the shape of an 'egg carton' that is depressed in the center. The obtained results are presented in Fig. 2. In this case, ABAGUS effectively identifies the portions of the parameter space that are defined by objective function (OF) below 20 and 0.1 for Rosenbrock and Griewank functions, respectively.

In addition, *ABAGUS*' performance is compared with standard Monte Carlo sampling. In this case, the test problem is based on a two-dimensional parabolic function with a single well-defined global minimum at (0,0) (Fig. 3). Monte Carlo analysis is performed using Improved Distributed Latin Hypercube Sampling[4] (IDLHS) which is implemented in MADS as well. Both techniques are tested to identify the fraction of the parameter space that is defined by an objective function less than 160; the acceptable portion of the parameter space is a circular domain centered around the global minimum. The area of this circle is about 5% of the explored parameter space. In this case, the problem is similar to identifying model predictions with a

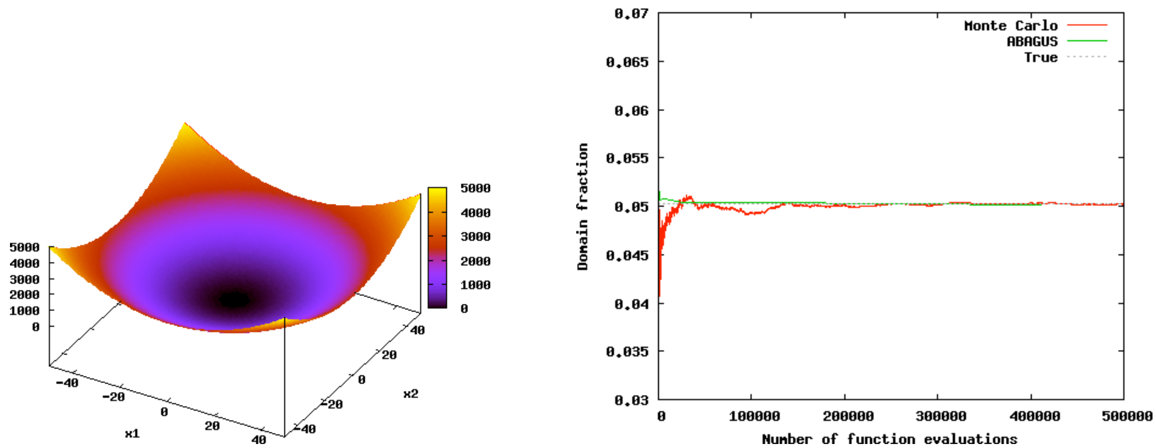

Figure 3: A 2D parabolic test problem (left) with global minimum at (0,0) applied to estimate convergence as a function of the number of model evaluations for the Monte Carlo and ABAGUS analyses.

probability of occurrence 5%. The estimated fractions of the parameter space as a function of the number of model evaluations are also presented in Fig 3. The results demonstrate that ABAGUS converges faster than the Monte Carlo analysis; ABAGUS needs about 20,000 evaluations to converge while the Monte Carlo method requires more than 140,000 evaluations.

4 SQUADS

Squads[13] is a newly developed adaptive hybrid optimization algorithm, providing a global optimization strategy with local optimization speedup. *Squads* combines an adaptive particle swarm optimization (global) strategy and a Levenberg-Marquardt (local) optimization strategy. In contrast with other existing hybrid optimization strategies, the global and local algorithms are adaptively coupled in *Squads*; the global and local optimizations steps are iteratively performed based on adaptive rules that depend on the optimization progress. The local optimization speedup is performed on selected particles from the swarm. The coupling of the global and local optimization techniques within *Squads* utilizes transformation of the parameter space within the local optimization speedup to enhance the local optimization performance near parameter boundaries. This allows the PSO strategy to be performed in bounded parameter space while Levenberg-Marquardt strategy to be performed in unbounded parameter space. *Squads* performs these transformation using trigonometric functions[13], which allows for seamless transitions between global and local optimization strategies.

The robustness and efficiency of *Squads* is compared to Levenberg-Marquardt[14], particle swarm optimization (PSO)[7], TRIBES[8] (an adaptive particle swarm optimization algorithm), and hPSO[15] (an alternative hybrid optimization algorithm). The comparisons are performed on two polynomial test functions the Rosenbrock and Griewank functions (Fig. 1). The analyses include two-, five and ten-dimensional versions of these functions.

The response function defined by the Rosenbrock function is comprised of a large valley with an ill-defined, shallow global minimum; the global minimum is difficult to identify by either local and global optimization strategies. In the two dimensional case, the test function is unimodal with a single global minimum; in cases with greater than two dimensions, there are multiple suboptimal local minima.

The multidimensional Griewank function is important for testing of hybrid optimization strategies because it becomes more difficult to minimize for global strategies as its dimensionality increases. However, although counterintuitive, the Griewank function becomes easier to minimize for local strategies as the dimensionality increases. Therefore, with the increase in dimensionality, it is expected that LM performance will improve while PSO, TRIBES and hPSO performance will decrease. For different parameter-space dimensionality, the performance of hybrid strategies will depend on how efficiently they adaptively balance between the local and global strategies. At low dimensionality (D=2), the hybrid strategies should benefit from the global strategy; at high dimensionality, the hybrid strategies should benefit from the local strategy.

The robustness of each algorithm is quantified by the fraction of optimization runs that identify the global minimum of the test function. The efficiency of the algorithms is quantified by statistical representations of the number of function evaluations necessary to reach the global minimum of the test functions on successful runs (Fig. 4). The performance of *Squads*, considering both optimization robustness and computational efficiency, is superior to the other algorithms.

5 CONCLUSIONS

ABAGUS provides a discretized global uncertainty analysis approach filling the gap between local and sampling-based global approaches. *ABAGUS* is an attractive alternative for complex problems where it is recognized that a local analysis is inappropriate, but for which a rigorous sampling-based global analysis is infeasible due to computational constraints. The results obtained from a single *ABAGUS* run can be applied to estimate simultaneously uncertainties, sensitivities, and correlations in and between model parameters and predictions (uncertainty analysis of model prediction is sometimes called predictive analysis). *ABAGUS* results can be applied for decision support based on a predefined performance metric or compliance criteria. *Squads* is a new adaptive global hybrid optimization strategy for computationally intensive inverse problems involving models representing the behavior of complex systems. The algorithm shows superior performance compared to other commonly used optimization algorithms. The newly developed algorithms are implemented in the code MADS.

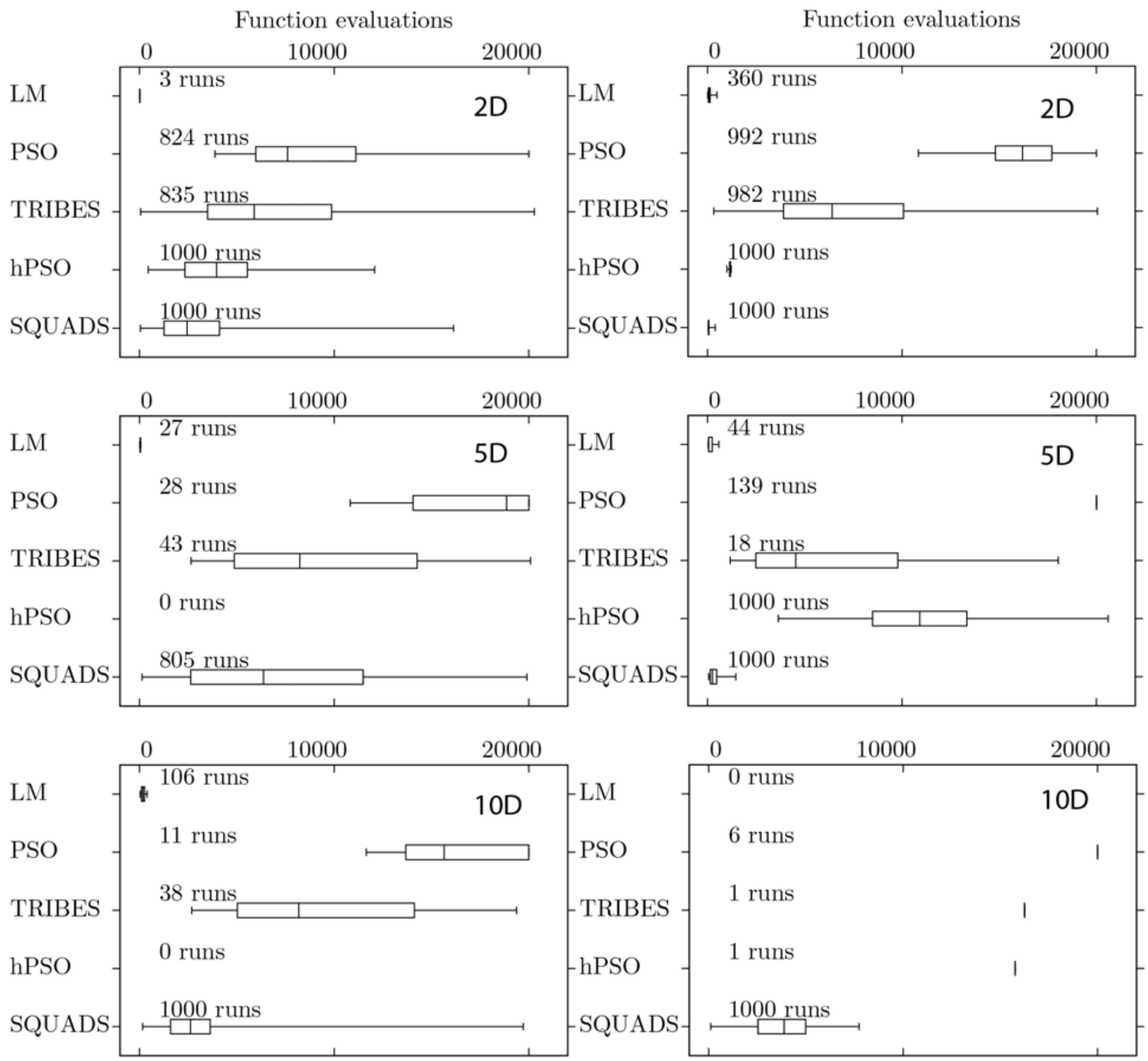

Figure 4: Boxplots of the number of function evaluations to reach the global minimum for the 2D, 5D, and 10D Rosenbrock (left) and Griewank (right) functions. The boxes represent the 25th to 75th percentile ranges, the bars inside of the boxes represent the median values, and the whiskers represent the min and max values. The number of successful runs out of 1000 for each strategy is stated above the boxes.

REFERENCES

[1] V.V. Vesselinov, D. Harp, "MADS: Model analysis and decision support in C", *http://mads.lanl.gov*, (2012).

[2] J. Doherty, "Ground Water Model Calibration Using Pilot Points and Regularization", *Ground Water*, *10.1111/j.1745-6584.2003.tb02580.x*, (2003).

[3] H. Wang and H. Wu, "Analytical Solutions of Three-Dimensional Contaminant Transport in Uniform Flow Field in Porous Media: A Library", *Front. Environ. Sci. Engin. China*, *10.1007/s11783-008-0067-z*, (2009).

[4] B. Beachkofski, and R Grandhi, *Improved Distributed Hypercube Sampling*, American Institute of Aeronautics and Astronautics Paper 2002-1274 (2002).

[5] M.K. Transtrum, B.B. Machta, and J.P. Sethna, "Geometry of nonlinear least squares with applications to sloppy models and optimization", *Phys. Rev. E* ***83***, *036701*, (2010).

[6] M. Clerc, *Particle Swarm Optimization*, ISTE, London, (2006).

[7] Paricle Swarm Central, *Standard PSO algorithm*, http://www.particleswarm.info/Standard PSO 2006.c, (2006).

[8] Y. Cooren, M. Clerc, and P. Siarry, "*Performance evaluation of TRIBES, an adaptive particle swarm optimization algorithm"*, Swarm Intelligence **3**,149–178 (2006).

[9] K. Beven and J. Freer, "*Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology"*, Journal of Hydrology, 10.1016/S0022-1694(01)00421-8 (2001).

[10] D. Harp, V.V. Vesselinov, "Contaminant remediation decision analysis using information gap theory", *SERRA*, *10.1007/s00477-012-0573-1*, (2012).

[11] D. Harp, V.V. Vesselinov, "An agent-based approach to global uncertainty and sensitivity analysis", *Computers & Geosciences*, *10.1016/j.cageo.2011.06.025*, (2011).

[12] J. Tsiombikas, *kdtree: A simple C library for working with KD-Trees*, http://code.google.com/p/kdtree/, (2009).

[13] V.V. Vesselinov, D. Harp, "Adaptive hybrid optimization strategy for calibration and parameter estimation of physical models", *Computers & Geosciences, submitted*, (2012).

[14] M.I.A. Lourakis, *LevMar: Levenberg-Marquardt nonlinear least squares algorithms in C/C++*, http://www.ics.forth.gr/~lourakis/levmar/, (2004).

[15] A. Leontitsis, *Hybrid Particle Swarm Optimization in Matlab*, http://www.mathworks.com/matlabcentral/fileexchange/6497-hybrid-particle-swarm-optimization, (2004).