A fast Evolutionary-based Meta-Modelling Approach for the Calibration of a Rainfall-Runoff Model

Soon-Thiam Khu[a], Dragan Savic[a], Yang Liu[a] and Henrik Madsen[b]

[a] *Centre for Centre for Water Systems, School of Engineering, Computer Science and Mathematics, University of Exeter, UK EX4 4QF. Email: {s.t.khu; d.savic; yang.liu}@ex.ac.uk*

[b] *DHI Water and Environment*

Abstract: Population-based search methods such as evolutionary algorithm, shuffled complex algorithm, simulated annealing and ant colony search are increasing used as automatic calibration methods for a wide range of water and environmental simulation models. However, despite the advances in computer power, it may still be impractical to rely exclusively on computationally expensive (time consuming) simulation for many real world complex problems. This paper proposed the use of meta-models to replace numerical simulation models for the purpose of calibration. Meta-models are essentially "model of the model". The meta-model used in this study is the artificial neural network and, when coupled with genetic algorithm, forms a fast and effective hybridisation. The proposed evolutionary-based meta-model reduces the number of simulation runs required in the numerical model considerably thus making the automatic calibration of computationally intensive simulation models viable. The new approach was developed and tested in the calibration of a popular rainfall-runoff model, MIKE11/ NAM, applied to the Treggevaede catchment in Denmark. Both the calibration and verification results for single objective calibration indicate that the proposed method is able to achieve the same or better calibration performance compared to published studies using traditional population-based search methods and yet required only about 40% of the simulation runs on average.

Keywords: evolutionary algorithms; meta-models; rainfall-runoff; artificial neural network; calibration

1. INTRODUCTION

Population-based search methods such as evolutionary algorithm (EA) (which includes genetic algorithms, evolutionary strategies, evolutionary programming etc), shuffled complex algorithm, and simulated annealing are powerful search algorithms that can be used for optimisation. These algorithms are increasing used as automatic calibration methods for a wide range of water and environmental simulation models, especially when there are a large number of calibration parameters and some, or all, of them are interacting with one another. However, the main weakness in using population-based search methods for automatic calibration is that they require a large number of fitness evaluation, thereby render them not suitable to calibrate computational intensive simulation models. It is not uncommon for large simulation models for run for up to an hour or more and with typical EA run requiring thousand (if not tens of thousands) of model evaluations, automatic calibration using EA for large simulation models may not be totally feasible.

Currently, there are two main approaches to resolve the problem using EA for model evaluation and calibration. They are: (i) using faster EA algorithms; and (ii) using more computing power. The first approach exploits the flexibility of EA to develop more efficient techniques requiring less function evaluations and hence, less model evaluations. Typical methods of this approach are: hybridization of EA with some form of heuristics (Deb and Beyer, 2001; Keedwell and Khu, 2003); enhancement to EA operators (reproduction and selection) (Salami and Hendtlass, 2003; Liong et al., 2001). The second approach uses the inherent parallel computing capability of EA and allows simultaneous multiple model simulation on multiple processors. (Kohlmorgen et al., 1999; Rivera, 2001).

However, there is a third method that can be effectively and effortless coupled with EA to enable the calibration of large water and environmental simulation models. To reduce the computational cost of model evaluations/ simulations, surrogate evaluation tools, i.e. meta-models, are used in place of the time consuming simulations. Meta-models, otherwise known as surrogate or approximate models, are essentially "model of the model" which may be used to approximate the simulation model. According to Emmerich et al. (2002), "a metamodel approximates a multivariate function using points that have already been evaluated…. and is considered to be a fast surrogate model to the exact evaluation model." A variety of meta-models exist (e.g. design of experiments, response surface methodology, Taguchi design, kriging, neural networks, multivariate adaptive regression splines) and Simpson et al. (2001) provides a comprehensive review of the use of meta-models for engineering design.

This paper proposed the use of meta-models to strategically replace numerical simulation models (the "Simulator") for the purpose of calibration. The meta-model used in this study is the artificial neural network (ANN) and, when coupled with genetic algorithm (GA), forms a fast and effective hybridisation. This paper starts with an overview of meta-models and its applications in engineering, specifically water resources and environmental engineering. GA and ANN are subsequently described followed by two examples of effective usage of evolutionary-based meta-models for parameter estimation / calibration. Finally section 5 will give concluding remarks and some discussions on future directions.

2. META-MODELS

Meta-models have been in existence for a fairly long period of time (Kleijen, 1975) and are widely used by the engineering design community to reduce the time require for full simulation. An extreme example is the use of meta-model in place of motor vehicle crash test simulations, where Ford Motor Company reports that one crash simulation on a full passenger car takes 36-160 hours (Gu, 2001).

The basic approach of using meta-model for design is as follows:

- Select a multivariate mathematical function (meta-model) which can be used to approximate the "Simulator";
- Run the "Simulator" for a small number of runs;
- Construct the meta-model and adjust the variables within the model to fit the run results from the simulator;
- Once the adjustments are complete, the meta-model is used in place of the "Simulator" for future evaluation of the new designs.

The above approach requires certain modification if the "Simulator" is constantly changing, such as the case during calibration. The modifications are:

- Make necessary adjustments by usually running the "Simulator" more times; and
- A mechanism to update the meta-model.

Regardless of the usage of meta-models, three steps are involved (Simpson et al., 2001):

1. choosing an experimental design for generating data;
2. choosing a mathematical model to represent the data; and
3. fitting the model to the observed data.

Each of the steps may have many options and the choice of option in each step give rise to different meta-models. For example, generating data using fractional factorial design and fitting the data onto a second order polynomial function using method of least squares regression gave rise to the meta-model known as "response surface methodology", while measured data may be fitted onto a network of artificial neurons using least squares with back-propagation giving rise to "artificial neural network" as a meta-model.

Meta-models have also been successfully applied to model a variety of water and environmental problems. Some examples are: the response surface method has been applied to predict numerical geophysical models (Tatang et al., 1997), reconstruction and interpolation of effluent plume in an estuary (Riddle et al., 2004) and calibration of urban drainage model (Liong et al., 1995); Kriging has been used to model spatio-temporal pollutant deposit trend through the atmosphere (Haas, 1998), spatial distribution of heavy metals in a river basin (Ouyang et al., 2002) and shallow water wave in an estuary (Gorman and Neilson, 1999); Artificial neural networks have been used to model the input-output behaviour of wastewater treatment plants (Belanche et al., 1999), deforestation simulation (Mas et al., 2004), prediction of pollutant trends in urban areas (Lu, et al., 2004); and many others applications.

3. EA-BASED META-MODELS

3.1 Genetic Algorithm

One of the most common EA is the genetic Algorithms (GAs). GAs are computationally simple yet powerful search algorithms based on the mechanics of natural selection and natural genetics, which combines an artificial survival of the fittest with genetic operators from nature. GAs mimic the adaptation of natural species and genetically evolve to suit their environment over many generations. Using this analogy, a mechanism involving selection, crossover, and mutation can be used to evolve a population of potential solutions towards improved solutions.

GAs are especially useful for complex optimisation problems where the analytical solutions are difficult to obtain and it has been used for water and environmental model optimisation and calibration since early 90s. However, one of the main obstacles when trying to apply GA in practical optimisation problems is the large number of function evaluations required.

3.2 Artificial Neural Networks

Artificial neural network (ANN) is a computing paradigm designed to mimic natural neural networks in the biological brain. ANNs are commonly thought of as universal approximators for function mapping. Multilayer perceptrons (such as feed-forward backpropagation algorithms) and radial basis functions (RBF) are commonly used ANNs. In this study, the RBF neural network has been suggested because its simple structure enables learning in stages, gives a reduction in the training time. A standard RBF network has a feed-forward structure consisting of two layers, a nonlinear hidden layer and a linear output layer, and uses a Gaussian function as activation function to transforming inputs. Clustering identifies the centre point and radius (i.e., mean and standard deviation) of the Gaussian function in each unit of the RBF network.

3.3 Integrating Meta-models with GAs

As stated in the introduction, one possible way of overcoming the problem of time consuming simulation in EAs (including GA) is to use meta-models in place of the simulation model. Many researchers, especially in engineering design, have examined strategies to integrate different meta-models with GA (Giannakoglou et al., 2001; Poloni et al., 2000; Ong et al., 2003).

The most direct way of integrating meta-models with GA is to replace the "Simulator" with the meta-model completely during evaluation of objective function in GA. However, in order to construct the meta-model, a small number of run of the "Simulator" is required. This is the experimental design mentioned in Section 2 and can be performed either using Taguchi method, Design of Experiments, response surface methodology or even using GA. Liong et al. (2001) detailed one such method using fractional fractorial design with central composite design to provide initial population for GA. Emmerich et al. (2002) used kriging as the meta-model and they found that Kriging provided the local error estimation which enable them to assess the reliability of the solutions. Giannakoglou et al. (2001) used radial basis function network as meta-model coupled with GA to optimise an airfoil shape design. Poloni et al. (2000) used a hybridisation of GA, ANN and local search method to optimise the design of a sailing yacht fin keel. The ANN acted as a surrogate model for 3D Navier-Stokes simulation of the fin keel while cruising.

Another potential usage of evolutionary-based meta-model is the evaluation of risk and uncertainty. Currently, different sampling approaches have been devised to perform fast and effective sampling. Monte-Carlo sampling (MCS) is commonly regarded as the most accurate approach but it requires thousands, if not tens of thousands, of model evaluation. Importance sampling, Metropolis algorithms, Latin Hypercube method etc., are fast alternatives but they are approximating the statistical properties of the MC samples. Recently, Khu and Werner (2003) proposed the use of meta-model (GA – ANN) to select regions of interest for sampling. Their method required only about 10% of the MCS method.

Despite the extensive works in evolutionary-based meta-models, little effort is place on overcoming the problem of "changing landscape". During the process of optimisation, the region of GA search will constantly change, and it is reasonable to assume that the meta-model will have to be suitability modified to account for such changes. As the search progresses, more information on the objective function will be obtained, and suitable mechanism should be implemented to utilise this additional information and update the meta-model. The example in the next section demonstrates such a scheme where a GA-ANN meta-model is used to calibrate a rainfall-runoff model and the meta-model is constantly but strategically updated with the latest information.

4. CALIBRATION OF A RAINFALL-RUNOFF MODEL

A fast evolutionary-based meta-model using an innovative hybridisation of GA and RBF is proposed for the automatic calibration of numerical simulation models. The GA is used to search for the optimal objective function in much the same way as any optimisation routine. The ANN is used to map (and adapt) to the response surface of the objective function and used as a fast surrogate for the NAM model at regular intervals. The concept of using RBF as surrogate models is not new. However, as the GA search progresses, the response surface of the objective function tends to change and therefore, the meta-model needs to be self-adapting to the changing landscape.

The hybrid algorithm (GA-RBF) is presented below and illustrated in Figure 1:

(1) Run GA for g number of generations of population size, p;
(2) Train RBF to map the response surface using (g * p) points generated by GA;
(3) Perform selection and reproduction in GA;
(4) Evaluate the new GA population using the trained RBF instead of NAM;
(5) Select m best individuals in the new population and evaluate the true fitness using the NAM model;
(6) Update RBF using the true fitness from (5);
(7) Perform steps (3) to (6) until the stopping criterion is met.

4.1 Application Example

The proposed meta-model was adopted for calibration of the MIKE 11/NAM rainfall-runoff model applied to the Danish Tryggevaelde catchment. The calibration parameters used are the same as those in Madsen (2000). This catchment has an area of 130km^2, an average rainfall of 710 mm/year and an average discharge of 240 mm/year. The catchment is dominated by clayey soils, implying a relatively flashy flow regime. For the calibration, a 5-year period (1 Jan. 1984−31 Dec. 1988) was used where daily data of precipitation, potential evapo-transpiration, mean temperature, and catchment runoff are available. For comparing the calibrate models, validation data covering the periods 1 Jan. 1979−31 Dec. 1983 and 1 Jan. 1989−31 Dec. 1993 were used. The following two of objective functions are used in this study:

Average Root Mean Squared-Error (RMSE) of low flow events:

$$F_1(\theta) = \frac{1}{M_l}\sum_{j=1}^{M_l}\left[\frac{1}{n_j}\sum_{i=1}^{n_j}[Q_{obs,i} - Q_{sim,i}(\theta)]^2\right]^{1/2} \qquad (1)$$

Average Root Mean Squared-Error (RMSE) of peak flow events:

$$F_2(\theta) = \frac{1}{M_p}\sum_{j=1}^{M_p}\left[\frac{1}{n_j}\sum_{i=1}^{n_j}[Q_{obs,i} - Q_{sim,i}(\theta)]^2\right]^{1/2} \qquad (2)$$

In Eqs. (1)−(2), $Q_{obs,i}$ is the observed discharge at time i, $Q_{sim,i}$ is the simulated discharge, M_p is the number of peak flow events, M_l is the number of low flow events, n_j is the number of time steps in peak/low event no. j, and θ is the set of model parameters to be calibrated. Peak flow events were defined as periods with flow above a threshold value of 4.0 m^3/s, and low flow events were defined as periods with flow below 0.5 m^3/s.

Preliminary optimisation runs showed that entire population converged around the global optimum after about 2000 model evaluations. Thus, for each test, a maximum number of model evaluations equal to 2000 were employed as a stopping criterion with the population size p=50 and total number of generations G=40. In this work, floating point representation was used together with uniform crossover. The crossover rate used was 0.9 and mutation rate = 0.1.

Sensitivity analysis was performed to provide some idea of the values of g and m to use and the values of g = 2 and m=15 was found to provide good results. Hence, the number of simulation runs for NAM model calibration using the hybrid method can be calculated by: $p * g + m * (G - g)$ = 50 × 2 + 15 × (40 − 2) =670, which is 33.5% of the required GA run.

4.2 Results and Discussions

A total of 10 random calibration runs were perform for each objective function (Eqns. (1) and (2)) and the evaluated performance statistics of the GA and the hybrid method (GA-RBF) are shown in Tables 1. Tables 1 also shows the validation results of applying the calibrated parameter set to the two different validation periods. The hybrid method (GA-RBF) gave very close results to those from GA (in terms of best, worst and mean RMSE) but required only about 40% of the GA simulation runs. Even though some of the calibration results indicated that GA-RBF were be better than GA, mixed but comparable results were obtained for the validation data sets. The small standard deviation indicated the hybrid method is very stable and able to reproduce consistent results.

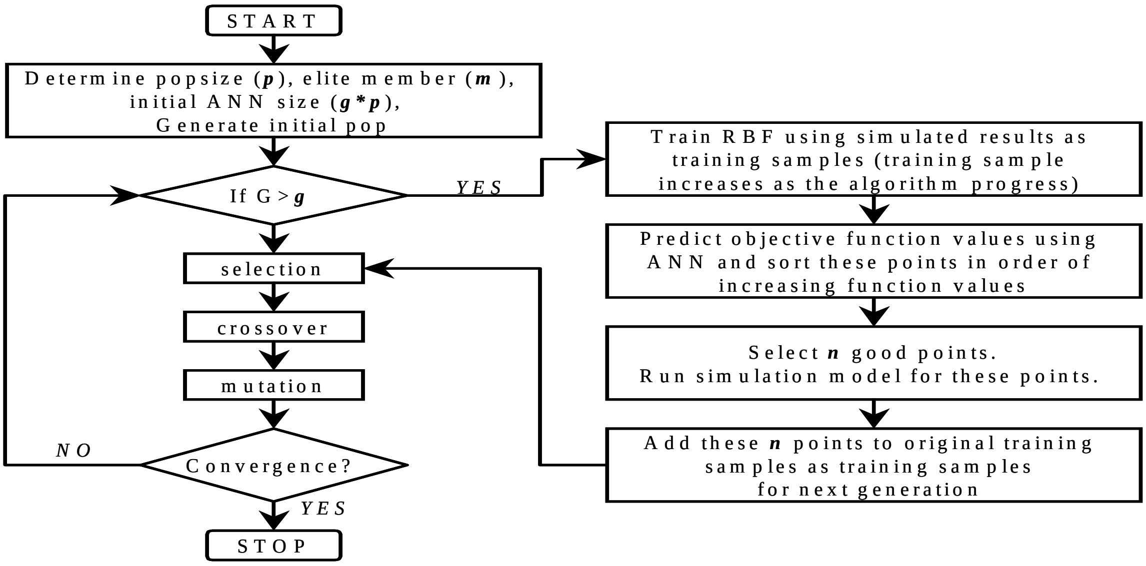

Figure 1: Flow chart of the evolutionary-based meta-model

Table 1: Calibration and validation results for low flow and peak flow RMSE

		Calibration data (1984−1988)		Validation data set 1 (1979−1983)		Validation data set 2 (1989−1993)	
		GA	GA−RBF	GA	GA−RBF	GA	GA−RBF
Low Flow RMSE (m^3/s)	Best RMSE	0.1345	0.1323	0.2065	0.1752	0.0986	0.1043
	Worst RMSE	0.1782	0.1697	0.2491	0.2516	0.1470	0.1519
	Mean	0.1543	0.1451	0.2215	0.2196	0.1217	0.1164
	STD	0.0119	0.0110	0.0140	0.0210	0.0166	0.0159
Peak Flow RMSE (m^3/s)	Best RMSE	1.1750	1.1687	1.1724	1.1836	1.0485	1.1325
	Worst RMSE	1.2378	1.2178	1.2564	1.2516	1.3945	1.3672
	Mean	1.2016	1.1966	1.2303	1.2165	1.2005	1.2386
	STD	0.0184	0.0175	0.0292	0.0227	0.0964	0.0798

5 CONCLUSIONS

This paper discusses the concept of meta-model and the integration between evolutionary algorithms and meta-models. It can be seen that there is significant advantage in using meta-models for water and environmental system simulation, design and calibration. However, one major problem for evolutionary-based meta-modelling is how to ensure that the meta-model is constantly relevant as the search progresses. To overcome this problem, a strategic and periodic scheme of updating the meta-model is proposed.

A novel evolutionary-based meta-model, using genetic algorithm and radial basis function neural network with dynamic updating, for hydrological model calibration has been proposed in this paper. It has been shown that the proposed method performed more efficiently when compared to GA. The results indicated that the proposed method was able to reduce the required simulation runs to 40% of GA while achieving comparable calibration and validation results. The results provided us with confidence that the proposed method is indeed a viable method to reduce the computation effort required in calibrating rainfall-runoff models while constantly updating the meta-model. The proposed methodology presents a viable option to calibration and optimise computationally intensive water and environmental simulation models.

Work is currently undergoing to (i) test the proposed algorithm on known different functions, (ii) extend the work to multiple objective functions; and (iii) extend the scheme to become even more adaptive with the changing landscape. Jin et al. (2002) addressed the dynamic landscape issue using "fuzzy" error rules to determine the frequency and timing of fitness approximation and it may be possible to incorporate such method with the proposed algorithm.

6 ACKNOWLEDGEMENTS

The financial assistance from ABPmer Ltd to Yang Liu is greatly appreciated.

7 REFERENCES:

Belanche, L.A., J.J. Valdes, J. Comas, I.R. Roda and M. Poch. Towards a model of input-output behaviour of wastewater treatment plants using soft computing techniques. *Env. Modelling & Software, 14*, 409-419, 1999.

Deb, K. and H.G. Beyer. Self-adaptive genetic algorithm with simulation binary crossover. Evolutionary Comp., 9(2), 197-221, 2001.

Emmerich, M., A. Giotis, M. Ozdemir, T. Back and K. Giannakoglou. Metamodel-Assisted Evolution Strategies. *Lecture Notes in Computer Sc. 2439*, 361-370, 2002.

Giannakoglou, K.C., A.P. Giotis, and M. Karakasis. Low cost genetic optimisation based on inexact pre-evaluations and the sensitivity analysis of design parameters. *Inverse Problems in Engrg., 9*, 389-412, 2001.

Gorman, R. M. and C. G. Neilson. Modelling shallow water wave generation and transformation in an intertidal estuary. *Coastal Engrg., 36(3)*, 197-217, 1999.

Haas, T.C. Statistical assessment of spatio-temporal pollutant trends and meteorological transport models. *Atm. Env., 32(11)*, 1865-1879. 1998.

Jin, Y., M. Olhofer and B. Sendhoff. A framework for evolutionary optimisation with approximate fitness functions. *IEEE Transactions on Evolutionary Computation, 6(5)*, 481-494, 2002.

Keedwell, E. and S.T. Khu. More choices in water system design through hybrid optimisation. *Proc. Int. Conf. on Comput. and Control for the Water Industry*, pp. 257-264, 2003.

Kleijnen, J.P.C. A comment on blanning's metamodel for sensitivity analysis: The regression metamodel in simulation. *Interfaces, 5(1)*: 21-23, 1975.

Khu, S.T. and M.G.F. Werner. Reduction of Monte-Carlo simulation runs for uncertainty estimation in hydrological modelling. *Hydrol. Earth Sys. Sc., 7(5)*, 680-692, 2003.

Kohlmorgen, U., H. Schmeck and K. Haase. Experiences with fine-grained parallel genetic algorithms. *Annals of Op. Res., 90*, 203-219, 1999.

Lu, W.-Z., W.-J. Wang, X.-K. Wang, S.-H. Yan and J.C. Lam. Potential assessment of a neural network model with PCA/RBF approach for forecasting pollutant trends in Mong Kok urban air, Hong Kong. *Env. Res., in press*, 2004.

Liong, S.Y., J. ShreeRam and W.T. Chan. Catchment calibration using fractional factorial and central composite designs-based response surface, *J. Hydr. Engrg., 121(8)*, 613-617, 1995.

Liong, S.Y., S.T. Khu and W.T. Chan. Derivation of Pareto front with genetic algorithm and neural networks. *J. Hydrol. Engrg., 6(1)*, 52-61, 2001.

Madsen, H. Automatic calibration of a conceptual rainfall-runoff model using multiple objectives, *J. Hydrol.*, 235, 276-288, 2000.

Mas, J.F., H. Puig,, J.L. Palacio and A. Sosa-Lopez. Modelling deforestation using GIS and artificial neural networks. *Env. Modelling & Software, in press*, 2004.

Ong, Y.S., P.B. Nair and A.J. Keane. Evolutionary optimisation of computationally expensive problems via surrogate modelling. *AIAA J., 40(4)*, 687-696, 2003.

Ouyang, Y., J. Higman, J. Thompson, T. O'Toole and D. Campbell. Characterization and spatial distribution of heavy metals in sediment from Cedar and Ortega rivers sub-basin. *J. Contaminant Hydrol., 54(1-2)*, 19-35, 2002.

Poloni, C., A. Giurgevich, L. Onesti, and V. Pediroda. Hybridization of a multi-objective genetic algorithm, a neural network and a classical optimzer for a complex design problem in fluid dynamics. *Comput. Methods in Applied Mech. and Engrg., 186*, 403-420, 2000.

Riddle, A.M., R.E. Lewis and A.D. Sharpe. Environmental quality information processor (EQUIP). *Env. Modelling & Software, 19(1)*, 57-62. 2004.

Rivera, W. Scalable Parallel genetic algorithm. *Artificial Intel. Rev., 16*, 153-168, 2001.

Salami, M. and T. Hendtlass. A fast evaluation strategy for evolutionary algorithms. *Applied Soft Comput., 2*, 156-173, 2003.

Simpson, T.W., J. Peplinski, P.N. Koch and J.K. Allen. Metamodels for computer-based engineering design: Survey and recommendations. *Engrg. With Computers, 17(2)*: 129-150, 2001.

Tatang, M.A., W. Pan, R.G. Prinn and G.J. McRae. An efficient method for parametric uncertainty analysis of numerical geophysical models. *J. Geophysical Res., 102*, 21925-21932, 1997.