Chapter 8

On the Optimization of Uncertainty, Complexity and Cost for Modeling Combined Sewer Systems

William James, Taymour El-Hosseiny and Hugh R Whiteley

This chapter introduces a tentative procedure for estimating the minimum level of complexity of a computational model for designing a cost-effective combined sewer overflow (CSO) system. In this approach, model complexity is related to a measure of model uncertainty, and to the design costs. Ultimately the goal is to determine the optimal (least cost) complexity for a model (version 4.3 of the U.S. EPA Storm Water Management Model, SWMM) which is itself uncertain.

Differences between modelers, and errors in datafiles could invalidate our proposed method. Thus we propose a procedure that first produces consistent models free of user-input error. Developed in the form of an expert system shell, our procedure uses a front-end rule-based (FERB) decision support system (DSS). This FERB-DSS comprises two functions, model development, and sensitivity analysis: first the user is led through the development of consistent input files; second a heuristic sensitivity analysis procedure is applied to selected model parameters. In a following step, calibrated parameters that consistently reproduce observations are determined. Finally, minimum uncertainty and corresponding cost production functions are derived, and from these parameters an optimal level of model complexity is defined. This chapter describes the final step, as well as the problem of determining a globally optimum set of calibration parameters for SWMM.

James, W., T. El-Hosseiny and H.R. Whiteley. 1998. "On the Optimization of Uncertainty, Complexity and Cost for Modeling Combined Sewer Systems." *Journal of Water Management Modeling* R200-08. doi: 10.14796/JWMM.R200-08.

© CHI 1998 www.chijournal.org ISSN: 2292-6062 (Formerly in *Advances in Modeling the Management of Stormwater Impacts*. ISBN: 0-9697422-8-2)

Using 1994 U.S. EPA policy as a guideline for design, our proposed method is applied to a complex and costly problem: CSO controls in Columbus OH. We conclude that analysis of uncertainty is essential when using models to design CSO controls.

8.1 Introduction

Engineering designs of stormwater and wastewater drainage systems are clearly becoming increasingly complex. Nowadays their objectives are to minimize problems caused by water drainage, to meet expanded and strengthened water quality standards, and to protect both public health and aquatic life. Many of the problems to be corrected are related to wet weather performance of these drainage systems. Wet weather problems may be obvious and recorded, like basement flooding and flooded roads, which result in health hazards, loss of local property values, and sometimes, loss of property. Or they may go unrecorded, despite their serious consequences. For example, when a sanitary sewerage system surcharges (hydraulic grade line rises above the sewer crown), due perhaps to a constriction downstream, suspended solids may deposit in the conduit network, reducing cross-section and capacity.

Intensive monitoring studies are rarely feasible in these cases, due to cost and the delay. On the other hand, computational drainage models are valuable tools for assessing drainage system response to various wet weather events, especially when evaluating alternative control strategies (WEF, 1994). Models can be tested against real or synthetic events, and input parameters can be optimized and verified against observations or against other design criteria. Once verified, a model may be used with an appropriate level of certainty for simulating unmonitored conditions, for design optimization of the sewerage system and for real-time control.

To solve even the identified problems, the first step of selecting the best modeling options is itself problematic. Concerns include:

- What array of models should be used?
- What is the extent of model applicability in the context of the study objectives?
- What level of design accuracy is achievable?
- What is the uncertainty of the model results?
- What investment of modeling effort is most cost-efficient?
- Is cost-efficiency an appropriate measure for optimizing an uncertain design?

Such questions beg the issue of model complexity, which affects the requirements for data collection, the model calibration effort, selection of the objective function (e.g. total flow, peak flow, duration of flow exceedance), and

selection of modeling options. Extremes of model complexity may lead to the user becoming mired in small details of process disaggregation, system discretization and model reliability, or so simplifying the modeling process that the exercise results in over- or under-, design. One approach to this optimization problem is to perform detailed sensitivity and error analyses of the models, based on specific criteria (e.g. weighted error). However, computational resources required for complex, continuous hydrologic models (e.g. simulation of 75 years of data) still present a problem, whether real or imaginary. Spectres of non-linearity and global optimization are inevitably raised, and are addressed later.

To demonstrate our approach, the SWMM model (Huber and Dickinson, 1988) was used in our study. SWMM simulates many aspects of urban hydrology and hydraulics including surface runoff and pollutant routing through a water channel or a collection system. Obviously, for reliable, efficient use, a complete understanding of the model and its limitations is required. But for the average user the SWMM interface presents a barrier, and care and attention is necessary. There is a clear role for intelligent interfaces such as PCSWMM97 (James and James, 1998).

In our study we introduced a method for interfacing SWMM with a knowledge base, and applied an error analysis to evaluate the credibility of the resulting model. Two general objectives were:

1. develop a FERB-DSS to interface with SWMM, and
2. optimize the modeling effort against model uncertainty using a systematic cost-effectiveness procedure (in other words, estimate the optimum value of model complexity considering reliability of model results and design costs).

In this chapter we do not cover the first objective, but focus on the latter objective.

8.2 Procedure for Determining Optimal Complexity

Cost estimation factors that make each case study unique include:

1. level of risk acceptable to a community;
2. number of sensitive areas (outstanding national resources, national marine sanctuaries, water that provides habitat for endangered species, waters with primary contact recreation, and water used for public water supply and shell fish); and
3. savings due to the management of damages.

Ultimately a relationship is required between system cost and model uncertainty, which can be further processed to produce the minimum cost. Cost is taken to be a combination of the following:

1. engineering fees to design alternative solutions;
2. construction costs of the selected alternative;
3. intangible costs;
4. damage costs due to uncertainty of the selected option.

Figure 8.1, not drawn to scale, depicts these relationships for a cost-effective design using arbitrary units. Total combined cost of the selected alternative is the sum of the individual unit cost at pre-defined levels of uncertainty. The cost-effective design is taken to be least-cost combination. Design costs can be estimated as a function of model components. Construction costs can be estimated as a function of design flow rates or storage volume, and rise with uncertainty, because of the tendency to over-design as uncertainty increases. Uncertainty cost arises from a failure of the design to achieve the correct optimum size thus incurring either excessive construction costs or excessive environmental and other intangible costs. However, assigning a dollar value to intangible costs is difficult due to its wide variation. Thus an index value was assigned to the intangible cost based on a weighting factor and the uncertainty. If we ignore intangible costs in the total costs, capital costs will likely control the EF, because the design costs are a small percent of the total. If intangible costs are included, total cost can be expected to reach a minimum (optimal) value at a certain level of complexity.

Figure 8.1 Cost optimization of CSO control option (arbitrary units).

Our proposed method may be said to derive the components schematized in Figure 8.1, and we tentatively proposed a procedure in point form as follows:

- Formulate all possible processes (e.g. snow melt, infiltration, run-off, storage) to be modeled and the associated parameters and variables.
- Estimate initial values for the model parameters.
- Select an objective function (OF) that serves the objectives of the

analysis. Here our objective is to determine a cost-effective model to design a CSO control facility, which is often related to storage volume, so we selected *total runoff volume* as the OF for the uncertainty analysis.

- Select a state-variable (SV) for input (e.g., high, medium, low rainfall intensity) and run the model.
- Apply a first-order sensitivity analysis to determine the sensitivity of model parameters selected for that SV, based on literature reviews and prior experience.
- Repeat steps 3 and 4 for all input SVs.
- Evaluate results from the sensitivity analysis and deactivate unnecessary processes.
- Compare model results to observed data (total volume of each runoff hydrograph) and determine the error in computing the OF.
- Select an error function (EF) as a measure of uncertainty. We used the first dimensionless form of the simple least square equation because it favors large flows.
- Calibrate model results to minimize differences between observed and computed OF based on sensitivity of model parameters.
- Sequentially assign more or less complexity (i.e., increase or decrease the number of modeled sub-spaces, watersheds or pipes) and estimate the error in the OF at each level of complexity.
- Calculate total costs of a control measure of CSO's using the U.S. EPA policy guidelines. Using model results at each level of complexity, the control facility is redesigned. Total costs are the sum of monetary, intangible costs and uncertainty costs. The monetary costs are the construction and engineering fees. Intangible costs are assumed to be an inverse function of the size of the control facility. Uncertainty costs are constructed from weighted functions for the extra cost due to deviation from the minimum optimal size as a result of model uncertainty.
- Add design office costs.
- Produce a relationship between total cost and model complexity.
- Estimate a reasonable level of complexity for minimum cost.

8.3 Application to Columbus, Ohio

For the 357,000 acre (1,500 km²) Columbus Metropolitan Area, the Division of Sewerage and Drainage (DOSD), collects, transports and treats sanitary wastewater and stormwater flows. In the center of the city, the combined sewer shed (CSS) area has been divided into sanitary sewer-sheds (SSs) and

runoff catchments (RCs), whose total area is about 5,285 acres (22 km²), about 1.5% of the current and future service area. They drain from east to west and flow in the main interceptor (OSIS) sewer that travels from north to south, parallel to the Olentangy River, where the river merges with the Scioto River. OSIS was planned to carry sewage from the combined sewer area in the central portion of the city and the outlying county sewer districts, which are served by separate sanitary sewer systems. It was designed in 1930 for a 30-year design period to serve a tributary area of 55,424 acres (232 km²), with estimated average DWF of 125 U.S. gal/capita.day (468 L/capita.day). Included was a provision for control of storm-sewage entering from the combined sewers via nineteen regulator chambers at the connections of the combined sewers and OSIS, designed to overflow during flows that exceed the capacity of OSIS. An adjustable sluice gate is installed at each regulator chamber to control the amount of flow entering OSIS. OSIS was designed such that the combined sewers would overflow an average of eight times per year. A stand-by tank was constructed at the end of OSIS and upstream of the POTW to provide partial treatment of the excess storm sewage flow.

Nearly 30 years after the end of the design period, during the 1980's, officials realized that the original design assumptions for many of the existing interceptor sewers no longer held. Current conditions were highly dynamic, unknown sources of excessive inflow and infiltration were present in the combined system, and water backed up into basements. The frequency of overflows from the combined sewer system was unknown. According to U.S. EPA requirements for long-term control plans for combined sewer systems, each municipality should adopt either the *presumption* or the *demonstration approach*. The presumption approach requires that the maximum number of overflows should not exceed four per year. A review of design reports for the CSS revealed that the north side of the central area of the City of Columbus was designed for eight overflows per year. Because of the complexity of the collection system, the city decided to develop a SWMM model to simulate the collection system and investigate its response to various hydrologic factors.

8.3.1 Discretization of sanitary sewer-sheds and runoff catchments

SWMM RUNOFF was used to generate runoff hydrographs at each RC. EXTRAN was used to route computed flows through the collection system. Table 8.1 summarizes all tasks related to each RC.

8.3.2 Monitoring Program

Monitoring was carried out for the study area by URS (1989). Their program included installation of a rain gage network and flow monitors. A flow monitor was installed at the end of each tributary area and upstream of its

Table 8.1 Runoff catchment data extraction task.

1.	Interceptor Level:
	Develop covering maps for combined sewer system
	Identify modeled sewers
	Get record drawings for modeled sewers
	Generate Pipe Ambiguity Record to clarify linkage
	Identify modeled storm sewers
	Identify interceptor sewer
	Identify sewers and areas discharging to separate storm system
	Get topographic maps
2.	Catchment Level:
	Transfer modeled nodes from sewer-shed base
	Delineate outer catchment boundaries
	Identify new nodes required for modeled sewers
3.	Subcatchment Level:
	Complete interior boundaries using small diameter sewers
	Identify nodes for subcatchment contribution
	Extract the following parameters for subcatchment
	Areas
	Width
	Percent Imperviousness (aerial photography)
	Ground slopes (average along overland flow)
4.	Modeled Sewers:
	Identify nodes using record drawing profiles
	Extract new or revised sewer segments using the identified nodes

regulator. Data were recorded at 15-minute intervals, and included simultaneous water depth and velocity. A rain gage network of 20 rain gages was constructed to monitor rainfall time series across the entire City of Columbus area. Rain gages consist of a computer as a data logger and a tipping bucket measuring 0.01 inch (0.25 mm) per tip.

8.3.3 Summary of Precipitation Observations:

A frequency analysis of rainfall data was carried out to determine (i) different levels of SVs for modeling purposes, and (ii) the relationship between rainfall and overflow frequency. This analysis used data collected between 1988-1994 for rain gage 14. Precipitation events were separated using a minimum inter-event time (MIT) of 4 hr based on a qualitative analysis using different values of MIT of from 3 to 7 hr (EMH&T, 1994). The MIT is an artificially-chosen minimum number of consecutive zero precipitation hours separating two storm events.

Storms varied in intensity from 0.1 in. (2.5mm) per day to 4.25 in. (106 mm) per day, but only storms above 0.2 in. (5 mm) per day were considered in this study. A total of 357 events were ranked, based on total rain, for years 1988-1994.

After reviewing the flow data, the recorded daily precipitation were classified into three state variable (SV) spaces:

SV Space	Rainfall Total (x)
Small	$0.2 < x < 0.5$ in ($5 < x < 12.5$ mm)
Medium	$0.5 < x < 1.0$ in ($12.5 < x < 25$ mm)
Large	$x > 1.0$ in ($x > 25$ mm)

The model was tested and calibrated against all SVs (i.e., small, moderate and large events) based on the above criteria for total rainfall. From the entire precipitation record, five representative events were then selected.

8.3.4 Model Development:

Major tasks of the modeling procedure were:

1. Develop a database that manages discretized data customized to fit the requirements of the selected model.
2. Select flow component(s) needs to be analyzed (e.g., DWF, base infiltration).
3. Identify special modeling concern based on API (e.g., diversion chamber, boundary condition, overflows, and storage).
4. Develop two model runs, one for DWF using SWMM TRANSPORT and another for rainfall-runoff using SWMM RUNOFF.
5. If tributary SS is modeled for wet conditions, develop Infiltration/Inflow (I/I) based on flow monitoring data to define rainfall dependent infiltration/inflow (RDII). I/I hydrographs are stored at predefined nodes within SWMM TRANSPORT.
6. Combine all flows at common nodes (i.e. DWF, Runoff, I/I).
7. Route combined flows using SWMM EXTRAN and assign valid boundary conditions.
8. Compare model results with observed data at selected nodes and conduits for a predefined OF. For first OF chosen was *peak flow* to assess capacity of the collection system. The second was *overflow volume* and its frequency, chosen to meet U.S. EPA policy.
9. Carry out sensitivity analyses for selected parameters to assist in model calibrations.
10. Estimate errors in the OFs.
11. Determine optimum estimates of model parameters and rerun the model.
12. Repeat steps 10 and 11 until differences between computed and observed OF become constant and no improvements occur in estimated error.
13. Display model results.

8.3.5 Model Development and Calibration:

The total acreage of SS is about 600 acres and of RC is about 270 acres, almost fully developed. The primary activities within each subcatchments are categorized as follows: single-family residential, multi-family residential, commercial, industrial, and open areas. To optimize model resolution (i.e. optimum number of RCs and SSs, and optimum number of calibration trials), three datasets: 27, 10 and 1 RCs, were set up. Figure 8.2 is the schematic for 27 RCs. Initial model parameters were discretized based on 27 local areas. Aggregation of model parameters was based on parameter type. For example the area parameter was aggregated by the summation of local RC and SS, percent of imperviousness was aggregated using the following equation:

$$\%imp_i = \frac{\sum (\%imp_j \times Area_j)}{\sum (Area_j)} \quad (8.1)$$

where:

- $\% imp_i$ = % imperviousness of aggregated area i
- $\% imp_j$ = % imperviousness of local area j
- $area_j$ = total acreage of area j

Width of the RC was based on the longest path of flow at each RC. Ground slope was recalculated based on average ground slopes after aggregation.

8.3.6 Sensitivity Analysis:

Sensitivity coefficients were derived to generate a sensitivity matrix that can be used in model calibration and error analysis, in an order that distinguishes between the orders of confidence in the estimated input values:

- Order 1: Parameter can be estimated with an expected size of error between 1-10% (e.g., diameter, pipe length, pipe slope).
- Order 2: Parameter can be estimated with an expected size of error between 10-20% (e.g., flow, ground slope, ground elevation).
- Order 3: Parameter with an expected size of error between 20-30% (e.g., infiltration parameters, pipe roughness).
- Order 4: Parameter with an expected size of error between 30-100% error (e.g., initial soil-water deficit).
- Order 5: Parameter cannot be measured and no level of confidence can be assumed (e.g. % imperviousness, depression storage subcatchment width).

Dunn and James (1985) recommended at least three values for each parameter to account for the expected range and distribution of estimates of parameters. Their approach indicates whether computed results are approximately linear or nonlinear, so that an appropriate sensitivity analysis method may be applied.

Sensitivity analysis was applied twice, once for a complex model with 27 RCs and SSs, and another for a less complex model with ten RCs and SSs. No sensitivity analysis was carried out for one RC due to its obliviousness of relief structures upstream of the last node. Sensitivity gradients were obtained for the following six parameters, which can be considered to be the least-confident parameters that influence the OF (*peak flow* or *overflow volumes*) (James and Robinson, 1981) and (Zaghloul, 1983): area, percent of imperviousness, ground slope, runoff catchment width, impervious depression storage, pervious depression storage. Peak flow was used as an OF to display sensitivity. James and Robinson (1981), Zaghloul (1983), and Lei and Schilling (1993) applied sensitivity analysis for the above parameters and showed that the sensitivity gradients for *peak flow* as an objective function yields the same gradients as for *flow volume*. In this study, *peak flow* was used as the OF for the sensitivity analysis but *flow volume* was used for the design of the CSO control facility. Dominant parameters were selected based on their sensitivity: area, % of imperviousness, subcatchment width. Sensitivity gradients were computed using normalized (relative) sensitivity so that comparisons can be made irrespective of parameter units:

$$R_s = \frac{\left(\frac{\partial O}{\partial p} \right)}{\left(\frac{O}{p} \right)} \quad (8.2)$$

where:

- R_s = relative sensitivity
- O = output objective function
- p = input parameters

The analysis showed that peak flow sensitivity increased with complexity (higher number of RCs). The relative sensitivity R_s , using percent of imperviousness as indicator decreased by 17% by increasing model complexity from 10 RCs to 27 RCs. The impact was higher using subcatchment width as indicator, where R_s decreased by 50%.

Figures 8.3 and 8.4 illustrate sensitivity gradients for different complexities and the following percent chances of the parameters around its mean value: +/-20%, +/-50%, and +/-70%. Results of sensitivity gradients for percent of imperviousness showed that as complexity (number of subcatchments) increases, a decrease in sensitivity occurs using % imperviousness and subcatchment width as indicators.

Figure 8.3 Impact of % imperviousness change on computed peak flow.

Figure 8.4 Impact of width change on computed peak flow.

Sensitivity analysis showed that *area* and *percent of imperviousness* were the most sensitive parameters in this study, as expected.

8.3.6 Model Calibration

Our developed model was calibrated against all SV spaces (i.e. small, moderate and large storm events), using one parameter at a time. Sangal and Bonema (1994) used this approach for a large number of urban watersheds in the U.S. to calibrate for both runoff volume and response time. Our analysis was applied for different complexities (10 and 27 RCs) until sufficient convergence was reached. Objective functions were minimized or maximized perturbing one parameter at a time while others were held constant. Ultimately for a system that behaves conformably, a set of parameters that can be used for the next trial is obtained. A calibration trial is a series of steps that improve the match between the observed and computed values for a selected objective function (OF):

1. Compute the difference between the observed and computed OF using initial model parameters.
2. Calibrate the OF using the most sensitive parameter. Linear interpolation is applied by increasing or decreasing the parameter based on its sensitivity. A new value of the parameter is derived.
3. If the match is good no further steps need to be applied. If a considerable error in the computation still exists, the second sensitive parameter is altered based on its sensitivity, without making any changes to the previous parameter.

Common concerns about this serial, heuristic method are that real systems are not linear, but non-linear, and that this search (optimization) may find a local but not necessarily the global optimum. In response, we believe that our approach (ranking dominant parameters and solving for their best values one at a time) succeeds for SWMM studies like ours because:

- SWMM probably behaves approximately linearly much of the time, though clearly not for overflow events, when it is discontinuous (yet we found that our method still works!).
- Non-linearity of the real system is irrelevant, rather it is the model that matters, since it is being calibrated.
- Initial estimates were kept within 10% of final values, so that there is little space for multiple peaks.
- The number of dominant parameters is very small, often no more than one, even when the datafile has hundreds of candidates.
- We used a larger number of observations for calibration than the number of dominant parameters to be calibrated.
- Even when sensitivity gradients were non-linear correct results were obtained.
- SWMM though a non-linear system is structured as a system of independent processes.

Our methodology was tested in a graduate course by assigning different sets of computed (instead of observed) input and output to the various graduate

students, and corrupting the input parameters by up to 20%. Every student solved the problem obtaining a perfect set of original input parameters. This is not proof, but evidence that the method has validity.

On the other hand, it may indeed not work for systems where unexplained interactions occur. These could derive from e.g. a serious coding bug whereby an otherwise independent process is affected by some unrelated variable (snow pack depletion in a remote catchment by this means could remove sediment in an unrelated sewer perhaps in an earlier month). Or it could be by other unexplained influences, perhaps those that sometimes intervene in our own coding. But for our routine and well-behaved, conformable model application, it appeared to work consistently. Further discussion is considered inappropriate here, since the calibration procedure is a small part of our overall method.

Our results showed that, for our limited studies, the higher complexity (i.e. 27 RCs) gives faster relative improvements in model error than is the case for lower complexity (i.e. 10 RCs) based on the slope of the curve. Also, most of the improvement in model error occurred at the second trial and extra trials showed little improvement. Uncertainty of model results used the same evaluation function (EF) used for different SV spaces.

8.4 Uncertainty Analysis

Now we estimate model uncertainty due to model errors after model calibration. *Flow volume* was selected as an OF, and the first dimensionless form of simple least squares was selected to calculate the error (e) in computing the OF:

$$e^2 = \left(\frac{Y - F(z)}{Y} \right)^2 \times 100 \quad (8.3)$$

Table 8.2 illustrates the rather wide range of calculated variances based on a dimensionless form of simple least squares for the error of *computed flow volume* based on the difference between computed and observed values. Figure 8.5 displays average error of computed flow volume at different complexities. Results shows that the change in model error for a model with 10 compared to 27 RCs is less than the change in model error for a model with 1 compared to 10 RCs. Improvement in model accuracy going from 1 to 10 RCs is much higher than for models going from 10 to 27 RCs. It is noted that much of the method up to and including Table 8.2 can be automated, given sufficient coding assistance.

For validation of the calibrated model, the rainfall event of July 21, 1988 was used. Computed error (variance) in estimating the total runoff volume was about 0.25.

Figure 8.5 Uncertainty of flow volume based on number of sub-catchments for different state variables.

Table 8.2 Error in computing flow volume for different number of subcatchments.

Date	e ²	e ²	e ²
	1 RC	10 RCs	27 RCs
06/16/88	0.3100	0.0400	0.0400
08/05/88	0.0730	0.0500	0.570
08/15/88	0.6200	0.3000	0.2000
08/20/88	0.9400	0.3700	0.2500
Average Error (σ^2)	0.4800	0.1900	0.1300
Average Error (σ)	0.6900	0.4300	0.3600

8.4.1 Alternative Selection and Cost-Benefits Optimization:

While the overall objective of CSO control is to protect or to improve the water quality of the receiving water body, selecting a control process and determining its size involves a specific performance target. The high variability of CSO and the complex nature of receiving waters (U.S. EPA, 1993) make it difficult to establish a relationship between CSO and water quality. The U.S. EPA recommended defining performance goals for CSO control, to wit either a *presumptive* or *demonstrative* approach, selected by the community. Examples are percent capture, overflow frequency, treatment level, first flush, and knee-of-the curve. In this study the *presumptive* approach was adopted, based on the expectation that water quality standards will be met if one of the criteria is achieved. The analysis is presented through sewer system characterization

Figure 8.6 Performance of CSO storage and storm size used as a design basis.

modeling and monitoring. Under both strategies (i.e., reducing number of overflows or increasing percent capture) all captured flow is to be stored in surface storage and treated at a POTW. An appropriate design storm must be defined.

8.4.2 Design Storm Selection for a Control Alternative:

A preliminary indication of the design storm size was developed. Rainfall analysis determined the number of storm events that have larger volumes than the design storm. It was also assumed that the storage unit used to control CSO is not emptied until the end of a storm. Any event volume that exceeds the design storm produces an overflow.

Figure 8.6 illustrates a relationship between the number of overflows per year and the design storm size for which storage is provided, based on rainfall records of the City of Columbus from 1/1/88 - 1/1/95. From observations and flow monitoring records, it was determined that a storm with a total volume of 1.1 inch (32 mm) causes 8 to 10 overflows per year.

Using the CSS performance illustrated in Figure 8.6, and the U.S. EPA presumption approach to set a limit of four overflows per year, the 1.4 inch (35 mm) storm event was selected as the design storm to size the storage facility. For the storm of July 20, 1988, the maximum intensity was 2.04 in/hr and total duration about 2 hours. This storm event caused overflows at all overflow locations and can be ranked as a five-year storm, based on the City of Columbus rainfall intensity-duration-frequency (IDF) curves.

8.5 Cost Functions for CSO Control

In this section we determine an optimum model complexity based on minimum cost of a storage facility to control CSOs. Design of the storage facility is determined using model results at different levels of model uncertainty due to different levels of model complexity.

Cost of the storage facility is estimated for the following items:

- CC construction cost of a CSO control facility of a given size.
- DC design costs (fees) for CSO control facility, design fees are a function of model complexity.
- EC engineering cost to be added to CC (e.g. DC plus construction inspection).
- PC total project costs for installation of a CSO facility; this is the sum of CC and EC.
- IC intangible (external) costs, which includes costs of intangible items that are present after a CSO facility of a certain size has been installed (e.g. present value of costs of future remediation of environmental damage, future property losses, economic losses due to degraded amenities such as beach closures and unattractive riverside parks, legal liabilities).
- TC total cost of a CSO facility of a given size; sum of PC and IC
- OC optimal cost of a CSO facility at a given site; OC is the minimum value of TC for that site and occurs at the optimal size of CSO facility for that site.
- XC the extra cost ($TC - OC$) incurred for a single project by the choice of a size of CSO facility differing from the optimal size.
- UC uncertainty cost, the weighted average of the extra costs incurred when a large number of similar facilities are installed with non-optimal size of facility chosen due to uncertainty in design procedures and due to any bias in the mean size.
- ATCU average total costs with a given level of uncertainty and no bias; it is the sum of OC and UC.
- ATCB average total costs with a given level of uncertainty and with bias; it is also the sum of OC and UC.

To estimate CC for CSO controls, we used costs compiled by the U.S. EPA (1992), which were extracted from actual construction costs, consultant-supplied data and other published data. Several assumptions are included in our cost estimating procedure:

1. U.S. EPA requirements for long-term plan to control CSOs is the minimum requirement to be maintained by any community. The presumption approach adopted in this study was assumed to achieve at least 85% capture.

2. Our estimation procedure to size the control facility is assumed to be unbiased for the different levels of model complexities. It is assumed that the model parameters for each level of complexity are around their mean values.
3. Due to high variability and complexity to estimate the intangible costs (IC), it is assumed that these costs are an inverse function of the constructed storage size multiplied by a coefficient.
4. Model complexity is represented as a cost function (ATCU and ATCB). The function penalizes model inaccuracy for simple coarse and inaccurate models based on determination of model uncertainty.

The design costs (DC) function penalizes costs for a very complex model due to detailed discretization. The estimation of the DC is based on involved tasks to develop a model.

The steps of the proposed procedure to estimate ATCU or ATCB of a CSO control facility are:

1. Develop a CC curve at different storage sizes using the cost function developed by the U.S.EPA (1992). The function used was:

$$\text{Construction costs} = 3.49 V^{0.82} \quad (8.4)$$

where:

Construction costs = U.S. million \$

V = volume in U.S. million gallon

2. Develop an EC curve as a function of the CC. It was assumed that the EC represents 20% of the CC.
3. Estimate project costs (PC) which is the sum of the EC and CC.
4. Develop an intangible costs (IC) curve for various storage sizes as an inverse function of the storage.

This function can be customized for different communities, depending on the needs and priorities. The following function was used:

$$IT = \frac{27.18}{(V + 0.25)^{0.82}} \quad (8.5)$$

where:

IT = Intangible Costs = U.S. Million \$

V = Volume in U.S. Million Gallon

5. Add the IC to the PC to include costs of the intangible items that occurs after building the facility. This cost represents the total costs (TC) of the project.

6. Determine the optimal cost (OC) of the CSO facility, which occurs at the minimum value of the TC. This represents the optimal design size with 100% certainty (i.e. zero error in the calculation of the objective function). Size (i.e. average size) is the average of a number of similar projects (say 100 projects).
7. Estimate the extra costs (XC) for this project due to associated uncertainty of model results. The XC is the difference between the TC and the OC.
8. Develop probability density functions (PDFs) for the storage mean value at different variance levels, corresponding to the variances of the model results for this project. In the absence of prior information, the PDF assumed in this study was the normalized PDF for the normal distribution (Bedient and Huber, 1992):

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-0.5\left(\frac{x-\mu}{\sigma}\right)^2} \quad (8.6)$$

9. Perform numerical checks on the PDFs (e.g. ensure that the area under the curve is equal to unity and that the peak value is located at the mean).
10. The PDFs are used as weighting functions to penalize model uncertainty. The weighting functions are multiplied by XC to produce weighted average (uncertainty costs) of the extra cost due to uncertainty in design procedures.
11. Integrate the area under the weighted average uncertainty cost curves for each uncertainty level to obtain the cost due to model uncertainty (UC).
12. Calculate the average total costs at a given uncertainty level (ATCU) by adding the UC at each level of model uncertainty corresponding to each level of model complexity.
13. Determine the minimum value of the ATCU, which would be the optimum level of model complexity.

The losses could be considered to be a reduction in the project benefits for the cost-benefit evaluation.

8.6 Results and Discussion

The economic analysis was applied to our selected model complexities (1, 10, 27 RCs). Figure 8.7 illustrates the cost curves for a CSO control facility as functions of the storage size. Figure 8.7 includes the following curves:

construction costs (CC),
 intangible costs (IC),
 project costs (PC), which is the sum of the EC and the CC, and
 total costs (TC), which represent the sum of the PC and the IC.

Figure 8.7 Cost function curves for CSO control facility.

The OC is the minimum value of the TC and it occurs in the vicinity of the intersection of the IC and the PC. The DC estimates are developed using common practices in Columbus, Ohio. Estimating the DC involves:

- data discretization,
- model development,
- mapping development, and
- surveying, and field verification.

DC is based on the assumption that monitoring data are available (i.e., observed flow and rainfall data). If these data are not available, extra costs should be added. Table 8.3 illustrates DC for the CSO control facility at different levels of complexity.

Table 8.3 Engineering fees for different model complexity.

Task	Design Costs (US \$1000)		
	1 RC	10 RC	27 RC
Data Discretization	2,000	8,000	16,000
Model Development	2,400	9,000	18,000
Mapping	2,500	10,000	20,000
Surveying & Field Verification	2,500	25,000	50,000
Design	50,000	50,000	50,000
Total	59,400	102,000	154,000

Figure 8.8 illustrates the PDF at different levels of model uncertainty due to different levels of model complexity. Uncertainties of model results are calculated based on average error of model results for different SV spaces (i.e. different rainfall intensities). The first dimensionless form of simple least squares was used. Table 8.4 lists these calculated variances and their corresponding costs at each complexity level (i.e. 1, 10, and 27 RCs).

Figure 8.8 Probability density functions for different uncertainty levels.

Table 8.4 Modeling error and corresponding cost (US\$1000)

# of RCs	Uncertainty	EC	CC	IC	UC	ATCU
	%	U.S. \$	U.S. \$	U.S. \$	U.S. \$	U.S. \$
1	69	1,774,000	8,872,000	10,048,000	2,152,000	22,846,000
10	43	1,748,000	8,738,000	10,190,000	1,343,000	22,019,000
27	36	1,709,000	8,543,000	10,404,000	1,064,000	21,719,000

Figure 8.9 illustrates ATCU, where: $ATCU = CC + EC + IC + UC$. Differences in total cost may appear to many readers to be disappointingly small.

For a model uncertainty of 36%, the minimum occurs using 27 RCS, which would be the optimum level of complexity. Due to the small contribution of the UC in ATCU, use of ACTU to determine optimal complexity is itself subject to uncertainty. It is suggested that the sum of UC and DC is more sensitive to design costs, assuming that the design is the main cause of these uncertainties. From another set of computations, at 43% uncertainty, the minimum sum of UC and DC was found to occur at a complexity of 10 RCs. Such uncertainties may appear to many readers to be disappointingly large.

Figure 8.9 Total costs of a CSO control facility at different levels of model uncertainty.

8.7 Conclusions

A tentative attempt has been made to evaluate the optimal level of model complexity with minimum overall cost. Many assumptions have been made. Parameter uncertainties are evaluated in a complex hydrologic and hydraulic model using first order analysis, considering alternative solutions for combined sewer overflow control, to achieve a design in accordance with established standards. The heuristic sensitivity analysis is described. Production functions for cost-effectiveness incorporated design costs for modeling analysis, construction, and intangible costs of the alternative designs. Model uncertainty is used as a relative and consistent basis for comparison purposes. Our function penalizes high model uncertainties.

From an initial application to the City of Columbus, OH, tentative conclusions can be drawn:

Uncertainty analysis in calibrating a hydraulic and/or hydrologic model provides much insight into the inherent reliability of the design and should form part of the modeling effort for complex studies.

The results of the sensitivity analysis confirm that good estimates of the most sensitive parameters (in this case, area and % imperviousness) must be used. Systematic sensitivity analysis for various state variable spaces allows the smallest number of dominant parameters to be identified and optimized.

In our limited study, model uncertainty *increases* as model complexity *decreases* (as we expected) for small and moderate storms of 1-2 inches (25-50 mm). However, for high hourly intensities such as 4 inches (100 mm), system

behavior was unpredictable. This may be due to system hydraulics (e.g., activation of back-up conditions and relief structures) which may have been poorly modeled.

In determining the optimal complexity, there was no difference between using either the total-costs-of-an-alternative or the design-costs-plus-the-uncertainty-costs.

Increasing the reliability of modeling by increasing model complexity, while it increases the analysis-and-design-engineering fee, incurs only a slight increase in the ultimate cost of the alternative. (In our application, when model complexity was increased by 170%, the total cost of an alternative increased by only 1.3%, while uncertainty was decreased by 16% in relative terms.)

The potential benefits of the procedure include:

- Evaluation of the uncertainty of the information presented by a consultant to support the findings of a model (i.e., judge the value of the information yielded by a computer model),
- Determination of the limitations of model usage,
- Provision of a tool that a municipality can use to estimate present and future needs based upon modeled (computed) and monitored (observed) information,
- Minimization of the sum of the direct monetary costs and the costs of intangible items considered when preparing a CSO control plan, and
- Provision of a criterion for determining the optimal level of model complexity to achieve a prescribed level of reliability.

Acknowledgments

Bill James collected the resources, and provided the research ideas, advisement, facilities and support funds through his NSERC research operating grant, and finally wrote this short chapter from Taymour El Hosseiny's very much longer dissertation. Taymour did all the hard work, using much of his experience with EMHT in Columbus OH, and presented it for a PhD degree at the U of Guelph. Hugh Whiteley contributed the approach used to incorporate intangible costs. For further information contact Bill James at the U of Guelph.

References

- Bedient, P.B. and W.C. Huber, 1992, "Hydrology and Flood Plain Analysis" Addison Wesley Publishing Company, Reading, MA.

- Dunn, A.R. and W. James, 1985, "Extended Expert Systems, Proceedings of the Stormwater and Water Quality Modelling Conference, sponsored by U.S. EPA and Ontario Ministry of the Environment, Toronto, Canada, published by CHI, Hamilton, ON.
- EMH&T, 1994, "Columbus Sewer Capacity Study", a report submitted to the City of Columbus.
- Huber, W.C. and R. Dickinson, 1988, "Storm Water Management Model Version 4, User's Manual", U.S. Environmental Protection Agency, Athens, Georgia.
- James, W. and James, W.R.C. 1998. PCSWMM98 Web pages at: www.chi.on.ca and also at: www.eos.uoguelph.ca/webfiles/james/
- James, W. And M. Robinson, 1981, "Standard Terms of Reference to Ensure Satisfactory Computer-based Urban Drainage Design Studies", Can. Journal of Civil Engineering, Vol. 8, pp. 294-303.
- Lei J. And W. Schilling, 1993, "Propagation of Model Uncertainty" Proceedings of Urban Storm Drainage, Niagara Falls, Ontario, Environment Canada's Great Lakes Cleanup, Ontario Ministry of Environment and Energy and ASCE, pp.465 - 470.
- Sangal, S. and S.R. Bonema. 1994. "A Methodology for Calibrating SWMM Models." *Journal of Water Management Modeling* R176-24. doi: 10.14796/JWMM.R176-24.
- URS, 1989, "Combined Sewer Overflow Study, Columbus, OH", a report submitted to the Division of Sewerage and Drainage, City of Columbus, OH.
- U.S. EPA, 1993, "Combined Sewer Overflow Control", Office of Research and Development, Center for Environmental Research Information, Cincinnati, OH, EPA/6-93/R-93/007.
- U.S. EPA, 1992, "CSO Needs Survey", in a report prepared for EPA entitled: "Cost Estimates for Selected Combined Sewer Overflow Control Technologies", Office of Water, Municipal Support Division, Washington, D.C.
- WEF, Water Environment Federation, 1994, "Combined Sewer Overflow (CSO) Control Policy", Final policy, 40CFR Part 122, pp. 9.
- Zaghloul N.A., 1983, "Sensitivity Analysis of the SWMM Runoff-Transport Parameters and the Effect of Catchment Discretization", Advanced Water Resources, Vol.6, PP 214-223.