HAL
open science

# Computationally efficient calibration of WATCLASS Hydrologic models using surrogate optimization

M. Kamali, K. Ponnambalam, E. D. Soulis

## To cite this version:

M. Kamali, K. Ponnambalam, E. D. Soulis. Computationally efficient calibration of WATCLASS Hydrologic models using surrogate optimization. Hydrology and Earth System Sciences Discussions, 2007, 4 (4), pp.2307-2321. hal-00298862

**HAL Id: hal-00298862**

**https://hal.science/hal-00298862v1**

Submitted on 18 Jun 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Hydrol. Earth Syst. Sci. Discuss., 4, 2307–2321, 2007
www.hydrol-earth-syst-sci-discuss.net/4/2307/2007/
© Author(s) 2007. This work is licensed
under a Creative Commons License.

Hydrology and Earth System Sciences Discussions

Papers published in *Hydrology and Earth System Sciences Discussions* are under open-access review for the journal *Hydrology and Earth System Sciences*

# Computationally efficient calibration of WATCLASS Hydrologic models using surrogate optimization

**M. Kamali, K. Ponnambalam, and E. D. Soulis**

University of Waterloo, Waterloo, ON, Canada

Received: 16 May 2007 – Accepted: 25 May 2007 – Published: 23 July 2007

Correspondence to: M. Kamali (mkamali@engmail.uwaterloo.ca)

## Abstract

In this approach, exploration of the cost function space was performed with an inexpensive surrogate function, not the expensive original function. The Design and Analysis of Computer Experiments(DACE) surrogate function, which is one type of approximate models, which takes correlation function for error was employed. The results for Monte Carlo Sampling, Latin Hypercube Sampling and Design and Analysis of Computer Experiments(DACE) approximate model have been compared. The results show that DACE model has a good potential for predicting the trend of simulation results. The case study of this document was WATCLASS hydrologic model calibration on Smokey-River watershed.

## 1 Introduction

Hydrologic model calibration has been a challenge for hydrologists for decades. A hydrologic model is a representation of a part of nature which is capable of reproducing some of its characteristics, such as streamflow. Generally, in a hydrologic model, many parameters are involved and the process of finding the suitable values for these parameters is called model calibration or parameter estimation.

The process of parameter estimation will be essentially an optimization process, in which a cost function will be optimized and the corresponding parameter set will be selected. The main issues that make this optimization process more complicated than a traditional optimization problem are: first, the cost function is not convex and smooth; second, the cost function is usually computationally very expensive; third, although most of parameters have physical meaning individually, the combination which give optimum results is not necessarily a reasonable answer.

One point of view in hydrologic model calibration is that there is just one set of the best parameters. Consequently, because of the nature of the problem the optimization methods used for finding this one point are more concentrated around population

based optimization methods such as Genetic Algorithm (Fraser and Burnell, 1970), Simulated Annealing (Metropolis et al., 1953) and Shuffle Complex Evolution (Duan et al., 1992). These processes usually start from a randomly chosen initial point and continue by searching the space for a better point until no better point is found. These methods are considerably successful in finding near global optima, yet there is no mathematical proof for any of them except for SA if the annealing process cools down very slowly (SA converges to a Monte Carlo), they do not offer any insight into the performance of the model. In other words, it is unlikely to discover a region where the model performs well and a region where it does not.

The other point of view claims that there are many good sets of parameters. Usually more than one set of good parameters are chosen and all of them will be reported as candidates. Since all hydrologic models, both conceptual and physically distributed, are based on some kinds of approximation and simplification, the second point of view seems more attractive.

This project attempts to discover regions where the model performs reasonably well. This paper compares Monte Carlo Sampling, Latin Hypercube Sampling (LHS) and loosely grided LHS along with DACE method for calibration of WATCLASS over Smokey Watershed. Using DACE as surrogate model, the process of finding the sets of best parameters will be done. By doing this, the computational cost will be reduced and the performance of the process will improve. The reminder of this document is organized as follows: a brief description of the Smokey-River Watershed will be followed by a brief introduction of WATCLASS.[1] .The theory of DACE and the experimental part will come next.

[1] During this research a new version of WATCLASS was released. Although this new version results were better than the version we worked with, the computational cost of the experiments did not allow us to repeat these experiments with the new version.

## 1.1 Smokey-river watershed

Smokey-River watershed is a part of Peace River watershed. With a drainage area of 3840 $km^2$, this river drains to the foothills of Rocky Mountains. This watershed is mostly alpine forest and is located in the northwest of Edmonton, Alberta, Canada.

## 1.2 WATCLASS

WATCLASS (Soulis and Verseghy, 2000) model is an integration of WATFLOOD Kouwen (1988) and CLASS Verseghy (1991), the former model is a hydrologic model and the latter model is a land scheme model.

WATFLOOD is based on the GRU (Grouped Response Unit), which for each grid takes a dominant landcover based on the largest vegetated area. This model takes the distribution of precipitation, interception and infiltration over the landscape. Class, Canadian Land Surface Scheme, is a landcover-based model with a modeling strategy similar to GRU. In this approach each grid cell is represented by a single landscape type. The properties of each landscape, however, are made of different landscape types. WATCLASS approach is very similar to GRU, but it is more developed in a sense that any landscape unit with its defined hydrological response could be considered. In addition to that, in order to accommodate the gradient needed for lateral flow a local slope is assigned to each of the landscapes.

Generally, a combination of an atmospheric model and a hydrologic model that satisfies Environment Canada standard was the main motivation behind coupling these two models. This coupling enables WATCLASS users to take advantage of strength of both models such as complete vertical processes modeled in CLASS and a well developed routing algorithm in WATFLOOD. The end target of development of WATCLASS is to construct an operational model for climate simulation and forecasting.

## 2 Surrogate optimization

In many real world problems, traditional optimization methods fail to work. In cases the objective function is not convex and the derivative is either expensive or not available , these methods are not very applicable. As mentioned before, WATCLASS calibration is one of these cases. Thus the use of other types of optimization methods such as approximate modeling approach becomes reasonable.

The idea of employing approximate model in optimization of expensive functions was first propounded by Jones et al. (1998) called Design and Analysis of Computer Experiments(DACE) (Jones, 1998). This idea has been customized and applied in variety of engineering problems such as shape optimization (Marsden et al., 2004), calibration of groundwater bioremediation models (Mugunthan, 2005) and assessment of parameter uncertainty in groundwater models (Mugunthan and Shoemaker, 2006).

In this method, it is supposed that simulations have been produced by a mathematical model and this model is the so called surrogate model. Based on simulated points, the surrogate model or approximate model is constructed. This approximate model imitates that part of the region of objective function and it will be used to predict the model optimum. This process is an iterative process and when the surrogate found the optimum with a good approximation the process is terminated. In other words, the optimum of the constructed approximate model will be compared to its corresponding simulated result, if the error was lower than the threshold, the model is good enough and this model could be used later, otherwise the model will be modified and another simulation will be performed. Once the approximate model prediction was satisfactory, it could imitate the expensive process of simulation.

## 2.1 Introduction to DACE

In optimization using DACE approach, it is assumed that observations are coming from a model and $\hat{y}$ could be modelled as follows:

$$y(x^i) = \sum_{i=1}^{n} \beta_h f_h(x^i) + \epsilon^i \qquad (i = 1, \ldots, n) \tag{1}$$

In this equation, n is number of inputs, $f_h(x)$'s are linear or nonlinear functions of $x$; the $\beta$'s are their corresponding coefficients, which are unknown; and $\epsilon^i$ are white noise, which is a normally distributed, independent error terms with zero mean and variance $\sigma^2$.

There is a subtle point worth mentioning here, that makes this approach very different from response surface approach. In this approach, it is assumed that errors are related or correlated, but not independent, and the correlation between error terms depends on their distances. In other words, the correlation is high when two points are close, and as they get farther apart, the correlation decreases. So, the approximate model will not be the same as a regression model. This particular type of stochastic process using spatial correlation functions is called Kriging Krige (1988). This method is named after a South African engineer, who originally developed this model to predict the location of ore reserves precisely. Commonly, the above stochastic process is called "DACE stochastic process model."

The main reason for treating error in this way comes from the nature of errors in computer codes, which stems from modeling errors not measurement or noise. The correlation between errors at $x^i$ and $x^j$ is defined as:

$$\mathrm{Corr}[\epsilon(x^i), \epsilon(x^j)] = \exp[-d(x^i, x^j)] \tag{2}$$

In Eq. (2), $d$ is the distance between two points and is defined as:

$$d(x^i, x^j) = \sum_{i=1}^{k} \theta_h |x_h^i - x_h^j|^{p_h} \qquad (\theta_h \geq 0, p_h \in [1, 2]) \tag{3}$$

In Eq. (3), $\theta_h$ and $p_h$ are two parameters that needed to be adjusted. $\theta_h$ is the parameter that controls the weight of the distance of two variables and $p_h$ is the parameter that controls the smoothness of the function in $h$ direction. These parameters will be estimated by maximizing the likelihood of the sample. As a result, based on known points we are able to predict unknowns.

Suppose the value of the function is unknown in a location and we wish to predict its value based on the values of known points. In the case of a linear predictor, we have

$$\hat{y}(x^i) = c(x)^T y_s \tag{4}$$

In this equation, $y_s$'s are function values of m known points; $c$ is the vector of coefficients. By determining $c$, it is possible to estimate the value of unknown points. Now, using Eq. (4) and Eq. (1), we can estimate the value of unknowns by minimizing the Mean Square Error (MSE) for the estimated points as follows:

$$\mathrm{MSE}[(\hat{y}(x))] = E[c(x)^T Y_s - Y(x)]^2 \tag{5}$$

Substituting and imposing unbiasedness constraint gives

$$\mathrm{MSE}[(\hat{y}(x))] = E[\epsilon^2 + c^T \mathbf{e}\mathbf{e}^T c - 2c^T \mathbf{e}\epsilon] \tag{6}$$

In Eq. (6), $\epsilon$ is a realization of $\mathbf{e}$; we have $[\mathrm{E}(\epsilon^2)]=\sigma^2$, $\mathrm{E}[\mathbf{e}\epsilon] = \sigma^2 r$ and $\mathrm{E}[\mathbf{e}\mathbf{e}^T]=\sigma^2 R$, where $r$ is the correlation vector between the error of unknown points and an untried point $x$, and $R$ is the correlation matrix between the error of the known points. Minimizing MSE will result in a set of regression and correlation function coefficients.

There are several choices for the basis function and correlation function. However, in this problem the polynomial function and exponential correlation function were selected.

## 3 Experiments and results

### 3.1 Decision variables

In this experiment, 14 parameters were calibrated. It is supposed that all of these parameters are uniformly distributed within their upper and lower bounds.The description and intervals of these parameters are in Table 1.

### 3.2 Experiment process

The error criterion that was chosen for evaluation of the suggested model was Nash-Sutcliffe Nash and Sutcliffe (1970) coefficient which is defined as following:

$$NS(x) = 1 - \frac{\sum_{i=1}^{n}(Y_{\text{observed}} - Y_{\text{simulated}})^2}{\sum_{i=1}^{n}(Y_{\text{observed}} - E(Y_{\text{observed}}))^2} \tag{7}$$

The time interval in Eq. (7) is from 1 to $n$. This case study is based on one daily reading of streamflow in period of 3 years. In this equation Y is the predicted value (streamflow here) and $E(Y_{\text{observed}})$ is expectation of $Y_{\text{observed}}$. NS number will be negative when the prediction of the model is very different from the observed value, and close or equal to one when the model prediction is perfect. Therefore, as NS numbers become closer to 1 model prediction becomes closer to observed values. In this paper, a NS number greater than or equal to 0.7 is considered acceptable.

In order to evaluate the approximate model, Monte Carlo sampling with 10 000 random samples was done. This way, the performance of the approximate model will be compared with the performance of the WATCLASS. In the second set of experiment, 6000 Latin Hypercube Samples McKay et al. (1979) were produced. In LHS, random samples are generated from a grided space, this ensures that samples are from the entire space.

The third set of experiments was focused on approximate model evaluation. This set had two parts. In the first part, 1000 LHS were generated on the input space and

their corresponding NS number was computed.In order to increase the concentration of points with good prediction (lower error values/higher NS coefficient), 10 best points were selected and in the close vicinity of these points another 500 simulations were performed. This 500 points were located in a hypercube that its edges extend up to the upper bound and lower bound of what that previously 10 selected parameter sets were located. In the second part of the third set of experiment, based on the 500 best results, an approximate model was constructed. This approximate model employed to predict points with NS numbers more than 0.70. So, 10 000 LHS were generated and their corresponding NS coefficients were evaluated by the approximate model. Then, among these 10,000 points 100 best points were selected. To evaluate the performance of the approximate model, these best points were simulated with WATCLASS. Then, the estimated results were compared to the simulated results. Deviation of predicted results (from approximate model) from the simulated results (WATCLASS) was the measure for goodness of the approximate model. After 4 iterations (400 points) of approximate model evaluation satisfactory results were achieved. The schematic of the algorithm is shown in Fig. 1. Efficiency of these processes are compared in Table 2. Since the complete process of approximate model construction relies on random sampling of the points, the process replicated 10 times and the mean and standard deviation of approximate model process is reported.

All of the simulations were performed on a "Shared Hierarchal Academic Research Computing Network" Called SHARCNET [2] This service allows the users to conduct hundreds of jobs in parallel.

Comparison of the results shown in Table 2 shows that Latin Hypercube Sampling was slightly more successful in finding points with high NS numbers than random sampling (Monte Carlo). Since in LHS all the space has been sampled, the approximate model was constructed based on this sampling method. This way, the entire space of

[2]SHARCNET based on SHARCNET website is "A multi-institutional high performance computing network that spans 16 leading academic institutions in south central Ontario, Canada." University of Waterloo is part of this network.

parameters will be explored.

In the first row of Table 2, the entire process of DACE approximate model is shown. In the entire process 1900 simulation performed, which only 400 of simulations were based on approximate model prediction. Comparison of LHS and DACE (firts part) shows that on average 24% of points have NS greater than 0.7, whereas in case of LHS method 15% of points have NS number greater than 0.7. In addition to that, standard deviation of the distribution is small.

The results are the same for the rest of NS number intervals, but the standard deviation is slightly higher correspondingly. However, all of 10 trials in this experiments have better results than LHS and as NS intervals moves to the higher intervals the capability of approximate model to find good points increases. In the second part of DACE, the NS number of 400 points that were simulated based on the approximate model is reported. In this part, 36% of the results are above the first threshold, which is 21% better than LHS. Among all identifying points, 29% were in the next threshold (NS greater than 0.71), but in LHS only 11% of points were in that interval.

Overall, DACE method outperformed previous methods in identifying good input sets. DACE model was used to map part of the input space that seemed to contain more acceptable outputs. Model precision increased in each iteration until the error criteria met. This experiment continued for 4 iterations. DACE has the potential of being adapted with any other model, for example neural network or spline can be one of the cases for the approximate model.

The other advantage of DACE is its adaptability with other optimization methods. This technique is not only permits user to employ other types of optimization methods such as GA or SA, but also can be combined with any of these methods in an efficient way. For instance, one of the main disadvantages of these population based techniques is being computationally expensive especially when the dimension of the space is high. However, if the input space shrinks by using DACE (smaller parameter intervals, in this case input interval became 50% of the original input interval), the computational cost will be definitely less.

## 4 Conclusions

he goal of this research was to evaluate the performance of DACE for hydrologic model calibration. Number of acceptable points corresponding to different methods were compared and the results are reported in Table 2.

This experiment showed promising results for calibration of a small watershed (in North American term). Application of DACE along with LHS reduced the computational cost of calibration process. For example, in this experiment, each WATCLASS simulation took 8 min. Therefore, finding 100 points with NS number greater than 0.7 using LHS method will take 5152 min if the process is completely serial, whereas using DACE for the same case will take only 3360 min. So that, the overall computational cost will be 35% less.

DACE could be combined well with clustering techniques and this way the modeling process will be done on the part of the space that contains acceptable points. This way, the computational expenses will be even less by being more concentrated on the region of interest and getting less sample from other part of the space. It could be also combined with dimensionality reduction techniques and work in a lower dimension. So, the dimension of the modeled space will be smaller (much smaller).

In general, this technique showed the potential of working with any expensive optimization problem, such as calibration of the hydrologic models over large watershed.

## References


Jones, D. R., Schonlau, M., and Welch, W. J.: Efficient Global Optimization of Expensive Black-Box Functions, J. Global Optimization, 13, 455–492, 1998. 2311

Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resour., 4, 1015–1031, 1992. 2309

Soulis, E. D., Snelgrove, K. R., Kouwen, N., Seglenieks, F., and Verseghy, D. L.: Toward Closing the Vertical Water Balance in Canadian Atmospheric Models:Coupling of the Land Surface

Scheme CLASS with the Distributed Hydrological Model WATFLOOD, Atmos.-Ocean, 38, 251–269, 2000. 2310
Fraser, A. and Burnell, D.: Computer Models in Genetics, McGraw-Hill, New York, 1970. 2309
Kouwen, N.: WATFLOOD: A micro-computer based flood forecasting system based on real-time weather radar, Can. Water Resour. J., 13, 62–77, 1988. 2310
Krige, D.: A statistical approach to some mine valuations and allied problems at the Witwatersrand, Can. Water Resour. J., 13, 62–77, 1988. 2312
Marsden, A. L., Wang, M., Dennis Jr., J., and Moin, P.: Optimal Aeroacoustic Shape Design Using the Surrogate Management Framework, Optimization and Engineering, 5, 2004. 2311
McKay, M., Conover, W., and Beckman, R. J.: A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code., Technometrics, 21, 239–245, 1979. 2314
Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E.: Equation of State Calculations by Fast Computing Machines., J. Chem. Phys., 21, 1087–1092, 1953. 2309
Mugunthan, P. and Shoemaker, C.: Assessing the impacts of parameter uncertainty for computationally expensive groundwater models., Water Resour. Res., 42, 2006. 2311
Mugunthan, Pradeep, S. C. A. R. R. G.: Comparison of function approximation, heuristic, and derivative-based methods for automatic calibration of computationally expensive groundwater bioremediation models, Water Resour. Res., 41, 1–7, 2005. 2311
Nash, J. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I A discussion of principles, J. Hydrol., 10, 282-290, 1970. 2314
Verseghy, D. L.: CLASS-A Canadian Land Surface Scheme for GCMs, I. Soil Model, Int. J. Climatol., 11, 111–133, 1991. 2310

**Table 1.** WATCLASS Parameter description.

| Parameter | Description | Lower Bound | Upper Bound |
|---|---|---|---|
| drnrow | Drainage index for layer 1 | 0 | 1 |
| wfcirow | An interflow index (mean of KSAT[a]) for layer 1 | 0 | 20 |
| sand11 | volume of sand in layer1 landclass1 | 0 | 100 |
| clay11 | volume of clay in layer1 landclass1 | 0 | 100 |
| sand12 | volume of sand in layer1 landclass2 | 0 | 100 |
| clay12 | volume of clay in layer1 landclass2 | 0 | 100 |
| sand21 | volume of sand in layer2 landclass1 | 0 | 100 |
| clay21 | volume of clay in layer2 landclass1 | 0 | 100 |
| sand22 | volume of sand in layer2 landclass2 | 0 | 100 |
| clay22 | volume of clay in layer2 landclass2 | 0 | 100 |
| drnrow | Drainage index for layer 2 | 0 | 1 |
| wfcirow | interflow index (mean of KSAT)for layer 2 | 0 | 20 |
| drnrow | Drainage index for layer 3 | 0 | 1 |
| wfcirow | interflow index (mean of KSAT)for layer 3 | 0 | 20 |

[a] Hydraulic Conductivity at Saturation

**Table 2.** Comparison of mean of number of simulations for different methods.

| | | Nash Coefficient | | | |
|---|---|---|---|---|---|
| Method | Number of simulations | >0.7 | >0.71 | >0.72 | >0.73 |
| Monte Carlo | 10000 | 1300 | 1011 | 380 | 18 |
| LHS | 5000 | 776 | 561 | 198 | 7 |
| DACE[a](First Part)mean value | 1900 | 450 | 360 | 148 | 12 |
| DACE(First part)standard deviation | 1900 | 2.5 | 5 | 19 | 6 |
| DACE (Second Part)mean | 400 | 144 | 118.5 | 48.8 | 12.6 |
| DACE(Second part)standard deviation | 400 | 2.6 | 4.8 | 10.5 | 12.9 |

[a] value with polynomial fit function

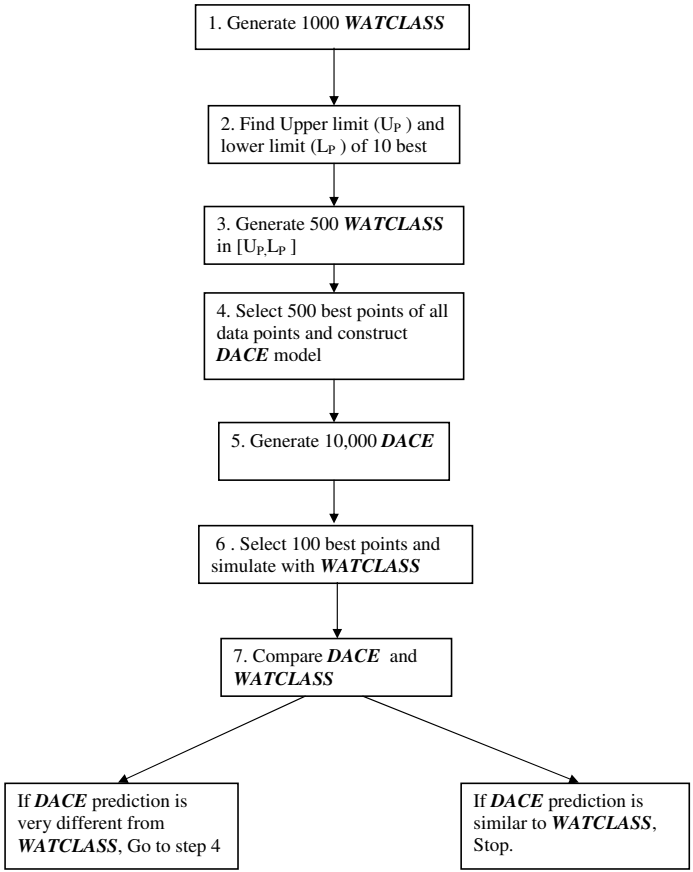


**Fig. 1.** DACE Algorithm.