World Academy of Science, Engineering and Technology
International Journal of Environmental and Ecological Engineering
Vol:2, No:2, 2008

Using Radial Basis Function Neural Networks to Calibrate Water Quality Model

Lihui Ma, Kunlun Xin, and Suiqing Liu

***Abstract*—**Modern managements of water distribution system (WDS) need water quality models that are able to accurately predict the dynamics of water quality variations within the distribution system environment. Before water quality models can be applied to solve system problems, they should be calibrated. Although former researchers use GA solver to calibrate relative parameters, it is difficult to apply on the large-scale or medium-scale real system for long computational time. In this paper a new method is designed which combines both macro and detailed model to optimize the water quality parameters. This new combinational algorithm uses radial basis function (RBF) metamodeling as a surrogate to be optimized for the purpose of decreasing the times of time-consuming water quality simulation and can realize rapidly the calibration of pipe wall reaction coefficients of chlorine model of large-scaled WDS. After two cases study this method is testified to be more efficient and promising, and deserve to generalize in the future.

***Keywords*—**Metamodeling, model calibration, radial basis function, water distribution system, water quality model.

I. Introduction

THE pursuit of a safe drinking water has long been and remains a major concern of public health officials and water-treatment operators since the recognition of waterborne disease by the end of the 19th century. Increased attention has been directed toward protecting the nation's health in addition to quenching its thirst[1]. Under driving force of complying with increasingly stringent governmental regulations and customer-oriented expectations, modeling water quality in water distribution systems has become a widely accepted tool in support of water supply planning, operations, and research.

Modern managements of WDS need water quality models that are able to accurately predict the dynamics of water quality variations within the distribution system environment. Such models would have possible applications in predicting water quality degradation problems, calibrating system hydraulics, designing water quality sampling programs, optimizing the disinfection process, evaluating the water quality aspects of distribution network and storage-reservoir improvement projects, and assessing alternative operational and control strategies for maintaining and improving water quality in distribution systems[2].

Water quality models need to be calibrated before they can be applied to solve system problems. Reference [3] had list that numerous investigators have emphasized the importance of calibrating a model. Effective water quality model demands accuracy of hydraulic models and its accurate parameters. So the calibration of hydraulic model is the basic of calibration of water quality model. Water distribution model calibration is typically accomplished by adjusting network parameters so that model results match field measurements. It can be categorized into trial and error and optimized methods, manual and automated methods, simple parameters and grouped parameters methods, rough-tuning or macro-calibration and fine-tuning or micro-calibration method based on different classified criterion.

In the past, Reference [4; 5] used genetic algorithm (GA) to calibrate the parameters of hydraulic models. Reference [6] provided Markov chain Monte Carlo calibration algorithm by incorporating spatial correlation into the parameter estimation framework for network model calibration. Reference [7] used artificial neural networks (ANN) to perform the hydraulic model calibration, which obtained pipe's roughness from pressures and flow rates. Reference [8] used fussy logic to capture knowledge from people who manually calibrate those networks, because process could become automatic in this way. Otherwise, calibration of water network model can be integrated with leakage distribution in the process of hydraulic calibration [9]. Reference [10] considered the uncertainties in measurement and estimation and provides a measure of the quality of the calibration. Reference [11] made full use of information from tracer studies as well as information from pressure surveys, which would be helpful to calibration of both hydraulic and water quality model.

The literatures presented above are only the part of calibration of hydraulic model. Though the underlying philosophy of water quality calibration is the same as that of hydraulic calibration, the water quality calibration will cost more computing time than calibration of hydraulic model, even reaching more than a hundred times. Newly, some programs to calibrate water distribution models have been available and have occasionally been used on real water systems, such as the Darwin Calibrator of Haestad Methods [12; 13] using a GA solver to optimize relative parameters, it is difficult to apply on the large-scale or medium-scale real system for long

Lihui Ma is with Beijing Construction Procuring and Contracting Trade Center, Beijing 100083, China (phone: 86-10-62344059; 86-0-13466385845; e-mail: powermlh@163.com).

Kunlun Xin is with College of Environmental Science and Technology, Tongji University, Shanghai 200092, China (phone: 86-21-65985869; e-mail: xkl@tongji.edu.cn).

Suiqing Liu is with College of Environmental Science and Technology, Tongji University, Shanghai 200092, China (phone: 86-21-65985869; e-mail: liusuiqing@mail.tongji.edu.cn).

computation time. We have to seek new method to calibrate the parameters of water quality model.

This paper is organized as follows. Section 2 provides the brief background discussion of this topic preparing for subsequent sections. It is followed by a specified description of new method in section 3. In section 4, a large-scale water distribution network applies this method to optimize the parameter of water quality model. The final section concludes this method and restates the key traps in the optimal process.

II. Background

A. Water Quality Model

Most water quality models make use of one-dimensional advective-reactive transport to predict the changes in constituent concentrations due to transport through a pipe. It can be written as follows.

$$\frac{\partial c}{\partial t} + v\frac{\partial c}{\partial x} = R(c) \tag{1}$$

Where c is the concentration of a constituent; t is time; v is the flow velocity; x is the distance and R represents the constituent reaction relationship. Water quality model used in this paper is based upon a parcel tracking algorithm. It tracks the change in water quality of discrete parcels of water as they move along pipes and mixes together at junctions between fixed-length time steps. In order to do this it needs to know the rate at which the substance reacts and how this rate might depend on substance concentration. Reactions can occur both within the bulk flow and with material along the pipe wall.

Bulk flow reactions are the reactions that occur in the main flow stream of a pipe or in a storage tank, unaffected by any processes that might involve the pipe wall. A water quality model simulates these reactions using n-th order kinetics, where the instantaneous rate of reaction (R in unit of mass/volume/time) is assumed to be concentration-dependent, given as:

$$R(c) = \mathrm{K_b} c^n \tag{2}$$

Where K_b is a bulk rate coefficient; c is reactant concentration (mass/volume) and n is a reaction order. K_b has units of concentration raised to the $(1-n)$ power divided by time. It is positive for growth reactions and negative for decay reactions. It also considers reactions where a limiting concentration exists on the ultimate growth or loss of the substance. In this case the rate expression for a growth reaction becomes:

$$R(c) = \mathrm{K_b}(c_L - c)c^{(n-1)} \tag{3}$$

Where cL is the limiting concentration. Thus, there are three parameters (Kb, c_L, and n) that are used to characterize bulk reaction rates. Different values of these parameters lead to different kinetic models. Bottle test is recommended for determining the bulk reaction coefficient such as chlorine decay factor. It provides a good baseline value and reference for constructing a water quality model[12; 14; 15].

In addition to bulk flow reactions, constituent reactions occur with material on or near the pipe wall. The rate of this reaction is dependent on the concentration in the bulk flow and pipe wall conditions, given as:

$$R(c) = \frac{A}{V}\mathrm{K_w} c^n \tag{4}$$

Where K_w is a wall reaction rate coefficient, (A/V) is the surface area per unit volume within a pipe. It converts the mass reacting per unit of wall area to a per unit volume basis. n is the wall reaction order taking value of either 0 or 1, so that the unit of K_w is either mass/area/time or length/time. Both K_w and n are site specific and need to be calibrated for water distribution pipes.

Water quality model with combined bulk and wall decay gives better results than a single decay coefficient. Reliability of wall demand Kw estimation results in very low model error [14]. Combined bulk and wall decay is adopted in this paper.

B. Metamodeling

Metamodeling research has been published since 1970, it has been a major research field during the last decade [16; 17]. The basic idea of metamodeling is to construct an approximate model using function values at some sampling points, which are typically determined using experimental design methods [18]. The main purpose of metamodeling is to reduce the cost, time, and amount of effort required during a simulation analysis. It is usually a supplementary model that can be alternatively used to interpret a more detailed model. The goals of metamodeling cover understanding, prediction, optimization, and verification and validation [19]. Metamodels are constructed in three stages, i.e. estimation, analysis and validation. Its process can involve both qualitative and quantitative factors [20].

The most popular metamodeling approach in simulation involves the use of parametric polynomial regression models in response surface methods [16], which is formulated as follow:

$$Y = \sum_{j=1}^{m}\sum_{k=1}^{q} \beta_{kj} p_k(X_j) \tag{5}$$

Where Y is the output vector; X is the input vector; m is the dimension of input vector; q is a polynomial order. Metamodel can be classified into parametric and nonparametric techniques [21]. The most popular types of metamodels are polynomial regression models, splines, kriging and neural networks [19]. The metamodel might model only a local portion of the simulated system or may encompass the complete simulated system, a global metamodel. Trade of between accuracy and computational expense and between local and global information must be considered when developing a simulation metamodel.

The use of metamodels as surrogates for WDS simulation models has been very rare to date. Reference [22] used ANN in place of the simulation model KYPIPE to calibrate pipe roughnesses. Roughness coefficients were optimized using a GA linked to the ANN. This metamodeling example in WDS optimization ran approximately at twice the speed of the hydraulic simulation model, suggesting that the ANN metamodeling technique has potential for increasing the

computational efficiency of WDS optimization. ANNs are used as a surrogate model to optimize the drinking water distribution [23]. The neural networks may be used to obtain precalibrations or guides for a manual calibration, but they are insufficient when used as unique calibration tools [5].

C. Radial Basis Functions

Neural networks have seen an explosion of interest over the last few years, and are being successfully applied across an extraordinary range of problem domains, in areas as diverse as finance, medicine, engineering, geology and physics. Neural networks are very sophisticated modeling techniques, capable of modeling extremely complex functions. In particular, neural networks are non-linear. It also keeps in check the curse of dimensionality problem which bedevils attempts to model non-linear functions with large numbers of variables. Neural networks can accommodate a combination of continuous variables and discrete numeric variables. Additionally, most neural network paradigms are global models, so a single neural network could be developed to model the entire simulation response surface. This differs from polynomial regression metamodeling, where the regression surface is fitted to a locality, i.e. a subset of the response surface.

Radial basis functions (RBF) were originally developed by Hardy to fit irregular topographic contours of geographical data [21; 24]. RBF networks enjoy the best approximation property among all feed-forward networks and which have produced excellent fits to arbitrary contours of both deterministic and stochastic response functions[25]. Radial basis function networks have an input layer, a hidden layer of radial units and an output layer of linear units. The RBF metamodel is based on radial basis functions using cones (circular hyperboloid), and it is mathematically represented as follows:

$$f(x) = \sum_{i=1}^{n} w_i \phi(\|x - x_i\|) \tag{6}$$

Where n is the number of sampling points, x is the vector of input variables, x_i is the center of basis function ϕ, $\|\bullet\|$ is any l_p norm (typically is Euclidean norm, this kind of norm is used in this study) and w_i is the unknown weighting coefficient. Therefore, an RBF is actually a linear combination of n basis functions with weighted coefficients. Some of the most commonly used basis functions include:

Thin-plate spline:

$\phi(r) = r^2 \log(cr^2)$, $0 < c \le 1$;

Gaussian:

$\phi(r) = e^{-cr^2}$, $0 < c \le 1$;

Multiquadric:

$\phi(r) = \sqrt{r^2 + c^2}$, $0 < c \le 1$;

Inverse multiquadric:

$\phi(r) = \dfrac{1}{r^2 + c^2}$, $0 < c \le 1$.

RBF can be expressed as matrix format:

$$\boldsymbol{f} = \boldsymbol{A}\boldsymbol{\lambda} \tag{7}$$

Where $f = [f(x1), f(x2), \ldots, f(xm)]^T$, $A_{ij} = \phi(\|x_i - x_j\|)$ $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$

The coefficient vector $\boldsymbol{\lambda}$ is obtained by solving Eqs (7). An RBF using the aforementioned highly nonlinear functions does not work well for linear responses. To solve this problem, we can augment an RBF by including a polynomial function such that.

$$f(x) = \sum_{i=1}^{n} w_i \phi(\|x - x_i\|) + \sum_{j=1}^{m} c_j p_j(x) \tag{8}$$

Where m is the total number of terms in the polynomial, and c_j ($j = 1, 2, \ldots, m$) is the corresponding coefficient.

The advantage of RBF is found to be the best for overall performance on accuracy, robustness, problem types, sample size, efficiency, and simplicity compared to response surface method (RSM), kriging method (KM), and multivariate adaptive regression splines (MARS), based on evaluations of the coefficient of multiple determination (R_2), relative average absolute error (RAAE), and relative maximum absolute error (RMAE) [26].

One of disadvantages of RBF is more expensive than RSM, because it uses a series of computationally expensive functions for a single model; therefore, it is less efficient in performing function evaluations. This drawback becomes apparent when solving multi-objective design optimization problems in which millions sometimes even billions of solutions need to be found in order to develop the Pareto Frontier. Another disadvantage of using RBF is that model fitness cannot be checked using ANOVA, because by definition an RBF passes exactly through all the design points [18].

III. Methodology

Water quality modeling of WDS is a time-consuming task, which has to solve the hydraulic equation firstly, i.e. the equations of continuity and energy, for transient analysis the equations of momentum are necessary. The water quality computing step is shorter than hydraulic one and the number of computing times will be more than hydraulic one. Additionally water quality analysis must take extended-period simulation, or it will be no-good for long distance pipelines as the constituent can't reach the relative node. From my previous experiences, a medium-scale network with 6000 nodes and 6000 pipes will cost near one minute time to make a 24-h simulation (15 min/step, 96 steps in all; normal computer configuration: 512M / 1.7G). Optimization algorithms may require a few hundred to several thousands of model simulations to converge to a unique set of parameters. For example, an optimization using GA with 30 populations and 100 generations needs 3000 min (50 h). In a word, water quality simulation in each iterative program of specified optimization belongs to costly functions that is computationally challenging.

Because of the enormous computational cost involved, an analyst is typically willing to perform only a small number of function evaluations when optimizing such costly functions. Our goal, then, is to develop global optimization algorithms

that produce reasonably good solutions with a very limited number of function evaluations. A more practical type of optimization method for computationally expensive function is one that is based on a metamodel presented as before (also known as response surface model or surrogate model). The purpose of the metamodel is to serve as an inexpensive approximation to the costly evaluation function that can help identify promising points for costly function evaluation.

Reference [27] presented a new framework, CORS (Constrained Optimization using Response Surface), for solving costly evaluation function optimization. The CORS method is a flexible framework and can be used to different fields. But it should be adjusted and reinforced according to the characters of applied objects.

Based on the framework of CORS method and the characters of water quality model of WDS, a new combinational algorithm, CORS-RBF-GA, is designed which combines the CORS framework, RBF network and GA to calibrate the parameters of water quality model. The flowchart of CORS is the same as the left side of fig 1 (step1-6), the RBF network is the water quality metamodel, GA is the optimizing method of metamodel. In addition, the algorithm of CORS-RBF-GA is also reinforce by using the generability checking and cluster analysis. The algorithm of CORS-RBF-GA is presented as follows.

Step 1 is to select initial evaluated points. Set and select a finite initial set of points $S = \{x_1, \cdots, x_k\}$ which are prepared for water quality simulation and fit the RBF network. S is a vector when only one parameter is calibrated, and a matrix with m row (the number of calibrated parameters) and k column (the number of points) when more than one parameter. The method of selecting evaluated points includes factorial design (grid), Latin hypercube and orthogonal arrays. The most known technique is the factorial design which requires fitness function evaluations at K_N design points, where N is the number of design variables and K the number of levels defined between the lower and upper bounds of each variable. Usually, K is 2 or 3, depending on the effects we want to model[28]. After experimental comparison, we found that the number of experimental points should be beyond 100 for more than two parameters that should be calibrated.

Step 2 is to do water quality simulation. Data from water quality simulation can be used directly or changed to other form, which typically are transformed into discrepancy between calculated data and measured data. These data will play a role as output of RBF network, while the evaluated points as input. Basis function employs the Gaussian function because MATLAB is based on this one. In this paper the predicted error, *f(x)*, is adopted to the output of RBF network and defined as follow in order to calibrated the parameters according to minimize the discrepancy. The data of boundary condition of WDS are manually treated firstly with the goal of focusing on this method singly.

$$f_i(x) = g_i(x) - m_i \quad i = 1, \cdots n \tag{9}$$

$$f(x) = \sum_{i=1}^{n} |f_i(x)| \tag{10}$$

Where $g_i(x)$ is the calculated value of No. i monitoring node; m_i is the measured value of No. i monitoring node; n is the total number of monitoring nodes. Note that the calculated values of monitoring nodes are zero in the early period of extended period simulation of WDS until the disinfectant pass through these nodes. So *f(x)* must start to accumulate from the time that is above water age of every node.

Step 3 is to save evaluated points and the value of accumulated error. Evaluated points are saved into $S = \{x_1, \cdots, x_k\}$ and the values of error into $E = \{f(x_1), \cdots, f(x_k)\}$. When the following optimization produces new data, S adds one point after the last point and E adds the evaluated value with respect to new added point. Note that To prevent oscillation in the RBF interpolation, the large error values should be replaced by the median of all available error function values [29].

Step 5 is to fit or update RBF. The RBF is an approximated metamodel as a surrogate of WDS, which is key part of the whole algorithm. Based on data set S and E, the former is input and the latter is output, new RBF network will be trained and updated in each iteration. For practical purposes, it should be intuitively clear that the rate of convergence is somehow dependent on how well the RBF model approximates the water quality model and also on how well we solve the optimization problem on the RBF network. In first time we should optimize the spread parameter of RBF which is so important that directly influence the degree of approximation. Quasi-Newton Methods or GA [30] can be used to solve this optimization. After the optimized spread parameter has been found a new RBF network should be train again.

Step 6 can use different optimization method, such as nonlinear programming or GA, to find minimized point in surface of RBF function which will be the next evaluated point. In former research we found that nonlinear programming is easy to make matrix is close to singular or badly scaled, it will result in optimization terminated early. GA is recommended as a universal optimization method. Objective function and constraint condition are presented as follows.

$$\text{Minimize } f'(x) \tag{11}$$

Subject to:

$$\left\| x - x_j \right\| \geq \beta\Delta \qquad j = 1, \cdots, k+i-1, \; x_j \in S \quad 0 \leq \beta \leq 1 \tag{12}$$

$$0.5 \leq \frac{x(i)}{x(k)} \leq 1.5 \qquad i, k = 1, \cdots, m \tag{13}$$

Where $\Delta_i = \max_{\tilde{x} \in D} \min_{1 \leq j \leq k+i-1} \| \tilde{x} - x_j \|$, $\bar{x}$ is the cover points defined beforehand which covers the whole hypercube domain. Function *f'* is the RBF. *x(i)* and *x(k)* are the elements of the vector of *x*, m is the dimension of *x*.

The purpose of the constraint is to drive the algorithm towards unexplored regions and prevent the algorithm from prematurely converging to some possibly undesirable points.

To be able to perform both local and global search in this scheme, we use β to control the distance range from high values for global search to low values to local search. Detailed proved process can be referred to [27]. Inequality constraints (13) restrict the grouping parameters (wall reaction coefficients), assuming that there are no great discrepancies among the parameters.

Step 4 and 7 are identification of convergence condition and obtainment of optimized values via optimization of RBF metamodel. According to following case study we found that many combinations of wall reaction parameters can have the near same influence on predictions of monitoring points in WDS. So we have to select the point as the final result that has the best generalization ability among the options. Step 8 utilizing the data from step 8 compare the generalization ability of given new evaluated points from RBF optimization, which computes individually the total predicted error of given period of water quality behaviors. Step 9 is the end of methodology.

This method is realized in MATLAB programming language, and RBF network and GA call functions of corresponding toolbox. Water quality modeling is operated using EPANET programmer's toolkit, the EPANET2.dll is incorporated into main program.

IV. Case Study

Example network is shown in Fig. 2 and used to apply the new method. It is a water utility in the southern of China. This network has 6909 nodes, no tank, and 7452 links, totaling 513 km of pipe. It is supplied from the four treatment plants. Hydraulic model is calibrated manually and automatically with GA. The extended period is set 72-h and time step is 15 min. Hydraulic demands and boundary conditions are set according to real operation data after manual treatment. This network includes six monitoring points which monitor residual chlorine, turbidity, pH and temperature, and are red quadrate in Fig. 2. This section will study two cases. Case 1 is to get whole response surface intuitively based on different bulk and wall reaction coefficient. Case 2 is real calibration of example network which bulk reaction parameter is obtained from bottle test as a baseline and wall reaction coefficients are divided into six groups.

A. Case 1

This case does not set the bulk and wall reaction coefficient does not make grouping, these are both arguments for the error surface. Limiting the number of parameters is to present the surface intuitively, for the space of more than three dimensions can't express simply using figure. The two parameters both range from -6 to 0, but the units are different. Unit of bulk reaction coefficient is 1/d and reaction order is 1st-order. Unit of wall reaction coefficient is meter/d and reaction order is 1st-order. The figures are showed as follows. The 144 data is obtained via water quality modeling and error calculation under the domain specified by bulk and wall reaction coefficients. All data are presented in Fig. 3 (a), for looking carefully we remove some edge data in Fig. 3 (b). The area of minimized total error is a line style and it means that different combination of wall and bulk reaction coefficient can obtain near same total error from Fig. 3 (c). In practical case, bottle test provides a good baseline value and reference for bulk reaction coefficient of a water quality model, and then we can find the best parameters for model.

B. Case 2

The selection of values for the candidate policies is problem dependent. This paper uses a uniform discretized lattice over the allowable ranges of the six input variables; however more sophisticated methods such as k-p designs, Latin hypercube sampling or Bayesian techniques may well improve neural network metamodels[31].

To enhance the accuracy of calibration, in this case we use cluster analysis method, hierarchical cluster analysis, to divide the whole pipes into six groups. Hierarchical cluster analysis attempts to identify relatively homogeneous groups of wall reaction coefficient based on selected characteristics, which include pipe diameters, pipe materials, construction time, flow rates, flow velocity, location (quantified by means of average coordinates). As SPSS provides a powerful statistical analysis, SPSS 14.0 is used to operate cluster analysis.

The initial evaluated points, total 128 ($2*2^6$) points, are obtained using simple grid method. Using this method presented in the paper result can be obtained soon and the process is shown in Fig. 4. Fig. 4(a) presents the whole process and the training data fluctuate acutely from No.1 to 128 reflects partly that the RBF network represents the whole approximated space. We select 100 data to train the RBF network and 28 data to optimize the spread parameter of RBF network. The following process shows the real optimization process in fig 4(b) that surges in the early stage and smoothes or coverage to the minimum in the latter.

In order to test the influences of different environment of optimization, we have compared four kinds of optimization and the results are shown in Fig. 5. First series is only optimization of two parameters that are bulk and wall reaction coefficients the same as case 1. This smooth line presents a few numbers of combinations of the two parameters, but the performance function, total error, is a little bit higher than other series. Second series represents that we fit the RBF network using given the spread parameter other than using optimal value. This kind of optimization fluctuates in the early stage because the RBF network needs longer time to enhance the fitting ability as increase of data. Third series represents the same data as fig 4 (b) and the optimization subjects to the inequality constraints (13) (six wall reaction coefficients, spread parameter optimization). The RBF network becomes stability is shorter because of the contribution of the spread parameter optimization. In fourth series, the inequality constraints (13) are not used. In the early stage the line surges a short time, then becomes gentle incline. But there are two unwilling juts in the range from No. 22 to 28. It testifies that some margin inputs, for example [-0.5, -6.0, -6.0, -3.5, -4.2, -6.0], will bring more total error because short of initial fitting points can not fit the RBF

surface more accurate in margin region while these can do well in the middle of network.

One of comparisons of predicted values and measured values is showed in Fig. 6. From the figure you can find the predicted line is smoother than the measured one, there are a big difference in the early time of every day from 1 to 9 o'clock. According to simply research we find the possible reason is the influence of flow velocity the same as other research findings [32]. The contrastive figure of one monitoring point is presented in Fig. 7. Another possible reason may be that the pipe water is still in the leading out thin branch pipe at the water quality monitering station, which will be no refilling chlorine into this pipe at this pierod. While the simulation concentration is upstream pipeline relative to this monitoring location in WDS (Fig. 8), this difference of position contribute to the simulating error. In order to solve this problem the deeper research should be executed in the future.

Comparison of computational times is listed in Table I, which compare EPANET linked to GA and the new method presented in this paper. The total computing time of RBF metamodeling is only 10.3 % of the one of EPANET linked to GA which is adopted normally. Every water quality simulation needs 3 min, and the sampling data are produced after water quality simulations based on every initial evaluated point. Thus sampling computing time should be 128×3min, namely 6 h and 24 min. Optimization time of the former method is calculated by assuming GA parameters adopted normal values, 100 generations and 20 populations. Generalization ability checking is carried out 12 d long simulation and finds the optimal value in the end.

TABLE I
COMPARISON OF COMPUTATIONAL TIMES

Methods	Sampling Data (h : min)	Training Spread Parameter	Optimization (h : min)	Generalization Checking (h : min)	Total(h : min)
GA			100:00 (100×20×3 min)		100:00
RBF metamodeling	6:24(128×3min)	1min	2:35 (50×3min + 50×6s)	1:30 (10×3×3min)	10:30

V. CONCLUSION

This paper presents a new method using RBF metamodeling as a surrogate to be optimized for the purpose of decreasing the times of time-consuming water quality simulation. The CORS-RBF-GA algorithm succeeds in calibrating the parameters of water quality model more efficiently than EPANET linked to GA. In order to assure the quality of optimization and avoid traps, key points should be restated and obeyed:

1. The sampling data should be adequate, in general above 100.

2. The spread parameter should be optimized; this will enhance the fitting level.

3. The constraints between two parameters should add to the constraint conditions of main optimization.

4. Generalization ability checking can decrease the influence of local target value, namely measured value.

Although the new method can solve the problem of computational time, there also are some works to be researched more deeply in the future, such as considering data uncertainty the same as former hydraulic calibration in the water quality calibration and quantification of other influence factors to bulk and wall reaction coefficients.

REFERENCES

[1] P. Boulos, T. Altman, P. Jarriage and F. Collevati. "Discrete Simulation Approach for Network Water Quality Models." Journal of Water Resources Planning And Management, vol 121(1), pp. 49-60, 1995.

[2] L. A. Rossman, P. F. Boulos and T. Altman. "Discrete Volume-Element Method for Network Water-Quality Models." Water Resources Planning and Management, vol 119(5), pp. 505-517, 1993.

[3] T. Walski, Z. Wu and W. Hartell. "Performance of Automated Calibration for Water Distribution Systems". Proceedings Of The 2004 World Water and Environmental Resources Congress, Salt Lake City, UT, USA,2004, ASCE.

[4] S. Lingireddy and L. E. Ormsbee. "Optimal Network Calibration Model Based on Genetic Algorithms". Proceedings of the 26th Annual Water Resources Planning and Management Conference, Tempe, Arizona, USA, 1999, ASCE.

[5] R. Gómez, D. Salas and J. Saldarriaga. "Several Configurations and Aggregation Levels in Neural Networks for Water Distribution Networks Calibration". Proceedings Of The 2004 World Water and Environmental Resources Congress, Salt Lake City, UT, USA,2004, ASCE.

[6] D. L. Boccelli and J. G. Uber. "Incorporating Spatial Correlation in a Markov Chain Monte Carlo Approach for Network Model Calibration". Proceedings of the 2005 World Water and Environmental Resources Congress, Anchorage, Alaska, USA,2005, ASCE.

[7] B. G. Van and B. Waanders. "Application of Optimization Methods to Calibration of Water Distribution Systems". Proceedings of The 2004 World Water and Environmental Resources Congress, Salt Lake City, UT, USA,2004, ASCE.

[8] D. Salas, R. Gómez and J. Saldarriaga. "Water Distribution Network Calibration with Fuzzy Logic". Proceedings of the 2004 World Water and Environmental Resources Congress, Salt Lake City, UT, USA,2004, ASCE.

[9] L. Ainola, T. Koppel, K. Tiiter and A. Vassiljev. "Water Network Model Calibration Based on Grouping Pipes with Similar Leakage and Roughness Estimates". 2000 Joint Conference on Water Resources Engineering and Water Resources Planning & Management, Minneapolis, Minnesota, USA,2000, ASCE ; L. E. Ormsbee and S. Lingireddy. "Calibrating Hydraulic Network Models " Amerian Water Works Association, vol 89 (2), pp. 42-50, 1997; M. Greco and G. D. Giudice. "New Approach to Water Distribution Network Calibration." Journal of Hydraulic Engineering, vol 125(8), pp. 849-854, 1999; K. E. Lansey, W. El-Shorbagy, I. Ahmed, J. Araujo and C. T. Haan. "Calibration Assessment and Data Collection for Water Distribution Networks." Journal of Hydraulic Engineering, vol 127(4), pp. 270-279, 2001.

[10] C. A. Bush and J. G. Uber. "Sampling Design and Methods for Water Distribution Model Calibration." Journal of Water Resources Planning and Management, vol 124(6), pp. 334-345, 1998; D. Y. Sumer, D. A. Sier, J. Roach, S. Schladweiler, U. Yenal and K. Lansey. "Impact of Data Collection and Calibration of Water Distribution Models on Model-Based Decisions". Proceedings Of The 2004 World Water and Environmental Resources Congress, Salt Lake City, UT, USA,2004, ASCE.

[11] P. J. Skipworth, J. Machell and A. Saul. "Explicit Consideration of Spatial Flowrate Distribution in Network Model Calibration". Proceedings of the World Water and Environmental Resources Congress, Orlando, Florida, USA,2001, ASCE. Z. Y. Wu. "Optimal Calibration Method for Water Distribution Water Quality Model." Environmental Science and Health Part A, vol 41, pp. 1-16, 2006.

[12] Z. Y. Wu. "Water Quality Model Calibration by Means of Fast Messy Genetic Algorithm". Proceedings of the 2005 World Water and

Environmental Resources Congress, Anchorage, Alaska, USA,2005, ASCE.
[13] S. H. Maier, R. S. Powell and C. A. Woodwar. "Calibration and Comparison of Chlorine Decay Models for a Test Water Distribution System." Water Research, vol 34(8), pp. 2301-2309, 2000.
[14] J. J. Vasconcelos, L. A. Rossman, W. M. Grayman, P. F. Boulos and R. M. Clark. "Kinetics of Chlorine Decay." American Water Works Association, vol 89(7), pp. 54-65, 1997.
[15] R. A. Kilmera, A. E. Smithb and L. J. Shuman. "Computing Confidence Intervals for Stochastic Simulation Using Neural Network Metamodels." Computers & Industrial Engineering, vol 36, pp. 391-407, 1999.
[16] D. J. Fonseca, D. O. Navaresse and G. P. Moynihan. "Simulation Metamodeling through Artificial Neural Networks." Engineering Applications of Artificial Intelligence, vol 16, pp. 177-183, 2003.
[17] H. Fang, M. Rais-Rohani, Z. Liu and M. F. Horstemeyer. "A Comparative Study of Metamodeling Methods for Multiobjective Crashworthiness Optimization." Computers and Structures, vol 83 pp. 2121-2136, 2005.
[18] J. P. C. Kleijnen and R. G. Sargent. "A Methodology for Fitting and Validating Metamodels in Simulation." European Journal of Operational Research, vol 120, pp. 14-29, 2000.
[19] S. Tunali and I. Batmaz. "A Metamodeling Methodology Involving Both Qualitative and Quantitative Input Factors." European Journal of Operational Research, vol 150, pp. 437-450, 2003.
[20] M. F. Hussain, R. R. Barton and S. B. Joshi. "Metamodeling: Radial Basis Functions, Versus Polynomials." European Journal of Operational Research, vol 138, pp. 142-154, 2002.
[21] S. Lingireddy and L. E. Ormsbee. "Neural Networks in Optimal Calibration of Water Distribution Systems." Artificial Neural Networks for Civil Engineers: Advanced Features and Applications, vol, pp. 53-76, 1998.
[22] D. R. Broad, G. C. Dandy and H. R. Maier. "A Metamodeling Approach to Water Distribution System Optimization". Proceedings Of The 2004 World Water and Environmental Resources Congress, Salt Lake City, UT, USA,2004, ASCE; D. R. Broad, G. C. Dandy and H. R. Maier. "Water Distribution System Optimization Using Metamodels." Journal of Water Resources Planning and Management, vol 131(3) pp. 172-180, 2005a; D. R. Broad, H. R. Maier, G. C. Dandy and J. B. Nixon4. "Estimating Risk Measures for Water Distribution Systems Using Metamodels". Proceedings of the 2005 World Water and Environmental Resources Congress, Anchorage, Alaska, USA,2005b, ASCE.
[23] S. Haykin. Neural Networks: A Comprehensive Foundation (2nd Edition), Prentice-Hall, Inc,1999.
[24] A. Shahsavand and A. Ahmadpour. "Application of Optimal Rbf Neural Networks for Optimization and Characterization of Porous Materials." Computers and Chemical Engineering, vol 29 pp. 2134-2143, 2005.
[25] R. Jin, W. Chen and T. W. Simpson. "Comparative Studies of Metamodelling Techniques under Multiple Modeling Criteria." Struct Multidiscip Optim, vol 23, pp. 1-13, 2001.
[26] R. G. Regis and C. A. Shoemaker. "Constrained Global Optimization of Expensive Black Box Functions Using Radial Basis Functions.Pdf." Journal of Global optimization, vol 31(1), pp. 153-171, 2005.
[27] K. C. Giannakoglou, D. I. Papadimitriou and I. C. Kampolis. "Aerodynamic Shape Design Using Evolutionary Algorithms and New Gradient-Assisted Metamodels." Computer Methods in Applied Mechanics and Engineering, vol 195 pp. 6312-6329, 2006.
[28] H. M. Gutmann. "A Radial Basis Function Method for Global Optimization." Journal of Global Optimization, vol 19(3), pp. 201-227, 2001.
[29] H. Sarimveis, A. Alexandridis, S. Mazarakis and G. Bafas. "A New Algorithm for Developing Dynamic Radial Basis Function Neural Network Models Based on Genetic Algorithms." Computers and Chemical Engineering, vol 28, pp. 209-217, 2004.
[30] R. A. Kilmer, A. E. Smith and L. J. Shuman. "Computing Confidence Intervals for Stochastic Simulation Using Neural Network Metamodels." Computers and Industrial Engineering, vol 36, pp. 391-407, 1999.
[31] R. M. Clark and R. C. Haught. "Characterizing Pipe Wall Demand: Implications for Water Quality Modeling." Water Resources Planning and Management, vol 131(3), pp. 208-217, 2005; R. M. Clark and R. C. Haught. "Characterizing the Pipe Wall Demand for Free Chlorine in Metallic Pipes". Proceedings Of The 2004 World Water and Environmental Resources Congress, Salt Lake City, UT, USA, 2004, ASCE.

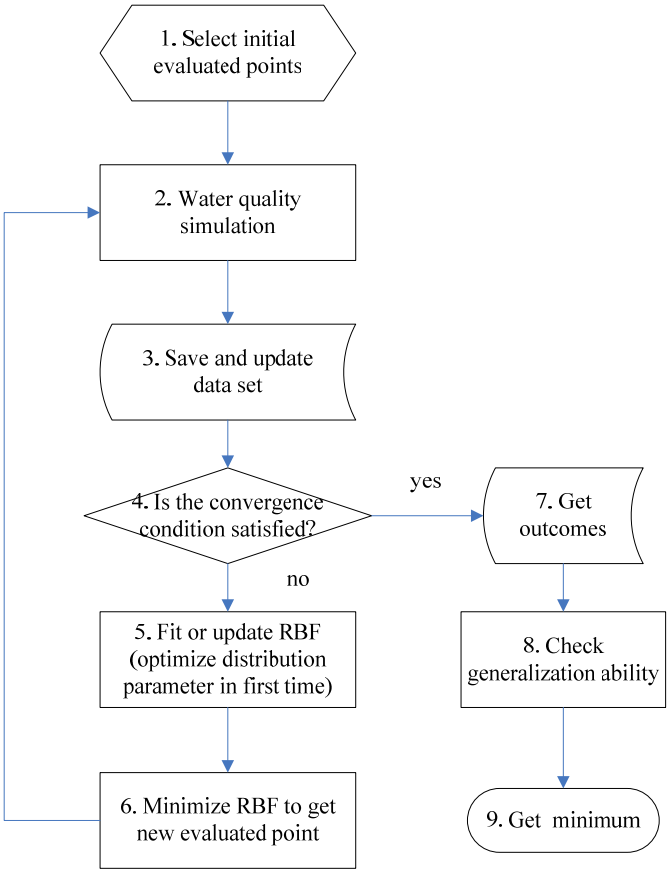


Fig. 1 Flowchart of calibration of water quality model of WDS

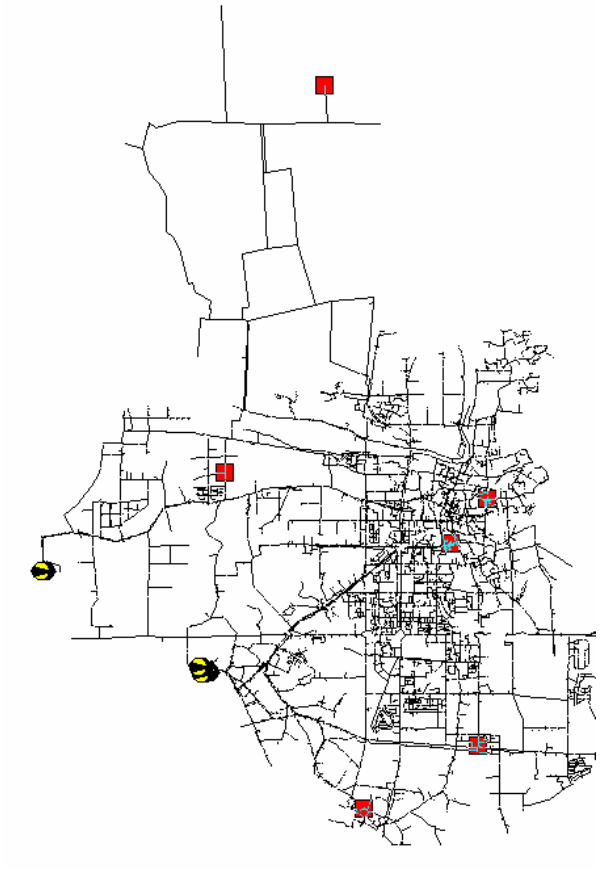
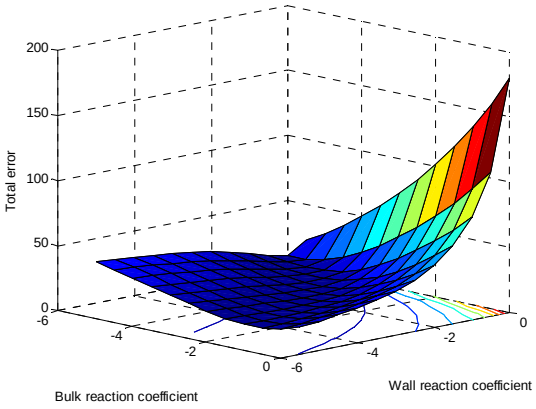
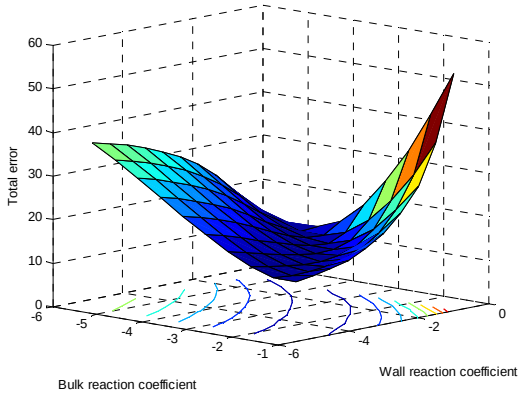


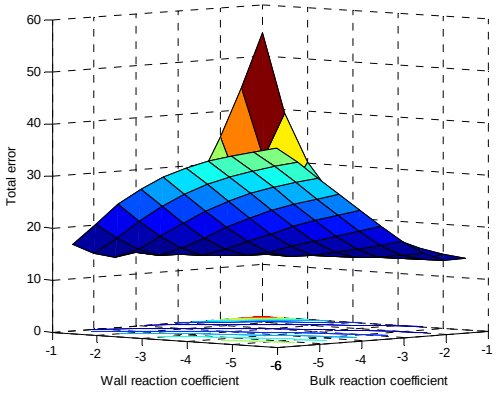
Fig. 2 Example network



(a)

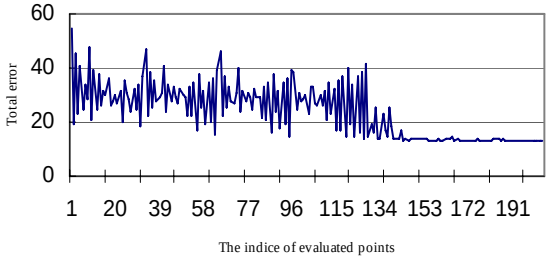


(b)

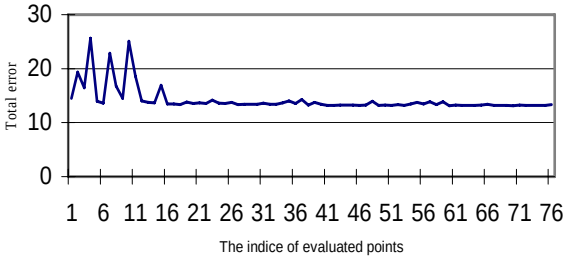


(c)

Fig. 3 Error surface with respect to bulk and wall reaction coefficients (a) All data; (b) Remove edge data; (c) Change view of figure (b)



(a)



(b)

Fig. 4 The process of optimization via RBF approximation (a) The whole process include initial training and optimization; (b) Optimization process

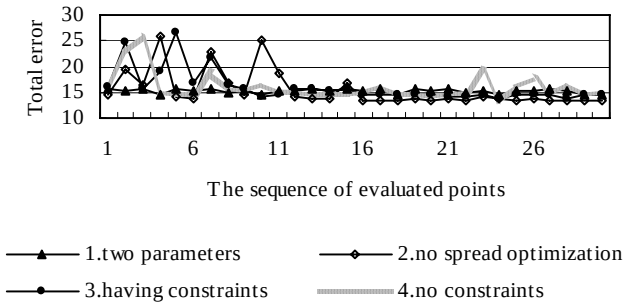


Fig. 5 Different optimization based on different conditions

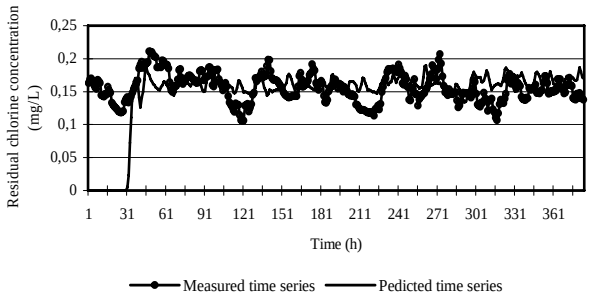


Fig. 6 One of comparisons of predicted values and measured values

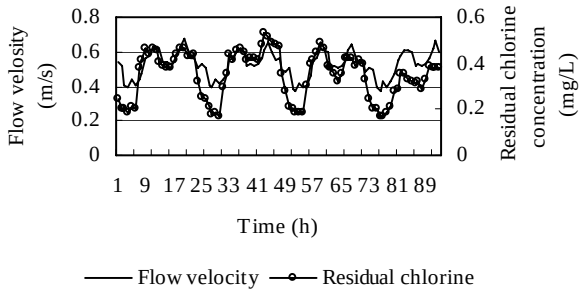


Fig. 7 Contrastive figure of residual chlorine concentration and flow velocity

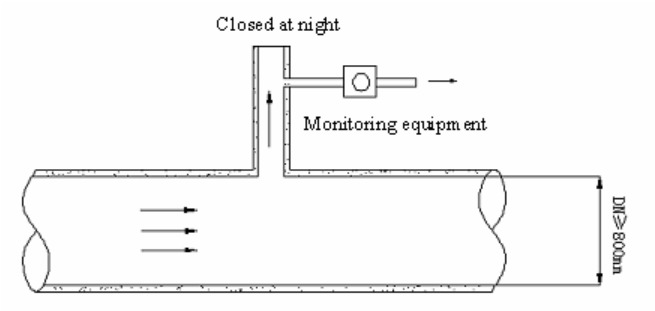


Fig. 8 The schematic plan of water quality monitoring location

Lihui Ma was born at Heibei province, China. Master degree of municipal engineering was awarded at Taiyuan University of Technology in Taiyuan of China in 2004, doctor degree of municipal engineering was awarded at Tongji University in Shanghai of China in 2007. Rresearch interests mainly concerned about the optimization of water distribution network. Dr. Ma has designed above ten water supply and sewage system projects in the past, assisted a few water utilities by developing informationization, and participated in a few national research projects.