

# **Using Genetic Algorithms and Particle Swarm Optimization for Optimal Design and Calibration of Large and Complex Urban Stormwater Management Models**

Misgana K. Muleta , Paul F. Boulous , Chun-Hou Orr , and Jun Je Ro

## ***Abstract***

Computer models are vital for the evaluation and management of urban drainage systems. Usefulness of these models, however, depends on how well they are calibrated. Properly calibrated models can be used to conceive, evaluate and compare various design improvement alternatives. Unfortunately, calibration and design of urban stormwater models, especially with the commonly used trial-and-error approach, are an expensive, time-consuming process and normally represent the most painful step of a modeling exercise. Their success depends mainly on the engineering expertise of the modeler and budget availability. The effort is complicated by the fact that these models normally necessitate the evaluation of a large number of variables and parameters in order to adequately describe the complex relationships that exist between rainfall, runoff, watershed characteristics, and system hydraulics in an urban setting. The trial-and-error evaluation of all calibration and design/improvement options is therefore unlikely to be practically feasible or manageable, and even knowledgeable modelers often fail to obtain good results. In this paper, a rigorous optimal calibration and design methodology is presented, which eliminates the need of the traditional trial-and-error technique. The optimal calibration and design problems are cast as nonlinear optimization problems and solved using genetic algorithms (GA) optimization and particle swarm optimization (PSO). The EPA storm water management model (SWMM5) is employed to perform hydrologic and hydraulic analyses. The optimal calibration model determines the set of calibration parameters that best matches field observations of flow, depth or velocity to accurately mirror actual system performance. The optimal design model determines the set of design parameters that best meets desired system performance requirements at minimum cost. Design parameters can include any combination of pipe slope and size, storage, pumping, and new piping. System performance criteria include explicit constraints on the maximum allowable depth to diameter ratio, minimum and maximum pipe velocities, and maximum head loss for force mains. The proposed optimal calibration and design models are demonstrated by application to an example urban

stormwater collection system. Enhancement of urban drainage system planning, design, operation and management is a principal benefit of the proposed methodology.

## 1. INTRODUCTION

Utilization of advanced and comprehensive urban drainage systems simulation models is the most viable means of evaluating system response to various operational, management and remedial strategies. However, no matter how advanced they may be, models are mathematical representation of the underlying physical processes. As such, they experience some aspect of conceptualization or empiricism, implying that model results may not always represent the reality to the desired level of accuracy. Therefore, it is imperative that computer simulation models pass through a rigorous calibration process before they are employed as a decision making aid in the planning, design, analysis, and management of urban drainage systems. Calibration generally refers to the process of estimating the values of some model parameters so that model predicted results closely match observed behaviors such as flow at one or more locations of the drainage system with the intention of using the estimated optimal parameter values for subsequent modeling. Calibration increases confidence in the model parameter values and predicted results.

Model calibration is a two-stage process. It involves the process of selecting or identifying model parameters that need to be calibrated, a stage referred to as parameter specification, and the process of estimating optimal values of the selected parameters, a stage known as parameter estimation (Sorooshian and Gupta, 1995). Parameter specification is basically the process of reducing the number of calibrable parameters to a manageable size. It is particularly warranted for simulation models with large number of parameters, which is typical for spatially distributed urban drainage models designed to reasonably represent the spatial heterogeneities encountered in topographic, land use, soil, and climate conditions of urban watersheds. Once decision is made regarding what parameters to calibrate, the next logical step of model calibration is parameter estimation. Traditionally, both parameter specification and parameter estimation are conducted using manual (i.e., trial-and-error) approach. Manual calibration of urban drainage models is a tedious, time consuming, expensive, and subjective process. In short, it is very unlikely for manual calibration to bear good results even when conducted by experienced modelers. As a result, calibration of urban drainage network models has generally been neglected or done haphazardly. Automation of these calibration processes could avoid the hurdles of manual procedure making calibration of urban drainage models significantly easier and less prone to errors.

At the parameter specification stage, optimal calibration of urban drainage models should involve the use of a sensitivity analysis model to systematically rank model parameters according to their effect on desired model outputs so that the most influential parameters are taken to the estimation stage. In this paper, we describe a global sensitivity analysis model that employs stepwise regression analysis on rank-transformed input-output data pairs in order to automate the specification stage of model calibration (Muleta and Nicklow, 2005). Latin hypercube sampling is used to generate input data from the assigned parameter distributions and ranges. Estimation of model parameters involves the use of optimization algorithm to determine the best-fit parameter values,

thus making model calibration fast, less subjective and more effective since it extensively searches for possible parameter combinations. Here, we describe application of two nature based global optimization algorithms, genetic algorithms (GA) and particle swarm optimization (PSO), at parameter estimation stage of urban drainage model calibration. The effectiveness of the optimal calibration approach is demonstrated using a typical urban drainage system.

Once properly calibrated, models can be effectively employed for various system management applications. In urban drainage systems, for example, mitigation of adverse impacts such as sewer overflows and the associated pollution problems requires the use of calibrated drainage system models to evaluate effectiveness of alternative remedial measures. For such capital improvement projects, the objective is to minimize costs of improvement alternatives while satisfying regional environmental regulations for stormwater quantity and quality control. The current practice of identifying cost-effective remedial solutions involves the conventional tedious trial-and-error evaluation procedure. However, given the vast number of possible combinations of design improvements, it is unlikely that the manual trial-and-evaluation procedure for screening alternatives would yield the optimal solution. In this paper, the problem of identifying least-cost remedial solution for urban drainage systems is cast as an optimization problem and solved using genetic algorithms and particle swarm optimization. The USEPA storm water management model (SWMM5) is employed to analyze hydrologic, hydraulic, and water quality of the watershed. The effectiveness of the optimal design methodology is illustrated using a sample drainage system.

## **2. THE URBAN DRAINAGE SIMULATION MODEL**

InfoSWMM (Boulos, 2005), the urban drainage simulation model used in this study, utilizes an extended version of the USEPA SWMM5 (Rossman, 2005) to solve urban drainage hydrology, hydraulics and water quality. InfoSWMM can model the entire land phase of the hydrologic cycle (i.e., surface runoff, vadose zone processes, groundwater flow, evaporation, and snow fall and snow melt) as applied to urban stormwater and wastewater collection systems. The model can perform single event or long-term (continuous) rainfall-runoff simulations accounting for spatial and/or temporal variability in climate, soil, land use, and topographic conditions of the urban watershed. Rainfall-runoff simulations could be conducted using several approaches including the non-linear reservoir method, Colorado Urban Hydrograph Procedure (CUHP), Santa Barbara hydrograph method, the Modified Rational Formula, and various synthetic unit hydrograph approaches including the Natural Resources Conservation Service (NRCS) methods, Delmarva, Snyder, Clark, and Espey methods. In addition to runoff quantity, InfoSWMM can simulate runoff quality including buildup and washoff of pollutants from primarily urban watersheds composed of developed and non-developed land uses.

Once runoff quantity and quality are simulated and wastewater loads at receiving junctions are determined, the routing portion of InfoSWMM can transport the flow using either steady, kinematic wave or dynamic wave routing, through a conveyance system of pipes, channels, storage/treatment devices, pumps, and hydraulic regulators such as weirs, orifices, and other outlet types. The model can accurately simulate flow conditions

of any level of complexity including those experiencing backwater effect, flow reversal, and pressurized flow. Because it simultaneously solves the solution for both water levels at nodes and flows in conduits, the model can be applied to any general network layout, even those containing multiple downstream diversions and loops. The model also offers advanced Real-Time Control (RTC) scheme for the operational management of hydraulic structures.

InfoSWMM is a complete decision support system embedding intrinsic modeling and optimization capabilities in GIS. It explicitly integrates ESRI ArcGIS geospatial model with the extended version of SWMM5 and other modules including automated dry weather flow generation and allocation, automated subcatchment delineation and parameter extraction, automatic model calibration, and an optimization capability for screening design alternatives to address every facet of urban drainage infrastructure management. The GIS interface allows seamless communication between the GIS and the various modeling applications. It provides the geospatial platform for developing models, calibrating them, screening design alternatives, and displaying results. The resulting decision support system effortlessly reads GIS datasets, extracts necessary modeling information, and automatically constructs, loads, designs, calibrates, analyzes and optimizes a representative urban drainage models considering hydrologic, hydraulic, and water quality performance requirements. It also makes it easy to run, simulate and compare various modeling scenarios, identify system deficiencies, and determine cost-effective physical and operational improvements to achieve optimum performance and regulatory compliance. These combined capabilities provide a consistent environment to plan, design, and operate safe and reliable urban drainage systems.

### **3. THE OPTIMIZATION ALGORITHMS**

Usage of proper search mechanism is one of the decisive factors in successful automatic model calibration. Needless to say, for urban drainage systems the relationship between model parameters and output variables is too complicated to derive simple (i.e. smooth and unimodal) mathematical functions and, hence, their derivative information. This implies that gradient based nonlinear optimization methods have little to offer for complex problems like calibration of distributed urban drainage models. As an alternative, heuristic search algorithms, such as GA and PSO, that require no derivative information about the objective function or constraints, have been practically proven to work well on nonlinear, nonconvex, and multimodal problems (Schwefel, 1995; Kennedy and Eberhart, 1995), and are utilized in this study.

GA (Holland, 1975) are designed to perform optimization by emulating the evolution process. They operate based on the preferential survival and reproduction of the fittest member of a population, where offspring inherit genetic information from parents while experiencing occasional mutation of genes. Though not ultimately guaranteed to locate global optima (Cieniawski et al., 1995), GA search a wide portion of the solution space and, thus, possess a better capability of locating optimal solutions. In fact, the majority of GA literature consistently demonstrates an ability to identify global or very near global optima for a range of complicated problems (Nicklow, 2000). GA use randomized and localized operators as opposed to deterministic rules. Each individual in

the population is represented by either a string or a set of real numbers encoding one possible solution. The performance of each individual in the population is measured by its fitness (goodness), which quantifies the degree of optimality of the solution. Based on their fitness values, individuals are selected for reproduction of the next generation. Each new generation maintains its original size. The selected individuals reproduce their offspring by mimicking gene operations of crossover and mutation. After a number of generations, the population is expected to evolve artificially, and the optimal or near optimal solution is ultimately reached.

Like GA, PSO is a population-based stochastic search algorithm that utilizes a population of candidate solutions in which individuals change their position with time to evolve to an optimal or near-optimal solution. PSO is inspired by the paradigm of birds flocking in search for food (Kennedy and Eberhart, 1995). The algorithm mimics the social behavior of a flock of birds in order to guide swarms of particles (i.e., solution candidates) towards the most promising regions of the multidimensional search space. It is grounded on the principle of exchanging information among the swarm particles (i.e., ensuring cooperation among alternative solutions) then competing the solution candidates to arrive at the optimal solution. This way, individuals interact with one another and benefit from the very best performance of any member of the entire population while learning from their own experience as well, and gradually the population members move into better regions of the problem space. Unlike GA, PSO has no evolution operators such as crossover and mutation. PSO has been applied to several science and engineering optimization problems and has been proven to be an efficient optimization algorithm (Kennedy and Eberhart, 1998; Jung and Karney, 2005). Compared to GA, the main advantages of PSO are that it requires fewer parameters to be adjusted and exhibits faster convergence characteristics.

#### **4. THE CALIBRATION MODEL**

The automatic calibration module comprises various approaches and algorithms including a grouping mechanism that lumps some model parameters based on the homogeneity of corresponding model components to reduce the number of calibration parameters, a global sensitivity analysis model to identify the most influential model parameters, two nature based advanced optimization algorithms (i.e., GAs and PSO) to determine optimal values of the calibration parameters so that model simulations closely match field observations, and a comprehensive network analysis model to uncover implications of various parameter sets on the desired model outputs considering hydrology and hydraulics of the drainage system. The calibration model considers flow, velocity, and or flow depth at one or multiple sites. In addition, depending on the objective of the calibration exercise, the modeler may choose to calibrate based on the entire time series of observed data, or only peak values (i.e., observations exceeding a specified value which is an ideal option if accurate simulation of only peak values is desired), or only low values if the intent is, for example, accurate simulation of low flows. The automatic calibration module offers a range of goodness-of-fit evaluation criteria, and multiple options are provided for use as stopping criteria. Result presentation includes time series plot, scatter plot, tabular report, and statistical summary.

Calibration parameters are classified into five distinct groups: subcatchment group, soil group, aquifer group, RDII group, and conduit group. The subcatchment group represents subcatchment parameters and groundwater parameters such as area, width, percent imperviousness, slope, and many more. The soil group contains infiltration parameters including Horton's parameters, Green-Ampt parameters, and Curve Number parameters. The aquifer group comprises aquifer parameters including porosity, wilting point, field capacity, and etc. The RDII (Rainfall Driven Infiltration and Inflow) group represents RDII tri-triangular unit hydrograph (Boulos, 2005) parameters including R, T, and K for short-term, medium-term, and long-term responses, and the conduit group represents conduit parameters including the Manning's roughness coefficient. Parameters belonging to each one of these five groups could vary spatially depending on the behavior of the individual element. For example, parameters in the subcatchment group could vary from one subcatchment to another depending on the shape, size, topography, and groundwater flow condition of the subcatchments. However, depending on similarity of elements, the modeler can force multiple elements to have same parameter value using the grouping mechanism. As an example, if two or more subcatchments have similar shape and size, one may decide to use the same width value for these subcatchments. Similarly, subcatchments with similar topographic condition may assume the same slope value. This grouping mechanism can significantly reduce the number of calibration parameters (decision variables) without sacrificing accuracy of the calibration exercise.

For sensitivity analysis, the calibration model utilizes stepwise regression (Muleta and Nicklow, 2005). As the name implies, stepwise regression involves a step-by-step construction of multiple regressions, starting with a simple regression model and building upon it. At each step, another input that is highly correlated to output uncertainty unexplained by the regression model is considered, until a stage appears at which inclusion of another input factor does not appreciably improve performance of the regression model. In brief, stepwise regression analysis involves the following tasks: (1) performing a correlation analysis between each of the input factors and an output variable, and selecting the input factor most related to the output; (2) constructing a simple regression model between the output variable and an input factor highly correlated to the output variable; (3) determining output uncertainty not yet explained by the current regression model, performing a correlation analysis between output uncertainty not yet explained by the regression model and all input factors (excluding those factors selected in previous steps), and identifying the highly correlated input factor; (4) constructing a multiple regression model between the output variable and the input factor(s) selected to this point; and (5) testing the performance of the regression model and the stopping criteria, and to repeat the process until the stopping criteria is satisfied.

Once parameters to be calibrated are identified, the calibration module embarks on searching for best-fit parameter set using either GA or PSO. As described previously, both GA and PSO start the search process with a population of candidate solutions. Each of these solutions is made up of a certain value for every calibrable parameter, where the values are randomly selected from the feasible range, defined in terms of the minimum and maximum bound specified for the individual parameters. Relative fitness of the candidate solutions is evaluated by comparing model simulations with observed data at one or more sites. For fitness evaluation, the calibration tool offers a number of criteria including the root mean square error, Nash-Sutcliffe efficiency criterion, modified

coefficient of efficiency, r-square, simple least square error, mean least square error, dimensionless root mean square error, dimensionless simple least square error, and deviation in total volume among observed and simulated values. Fittest alternatives (i.e., solutions that bring model simulations closer to observed data) are given better chance to participate in creating new solution candidates following the respective algorithm's (i.e., GA or PSO) search mechanism described previously. The process of determining relative fitness of solution candidates and creating new alternatives continues until the stopping criterion is satisfied. The automatic calibration module described here utilizes one or more of the following criterion to terminate a search process: (1) fitness value exceeding a specified threshold (e.g., r-square greater than 0.99, root mean square error less than 0.05); (2) the sum of square of differences between observed and simulated values falling below a specified threshold; (3) reaching the maximum number of model simulations allowed; and (4) failing to achieve specified improvement in the fitness criterion within specified number of consecutive generations.

## **5. THE DESIGN MODEL**

Control of sewer flooding is vital to reducing risks to public health and protecting the environment from water pollution. Remediation of the adverse impacts of urban drainage systems such as sewer overflows is commonly achieved by constructing stormwater detention ponds, increasing size and slope of pipes to enhance conveyance capacity, and increasing pumping capacity. The objective is to determine the least-cost design solution that meets regional environmental regulations and can be realized using optimization technology. The automatic design methodology eliminates the need for the tedious trial-and-error approach and uses GA to screen the improvement alternatives. The optimal solution is selected from a combination of pipe upsizing and alignment, storage size, and pump capacity. The design constraints represent desired system performance criteria and include values for maximum depth to diameter ratio, minimum and maximum pipe velocity, maximum head loss in force mains (i.e., pipes always flowing under pressure), and minimum and maximum downstream pipe offset (i.e., slope). The solution is determined from user-specified eligible options such as shape and maximum possible storage size, minimum and maximum possible pipe sizes and pipe offsets, and a list of available pump curves. In addition, the optimal design module ensures that impractical engineering solutions such as pipe sizes decreasing in the downstream direction are discarded.

Improvement unit costs (i.e., cost per pipe length, pump cost, and/or cost per volume of storage) can be specified for each type of candidate improvement (construction of a new sewer, building additional storage, and increased pumping capacity) alternatives. These cost data can vary for each option depending, for example, on the size and location of the conduit, size and location of detention pond, and size and type of pump. Any number of unit costs can be designated for each sewer size so that estimated costs based on urban versus rural areas, high traffic versus low traffic areas, and other pipe installation conditions can be accurately represented. The portion of the system to be designed may also be specified including the entire system, any particular sub-division or geographical area. Conduits, storages, and pumps selected for an optimization run are

lumped into distinct groups for each type of facility based on the cost. Separate facility groups are then identified with the appropriate cost code and allowable improvement actions. Following the completion of the optimization run, the identified top ten solutions are presented. Output results include recommended replacement sewers, additional or increased storage size, and modified pumping capacity for increasing the hydraulic capacity of existing system.

## 6. APPLICATIONS

The optimization methods are demonstrated using an actual stormwater system containing 208 junctions, 214 pipes, 1 outfall, 2 pumps, and 5 storage nodes. Rainfall driven infiltration and inflow (i.e., RDII) is the major source of wet weather flow for the system. The tri-triangular, also called RTK, unit hydrograph approach to estimate RDII into a sewer system is used in this example. The RTK hydrograph contains three triangular hydrographs, one for a short-term response, one for an intermediate-term response, and one for a long-term response. Each of the three hydrographs is defined by three parameters (i.e., R, T, and K). R represents the fraction of rainfall volume that enters the sewer system, T represents the time from the onset of rainfall to the peak of the hydrograph in hours, and K represents the ratio of time to recession of the hydrograph to the time to peak. Therefore, estimation of RDII at a sewer collection system node requires definition and calibration of up to nine parameters. Since flow measurements are available at every receiving node, the calibration model is applied using flow entering a junction node. All nine parameters of the RTK hydrograph are calibrated. The sensitivity analysis model was not utilized for this example based on the notion that optimizing for nine parameters is manageable.

The calibration problem was solved using GA optimization. A population size of 100 was used and the maximum generation was set to fifty. A time series plot of the calibration results is given in Figure 1. As shown in the figure, excellent agreement between model predicted values and field measurements is obtained.

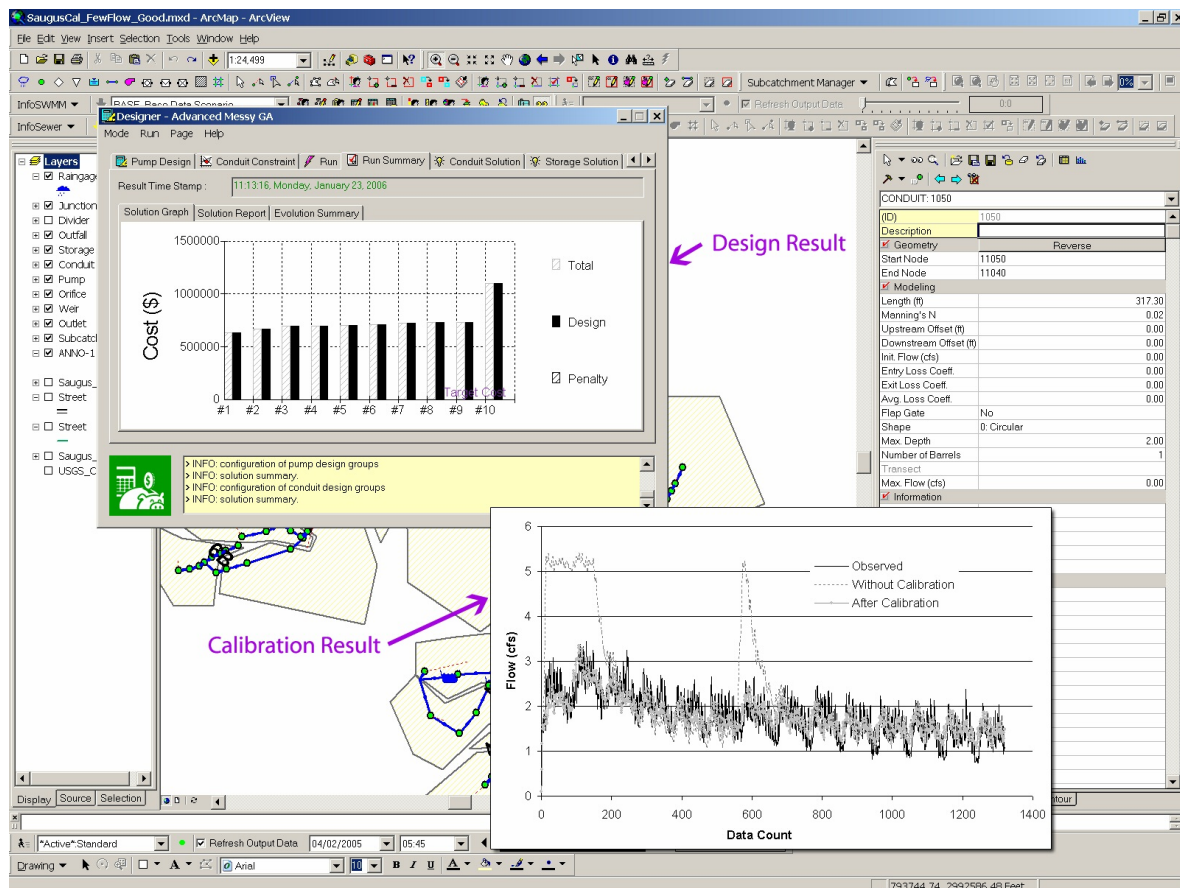
The optimal design model was then applied to identify cost-effective design improvements so that capacity of the existing treatment plant is not exceeded. Design results that show the costs incurred by the top ten proposed solutions are also given in Figure 1.

## 7. CONCLUSION

Computer based network simulation models provide the most viable and effective means of evaluating urban drainage system performance, identifying deficiencies, and assessing the reliability of proposed design improvement alternatives. However, these models must first be properly calibrated to ensure that they predict system behavior with sufficient accuracy. Historically, most urban drainage model calibration and design/rehabilitation attempts have typically employed a manual trial-and-error approach. Such an approach often proves time consuming and frustrating when dealing with most typical urban stormwater systems. In this paper, comprehensive models for use in automatic calibration and design are developed using optimization theory. The automatic calibration model adjusts model parameters to reduce discrepancies between measured data and model



results. The automatic design model determines the optimal combination of pipe upsizing and alignment, storage size, and pump capacity to meet desired system performance requirements at lowest cost. Genetic algorithms and particle swarm optimization are coupled with the USEPA storm water management model (SWMM5), and are used to determine optimal values of calibration and design parameters. The models are embedded within GIS (ArcGIS) to provide a powerful geospatial decision support system for developing complex urban drainage system models, calibrating them, and optimizing them to achieve improved performance and lowest cost. The resulting decision support system should prove useful to any wastewater utility attempting to optimize its capital improvement program.



**Figure 1: Optimal Calibration and Design Results.**

## 8. REFERENCES

- Boulos, P.F. (2005) 'InfoSWMM Suite – Users Guide', MWH Soft, Inc., Pasadena, CA.  
 Boulos, P.F., and Ormsbee, L.E. (1991) 'Explicit Network Calibration for Multiple Loading Conditions', *Journal of Civil Engineering Systems*, 8, 153-159.

- Cieniawski, S.E., Eheart, J.W., and Ranjithan, S. (1995). "Using genetic algorithms to solve a multiobjective groundwater monitoring problem." *Water Resour. Res.*, 31(2): 399-409.
- Holland, J.H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI.
- Jung, B.S., and Karney, B.W.(2005). Benchmark Tests of Evolutionary Computational Algorithms. *Proceedings of the 2005 Conference of the Environmental and Water Resources Institute, ASCE*. Anchorage, AK .
- Kennedy, J. and Spears, W. M. (1998) Matching algorithms to problems: an experimental test of the particle swarm and some genetic algorithms on the multimodal problem generator. In *Proc. 1998 IEEE World Congress on Computational Intelligence*, pp. 74--77, Anchorage, AK.
- Kennedy, J., and Eberhar, R.C. (1995). Particle swarm optimization. *Proceedings of the IEEE International Conference on Neural Networks*, Perth, Australia, IEEE Service Center, NJ, IV: 1942-1948.
- Muleta, M.K., and Nicklow, J.W. (2005) "Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model." *J. of Hydrology*, Elsevier, 306: 127-145.
- Nicklow, J.W. (2000). "Discrete-time optimal control for water resources engineering and management." *Water International*, 25(1): 89-95.
- Nicklow, J.W., Boulos, P.F., and Muleta, M.K. (2004) 'Comprehensive Sewer Collection Systems Analysis Handbook for Engineers and Planners', 1st ed., MWH Soft, Inc. Publ., Pasadena, CA, 273 pp.
- Rossman, L.A. (2005). "Stormwater Management Model User's Manual, Version 5", U.S. EPA, Cincinnati, OH.
- Shi, Y.H., and Eberhart, R.C.(1998). A modified particle swarm optimization, *Proceedings of the 1999 congress on evolutionary computation*, Piscataway, NJ, 1945-1950.
- Sorooshian, S. and Gupta, V.K. (1995). "Model calibration" In Singh, V.P. (editor), *Computer models of watershed hydrology*. Water Resources Publication, Highlands Ranch, CO.
- Schwefel, H-P (1995). *Evolution and optimum seeking*. Wiley, NY.