Auto-Calibration of Hydrological Models Using High Performance Computing

Vimal Sharma[a]**, David Swayne**[a]**, David Lam**[a]**, William Schertzer**[a]

[a]Computing Research Laboratory for Environment, University of Guelph, Guelph, ON, Canada.
vsharma@uoguelph.ca

Abstract: Hydrological models have been increasing in complexity over the years. These models rely on their calibration to simulate real world conditions as close as possible. Calibration is a tedious and time-consuming process. An auto-calibration algorithm (SCE-UA) developed by Duan et al. [1992], has been successfully used in hydrological modeling area. This is a serial algorithm and as complexity of the models to be calibrated increases the computational cost, also significantly increases. In this study, a parallel version of the algorithm developed is used for testing of two simple hydrological models. The results show that parallel version of the algorithm can be successfully used to calibrate complex hydrological models.

1. INTRODUCTION

Manual calibration is the most commonly used approach for hydrological modeling. It is tedious, time consuming and the success of it depends on the expertise of the modeler with prior knowledge of the watershed being modeled (Eckhardt and Arnold [2001]). Automatic calibration involves the use of a search algorithm to determine best-fit parameters. It is highly desirable as it is faster, less subjective and due to extensive search of parameter possibilities, can give better results as compared to the results using manual calibration.

The two most important stages of calibration are parameter specification and parameter estimation. In the first stage, parameter specification, the parameters that need to be adjusted are selected, and in the parameter estimation stage, the optimal or near optimal values for the parameters are found (Sorooshian and Gupta [1995]).

There has been work done on the auto calibration methodology. Duan et al. [1992] developed Shuffled Complex Evolution (SCE-UA) algorithm for auto calibration. This algorithm is based on genetic algorithm technique. It combines the best features of multiple complex shuffling and competitive evolution based on simplex search method (Muttil and Liong [2004])
This auto calibration algorithm has been applied to one of the commonly used hydrological model, SWAT (Soil and Water Assessment Tool). Major components of the model include weather, hydrology, soil temperature, plant growth nutrients, pesticides, and land management (Arnold et al. [1998]). The details of the model are presented by Neitsch et al. [2002]. This model has been validated for several watersheds (Bosch et al. [2004]; Saleh et al. [2000]; Saleh and Du [2004]; Spruill et al. [2000]; Peterson and Hamlett [1998]).

Muleta and Nicklow [2005], presented a methodology for sensitivity and uncertainty analysis coupled with automatic calibration of SWAT for a watershed in southern Illinois. In that study, the parameter specification stage consisted of screening, parameterization and parameter sensitivity analysis. The parameter sensitivity analysis was conducted using a Monte Carlo based approach using Latin hypercube sampling along with a stepwise regression analysis of rank transformed input-output data pairs. A genetic algorithm was used for parameter estimation. The uncertainty analysis was done using Generalized Likelihood Uncertainty (GLUE) method. The results of the simulation indicated successful use of the automatic approach.

Eckhardt and Arnold [2001], used Shuffled Complex Evolution (SCE-UA) algorithm for parameter estimation using a modified version of SWAT (SWAT-G, modified for conditions on a central German watershed). The results showed a successful calibration of the model using this technique, but it took 6 days to optimize the parameters using an IBM RS/6000 workstation. A similar approach for calibration was used by Di Luzio and Arnold [2004] with SWAT. Muttil and Liong

[2004] showed that using a scheme to systematically, instead of randomly, generate the initial population as in the case of the SCE-UA algorithm increased the robustness of the algorithm.

In view of the literature cited above, in this study a parallel version of the existing serial algorithm was developed using high performance computing to significantly decrease the computation time. This parallel version of the algorithm was applied to two simple hydrological models namely, DYRESM-NWRI, and GAMES.

The current serial algorithm and parallel version developed is presented in section 2 followed by a brief description of hydrological models in section 3. Testing of the parallel model with simple functions is described in section 4. The results are presented and discussed in section 5.

2. SHUFFLED COMPLEX EVOLUTION ALGORITHM

The Shuffled Complex Evolution (SCE) algorithm was developed by Duan et al. [1992], using the best features of multiple complex shuffling and competitive evolution based on the simplex search method. The algorithm involves the following steps:

1. Sample points are generated from the feasible space (using upper and lower bounds of the parameters). The criterion function values are computed using these sampled points.
2. The sampled points are sorted and ranked in ascending order based on the criterion function values. This will result in the smallest criterion function value generating parameters at the top of the sampled parameter list.
3. The sampled points are partitioned into complexes with predefined size of the complex population.
4. Each complex is evolved independently a predefined number of times. The evolution of the complexes takes place using three types of evolution steps, namely: reflection, contraction and mutation.
    In the reflection step, the worst point in the sub-complex is reflected through the centroid of the other points. Since the reflected point has a lower criterion value than the worst point, the worst point is discarded and replaced by the new point. Thus an evolution step is completed. If the reflection step does not improve the criterion value, a contraction step in evolution is tried. In the contraction step, the new point lies

halfway between the worst point and the centroid of the other points.
    If after the reflection step, the criterion value is outside the feasible parameter space, the mutation step is initiated. It is also used when both reflection and contraction steps fail to improve the criterion value. In a mutation step, a point is randomly selected in the feasible parameter space to replace the worst point of the sub-complex.
5. The evolved complexes from the previous step are combined into a single sample population. The sample population is sorted in order of increasing criterion value. Steps 3 to 5 are repeated until conditions as defined in step 6 are met.
6. The loop is stopped if the number of evolutionary steps has exceeded a predefined value or the criterion value has not improved by a predefined percentage in a predefined number of steps.

Further details of the algorithm can be obtained in Duan et al. [1992].

In the parallel version of the SCE-UA algorithm developed involves the following steps (Sharma et al. [2006]):

1-2 Same as serial algorithm described above.
3. The root node partitions the complexes and distributes the complexes on to multiple processors
4. Complexes evolve on multiple processors instead of a single processor.
5. The evolved complexes from multiple processors are sent back to the root node to be combined into a single population.
6. The root node checks for the loop criteria and continues or stops accordingly (Figure 1).

3. HYDROLOGICAL MODELS

In this section, a brief description of the models used for testing is presented.

3.1 DYRESM-NWRI

Dynamic Reservoir Simulation Model (DYRESM) is a one-dimensional hydrodynamic model for predicting the vertical distribution of temperature, salinity and density in lakes (Imberger and Patterson [1980]; Antenucci and Imerito [2002]). It has been developed by Centre for Water Research (CWR), Australia. The lake is modeled as a series of horizontal layers of uniform property by variable
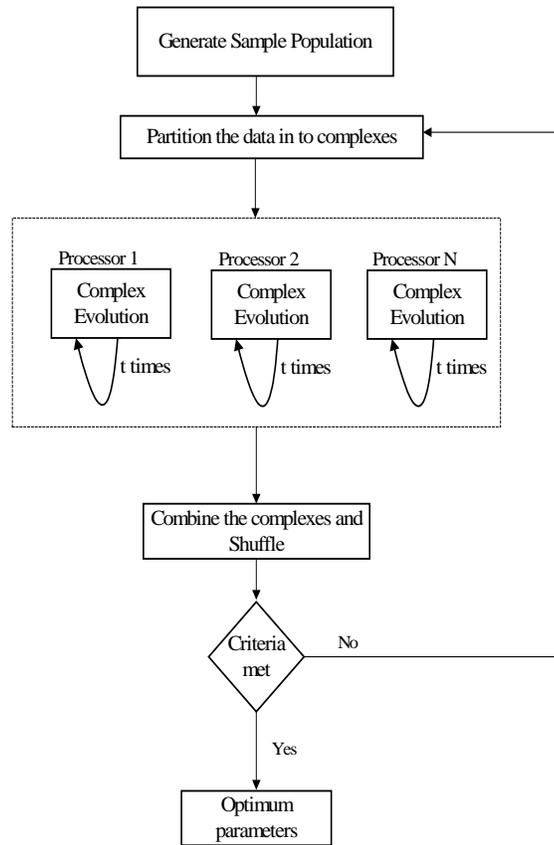
Figure 1. Simplified representation of the Parallel SCE-UA algorithm

thickness. The model is based on a Lagrangian layer scheme meaning that the series of horizontal layers are adjusted to stay within user-defined limits instead of fixed grid approach.

The primary driving mechanisms for DYRESM are the surface heat, mass and momentum exchanges. These processes are responsible for majority of energy requirements for heating, mixing and stratifying the lake.

In this study, a version of DYRESM modified at National Water Research Institute (NWRI), Burlington, was used. In this version, a single horizontal layer models the lake instead of multiple horizontal layers. This modification to the model enables it to be used for small lakes, which do not have detailed vertical profile data.

3.2 The GAMES Model

The Guelph model for evaluating effects of Agricultural Management systems on Erosion and Sedimentation (GAMES), based on Universal Soil Loss Equation (Wischmeier and Smith [1978]), was developed for watershed management (Rudra et al. [1986]). It predicts soil loss by erosion and the delivery of suspended solids from the fields to the streams. GAMES demonstrates areas within a watershed that are critical sediment sources and also provides a method to evaluate various soil conservation practices (Dickinson et al. [1987]; Dickinson et al. [1990]). The watershed used for analysis with GAMES must be discretized into field-sized elements with homogeneous characteristics of land use, soil type, and slope class. The model can be used for seasonal or annual assessments, depending on the selection of input parameter values. The sediment delivered from each cell to the watershed's stream channels is calculated from a delivery ratio for each cell based on the field cell's characteristics. The delivery ratio calculations require parameter 'α', which needs to be calibrated.
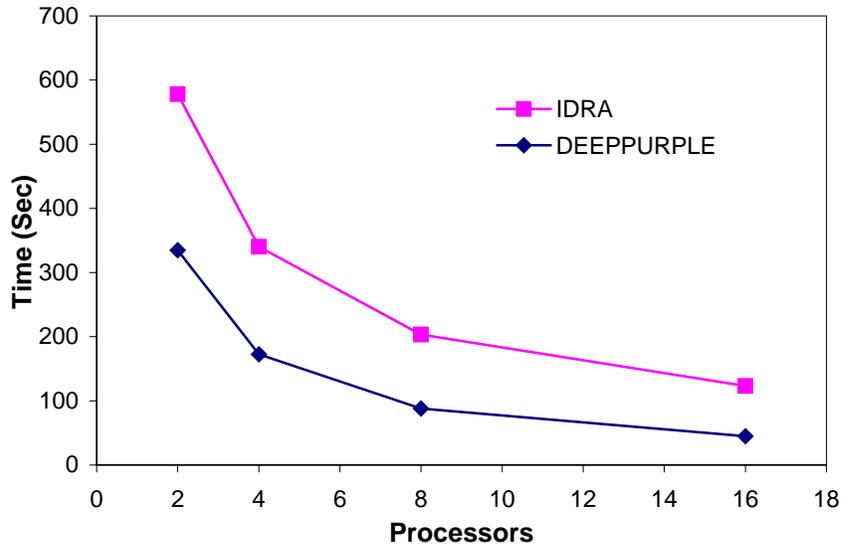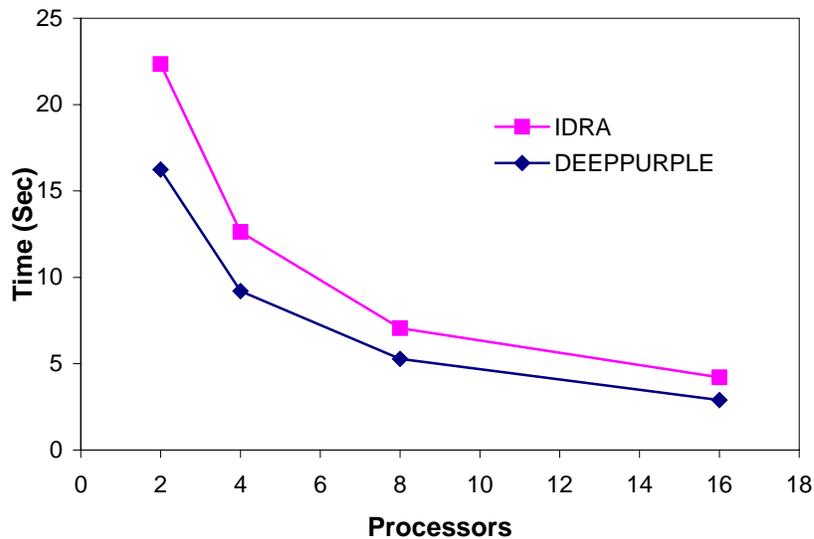
Figure 2.Timings obtained for DYRESM

Figure 3.Timings obtained for GAMES

4. TESTING OF THE PARALLEL ALGORITHM

The parallel SCE-UA algorithm developed was implemented and tested on the high performance computing (HPC) Idra and Deeppurple cluster of SHARCNET (The Shared Hierarchical Academic Research Computing Network). The clusters Idra and Deep purple within SHARCNET contain 128 and 48 processors respectively, with Compaq Alpha architecture.

The choice of the clusters, Deeppurple and Idra was merely based on the ease of availability of the processors. Deeppurple cluster is running Red Hat Linux 7.2 and Idra is running the Tru64 operating system.

The parallel version of the algorithm was tested using two simple hydrological models, namely DYRESM (Dynamic Reservoir Simulation model) and GAMES (The Guelph model for evaluating effects of Agricultural Management systems on Erosion and Sedimentation). The objective of the

testing of the algorithm was different for the two models.

The objective of the GAMES simulation was to calibrate the alpha parameter used to calculate the delivery ratio in the model. This was done to test the parallel version of the algorithm against a simple hydrological model before it can be applied to more complex hydrological models.

The objective of DYRESM-NWRI simulations was to find the lake depth corresponding to known parameters (i.e. surface temperature). This simulation can be used to find the unknown parameter (lake depth) by correlating it with the known parameters (i.e. AVHRR data source) like surface temperature, long wave and solar radiations. The algorithm simulates different lake depth values, which in turn are used for computation of known parameters by running the lake model. The simulated and observed values are compared and the lake depth value, which results in closest simulated known parameters, is chosen as the optimum lake depth. This simulation has an important application in area of regional climate change modeling. The regional climate change studies usually require simulation of millions of small lakes (especially in case of Canada) where the data for lake depth may not be available.

The predefined parameters, like number of members of the complexes, etc., for the SCE-UA algorithm were computed using the relationships recommended by Duan et al. [1992]. The modified version successfully attained the global optimum value for the test functions.

5.   RESULTS AND DISCUSSION

The hydrological models (DYRESM-NWRI and GAMES) were run on both the SHARCNET clusters (Idra and Deeppurple). This was done to test for replication of the results on different clusters. The sum of square of difference between the target and simulated was used as selection function. The specifications for simulation run are listed in Table 1.

The DYRESM-NWRI dataset of 100 days with time step of 15 minutes for a point in Lake Erie was used as test data. The target data set was obtained by running the hydrological model for 0.95 m lake depth.

The parallel version of the SCE-UA was run and it took approximately 2400 calls to DYRESM-NWRI to achieve the 0.01 tolerance level of 0.95m lake depth. The system timing results for DYRESM-NWRI on both the clusters are presented in Figure 2.

The data for Startford/Avon upper watershed was used for simulation of GAMES. The data set consisted of daily inputs for 471 days. The results for timings obtained for GAMES model with varying number of processors are presented in Figure 3. The target parameter 'α' value was 2.63. It took approximately 1800 calls to GAMES to obtain the target value.

The SHARCNET being a shared cluster, these results were obtained by running the algorithm multiple times to obtain representative mean system time values.

Function	Target parameter	Parameter range	Optimum value
DYRESM-NWRI	Depth	[0.5,1.0]	0.95
GAMES	Alpha	[1,10]	2.63

Table 1. Test models and their specifications

The results show similar trend on both the SHARCNET clusters. It can be seen from the results as the number of complexes per processor decreases, the timing value moves towards the minimum computation and communication time required to run the algorithm. It can also be observed from the results as the complexity of the hydrological model increases (GAMES to DYRESM), the computation time also significantly increases despite the calibration of only single parameter in each case. The algorithm can be used to calibrate as many parameters as needed but it is at the expense of increased computation time with increase in number of parameters. Therefore, use of the parallel instead of serial version in such cases, may result in significant reduction of the computation time.

The successful application of the parallel algorithm to two hydrological models shows that this algorithm can be used for more complex models such as SWAT. Eckhardt and Arnold [2001] found that it took about 18,000 calls to SWAT before the globally optimum 18 calibrated parameters were found. The results were based on a three-year daily hydrological dataset.

For comparison to Eckhardt and Arnold [2001], single run of SWAT were executed for a dataset. The dataset consisted of two-year daily values required to run SWAT for a watershed of size approx. 52.6 km^2. It took about 9.1 sec for single run of SWAT on Deeppurple for this dataset. This shows that parallel version of the algorithm would be highly desirable as compared to the serial version to significantly reduce the time required to calibrate complex hydrological models.

6. CONCLUSIONS

In this study, a modified version of existing serial SCE-UA algorithm is used to calibrate GAMES and find missing lake depth for DYRESM-NWRI. It parallel algorithm has been tested two SHARCNET clusters. The results indicate that the parallel version of the algorithm will reduce the computational time required for automatic calibration of complex hydrological models.

7. ACKNOWLEDGEMENTS

8. REFERENCES

Antenucci, J., and Imerito, A. DYRESM- Science Manual, Centre for Water Research, The University of Western Australia.2002.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R. Large area hydrologic modeling and assessment. I. Model development, *Journal of the American Water Resources Association*, 34(1), 73-89, 1998.

Bosch, D. D., Sheridan, J. M., Batten, H. L., and Arnold, J. G. Evaluation of the SWAT model on a coastal plain agricultural watershed, *Transactions of the ASAE*, 47(5), 1493-1506, 2004.

Di Luzio, M., and Arnold, J. G. Formulation of a hybrid calibration approach for a physically based distributed model with NEXRAD data input, *Journal of Hydrology*, 298(1-4), 136-154, 2004.

Dickinson, W. T., R.P.Rudra, and G.J.Wall, Identification of Soil Erosion and Fluvial Sedimentation Problems, *Hydrologic Processes*, 1, 111-124, 1987.

Dickinson, W. T., R.P.Rudra, and G.J.Wall , Targetting Remedial Measures to Control Non-point Source Pollution, *Water Resources Bulletin*, 26(3), 499-507, 1990.

Duan, Q., Sorooshian, S., and Gupta, V. Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff models, *Water Resources Research*, 28(4), 1051-1031, 1992.

Eckhardt, K., and Arnold, J. G. Automatic calibration of a distributed catchment model, *Journal of Hydrology*, 251(1/2), 103-109, 2001.

Imberger, J., and Patterson, J. C. Dynamic Reservoir Simulation Model - DYRESM: 5, *Transport Models for Inland and Coastal Waters, Academic Press New York. 1981. Proceedings of a Symposium on Predictive Ability*, August 18-20, 36, 1980.

Muleta, M. K., and Nicklow, J. W. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model, *Journal of Hydrology*, 306(1-4), 127-145, 2005.

Muttil, N., and Liong, S.-Y. Superior Exploration–Exploitation Balance in Shuffled Complex Evolution, *Journal of Hydraulic Engineering*, 130(12), 1202-1205, 2004.

Neitsch, S. L., J.G.Arnold, J.R.Kiniry, J.R.Williams, and K.W.King , Soil and Water Assessment Tool Theoretical Documentation - Version 2000, TWRI Report TR-191, Texas Water Resources Institute, College Station, TX..2002.

Peterson, J. R., and Hamlett, J. M. Hydrologic calibration of the SWAT model in a watershed containing fragipan soils, *Journal of the American Water Resources Association*, 34(3), 531-544, 1998.

Rudra, R. P., W.T.Dickinson, D.J.Clark, and G.J.Wall , GAMES - A Screening Model of Soil Erosion and Fluvial Sedimentation in Agricultural Watersheds, *Canadian Water Resources Journal*, 11(4), 58-71, 1986.

Saleh, A., Arnold, J. G., Gassman, P. W., Hauck, L. M., Rosenthal, W. D., Williams, J. R., and McFarland, A. M. S. Application of SWAT for the Upper North Bosque River Watershed, *Transactions of the ASAE*, 43(5), 1077-1087, 2000.

Saleh, A., and Du, B. Evaluation of SWAT and HSPF within BASINS program for the Upper North Bosque River watershed in central Texas, *Transactions of the ASAE*, 47(4), 1039-1049, 2004.

Sharma, V., D.A.Swayne, D.Lam, and W.Schertzer , Parallel Shuffled Complex Evolution Algorithm for Calibration of Hydrological Models, *High Performance Computing Symposium, May 14-17,* St. John's, New Foundland,2006.

Sorooshian, S., and Gupta, V. K. *Computer Models of Watershed Hydrology*. V. P. Singh, ed., Water Resources Publication, Highlands Ranch, CO,1995.

Spruill, C. A., Workman, S. R., and Taraba, J. L. Simulation of daily and monthly stream discharge from small watersheds using the SWAT model, *Transactions of the ASAE*, 43(6), 1431-1439, 2000.

Wischmeier, W. H., and Smith, D. D. Predicting rainfall erosion loss – A guide to conservation planning, *Agriculture Handbook No. 537. Washington D. C. : USDA-ARS*, 1978.