Rules *for* Responsible Modeling

by William James

4th Edition

CHI

Rules for RESPONSIBLE MODELING

4th edition

William James

Published by CHI
Guelph, Ontario.

National Library of Canada Cataloguing in Publication Data

James, William, 1937-
Rules for responsible modeling

4th ed.
Includes bibliographical references and index.
ISBN **0-9683681-5-8**

1. Water quality – mathematical models. 2. Hydrologic models. 3. Hydraulic models. 4. Municipal water supply.
I. Title. II Title: Responsible modeling.

TC164.J34 2001 628.1’01’5118 C2001-900212-2

CHI Catalog Number: R184

Reference Data:

Rules for Responsible Modeling, 4th edition. xiv + 303 pp.
James, William, 1937- ,
Compiled and published by CHI, Guelph, Ontario, Canada.
Cover design by Rob James.

Other CHI publications:

User's guide to SWMM. 10th Edition. William James, Wayne C. Huber, Lewis E. Rossman, Robert E. Dickinson, Robert E. Pitt, W. Robert C. James, Larry A. Roesner and John A. Aldrich. ISBN: 0-9736716-1-0. Sep. 2005. 942 pp. Expanded, updated, edited and indexed manual to the USEPA Stormwater Management Model.

Monograph Series – Proceedings of the SWMM Users Group:

Effective Modeling of Urban Water Systems Monograph 13, Proceedings of the 2004 Conference on Stormwater and Urban Water Systems Modeling. ISBN 0-9736716-0-2
Innovative Modeling of Urban Water Systems Monograph 12, Proceedings of the 2003 Conference on Stormwater and Urban Water Systems Modeling. ISBN 0-9683681-9-0
Practical Modeling of Urban Water Systems Monograph 11, Proceedings of the 2002 Conference on Stormwater and Urban Water Systems Modeling. ISBN:0-9683681-7-4
Best Modeling Practices for Urban Water Systems Monograph 10, Proceedings of the 2001 Conference on Stormwater and Urban Water Systems Modeling. ISBN:0-9683681-6-6
Models and Applications to Urban Water Systems Monograph 9, Proceedings of the 2000 Conference on Stormwater and Urban Water Systems Modeling. ISBN:0-9683681-4-X
Applied Modeling of Urban Water Systems Monograph 8, Proceedings of the 1999 Conference on Stormwater and Urban Water Systems Modeling. ISBN:0-9683681-3-1
New Applications in Modeling Urban Water Systems Monograph 7, Proceedings of the 1998 Conference on Stormwater and Related Modeling: Management and Impacts. ISBN:0-9697422-9-0
Advances in Modeling the Management of Stormwater Impacts Volume 6, Proceedings of the 1997 Conference on Stormwater and Water Quality Management Modeling. ISBN:0-9697422-8-2
Advances in Modeling the Management of Stormwater Impacts Volume 5, Proceedings of the 1996 Conference on Stormwater and Water Quality Management Modeling. ISBN:0-9697422-7-4
Advances in Modeling the Management of Stormwater Impacts, Proceedings of the 1995 Conference on Stormwater and Water Quality Management Modeling. ISBN:0-9697422-5-8
Modern Methods for Modeling the Management of Stormwater Impacts, Proceedings of the 1994 Conference on Stormwater and Water Quality Management Modeling. ISBN: 0-9697422-X-X
Current Practices in Modelling the Management of Stormwater Impacts Proceedings of the 1993 Conference on Stormwater and Water Quality Management Modeling. ISBN: 1-56670-052-3
New Techniques for Modelling the Management of Stormwater Quality Impacts, Proceedings of the 1992 Conference on Stormwater and Water Quality Management Modeling. ISBN: 0-87371-898-4

All volumes may be ordered from CHI
fax: 519-489-0695 tel: 800-891-8447 *or* 519-767-0197
email: *info@computationalhydraulics.com*,
web: *www.computationalhydraulics.com*

Preface

Richard Feynman prided himself on being able to devise ways to explain even the most profound ideas to beginning students. "Simple things have elementary demonstrations", he said.* He would surely have approved the intent of this book.

The book summarizes my personal view of the way in which deterministic modeling for urban water systems should be done. It covers the purpose, methods, and use in long-term (meaning ca. 100-year) modeling of environmental and ecosystems impacts for design and research. Further, it provides the background for the 2003 and earlier versions of PCSWMM, a program whose development I have overseen for the past 25 years. Thus the book reviews our recent implementation of optimal complexity, information management, model reliability, error analysis, parameter optimization, model uncertainty, sensitivity analysis and heuristics. Approaches using principles of fuzzy reasoning and genetic algorithms are developed in order to reduce, the amount of computing required. Finally a framework and some simple rules for "responsible" modeling are presented. The book could be sub-titled *towards reliable high-res fuzzy hydrology*.

You might benefit from reading this book if you develop, run or use the results of, any complex, continuous modeling of water resources systems (complex in the sense of a fine spatial resolution, or a fine time-resolution, and implying a large number of environmental parameters).

Graduate work by a number of my past students forms part of this work. They include: Ed Siqueira, Benny Wan, Andy Chan, Tony Kuch, Taymour El Hossieny, Alan Dunn, Peter Nimmrichter, Ron Scheckenberger, Mark Stirrup and Mark Robinson. Other people who deserve mention are Don Waye, for use of his data files and for reviewing an early draft, and Rob James who wrote PCSWMM2003 and earlier versions, and so made these approaches workable. My earlier graduate students who worked on related research include: Jenny Wang, Mike Gregory, Tom Roberts, Ali Unal, Karen Dennison, and Zvi Stifter.

William James, Guelph.

* Goodstein and Goodstein, 1996: *Feynman's lost lecture*, Norton and Co.

Things should be made as simple as possible, but not any simpler.

- Albert Einstein

Contents

Chapter 1

INTRODUCTION - HOW MODELS RESOLVE PROBLEMS

"All models are wrong, though some may be said to be useful."
G.E.Box

It's not enough to merely know when a model may be said to be useful - it's important to know how reliable it is. *WJ*

Models are used to resolve problems.

A problem is *a difficult and doubtful matter requiring solution*; it is something that is *hard to understand, or accomplish, or deal with* (Oxford Dictionary of Current English). Thus *problems* have to do with *complexity*, and since models (the word *model* is synonymous with *concept*) are used to help resolve problems, *models* are inherently tied up with complexity. Throughout this book, we develop general principles of responsible modeling, and they are applied with specific relevance to urban water systems, particularly the decision support system known as PCSWMM. Deterministic models of urban water systems are extremely useful, and their ability to find solutions to certain important problems is now beyond dispute. Applied in planning, engineering design, operation and in settling legal claims, discussion today revolves more around the most valid interpretation of their output, for example their accuracy or uncertainty, than it does about their use in the first place. Written to help in that discourse, this book will hopefully encourage a more responsible use of models.

Many of the modeling methods used in the past have been so unreliable that their continued use now borders on dishonesty. For the design of their interventions in the landscape, engineers have persistently and understandably preferred to use expedient driving input functions, such as short design storms, and simple unit-response functions, such as single-valued unit hydrographs. Such approaches, however, carry very little information beyond the simple estimates of water flows. About the important further impacts of engineers' landscape interventions, these expedient models had little to contribute. Engineers used (and continue to use) them because, until recently, they tended to ignore water quality and ecosystem issues. Another problem with our past modeling practice was our widespread disinterest in applying the concepts of sensitivity analysis, uncertainty, parameter optimization, error analysis, validation, verification, discretization error, disaggregation error, propagated error, structural error, aggregation, best model complexity, model reliability, and the associated use of long-term time series.

This book sets out a heuristic method relevant to ecosystem concerns. Heuristic means that the methodology relates to *exploratory problem-solving techniques, which use self-education ideas* (Webster's 3rd New International Dictionary, 1981 ed.). Evaluation of ranked independent sensitivity and its further analysis to improve performance of a model is an example of a heuristic process.

Material in this book is aimed at students, engineers, planners, and researchers who are interested in reliable, honest and intelligent modeling. It has little or nothing do with simplistic models that anchor users' imaginations to those past, single-purpose practices hinted at above.

> A model is used to help select the best among competing proposals. It is fundamentally irresponsible and unethical for modelers not to interpret the inherent uncertainty of their model output.

1.1 What is a model?

A *model* is a concept (or object) that is used to represent something else. It is reality simplified to a form we can comprehend. Natural reality may be infinitely complex (and chaotic), but a model is a simple view of it. A model helps us to deal with complexity.

There are two cases for which one would not have any need of a model:

1. If you knew everything, you would not need to model anything - you would be able to simply state the solution out aloud and the problem would be solved...
2. If you had all the data that one could imagine, both for the *as-is* and all possible *what-if* or *to-be* scenarios, you could also simply look up a solution in your infinite dataset, without having any recourse to a model...

But neither option is remotely plausible, and so models remain an essential method of approximation for engineering planning and design, especially when designing solutions to complicated problems with arbitrary, 3-D geometry (such as natural topography).

Model-building, problem-solving, critical thinking, engineering design, scientific enquiry, and learning are all very similar processes – all use models, and have a similar formal logical structure, wherein the model improves with experience and intellectual effort.

Mathematical models comprise equations, inequalities, constraints, functions, variables, parameters and constants, and can be programmed and the program executed for a single set of input time series (called a *run*). The resulting model can also be tested for *sensitivit*y, wherein several runs are executed for various changes in one or more input environmental parameters. *Parameter optimization* is an extension of such a systematic series of runs, whereby we endeavor to find the best value of key parameters. These three functions (a run, sensitivity analysis, and calibration) are all that can be done with the models discussed in his book (a fact that should encourage any readers who find the topic daunting).

Deterministic models make use of the limited, known, scientific knowledge and the limited, observed data available. Deterministic models require an understanding of the underlying physics, chemistry, biochemistry, biology and ecology – they "stimulate the little gray cells", rather than allow us to avoid the cerebral discomfort of having to figure it out. Models based simply on observations, however, including many regression, statistical or stochastic models, bypass the need to understand the processes involved. Avoidance of critical thinking is the danger of statistical models. The latter are termed "black boxes" for good reason.

For the record, SWMM and PCSWMM contain routines that are both deterministic and statistical; however we classify them as deterministic because most of the processes follow physical explanations.

Thus in the context of this book, *model* is usually taken to mean *deterministic urban water systems model* (WSM). Strictly speaking, it is what results when a generalized, executable computer code or program has been attached to a specific, hydro-topographic input-data-file of environmental parameters. Once tied directly to a tract of landscape, whether existing (*as-is*), or a proposed development (*to-be*), or a long-past condition (*as-was*), it becomes a model of that tract. The point is significant because it is the data file that determines which processes are to be dominant, deleted, or rendered relatively inactive (dormant). It follows that *only this local model* can be analyzed for sensitivity, uncertainty, parameter optimization, and error (we distinguish between error and uncertainty later).

Deterministic models of urban water systems considered in this book, represent relevant water transport and transformation processes in executable computer code, in which the processes simulated are related to one another according to accepted, scientific laws. Simulated processes may be aggregated (subsumed into fewer, broader processes, with fewer parameters that are less certain) and/or disaggregated (broken down into component processes, with more parameters that are less uncertain). Aggregation and disaggregation (or discretization) are covered in more detail in the next chapter.

> Do not test the generalized program *per se* (absent the input data file) for sensitivity, parameter optimization, or error, because individual applications are likely to be radically different. Values of parameters in the input datafile determine which processes will be dominant or dormant. Relative parameter values change both the model sensitivity and the model uncertainty. Each model application (including the applicable input data file) must be separately tested over the relevant range of model states.

Programs like the U.S.EPA Storm Water Management Model SWMM (Huber and Dickinson, 1988), the U.S.EPA Hydrological Simulation Program in Fortran (HSPF) (Johansen et al., 1984) and the U.S.EPA Water Analysis Simulation Program (WASP) comprise hundreds of files, hundreds of routines, scores of processes, and tens if not hundreds of thousands of lines of source code. Input data files could potentially run to thousands of pages and involve hundreds of thousands of input parameters. WASP, HSPF and SWMM are loosely referred to as *models* rather than *programs* in the literature. SWMM is not a simple, single model, as its name implies, but many separate executable and source-code programs and data files. In this sense, WASP and HSPF, having the word *program* in their acronyms, are more correctly named.

Packages covered in this book have been supported and have enhanced lives over scores of years. Indeed, WSMs are becoming both complex and very long-lived:

- *integrating* over time encyclopedic knowledge of component processes, as knowledge-bases expand;
- *applying* it to large databases describing the physical drainage system, (perhaps covering in GIS detail; for example all of SE Michigan), as the databases build over time;
- *examining* perhaps thousands of arrays of combinations of perhaps hundreds of so-called best management practices (BMPs), as these proposals are put forward from time to time; and
- *re-running* each array for (say) 75 years of hydro-meteorologic data.

Precisely this scenario is being increasingly attempted, and hence model reliability issues are becoming more important than ever. In the next few sections we cover the design and development of deterministic models from basic Newtonian mechanics.

1.2 Deterministic models in design and problem-solving

Model users find the problem-solving experience to be more iterative than linear, like engineering design. In the following 9-step list for a design procedure, many details have been omitted (e.g. the program code is assumed to have been developed):

1. review and re-state the problem
2. construct the *as-is* model input data set
3. select model performance evaluation criteria
4. select an objective function
5. calibrate and evaluate the model
6. satisfied? If no - go back to 1; If yes - continue to 7
7. model several theoretical or *to-be* situations
8. select the likely best alternative
9. report and present the best solution and its uncertainty.

Steps involved in developing a model are also similar to the steps involved in design (or in scientific enquiry, critical thinking, or just learning). It is the methodology of systematic enquiry into (and subsequent explanation of) any complex set of real problems. Your choice of terms may differ from mine, but in my own words, a simplified linear model building sequence may be set out in 12 steps as follows:

1. state the problem
2. identify the significant factors and processes
3. estimate the accuracy required or achievable
4. derive or obtain a suitable mathematical description for each process
5. write the code for these equations and link the codes
6. evaluate the program
7. collect field data
8. develop an *as-is* input dataset
9. evaluate and optimize the model performance by changing model run control and input environmental parameters
10. use the model to infer the behavior of various *to-be* scenarios/solutions
11. select an optimum solution
12. report and present the best solution and its uncertainty

The next section offers more discussion of model development, using a simple pond or containment volume as an example.

1.3 Steps in model construction: Newtonian mechanics, partial differential equations, finite difference equations and computer programs

Deterministic models are developed by following a fairly consistent method of engineering mechanics, for example:

Step 0: Instant spark of genius in the mind of the model builder
Step 1: consideration of the underlying physical laws,
Step 2: approximation by simple partial differential equations (pde's),
Step 3: approximation of the pde's into finite difference equations (fde's), and
Step 4: programming these fde's in a suitable computer language and generating executables.
Step ∞: Debug, improve and explain the code is an endless loop.

As we follow these steps we may find that, like Topsy, the internal mathematical and logical expressions just grow, and become increasingly complex. When building the basic blocks of the model, we usually apply Newtonian mechanics - by which we mean the technique of developing a mathematical relationship derived from one or both of the great principles of conservation of mass, and of energy (or momentum). For a control volume, e.g. an elementary parallelepiped (a small rectangular block), we consider fluxes in and out and the change of quantity within. Scales may vary widely, e.g. from the Great Lakes at one extreme to a laboratory flask at the other.

In the case of a large pond, the inflow is an input hydrograph, supplied from some external analysis or by observation, the outflow is the computed response hydrograph, the surface area is an environmental parameter, given by measurement probably of contour maps. The *state* of the model is given by the instantaneous water levels in the pond, and the computed water surface elevation z is called the *state variable*. *Initial conditions* of the model must be specified at the outset of the calculation, namely the starting depth of the water level in the pond at the start of the simulation. If rain stopped forever, or fell at a constant rate forever, the models would reach a *steady state*. The size or discharge rating curve for the outlets from the pond may be considered to be a *boundary condition*.

Consider the simple example of a pond or water storage unit. By applying the Principle of Conservation of Matter (called the continuity condition or equation) to water flows, we get:

Step 1 In words, for a control volume:

inflow - outflow = rate of change of mass within

This seems to be a simple idea that does not need many words (just eight to express in one short line).

Step 2 In simple mathematical notation:

$$I - O = \frac{dS}{dt}$$

where:

I	=	inflow
O	=	outflow
S	=	volume of water
t	=	time

We assumed that the water is homogeneous, nevertheless the mathematical formulation required a few more lines.

Step 3 For an arbitrary storage area, such as a natural pond:

$$I - O = A(z).\frac{dz}{dt}$$

where:

A	=	horizontal surface area
z	=	elevation of the water surface

and

$$\frac{dS}{dz} = A = f(z)$$

Or in finite differences:

$$0.5(I_1 + I_2) - 0.5(O_1 + O_2) = 0.5(A_1 + A_2)\frac{z_2 - z_1}{dt}$$

where the subscripts 1 and 2 denote the beginning or end of a time step *dt* respectively. In finite differences the algebra has already grown a little more complex, though still an extremely simple equation.

The resulting simple algebraic expressions can usually be re-arranged explicitly or, if implicit, solved iteratively, for an unknown of interest, and programmed, for instance in a spreadsheet. The response function (e.g. the outflow O_2) is subsequently analyzed for the required objective function (e.g. the peak flow O_{max}). The numerical treatment required to do this adds considerably more complexity.

Step 4 Finally the numerical algebra is translated into a programming language such as FORTRAN. In SWMM4, level-pool routing is accomplished in several modules, including STORAGE-TREATMENT. That module, of course, includes dozens of processes other than the central reservoir routing routine. Written in FORTRAN, it is several thousand lines long. In Table 1.1 below, a small part of the code has been abstracted, just two pages (*snip* indicates where large portions of code have been deleted). Though only the few lines that govern the use of the above equation are listed, the complexity of the code is evident. The extracted code is interesting because it illustrates how the SWMM4 source code was structured and annotated (written in 2002). Of course to follow the code in detail, programmers need to examine the coding for the entire module. Obviously, our initial simple statement has now "growed" to many pages.

Table 1.1: A glimpse into the STORAGE-TREATMENT source code (SWMM44GU)

```
      SUBROUTINE UNIT(I)
C     STORAGE/TREATMENT BLOCK
C     CALLED BY CONTRL NEAR LINE 173
C=======================================================================
C     THIS ROUTINE COMPUTES MIXED, PLUG FLOW, AND PARTICLE SIZE REMOVAL.
C
snip
C=======================================================================
C     Here for detention units.
C=======================================================================
snip
C=======================================================================
C     Here, set up SI arrays for rating curve - power function outflows.
C=======================================================================
 2300 IF(IOUT(I).LE.1) THEN
         DO 2310 MM = 1,JINT
         IF(IQRS.GT.0) SAO(MM) = SQQRS(I,MM)*DS/2.0
         IF(IQRS.LE.0) SAO(MM) = SQQOU(I,MM)*DS/2.0
         SAT(MM) =     SST(MM) + SAO(MM)
         IF(SAO(MM).LE.0.0) MIN = MM
 2310    CONTINUE
         ENDIF
C
      IF(QQIN(I).GT.0.0) QRESL(I)=0.0
      STERMS = (QQIN(I)-(QOUTL(I)+QRESL(I))/2.0)*DS+WARN(I)
      IF(IQRS.LE.0.AND.STERMS.LE.SAT(MIN).AND.QOUTL(I).GT.0.0)
     +             STERMS = SAT(MIN)
      IF(IQRS.GT.0.AND.STERMS.LE.SAT(MIN).AND.QRESL(I).GT.0.0)
     +             STERMS  = SAT(MIN)
C=======================================================================
C     Check for exceedance of unit volume.  If so, bypass extra (FLOOD).
C=======================================================================
      IF(STERMS.GT.SAT(JINT)) THEN
         FLOOD = STERMS-SAT(JINT)
         IF(NP.GT.0) THEN
            DO 2410 IP = 1,NP
            PMBY(I,IP) = PMBY(I,IP) +
     +             FLOOD*PMIN(IP)/(QQIN(I)*DS)
```

```
 2410         PMIN(IP)   = PMIN(IP)*
     +                   (1.0-FLOOD/(QQIN(I)*DS))
              ENDIF
            QQBY(I)    = QQBY(I)+FLOOD/DS
            QQIN(I)    = QQIN(I)-FLOOD/DS
            STERMS     = SAT(JINT)
C=================================================================
C#### WCH, 12/5/94.  Keep some statistics on volume-exceedance bypasses.
C     Note, reset JFLOOD counters at end of Sub. CONTRL.
C     Index 1 for monthly, 2 for annual and 3 for total simulation.
C     JFLOOD records every bypass.  JFDAY records days with bypasses.
C=================================================================
            JFLOOD(I,1) = JFLOOD(I,1) + 1
            JFLOOD(I,2) = JFLOOD(I,2) + 1
            JFLOOD(I,3) = JFLOOD(I,3) + 1
            IF(JULDAY.GT.JDLAST(I)) THEN
              JFDAY(I,1) = JFDAY(I,1) + 1
              JFDAY(I,2) = JFDAY(I,2) + 1
              JFDAY(I,3) = JFDAY(I,3) + 1
              JDLAST(I)   = JULDAY
              ENDIF
            IF(JFLOOD(I,3).EQ.1) WRITE(N6,2430)
     +                    I,MONTH,NDAY,NYEAR,JHR,MINUTE
C
C####                     JFLOOD  = JFLOOD + 1
C####                     IF(JFLOOD.EQ.1) WRITE(N6,2430)
C####    +                     I,MONTH,NDAY,NYEAR,JHR,MINUTE
            ENDIF
        IF(STERMS.LE.0.0001) STERMS = 0.0
C
        IF(IOUT(I).GE.2.AND.IQRS.LE.0) GO TO 2510
C=================================================================
C     Interpolate knowing RHS of continuity eqn. ==> storage-indication
C     routing.
C=================================================================
        CALL INTERP(SAT,SAO,JINT,STERMS,SO2DT2)
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
 2510             STORE    = STERMS-SO2DT2
        IF(STORE.LE.0.001)    STORE    = 0.0
        IF(STORE.GT.VMAXS(I)) VMAXS(I) = STORE
        IF(IQRS.LE.0) THEN
C==================================================================
C     Calculation of treated outflow rate.
C==================================================================
            QOUT    = SO2DT2*2.0/DS
            QQOU(I) = QQIN(I)+(WARN(I)-STORE)/DS
            IF(QQOU(I).LE.0.0001) QQOU(I) = 0.0
            ELSE
C==================================================================
C     Calculation of residual outflow rate.
C==================================================================
            QRES    = SO2DT2*2.0/DS
            QQRS(I) = (WARN(I)-STORE)/DS
            IF(QQRS(I).LE.0.0001) QQRS(I) = 0.0
            ENDIF
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
        CALL INTERP(SST,SAR,JINT,STORE,AREA)
C==================================================================
```

```
C    Compute evaporation loss.
C====================================================================
     IF(METRIC.EQ.1.AND.STORE.GT.0.0) QQEV(I)=AREA*E(MONTH)/1036800.0
     IF(METRIC.EQ.2.AND.STORE.GT.0.0) QQEV(I)=AREA*E(MONTH)/8.64E7
     STORE1 = STORE - QQEV(I)*DS
     IF(STORE1.LE.0.0) THEN
        QQEV(I) = STORE/DS
        STORE1  = 0.0
        ENDIF
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
     CALL INTERP(SST,SDE,JINT,STORE1,DEPTH)
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
 2700 IF(NP.LE.0)   GO TO 5000
     IF(IROUTE(I)) 4000,4000,3000
C====================================================================
snip
```

Following development of the source code, the program is debugged and compiled into executables and then linked into an executable (.EXE) program that can be run on certain hardware and operating systems, often a task in itself. When run, the program computes a solution that steps out in time, a class of problem referred to as *initial value problems*. At each step there is an iteration to solve the equations.

Summary: We have conveniently dodged the vast subject of numerical analysis, which fits between Step 3, derivation of the fde's, and Step 4, development of the executable program. Nevertheless the above description describes how our models are typically built. It is a creative endeavor often fraught with error from beginning to end, and results in very long codes (SWMM44GU has about 50,000 lines of code including 15,000 comment lines, while the shell PCSWMMGIS2002 alone probably runs to over 200,000 lines).

The creative art in programming models is not well appreciated by the public, who may consider engineers to be pedestrian and unimaginative. In the long, dry code, that initial, conceptual, brief spark of creative genius is key.

1.4 When to use a model

Models are used to supply or fill in missing data or information. As a design tool, a WSM for example usually provides credible information about the relative performance of alternate arrangements of storm water management facilities. Analysis of the information provides knowledge of the various systems. As stated earlier, if the engineering community had innate knowledge of the long-term environmental performance of proposed large, complex, urban drainage and pollutant removal systems (and if the engineers were infallible), then modeling would of course be unnecessary. But urban storm water drainage systems are sometimes both extensive, incorporating more conveyances and devices than can be comprehended, and sometimes exceedingly complicated, starting with atmospheric pollutant and storm processes at the top and ending with effluents mixing in creeks and at lake outfalls at the bottom, perhaps hundreds of kilometers away. Furthermore, there are long-term water quality and environmental impacts that are difficult to predict, such as water quality violations at various places in the drainage system, or the gradual displacement of cold-water aquatic biosystems at various places.

For the very large number of calculations needed to deal with such complexity, large-scale deterministic surface water quality models are better (in the sense of being more reliable) than manual methods. Thus decision makers routinely deal with increasingly complex problems of water management by means of WSMs. Critics however feel that some water managers may believe that the outputs from their WSMs are perfectly reliable. Other critics, on the other hand, worry more about the interpretation of the model output, i.e. the claims made on behalf of WSMs, than they do about the structure or processes of the WSMs themselves. Bergstrom (1991) calls this faulty interpretation *unethical modeling*. Since all models are, in a sense, wrong, modelers need to clearly report the reliability of their computed results. We should always be concerned about the *interpretation* of the model output.

Report the uncertainty, first and foremost.

> Computed and observed time series are more ethically represented as smudges than single-valued lines.

1.5 How a model relates to design problems

Between

1. design problems,
2. model uncertainty and
3. output interpretation,

there exist fuzzy relationships that are the central concern of this book. This book provides simple rules and robust methods are provided for reliable, long-term, deterministic modeling of optimal complexity.

We start with a clear statement of the design problems to be addressed by the modeling exercise, and end with a meticulous interpretation of the model complexity and results, their uncertainty, and their relationship to the initial design questions.

Stated in the most general form, the design problem may be formulated as: *find the optimum cost-effective array of water management solutions to solve* *certain design problems*. Problems may at the outset be formulated as a list of specific, narrow questions (James and Robinson, 1981 a, b), each with a question mark, e.g.:

- *what is the cheapest, or smallest, stormwater sewer or conveyance necessary to convey flows from Foxran Developments?*
- *what is the cheapest, or smallest geometry, culvert required at Cross St.?*
- *what will be the mean annual flooding and pollution damages (costs) downstream as a result of the Foxran Developments?*
- *what size, or volume, of detention storage is necessary at Cross St. to ensure that there will be no net increase in flux of stormwater or associated pollutants into Mud Creek as a result of the Foxran Estates development?*
- *what capacity of pump station is required to eliminate all combined sewer overflows to Parkside Beach from the Cross St interceptor?*
- *what is the maximum quantity, or flow rate, of sewage treatment plant effluent, or stormwater, that can be added to Mud Creek at Cross St. during low flows in July, without violating federal and local water quality objectives?*

Such questions help us to define the *design objectives*. Other conventional questions in typical engineering applications are suggested in James (1993b); suggestions for questions related to eco-sensitive design are given later in this book.

Concise and precise questions should be formally established by co-operative discussions between the client, engineering committee, and the modelers. When design questions are formulated simply in this manner, it is much easier to relate the design problems to the required model output, to select the best objective functions and evaluation criteria, and to provide quantitative solutions with appropriate uncertainty. When the design objectives are imprecise, fuzzy approaches may be better. Examples of fuzziness in design are given later.

> Design objectives must be simplified and related to the computed output and objective functions. The model must include code that adequately describes all significant processes.

For design, it has become common to integrate many existing, related, process codes into large executables, and to apply these programs to the *as-is* (existing) scenario (often a rural landscape), *to-be* (proposed) scenarios or urban landscape interventions, and the *as-was* (original pre-development) scenario or natural landscape. When estimating what the original pre-development flows would have been, it is of course impossible to calibrate the model for the bygone conditions. *To-be* scenarios similarly cannot be directly calibrated. While the *as-is* scenarios can be calibrated, the *to-be* and *as-was* must be inferred. Using these three different sets of scenarios, the best-informed decisions can be made.

Unfortunately, where no suitable code is available (e.g., for a novel, proposed BMP), model users must often resort to "fudging" the input for a similar, but existing, process to approximate the new processes, so far as they are known or can be guessed. This is where output interpretation becomes especially problematic, and where model critics may successfully impute unreliability into the whole modeling exercise. Rather than “fudging it” it would be better of course to use a model that already incorporates the required, specific code, or to write it from scratch.

For sustainability concerns, long-term modeling, of the order of 50 to 100 years, say three generations, is more credible. Several writers (Gore, 1992; Brundtland, 1987) contend that society understands and is concerned about changes over three generations. Families have a good memory of this length, because grandparents talk to grandchildren. Three-generations-modeling (3GM) can be used to stress the importance of sustainability. Of course not all processes can be reliably modeled over this time span, but urbanization, sediment accumulation in storages, operating and maintenance costs, and fluvial morphology, are all examples of important processes that can be usefully estimated over such long terms.

1.6 Concluding remarks

Continuous, long-term modeling requires informed management and reporting of the model uncertainty. Interpretation of the model output is of concern to everyone touched by the consequences of the model use. Often, deterministic models *per se* are condemned because of wild claims made by users. Interpretation of all computed responses should be assiduously linked to the original design problems, as well as the uncertainty.

In fact, it boils down to a question of professional ethics and responsibility.

Surface water quality models and computer systems have progressed to the point where modeling for a period of three generations (3GM) is a serious design option. Over time, open source WSMs like SWMM have more and more code appended, for example dealing with BMPs designed to meet new and emerging storm water management practices and policies. Sometimes it is hard to keep up-to-date. User-group list servers and meetings are a vital way of keeping informed about modeling advances contributed by various subscribers, such as additional code. Readers are encouraged to subscribe to relevant list servers, like SWMM-USERS, and to attend the regular Urban Water Systems (formerly the Storm water and Water Quality Management) modeling conferences that are held annually both in Canada and the U.S.A. For more details, see: *www.chi.on.ca/confsem.html*.

In the next chapter we proceed to describe spatial discretization and temporal resolution. Disaggregation is covered by describing its complement – aggregation. Concepts of model complexity are also introduced.

References

Bergstrom, S. 1991. Principles and confidence in hydrological modelling. Nordic Hydrology, 22: 123-136.

Brundtland, H., 1987. Our Common Future - World Commission on Environment and Development, United Nations Brundtland Commission 1987.

Gore, A., 1992. Earth in the Balance - Ecology and the Human Spirit. Houghton Mifflin Company, New York, New York, 1992, 408 pp.

Huber, W.C. and Dickinson, R.E 1988. Storm water management model user's manual, version 4. EPA/600/3-88/001a. (NTIS PB88-236641/AS) U.S.EPA, Athens, Ga.

James, W. 1993b Stormwater Management Modelling Conceptual Workbook. Pub. by CHI Guelph. 200 pp.

Johanson, R.C., Imhoff, J.C., Kittle, J.L., and Donigian, A.S. 1984. Hydrologic simulation program - FORTRAN (HSPF): User manual. U.S.EPA. Report EPA/600/9-80-015.

Bibliography

Andersson, Lotta. 1992. Improvements of runoff models - what way to go? Nordic Hydrology, 23:315-332.

Crawford, N. 1993. The next major breakthrough... Paper presented at the U.S.EPA workshop on Developing dynamic watershed modelling capabilities for the Great Lakes tributaries. Heidelberg College, Tiffin, Oh. August 4-6, 1993. 8pp.

Huber, W.C. 1993. Methods vs. parameters: the importance of parameter estimation and sensitivity analysis in hydrology. In: Water resources education: a lifetime of learning. AWRA. 1993 pp151-158.

James, W., 1993a Need for 3-generation stormwater quality impacts models. USEPA Invited Conference on Model-building for Great Lakes Water Quality problems of the 2000's. Heidelberg College, Tiffin, OH. Aug 4-6. *unpublished; copy available from author*

Klemes, V.1987. Empirical and causal models in hydrologic reliability analysis. Procs NATO Institute on Engineering reliability and risk in water resources, Tucson, Ariz. May19-Jun2. 1985. ISBN 90-247-3492-4. pp391-403.

Orlob, G.T. 1992. Water-quality modelling for decision making. J. Water Resources Planning and Management. 118(3): 295-307.

Thompson, D.B. and Westphal, J.A. 1989. Tilt at the windmill of runoff simulation. Proc of the Int Conf on catchment runoff, U of Virginia, Charlottesville. 1989. pp176-185.

Chapter 2

HOW TO DISCRETIZE AND DISAGGREGATE, AND MORE ON COMPLEXITY

Nature has some sort of arithmetical-geometrical co-ordinate system, because nature has all kinds of models. What we experience of nature is in models, and all of nature's models are so beautiful.

Universe to each must be
All that is, including me.
Environment in turn must be
All that is, excepting me.

Thinking is momentary dismissal of irrelevancies.

- R. Buckminster Fuller

2.1 Introduction

Subjects for this chapter include consideration of *spatial and temporal discretization*, and *process disaggregation*. Both concerns are closely tied to the concept of model complexity and the development of input data files. Spatial discretization is the term preferred in urban water systems modeling, and is known in some other types of model as compartmentalization. It is closely allied with the graphical management of spatial data (or geographical information systems, GIS), which is one essence of urban water systems modeling. The other essence, graphical time series management system, is discussed later. We introduce all these issues here.

2.2 The procedure of discretization

Basic connectivity of the drainage system must be replicated, and this is the most important consideration for model discretization. Simply modeling the connectivity itself carries significant information. By methodically analyzing the system, and having the constituent lags systematically proportioned in relative terms, the drainage connectivity itself ensures that hydrographs and pollutographs are properly structured. Volumes of pollutants and water are generated, transported, changed and arrive at the outlet in the correct sequence.

Let us consider the relative representation of a quantity in space, such as elevation or gray-scale for example the four images in Figure 2.1. Each is from the same original (aspect ratio 5:6) and they are rendered in four levels of spatial detail, approximately $4x10^4$, $1.6x10^3$, $4x10^2$ and $1x10^2$ sub-spaces respectively, or 400, 16, and 4 times the detail of the simplest. It is the relative values of intensity of each pixel that provide the clearly understood image - absolute values are less important. If rendered in a TV signal, the pixels would be scanned line by line and the signal transmitted as a linear time series. Similarly, if the sub-spaces were sub-catchments, the runoff flow hydrograph from a uniform stationary storm would be formed in an analogous time series. Sensitivity analysis shows that increasing the spatial resolution itself accounts more and more for the general shape of the time series, without being influenced by the absolute values of the parameters. In other words the sub-space intensity becomes systematically less important – its relative sensitivity decreases with complexity.

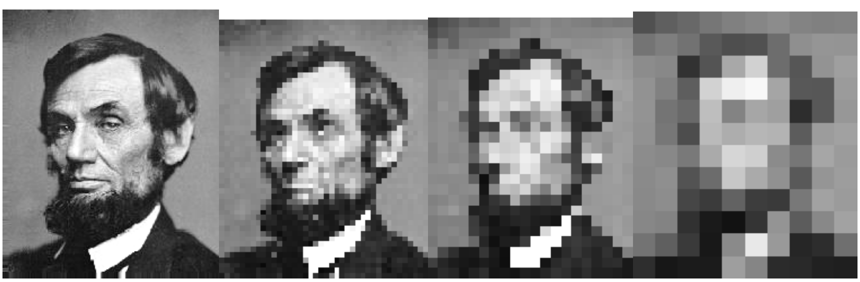


Figure 2.1 From left to right, number of horizontal pixels or uniform sub-spaces is 192, 40, 20 and 10. If you study the pictures while crunching up your eyes, two become clearer, and the two extremes are unaffected, or become less clear.

When choosing the number of elements incorporated into a water systems model, the most significant elements should first be identified, perhaps through a site inspection. Significant elements may include, for example, dams, ponds, storage tanks, pump wells and pump stations, the largest diameter conveyances, the tributaries, diversions, outfalls, weirs and gates. These are elements that are expected to significantly impact the computed response. In later iterations, elements at the next (lower) level of significance may be chosen. Listed in the tables below in mind-numbing detail are various elements for urban water systems.

Concepts of discretization are explained in very general terms in this section. Consider a system of connected great lakes which have rain as input and stream flow as output, and whose water levels change from time to time. Various models can be built at various resolutions, for example:

Simplest model: One could consider a simple model in which all of the North American Great Lakes act as one single tank (call this model GL1; here *dx, dy* would be of the order of 1000 km or more).

Less simple model: Or one could consider a networked system of six lakes, linked sequentially: Lake Superior flows into Lakes Michigan and Lake Huron/Georgian Bay, which flow into Lake St. Clair, which flows into Lake Erie, which in turn flows into Lake Ontario, which finally outflows at Kingston, Ontario - call this model GL6; now *dx, dy* are about 100 km. We could argue that GL6 is six times more complex than GL1. Adding Lake Nipigon and a few other large lakes would soon bring us up to GL10, 10^1 times more complex.

More detailed model: Now consider modeling a thousand river reaches, creeks, drains, channels and storm sewer pipes in each part of each one of the Great Lakes' sub-catchments (for example in the city of Guelph). This more detailed sub-model may be denoted GL1K, and *dx* will now be about 100 m. Such a model could be said to be 10^3 times more complex than GL1. If 1000 such sub-models were structured together in the framework of the Great lakes, one would soon create GL1M, 10^6 times more complex than GL1. Incidentally such GL1K models normally comprise a dendritic (tree-structured) network like GL6 above, but occasionally some of the drains are cross-connected, and the drainage system is then said to be loop-structured.

Discretization of the study problem proceeds by ensuring easy parameterization, which requires prior knowledge of the science underlying the model (i.e. the constitutive equations). In other words, we are searching for spatially averaged parameters required for the mathematical equations in the model. It is easier to average the parameters if the sub-space is homogeneous with respect to the particular set of parameters. Hence, uniform sub-spaces are chosen. This approach favors discretization into subspaces already characterized, for example in a GIS or assets management database. Homogenous land-uses are an example of a basis for discretization for simulating the build-up of surface particulates. If the sub-catchment being “parametized” is heterogeneous, modelers or model builders may resort to area-weighted calculations to determine the spatially-averaged land use.

A description of how to discretize is an essential starting point for novice modelers. Start out by working with soft pencil on a large paper print of a contour map or drawing of the area, or equivalently on your monitor using AutoCad or GIS programs.

1. First mark the location of the study problem in the drainage channel,
2. then demarcate the overall divide,
3. sketch in the field of surface overland flow vectors within that watershed,
4. proceed to mark all significant channels,

5. mark major hydrological components (make a field inspection),
6. delineate sub-catchments for each tributary, potential storage unit, flow gauging station, and other potential BMPs, and
7. build a schematic of the connections between all the identified elements. Later when the input data file is constructed, and a graphic schematic of the connected system taken from the data file, the latter should be compared visually against the original contour drawings.

> The drainage connectivity must be correctly replicated. Improving the spatial resolution alone informs the computed response and reduces the sensitivity to parameter uncertainty.

In the previous chapter we covered the development of deterministic models from basic Newtonian mechanics, and introduced the concepts of spatial discretization and temporal resolution. In the next few sections we further describe disaggregation by its complement, aggregation, and introduce concepts of model complexity.

2.3 Introduction to process disaggregation

Disaggregation means adding more high speed processes (i.e. processes that are characterized by shorter reaction times or periodicities, such as the passage of a storm cell over a sub-catchment, for example). Dissaggregation is thus a by-product of finer spatial resolution, since finer spatial resolution requires a smaller time step and thus further (higher-speed) processes to be activated. Time scale is related to the spatial scale through the impulse response function (for example, an instantaneous unit hydrograph for a given sub-catchment). For instance, if the impulse response function lasts 60 minutes, and 20 points are required to define it sufficiently accurately, then a time step of 3 minutes is indicated. (The discretization error is indicated by the maximum vertical intercept over a 1.5 minute time step, assuming central differences.)

Another type of process addition by finer discretization is exemplified by rain storm dynamics. When only one rain gage and one catchment is involved, the storm appears to be an areally uniformly distributed hyetograph from a stationary storm. When several gages and sub-catchments are involved, however, the storm appears at different times across the catchment. In this way a storm is seen to move dynamically across the model catchment. If moving down the drainage system, the storm may track synchronously with the gathering flood wave. In such a case the flood wave is dynamically amplified. It would be attenuated if the storm tracks in the reverse direction, up the drainage system. Thus increased discretization has added a new process: dynamic storm and flood synchronization. Even more interesting, for certain drainage planforms, a storm comprising several rain cells could theoretically resonate with the drainage dynamics.

Many modelers advocate model parsimony, meaning that one should consider the simplest useful combination of processes and sub-spaces. Aggregation of sub-spaces such as pipes and conveyances into fewer, longer units can be facilitated by GIS utilities such as PCSWMMGIS, and proceeds according to chosen rules. Aggregation of processes in models such as SWMM is accomplished by deleting a process and then re-calibrating a remaining process to account for the now unexplained process. Usually process aggregation will increase the variance between computed and observed responses and vice versa.

It is a fundamental tenet of this book that the variance can be reduced by systematically explaining or including more and more relevant processes.

2.4 Introduction to temporal resolution

Choosing a coarse spatial resolution determines to some extent the time step, through the speed of the processes involved. As indicated above, the higher the spatial resolution, the faster the frequency of the processes involved, and the finer the time resolution (dt) should be.

For example, in the Great Lakes system, outflow varies only slightly on a weekly basis. Lake levels respond to seasons and to cycles of years, i.e. a series of wetter years. Coarse models consider only longer frequency or slow, damped processes such as evaporation. Individual storms have no measurable impact. Hence we could build a GL1 model with dt = 1 month.

On the other hand, small gutters and pipes in Guelph respond strongly to short thunderstorms, even if the storm lasts only 10 minutes. Here we could build a GL1M model with dt = 1 minute. Flow in a surcharged pipe is a higher frequency process that must be modeled in GL1M with dt = 1 to 10 s. Choosing the time step controls the processes that will be involved. E.g. little ripples on the surface of the water have a natural frequency of about 10 Hz. They would be averaged to zero if dt = 1 minute or more. Similarly, a short sharp thunderstorm would have a negligible rain intensity if averaged out over a month. We will have much more to say about this later.

2.5 More introductory discussion of model complexity

From our previous discussion, without any formal proof, we established by simple definition that models that require fewer input parameters are inherently less complex - in other words, the number of input variables in the input data file is indicative of the model complexity. This is not a widely accepted view of complexity, but it will suffice for the meantime.

We are developing rules for responsible modeling for urban water resources systems. Now let us define a few more terms.

A *system* is taken to mean a representation of a situation, and comprises an assembly of *elements* arranged in an organized whole. A water resources system is an assembly of engineered and natural physical elements such as ponds, channels, creeks, dams, pumps, weirs, diversions, and the like. Comprehensive lists of elements are available on my website at:

/-- /homepage/Teaching/661/urban_water_infrastructure.htm
where /--/ denotes: http://www.eos.uoguelph.ca/webfiles/james

Listed in tables in the next few pages are many elements that will not normally be modeled as separate elements. However they are listed here, not to frighten or stun the reader, but rather to give an idea of their formidable variety. From these tables one could proceed to methodically build a schematic of the drainage system. Sources of the information are provided in the tables, which are entitled:

Table 2.1: General introductory list
Table 2.2: Conventional water supply systems: water supply wells, dams, water distribution systems, water storage, fire control, water quality in distribution networks, network analysis, transients.
Table 2.3: Conventional sanitary collection systems: usual infrastructure, pump stations, water quality transport modeling.
Table 2.4: Conventional storm water systems: normal storm water infrastructure, best management practices.

Some of the tabulated elements may be represented in your model by sub-models containing code for the processes relevant to that element. All elements - as represented by process code - are *concepts built on assumptions*. Often the underlying assumptions and their uncertainty are not readily available in the model documentation.

Representation of a process is described by a simple noun phrase. Model users have come to depend more upon the presence or otherwise of the phrase, than knowledge of its fundamental uncertainty. Nevertheless, its existence, informed observers agree, is worth assuming in order to gain necessary insight. Behavior of the chosen process should cause a change in a significant attribute of your overall model. If it does not, it usually is reasonable to switch off the process in the model by inserting a zero for one or more parameters; this is sometimes called *zeroing out* insensitive processes.

Component processes require specific input parameters; the more detailed the process description, the more parameters required. Inspection of the parameters active in the input data file informs the user about the processes simulated in the model. But it does not provide full details of the internal *structure* of the model. For instance, while many processes act independently, many others are interdependent. A relationship is said to exist between X and Y if the behavior of either is influenced by the other. Such relationships may involve the flow of materials, information or energy.

Table 2.1: General introductory list.
[the very large array of infrastructure used in urban water systems]

1. Water supply sources	Artesian well Pumping well Bored, clear, and drilled wells
2. Intakes	Submerged crib intake Screened pipe intake Reservoir and Lake intakes
3. Storage	Reservoirs and dams Elevated storage tanks (steel and concrete) Ground level storage tank Fire and emergency storage
4. Pumps	High lift pump Low lift pump Pumphouses
5. Chambers	Air relief chambers Manholes Valve chambers Valve boxes and stems
6. Anchors	Thrust Restraint Thrust block Friction anchor
7. Pipes – usage types	Sanitary sewer Storm sewer Forcemains Watermain
- materials	Asbestos cement pipe Cast iron pipe (un) lined Concrete pipe Ductile iron pipe (un) lined Flexible and rigid pipes Milled steel (un) lined Polyvinyl chloride pipe
- operational types	Heated pipelines Insulated pipes
8. Joints	Cast coupling Flanged joints Restrained joints Service connections Victaulic coupling Welded joints
9. Fittings	Bends Tees Crosses Reducers

	Couplings Sleeves Vents Hydrants Water meters
10. Valves	Air relief valve Ball valve Butterfly valve Control valve Gate valve Drain and scour valves Pressure control valves Check and reflux valves Foot valves Tide gates and valves

Table 2.2. Conventional water supply systems
[water supply wells and dams, water distribution systems including network analysis and water quality issues, water storage tanks and reservoirs, fire control devices, analysis and control of transients.]

*1 From *Pressure Pipeline Design for Water …*	*2. From *Water Supply and Sewerage:*
Air chamber	Air piping
Air release valve	Aqueduct
Asbestos cement pipe	Aquiclude
Bends	Aquitard
Blow-off hydrants	Artesian well
Cast iron pipe	Backflow preventer
Cavitation	Double check valve
Centrifugal pumps	Belt filter
Check valves	Blowoff branch
Concrete pressure pipe	Broad crested weir
Concrete thrust block	Chlorinator
Continuous pipelines	Clarifier
Control valves	Clear well
Corrosion protection	Compressor, centrifugal
Dead-end pipes	Dam
Displacement pumps	Diffuser
Distribution pipeline blow-offs	Distribution system
Dynamic pumps	Dosing tank
Excess flow activated valves	Filter
Fiberglass pipe	Fire hydrant
Fire hydrants	Fire pump
In-line surge suppression	Gooseneck
Isolation valves	Joints
Non-slam check valves	Mains
Pipe casings	Membrane filter
Pipe	Meter
Pipe fittings	Motor pumper
Pipe joints	Orifice plate
Pipe materials	Pipe
Pipelines	asbestos cement; cement lining;
Polyethylene pipe	corrugated metal; cast iron;
Polyolefin pipe	plastic; steel; threaded
Polyvinyl chloride pressure pipe	Pipe joints
Pump station	ball; bell and spigot
Pumps	Pump
Relief valves	Pumping station
Steel water pipe	Reservoir

Surge anticipation valve	Siphon
Surge relief valve	Valves
Surge tank	air relief; air-vacuum;
Surge valve	backflow preventing; box; butterfly;
Thrust block	check; foot; gate; vacuum breaker;
Thrust restraint	Venturi
Seismic valve	Weirs
Vented surge tank	Contracted weir
Water Hammer	Rettger weir
Water meter	Sharp crested weir
	Submerged weir
	Suppressed weir
	Triangular weir
	V-notch weir

*Two books were used to list components of water supply infrastructure:
1. American Society of Civil Engineers, 1992. Pressure Pipeline Design for Water and Wastewater. American Society of Civil Engineers, New York.
2. McGhee, Terrence J. 1991. Water Supply and Sewerage. McGraw-Hill Inc., New York.

Table 2.3. Conventional sanitary collection systems
[usual infrastructure, pump stations, water quality transport modeling.]

The list was taken from: American Society of Civil Engineers, 1982. Gravity Sanitary Sewer Design and Construction. American Society of Civil Engineers, New York.

Air jumper
Asbestos cement pipe
Backflow preventer
Backwater gate
Bends
Bituminous pipe joints
Cast iron pipe
Cement mortar pipe joints
Check valve
Concrete arches
Concrete pipe
Conduits (shape, material, size)
Corrugated metal pipe
Drop manholes
Ductile iron pipe
Elastomeric sealing compound
Encased pipe
Flap gates
Flexible pipe
Float well
Gasket pipe joints
Hydraulic jumps
Junction
Linings
Manhole
Mastic pipe joints
Metering devices
Open channel
Pipe
Plastic sewer pipe
Polyethylene pipe
Polyvinyl chloride pipe

Precast concrete tunnel lining
Push-on pipe joint
Reinforced plastic mortar
Rigid pipes
Sealing band joints
Shallow manhole
Siphon
Steel pipe
Thermoset plastic pipe
Thermoplastic pipe
Weirs

Table 2.4. Conventional stormwater systems
[normal stormwater infrastructure, best management practices]

*1. From *Municipal Storm Water Management*:	*2. From *Design and Construction of Urban Stormwater Management Systems:*	*3. From *Design and Construction of Sanitary and Storm Sewers*:
Baffled outlets	Acrylonitrile-butadiene-styrene pipe	Air jumpers
Best management practices	Apron	Asbestos cement pipe
Catch basin inserts	Artificial channels	Bar racks
Channels	Asbestos cement pipe	Bends
Combination curb and gutter	Baffle pier	Cast iron pipe
Combination inlet	Box inlet drop structures	Catch basin
Culverts	Cast iron pipe	Channel
Curb and gutter	Cement mortar pipe joint	Check valve
Drop inlet	Centrifugal pump	Clay pipe
Filter strip/ flow spreader	Channel lining	Cleanouts
Flume	Channels	Concrete pipe
Grate inlet	Check dam	Corrugated metal pipe
Gutter	Chute block	Culverts
Hydraulic jump	Cleanout structure	Ductile iron pipe
Improved inlets	Combination inlets	Ejector
Infiltration trenches	Concrete arch	Flap gate
Inlets	Concrete cradle	Flexible pipe
Oil/grit separators	Conduits	Force main
Orifice meters and nozzles	Corrugated aluminum pipe	Hydraulic jump
Shoulder gutters	Corrugated steel pipe	Inlet
Sills	Culverts	Joints
Slotted drain inlet	Curb inlets	Junctions
Venturi meter tubes	Detention basins/ponds	Leaping weirs
Weirs	Drop inlet	Manhole
broad crested	Ductile iron pipe	Meter
inverted triangle	Encased pipe	Open Channel
ogee shapes	Filter	Overflow weir
sharp crested	Flap gate	Plastic pipe
trapezoidal	Float operated gate	Pump
V-notch	Flow splitter	Centrifugal
	Gasket pipe joints	Non-clogging
	Grate inlet	Relief overflow
	Head wall	Retention basin
	Inlets	Siphon
	Junction box	Steel pipe
	Leaping weirs	Suction piping
	Lining	Tide gate

	Manholes	Venturi
	Open channel drops	Weirs .
	Open channels	
	Outlets	
	Overflow structures	
	Pipe	
	Pipe joints	
	Pipe materials	
	Plastic sewer pipe	
	Porous check dams	
	Pressure conduits	
	Rigid pipe	
	Riprap	
	Screw pumps	
	Side overflow weirs	
	Siphon	
	Slotted drain inlets	
	Spillway	
	Storm drain	
	Swales	
	Thermoplastic & thermoset pipe	
	Tipping gate regulator	
	Trash racks	
	Trench box	
	Vertical pumps	
	Volute pumps	
	Vitrified clay pipes	
	Weirs	

*Three textbooks were used to list components of water supply infrastructure:
1. Debo, Thomas N. and Reese, Andrew J., 1995. Municipal Storm Water Management. Lewis Publishers, London.
2. American Society of Civil Engineers, 1992, Design and Construction of Urban Stormwater Management Systems. American Society of Civil Engineers, New York.
3. American Society of Civil Engineers. 1986. Design and Construction of Sanitary and Storm Sewers, New York.

Some attributes of the processes are called *state variables*, and their relationships may be plotted in state space. The water level in a dam or in the ground indicates the amount of moisture currently stored in the watershed, and is computed as the simulation proceeds. Output from the model may be a time series of the outflow hydrograph. Water level and water storage are candidate examples of state variables. If the state variables map on a one-to-one or simple basis, the system is said to be deterministic.

Structure is the term used to describe how the component processes can be related to each other. If the system attains a steady-state or dynamic equilibrium the state is termed *homeostasis*. If the system is complex and the system is fuzzy, it is said to be poorly structured. The concept of the tendency for a system to move toward greater disorder is termed *entropy*, and this includes the tendency of a simple, organized or engineered system to break down and revert to a natural system whose components are characterized by infinitely complex inter-relationships. A system that comprises interacting processes that is difficult to appreciate as a whole has been technically defined as a *mess*. Here the terms *disorder* and *mess* are emotionally toned, perhaps unfortunately, for they do not seem to reflect an appreciation of the natural order.

A *black box* model lumps all the systems processes into a single transfer function, not based on the underlying physics. Models that we deal with here, (e.g. SWMM, EPANET, HSPF and WASP) are generally deterministic, nevertheless some of their sub-models are empirical, and act as black boxes. Many of the above elements may have empirical or pseudo-empirical equations, which act as black boxes. Take for instance the case of the so-called rational formula (dating from the late 19th century) for peak flow

$$Q_p = ciA$$

where

c = runoff coefficient, an empirical factor
i = rainfall intensity
A = sub-catchment surface area.

This equation has no basis in physics, and it is not possible to create or obtain a surface for which it may be applicable – the runoff coefficient c subsumes a myriad of unexplained processes. Nevertheless, probably because of its simplicity and usefulness, c has been the subject of acute and chronic investigation, and many recommendations have been published, of which several were gathered and plotted in Figure 2.2 by R Pecher.

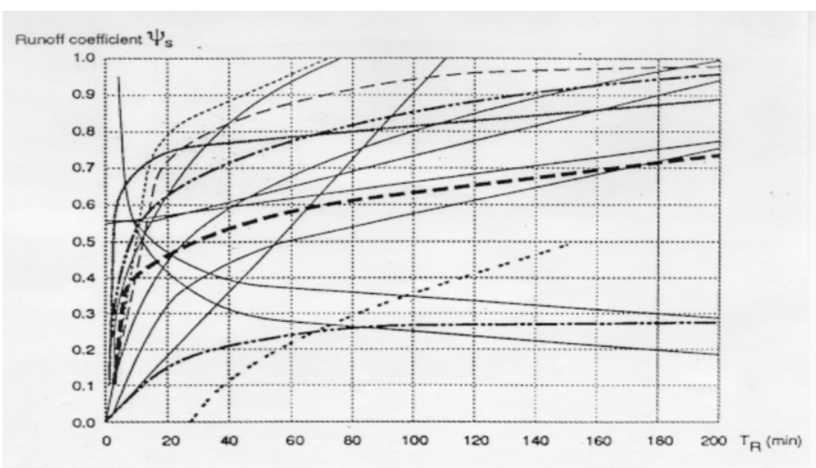


Figure 2.2 Runoff coefficients for the so-called rational formula.

He searched the literature for and obtained relationships between the runoff coefficient and the runoff duration from 15 different sources, plotting them as Figure 1 on p. 5 of his dissertation "Der Abflussbeiwert und seine abhägigkeit von der Regendauer", (Berichte aus de Institut für Wasserwirtschaft und Gesundheits-ingenieurwesen Technische Hochschule München Nr. 2, 1969, ca 200pp.). What conclusion can be drawn from Figure 2.2, other than that *c* varies widely and unpredictably with duration? Surely, that the given relationship is poorly structured, or fuzzy, perhaps even a mess, since, in the absence of a better explanation of the processes involved, evidently *any value is as good as any other*.

Reporting this variance of *c* in those models that use it has to my knowledge never been done. It seems to be virtually unknown in the English literature on the subject. In this sense a simplistic sub-model increases a difficulty when using the model, because its mathematical formulation is likely incorrect physically, and evidently meaningful values of the empirical coefficient *c* cannot be acceptably fudged. Ideally, the inherent scatter in all sub-models should be incorporated in the computations in the model, and reported (the uncertainty can probably be uncovered in the original publications).

Complexity arises when there is a large number of processes, when the relationships are non-linear, and where there are *nonholonomic* constraints (e.g. if a pump station in a drainage system switches on or off without regard to the system as a whole).

We may think of model complexity C as the sum

$$C \equiv \sum_{m=1}^{N_M} \sum_{s=1}^{N_S} \sum_{p_r=1}^{N_{pr}} Np_{a_{p,s,m}}$$

where:

N_m = the number of modules active in the model,
N_s = no. of sub-spaces modeled in each module,
N_{pr}= no. of processes modeled in each sub-space,
Np_a = no. of input parameters required for each process.

> Model complexity is related to the total number of uncertain input parameters.

C is independent of the number of time steps in the input time series (e.g. rainfall or inflows) that drives the system model. For instance, a rain time series may be 100 y long at 1 minute time steps, comprising about $3.15\text{x}10^9$ time steps, while the system model itself may have just six elements each requiring (say) ten parameters to describe their geometry. Such a system model may be said to have a complexity of 60, a relatively small number, even though the driving input function is extremely large.

There is another hierarchy of complexity that depends upon how well-known the processes are, and how they depend upon earlier levels of understanding. In this approach, physical systems are the simplest. SWMM hardly rises above the incorporation of purely physical concepts. Chemical systems are more complex since the underlying physics must first be correctly modeled. HSPF is an example. Biological (and human-medical) systems are yet more complex, depending as they do upon correct chemistry. WASP is a simple example of this class of model. Even more complex are problems of animal behavior (or human psychology), which depend on correct biological models, and most complex in this hierarchy are ecological and sociological systems. Aesthetics, which may be defined as the many auditory and visual perceptions that accompany a certain feeling of value, is an example of a complex system. We do not deal with such complexity here.

As mentioned earlier, one of the difficulties faced in this book is that of attaining a parsimonious model. Parsimony is circumscribed by the fact alluded to above, that the whole system is usually greater than the sum of the parts – in other words the basic structural connectivity should be replicated. This statement is not deduced, it is simply another rule of thumb based on engineering experience. Engineers naturally build on their collective experience - they are infamous for dealing in *rules of thumb, techniques, methodologies* and *philosophies*. These terms also form another kind of hierarchy of complexity: a rule of thumb seems to mean a very narrow and pointed judgment based on empirical evidence, a technique is a precise, specific program of action that produces a standard result. A methodology lacks the precision of a technique but is a firm guide to action. A philosophy is a broad non-specific guideline for action.

Design for ecosystem concerns is an example of an engineering philosophy. Integrated SWMM-WASP water quality modeling is an example of an engineering methodology. Continuous water quality modeling in SWMM is an example of an engineering technique. We will discuss all three concerns, and derive some rules of thumb.

When several system models interact, the system that sits above it in a hierarchy of control, is called a *metasystem* and its control words constitute a *metalanguage*. PCSWMM is an example.

2.6 Brief introduction to data collection

Data collection, not the focus of this book, is often more expensive than modeling. Model costs are usually inclusive of the associated cost of the field data collection program. Normally historic or nearby data are sufficient for design or inferential use of models, but local field data is required to prove, validate and calibrate your model. Data gathering is closely allied to model building. Both follow a procedure of discovery, and they should proceed in harmony. Data collection can be managed by conducting sensitivity analysis on your model, to discover which processes are most likely to influence your computed results, and thus which data are most important to your study.

Clearly, collecting data which relate to processes and parameters that are insensitive is a waste of time and money.

Strict quality controls are essential in the field and laboratory. Faulty data often lead to wrong model interpretation, and hence wrong designs. Most agencies provide standards for data collection, analysis and presentation. Moreover, it's essential to know the accuracy of your data. Always plot field observations as bars showing the expected range or uncertainty. We discuss the uncertainty of field data collected for SWMM input files later.

Field data uncertainty must be estimated in order to estimate model performance. Observed data if sufficient and good can be processed to detect trends and calculate basic statistics before using it in deterministic modeling. Uncertain data should be analyzed as it is gathered for: i. mean, ii. median, iii. variance, and iv. correlations All can be easily computed in the field.

> Data gathering is closely allied to model building in that both follow a procedure of discovery. They can proceed in harmony, if managed by sensitivity analysis, to discover which processes, and thus which data, are most important.

2.7 Introduction to graphical data management

Any data which can be referenced to geographical locations can be managed by a host of geographical information systems (GIS), of which PCSWMM GIS is one. PCSWMM GIS acts as an independent GIS, providing a graphical plan view editor for quickly creating, editing and/or querying the physical entities of a SWMM model and their attributes, using any coordinate system. But PCSWMM GIS can also support linkages to existing GIS and Facilities Management Systems (FMS) and Information Management Systems (IMS) databases, spreadsheets and text files, with full Structured Query Language (SQL) support. Users can import from and export to the Runoff, Transport and Extran modules of SWMM. SWMM model results can be dynamically displayed as a layer and links are provided to other PCSWMM tools. The following are three of the main attributes:

1. SWMM models

SWMM entities can include conveyances (conduits), nodes and subcatchments. Entities can be graphically created with simple mouse clicks and/or imported from existing SWMM files, many database formats, spreadsheets and other file types. Entity attributes are edited through linked tables and tools for the reduction of model complexity are provided with the Aggregation wizard.

PCSWMM GIS features intelligent connectivity checking and reporting. Runoff, Transport, or Extran input data files can be updated with a single click.

2. Display background images
Most common map and image formats can be displayed as backgrounds to the SWMM entity layers. For example, you may wish to display multiple tiled TIFF images of a USGS topographic map, an AutoCAD DXF vector drawing file of city streets, and a layer containing objects representing the results of a point database query, all overlain by your Extran and Runoff SWMM model layers. Supported formats include ArcView shape files, AutoCAD DXF and DWG files, MapInfo, and Microstation, as well as raster image support for TIFF, JPG, BMP and more.

3. Animate model results
PCSWMM GIS also features dynamic thermometer-style playback of computed Extran model results. In addition, PCSWMM GIS provides path and element selection for the display/analysis of computed results in other PCSWMM tools (e.g. the Dynamic Hydraulic Grade line tool).

2.8 Summary and closing remarks

To discretize a real watershed, identify the problem watershed and its most significant elements. Discretize so that parameterization does not require look-up tables, substitution into equations, or the use of nomographs etc. Starting from the location of the problem in the drainage channel, sketch in the field of surface overland flow vectors, demarcate the overall divide, mark all significant channels and major hydrological components, and delineate sub-catchments for each tributary and storage unit. Number and measure them appropriately, and construct a schematic depicting the sub-catchment and conveyance connectivity. Finer discretization usually requires a finer time step. A large number of elements that may be considered in the systems analysis procedure are tabulated in this chapter.

Process disaggregation proceeds by incorporating additional processes, while its complement, aggregation, proceeds by averaging over larger areas, absorbing detailed processes into fewer, coarser processes, and recalibrating those coefficients. Model complexity is dependent on available field data, and the importance of estimating the uncertainty of field data is paramount.

GIS systems are appropriate for the management of most physical field data (excluding the input and output time series) required for urban systems models. Attributes of PCSWMMGIS are reviewed in the chapter.

References

American Society of Civil Engineers, 1982. Gravity Sanitary Sewer Design and Construction. American Society of Civil Engineers, New York.

American Society of Civil Engineers. 1986. Design and Construction of Sanitary and Storm Sewers, New York.

American Society of Civil Engineers, 1992. Pressure Pipeline Design for Water and Wastewater. American Society of Civil Engineers, New York.

American Society of Civil Engineers, 1992, Design and Construction of Urban Stormwater Management Systems. American Society of Civil Engineers, New York.

Debo, Thomas N. and Reese, Andrew J., 1995. Municipal Storm Water Management. Lewis Publishers, London.

McGhee, Terrence J. 1991. Water Supply and Sewerage. McGraw-Hill Inc., New York.

Pecher, R. 1969. Der Abflussbeiwert und seine abhägigkeit von der Regendauer", (Berichte aus de Institut für Wasserwirtschaft und Gesundheits-ingenieurwesen Technische Hochschule München Nr. 2, 1969, ca 200pp.)

Bibliography

Angulo, C.E. Puente, and Bras, R.L. 1984. Non-linear filtering, parameter estimation and decomposition of large rainfall-runoff models. Report 297, (R-84-13; OSP 89310) Ralph M. Parsons laboratory, MIT. Oct 1984.

James, W. and Robinson, M.A. 1985. Required Time and Space Resolution for Continuous Dynamic Storm and Runoff Modelling and for Associated Field Data. ASCE Hydraulics Specialty Conf. on Hydraulics and Hydrology in the Small Computer Age, Orlando, pp. 1160-1165, Aug 1985.

Robinson, M. and James, W., 1984. "Continuous Variable Resolution Stormwater Modelling on a Microcomputer", Proc. of the Stormwater and Water Quality Modelling Conf., Sponsored by USEPA and Ontario Ministry of Environment, Burlington, Ontario, published by CHI Hamilton, pp. 59-74, Sept 6-7, 1984.

Zaghloul, N.A. 1981. SWMM model and level of discretization. J. of the Hydraulics Div. ASCE. 107(11):1535-1545. *see also Marsalek above.*

This page intentionally left blank.

Chapter 3

RELIABILITY OF INPUT PARAMETERS, GOOD INPUT DATA FILES AND FILE MANAGEMENT.

[Here we] address errors of the third kind. You've never heard of them? But you've made them; we make them every time we write a computer program to solve a physical problem.
Errors of the first kind are grammatical... The compiler finds them.
Errors of the second kind are mistakes in programming… We have to find them [before the program runs].
Errors of the third kind are the ones we haven't found yet!

– Forman S. Acton

3.1 Introduction

For long term continuous water quality modeling, field data comes in two main classes:

1. hydro-meteorological time series (TS), usually collected and archived by governmental and quasi-governmental agencies, and
2. field or environmental parameters, usually collected by a team associated with the modeling.

Further, two types of hydro-meteorological time series are required:

1. good short term TS for calibration, and
2. credible long term input TS for design/inference.

Also, two types of environmental data are required:

1. good details of the as-is scenario to be calibrated, and
2. credible data for each potential scenario that is to be evaluated.

This chapter describes the likely accuracy and reliability of these categories of data. For the various types of data, classes of reliable input parameters and times-to-estimate them are suggested. Well-organized datafiles are exemplified and it is shown how well- managed data files may be referenced graphically in a GIS. The subject of data management utilities is itself weighty and is only covered fleetingly. Graphical representation of geo-referenced data (GIS), and its implementation in PC-SWMMGIS is described in some detail.

3.2 Short-term time series dataset for calibration.

Contrary to what is evidently a widespread belief, it is unnecessary to have a long observed response function TS (e.g. runoff) to develop a reliable, continuous model. On the contrary, there need only be sufficient, good, observed time series to cover the number of events needed for calibration or parameter optimization, and a long, credible observed input function TS (e.g. rain). Because dry-weather inter-event records can be obtained at virtually no increase in cost, the total field data collection effort (for calibration) is not significantly different for continuous models than it is for event models. Table 3.1 sets out a basis for organizing the different types of required input data. There are two different classes of data quality: 1. that required for calibration, which must be as accurate as is reasonably possible, and 2. data for design inference, which must be acceptable (credible) to all parties involved in the project. Only the *as-is* dataset or scenario can be calibrated. Three design scenarios are shown in Table 3.1, even though many more will normally be required.

Table 3.1: Different types of required input data

Input and output function TS		Input parameters			
Short-term IF & OF for calibration	Long-term IF for inference	As-is Scenario 0 for calibration	To-be$_1$ Scenario 1 for inference	To-be$_2$ Scenario 2 for inference	To-be$_3$ Scenario 3 for inference
accurate	credible	accurate	credible	credible	credible

TS = time series; OF = output function (e.g. runoff); IF = input function (e.g. rain)

Sometimes, when good data is available only in nearby drainage systems, or elsewhere in the system, two models are developed, one for the area with good data, and another for the area deficient in data. Of course both models must be similar and for the same hydro-meteorological region. Optimized parameters are then transposed to the model lacking data.

In both cases, the observed time series must be searched over, say, a span of a year or two, for the requisite events for calibration. Also in both cases, the accuracy or uncertainty of the observed TS must be determined and reported.

3.3 Long-term dataset for inference.

At least one long input function is required for the inferential runs, when the alternative arrays of BMPs are to be compared. This long input TS need only be credible and can be transposed from a station in the same hydro-meteorologic region, or it may be generated synthetically from a reasonably long observed TS in the same hydrologic region. Artificial rainfall generation is covered later in this book.

Consider an implicit design problem: to find the optimum cost-effective array of BMPs. Since the continuous input TS will be used for comparing the long-term performance of various scenarios, it is only necessary for the long-term input TS to be plausible. The test is: *could this input time series have reasonably occurred at this point?* The solution to the design problem, usually accepted by stakeholders as a credible position, may be stated as follows: *Could the long-term input rainfall time series that in fact occurred elsewhere, have reasonably occurred at the design problem area? If yes, then if all plans examined had really existed over this time, would the plan identified would have been the most cost-effective?*

Once the long term continuous rain record has been developed (i.e., processed, transposed, or generated) it is available for all studies in that hydrologic region, much like standard design storms are. It is not often recognized that for the long-term continuous model input, the basic information is exactly the same as that used in the first place to derive the design storms for that area. In other words, unless the local design storms were in fact arbitrary, no additional data collection effort is required for continuous SWMM4 modeling than that required for the so-called rational equation. Of course, the derivation procedure for design storms always results in a loss of information because the intrinsic variability of rain is reduced to simplistic representations.

> A long-term observed dataset is required only for rainfall, and may be transposable (i.e., it need not have been observed at the precise design problem location).

3.4 Categories of uncertain data

Certain physical environmental data in the input file remain more-or-less constant through the run. We call these variables "parameters", and the act of choosing values for them, parameter estimation or parameterization.

We may categorize input parameters into five groups:

B. *Binary:* Input that acts like switches, or run control input parameters that may be either completely correct (in the sense that the datafile runs correctly) or completely incorrect (the datafile crashes or does not run successfully).

1. *Parameters that can be measured with almost total certainty*: examples are catchment areas, pipe diameter, and other conduit geometries, conduit slopes and elevations, conduit

lengths, storage areas, spillway geometry, and others. These are not good candidates for calibration.

2. *Parameters that can be estimated with a high degree of certainty in the field, design office or laboratory*. These include: % imperviousness, Manning's n for pipes, particle size distributions, temperature, flow, concentrations and others. These can and should be modified within narrow limits during calibration.
3. *Parameters that cannot be easily measured in the field or laboratory* like infiltration rates, rainfall, pollutant build-up, pollutant wash-off, and many others. These are good candidates for calibration.
4. *Parameters that cannot be measured with any certainty at all*, such as mean ground slope in some situations, catchment widths, recovery of infiltration capacities, and several others. These are obviously prime candidates for calibration.

For these categories, uncertainty varies in broad ranges:

- B. uncertainty is either null or complete,
- 1. uncertainty is small, a few percentage points, perhaps 5-10%,
- 2. perhaps 10-25%,
- 3. perhaps 25-50%, and
- 4. who knows? If any estimate can be made at all, perhaps assume 50-100% as a starting point.

Some models, graphical user interfaces (GUIs) and decision support systems (DSSs) such as PCSWMM account for the expected range and distribution of estimates of parameters, and request the user to input three values for each parameter (Dunn, 1986; Dunn and James, 1985) indicating the range: i.e. expected or mean value, and low and high values, near the extremes. Estimates are based on the user's idea of what the best estimates by a number of fellow users might be. Estimates by novices and users from very different areas may be expected to vary widely. This is also covered in more detail later. To estimate uncertainty for every parameter in your input data files, work from a consideration of the source of information, and the limitations of the original instruments and methods.

For SWMM4, some of the required input data are listed in Tables 3.2 to 3.24. *U* is the estimated uncertainty class, given by the above classes. *T* is the estimated time in minutes (1, 2, 5 or 10) required to find and enter the data when building the datafile for the first time, averaged on a per entry basis. Clearly the estimates of *U* and *T* are subjective, and dependent on the highly variable experience of the model user, as well as the local availability of source data. Cursory estimates are provided in these tables, and readers are advised to carefully consider their own experience and local data availability before relying on these estimates. Similar tables should be generated for other WSMs.

Contents of the next 23 tables are given in the next table.

Table 3.2A. Contents of the following 23 SWMM4 input tables.

Table	Module	Parameters
3.2	RUNOFF	run control
3.3		rain and ET
3.4		snowmelt
3.5		hydrology
3.6		hydraulics
3.7		water quality & erosion
3.8	STORAGE	run control
3.9		input function
3.10		hydraulics
3.11		water quality
3.12		cost
3.13	TRANSPORT	run control
3.14		conveyance geometry
3.15		water quality
3.16		storage and flow dividers
3.17		infiltration-inflow
3.18		dry weather flow
3.19	EXTRAN	run control
3.20		conveyance and bridge details
3.21		storage units
3.22		orifices, weirs and pumps
3.23		tides
3.24		input functions

Table 3.2 User-defined RUNOFF module job control parameters of binary uncertainty and their estimate time T in minutes.

line	*Input Variables*	*Meaning*	*T*
$	$RUNOFF	Call the RUNOFF module	1
A1	TITLE	Two title lines printed on output.	2
B1	METRIC	Metric input-output.	1
	ISNOW	Snowmelt parameter	1
	NRGAG	Number of hyetographs (rain gages)	5
	INFILM	Choice of infiltration equation	5
	KWALTY	Quality (or erosion) simulated	1
	IVAP	Evaporation parameter	1
	NHR	Hour of day of start of storm	5
	NMN	Minute of hour of start of storm	5
	NDAY	Day of month of start of simulation	5
	MONTH	Month of start of simulation	5
	IYRSTR	Year of start of simulation	5

	IVCHAN	Optional evaporation on channels	1
B2	IPRN(1)	Print control for SWMM4 input	1
	IPRN(2)	Print control for Runoff Block graphs	1
	IPRN(3)	Print control for output of 'Totals'	1
	IRPNGW	Print ground water error messages	1
	NOHEAD	Reprint headers after every 50 lines	1
	LANDURR	Percentage of each land use	10
B3	WET	Wet time step (seconds).	5
	WETDRY	Transition between wet and dry	5
	DRY	Dry time step (seconds).	5
	LUNIT	Units of LONG (simulation length).	1
	LONG	Simulation length (units from LUNIT).	1
B4	PCTZER	% impervious area with zero detention	1
	REGEN	Horton infiltration regeneration	5
M1	NPRNT	No. channels/pipes/inlets to be printed	1
	INTERV	Print Control.	1
M2	NDET	Number of detailed printout periods.	1
	STARTP(1)	First starting printout date	1
	STOPPR(1)	First stopping printout date.	1
	.	.	.
	STARTP(NDET)	Last starting date.	1
	STOPPR(NDET)	Last stopping date.	1
M3	IPRNT(1)	channel/inlets for which flows and	1
	.	.	.
	IPRNT(NPRNT)	concentrations are to be printed.	1
M4	MDEEP	Number of depth locations for printout	1
	KDEEP(1)	First conduit selected	1
	.	.	.
	KDEEP(MDEEP)	Last conduit selected	1
M5	MSUBC	No. of subcatchments for printout	2
	ISUBC(1)	First subcatchment selected	1
	ISUBC(MSUBCT)	Last subcatchment selected	1
$	$ENDPROGRAM	End your input data set	1

Table 3.3 RUNOFF rain and evaporation input parameters, their uncertainty class U and estimate time T in minutes.

line	*Input Variables*	*Meaning*	*U*	*T*
D1	ROPT	Precipitation input option	B	10
E1	KTYPE	Type of precipitation input.	B	5
	KINC	Number of precipitation values	B	5
	KPRINT	Print control for precipitation input.	B	1
	KTHIS	Variable THISTO option	B	1
	KTIME	Precipitation time units	B	1
	KPREP	Precipitation unit type	B	1
	NHISTO	No. of data points for each hyetograph	B	5
	THISTO	Time interval between values.	B	5
	TZRAIN	Initial time of day of precipitation input	B	5
E2	WTHIS(1,1)	Start time for first variable precipitation	B	2
	WTHIS(1,2)	End time for first variable precipitation	B	2
	WTHIS(1,3)	Length of THISTO for 1st precip. interval	B	2
	.	.	.	.
	WTHIS(KTHIS,1)	Start time for last variable precipitation	B	2
	WTHIS(KTHIS,2)	End time for last variable precipitation	B	2
	WTHIS(KTHIS,3)	Length of THISTO for the last precip.	B	2
E3	RAIN(1)	Rainfall intensity, first interval	2	1
	.	.	.	.
	RAIN(KINC)	Rainfall intensity, last interval	2	1
	Or			
	REIN(1)	Time of first precipitation	1	1

	REIN(2)	Precipitation for first interval	2	1
	.	.	.	.
	REIN(2NRAG+1)	Precipitation, last rain gage	2	1
F1	VAP(1)	Evaporation rate for month 1	3	2
	.	.	.	.
	VAP(12)	Evaporation rate for month 12	3	2
	Or			
	NVAP(1)	Start year of evaporation:	B	2
	NVAP(2)	No. of months of evaporation data	B	2
F2	EVAP	Monthly evap. max of 600 values	3	20

Table 3.4 Snowmelt input data, their uncertainty class U and estimate times T in minutes.

line	*Input Variables*	*Meaning*	*U*	*T*
C1	ELEV	Average watershed elevation	1	5
	FWFRAC(1)	Ratio of free water holding capacity	3	10
	FWFRAC(2)	Ratio of free water holding capacity	3	10
	FWFRAC(3)	Ratio of free water holding capacity	3	10
	SNOTMP	Dividing temperature between snow and rain	2	5
	SCF	Snow gage catch correction factor.	3	5
	TIPM	Weight for antecedent temperature index	3	5
	RNM	Ratio of negative melt coefficient	3	10
	ANGLAT	Average latitude of watershed, degrees north	1	5
	DTLONG	Longitude correction	1	5
C2	NUMB	number of months with wind speed data.	B	5
	MONTH	Integer number of first month	B	1
	WIND(MONTH)	Average wind speed for first month	3	5
	.	.	.	.
	MONTH	Integer number of last month	B	1
	WIND(MONTH)	Average wind speed for last month	3	5
C3	ADCI(1)	Fraction of impervious area covered by snow	2	5
	ADCI(2)	Value of ASC for AWESI = 0.1.	2	5
	ADCI(3)	Value of ASC for AWESI = 0.2	2	5
	.	.	.	5
	ADCI(9)	Value of ASC for AWESI = 0.8	2	.
	ADCI(10)	Value of ASC for AWESI = 0.9	2	5
C4	ADCP(1)	Fraction of pervious area covered by snow	2	5
	ADCP(2)	Value of ASC for AWESI = 0.1.	2	5
	ADCP(3)	Value of ASC for AWESI = 0.2	2	5
	.	.	.	.
	ADCP(9)	Value of ASC for AWESI = 0.8	2	5
	ADCP(10)	Value of ASC for AWESI = 0.9	2	5
C5	DTAIR	Time interval for air temperatures	B	1
	NAIRT	Number of air temperatures read	B	2
	TAIR(1)	Air temperature during time interval 1	2	1
	.	.	.	.
	TAIR(NAIRT)	Air temperature during time interval NAIRT.	2	1
I1	JK1	Subcatchment number or name	B	2
	SNN1	Fraction of impervious area	2	5
	SNCP(N)	Fraction of pervious area	2	5
	WSNOW(N,1)	Initial snow depth of impervious area	3	5
	WSNOW(N,2)	Initial snow depth on pervious area	3	5
	FW(N,1)	Initial free water on snowed impervious area	4	5
	FW(N,2)	Initial free water on snowed pervious area	4	5
	DHMAX(N,1)	Melt coefficient on snowed impervious area	3	5
	DHMAX(N,2)	Melt coefficient on snowed pervious area	3	5
	TBASE(N,1)	Snow melt base temp snowed impervious area	2	2
	TBASE(N,2)	Snow melt base temp for snowed pervious area	2	2

I2	JK2	Subcatchment number or name	B	2
	WSNOW(N,3)	Initial snow depth on bare impervious area	3	5
	FW(N,3)	Initial free water on bare impervious area	3	5
	DHMAX(N,3)	Max melt coefficient on bare impervious area	3	5
	TBASE(N,3)	Snow melt base temp for bare impervious area	2	5
	DHMIN(N,1)	Min melt coef snow covered impervious area	3	5
	DHMIN(N,2)	Min melt coef for snow covered pervious area	3	5
	DHMIN(N,3)	Min melt coef for snow on bare impervious area	3	5
	SI(N,1)	Snow depth on snowed impervious areas	3	5
	SI(N,2)	Snow depth on snowed pervious areas	3	5
	WEPLOW(N)	Redistribution depth on impervious area	3	5
	SFRAC(N,1)	Fraction transferred to snowed impervious area.	3	5
	SFRAC(N,2)	Fraction transferred to snowed pervious area	3	5
	SFRAC(N,3)	Fraction to pervious area in last catchment.	3	5
	SFRAC(N,4)	Fraction transferred out of watershed.	3	5
	SFRAC(N,5)	Fraction immediately melted on bare imp, area.	3	5

Table 3.5 RUNOFF module hydrology input parameters, their uncertainty and estimate times.

Input Variables	*Meaning*	*U*	*T*
JK	Hyetograph number	B	1
NAMEW	Subcatchment number or name.	B	1
NGTO	Channel/inlet number for drainage.	B	1
WIDTH	Subcatchment width	4	10
AREA	Area of subcatchment	1	10
%IMP	% directly connected area	3	10-30
SLP	Ground slope	2-4	10
IMPN	Imp area Manning’s n	2	5
PERVN	Pervious area Manning’s n	4	5
IDS	Imp area depression storage	3	5
PDS	Pervious area depression storage	4	5
WLMAX	Maximum initial infiltration rate	4	5
WLMIN	Minimum (asymptotic) infiltration rate	4	5
DECAY	Horton's infiltration decay rate	4	10
SUCT	Average capillarity suction	3	5
HYDCON	Saturated capillarity suction	3	5
SMDMAX	Initial moisture deficit for soil	4	5
RMAXINF	Maximum infiltration volume	4	10
IFLOWP	Overland flow plain to another	3	5
PZ	% imp area with no depression storage	2	5
NMSUB	Subsurface subcatchment indicator variable	B	2
NGWGW	inlet, channel or pipe for subsurface drainage	B	2
ISFPF	for printing soil moisture, water table, outflow	B	2
ISFGF	for graphing soil moisture, water table, outflow	B	2
BELEV	Elevation of bottom of water table aquifer	3	2
GRELEV	Elevation of ground surface	2	2
STG	Elevation of initial water table stage	3	5
BC	Elev of channel btm or for groundwater flow	2	2
TW	Channel water influence parameter	3	5
A1	Groundwater flow coefficient	4	5
B1	Groundwater flow exponent	4	5
A2	Coefficient for channel water influence	4	5
B2	Exponent for channel water influence	4	5
A3	Coefficient	4	5
POR	Porosity expressed as a fraction	2	5
WP	Wilting point expressed as a fraction	3	5
FC	Field capacity expressed as a fraction	3	5
HKSAT	Saturated hydraulic conductivity.	3	5
TH1	Initial upper zone moisture as a fraction	4	5
HCO	Hydraulic conductivity vs. moisture content	4	10
PCO	Average slope of tension versus soil	3	10
CET	Fraction of maximum ET rate	3	5
DP	Coefficient for unquantified losses,	4	5

DET	Maximum depth of lower zone transpiration	3	5

Table 3.6 RUNOFF module hydraulics input parameters, their uncertainty and estimate times.

Variable	*Meaning*	*Source of Information*	*U*	*T*
IIRDII	compute I/I response from rainfall	User defined	B	2
TSTEP	Time interval for rainfall,	User defined	B	2
TSTEP2	Time step for computation of I/I	User defined	B	5
NRDHYET	Number of hyetograph (rain gage)	User defined	B	1
RDIIT	Time to peak	User defined	2	2
RDIIK	Ratio of recession limb to time to peak	User defined	2	2
DSTORE	Maximum initial abstraction	User defined	2	5
STORAGE	Initial storage	User defined	2	5
DREC	Recovery rate for storage	User defined	3	5
NAMEG	Channel/pipe number or name.	User defined	B	2
NGTO	Ch/pipe/inlet number/name for drainage.	User defined	B	2
NPG=NP	Type of Channel	Maps or measurement	B	2
GWIDTH	Bottom width of trapezoidal channel	Maps	2	2
GLEN	Length of channel	Map or aerial photo	1	2
G3	Invert slope	Maps	3	2
GS1	Left-hand side slope	Maps or measurement	3	2
GS2	Right-hand side slope	Maps	3	2
G6	Manning's roughness coefficient	Maps	2	5
DFULL	Depth of channel when full	Maps + measurements	2	2
GDEPTH	Starting depth of pipe/channel	Flowrate measurements	4	2
WTYPE	Type of weir/orifice	Drawings or field	B	2
WELEV	Elevation of weir	Maps	2	5
WDIS	Discharge coefficient of weir or orifice	Tables	2	5
SPILL	Weir length (e.g. width of spillway)	Drawings	1	5
SEWAREA	Sewered area	Maps	2	10
RDIIR(1)	Fraction of first UH to total	Measurements	3	10
RDIIR(2)	Fraction of second UH to total	Measurements	3	10
RDIIR(3)	Fraction of third UH to total	Measurements	3	10
ICURVE	Indicator for UHs	User defined	B	2

Table 3.7 SWMM4/RUNOFF input parameters for water quality and erosion, their uncertainty and times to estimate (minutes).

Line	Input Variables	Meaning	U	T
	IMUL	Trigger multiple land uses	B	5
	NQS	Number of quality constituents	B	5
	JLAND	Number of land uses	B	5
	IROS	Erosion simulation parameter	B	5
	IROSAD	Option to add erosion to constituent	B	5
	DRYDAY	Number of dry days prior to storm.	B	5
	CBVOL	Av catchbasin storage volume	1	5
	DRYBSN	Dry days required to recharge	B	5
	RAINIT	Highest av 30-minute rainfall intensity	2	10
	REFFDD	Street sweeping efficiency (removal)	3	5
	KLNBGN	Day of year street sweeping begins	2	5
	KLNEND	Day of year street sweeping stops	2	5
	LNAME(J)	Name of Land use.	B	2
	METHOD(J)	Buildup equation type for D&D.	B	5
	JACGUT(J)	Functional dependence of buildup	B	5
	DDLIM(J)	Limiting buildup quantity.	4	10
	DDPOW(J)	Buildup parameters Power or exponent.	4	10
	DDFACT(J)	Buildup coefficient.	4	10
	CLFREQ(J)	Street sweeping interval, days.	3	10
	AVSWP(J)	Availability of street for sweeping	3	10
	DSLCL(J)	Street sweeping days since cleaning	2	2
	PNAME(K)	Constituent name.	B	2
	PUNIT(K)	Constituent units.	B	2

	NDIM(K)	Type of units.	B	2
	KALC(K)	Type of buildup calculation.	B	5
	KWASH(K)	Type of washoff calculation	B	5
	KACGUT(K)	Functional dependence of buildup	B	5
	LINKUP(K)	Linkage to snowmelt	B	2
	QFACT(1,K)	First buildup parameter, e.g., limit.	4	10
	QFACT(2,K)	2^{nd} buildup parameter	4	10
	QFACT(3,K)	3^{rd} buildup parameter, e.g. coefficient.	4	10
	QFACT(4,K)	4^{th} buildup parameter	4	10
	QFACT(5,K)	5^{th} buildup parameter	4	10
	WASHPO(K)	Power (exponent) for runoff rate.	4	10
	RCOEF(K)	Coefficient.	4	10
	CBFACT(K)	Initial catchbasin concentration.	4	5
	CONCRN(K)	Concentration in precipitation.	3	10
	REFF(K)	Street sweeping efficiency	3	10
	REMOVE (K)	Removal fraction for overland flow	4	10
	QDECAY (K)	First-order decay coef for pollutant	4	10
	KTO	Number of constituent .	B	5
	KFROM	Number of constituent .	B	5
	F1(KTO,KFROM)	Fraction of constituent KFROM added	3	10
J5		Groundwater concentration		
	TS		4	10
	BOD5		4	10
	TN		4	10
	NO2+NO3		4	10
J6		Constant I/I concentration		
	TS		4	10
	BOD5		4	10
	TN		4	10
	NO2+NO3		4	10
	NGNAME	Name of channel/pipe	B	1
	GREMOVE(1)	Removal fraction for first pollutant	4	10
	GREMOVE(2)	Removal fraction for second pollutant	4	10
	GREMOVE(NQS)	Removal fraction for NQS pollutants	4	10
JJ		Multiple land uses per subcatchment.		
	N=NAMEW	Subcatchment number or name	B	1
	ERODAR	Area of subcatchment for erosion	1	5
	ERLEN	Flow distance over erodible area	2	5
	SOILF	Soil factor 'K'.	3	10
	CROPMF	Cropping management factor 'C'.	3	10
	CONTPF	Control practice factor 'P'.	3	10
	NAMEW	Subcatchment number or name.	B	1
	KL	Land use classification.	B	2
	BASINS(N)	No .of catchbasins in subcatchment.	B	1
	GQLEN(N)	Total curb length within subcatchment	1	10
	PSHED(1,N)	Initial loading, first constituent.	4	10
	.	.		
	PSHED(10,N)	Initial loading, tenth constituent.	4	10
	PLAND(1,N)	land use 1, fraction of subcatchment N	1	5
	PLAND(2,N)	land use 2, fraction of subcatchment N	1	5

Table 3.8 List of SWMM4/STORAGE run control input parameters of binary uncertainty and their times to estimate T

Input Variables	Meaning	U	T
A1	First title line	B	1
A1	Second title line	B	1
NOTAPE	Input data source	B	2
JNS	External element number from the external	B	1
NDT	Total number of simulation time steps.	B	2
DS	Size of time step.	B	2
NU	Number of storage/treatment units	B	1
NP	Number of pollutants routed	B	1
ICOST	Cost calculations performed?	B	1

METRIC	Metric input-output.	B	1
SAREA	Service area	1	2
IDATE	Date at beginning of simulation	B	2
TIME	Time at beginning of simulation	B	2
ISUM	Summary print control parameter	B	1
IDET	Detailed print control parameter	B	1
NPR	Number of detailed print periods.	B	1
ISTART(1)	First detailed print period starting	B	1
IEND(1)	First detailed print period ending	B	1
ISTART(NPR)	Last detailed print period starting date.	B	1
IEND(NPR)	Last detailed print period ending date.	B	1

Table 3.9 List of SWMM4/STORAGE input function parameters and their uncertainty U and times to estimate T in minutes

Input Variables	Meaning	U	T
E(1)… E(12)	Evaporation rate, January - December	3	10
TCAR	Time of day, decimal hours	B	2
QCAR	Flow entering S/T plant (at unit 1).	2	1
PCAR(1)	Conc. of pollutant 1 entering S/T plant sed only	4	2
PCAR(2)	Conc. of pollutant 2 entering S/T plant (unit 1).	4	2
PCAR(3)	Conc. of pollutant 3 entering S/T plant (unit 1).	4	2

Table 3.10 List of SWMM4/STORAGE hydraulics input parameters and their uncertainty U and times to estimate T in minutes

Input Variables	Meaning	U	T
IDENT(1)	Detention modeling parameter	B	2
QMAX(I)	Max inflow (above which bypass occurs)	1	2
QRF(I)	Residual flow as a fraction of the inflow.	2	5
IDIREC(I,1)	Unit number to which bypass is directed	B	2
IDIREC(I,2)	Unit number for treated outflow	B	2
IDIREC(I,3)	Unit number for residuals stream	B	2
IROUTE(I)	Pollutant routing parameter	2	5
IOUT(I)	Treated outflow routing parameter	2	5
IDRAW(I)	Residuals stream draw-off scheme	B	5
IRES(I)	Residual stream routing parameter	B	5
ALEN(I)	Travel length for plug flow.	2	10
AMAN(I)	Manning's roughness coef. for detention units	1	2
SDEPTH(I,MM)	A unit depth	1	2
SAREA(I,MM)	Surface area corresponding to the above depth	1	2
SSTORE(I,MM)	Volume corresponding to the above depth	1	2
SQQOU(I,MM)	Treated outflow at the above depth	1	2
SQQRS(I,MM)	Residuals stream flow at the above depth	1	2
C1	Depth-treated outflow equation coefficient, C1.	1	2
D0	Depth below which no treated outflow occurs	1	2
C2	Depth-treated outflow equation coefficient, C2.	1	2
DSTART(I,1)	Depth at which pumping rate begins	1	2
DSTART(I,2)	Depth at which pumping rate begins	1	2
QPUMP(I,1)	Pumping rate when depth >_ DSTART(I,1)	1	2
QPUMP(I,2)	Pumping rate when depth >_ DSTART(I,2)	1	2
DSTOP(I)	Depth below which all pumping stops	1	2
C3	Depth-residual flow equation coefficient.	1	2
D1	Depth below which no residual flow occurs.	1	2
C4	Depth-residual flow equation coefficient.	1	2
WARN(I)	Tot vol of water in unit at the start of simulation	1	5

Table 3.11 List of SWMM4/STORAGE water quality input parameters and their uncertainty U and times to estimate T in minutes

Input Variables	Meaning	U	T
IPOLL(1)	Pollutant 1 selector.	B	2
NDIMI(1)	Dimensions for pollutant 1.	B	2
IPART(1)	Particle distribution parameter	B	5
PNAME1(IN,1)	Pollutant 1 name.	B	1
PUNIT1(IN,1)	Pollutant 1 dimension label.	B	2
IPOLL(2)	Pollutant 2 selector.	B	2
NDIMI(2)	Dimensions for pollutant 2.	B	2
IPART(2)	Particle distribution parameter.	B	5
PNAMEI(IN,2)	Pollutant 2 name.	B	1
PUNITI(IN,2)	Pollutant 2 dimension label.	B	2
IPOLL(3)	Pollutant 3 selector.	B	2
NDIMI(3)	Dimensions for pollutant 3.	B	2
IPART(3)	Particle distribution parameter.	B	5
PNAMEI(IN,3)	Pollutant 3 name.	B	1
PUNITI(IN,3)	Pollutant 3 dimension label.	B	2
NVS	Classification parameter	B	2
NNR	No. of particle size ranges or settling velocities	B	5
RAN(1,1)	Lower bound of size or velocity range 1	1	5
RAN(1,2)	Upper bound of size or velocity range 1	1	5
RAN(NPR,1)	Lower bound of size or velocity range NPR	1	5
RAN(NPR,2)	Upper bound of size or velocity range NNR	1	5
SPG(1)	Specific gravity for particles in size range 1.	1	5
SPG(2)	Specific gravity for particles in size range 2.	1	5
SPG(NNR)	Specific gravity for particles in size range NNR.	1	5
TEMP(1)… TEMP(12)	Waste stream temperature, January - December	3	10
PSD(IP,1)	Fraction of pollutant IP in range 1.	3	10
PSD(IP,2)	Fraction of pollutant IP in range 2.	3	10
UNAME(I,ID)	Name of unit.	B	1
RMX(I,IP)	Maximum removal fraction (<= 1.0).	2	5
INPUT(I,IP,1)	Program variable for equation variable X1.	2	10
INPUT(I,IP,2)	Program variable for equation variable X2.	2	10
A(I,IP,1)	Value of coefficient a1.	2	5
A(I,IP,2)	Value of coefficient a2.	2	5
PSC(I)	Critical particle size or settling velocity	3	10
NPSL(I)	Pollutant responsible for sludge generation.	B	5
SLDEN(I)	Concentration of pollutant NPSL(I) in sludge.	4	10
SLDMAX(I)	Maximum sludge depth, ft [m]. A warning	1	2
PCO(I,1)	Conc. of pollutant 1 in the unit at start	4	2
PCO(I,2)	Conc. of pollutant 2 in the unit at start	4	2
PCO(I,3)	Conc. of pollutant 3 in the unit at start	4	2

Table 3.12 List of SWMM4/STORAGE cost input parameters and their uncertainty and times to estimate T in minutes

Input Variables	Meaning	U	T
KPC(I,1)	Type of cost variable for initial capital cost	B	10
CC(I,1)	Initial capital cost equation coefficient, a.	2	10
CC(I,2)	Initial capital cost equation coefficient, b.	2	10
KPC(I,2)	Type of cost variable for O&M costs	B	5
CC(I,3)	O&M cost equation coefficient, d.	2	5
CC(I,4)	O&M cost equation coefficient, f.	2	5
CC(I,5)	O&M costs equation coefficient, h.	2	5

Table 3.13 TRANSPORT run control and input function parameters and their uncertainty class and approximate times to estimate in minutes

line	parameter	Meaning	*U*	*T*
A1	Title1	Title for the job	B	1
A1	Title2	Title for the run	B	1
B0	ISLOPE	Default slope of ft/100 ft or m/ 100 m	B	2
	ITRAP	Input a trapezoid side slope	B	2

	IFLIP	Default input of flow/pollutants on line R1	B	2
	INFLEW	NNYN inflow hydrographs for printing	B	2
	IDETAIL	Changes in the operation of transport	B	2
B1	NDT	Number of time steps.	B	5
	NINPUT	No. of non-conduit elements with data input in data group R1	B	2
	NNYN	Number of non-conduit elements with input hydrographs	B	2
	NNPE	Number of non-conduit elements with printout(s).	B	2
	NOUTS	Number of non-conduit elements to be placed on the interface file	B	2
	NPRINT	Suppress most error messages (recommended).	B	1
	NPOLL	Number of water quality constituents (Max of 4).	B	5
	NITER	Number of iterations used in routing subroutine	B	5
	IDATEZ	Starting date of storm; year/month/day.	B	5
	METRIC	Units.	B	1
	INTPRT	Print interval for input and output tables	B	2
B2	DT	Time step size, seconds	B	5
	EPSIL	Allowable error for convergence.	B	2
	DWDAYS	Total number of dry-weather days prior to simulation	B	5
	TZERO	Starting time of storm in hours.	B	5
	GNU	Kinematic viscosity of water	1	1
	TRIBA	Total catchment area	1	1
B3	NCNTRL	Input control	B	1
	NINFIL	Sewer infiltration inflows.	B	5
	NFILTH	Estimate dry-weather sewage inflow	B	5
	NDESN	Hydraulic design routine.	B	1
B4	IBFF 1	Multiplication factor for monthly base flow	1	2
	IBFF 2	Multiplication factor for monthly base flow	1	2
	IBFF 3	Multiplication factor for monthly base flow	1	2
	NUMBFF	Number of base flows input	B	2
	BFFMO(1)..	First monthly base flow	2	5
	.	.	.	.
	BFFMO(NUMBFF)	last monthly base flow	2	5
H1	JN(1)...JN(NOUTS)	Non-conduit elements transferred to subsequent blocks.	B	2
H2	TSTART	Time of day (hours) at which linkage file output should begin.	B	2
	IDEP	Output volumes, depths and velocities at every time step	B	2
	NWAS	Number of WASP segments in data group H3.	B	2
H3	KK	Transport external element number or name.	B	2
	JUNSEG	Corresponding WASP segment number, an integer.	B	2
I1	NORDER(1)	Non-conduit elements into which hydrographs and pollutographs	B	2
	NORDER(NINPUT)	enter the system using group R1.		
I2	NSURF	Number of conduit (and/or storage) elements.	B	2
	JSURF(1)	Conduit elements	B	2
	JSURF(NSURF)			
J1	NYN(1)...NYN(NNYN)	non-conduit elements printed at the end of the simulation.	B	2
	NPE(1) .....NPE(NNPE)			
	DINFIL	Base dry weather infiltration	2	10
	GINFIL	Groundwater infiltration	4	10
	RINFIL	Rainwater infiltration	4	10
	RSMAX	Peak residual moisture	4	10
	CPINF(1)	Concentration of quality constituent # 1.	4	10
R1	PE2(1) .... PE2(4)	time in hours flow (concentrations)	3	5

Table 3.14 TRANSPORT geometrical conveyance input parameters and their uncertainty class and approximate times to estimate in minutes

line	parameter	Meaning	*U*	*T*
C1	NKLASS	Number of user-defined sewer cross-sectional shapes.	B	1
	KPRINT	Control for printing flow routing parameters for all shapes	B	1
D1	SHAPENAME1	First shape name.	B	1
	SHAPENAME2	Second shape name.	B	1
D2	NN(17)	No. of DNORM values for shape 1	B	2
	NN(18)	Number of DNORM values for shape 2	B	2
D3	MM(17)	Number of QNORM values for shape 1	B	2
	MM(18)	Number of QNORM values for shape 2	B	2
D4	ALFMAX1	Value of A/Afull corresponding to maximum Q/Qfull for shape 1	1	2

	ALFMAX2	Value of A/Afull corresponding to maximum Q/Qfull for shape 2	1	2
D5	PSIMAX1	Maximum Q/Qfull for shape 1	1	2
	PSIMAX2	Maximum Q/Qfull for shape 2	1	2
D6	AFACT1	Factor used to determine full flow area for shape 1.	1	2
	AFACT2	Factor used to determine full flow area for shape 2.	1	2
D7	RFACT1	Factor used to determine full flow hydraulic radius for shape 1.	1	2
	RFACT2	Factor used to determine full flow hydraulic radius for shape 2.	1	2
D8	DNORM	data at 8 values per line	1	10
D9	QNORM	data at 8 values per line	1	10
E1	NOE	Element number.	B	1
	NUE(1)	First of three possible upstream elements (number or name).	B	2
	NUE(2)	Second of three possible upstream elements.	B	2
	NUE(3)	Third of three possible upstream elements.	B	2
	NTYPE	Element type	B	1
	DIST	Element length or inflow	1	1
	GEOM1	First characteristic dimension	1	1
	SLOPE	Invert slope	2	2
	ROUGH	Manning's roughness	1	1
	GEOM2	Second characteristic dimension	1	1
	BARREL	Number of barrels	B	1
	GEOM3	Third characteristic dimension	1	1
	KGEOM	Third characteristic dimension for named conduits	1	1
E2/NC	XNL	n for the left overbank.	1	1
	XNR	n for the right overbank.	1	1
	XNCH	n for the channel.	1	1
E3/X1	SECNO	Cross section identification number.	B	1
	NUMST	Total number of stations on the following	B	1
	STCHL	Station of the left bank of the channel	B	2
	STCHR	The station of the right bank of the channel	B	2
	XLOBL	Not required for Transport (enter 0.0).	B	1
	XLOBR	Not required for Transport (enter 0.0).	B	1
	LEN	Length of channel reach represented by this cross section	2	1
	PXSECR	Factor to modify the horizontal dimensions for a cross section.	B	2
	PSXECE	Constant to be added (+ or -) to E4 (GR) elevation data	B	2
E4/GR	EL(1)	Elevation of cross section at STA(1)	1	1
	STA(1)	Station of cross section 1	1	1
	EL(2)	Elevation of cross section at STA(2).	1	1
	STA(2)	Station of cross section 2.	1	1

Table 3.15 TRANSPORT water quality input parameters and their uncertainty class and approximate times to estimate in minutes

line	parameter	Meaning	*U*	*T*
F1	KPOL	Constituent selector from interface file.	B	2
	PNAME	Constituent label if KPOL = 0.	B	2
	PUNIT	Constituent units if KPOL = 0.	B	2
	NDIM	Dimensions of water quality constituent.	B	2
	DECAY	First order decay coefficient, 1/day.	2	10
	SPG	Specific gravity. If SPG > 1.0 scour/deposition is modeled	1	5
	PSIZE(2)	Particle size, mm.PSIZE(1) is automatically set to 0.0 mm.	1	5
	PGR(2)	Percent greater than (%).PGR(1) is automaticallyset to 100.	1	5
	PSIZE(3)	Particle size, mm.	1	5
	PGR(3)	Percent greater than (%).	1	5
	PSIZE(4)	Particle size, mm.	1	5
	PGR(4)	Percent greater than (%).	1	5
	PSIZE(5)	Particle size, mm.	1	5
	PGR(5)	Percent greater than (%).	1	5
	PSDWF	Maximum particle size contained in dry-weather flow input;	2	5

Table 3.16 TRANSPORT storage and flow divider element input parameters and their uncertainty class and approximate times to estimate in minutes

line	parameter	Meaning	*U*	*T*
G1	LOUT(IS)	Outflow routing parameter.	B	2
	TSDEP	A unit depth	1	2
	TSAREA	Surface area corresponding to TSDEP	2	5
	TSTORE	Volume corresponding to TSDEP	2	5
	TSQOU	Outflow corresponding to TSDEP	2	5
G2	TSDEP	A unit depth	1	2
	TSAREA	Surface area corresponding to TSDEP	2	5
	TSTORE	Volume corresponding to TSDEP	2	5
	TSQOU	Outflow corresponding to TSDEP	2	5
G3	A1(1)	Depth-outflow equation coefficient.	1	5
	DO(1)	Minimum depth for outflow	1	5
	A2(1)	Depth-outflow equation exponent.	1	5
	A1(2)	Depth-outflow equation coefficient.	1	5
	DO(2)	Minimum depth for outflow	1	2
	A2(2)	Depth-outflow equation exponent.	1	5
	GEOM3	External element into which flows outflow from the second outlet.	B	2
G4	TDSTAR(1)	Depth at which TQPUMP(1) begins	1	2
	TDSTAR(2)	Depth at which TQPUMP(2) begins	1	2
	TQPUMP(1)	First pumping rate	1	2
	TQPUMP(2)	Second pumping rate	1	2
	TDSTOP	Depth below which all pumping stops	1	2
G5	STORL	Total volume of water at start	1	2
	PTCO(1)	Concentration of quality constituent # 1 at start.	4	5
	PTCO(NPOLL)	Concentration of quality constituent # NPOLL	4	5
G6	INSTRING	40-character description of diversion structure	B	1
G7	SPLITIN(1)	Inflow (First value should be 0.0)	1	2
	SPLITOUT(1)	Undiverted outflow (First value should be 0.0)	1	2
	SPLITIN(2)	Inflow	1	2
	SPLITOUT(2)	Undiverted outflow	1	2

Table 3.17 TRANSPORT conduit infiltration/inflow input parameters

line	parameter	Meaning	*U*	*T*
K1	DINFIL	Base dry weather infiltration, cfs [cms].	4	10
	GINFIL	Groundwater infiltration, cfs [cms].	4	10
	RINFIL	Rainwater infiltration, cfs [cms].	4	10
	RSMAX	Peak residual moisture, cfs [cms].	4	10
	CPINF(1)	Concentration of quality constituent # 1.	4	10
K2	MONTH(1)-(12)	Monthly degree decay: Jan-Dec	4	10

Table 3.18 TRANSPORT dry weather flow (DWF) input parameters

line	parameter	Meaning	*U*	*T*
L1		Account for daily sewage flow variations: Sunday-Saturday	4	5
L2		Account for daily sewage BOD5 variations: Sunday-Saturday	4	10
L3		Account for daily sewage SS variations. Sunday-Saturday	4	10
M1		correct daily average sewage flow to hourly flows	4	5
M2		correct daily average sewage BOD5 to hourly BOD5	4	10
M3		correct daily average sewage S.S. to hourly S.S.	4	10
M4		correct daily average sewage t.colif. to hourly t.colif.	4	10
N1	KTNUM	Total number of sub-areas within a given study area.	B	5
	KASE	Estimate sewage quality from treatment plant records.	B	5
	NPF	Total number of process flows within the study area.	B	2
	KDAY	starting day of the week.	B	2
	CPI	Consumer Price Index.	2	10
	CCCI	Composite Construction Cost Index.	2	10
	POPULA	Total population in all areas, thousands.	1	10
O1	ADWF	Total study area average sewage flow	2	5
	ABOD	Total study area average BOD	4	10
	ASUSO	Total study area average SS	4	10
	ACOLI	Total study area average coliforms	4	10
O2	TOTA	Total study area	1	2
	TINA	Total contributing industrial area	1	2

	TCA	Total contributing commercial area	1	2
	TRHA	Total contributing high income area	2	2
	TRAA	Total contributing average income area	2	2
	TRLA	Total contributing low income area	2	2
	TRGGA	Total area with garbage grinders	3	5
	TPOA	Total park and open area	1	2
P1	INPUT	Inlet junction/node/manhole number.	B	1
	QPF	Average daily process flow	2	2
	BODPF	Average daily BOD of process flow	4	5
	SUSPF	Average daily SS of process flow	4	5
Q1	KNUM	Sub-area number.	B	1
	INPUT	External number of the junction/node/manhole for inflow	B	1
	KNUM	Sub-area number.	B	1
	KLAND	Predominant land use within sub-area	B	1
	METHOD	Indicates whether water usage is metered	B	1
	KUNIT	Parameter indicating units for water usage estimates	B	1
	MSUBT	Subtotal printed after each sub-area?	B	1
	SAQPF	Total industrial process flow originating within sub-area KNUM,	2	5
	SABPF	BOD contributed from industrial process flow	4	10
	SASPF	SS contributed from industrial process flow	4	10
	WATER	Measured winter water use for sub-area	3	10
	PRICE	Cost of the last thousand gal [10 m3] of water per billing period	1	5
	SEWAGE	Measured average sewage flow from entire sub-area	1	5
	ASUB	Total area within sub-area KNUM, acres [ha].	1	2
	POPDEN	Population density within sub-area KNUM, persons/acre [pers/ha]	1	2
	DWLNGS	Total number of dwelling units within sub-area KNUM.	1	2
	FAMILY	Number of people living in average dwelling unit	1	2
	VALUE	Market value of average dwelling unit, thousands of dollars.	2	5
	PCGG	Percentage of dwelling units possessing garbage grinders	3	5
	XINCOM	Income of average family, thousands of dollars per year.	2	10

Table 3.19 EXTRAN run control input parameters

line	Parameter	Description	U	T
A1	Title1	Title	B	1
A1	Title2	Title	B	1
B0	ISOL	Solution technique parameter	B	2
	KSUPER	Use minimum of normal flow and dynamic flow	B	2
	KREDO	Use the last computed heads and flows during dry periods.	B	1
	TOLCS1	Max steady state flow imbalance	B	2
	QLOWCS	Maximum steady state outflow	B	2
	TOLCS2	Maximum change in flow	B	2
B1	NTCYC	Number of time-steps desired.	B	2
	DELT	Length of time-step.	B	2
	TZERO	Start time of simulation.	B	2
	NSTART	First time-step to begin print cycle.	B	1
	INTER	Interval between intermediate print cycles during simulation.	B	1
	JNTER	Interval between time-history summary print	B	1
	JREDO	Hot-start file manipulation parameter.	B	2
	IDATZ	Initial date of simulation, 6 digits, yr/mo/day.	B	1
B2	METRIC	U.S. customary or metric units for input/output.	B	1
	NEQUAL	Use an equivalent pipe	B	1
	AMEN	Default surface area for all manholes.	2	2
	ITMAX	Maximum number of iterations	B	1
	SURTOL	Fraction of av flow for convergence criterion for surcharge	1	1
B3	NHPRT	Number of junctions for detailed printing of head output	B	1
	NQPRT	Number of conduits for detailed printing of discharge output.	B	1
	NPLT	Number of junction heads to be plotted	B	1
	LPLT	Number of conduits for flows to be plotted	B	1
	NJSW	Number of input junctions	B	1
B4	JPRT(1)	First junction number/name for detailed printing	B	1
B5	CPRT(1)	First conduit number/name for detailed printing	B	1
B6	JPLT(1)	First junction number/name for detailed plotting	B	1
B7	KPLT(1)	First conduit number/name for detailed plotting of conduit flow.	B	1

B8	NSURF	Number of conduit upstream/downstream elevation plots.	B	1
	JSURF(1)	First conduit number/name for plotting.	B	1
B9	NOFLOW	Number of conduits	B	1
	NOFDUP	Number of additional conduits to write flows for.	B	1
	IFINTER	Number of time steps to output flows.	B	1
	FLOWMIN	Minimum flow	1	1
	FLOWOUT(1)	First conduit number/name for writing flows to ASCII file	B	1
	FLOWOUT(2)	Continue for the number of conduits defined by NOFLOW	B	1
	FLOWDUP	Additional conduit ID for which flows are to be written	B	1
	FLOWREF	Reference conduit	B	1
BA	JHEAD	Eliminate intermediate headers in output tables	B	1
	JP10	Print all 10 characters and digits at all locations in program.	B	1
	IWLEN	Irregular section conduit length are entered at two locations	B	1
BB	JELEV	Input elevations instead of depth for variables ZP1 and ZP2	B	1
	JDOWN	Use either minimum of normal or critical depth	B	1
	IPRATE	Use default of three PRATE/VRATE pairs for pump inputs	B	2
	IM2	Procedures for computing characteristic conduit parameters	B	2
	IPIPESED	Control of sediment depth on C1 line	B	1
BC	ICONTER	Number of time steps	B	1
BD	NUMBFF	The number of monthly base flow factors	B	1
	BFFMO (1)….	Monthly base flow factors separated by space	3	5
	BFFMO (2)		3	
	BFFMO (NUMBFF)		3	
BE	IBESTART	The simulation cycle for which output will start.	B	2
	IBEND	The simulation cycle for which output will end	B	2
BF	WTSTART	Time of day at which linkage file output	B	2
	IDEP	Output volumes, depths and velocities control	B	1
	NSTEPW	Number of Extran time steps per WASP time step	B	2
	IVCALC	Volume computation control	B	2
BG	IWASPSEG	WASP segment number	B	2
	ICONSEG	Extran conduit used to compute WASP segment	B	2
	NJUNSEG	Number of Extran time steps per WASP time step.	B	2
	JUNSEG (1)	Extran junctions that correspond to thisWASP segment	B	2
	JUNSEG (NJUNSEG)		B	2
BH	FROMSEG	"From" WASP segment	B	2
	TOSEG	"To" WASP segment	B	2
	NCONSEG	Number of parallel Extran conduits	B	2
	ICONSEG	Extran conduit (s) that correspond to this WASP interfacial flow	B	2
BZ	IDUMP	Parameter to control writing of hydraulics file	B	1
	DTHYD	Time step for hydraulics output file	B	1
	HYDSTR	Start time for hydraulic output file	B	1
	IVCALC	Compute volume control	B	1

Table 3.20 EXTRAN input parameters for geometrical detail of conveyances and bridges.

line	Parameter	Description	U	T
C1	NCOND	Conduit number (any valid integer), or conduit name	B	1
	NJUNC(1)	Junction number at upstream end of conduit, or junction name	B	
	NJUNC(2)	Junction number at downstream end of conduit, or junction name	B	1
	QO	Initial flow.	2	1
	NKLASS	Type of conduit shape.	B	1
	AFULL	Cross sectional area of conduit	1	2
	DEEP	Vertical depth (diameter for type 1) of conduit	1	1
	WIDE	Maximum width of conduit.	1	1
	LEN	Length of conduit.	1	1
	ZP(1)	Distance of conduit invert above junction invert at NJUNC(1)	1	2
	ZP(2)	Distance of conduit invert above junction invert at NJUNC(2).	1	2
	ROUGH	Manning coefficient (includes other losses).	1	2
	STHETA	Slope of one side of trapezoid.	2	2
	SPHI	Slope of other side of trapezoid.	2	2
	ENTK	Entrance loss coefficient.	2	5
	EXITK	Exit loss coefficient.	2	5
	OTHERK	Additional loss coefficient	3	5
CS	IREAD	Alphanumeric indicator variable	B	2

C2 or NC	XNL	n for the left overbank.	1	2
	XNR	n for the right overbank.	1	2
	XNCH	n for the channel.	1	2
C3/X1	SECNO	cross section id number	B	1
	NUMST	Total number of stations on the following C4 (GR) data lines	B	1
	STCHL	The station of the left bank of the channel.	B	1
	STCHR	The station of the right bank of the channel.	B	1
	XLOBL	Not required for EXTRAN (enter 0.0).	B	1
	XLOBR	Not required for EXTRAN (enter 0.0).	B	1
	LEN	Length of channel reach represented by this cross section.	2	2
	PXSECR	Factor to modify the horizontal dimensions	B	2
	PSXECE	Constant to be added (+ or -) to elevation data	B	2
C4/GR	EL(1)	Elevation of cross section at STA(1).	1	2
	STA(1)	Station of cross section 1	1	2
	EL(2)	Elevation of cross section at STA(2).	1	2
	STA(2)	Station of cross section 2.	1	1
C5	BRDGNO	Bridge identifier number.	B	1
	NUMHN	Number of Mannings-n station pairs on C6 lines.	B	1
	NUMST	Number of elevation-station pairs on C7 lines.	B	1
	NMPIER	Number of piers on C8 lines.	B	1
C6	VMAN(1)	Mannings n from left of bridge to station STMAN(1)	1	2
	STMAN(1)	Station of first change in n or location of first smooth pier.	1	2
	VMAN(2)	Manning's n from STMAN(1) to STMAN(2)	1	2
	STMAN(2)	Station of second change in n or location of second smooth pier.	1	1
C7	ELSTA(1,1)	Elevation of first cross-section point	1	2
	ELSTA(2,1)	Station of first cross-section point	1	2
	LSTA(1,2)	Elevation of second cross-section point	1	2
	ELSTA(2,2)	Station of second cross-section point	1	2
C8	PIERW(1)	Pier width of first pier	1	2
	PCLSTA(1)	Centerline station of first pier	1	2
	CHORDL(1)	Low chord elevation at first pier	1	2
	PIERW(2)	Pier width of second pier	1	2
	PCLSTA(2)	Centerline station of second pier	1	2
	CHORDL(2)	Low chord elevation at second pier	1	2
CS	IREAD	Alphanumeric indicator variable	B	1
D1	JUN	Junction number (any valid integer), or junction name	B	1
	GRELEV	Ground elevation.	1	2
	Z	Invert elevation.	1	2
	QINST	Net constant flow into junction.	2	2
	YO	Initial depth above junction invert elevation.	2	2
	XLOC	Junction X location	B	1
	YLOC	Junction Y location	B	1
D2	NTHRESH	Number of elevation thresholds entered for each junction	B	1
	THRESHID (1)	Name and ID for first elevation threshold	B	1
	THRESDID (2)	Name and ID for second elevation	B	1
	THRESDID (N)	Name and ID for last elevation	B	1
D3	JUNTHR	Junction number or name	B	1
	THRESH (1)	First elevation threshold to be include in output table	B	1
	THRESH (2)	Second elevation threshold to be include in output table	B	1
	THRESH (3)	Last elevation threshold	B	1
E0	NVSPR	Do not print echo of all E1-E2 input data, only print summary.	B	1

Table 3.21 EXTRAN input parameters for storage units

line	Parameter	Description	U	T
E1	JSTORE	Junction number containing storage facility, or junction name	B	1
	ZTOP	Junction crown elevation	1	2
	ASTORE	Storage volume per foot (or meter) of depth	1	2
	NUMST	Total number of stage/storage area points on following E2 data lines	B	2
E2	QCURVE(1,1)	Surface area of storage junction at depth point # 1,	1	2
	QCURVE(2,1)	Depth above junction invert at point 1	1	2

Table 3.22 EXTRAN input parameters for orifices, weirs and pumps

line	Parameter	Description	U	T
F1	NJUNC(1)	Junction number containing orifice, or junction name	B	1
	NJUNC(2)	Junction number/name to which orifice discharges	B	2
	NKLASS	Type of orifice.	B	2
	AORIF	Orifice area. Required just for circular orifices.	1	2
	CORIF	Orifice discharge coefficient.	1	5
	ZP	Distance of orifice invert above junction invert	1	2
	DEEPO	Vertical depth or height of rectangular orifice opening	1	2
	WIDEO	Horizontal width of rectangular orifice opening.	1	2
F2	NTIME	Number of data points to describe time history of orifice	B	2
	VORIF(1,1)	First time that orifice coefficient and area change	B	2
	VORIF(1,2)	First new value of orifice discharge coefficient.	1	2
	VORIF(1,3)	First new value of orifice area.	1	2
F3	ICNODE	Control node at which the depth controls the gate	B	2
	OOPEN	Depth at control node at which the gate opens.	1	2
	OCLOSE	Depth at control node at which the gate closes	1	2
	OCAREA	Close orifice area	1	2
	ORATE	Time required to move gate	1	2
	IDIR	Flow direction control	B	2
	IOPRNT	Controls intermediate output summary	B	1
F4	ICNODE	Control node at which the depth controls the gate	B	2
	OOPEN	Depth at control node at which the gate is fully open	1	2
	OCLOSE	Depth at control node at which the gate is fully closed	1	2
	OCAREA	Close orifice area.	1	2
	ORATE	Time required to move gate	1	2
	IDIR	Flow direction control	B	2
	IOPRNT	Controls intermediate output summary	B	1
G1	NJUNC(1)	Junction number at which weir is located, or junction name	B	2
	NJUNC(2)	Junction number to which weir discharges, or junction name	B	2
	KWEIR	Type of weir.	B	2
	YCREST	Height of weir crest above invert.	1	2
	YTOP	Height to top of weir opening above invert	1	2
	WLEN	Weir length, ft [m]. Not used for V-notch weirs.	1	2
	COEF	Coefficient of discharge for weir.	1	5
	ISUBEQ	Submergence equation	B	5
	ENDCON	Number of end contractions	B	1
	THETAV	Angle in degrees of V-notch or trapezoidal end sections	1	2
	COEF2	Coefficient for triangular portion of trapezoidal weir	1	2
	HEADW	Water elevation on upstream side of weir above weir crest	1	2
H1	IPTYP	Type of pump.	B	5
	NJUNC(1)	Junction number being pumped, or junction name	B	2
	NJUNC(2)	Pump discharge goes to this junction number, or junction name	B	2
	NRATES	Number of PRATE/VRATE pairs.	B	2
	PRATE(1)	Lower pumping rate.	1	2
	PRATE(2)	Mid-pumping rate.	1	2
	PRATE(3)	High pumping rate.	1	2
	VRATE(1)	If IPTYP = 1 enter the wet well volume for mid-rate pumps to start	1	2
	VRATE(2)	If IPTYP = 1 enter the wet well volume for high-rate pumps to start	1	2
	VRATE(3)	If IPTYP = 1 enter total wet well capacity.	1	2
	VWELL	If IPTYP = 1 then enter initial wet well volume.	1	2
	PON	Depth in pump inflow junction to turn pump on.	1	2
	POFF	Depth in pump inflow junction to turn pump off.	1	2
	PONDELAY	Time for pump rate to increase from zero to full capacity	1	2

Table 3.23 EXTRAN input parameters for tides

line	Parameter	Description	U	T
I1	JFREE	Number/name of outfall junctions without tide gate	B	1
	NBCF	Type of boundary condition	B	2
I2	JGATE	Number/name of outfall junctions with tide gate	B	1
	NBCG	Type of boundary condition	B	2
J1	NTIDE	Boundary condition index.	B	2

J2	A1	First tide coefficient.	1	5
	W	Tidal period, hours. Required only if NTIDE(I) = 3 or 4.	1	5
	A2	Second tide coefficient.	1	5
	A3	Third tide coefficient.	1	5
	A4	Fourth tide coefficient.	1	5
	A5	Fifth tide coefficient.	1	5
	A6	Sixth tide coefficient.	1	5
	A7	Seventh tide coefficient.	1	5
J3	KO	Type of tidal input.	B	5
	NI	Number of information points.	B	5
	NCHTID	Tide information print control.	B	5
	DELTA	Convergence criterion for fitting the tidal function	B	5
J4	TT	Time of day, first information point, hours.	B	5
	YY	Tide/stage at time above	1	5

Table 3.23 EXTRAN input parameters for input functions

line	Parameter	Description	U	T
K1	NINC	Number of input nodes and flows per line on line K3.	B	2
K2	JSW	First input node number/name for line hydrograph	B	2
K3	TEO	Time of day, decimal hours.	B	2
	QCARD(1)	Flow rate for first input node.	1	2

In the next section we show how detailed information on the sources of input data and their probable uncertainty can be archived by annotating your input data file.

> Assess in broad categories the uncertainty associated with the estimates of all parameters.

3.5 Some samples of well-documented data files

Occasionally one encounters large and comprehensively managed input data files, perhaps hundreds of pages long, if not more. In this section we have included an extensive excerpt from a real data file, to illustrate how input data files can and should be annotated. For the following input data file I am grateful to Don Waye of Annandale, Virginia. Note also that Don's file is formatted to work only with his own modification of the SWMM4 engine (which runs under the PCSWMM GUI). Excerpts are taken from the Extran portion of his Runoff-Extran model (4MR EXTRAN Baseline 8-7-00, Q100.dat).

Figure 3.3, an annotated screen dump, shows how the GIS utility in PCSWMM accommodates comments, which can be reinforced with true-scale visuals. In fact GIS provides excellent facilities for annotating input data, as described at the end of this chapter.

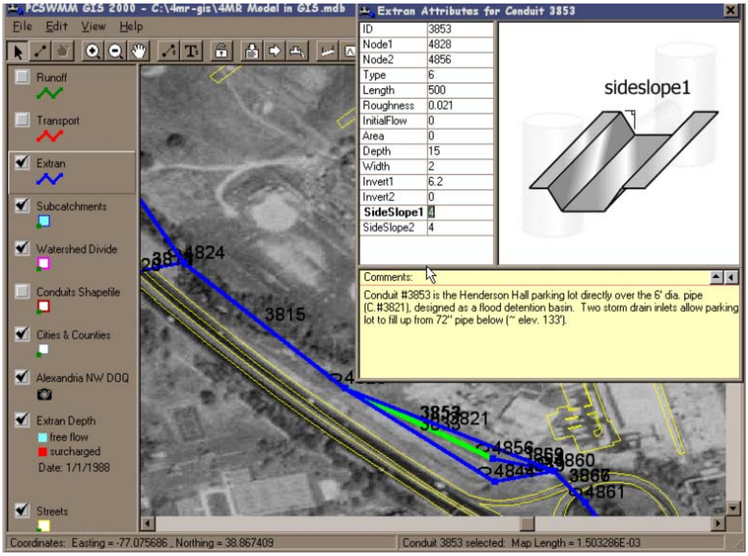


Figure 3.3. PCSWMM screen dump of an annotated graphic.

The following excerpts are taken from Waye's EXTRAN input data file.

```
*****************************************************************************
*****************************************************************************
***                                                                       ***
***                         FOUR MILE RUN MODEL                           ***
***                   MODIFIED PC-SWMM EPA VERSION 4.4                    ***
***                              BY NVPDC                                 ***
***                                                                       ***
*****************************************************************************
*****************************************************************************
***                                                                       ***
***                    <File Name:   4MR100.IN>                           ***
***                    <File Date:   January 18, 2001>                    ***
***                                                                       ***
*****************************************************************************
*****************************************************************************
*                                                                           *
*  File Documentation (all improvements made by Don Waye, NVPDC,            *
*                                                       unless noted):      *
*                                                                           *
*    1. Replaced each RUNOFF drain with EXTRAN equivalent.                  *
*                                                                           *
*    2. Added EXTRAN node info based on RUNOFF pipe slopes.                 *
*                                                                           *
*    3. All numbers in the 6000's are data from RUNOFF converted to EXTRAN. *
*                                                                           *
*    4. It represents watershed conditions as of the 1st qtr. of 1991.      *
*                                                                           *
*    5. It simulates the four hour, 100 year storm.                         *
```

```
*    6. It includes peak hydrograph shaving in all subcatchments, just like *
*       the mainframe model.                                               *
*                                                                          *
*    7. Changes made after 5/96 were not made to the mainframe model.      *
*                                                                          *
*    8. The 10th Street South drainage channel (Bailey's Branch) represen- *
*       tation is greatly improved (better subcatchment boundaries, EXTRAN- *
*       modeled drainage) over the mainframe model.                        *
*                                                                          *
*    9. Doctor's Run detail has been added, as per request by Arlington Co. *
*       DPW, May 12, 1994.  Much greater detail and significant pipe slope *
*       corrections were added in Dec. 1996.  Neither the 1994 nor 1996    *
*       details were added to the mainframe model.  (DW, 12/96)            *
*                                                                          *
*   10. Nauck Branch has been hydraulically reconnected to the model, and  *
*       Conduit 2102 has been disconnected from the model. (DW, 6/96)      *
*                                                                          *
*   11. All drainage elements in Arlington's "Virginia Highlands" Branch   *
*       have been corrected for conduit length, equivalent diameters, and  *
*       inverts.  Additionally, new Conduits 2918, 2922 & 2926 have        *
*       replaced existing Conduit 2916.  New Nodes (6918, 6922 & 6926) for *
*       these conduits have also been added, & old Node 6916 was removed.  *
*       (DW, 7/96)                                                         *
*                                                                          *

        snip

*                                                                          *
*  As of June 29, 2000, this is the BEST representation of the Four Mile   *
*  Run Model available on the PC that runs without errors. This model      *
*  replicates results of the mainframe model.  See also file 4MR-DS96.IN1  *
*  for subcatchment storage node methodology.                              *
*                                                                          *
****************************************************************************
*                                                                          *
*  Because of the large number of EXTRAN elements in this data set (over   *
*  500!), a specially compiled version of SWMM was used. It is called      *
*  "NVSWMM44.EXE", and is located on the hard drive in the "C:\SWMM 4.4\   *
*  NVPDC Code\" directory.  A batch file called 4MR.BAT has been set up    *
*  that automates use of NVSWMM44.EXE. Just type "4MR" at a DOS prompt to  *
*  access this model executable.                                           *
*                                                                          *
*  This data set generates an acceptable SWMM output data set with         *
*  significantly reduced continuity errors over pre-"RUNOFF-to-EXTRAN"     *
*  converted data sets. It was necessary to make these changes in order to *
*  represent detention storage in an appropriate, realistic, and explicit  *
*  manner.                                                                 *
*                                                                          *
****************************************************************************
*                                                                          *
*  Twelve optional, but recommended, steps remain:                         *
*                                                                          *
*    1. Replace the three remaining "hydrograph input" subcatchments using *
*       aerial photos and local detailed storm drain maps or development   *
*       plats.  Be sure to represent any detention storage for these sub-  *
*       catchments via E1 storage nodes or peak hydrograph shaving, as     *
*       appropriate.  To date, four subcatchments have been reinstated.    *
*       The three remaining subcatchments to be reinstated are:            *
*        Old #'s  318,  319 &  360.                                        *
*       (New #'s 1617, 1621 & 1725, respectively; they're all in Alexandria)*
*                                                                          *
*    2. Replace any remaining input hydrographs represented by the "K1"    *
*       thru "K3" data section in EXTRAN.                                  *
*                                                                          *
*    3. Add the E1 storage nodes to EXTRAN. Review method with Bill Frost  *
*       of Arlington DPW (and others, if necessary). Document method.      *
*       Delete S1--Hydrograph Shaving Option Inputs--at that point.        *
```

```
*                                                                          *
*    4. Improve Green/Ampt infiltration modeling in RUNOFF Block, Group H   *
*       (subcatchments).  Derive Green-Ampt parameters from new NRCS soils  *
*       data.  Refer to Water Resources Bulletin (AWRA), Feb. 1996,         *
*       pp. 125 - 136 for a very good & simple method for doing this.  SWMM *
*       documentation indicates that my SDMAX values are fine for continuous*
*       simulation mode, but make the soil behave too dry for single-event  *
*       mode, thus preventing early runoff.  For single event mode, it is   *
*       recommended that SDMAX be reduced.                                  *
*                                                                          *

        snip

*                                                                          *
*  After these changes have been made, the conversion process should be     *
*  documented for posterity and reference, as well.                         *
*                                                                   -DW  *
****************************************************************************
****************************************************************************
**                                                                        **
**=======================> CALL TO EXTRAN BLOCK:                           **
**                                                                        **
****************************************************************************
****************************************************************************
*
$EXTRAN
*
*  TWO TITLE CARDS (up to 76 spaces per line):
A1  ' Analysis of VDOT mods to 244/27 interchange. '
A2  ' Existing conditions.  Run date = July 20, 2000. '
*
*  All nodes from the former Runoff block have been renamed as the 6000 series.
*

        snip

*
*  Note: Manning's "n" adjusted to account for head losses explicitly modeled
*        with WREM (Mainframe Model) in Conduits 3131, 3139, 3141, 3149,
*        3161, 3245 & 3249.
*
*                         Conduit
* Conduit  u/s   d/s Init. |  X-sec. Depth  Width  Length  Vert.Offset  Mann.
Side Slopes Ent Exit Other
*     #   Node  Node   Q Type Area   (ft)   (ft)   (ft)    u/s     d/s   "n"
A     B    K    K    K
*   ----  ----  ----  --- --  ----   ----   ----   -----   ---     ---  ----   -
--   ---  ---  ---  ---
* Begin upper mainstem, Four Mile Run
C1  2100  6100  4102  0.0  1   0.0   4.75    0.0   1400.    0.0    0.0  .013
0.0   0.0  .00  .00  .00
C1  3101  4100  4102  0.0  6   0.0    14.    10.    690.    0.0    0.0  .124
1.0   1.0  .00  .00  .00
* C.#3103 is a triple 8'x8' box culvert under I-66.
C1  3103  4102  4104  0.0  2   0.0     8.    24.    788.    0.0    0.0  .013
0.0   0.0  .00  .00  .00
* C.#3105 side slopes changed from 2:1 to 1:1.  -DW, 11/98
C1  3105  4104  4106  0.0  6   0.0    13.     8.    597.    0.0    0.0  .075
1.0   1.0  .00  .00  .00
*  Begin Trammell Branch (Falls Church, upper mainstem, Four Mile Run)
* C.#2104 comes in as a 75" x 50" "squashed" ellipse CMP adjacent to the two
4'x10' BC's
* on south side.  -DW, 10/98
C1  2104  6104  4108  0.0  1   0.0   4.67    0.0   3313.    0.0    0.0  .022
0.0   0.0  .00  .00  .00
*  End Trammell Branch
* Continue upper mainstem, Four Mile Run
```

```
* Conduit 3107 is the double 4'x10' box culverts under the Church Courts Apts.
in Arl.
C1  3107  4106  4108  0.0  2   0.0     4.    20.    710.    0.0    0.0  .015
0.0   0.0  .00  .00  .00
* Conduit 3111 represents surface flooding (overland relief channel) through the
Church
* Courts Apts.  *** DO NOT MODIFY, ELSE CONTINUITY ERRORS MAY COMPOUND! ***  -
DW, 12/98
C1  3111  4106  4108  0.0  6   0.0     4.    75.    750.    5.1    5.0  .025
3.0   3.0  .00  .00  .00
*

         snip

*
****************************************************************************
*  JUNCTION (NODE) DATA IN INPUT GROUP D1                                  *
*                                                                          *
* Notes:                                                                   *
*                                                                          *
*  1) Some nodes in the 4000 series have ground elevations raised by 5 to   *
*     10 feet higher than actual ground elevation.                         *
*                                                                          *
*  2) Inverts for the 6000 series were determined by multiplying the length *
*     times the slope and adding it to the invert directly downstream,     *
*     unless otherwise noted.                                              *
*                                                                          *
*  3) If the line ends with 3 asterisks (***), then the following applies:  *
*     Drainage was originally modeled in the Runoff Block, but when these   *
*     sections were transferred over to the EXTRAN block, the model        *
*     generates ground elevation errors.  New ground elevations were       *
*     determined by adding the pipe's diameter plus one (1) foot to the    *
*     invert.  (Christina Kollay, June 1994)                               *
*                                                                          *
*  4) Added from 0.4 to 1.0 cfs of constant inflow at 55 strategic nodes in *
*     the watershed to help simulate groundwater contributions to baseflow. *
*     -DW, 6/97                                                            *
*                                                                          *

         snip

D1  6249  252.2   240.0   0.0   0.0
*
* 6253 drains Subcat. #1253.  GRELEV increased from 241.9 for flood
* relief channel.  -DW, 3/27/98
D1  6253  247.0   234.0   0.0   0.0
* !!! E S T I M A T E D !!!
*
* 6259 drains Subcat. #1259.  -DW, 3/27/98
D1  6259  250.0   243.0   0.0   0.0
* !!! E S T I M A T E D !!!
*
* 6265 drains Subcat. #1265.  It's also at the upper end
* of the box culvert under Rte 50 (Conduit #3271)  -DW, 3/27/98
D1  6265  205.0   174.0   0.0   0.0
* !!! E S T I M A T E D !!!
*
* N.#4256 GRELEV was increased from 241.5' to accomodate flood relief
* channel #2253.  -DW, 4/00
D1  4256  244.6   231.0   0.0   0.0
*

         snip

F1   4140    4152     2    33.2     0.60    0.0
*
* Orifice 4240 is a small (2' dia.) outlet just west of Ballston Beaver Pond
```

```
* for VDOT I-66 detention storage between on-ramp & off-ramp.  A field check
* on 7/17/97 revealed that the orifice is 85% clogged.  Consider modeling the
* effects of the clog.  A double 8' x 6' box culvert runs from 4244 to 4248.
* See also Weir #4240.
*F1   4240    4244     2    3.14     0.60    0.0
* The line below adjusts for the clogged orifice.
F1   4240    4244     2    0.47     0.60    0.0
*
   snip

J2  4.8
*
*+-----------------------------------------------------------------------+
*| Note 1: 11/29/95, -DW; Add a representative "bypass" (hydraulic overload)|
*| input hydrograph to simulate bypass flows at Arlington's WWTP on lower   |
*| Four Mile Run (Node #4916).  Normal, constant WWTP flows are represented |
*| in Group D1 Junction inputs (29.9 MGD = 46.26 cfs).                      |
*|                                                                          |
*| Note 2: 11/29/95, -DW; My goal is to eliminate all input hydrographs in  |
*| Group K. Most of these are associated with old "site studies." See the   |
*| appropriate paper files for alternative representation.  Or simply re-   |
*| determine imperviousness and any other pertinent subcatchment factors    |
*| affected by the site's development.                                      |
*+-----------------------------------------------------------------------+
*
*  Input hydrographs in Group K1 - K3
*   NINC (# of input nodes per line in K2 & # of flows per line in K3)
K1   10
*
*  K2 NOTE: Be sure that NJSW agrees with the total number of input hydro-
*           graphs provided in the K2 input group. NJSW is found in the B3
*           input group!
*
*  Node #4332 is probably associated with Doctor's Hospital (now Vencor Hosp.)
*             Stor. vol. = 22,970 cf for 15.68 ac.  See 3rd Qtr. 1979 Report.
*  Node #4636 may be associated with either Douglass Park or Concord Mews on
*             N. side of Four Mile Run near Four Mile Run Drive and Walter
*             Reed Drive.  Its removed 15-minute inputs were: 0.00/0.21/1.87/
*             4.83/6.93/8.52/14.28/23.19/85.45/191.17/112.19/45.90/40.59/
*             40.59(again!)/39.70/38.60/11.82/0.00/0.00/0.00/0.00//
*  Node #4740 is probably associated with the 90,700 cf det. storage pond in
*             the Marlboro Subdivision in Alexandria (Subcat. #1725).
*  Node #4604 is associated with missing subcatchments 1621 & 1617 and Park
*             Center detention storage.
*  Node #4332 is probably associated with the 22,970 cf storage from No. Va.
*             Doctors Hospital (now Vencor Hospital) from the last half of
*             1979 (611 S. Carlin Springs Road, Arl.)
*

        snip
```

Input data files should be carefully and comprehensively annotated to record the origins and purposes of the data, and its limitations. There is no limit to the amount of useful data that should accompany the data.

3.6 Urban water systems GIS concepts

Models of urban water systems apply basic equations of computational hydraulics to real, natural and engineered or modified environments. Elements of real water systems are placed in space, and their locations are fixed by global co-ordinates and positioning systems. Fixing the elements in the model by means of their spatial co-ordinates allows all the information related to the water system and its elements to be efficiently stored, located, recalled, manipulated and displayed.

In this section we briefly mention some aspects of spatially related information systems. Many GIS programs are available, but in this book we use the associated decision support system PCSWMM, which has two main focuses (the GIS code in PCSWMM is also used in other well-known programs):

1. The first is as a standalone GIS for graphically creating, editing and/or querying SWMM entities and attributes, displaying these SWMM layers with background layers and dynamic model results, and exporting data to SWMM input files.
2. The second focus is an interface between a GIS/IMS and SWMM. PCSWMM interfaces directly with the underlying database(s). ESRI, MapInfo and AutoCAD layers/themes can all be displayed. Its look and feel is most similar to ArcView.

Data is extracted directly from the GIS's underlying database (MS Access or ODBC compliant) using SQL queries and setup in the PCSWMM internal database (MS Access) for data tweaking into a suitable model (i.e. element aggregation, etc.). External access to the PCSWMM internal database is also possible.

Raster images and vector drawings (*.bmp or meta-files) can be used as backgrounds with completely accurate scaling (as long as one has geo-referenced files). Supported raster file formats include: .TIF, .BMP, .JPG, .PCX, .DIB, and .TGA. Other (vector) layer formats supported are: MapInfo MIF, MicroStation DGN, AutoCAD DXF, AutoCAD DWG, and ESRI Shape File.

The interface can be used to automate data input such as watershed areas, land use, and soils-related parameters such as runoff curve numbers or infiltration rates. Node, conduit, and subcatchment data are pulled from the underlying GIS database (ODBC) into an intermediate database for processing into a useful model. This data is then exported to a SWMM4 input file (Runoff, Transport or Extran). Supported parameters are listed below. Watershed area and conduit lengths can be calculated by PCSWMM.

Attributes supported by PCSWMM are:

Node Attributes:

ID,
easting (or longitude)
northing (or latitude)
ground elevation (Extran only)

invert elevation (Extran only)
inflow
initial depth (Extran only)
up to 4 pollutant concentrations in inflow (Transport only)

Conduit Attributes:
ID
upstream node
downstream node
type,
length,
roughness,
up to three measurements for determining cross-sectional shape (depending on type),
invert elevation at each end (Extran only),
side slopes (if applicable),
initial flow (Extran only),
initial depth (Runoff only)

Subcatchment Attributes:
ID
inlet node
width
area
impervious area
slope
roughness of impervious area
roughness of pervious area
storage in impervious area
storage in pervious area
3 infiltration parameters (for Horton's or Green-Ampt equation)

PCSWMM can calculate conduit lengths and subcatchment areas from the map display. Also, the Aggregation wizard calculates equivalent parameters when aggregating two or more conduits to a single conduit. As for updating SWMM4 input data files, one simply associates an input file with a SWMM4 layer (RUNOFF, TRANSPORT or EXTRAN) and clicks on the 'Export to Input File' menu item. The selected elements are written to the input file.

PCSWMM cannot determine input parameters directly from shape files. However, as stated above, the program can calculate conduit lengths and subcatchment areas from the SWMM4 layers. Other attributes must be filled in by the user, imported from the underlying GIS attribute database, or imported from spreadsheets or other databases. True GIS (ArcView and/or ArcInfo for example) products are best suited for other types of attribute determination (e.g.

subcatchment slope, percent imperviousness, etc.). Once these attributes have been determined and exist in the GIS, PCSWMM facilitates the data extraction and model development.

3.7 Summary

This chapter describes both the hydro-meteorological TS data and environmental parameter data. Good short term TS data are required for calibration, and credible long term TS for design/inference. Separate environmental data files are required for the *as-is* and each *to-be* scenario. The associated uncertainty is suggested. Excerpts of annotated input for SWMM4 RUNOFF and EXTRAN are given. Graphical representation of geo-referenced data (GIS), and its implementation, is covered in more detail than in the previous chapter.

References

Dunn, A.R., 1986. Feasibility of a Real-Time Auto Sensitivity and Error Analysis Framework for Fuzzy Stormwater Modelling on Personal Computers, M.Eng. Thesis, McMaster U., May 1986, 130pp.

Dunn, A.R. and James, W., 1985. Extended Expert Systems. Proc. of the Stormwater and Water Quality Modelling Conf., sponsored by USEPA and Ontario Ministry of the Environment, Toronto, Canada, published by CHI, Hamilton, pp. 373-380, Dec 5-6, 1985.

Bibliography

Bhaskar, N.R., James, W.P. and Devulapalli, R.S. 1992. Hydrologic parameter estimation using geographic information system. J. Water Resources Planning and Management. 118(5):492-512.

Dennison, K.D. and James, W. 1985. "A Database Environment and Sensitivity Framework for Continuous Water Quality Modelling", Proc. of the Stormwater and Water Quality Conf., sponsored by USEPA and Ontario Ministry of Environment, Toronto, Ont. published by CHI Hamilton, pp. 101-124, Dec 5-6, 1985.

Leavesley, G.H., and Stannard, L.G. 1989. A modular watershed modelling and data management system. Flood forecasting Symposium and Workshop. Mar-Apr 1989 NWS. pp71-95

Sircar, J.K. 1986. Computer Aided Watershed Segmentation for Spatially Distributed Hydrologic Modelling. Ph.D., Univ. of Maryland,

Stephenson, D., 1989. Selection of stormwater model parameters. J. of Environmental Engrg. 115(1): 210-220.

Chapter 4

OPTIMAL MODEL COMPLEXITY

> ***Complexity****: in science, field of study devoted to the process of self-organization. .... Complexity looks for the mathematical equations that describe the middle ground between equilibrium and chaos ... Much of the progress in the field can be attributed to advances in nonlinear dynamics, in the power of computers and in computer graphics, and in adaptive programs and fuzzy logic.*
>
> - The Columbia Encyclopedia, Sixth Edition. 2001.

4.1 Introduction

In this chapter we discuss model complexity, and how to determine the best complexity to solve a design problem at a given uncertainty. This usually boils down to finding the smallest suitable number of parameters and sub-spaces (parsimony).

Two types of model complexity may be considered:

1. complexity caused by process disaggregation, where unnecessary processes are activated, perhaps at an unnecessarily fine time resolution. HSPF is an example of a WSM with such a reputation (though when used for certain simple, design problems, it can be effectively simplified) and
2. complexity caused by spatial discretization, where an unnecessarily large number of sub-spaces such as sub-catchments, channels, pipes, etc are modeled.

Because of the cost of collecting, monitoring, analyzing and abstracting data for a large number of parameters, and variables, it seems desirable to reduce both process disaggregation and spatial discretization to the minimum needed for sound design. In other words, we seek model parsimony.

On the other hand, to improve the accuracy and reliability of the model it is desirable to include as many relevant processes, at as fine a spatial resolution as is reasonable. By no means, however, is it agreed in the literature that model reliability continues to increase with model complexity, evidently because of the difficulty of getting good parameter estimates, and their combined effect on the computed response (Seo,1991).

> Model complexity depends on both the level of discretization (compartmentalization) and the number of processes active (disaggregation).

Optimal model complexity is very much dependent on the evaluation function chosen, e.g. for cost-effectiveness. To find the minimum, the function should exhibit a minimum when plotted against a number representing the complexity. Also, the function should penalize model inaccuracy for very simple, coarse and inaccurate models, and penalize cost for very large and complex models. Unfortunately this introduces a degree of subjectivity. It is possible to estimate design-office costs as a function of collecting various types of field data, but it is not nearly so easy to place a dollar value on model reliability. The evaluation function may be written:

$$\Phi = f(\rho) + f(\varepsilon)$$

where:

ρ = design office costs, including field data collection, and
ε = model error.

One would expect Φ to be large for unacceptable error, smallest when the model achieves the requisite accuracy, and increase with large complexity, due to cost. Figure 4.1 depicts the relationship sought (shapes will depend on the functional relations chosen).

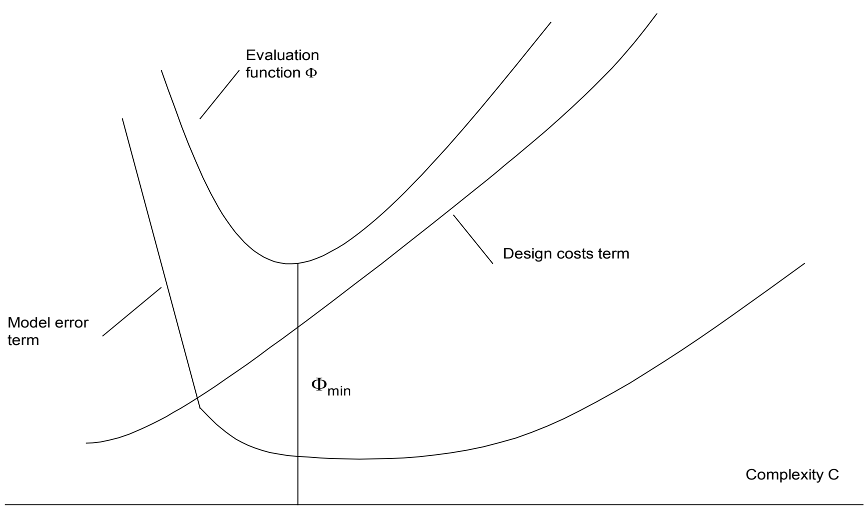


Figure 4.1: Model cost effectiveness as a function of complexity.

Systematic optimization of model complexity has evidently not been applied to WSMs in practice. The topic has been explored by Qaisi (1985) and Seo (1991), who both found that, as parameter uncertainty increases, a less complex model may be chosen as optimal. In other words, where there is a large degree of parameter uncertainty, relatively simple models should be used. Extending the argument, it is also evident that relatively simple models may be selected in situations where field data are seriously lacking. This intuitively correct result seems to be effectively and widely practiced in storm water management modeling.

Today the situation is becoming increasingly complicated by the ready availability of 1. advanced models, and 2. remotely-sensed information such as topography and hydrometry. Moreover, in developed areas, facilities and asset management databases may be inexpensively and easily imported into programs like PCSWMMGIS. These factors combine to reduce the costs of complex models, and will continue to do so for the foreseeable future. However in many undeveloped areas, much of the required data remains elusive, including even TS like rainfall. It seems intuitively obvious that, since the minimum desirable complexity is a function of the availability of required data, this target complexity will continually move upwards as time goes by and more data is collected, organized and distributed, perhaps over the web. Extrapolating the argument to its limit, we can look forward to using tens of thousands of sub-catchments and pipes in solving routine design problems.

4.2 Earlier work on complexity

Among the four important principles of model selection espoused by Yan (1990) two emphasize the problem of complexity:

1. *Physical Principle:* If some natural characteristics and physical relationships of the system are known, the model which describes these natural characteristics and relationships most accurately, should be selected.
2. *Simplicity Principle:* For the same level of predictive exactness, the simplest model should be selected.

There has been some support for the idea that models should be developed by proceeding from the simple to complex (Nash and Sutcliffe, 1970; Bergstrom, 1991; Biswas, 1976). In this strategy, complexity is systematically increased so long as the computed output is significantly improved. Bergstrom evidently used visual inspection of scatter plots to evaluate model performance, and found that the point of no model improvement was reached surprisingly soon.

An early study on the optimal degree of lumping in applications of SWMM was carried out by Proctor et al. (1976). Their work was not related to water quality, continuous modeling, or sensitivity and error analysis (S&EA). They developed some simple rules for reducing the number of conduits modeled, artificially compensating for the resulting loss of conduit storage. Their methods were applicable in the era before personal computers.

One method advocated from time to time in the literature for logically reducing the number of parameters to be tested is to carefully examine the formulations of all the basic equations, and, where several parameters are integrated into a single expression, to vary just one of those constituent parameters. For instance in SWMM-RUNOFF the characteristic sub-catchment width W, the sub-catchment slope S and the Manning roughness coefficient n are all combined into one FORTRAN variable WCON in the code. Varying just one of them is, in the opinion of the present writer, risky, because programmers and model developers frequently borrow current values of input parameters for insertion into code for unrelated processes elsewhere in the program. When certain parameters are eliminated from consideration, these parameters may be overlooked in the S&EA for other processes.

A case in point is the study by Baffaut (1988) who at that time was only interested in S&EA for quantity-oriented flow processes in the RUNOFF module of SWMM. She found that only the above three parameters in RUNOFF were combined in such a way and concluded that an appropriate change in any of these three parameters would cause an equivalent variation in the computed result. Similar conclusions had been made previously by Cunningham and Huber (1987) and Jewell et al. (1978). In fact, the slope is also used in the universal soil loss erosion algorithm, and the width may be used in the snow pack areal depletion processes.

It seems unnecessarily complicated to arbitrarily select certain parameters for S&EA in one process and not another, and so this artifice has not been adopted in PCSWMMGIS, especially since so few parameters may be eliminated.

Bergstrom (1991) considered the WMO (1975, 1986) model studies on snowmelt modeling, covering a span from simple degree-day methods to complete energy balance models. The significant conclusion from these studies was that it was not possible to rank the models in terms of their performance. Model complexity could not be related to the accuracy of the computed output.

Many writers have argued that model performance deteriorates as model complexity increases. Andersson (1992), whose models were also dominated by phase-change (snowmelt and evaporation) processes, postulates that model improvement cannot be achieved by increasing the complexity of component sub-routines, unless problems of spatial resolution [of rainfall in our case] and discretization of physiographic parameters are also resolved. Andersson did not provide guidelines for the level of complexity at which performance deterioration may be expected.

Other modelers have published summaries of studies on discretization, e.g. Litchfield and Warwick (1993), who found that the coarsest level of spatial aggregation was unsatisfactory, while two others were adequate. They used four events for calibration, and studied a 86-acre area. The coarse model used only one hydrologic unit; the medium 4 subcatchments, 11 pipes and 8 manholes; and the fine model 30 sub-catchments, 23 inlets, 36 pipes and 14 manholes. Readers should examine the paper to satisfy their own curiosity, because they may well interpret the results to be inconclusive, depending on their own particular model objectives.

That the literature should indicate inconclusive results for various levels of discretization, is certainly an important finding. Ultimately the optimum or simplest level of discretization will be determined by the design problem objectives. For a simple real-time control problem, where alternative design scenarios are not being tested, and where a simple YES/NO decision must be made in limited time, a very coarse operational model may suffice.

> Rule: In determining the best level of complexity, test simple models first, proceeding to more complex, until the required accuracy of the computed response function is achieved. Use the least number of processes, discretized spaces, and the biggest time step that delivers the required uncertainty.

4.3 Choosing the best complexity

Over the last century the engineering profession has been preoccupied with the development of more detailed deterministic descriptions. In fact choosing the best model for a management situation is often the problem of choosing the appropriate level of detail and complexity that will give the best overall project outcome. Frequently there is discussion as to whether a simpler or a more complex model is appropriate in each model application. Qaisi (1985) says that it is very important to know the appropriate level of complexity required to describe a given system with minimum level of uncertainty. The most common characteristic used to compare models is complexity (John and Flynn, 2000).

Choosing the best model for an application is one of the most difficult aspects of the modeling process, and has a major effect on the success of the project. However the criteria given for selecting the level of detail or complexity seem to consist almost entirely of vague principles and general guidelines. There is a need for a more quantitative method to allow comparison among models.

In current engineering design practice, decision-making is generally based on qualitative characteristics of models, such as those listed in Table 4.1.

Table 4.1 - Advantages and disadvantages of simplistic and complex models. Compiled and adapted from Brooks and Tobias (1996).

Simple Model	Complex Model
May be unrealistic	Requires considerable resources to be built
Less portable	Give more detailed results
Easier to understand	More difficult to understand
Runs on any computer	Demands better hardware
Less likely to contain errors	More likely to contain errors
Easier to verify	Harder to verify
Requires less input data	Requires more input data
Easier to modify in case the design objectives change	Allows the investigation of the effect on the system of many factors
Allows quicker results	Offers more flexibility to the model user, adapting to different environments
Requires less skilled modeler	Requires skilled modeler

4.4 Distinguishing between level of detail and complexity

There is no universal definition of either *level of detail* or *complexity*. Although these terms are often used synonymously, they are differentiated here. In hydrological and water quality modeling, complexity is more commonly used than level of detail but is used variously, with extensive discussion about its meaning.

Level of detail is a description of the system, related to the number of system elements included in the model. As shown in Table 4.2, the Source Loading and Management Model (SLAMM) for examining the behavior of runoff and storm water quality in urbanized areas) describes land use in a watershed by separating areas into different categories, such as roofs, driveways and streets. In comparison with a model that describes these elements of the system as one single category, SLAMM would be considered to be more detailed. However a model that has a higher level of detail should not necessarily be considered to be more complex.

Table 4.2 - Land use in the watershed - Sample input data for SLAMM

Source Area	%
Roofs (pitched, directly connected)	14
Roofs (pitched, draining to lawns)	1
Driveways (directly connected)	6
Driveways (draining to lawns)	4
Sidewalks (all directly connected)	20

Street (inter. texture, extensive parking)	20
Street (inter. texture, medium parking)	10
Street (inter. texture, light parking)	5
Small landscaped areas (lawns, clayey soils)	20
TOTAL	100

Complexity usually refers to a structural property of the model (Brooks and Tobias, 1996). Brooks and Tobias (1996) define complexity as a measure of the number of constituent parts and relationships in a model and argue that complexity should differ from level of detail. Gan et al. (1997) ordered models in the terms of their complexities based on the number of parameters involved. James (2000) as a practical matter defines model complexity as the total number of environmental parameters in the input data-file.

The rational method, for example, comprises a simple equation to represent runoff occurring in a limited area as a function of rain intensity, land use and area size (just two environmental parameters). Most currently-used models allow spatial discretization, and inclusion of processes such as infiltration, evaporation and snowmelt, requiring more input data and a greater understanding of individual hydrological processes, and have a higher level of complexity than does the simple rational method.

In modeling dissolved oxygen in streams, for example, the modeler can work with the different levels of complexity, simple and arbitrary qualitative numbers assigned by the authors of WASP (Ambrose et al., 1991) as shown in Table 4.3. At the lowest level of complexity, level 1, just one source of oxygen (reaeration) and one sink (biochemical oxygen demand) are considered in the dissolved oxygen budget. The higher levels of complexity include other factors affecting dissolved oxygen concentrations, such as nitrogenous biochemical oxygen demand, nitrification and eutrophication. The inclusion of each of these factors increases the model complexity. In modeling certain water constituents like nitrate, a minimum level of complexity of 2 is necessary, because dissolved oxygen and BOD must be modeled before nitrate.

Table 4.3 - Complexity levels C in modeling dissolved oxygen due to the number of processes modeled. Source: Ambrose et al. (1991)

C	Explanation
1	"Streeter-Phelps BOD-DO "
2	"Modified Streeter-Phelps" with nitrogenous biochemical oxygen demand
3	Linear DO balance with nitrification
4	Simple eutrophication
5	Intermediate eutrophication

Each process modeled requires a certain amount of input data. Evaporation, for instance, is a function, amongst other factors, of air temperature, relative humidity, solar radiation, and wind speed. The more factors that affect an individual process, the more complex the model becomes.

Brooks and Tobias (1996) believe that overall complexity is a combination of three elements: the number of components, the structure of the connections (which components are related), and the nature of the connections (the complexity of the calculations determining the relationships). These elements are also called size, connectedness, and calculation complexity, respectively. El-Hosseiny (1996) approached the problem of complexity in the context of combined sewer overflow modeling, dividing complexity into two groups: complexity due to process disaggregation and complexity due to spatial discretization.

4.5 Objective measures of model complexity

How does one objectively classify a model as simple, intermediate or complex? Evidently there is a need for a quantitative method of measurement that allows the user to select the model that better suits the project objectives, the required uncertainty, and field data availability. Surprisingly, given the widespread use of mathematical and computer models in science, there is a scarcity of published research on optimal model complexity (Brooks and Tobias, 1996). Ranking models by complexity could help model users select the model that best suits the project objectives.

Complexity is not an inherent attribute of the computer program; rather it is an attribute of the implementation of the program. In other words, complexity depends on the number of subspaces and processes activated by the particular input data file. Of course, the maximum achievable complexity is subject to the limitations of the code in the program.

Some measures of complexity have been proposed. One type is classed as graphical, based on the principle that models can be specified as a number of connected components that can be represented as a graph. However, for any given model, there are likely to be several possible graphs and many alternative measures of complexity. Brooks and Tobias (1996) point out that differences in the complexity of the models may be due to differences in the complexity of the calculations, and so graph theory measures may be inappropriate or perhaps be combined with other measures.

Brooks and Tobias (1996) discuss an alternative approach to graph theory using concepts from information theory, wherein information entropy indicates a measure of complexity, justified by the argument that entropy measures the amount of uncertainty, and that a more complex model is more difficult to understand, and therefore more uncertain. This measure is applicable in very limited cases when it is possible to estimate the probability of each system state.

In computer science many measures (usually termed metrics) of the size and complexity of the code have been proposed, and these can be used to measure the complexity of a model in the form of software. Most metrics are based on counting the occurrence of particular items in the code, such as the number of decision points, or simply the number of lines of code. Many software metrics, however, are partly dependent on the programming language used, whereas measurement of model complexity should be independent of specific model implementation.

The purpose of a complexity measure is to characterize the model so that this information can aid in the determination of the best model. It would be advantageous if there was a single, system-independent definition, a measure of complexity covering all aspects of a model, including level of detail and applicability to all conceptual models. However, it is apparent that no such generally accepted definition or measure exists. As a result, the term complexity itself is a source of confusion due to its usage in many different contexts. To what extent does the addition of modeling processes such as evaporation or snowmelt increase complexity? Is this increment the same for each and every hydrological process? Which hydrological processes most increase complexity? The answer to this type of question presupposes the existence of an agreed measure of complexity. According to Brooks and Tobias (1996) the best approach in measuring complexity would be to identify more specific model attributes such as size, connectedness, and calculation complexity, and to devise measures for these. They recommend that such measures and the type of measurement scale should match our intuitive notion of the nature of the attribute, which is not always the case with software metrics.

The level of detail and level of system process disaggregation (together comprising the complexity) of a model are widely recognized as having an important effect on model performance. However, the particular elements of model complexity which have the greatest effect on each performance element have not been identified.

There is a lack of studies on the balance between simplicity and complexity in model computation of urban drainage structures and the relationship is poorly understood (Brooks and Tobias, 1996). They believe that a universal and unique definition is very unlikely to be found, whereas a general statement for particular types of models or for particular circumstances may be possible. If a single absolute measure of model performance cannot be obtained, perhaps a meaningful comparison of alternative models in similar circumstances is possible.

Various authors have evaluated and compared the performance of hydrological and water quality models of different complexities. In general model performances are measured by fitness criteria. The use of mathematical models to compute runoff from rainfall has been addressed by Loague and

Freeze (1985). In their study, modeling efficiencies for three event-based rainfall-runoff models on three sets of data from small upland catchments were calculated and compared: a regression model, a unit hydrograph model, and a quasi-physically based model. Model performances were assessed for a verification period that was carefully distinguished from the calibration period. The results showed poor forecasting efficiencies for all models on all data sets. The authors do not point to any one modeling approach as being superior to all others for all catchments. However, they suggest that simpler and less data-intensive models in the form of regression and unit hydrograph may provide as good or better computations than a more physically-based model.

John and Flynn (2000) compared the more complex phosphate transport and assimilation in a micro-algae model (PIM - Phosphate Interaction Model) with a conventional, simpler model. The authors argue that the use of a complex model rather than a simple model parameterized on a full data set depends on the time frame of interest.

Complexity is not always a model attribute but an option of the model user in many cases. Many models allow the variation of complexity by adding or subtracting connections. El-Hosseiny (1996), analyzing combined sewer overflow modeling, suggests that a sensitivity analysis of all parameters should be performed before making a decision to remove inactive processes. Known simplification methods include: dropping unimportant parts of the model, replacing part of the model with a random variable, coarsening the range of values taken by a variable, and grouping parts of the model together. Considerable effort may be required to simplify an existing model.

In addition to the factors listed in Table 4.3, in hydrological and water quality modeling the model choice is often based on the availability of data.

> When comparing different complexities of the same program, a simple measure of complexity is the total number of uncertain parameters in the input data file.

4.6 Data availability and model complexity

El-Hosseiny (1996) argues that, as a general rule, the simplest model that achieves the goals of the study should be used. Brooks and Tobias (1996) believe that it is important to match the complexity of the model with the modeling objectives and the available data. Nihoul (1998) questions how refined an ecosystem's description must be in order to be adequate. He suggests selecting a limited number of state variables, which must be known to provide a satisfactory picture of the system, based on the database available.

Professionals and researchers have frequently reported a scarcity of water quantity/quality databases for hydrological and water quality modeling purposes (Siqueira, 1996; Von Sperling, 1995). Elert et al. (1999) mentioned that for biosphere modeling input data are often scarce and frequently data from the literature, from laboratory experiments and from different sites are used. In most cases they have used the limited data available and indicated the uncertainty involved in using the model as a management tool. Beven (1993) affirms that an assessment of the value of data in reducing uncertainty has not received sufficient attention in the hydrological literature. In the application of distributed hydrological models there seems to be never enough data. Beven (1993) suggests that a proposal to evaluate data in terms of reducing model uncertainty would surely provide a compelling case for research funding for field work. John and Flynn (2000) question whether adding complexity to a simple phosphorus model would justify the additional experimental analysis required to parameterize it. Problems arise when model parameters have to be estimated for model application. The higher the complexity of the model, the more input data it requires.

Frequently there is a discussion as to whether a simple or a more complex model is appropriate. Is there an ideal level of model complexity according to the size of the database available for which the uncertainty is acceptable? The reason why some model users opt for using a simple model is because they require less input data. Is this an adequate procedure? The quality of the conclusions arising from using the model depends upon the quality of data, which is a combination of their accuracy, the extent to which they address the modeling objectives and their quantity.

Complex models generally require more input data, which are often not available. On the other hand, the use of modern data acquisition systems have rapidly increased the availability of environmental information in the form of geographical information systems (Masuda and Ichikawa, 1999; Sakai et al., 1999), radar rainfall measurements, and continuous water quality sampling. The watershed inventory for watershed runoff modeling includes information such as physical features and landforms, climate, soil characteristics, stream flow, groundwater, water quality, land use. Simpler models do not offer in their structure the ability to incorporate the advances of modern data acquisition.

4.7 Complexity, reliability and uncertainty

Various authors (Qaisi, 1985; Seo, 1991) based their lake studies on the belief that there is an optimum level of complexity for lowest overall uncertainty. It is often assumed that the reliability of a model will increase with model complexity up to a certain level, and beyond this, reliability decreases with complexity. This form of the complexity relationship has never been demonstrated for surface water models, but was confirmed for lake chemicals

(Qaisi, 1985). However, he gave no scale on the x axis (Figure 4.2) and his complexity was merely qualitative. None of the authors quantified complexity. This relationship can be best understood by comparing Figure 4.3A with Figure 4.3B.

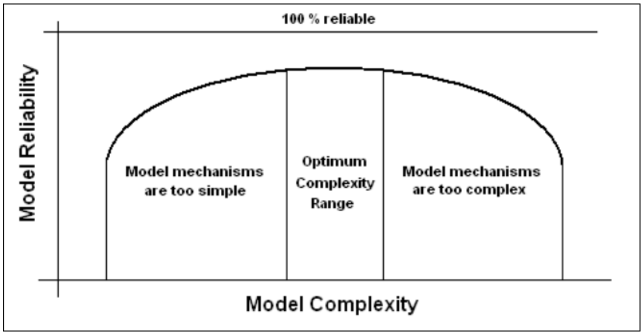


Figure 4.2 - Relationship between model complexity and reliability

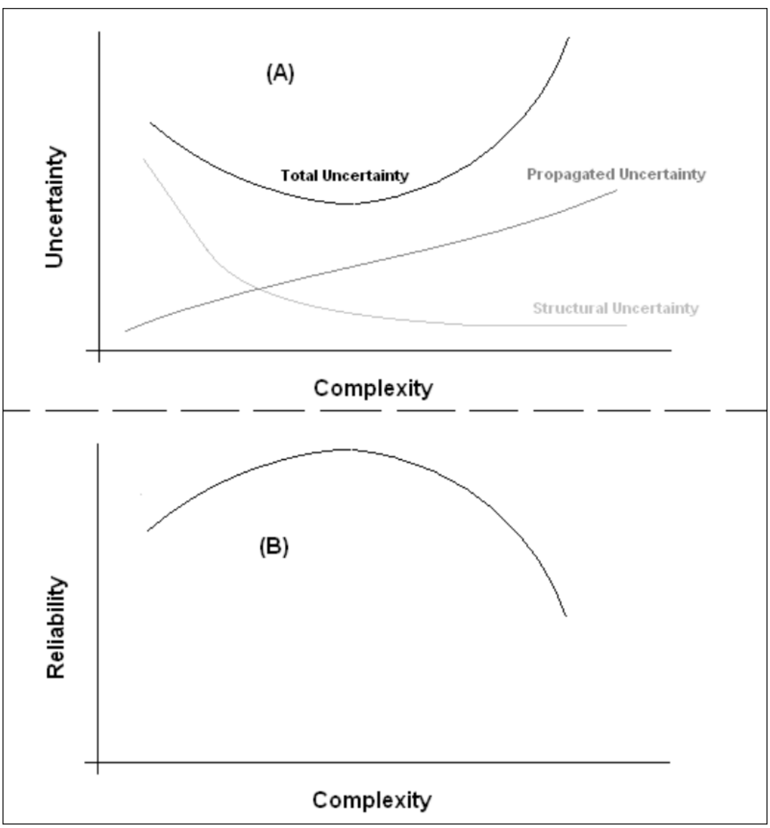


Figure 4.3 – Conceptual relationship between model complexity, reliability and uncertainty.

Qaisi (1985) classified uncertainty into two groups: (a) calculated and (b) structural uncertainty. These two types are illustrated in Figure 4.3A. In order to calculate propagated and structural uncertainty, Elert et al. (1999) evaluated how uncertainty varies with model complexity. A total of seven modeling teams participated in the study using 13 models computing radio nuclide transport in the biosphere. The techniques used for the uncertainty calculations differed substantially between the modeling groups. A comparison of computation between the complex models and the simple models for this scenario showed that the uncertainty in the model computations does not have a simple relationship with complexity, differing slightly from other authors' descriptions. The inclusion of uncertainty in meteorological input data, for example, tended to have a greater effect on the results than the choice of model. This indicates that an important factor affecting uncertainty is the quality and quantity of available input data.

The assessment of uncertainty of a model is preceded by a protocol which includes definition of a design question (objective function) for a specific watershed, estimation of parameters, sensitivity analysis and calibration. The following sections describe how different authors overcame the problems associated with these protocol steps.

Tiscano-Lopez et al. (1995) calculated uncertainty involved in computing peak flow rate, runoff volume and sediment yield. They concluded that errors were largest in the model components with the highest level of aggregation associated with runoff volume, intermediate error with peak runoff and the largest error with sediment yield.

According to Elert et al. (1999) the apparent advantage of using a more complex model, which may give a better representation of reality, may be overshadowed by the introduction of greater data uncertainty. Qaisi (1985) concluded that in a river-modeling situation where the degree of uncertainty in model parameters is large, a model with a lower level of complexity is more appropriate to evaluate system response. On the other hand Cox (1999) has shown that risk models for biological applications with greater intrinsic or structural complexity (meaning complexity that cannot be eliminated by reformulation of the model in terms of its reduced quantities) lead to better-informed, and hence more certain risk estimates. Even poor information (in this case relevant to estimating internal doses) leads to better estimates when information is included than when it is ignored.

4.8 Relationship between complexity and available data

Theoretically, the more information available about a certain system being modeled, the less uncertainty is involved. A longer set of calibration data should achieve a better calibration because model parameters should be more accurately calibrated. Seo (1991) recommended considering data availability as part of the procedure for appropriate model selection. Figure 4.4 is an attempt to illustrate the relationship between data acquisition, data availability, model complexity and uncertainty.

The relationship expressed in Figure 4.2, seems to be valid for a fixed amount of information available. However data are constantly being collected. Figure 4.4 illustrates what might possibly happen to uncertainty in time with data acquisition, based on the literature reviewed above. The gap A in Figure 4.4 shows the remaining uncertainty which could be attributed to, for example, model structural uncertainty.

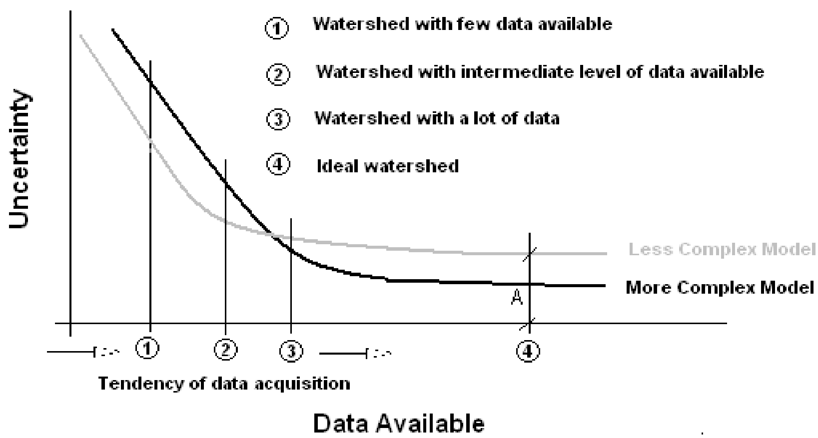


Figure 4.4 - Relationship between data acquisition, data availability, model complexity and uncertainty.

Model results used in calculating uncertainty are a direct function of the data available for modeling. Qaisi (1985) studied two situations in regard to the availability of data: i) where there is a large degree of uncertainty in the model parameter values and ii) where there is a small amount of uncertainty. The analysis included a graph, mapping reliability against complexity similar to Figure 4.2.

Seo (1991) concluded that the optimal order of complexity is almost the same regardless of the range or variation in the observed data. However we may pose the following three questions: 1. When and which additional measurements are advisable to minimize future model uncertainties? 2. Will better data acquisition lead to the use of more complex models? 3. Does a simple hydrological model (which describes a watershed as a single spatial element) describe the real system better than a complex model (when the watershed is discretized into more and various subcatchments)? It seems that the arguments used by some authors for using simpler models are questionable.

It is known that calibrated parameters are data dependent. Elert et al. (1999) suggest that uncertainties in the input data have a greater effect on the hydrological modeling results than the choice of model.

Many hydrological models offer GIS functionality, for example. Uncertainties in GISs are attributed to data and to models. According to Karimi and Hwang (1996) little work has focused on analyzing the uncertainty in specific application models used within GISs. They recommend categorizing models into groups that share common characteristics and then developing generalized techniques for handling these uncertainties. They believe that one reason for this lack of attention is that the application models used in GISs are numerous and from many diverse areas, making the management of uncertainty difficult. In current GISs, models are usually used without analyzing their uncertainties. Consequently, users have no knowledge about the quality of the information they receive, so the accuracy of the decisions made using this information is unknown.

4.9 Concluding remarks

Interesting and important as optimal complexity is, there is still no universal agreement on what the concept really means. Even though we have adopted a very pragmatic definition, one that dovetails with other concepts raised in this book, we failed to derive a method or determining optimal model complexity. Optimal complexity seems to be related to the purpose of the study, the questions and problems being faced, the availability of computer utilities and organized field data, uncertainties of the field parameters, and available study resources (models, time, experienced personnel, and money). Certainly the optimal complexity is dependent on the quality of the user, on the model/user interface, and on the quality of the model information support system. It is clear also that heterogeneous areas generate more complex models.

Optimal complexity is evidently a moving target. It will differ seasonally, and over time as an area develops, and with the interest and capabilities of different users. Over time, user communities grow in number, experience and competence, and web-based methods help speed this along.

We might foresee a need for automatically varying model complexity, even within a single study or run.

In the rest of this book we discuss in more detail many of these issues.

References

Ambrose, R.B. et al. 1991 WASP4 - A hydrodynamic and water quality model - model theory, user's manual, and programmers's guide. Georgia, EPA, 1991.

Andersson, Lotta. 1992. Improvements of runoff models What way to go? Nordic Hydrology, 23:315-332.

Baffaut, C. 1988. Knowledge techniques for the analysis of urban runoff using SWMM. Ph.D. thesis. Purdue U. 321 pp.

Bergstrom, S. 1991. Principles and confidence in hydrological modelling. Nordic Hydrology, 22: 123-136.

Beven, K. 1993. Prophecy, reality and uncertainty in distributed hydrological modeling. Advances in Water Resources, vol.16, p.41-51

Biswas, A.K. 1976 Mathematical modelling and water resources decision making. In: Biswas (Ed), Systems approach to water management. McGraw-Hill NY pp398-414.

Brooks, R.J. and Tobias, A.M. 1996. Choosing the best model: level of detail, complexity, and model performance. Mathematical and Computer Modeling, vol.24, n.4, p.1-14

Cox, L.A. 1999. Internal dose, uncertainty analysis, and complexity of risk models Environment international, v.25, n.6/7, p.841-852

Cunningham, B.A., and Huber, W.C. 1987. Economic and predictive reliability implications of stormwater design methodologies. Pub No 98, U of Florida Water Resources Research Center. Gainesville, FL.

Elert, M. ;Butler, A.; Chen, J.; Dovlete, C. ; Konoplev, A; Golubenkov, A.; Sheppard, M.; Togawa, O.; Zeevaert, T. (1999) Effects of model complexity on uncertainty estimates. Environmental Radioactivity, v.42, p.255-270

El-Hosseiny T. 1996. Uncertainty optimization for combined drainage systems. Guelph, 281p. Thesis (Ph.D.) - School of Engineering, University of Guelph.

Gan, T.Y.; Dlamini, E.M.; Biftu, G.F. (1997) Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. Journal of Hydrology, v.192, p.81-103

James, W. 2000. GIS, data management, model complexity, catchment discretization and process disaggregation (PCSWMMGIS). http://www.eos.uoguelph.ca/webfiles/james/homepage/Teaching/661/wj661M3W00.html

Jewell, T.K., Nunno, T.J., and Adrian, D.D. 1978. Methodology for calibrating stormwater models. Proc. Stormwater and water quality model users group meeting, May 4-5 1978. USEPA 600/9-78-019 pp 125-169.

John, E.H. and Flyn, K.J. 2000. Modeling phosphate transport and assimilation in microalgae; how much complexity is warranted? Ecological Modeling, .125, p.145-157

Karimi, H.A. and Hwang, D. (1996) Toward managing model uncertainty in GISs: an algorithm for uncertainty analysis of air quality advection models. Geomeatica, v.50, n.3, p.251-260

Litchfield, J.B. and Warwick, J.J. 1993. The effect of spatial data aggregation on urban stormwater quality modeling. Preprint for 1993 Runoff quantity and quality modelling conf. ASCE. Reno, Nev. Nov.1993.5pp.

Loague K. M. and Freeze, R.A.. 1985. A comparison of rainfall-runoff modeling techniques on small upland catchments. Water Resources Research, vol.21, n.2, p.229-247

Masuda, T. and Ichikawa, A. 1999 Evaluation of dry-weather flow and pollutant loads in a Japanese river basin. In: Eighth International Conference on Urban Storm Drainage. Sydney, Proceedings. n.4, p.1697-1688

Nash, J.E and Sutcliffe, J.V. 1970. River flow forecasting through conceptual models Part 1 - a discussion of principles. J. of Hydrology, 10:282-290.

Nihoul, J.C. 1976. Applied mathematical modelling in marine science. Applied Math. Modelling Vol 1:3-8.

Qaisi, K.M. 1985. Uncertainty analysis of stream dissolved oxygen models. Ph.D. thesis. U. of Michigan Ann Arbor. 234 pp.

Proctor Redfern and MacLaren, J.F. 1976. Storm water management model study, Vol 1 Final report. Ontario Ministry of the Environment, Toronto.

Sakai, A.; Sumiyama, M.; Genyong, L.; Shengping, Z. 1999. Estimation of nonpoint source runoff form urban area and its control measures. In: Eighth International Conference on Urban Storm Drainage. Sydney, 1999. Proceedings. n.1, p.143-150

Seo, Dong-il. 1991. Optimal complexity analysis of total phosphorus models. Ph.D. thesis. U. of Michigan Ann Arbor. 173 pp.

Siqueira, E.Q. 1996 Aplicação do modelo de qualidade de água (Qual2e) na modelação de oxigênio dissolvido no Rio Meia Ponte (GO). São Carlos, 1996. 90p. Thesis (M.Sc) - Escola de Engenharia de São Carlos, SHS, Universidade de São Paulo.

Siquiera, E.Q. 2001 PhD research proposal. U of Guelph.

Tiscareno-Lopes, M.; Claveran, R.; Weltz, M.A.; Kidwell, M.R. (1995) An approach to representing spatial variability when evaluating model uncertainty in process-based watershed models. In: Proceedings of a Boulder Symposium – IAHS Publications, v.226, p.199-206

Von Sperling, M. 1995 Dificuldades no Cumprimento Integral dos Padrões de Oxigênio Dissolvido em Cursos D'água. Necessidade de uma Abordagem Alternativa. In: 18º Congresso Brasileiro de Engenharia Sanitária e Ambiental, Salvador, Bahia, 17-18 set. 1995. Anais.

Yan, J. 1990. Parameter estimation for multipurpose hydrologic models. Ph.D. thesis. Oklahoma State U. 184 pp.

Bibliography

Flood, R.L.; Carson, E.R. 1993 Dealing with complexity: an introduction to the theory and application of systems science. New York, Plenum, 1993. 280p.

This page intentionlly left blank.

Chapter 5

CONTINUOUS MODELS

When we mean to build,
We first survey the plot, then draw the model;
And when we see the figure of the house,
Then must we rate the cost of the erection.
- William Shakespeare
King Henry IV. Part II. Act i. Sc. 3.

5.1 Summary

Proceeding from a discussion of human “populution”, this chapter describes some dimensions of ethical design for sustainable ecosystems. Arguments are presented that event modeling, and its associated design methodology, at best contributes to the destruction of aquatic ecosystems. The principal (if not the only) argument in favor of design storm methods is design economy, but cheap storm water drainage design is an avoidance of the consideration of inevitable long-term ecological impacts. It is tantamount to a deliberate decision to remain ignorant of the impacts of urban drainage system design. Eco-sensitive design, on the other hand, demands the adoption of continuous modeling.

5.2 Introduction

In the surface water quality modeling litany, models that disregard all dry-weather processes are referred to as event models, while models that also include code for processes that are active during dry weather, such as pollutant build-up, evapo-transpiration, storage depletion, recovery of loss rates, and so on, are termed continuous models. Continuous models also usually include processes associated with winter seasons.

With event modeling the difficulty is that every model run is governed by arbitrarily assumed startup conditions. These assumptions are normally not subject to careful modeling scrutiny, such as sensitivity analysis, calibration, and error analysis. Event models are obviously only run for short durations. To the extent that the effect of these initial conditions persists through the model run, the computed responses may be unreliable. There is a great deal of hydrology literature asserting that these start-up effects are indeed important (James and Robinson, 1986a, provide a review).

Event modeling evolved in bygone times before computing, and it is simply no longer appropriate to adopt such simplistic methodology (James and Shivalingaiah, 1986). In this present chapter, modeling is taken to be continuous, except in so far as short runs for both dry and wet events, and events that are combinations thereof, are recommended for analysis of sensitivity, calibration and error. (In these cases there is no start-up error, because the initial state is given by the observed record.)

Southerland (1982) examined the types of analyses that should be performed with continuous models for non-point-source (NPS) control strategies to meet downstream water quality goals. BMPs simulated included multipurpose detention basins and infiltration trenches for the time span 1980 to 2005 for a 479 sq km catchment in suburban Washington, DC. She used a precursor of HSPF. Her work convincingly demonstrated that assessments of NPS control strategies require the concentration-frequency information that only continuous models can provide, and, more ominously, that disparate source controls may produce more water quality violations than an uncontrolled catchment, due to synergistic concentration releases. Again, only continuous models can provide the necessary information.

The major difficulty associated with continuous models, on the other hand, relates to the copious amounts of input, state, and output information that must be managed. Most continuous models provide code for this purpose, but inexperienced continuous model users can soon run out of computer storage capacity. Although storage capacity for the computed response functions is probably the most stringent constraint, requiring gigabytes for moderate design problems, the literature has focused rather on computer time, as there are limits on the duration of the simulation that inexpensive workstations can handle in the span of a routine day in the design office. In our research, we consider runs that exceed eight hours on an inexpensive desktop microcomputer, not to be cost-effective (Kuch and James, 1993; Kuch, 1994).

Many continuous model studies are being reported, and increasingly so. Chaudhury (1992) built a continuous SWMM model of the City of Providence, Rhode Island, using the RAIN, RUNOFF and EXTRAN blocks to estimate the CSO loadings to the Providence River and Upper Narragansett Bay. At the other end of the scale, Wigmosta (1991) at the University of Washington developed a continuous hydrologic model for a 37-hectare forested catchment, and a 17-hectare urbanized catchment, obtaining excellent results for storm peaks using a 12 month record of 15-minute data. James (1993), citing work by Kuch (1994), [it's not often one gets a chance to write dates like that] argues that 75-year continuous modeling has now become feasible, indeed desirable, in order to address concerns of sustainability. In a landmark case, the Supreme

Court of Canada upheld a decision in favour of downstream riparian owners suffering fluviological impacts resulting from the urbanization of a large city. The arguments that helped to convince the judges were based on continuous modeling, whereas the losing side based their analyses on event hydrology. Event hydrology cannot be used to evaluate fluvial morphology downstream (James and Robinson, 1986).

> Event hydrology is thought to be a cheaper design method. It cannot account for ecosystem concerns. Continuous hydrology allows consideration of aquatic environments.

5.3 "Populution" issues

As a first step in building our argument in support of continuous modeling, we acknowledge the need for eco-sensitive design. In this section, we attempt to build a case against anthropocentrism in engineering design. The case is based on empirical evidence that engineering design very often leads to landscape interventions that increase the human population locally, and that the process seems to have no imposed limits.

Writing for the moment as a drinker of water downstream of major US conurbations, I am perhaps qualified to challenge some premises of modern engineering, viz. that further growth and development is necessarily and unconditionally good. I question whether we humans are not already too numerous in North America. Overall, the population density in the US part of North America is probably about an order of magnitude greater than that of the Canadian part. Significant problems arise when populations of 25 thousand (say) to 25 million people congregate in urban areas (Mexico City's population will soon be comparable to that of Canada). Modelers must, at some point, probably sooner rather than later, ask a leading question: are we doing enough in North America about containing the numbers of humans and domestic animals that directly cause a deterioration in the natural environment? When will there be development enough? Is a population of 7 million sufficient in SE Michigan, or Southern Ontario? Why not 70 or 700 million? Or is 700 thousand more reasonable? Why not 70 thousand for that matter? These questions may be too big for engineers to answer, but we are involved because we write the computer models that relate pollution to population activities. Or models lead to development controls. The basis for the answer to these questions of carrying-capacity vs. sustainability seems to invoke our so-called sustainable quality-of-life. I have not met anyone who feels that surface water quality, and her/his quality of life, will be better if the local region developed from housing and employing, say, 7 thousand, to, say, 7 million. In fact the opposite seems clearly to be the human experience. This truth is evidently not universally acknowledged by the councils of our elected representatives. On the contrary, they make indirect use of the results of our water-environment models to encourage further development, and thus population growth.

In this particular respect our models are not functioning as they should; since our models do not explicitly examine carrying capacity, they do not test all reasonable alternative management strategies. It is the large numbers of us people, especially the very large numbers in our North American cities, that destroy natural habitats. Pet and farm animal populations are also causing serious problems at several locations in our part of the globe. Processes now incorporated in our storm water and water quality models do not make the connection (between population carrying capacity or land-use controls on the one hand, and aquatic ecosystem changes or pollution on the other) very obvious. Modelers need to be able to convince decision-makers that negative growth is a valid strategy that ought to be routinely examined.

> Negative growth is a BMP that warrants attention.

5.4 Coming to terms with sustainability issues

In this section we attempt to develop the argument that continuous modeling is ecosystem-friendly, and that certain concerns can be met, especially if the simulation period is very long, say 75 years or so.

I venture to suggest that the biggest water environment problem in North America is loss of natural aquatic habitat. In my judgment this is determined by three main factors:

1. Modulus of the rate of change (and magnitude) of the mean discharge, $\left|\frac{\partial \overline{Q}}{\partial t}\right| > 0$ and becomes relatively very high as a result of urbanization. During flood events, the rapid increase in stage washes away ecosystems and changes the fluvial morphology. Bank-full flows are increased by perhaps two orders of magnitude over three generations. And during dry events, it dries out channel beds and riparian areas, increasing pollutant concentrations, soon displacing the original cold water ecosystems. Flow variations kill relentlessly. They are sensitive to most anthropogenic activities.
2. Sediment and sediment-attached pollutants, and their impacts on the substrate, the habitat, and aquatic organisms. Sediment loads

are also sensitive to increased anthropogenic activities, including, of course, flow variations.

3. Thermal enrichment is a factor which is seldom considered. It is a result of tracts of land being cleared of indigenous vegetative canopy, and covered by black asphalt. Our urban areas act as solar receptors with excellent water conveyance, a deadly combination, because water is an excellent conductor of heat. Most population centers with say 50,000 people or more, when located on small creeks, have so severely degraded the aquatic thermal environment that riparian habitat immediately downstream becomes totally changed, e.g. from cold water fisheries to a coarse warm water system. Of course, this effect depends on location, the relative size of the creek catchment and the area developed. The concern is fundamentally important in Canada.

Such sustainability and eco-restoration factors are first and foremost computable by continuous modeling, or period-of-record modeling.

Lack of progress in continuous modeling methodology has been truly amazing. There is still copious event modeling. Some modelers argue that event modeling still makes sense, being easier than continuous modeling, or cheaper. It can be demonstrated that none of these assertions is true anymore. Norm Crawford published his dissertation on the Stanford Watershed Model in 1962, which is more than a generation ago, 40 years is a long time given modern science and technology. In this time, continuous modeling has again and again proven effective, reliable and cost-effective.

Originally proposed in the late 19th century, and developed in the first half of the 20th century, long before computers, the need for event modeling has become obsolete. The term implies that the (only) event is a wet event; everything else is an non-event, nothing happens between rainfalls. Yet many processes that are active during dry periods are extremely important: recovery of infiltration capacity, recovery of storage, build-up of pollutants, to name a few. Some models have been developed recently that set up a series of alternating wet and dry events, as a kind of compromise with tradition. But this also leads to complications: how can the fuzzy periods between wet and dry events be accommodated by such hydrology? For significant periods of natural time there is very light precipitation and/or saturated humidity levels; during these periods it is not clear whether pollutants are building-up or washing-off. Process disaggregation into an artificial wet/dry dichotomy is unnecessary today, given modern computing.

Of course, continuous modeling is encumbered with difficult data management. Fortunately PCSWMMGIS interfaces with GIS. Aslo, time series management systems such as HEC-DSS and ANNIE, are available in the public domain.

Resisting the pursuit of the simplistic is not a trivial challenge. Notwithstanding the need for model parsimony, my position is that simplistic representation is not somehow a worthy end in itself. Subsuming complexity into a simple fudge-factor is sometimes anti-intellectual, if the factor has no physical explanation, and is often difficult to explain in legal or confrontational situations. The argument that simplistic models are easier to understand or apply is not true. The notion, that the most simple representation possible is always the most desirable was spawned from BC (before computing) era. We ought not to support it these days. Only if all potentially relevant processes are represented in our models, can modeling can help us determine which processes are important.

Three generation modeling (3GM) with SWMM is readily accomplished, although we still need quality code for a six-minutely rainfall generator. It is not unusual in North America to find rain records of 50 years and longer, although the available time step is rather coarse. It seems likely that 3GM will make a difference to the way we do business, and thus to our environments and ecosystems.

5.5 Some eco-ethical concerns

In this section we proceed to some new dimensions of eco-sensitive design, not yet associated with continuous modeling, but impossible to reconcile with event methodology.

Apart from human population control, or reduction, there can be no activity more urgent to the future of the world than to reverse the loss of natural habitat (Kennedy, 1993). Habitat loss is the result of anthropogenic activity in watersheds, such as urbanization. Lazaro (1990) describes the physiology, anatomy and the relentless morphology of cities, with their consequent degradation of aquatic environments. By urbanization, we mean the concentration of people into urban settlements, and the change in land-use first from indigenous, original forest or prairie, to rural or agricultural, then to urban commercial and industrial. All land-use changes affect the hydrology of an area, but urbanization is by far the most forceful (Leopold, 1968). There is no doubt that the contemporary population intensification into urban areas will continue through the next few generations, and that the associated hydrological problems of aquatic habitat destruction will become increasingly more acute.

There also can be no doubt that conventional design-storm, or event-hydrology methodology, focuses on extreme simplification, precisely to reduce the cost of the design phase of landscape interventions. Simplistic methods have been strongly supported by some pro-development forces. Land developers, construction industries, government departments of agriculture, and consulting engineers believe that they will profit from quick, inexpensive design studies. They advocate extremely fast, short-cut design, design that is not concerned with, but may even denigrate, the study of long-term ecosystem impacts. Design costs, whether cheap or expensive, are eventually passed on to homeowners, of course. Recent consumer surveys, however, suggest that homeowners in North America are prepared to share the higher cost of eco-sensitive design.

Long-term continuous water quality modeling on the other hand leads naturally to consideration of impacts on aquatic ecosystems. Necessary information is brought into focus in the foreground. As an example, salmonid (cold water fish) habitat requires seasonal and growth-phase upper and lower limits on flow depth and flow velocity, a continuous low upper limit on turbidity (for sight feeders), a mean summer low upper limit on mean water temperature, very small deposition of suspended material in spawning areas and seasons, and all this in addition to the established limits on transported chemicals. Event hydrology does not provide such information, whereas output from continuous SWMM, for example, begs to be fed into the In-stream Flow Incremental Method (IFIM) for evaluating fish habitat (Navarro et al., 1994).

Features of an unpolluted river are: (a) ecosystem diversity, (b) flow and water quality stability, and (c) self-purification. In a few minutes it is possible to collect two dozen visible plant and animal species, and 100 microorganisms. Proponents of ecological sustainability regard nature in this complexity not just as a set of limiting concentrations, but as a better model for the design of housing, townships, neighbourhoods and regional economies. Sustainability depends upon replicating the structure and function of natural systems with their far-reaching inter-connectedness. Orr (1992) suggests a number of concerns for design that is sensitive to sustainable eco-systems:

- the living world should be the matrix for all design,
- design should follow the laws of life,
- biological equity must determine design,
- design must reflect bioregionality,
- projects should use renewable energy systems,
- design should integrate living systems,
- projects should heal the planet, and
- design should follow a sacred ecology

Advocates of cheap design will raise objections to the above position. And while they would agree that traditional design has led to numerous minor eco-disasters, such as the loss of cold-water fisheries locally, it is true that these have been more the result of ignorance than of deliberate eco-vandalism. Conventional, narrow design does not care about downstream ecosystems; the case was neatly posited by Field and Lager almost two decades ago (1975): …simply stated, the problem is as follows. *When a city takes a bath, what do you do with the dirty water?* The question suggests that there is no alternative but to degrade aquatic and riparian ecosystems downstream of urban creeks.

Serious questions should be asked not only about the quality of the designs proposed, but also about the quality of the design study itself. By now we all know that simple engineering solutions have very often led to ecosystem nightmares. Wedell Berry (cited by Orr, 1992) states that a bad solution is bad because it acts destructively upon the larger patterns [of nature] in which it is contained. It acts destructively upon those patterns, most likely, because it is formed in ignorance or disregard of them. On the other hand, a solution is good if it is in harmony with those larger patterns. As I understand it, and further paraphrasing Berry, this means that a good design will have to meet a number of requirements more appropriate to a post-modern society. Among these a number may be mentioned, even though few of them, if activated at all, could be attributed directly to continuous modeling philosophy. Among other requirements, a good design will:

- accept the limits of the discipline of engineering;
- improve and restore the natural balances and bio-diversity;
- correct the human behaviour that caused the problem to the ecosystem;
- imitate the structure of the natural, native or indigenous system;
- be good for all parts of the natural system;
- not enrich one individual or group to the distress or impoverishment of another; and
- be in harmony with good character, cultural value, and moral law.

(Not all points have been listed; only the first five would probably be implicated by continuous modeling, assuming that the continuous output would be further processed by aquatic ecologists; the last two distinctly relate to post-modern society, and are included from the larger list out of interest; the ranking is by the present writer.)

The point is that continuous modeling makes it difficult to avoid ecosystem concerns, while the use of event hydrology makes it difficult to consider them. This point has also been developed by Abbott (1993). Opting for event hydrology, then, given the present computational environment, is akin to a deliberate decision to choose ignorance all-the-way-down-the-road, to invite eco-disaster. In this sense, perhaps we may say that it has become simple-minded.

Modelers should follow a sacred ecology.

5.6 Processes relevant to ecosystems

In this section, we mention by way of example, four sets of processes or procedures that are not commonly found in popular event models used in design applications for managing the impacts of urban stormwater. All four rank at the top priority for future enhancements, in the writer's judgment, and the discussion helps develop an argument for new code that may be added to existing codes, perhaps through suitable shells.

5.6.1 Ecosystem restoration

In my judgment, the most important BMPs (so-called) have not yet been coded into the widely-distributed WQMs, and hence these BMPs appear infrequently in water pollution control strategies. The most important BMPs are those which redirect us back to the *as-was* ecosystems (which existed before so-called "white-man's" agriculture or urbanization), BMPs which by their design objectives seek to direct runoff back into the ground; to stop or remove pollutants at the source; and to add feed-stock for aquatic ecosystems. Deciduous canopy is an important BMP in urban areas, to cool heated cities. Similarly, infiltration BMPs, particularly those with sand filters, would help reduce the loss and destruction of cold water fish habitats. We have not yet developed design methods for BMPs that add the "discomforts" that are essential for aquatic life, such as aquatic larval stages of blood-sucking flies.

5.6.2 Thermal enrichment

Temperature modeling is very inadequate in our existing models. Not even HSPF can model heat accumulation in blacktop paving and roof tiles, nor do any urban runoff models cover the wash-off of thermal energy from paving and roofs. Temperature is the number one determinant of ecosystem types downstream of urbanization. Available models do not deal with solar thermal enrichment, yet all aquatic chemical and biochemical processes are temperature-dependent. Areas such as Detroit are effectively enormous, exposed, black-body solar receptors, and rainfall/runoff is an efficient transporter of heat from hot pavement and roofs. In this discussion, thermal enrichment is taken to mean the increase in thermal energy, carried from the existing or proposed development to receivers, over what it would have been had there been no significant anthropogenic activity. Activity includes agriculture and clearance of deciduous canopy. Enrichment should not merely denote potency in terms of existing conditions - sediment constituents, for example, because the sediments have been seriously degraded since the time of the original mixed forest canopy.

5.6.3 Decision support systems

What are the requirements of decision support systems (DSS) that favor continuous modeling? DSS such as PCSWMM should provide:

- spatial data and facilities management systems,
- time series management systems,
- communications to remote, integrated, distributed databases,

- long-term six-minutely rain generators,
- analysis of the number and duration of exceedances and deficits in the output time series,
- presentation tools for very long time series analysis,
- data compression for long duration time series (e.g. 75 years of 6-minute data for hundreds of stations and dozens of chemicals),
- integration with downstream, dependent models such as fluviology,
- sensitivity analysis,
- calibration,
- error analysis,
- tools to display model reliability and optimal complexity, and
- GUI WIMP presentation tools that display the inherent fuzziness of the computed output.

There seems to be no doubt then that our deterministic surface water quality models could benefit from more statistical tools. For design purposes, very-long-term input functions may be synthesized from shorter records using stochastic rainfall generators, for example. Considering the inherent variability of rainfall, more statistical manipulations are necessary to build 75 years of data, and to present the results of 75 years of flow and many pollutants at many points. End-users of 3G modeling could not comprehend such results without a fair amount of statistical manipulation. The melding of statistics with deterministic models will have another advantage: it will facilitate the removal of the artificial distinction between data gatherers (field personnel) on the one hand and data consumers (modelers) on the other (or between monitoring people and analytical laboratory people on the one hand, and the modelers on the other). Provided such models incorporate suitable sensitivity analyses, very long term complex deterministic models with comprehensive statistical tools provide useful management tools for data collection programs. They are the best means for filling in missing data. Ranking the sensitive parameters helps rank priority for selecting chemicals, sampling frequencies and locations, and accuracies of determination.

5.7 Concluding discussion

Principal among the purposes of the modeling effort, is to design an optimum array of BMPs - landscape interventions that divert, store or treat urban stormwater runoff to remove pollutants, reduce flooding or provide other amenities. The recommended way to design them is by using continuous 3GM. In any case, adoption of a long-term time-series should not be the insurmountable challenge that it seems to have been: because of:

- the modern wide distribution of inexpensive computers;
- freely available knowledge and information on continuous modeling;

- the urgent need to develop ecosystem sensitive-methods; and
- the informed engineering community, itself the product of excellent higher educational institutions, and an informed society;

there can no longer be any case for event-hydrology methods.

The old argument that continuous modeling is not computable is no longer true: For continuous, fuzzy modeling, appropriate decision support systems are recommended: PCSWMM is such a decision support system, involving:

- error analysis - the computation of the likely error that a computed response may incur;
- dissaggregation - the degree to which the physical components of a system are modeled by increasing the number of defined processes;
- discretization - the number of spatial components selected to represent the physical system that has been dissaggregated into processes, and the degree to which the physical parameters are averaged (lumped) spatially and temporally;
- model complexity - a measure of the number of parameters in the model. Models should be neither too complex nor too simple for the problem and problem-solving environments. By environment we mean the space with its objects that surrounds a thing that is considered to be more important.

Finally, in keeping with our philosophy of continuous modeling for ecosystem concerns, we have to redefine potency factors and enrichment, so that we can relate our results back to pre-forest-clearance conditions, since impacts of the change from indigenous forest to agriculture were often nearly as bad as the change from agriculture to urbanization.

> Remember to never use event hydrology for design.

References

Chaudhury, R.R. 1993. Stormwater modelling of a combined sewer system. Ph.D. thesis. U. of Rhode Island. 707 pp.

Field, R. and Lager, J.A. 1975. Urban runoff pollution control - state of the art. J.Environmental Eng'rg, ASCE EE1 pp.107-125.

James, W. 1993. Need for 3-generation stormwater quality impacts models. USEPA Invited Conference on Model-building for Great Lakes Water Quality problems of the 2000's. Heidelberg College, Tiffin, OH. Aug.4-6. copy available from author.

James, W. and Robinson, M.A. 1986a. Continuous Deterministic Urban Runoff Modeling. Int. Symp. in Urban Drainage Modeling, Dubrovnik, Yugoslavia, published by Pergamon Press, London, pp.347-378, Apr. 1986.

James, W. and Robinson, M.A. 1986b. Continuous Modeling of Downstream River Erosion caused by 40 Years of Urbanization. Water Forum '86, ASCE, Long Beach, pp.1710-1717, Aug.4-6, 1986.

James, W. and Shivalingaiah, Boregowda. 1986. Continuous Mass Balance of Pollutant Build-up Processes. Proc. NATO Advanced Research Workshop on Urban Runoff Pollution, Montpellier, France. Aug.25-30, 1985, published by Springer-Verlag in Urban Runoff Pollution (Series G. Vol.10), pp.243-271,1986.

Kennedy, P. 1993 Preparing for the 21st century. Random Hse, ISBN 0-394-58443-0, 445pp.

Kuch, A. J. 1994. A sensitivity analysis shell for complex continuous water quality management modelling. M.Sc. Thesis. Univ of Guelph. ca.120 pp. January 1994.

Kuch, A.J. and James, W. 1993. A sensitivity analysis tool for three generation continuous urban storm water quality management modeling. Runoff Quantity and Quality Model Group Conf, Sponsored by ASCE. Reno NV, Nov.8-9, 1993, in press.

Lazaro, T.R. 1990. Urban hydrology - rev ed. Technomic. ISBN 87762-547-6. 250pp.

Leopold, L.B. 1968. Hydrology for urban land planning - a guidebook on the hydrologic effects of land-use. USGS. Circ. No. 554. 18pp.

Navarro, J.E., McCauley, D.J. and Blystra, A.R. 1994. Instream flow incremental methodology (IFIM) for modeling fish habitat. In: Current practices in modeling the management of stormwater impacts. Lewis Pub. ISBN 1-56670-052-3 pp.1-22.

Orr, D.W. 1992. Ecological literacy. State U. of NY. ISBN 0-7914-0874-4 224pp.

Southerland, E. 1982. A continuous simulation modelling approach to nonpoint pollution management. Ph.D. thesis. Virginia Polytechnic Institute and State U. 165 pp.

Wigmosta, M.S. 1991. Modelling and monitoring to predict spatial and temporal characteristics in small catchments. Ph.D. thesis. U of Washington. 202 pp.

Bibliography

Abbot, M.B. 1991. Hydroinformatics: information technology and the aquatic environment. Avebury Technical. ISBN 1-85628-832-3. 157pp.

Donigian, A.S. and Linsley, R.K. 1979. Continuous simulation for water quality planning. Water Resources Bulletin, 15(1): 1-16

James, W., in press. Channel and Habitat Change Downstream of Urbanisation. Eng'rg Fndn Conf. on Urban Runoff and Receiving Streams. Crested Butte, Colorado. Aug. 4-9, 1991.

Klemes, V. 1987. Empirical and causal models in hydrologic reliability analysis. Procs NATO Institute on Engineering reliability and risk in water resources. Tucson, Ariz. May 19-Jun 2, 1985. ISBN 90-247-3492-4. pp.391-403.

Robinson, M. and James, W., 1984. “Continuous Variable Resolution Stormwater Modelling on a Microcomputer”, Proc. of the Stormwater and Water Quality Modelling Conf., Sponsored by USEPA and Ontario Ministry of Environment, Burlington, Ontario, published by CHI Hamilton, pp. 59-74, Sept 6-7, 1984.

Thompson, D.B. 1989. Determining parameters for a continuous simulation model by estimation, synthetic calibration and analytic calibration. Ph.D. thesis. U. of Missouri-Rolla. 261 pp.

Young, P.C. 1981., Parameter estimation for continuous-time models - A survey, *Automatica, 17*(1):23-39, 1981.

Chapter 6

INTRODUCTION TO GENERATION OF LONG-TERM RAIN INPUT

For many years I was self-appointed inspector of snowstorms and rainstorms, and did my duty faithfully, though I never received one cent for it.
- Henry David Thoreau.

In the Spring I have counted one hundred and thirty-six kinds of weather inside of twenty-four hours.
– -Mark Twain.

6.1 Introduction

This chapter is a preliminary introduction to the topic, and has been abstracted from Chan (1998). PCSWMM does not include a rain generator, but we plan to include one in the next version. This chapter will be completed when more details are available.

Distributions of rainfall intensities or depths over time and space are important for modeling urban water systems. Precipitation data forms the input function for most hydrologic models, and is typically derived from:

1. synthetic design storms; or
2. historical rainfall records; or
3. synthetic rainfall data generated by a special mathematical model.

The use of synthetic design storms, such as those derived from intensity-duration-frequency (I-D-F) averaged intensities and artificial storm distributions, is still prevalent in design offices. As stated earlier, such methods simplify the design process and minimize costs but at the expense of the environment (Ormsbee, 1989). Synthetic design storms generate single rainfall events for some averaged intensity and duration based upon a statistical analysis of observed storms. The drawbacks to the single-event design storm approach are that dry-weather processes are not modeled and that antecedent conditions must be assumed at start-up (James, 1994a; Ormsbee, 1989; and Robinson and James, 1981).

An alternative to using synthetic design storms and event hydrology is to use observed or generated rainfall data for continuous modeling. Continuous modeling using long-term rainfall data is becoming increasingly necessary to infer the evolution of water quality and of the responses of the aquatic environment to anthropogenic inputs (James and Robinson, 1982; and Yevjevich, 1991).

Long-term rainfall data, generally in the order of 70 to 100 y, is available from most Canadian urban centers. However, the temporal resolution of this data is generally of a daily time-step and much too coarse for any small watershed hydrology. In order to properly gage the response of a watershed, the time-step of the rainfall data should be much less than the time of concentration of that watershed (Ormsbee, 1989). Ten to thirty years of hourly data is also available but is still too coarse for hydrologic modeling. Data of even smaller time-steps is available (15 minutes and tipping bucket time-between-tips) but the typical duration of record is often 10-15 y. A desirable time-step for hydrologic modeling would be small enough to model the catchment response properly and yet not be too small to require enormous data storage requirements. The desired record length for rainfall data is in the range of 75 to 100 y. This is in order to properly compare design alternatives over the expected lifespan of the structure; account for probable variations in rainfall intermittency and scale; and because the living memory of catastrophic events and of environmental change is passed down through three generations of the population (James, 1994b). 75 y of 15-minute (or smaller) historic rainfall data will not become available for another 60y. Rainfall models (generators) have been developed in an attempt to remedy the paucity of useable rainfall data.

There is a plethora of rainfall models and more are continually being developed. Stochastic models have been the focus of rainfall modeling research for the last few decades. Stochastic models take statistical properties of historical rainfall data and probabilities to generate long-term rainfall data records. Deterministic or physically-based rainfall models (also called dynamic models) attempt to predict rainfall using physical laws incorporating meteorological information as input and following the principles of conservation of mass, momentum, and energy (Foufoula-Georgiou and Krajewski, 1995). Because of the sophistication and complexity of atmospheric physics and the breadth of associated parameters, physically based rainfall models have been limited to mesoscale and general circulation models (GCM). Stochastic and deterministic rainfall models can be purely in the temporal domain, purely spatial, or a combination of both (where rainfall is simulated in both time and space). Disaggregation models take historical data of large time steps and increase the temporal resolution to obtain finer time steps while attempting to maintain the same statistical properties of the original data. Most disaggregation models that have been developed generate hourly rainfall data from daily rainfall data (Hershenhorn and Woolhiser, 1987) or generate data with smaller time steps (1-30 minute) data from hourly data (Ormsbee, 1989).

There is, however, currently no model that can directly disaggregate daily rainfall into 1-15 minute rainfall. Also, available disaggregation models cannot extend the length of the rainfall data, they only artificially enhance the time resolution of the data.

Most stochastic models attempt to model occurrence and duration of both wet and dry events (intermittency) coupled with another algorithm to determine precipitation amounts for the wet events. These increasingly complex models do not, however, provide any explanations relating to processes that drive the rainfall phenomenon (Yevjevich, 1991). Rainfall is a complex atmospheric process that is affected by incident solar radiation, sunspot activity, lunar orbits, the El Niño-Southern Oscillation (ENSO) phenomenon, wind circulation patterns, etc. Models based on probabilistic outcomes and parameter fitting without any correlation to physical processes are insufficient and dishonest (Klemeš, 1986).

Within the stochastic modeling approach, there is interest in developing a general theory to characterize the internal structure of rainfall. By examining the internal structure of rainfall, hydrologists attempt to describe the link between the observed rainfall and the mechanisms responsible for its occurrence. Estimating the properties of the underlying processes that lead to an observed series is termed TS analysis (Bras and Rodriguez-Iturbe, 1985). TS analysis models of rainfall attempt to analyze trends, periodicities, and frequencies in order to describe, explain, and simulate rainfall rates. TS analysis is also being used to correlate climate patterns such as between rainfall and other physical phenomena.

Wavelet analysis is a new tool that can be used for time-analysis of multi-frequency, multi-scale signals that may be non-stationary or non-sinusoidal in structure (Bradshaw & McIntosh, 1994). Wavelets and wavelet transforms offer many advantages over traditional Fourier transforms, which are inadequate for analyzing non-stationary, variable, and intermittent data such as rainfall.

6.2 Review of rainfall modeling and analysis techniques

There are many rainfall model and analysis tools, each utilizing different temporal and spatial scales (Berndtsson and Niemczynowicz, 1988). For the sake of brevity, only a few model types from each approach are briefly mentioned here.

The common origins of all rainfall models can be traced back to the late 19th century where engineers required design flows for hydraulic structures such as sewers, open channels, and reservoir systems (Todini, 1988). The main focus for urban hydrologists has been to determine the return period, or return frequency, of extreme hydrologic events (Niemczynowicz, 1994). I-D-F curves were developed as a packaged representation of recorded rainfall data for a geographic region. These I-D-F curves provided average rainfall intensities as a function of storm duration and required design frequency. Synthetic storm patterns (Keifer and Chu, 1957) were developed to convert these average rainfall intensities into representative rainfall hyetographs. The result is that statistical information inherent in the observed rainfall record is lost when the I-D-F curves are generated and generic storm patterns are produced. I-D-F curves, synthetic storm patterns, and the accompanying rational method runoff computation method were developed at a time when computers in design offices were uncommon and simplicity ruled. However, these design storms only represent a portion of the total rainfall volume of observed rainfall and when coupled with a storm pattern such as the Chicago design storm, give an unrealistically high peak rainfall intensity (Arnell et al., 1984).

A more accurate alternative to using I-D-F curves is to use the rainfall data that the I-D-F curves were generated from. As one long-time proponent of continuous modeling says, "instead of using the 1-in-50-y storm, use one 50-y storm" (James, 1994a). The use of historical data lends more credibility in the eyes of the general public. However, existing hydrologic records are not sufficiently long to provide many important statistics such as the 1-in-100 y return period.

This prompted the development of synthetic rainfall models to provide rainfall records of long duration so that they may be used to determine properties such as the 100-y return period with a certain degree of certainty (Fiering and Jackson, 1971). Early work in rainfall modeling comprised attempts to generate long-term daily rainfall data that could be converted to stream flow data in an effort to aid flood frequency estimation (Franz, 1970; Fiering and Jackson, 1971; Ott, 1971; and Wyrick, 1974). Stochastic models have been the primary approach in rainfall modeling for small watersheds. Deterministic or physically based precipitation modeling is more suited to the realm of atmospheric circulation models and to mesoscale hydrology due to its complexity and parameter requirements.

> Instead of using the 1-in-50-y storm, use one 50-y storm.

6.3 Stochastic models

Stochastic models to generate long-term monthly, daily, and hourly point rainfall data have been available since the 1960s. Most stochastic rainfall processes are concerned with the concept that hydrologic series are random processes that follow probability distributions. Markov type models have been popular with many stochastic rainfall modelers in the past due to its ease of application and non-parametric nature but suffer from the inability to reproduce persistence and intermittency of the rainfall process (Lall et al., 1996). Markov chain models simulate the occurrence of wet and dry periods and rainfall amounts within the wet periods. Pattison (1965) developed an hourly rainfall model using a first order Markov chain for wet periods to determine the state (wet or dry) of the current hour based on the state of the previous hour. The model switched to a sixth order Markov chain model during dry periods to determine the state of the current hour based upon the six preceding hours. Seasonal variability of the rainfall characteristic was incorporated into these Markov chain models by developing parameters that varied on a monthly basis (Franz, 1970). Pattison's model tended to over-estimate the average inter-event duration, which were generally longer than normally found in nature. Kline and McFarland (1988) used a two-state second order and two-state third order Markov chain model to generate occurrence of daily precipitation events depending on some threshold rainfall value. In determining the seasonal variation in the parameters, they used a Fourier series to describe the periodic fluctuations. Tan and Sia (1997) investigated a variation on rainfall modeling using a lag-1 Markov chain model to generate an event-based data series similar to time-between-tips recorded from tipping-bucket rain gage data. This method eliminates the need to generate both wet and dry events as well as the depth of rainfall in the wet event. Only time values representing time between tips (TBTs) need to be generated as each tip represents a minimum depth increment, usually 0.2mm.

The Neyman-Scott (N-S) and Bartlett-Lewis (B-L) models are clustered point process models that have been applied to rainfall modeling. Point processes are a series of events that occur randomly in time (Chatfield, 1996). A point process rainfall model generates storm origins from a Poisson process. These storms have a single rain cell representing a rectangular pulse of rainfall intensities that is randomly generated and independent of other rain cells. The intensities and duration of the pulse is randomly distributed. Storm cells overlap one another such that the rainfall intensity at any time is the sum of all pulse intensities (Cowpertwait, 1991). Clustered point processes were developed to overcome the inability of point process models to describe the statistical structure of rainfall over different scales (Foufoula-Georgiou and Krajewski, 1995). Cowpertwait (1991) investigated the N-S model performance for use in simulating temporal rainfall and found that the model performed well except for the month of June, which did not have good daily and hourly autocorrelations.

Franz (1970) used a multivariate normal generator to synthesize long-term hourly rainfall with four different rainfall seasons for multiple gages. Ott (1971) expanded on Franz's stochastic model by using five rainfall seasons and cumulative distribution functions of storm length, total rainfall depth, and inter-storm length to generate synthetic stream flows. Gumbel Extreme-Value distribution, log-Gumbel, Normal, lognormal, log-Pearson Type III, and Beard's method were used in a frequency analysis to compare peak stream flows.

Srikanthan and McMahon (1983a, 1983b, and 1985) developed annual, monthly, daily, hourly, and six-minute stochastic rainfall and evaporation models primarily for water-balance/crop-growth models. The hourly rainfall model used a transition probability matrix (TPM), one for each month, to determine the daily state (wet or dry), and a second order Markov chain to determine the probability of rainfall occurrence throughout the day. Hourly TPMs were then used to generate hourly rainfall depths. The six-minute rainfall model used a daily TPM distinct for each month, having seven different states. Within each day, an hourly model consisting of a two-state second order non-stationary Markov chain with probabilities that vary with the time of day to determine if the hour is wet or dry. If the hour is wet, one of four types of wet hour is chosen using an hourly TPM corresponding to the month. Rainfall depths were then generated for six-minute intervals during the wet hour. The number of parameters for this approach has been purported to be in the range of 5000 to 6000 (Hershenhorn and Woolhiser, 1987).

6.4 Disaggregation models

Another method of generating rainfall data is to take an existing rainfall record of a certain duration and coarse time-step resolution and increase the time-increment resolution. Disaggregation models accomplish this by taking historical or simulated rainfall data of large time-steps and breaking down the series into a sequence of values with shorter time-steps by approximations while keeping the statistics of the disaggregated and the observed TS consistent. Disaggregation is useful for long-term continuous modeling where the rainfall record length may be of sufficient duration but the temporal resolution is too large and therefore inappropriate for small watershed modeling applications. Hershenhorn and Woolhiser (1987) proposed a daily disaggregation model that simulated individual storms by determining the number of rainfall events per day, the amount, duration, and when the storms occurred given the daily rainfall amounts for the preceding, current, and following days. However, the procedure is described as parameter-intensive (Hipel, 1985), complicated, and not appropriate for common engineering problems (Koutsoyiannis, 1994). Koutsoyiannis (Koutsoyiannis and Xanthopoulos, 1990; and Koutsoyiannis, 1994) proposed a disaggregation method that assumed a Markovian structure with a gamma type distribution function. However, the level of disaggregation using this method was limited to transforming monthly rainfall data to hourly data. Ormsbee (1989) proposed a method of disaggregation where hourly rainfall is disaggregated into data of sub-hourly intervals (1-30 minutes). Ormsbee developed a stochastic and a deterministic approach for both discrete and continuous models and found that the continuous deterministic approach to be less computationally intensive. Ormsbee's continuous deterministic disaggregation model cycles through the observed record comparing consecutive three-hour rainfall observations (*t-1, t,* and *t+1*) computing variable probability density functions (PDFs) based upon the rainfall intensity sequences about the central hour. The PDF is composed of two line segments that have equal angles to the horizontal. The drawback is that it is only applied to hourly rainfall and there is often a limited duration of record.

References

Arnell, V., P. Harremoës, M. Jensen, N. B. Johansen, and J. Niemczynowicz. 1984. Review of Rainfall Data Application for Design and Analysis. *Water Science and Technology*. Vol. 16. P. 1-45.

Berndtsson, R., and J. Niemczynowicz. 1988. Spatial and temporal scales in rainfall analysis: Some aspects and future perspectives. *Journal of Hydrology*. 100: 293-313.

Bradshaw, G. A., and B. A. McIntosh. 1994. Detecting Climate-Induced Patterns using Wavelet Analysis. *Environmental Pollution*. 83(1-2): 135-142.

Bras, R. L., and I. Rodriguez-Iturbe. 1985 *Random Functions and Hydrology*. Addison-Wesley Publishing Company, Massachusetts.

Chan, A. M-C. 1998 Towards the generation of long-term, fine time-resolution rainfall data using wavelet analysis. M.Sc. (project report). Univ. of Guelph. 174 pp + diskette.

Chatfield, C. 1996 *The Analysis of Time Series: an Introduction*. Chapman & Hall, London,

Cowpertwait, P. S. P. 1991. Further Developments of the Neyman-Scott Clustered Point Process for Modeling Rainfall. *Water Resources Research*. 27(7): 1431-1438.

Fiering, Myron B., and Barabara B. Jackson. 1971. *Synthetic Streamflows*. Water Resources Monograph 1. American Geophysical Union. Washington, D.C.,

Foufoula-Georgiou, E., and W. Krajewski. 1995. Recent Advances in Rainfall Modeling, estimation and forecasting. *Reviews of Geophysics*. 33(suppl. B):1125-1137.

Franz, D. D. 1970 *Hourly Rainfall Synthesis for a Network of Stations*. Department of Civil Engineering, Stanford University, Stanford, California, Technical Report No. 126,.

Hershenhorn, J., and D. A. Wollhiser. 1987. Disaggregation of Daily Rainfall. *Journal of Hydrology*. 95(3): 299-322.

Hipel, K. W. 1985. Time Series Analysis in Perspective. *Water Resources Bulletin*. 21(4): 609-624.

James, William. 1994a *Rules for Responsible Modelling*. First Ed., Computational Hydraulics International, Guelph, Ontario,

James, William. 1994b. On Reasons Why Traditional Single-Valued, Single-Event Hydrology (Typical Design Storm Methodology) Has Become Simple-Minded, Dishonest and Unethical. *Urban Hydrology & Hydraulics Workshop*. US Army Corps of Engineers Hydrologic Engineering Center, Davis, California, September 13 to 15.

James, W. and Robinson, M.A. 1982 *Continuous Models Essential for Detention Design*. Conf. on Stormwater Detention Facilities Planning Design Operation and Maintenance. ASCE, Henniker, New Hampshire, U.S.A. pp. 163 175, Aug 16,.

Keifer, C. J., and H. H. Chu. 1957. Synthetic Storm Pattern for Drainage Design. *Journal of the Hydraulics Division*. 83(HY4), 1332.1-1332.25.

Klemeš, V. 1986. Dilettantism in Hydrology: Transition or Destiny? *Water Resources Research*. 22(9): 177S-188S.

Kline, K. S., and M.7 J. McFarland. 1988. An Event-Oriented Precipitation Model. *International Winter Meeting of the ASAE*, Paper No. 88-2519. Hyatt Regency, Chicago, Dec. 13 to 16,

Koutsoyiannis, D. 1994. A Stochastic Disaggregation Method for Design Storm and Flood Synthesis. *Journal of Hydrology*. 156(1-4): 193-225.

Koutsoyiannis, D., and T. Xanthopoulos. 1990. Dynamic Model for Short-Scale Rainfall Disaggregation. *Hydrological Sciences Journal*. 35(3): 303-322.
Lall, U., B. Rajagopalan, and D. G. Tarboton. 1996. A nonparametric wet/dry spell model for resampling daily precipitation. *Water Resources Research*. 32(9):2803-2823.
Niemczynowicz, J. 1994. New Aspects of Urban Drainage and Pollution Reduction - Towards Sustainability. *Water Science and Technology*. 30, 269-277.
Ormsbee, L. E. 1989. Rainfall Disaggregation Model for Continuous Hydrologic Modeling. *Journal of Hydraulic Engineering*. 115(4): 507-525.
Ott, R. F. 1971 *Streamflow Frequency using Stochastically Generated Hourly Rainfall*. Dept. of Civil Engineering, Stanford University, Stanford, California, Tech. Rep. 151.
Pattison, A. 1965. Synthesis of Hourly Rainfall Data. *Water Resources Res.* (1) 489-498.
Robinson, M.A. and W. James. 1981. *Continuous SWMM Quality Modelling for the City of Hamilton Using Atmospheric Environment Service Data*. Proc. of the Conf. on Water Quality and Stormwater Management Modelling. Niagara Falls, Ontario, pp. 469 492,
Srikanthan, R., and T. A. McMahon. 1983a. Sequential Generation of Short Time-Interval Rainfall Data. *Nordic Hydrology*. 14(5): 277-306.
Srikanthan, R., and T. A. McMahon. 1983b. Stochastic Simulation of Daily Rainfall for Australian Stations. *Transactions of the ASAE*. 26(3): 754-759, 766.
Srikanthan, R., and T. A. McMahon. 1985. *Stochastic Generation of Rainfall and Evaporation Data*. Australian Water Resources Council, Technical Paper No. 84. Department of Resources and Energy, Australian Water Resources Council Research Project No. 80/149. Australian Government Publishing Service,
Tan, S. K., and S. Y. Sia. 1997. Synthetic Generation of Tropical Rainfall Time Series using an Event-Based Method. *Journal of Hydraulic Engineering*. 2(2): 83-89.
Wyrick, R. G. 1974. *A Daily Stochastic Precipitation Generator for Continuous Streamflow Models*. Paper Presented at the 55th Annual Meeting of the American Geophysical Union. Washington, D.C., April 9, 1974.
Yevjevich, V. 1991. Tendencies in Hydrology Research and Its Applications for 21st Century. *Water Resources Management*. 5(1): 1-23.

Bibliography

Allerup, P. 1996. Rainfall Generator with Spatial and Temporal Characteristics. *Atmospheric Research*. 42: 89-97.
Arnold, J. G., and J. R. Williams. 1989. Stochastic Generation of Internal Storm Structure at a Point. *Transactions of the ASAE*. 32(1): 161-167.
Austin, T. A., and B. J. Claborn. 1974. *Statistical Model of Short-Duration Precipitation Events*. Proceedings, Symposium on Statistical Hydrology. Tucson, Arizona, August 31 to September 2, 1971, ARS-USDA No. 1275. pp. 356.
Berndtsson, R., K. Jinno, A. Kawamura, M. Larson, and J. Niemczynowicz. 1994. Some Eulerian and Langrangian Statistical Properties of Rainfall at Small Space-Time Scale. *Journal of Hydrology*. 153: 339-355.
Bo, Z., S. Islam, and E. A. B. Eltahir. 1994. Aggregation-Disaggregation Properties of a Stochastic Rainfall Model. *Water Resources Research*. 30(12): 3423-3435.
Bonser, J. D. *A Probabilistic Spatial-Temporal Model of Storm Rainfall*. Ph.D. Thesis, University of Waterloo, 1986.
Brown, B.G., R.W. Katz, and A. H. Murphy. 1985. Exploratory Analysis of Precipitation Events with Implications for Stochastic Modeling. *Journal of Climate and Applied Meteorology*. 24(1): 57-67.

Chin, D. A. 1995. A Scale Model of Multivariate Rainfall Time Series. *Journal of Hydrology (Amsterdam)*. 168(1): 1-16.

Chulsang-Yoo, J. B. Valdes, and G. R. North. 1996. Stochastic Modeling of Multidimensional Precipitation Fields Considering Spectral Structure. *Water Resources Research*. 32(7): 2175-2187.

Cooke, R. A., and S. Mostaghimi. *Comprehensive Generator for Correlated Rainfall Properties and Spatial Rainfall Patterns*. Northeast Agri/Bio-Engineering Conference (NABEC) Paper No. 94233. American Society of Agricultural Engineers. Held at the University of Guelph, Guelph, Ontario, Canada, July 24 to 27, 1994.

Cowpertwait, P. S. P., P. E. O'Connell, A. V. Metcalfe, and J. A. Mawdsley. 1996. Stochastic Point Process Modelling of Rainfall. II. Regionalisation and Disaggregation. *Journal of Hydrology*. 175(1-4): 47-65.

Elsner, J. B., and A. A. Tsonis. 1993. Complexity and Predictability of Hourly Precipitation. *Journal of the Atmospheric Sciences*. 50(3): 400-405.

Entekhabi, D. Rodriguez-Iturbe, I. Eagleson, P. S. 1989. Probabilistic Representation of the Temporal Rainfall Process by a Modified Neyman-Scott Rectangular Pulses Model: Parameter Estimation and Validation. *Water Resources Research*. 25(2): 295-302.

Farres, P. J. *A New Approach to the Simulation of Rainfall*. Seminar Papers No. 28, University of Newcastle upon Tyne, Dept. of Geography, 1975.

Franz, D. D. 1974. Hourly Rainfall Generation for a Network. In: *Proceedings, Symposium on Statistical Hydrology*. Tucson, Arizona, August 31 to September 2, 1971, ARS-USDA No. 1275. pp. 147.

Gregory, J. M., T. M. L. Wigley, and P. D. Jones. 1992. Determining and Interpreting the Order of a Two-State Markov Chain: Applications to Models of Daily Precipitation. *Water Resources Research*. 28(5): 1443-1446.

Gyasi-Agyei, Y. and G. R. Willgoose. 1997. A Hybrid Model for Point Rainfall Modelling. *Water Resources Research*. 33(7): 1699-1706.

Istok, J.D., and L. Boersma. 1989. Stochastic Cluster Model for Hourly Precipitation Data. *Journal of Hydrology*. 106(3): 257-285.

Katz, R. W., and M. B. Parlange. 1995. Generalizations of Chain-Dependent Processes: Application of Hourly Precipitation. *Water Resources Research*. 31(5): 1331-1341.

Kavvas, M. L., and J. W. Delleur. *The Stochastic and Chronologic Structure of Rainfall Sequences, Application to Indiana*. Purdue University Water Resources Center Technical Report No. 57, Purdue University, West Lafayette, IN., 1975.

Kedem, B., and L. S. Chiu. 1987. Are Rain Rate Processes Self-Similar? *Water Resources Research*. 23: 1816-1818.

Kedem, B., and L. S. Chiu. 1987. Lognormality of Rain Rate. In: *Proceedings of the National Academy of Sciences USA*. Vol. 84, No. 4, P 901-905.

Koepsell, R. W., and J. B. Valdes. 1991. Multidimensional Rainfall Parameter Estimation From Sparse Network. *Journal of Hydraulic Engineering (ASCE)*. 117(7): 832-850.

Koutsoyiannis, D., and A. Manetas. 1996. Simple Disaggregation by Accurate Adjusting Procedures. *Water Resources Research*. 32(7): 2105-2117.

Krstanovic, P. F., and V. P. Singh. 1993a. A Real-Time Flood Forecasting Based on Maximum-Entropy Spectral Analysis: I. Development. *Water Resources Management*. 7(2): 109-129.

Krstanovic, P. F., and V. P. Singh. 1993b. A Real-Time Flood Forecasting Based on Maximum-Entropy Spectral Analysis: II. Application. *Water Resources Management*. 7(2): 131-151.

Lindner, M. A. *A Program Suite for the Statistical Analysis and Comparison of Historical and Synthetic Hydrologic Records*. School of Civil Engineering, University of New South Wales, Report No. 153. 1978.

Marien, J. L., and G. L. Vandewiele. 1986. A Point Rainfall Generator with Internal Storm Structure. *Water Resources Research*. 22(4): 475-482.

Mellor, D. 1996. The Modified Turning Bands (MTB) Model for Space-Time Rainfall. I. Model Definition and Properties. *Journal of Hydrology*. 175(1-4):113-127.

Mellor, D., and P. E. O'Connell. 1996. The Modified Turning Bands (MTB) Model for Space-Time Rainfall. II. Estimation of Raincell Parameters. *Journal of Hydrology*. 175(1-4):129-159.

Morris, C.D.1984. A Stochastic Model for a Small-Time-Interval-Intermittent Hydrologic Process. *Journal of Hydrology*. 68: 247-272.

Nguyen, V. T. V., and G. R. Pondey. 1994. Estimation of Short-Duration Rainfall Distribution using Data Measured at Longer Time Scales. *Water Science and Technology*. 29: 39-45.

Nguyen, V.T.V. 1984. Stochastic Characterization of Temporal Storm Patterns. Water Science and Technology. 16(8): 147-153.

Nguyen, V.T.V. 1988. Stochastic Analysis of Rainfall for Urban Drainage Studies. In: *Hydrological Processes and Water Management in Urban Areas*. Proceedings of the International Symposium 24-29 April 1988, Duisburg, West Germany. International Hydrological Programme of UNESCO Publication 1988. P 77-84.

Nguyen, V-T-V., and J. Rousselle. 1981. A Stochastic Model for the Time Distribution of Hourly Rainfall Depth. *Water Resources Research*. 17(2): 399-409.

Nicks, A. D., and L. J. Lane. 1989. Chapter 2. Weather Generator. In: *USDA-Water Erosion Prediction Project (WEPP)*. eds. L. J. Lane and M. A. Nearing. National Soil Erosion Research Lab, Report No. 2, West Lafayette.

Olsson, J., J. Niemczynowicz, and R. Berndtsson. 1993. Fractal Analysis of High-Resolution Rainfall Time Series. *Journal of Geophysical Research – Part D – Atmospheres* – Printed Edition. 98(12): 23265-23275.

Olsson, J., J. Niemczynowicz, R. Berndtsson, and M. Larson. 1992. An Analysis of the Rainfall Time Structure by Box Counting – Some Practical Implications. *Journal of Hydrology*. 137(1-4): 261-277.

Onof, C., and H. S. Wheater. 1995. Modelling of Rainfall Time-Series using the Bartlett-Lewis Model. *Proceedings of the Institution of Civil Engineers – Water Maritime and Energy*. 112(4): 362-375.

Onof, C., and H. S. Wheater. 1996. Modelling of the Time-Series of Spatial Coverages of British Rainfall Fields. *Journal of Hydrology*. 176(1-4): 115-131.

Padmanabhan, G., and J. W. Delleur. 1978b. *Generated Urban Storm Drainage Quantity and Quality Data*. Purdue Water Resources Research Center Technical Report No. 108, Purdue Univ. Lafayette, IN.

Padmanabhan, G., and J. W. Delleur. *Statistical and Stochastic Analysis of Synthetically Generated Urban Storm Drainage Quantity and Quality Data: Systematic Development of Methodologies in Planning Urban Water Resources for Medium Size Communities*. Purdue University Water Resources Research Center, Technical Report No. 108, Purdue University, West Layfeyette, IN, 1978a.

Paturel, J. E., and B. Chocat. 1994. Analysis of the Rain Phenomenon at Fine Spatial Scale. *Water Science Technology*. 29(1-2): 31-37.

Ramachandra-Rao, A., and B. T. Chenchayya. Probabilistic Analysis and Simulation of the Short Time Increment Rainfall Process. Purdue University *Water Resources Research* Center Technical Report No. 55. Purdue University. West Lafayette, IN, 1984.

Robin, M. J. L., A. L. Gutjahr, E. A. Sudicky, and J. L. Wilson. 1993. Cross-Correlated Random Field Generation with the Direct Fourier Transform Method. *Water Resources Research*. 29(7): 2385-2397.

Rodriguez-Iturbe, I., and P. S. Eagleson. 1987. Mathematical Models of Rainstorm Events in Space and Time. *Water Resources Research*. 23(1): 181-190.
Schertzer, D., and S. Lovejoy. 1995. From Scalar Cascades to Lie Cascades: Joint Multifractal Analysis of Rain and Cloud Processes. In: *Space and Time Scale Variability and Interdependencies in Hydrological Processes*. (Cambridge University Press), 1995, Pp 153-174.
Schilling, W. 1984. Real-Time Estimation and Forecasting of Spatially Distributed Areal Rainfall. *Water Science and Technology*. 16: 327-348.
Shaw, S. R. 1983. An Investigation of the Cellular Structure of Storms. *Journal of Hydrology*. 62(1-4): 63-79.
Todorovic, P., and V. Yevjevich. *Stochastic Process of Precipitation*. Hydrology Papers, No. 35, Colorado State University, Fort Collins, CO, 1969.
Walker, J. 1985. The Amateur Scientist: Searching for Patterns of Rainfall in a Storm. Scientific American, 252(1): 112-119.
Wang, Chaoying. *Stochastic Partial Differential Equations in Simulation of Random Fields and Applications in the Multidimensional Model of Rainfall*. M.A.Sc. Thesis, University of Waterloo, 1992.
Wiser, E. H. 1974. A Precipitation Data Simulator using a Second Order Autoregressive Scheme. In: *Proceedings, Symposium on Statistical Hydrology*. Tucson, Arizona, August 31 to September 2, 1971, ARS-USDA No. 1275. pp. 120-134.
Woolhiser, D. A., T. W. Econopouly, and D. R. Davis. 1989. Disaggregation of Daily Rainfall. In: *Proceedings of the International Conference on Channel Flow and Catchment Runoff: Centennial of Manning's Formula and Kuichling's Rational Formula*. pp. 18-26.

This page intentionally left blank.

Chapter 7

DYNAMIC RAIN SYSTEMS

To talk of the weather it's nothing but folly
For when it rains on the hill, it shines in the valley,
- Michael Denham, *Proverbs*

A sunshiny shower Won't last half an hour.
- Michael Denham, *Proverbs*

Oh what a blamed uncertain thing
This pesky weather is!
It blew and snew and then it thew
And now, by jing, it's friz!
- Philander Johnson.

The use of a web-based group decision support system (DSS) for the design of benign urban water systems is presented. First we present an approach for estimating the mean speed and direction of a multi-cellular rainstorm using rate-of-rain gages. Code for spatial analysis of storms and their cells is described. Then we present a Monte Carlo analysis of rain storms observed in Toronto and Hamilton, Ontario, in which we demonstrate the robustness of the analysis with respect to observational uncertainty.

We draw heavily on the work of two former graduate students: Ron Scheckenberger and Peter Nimmrichter (for more details of their studies see: www.eos.uoguelph.ca/webfiles/james). First we present an approach for estimating the mean speed and direction of a multi-cellular rainstorm (Scheckenberger, 1983; Nimmrichter, 1986). Then we describe code for spatial analysis developed by Rob James for a web decision support system (PCSWMMGIS, see www.chi.on.ca). Methodology is illustrated in Figures 7.1 to 7.6 (James and James, 1996).

7.1 Introduction

An important function of a rainfall analysis package is, for model calibration purposes, generation of storm-cell vectors (speed and direction) for use in computing sub-spatial-mean and time-step-averaged input for urban water systems models. This function is considered to be essential for accurate water quality modeling, especially where the rainstorm moves synchronously with the drainage flood.

Credible rain data is fundamentally important for determining the water quality impacts of projected urban development and/or remediation. Distributions of rainfall intensities or depths over time and space are of particular importance to the hydrologic modeling of urban water systems. Still prevalent in design offices, is the use of stationary synthetic design storms, such as those derived from intensity-duration-frequency-averaged intensities and artificial storm distributions (Ormsbee, 1989). Drawbacks to the single-event design storm approach include the fact that dynamic effects of storm movement and coverage are ignored (James, 1994; Ormsbee, 1989). An alternative to using stationary event hydrology is to analyse observed rainfall data for kinematic spatial storm coverage, and then to compute spatial and time averages as the storm tracks across the urban area. Here, we analyze data sets from various meteorological stations in Edmonton, Canada, for spatial kinematics.

> Rain storm movement is considered to be essential for runoff model calibration, because the errors invoked by failing to do so are commensurate with other the other uncertainties involved, modeling and observational.

7.2 Kinematics of storm cells

In the Rainpak algorithm, storm vectors are produced by a geometric analysis of optimum lags between cell occurrences throughout the observation network (Scheckenberger, 1983; Nimmrichter, 1986). As shown in Figures 7.1 to 7.6, optimal lags are computed by serial lag cross-correlations between two hyetographs (Marshall, 1980). Using a method that assumes that storm cell planforms are regular and conformable, storm cell velocities are computed from the simple geometric relations given below in equations (1) to (3) (Scheckenberger, 1983). Correlograms, vectors and rosettes of means are plotted interactively (James et al., 2000).

Thunderstorms display intense, short-duration rainfall, rapidly varied spatially. High variability in both time and space is possibly explained by turbulence effects at all scales, meso to micro. Certain features, however, even for multi-cellular storms, may be identified, such as cell peak intensity and duration, and these show some degree of persistence as the storm system tracks across an urban area. Experience shows that the drawback to tracking an identifiable feature is the subjectivity involved. RAINPAK in PCSWMMGIS estimates the speed and direction of a cell by computing the likely lag time for a particular cell's overall hyetograph shape to track across two identified rain gauges. Subjectivity is avoided by computing the serial lag correlation and taking the optimal lag as the time of travel, provided that the optimal correlation is good (e.g. >0.85). Determinations with a lower correlation coefficient are ignored. Underlying the method is the assumption that cell shapes are, on the average, simple and conformable.

Cartesian co-ordinates of at least three rain gauges and the relative time for the general hyetograph shape to track over each gauge are required to compute one direction and speed of a cell. Every combination of three rain gauges produces a cell velocity vector. Because so many vectors are computed, a mean storm vector is determined by ignoring all determinations that are more than two standard deviations from the vector mean. Vectors are determined as follows (Scheckenberger, 1983).

Let: (x, y) be the Cartesian co-ordinates representing three rain gauge locations denoted 1, 2, 3,

(x^T, y^T) be the transformed co-ordinates where the x axis represents the storm direction,

α be the angle of the transformation,

v_s be the cell speed dx/dt, and

$\Delta t_{(2-1)}$ be the time of travel between gauges 1 and 2.

Then

$$v_s = (x_1 \cos\alpha + y_1 \sin\alpha - x_2 \cos\alpha + y_2 \sin\alpha)\Delta t_{2-1} \tag{7.1}$$

$$v_s = (x_2 \cos\alpha + y_2 \sin\alpha - x_3 \cos\alpha + y_3 \sin\alpha)\Delta t_{3-2} \tag{7.2}$$

The mean of these two velocities is taken.

Now set

$$A = \frac{\Delta t_{(2-1)}}{\Delta t_{(3-2)}}$$

so that, after some manipulation,

$$\alpha = \tan^{-1} \frac{(x_2 - x_1 + Ax_2 - Ax_3)}{(Ay_3 - Ay_2 + y_1 - y_2)} \tag{7.3}$$

Comprehensive tests on this methodology were carried out and reported by Nimmrichter (1986) and James and James (1997). Implementation for a dataset from Edmonton is depicted in Figures 7.1 to 7.6.

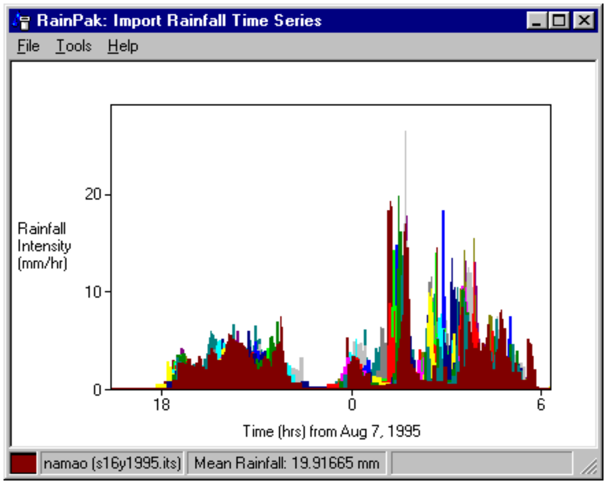


Figure 7.1 Derived 12-h 1-min increment hyetographs for all 20 tipping bucket rain gauges for Aug 7 1995 (rain data courtesy City of Edmonton).

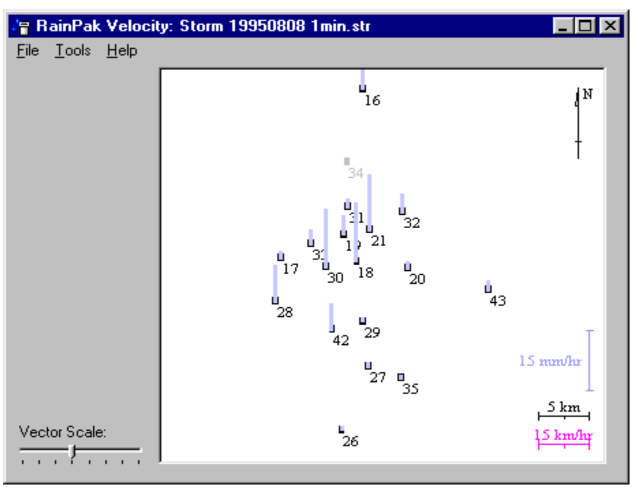


Figure 7.2 Hyetograph playback at selectable speed and scale (rain data courtesy City of Edmonton).

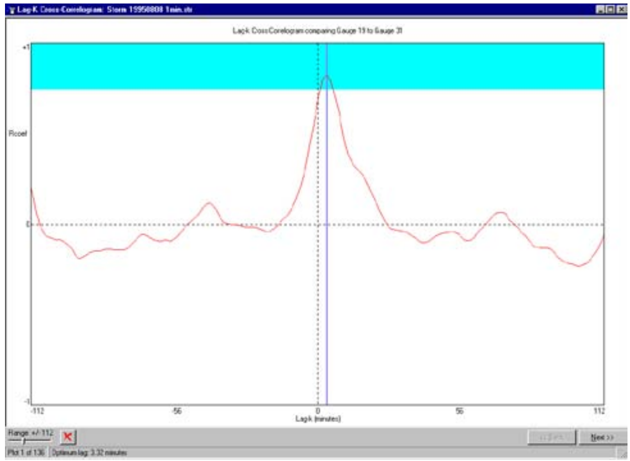


Figure 7.3 Auto-correlogram showing optimal lag (rain data courtesy City of Edmonton).

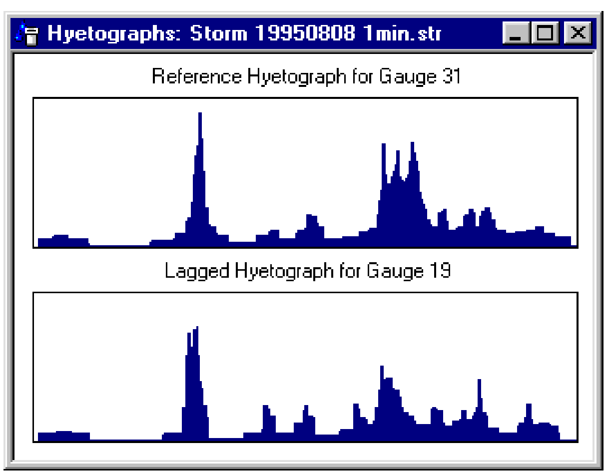


Figure 7.4 Two selected hyetographs auto-lagged for maximum serial correlation (rain data courtesy City of Edmonton).

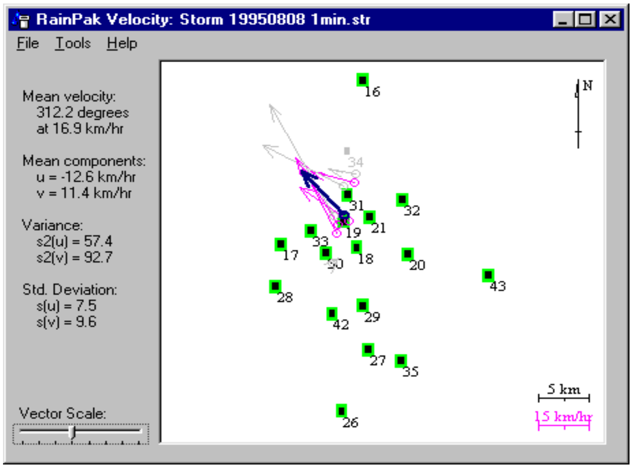


Figure 7.5 Computed cell velocities for selected storm cell hyetographs (heavy = vector mean) (rain data courtesy City of Edmonton).

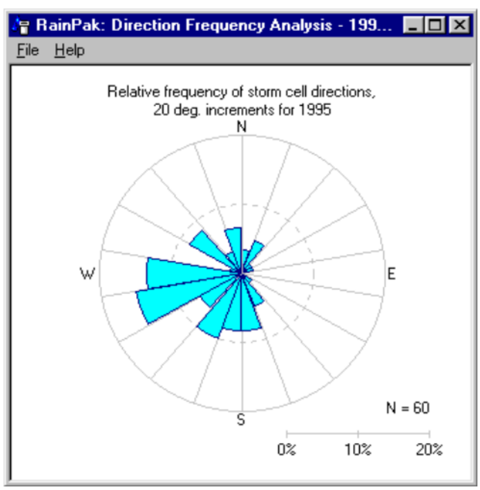


Figure 7.6 Relative frequency of collected computed storm velocities for all cells for 1995 (rain data courtesy City of Edmonton).

7.3 Uncertainty

Ground level rain gages provide the information from which the Rainpak computes storm cell velocity. Errors in rainfall measurement are many, resulting from improper calibration, malfunctions, poor location of rain gages, equipment limitations, and from data collection and reduction. Erroneous data may produce amplified errors in computed results, which in turn may cast doubt on the accuracy of the computations. Robustness of the velocity algorithm in relation to various rain gage errors is reported next.

For objective evaluation of the effect of input errors on the velocity analysis, a number of different approaches are available, each with individual strengths and weaknesses: 1. Monte Carlo simulations, 2. statistical derivations, and 3. first-order analysis. In simple applications the three methods have produced similar results. The advantage of the Monte Carlo method is that it is easy to understand and simple to apply to problems in which some variation of the input variables is expected. In this method a series of random, independent events is generated (Sobol, 1974). The average of the results of all trials provides a better understanding of the estimate of the objective function. The drawback of this method is the extensive computing resources required in comparison with the other two methods.

In his study, the hyetograph analysis used by Nimmrichter (1986) was based on estimated time-of-peak rain-rate, the same method used by Scheckenberger (1983). However the present study avoids searching for maxima, and uses direct determination of times-of-travel $\Delta t_{(2-1)}$, $\Delta t_{(2-3)}$ and $\Delta t_{(3-1)}$ between the rain gages. Note however that the same velocity algorithm used in PCSWMM-RAINPAK was used by Nimmrichter. It is the robustness of that algorithm that is the focus of this section.

7.4 Sensitivity of storm velocity to timing error

Nimmrichter wrote program STSEN to generate a series of perturbations, following either a uniform or normal distribution, and some of the results of his series of 136 runs are described in the following sections. Observed hyetographs for Toronto and Hamilton were used, as shown in Table 7.1. Each run usually comprised 1000 trials. A trial consists of 1. the generation of a random "timing error" sequence, 2. modification of the observed rainfall time-of-travel using the "timing error" sequence and 3. running the velocity algorithm, using the fictitious time-of-travel dataset. This amounted to a total of 136,000 runs of the velocity algorithm.

Table 7.1 Storm cells and tests for Monte Carlo timing error analysis

City of Toronto			
Test	Δt (min)	Storm	# Trials
1	5	19790610 Cell 1	1000
2	5	19770706 Cell 1	1000
3	5	19750823 Cell 3	1000
4	5	19800619 Cell 2	1000
5	5	19790630 Cell 2	1000

Hamilton-Wentworth Region			
Test	Δt (min)	Storm	# Trials
1	1	19810811 Cell 1	1000
2	1	19820728 Cell 2	1000
3	1	19820929 Cell 1	1000
4	1	19830729 Cell 3	1000
5	1	19830801 Cell 1	1000

STSEN automatically altered the time-of-travel between each pair of rain gages by plus and minus 5 min. This was repeated for every rain gage and the resultant velocity vector and its standard deviation were determined. This analysis may be useful if there is a suspected problem of synchronization of rainfall records (admittedly less likely these days where networked systems are more common).

To determine the effect of timing error on the resultant storm velocity vector, the fictitious time-of-travel was assumed to follow a normal distribution with a mean equal to the observed time-of-travel and various standard deviations were used to define different levels of confidence in the records (note that the time-of-travel for this Monte Carlo analysis was taken to occur in the middle of the time step i.e. all observed times-of-travel were shifted by 2.5 min for the City of Toronto network and by 0.5 min for Hamilton-Wentworth network). Levels of confidence of 80, 85, 90, 95, 96, 97, 98 and 99 % were tested.

The first phase of the sensitivity analysis indicated that the information content of the rainfall record is different for every storm event. Similar perturbations on the timing of different cells affect velocity estimates differently. A generalization about the variability of average cell velocity estimates cannot be made because the resultant distribution of estimates is evidently different for all storms.

7.5 Sensitivity of storm velocity to timing resolution

Apart from any errors that may occur due to the operation of the gages, the analysis of storm kinematics requires accurate times-of-travel but (in the case of the Toronto data) a time step of 5 min gives information no better than 5 min.

The results of the trials are given in Table 7.2 and Figures 7.2 and 7.3. Results indicate that 1. not knowing the true time-of-travel of a storm event can cause wide variations in the possible estimates of the cell velocity and again 2. the information content of the rainfall record for different storms is different.

Ten tests using uniformly distributed perturbations were carried out on the storm of 19790610 (Cell #1). The results of these tests, shown in Table 7.3, indicate that the variability of the statistical parameters over many identical tests is minor.

Table 7.2 The effect of timing resolution on Rainpak velocity estimates.

STORM	CELL	α	VAR	STD	SKEW	SPEED	VAR	STD	SKEW
City of Toronto									
19790610	1	289	117.57	10.84	-0.36	48.0	47.51	6.89	1.12
19770706	1	5	18.94	4.35	-0.38	24.3	28.35	5.32	6.42
19750823	3	191	23.36	4.83	0.65	17.6	2.62	1.62	0.47
19790630	2	137	26.52	5.15	0.01	31.5	17.18	4.14	1.32
19800610	2	212	142.70	11.95	0.30	34.9	25.49	5.05	2.02
Hamilton-Wentworth Region									
19810811	1	245	20.86	4.57	0.39	52.2	38.77	6.23	0.96
19820728	2	319	18.17	4.26	0.98	57.5	14.11	3.76	0.76
19820920	1	10	1.18	1.09	-0.01	15.1	0.07	0.26	-2.89
19830729	3	321	0.73	0.85	418.44	30.8	2.74	1.66	0.97
19830801	1	236	5.59	2.36	3.54	51.5	4.68	2.16	1.02

NOTE: All tests were conducted using 1000 trials.

Tests were also conducted to determine the effect of the number of trials on the statistical parameters. The results of' these tests are given in Table 7.3 and illustrated in Figures 7.4 and 7.5. Even at low numbers of trials (i.e. 10), evidently a reasonable estimate of the cell velocity can be achieved.

One difficulty with the 5 min timing resolution on rainfall records is that, in many cases, times-of-travel at more than three gages is identical, making calculation of the cell speed impossible (unless of course the cell was truly stationary). Nimmrichter then modified the rainfall timing slightly so that no three gages have the same times-of-travel. The above testing seems to indicate that even if three times-of-travel are identical (as with the storm of 19790610) a reasonable estimate of the storm speed can be made using a small number of trials in a Monte Carlo simulation. This means that the analyst could use the rainfall record directly, avoiding subjectivity in the estimates of the cell speed. This will obviously be useful in processing an extended continuous rainfall record. For this reason, the whole Rainpak analysis is done automatically.

Table 7.3 Variability of statistical parameters.

Δt (min)	TEST	DIR	VAR	STD	SKEW	SPEED	VAR	STD	SKEW
5	1	289	117.57	10.84	-0.36	48.0	47.51	6.89	1.12
5	2	289	112.01	10.60	-0.38	47.9	39.51	6.29	0.83
5	3	289	105.26	10.26	-0.14	47.9	44.42	6.66	1.37
5	4	289	116.99	10.82	-0.36	47.7	40.49	6.36	1.53
5	5	289	110.07	10.49	-0.48	47.9	43.49	6.59	1.12
5	6	289	119.11	10.91	-0.45	48.1	52.64	7.26	1.44
5	7	289	112.99	10.63	-0.49	47.9	42.01	6.48	1.26
5	8	289	129.10	11.36	-0.49	48.2	45.69	6.76	1.47
5	9	289	119.22	10.92	-0.45	48.4	51.37	7.17	1.48
5	10	289	118.32	10.88	-0.39	48.2	42.46	6.52	0.92

NOTE: All tests were conducted using 1000 trials.

The following units are applicable to the above Table.

^{o}N for direction (DIR)	km/h for speed (SPD)
$^{o}N^2$ for variance (VAR)	$(km/h)^2$ for variance
^{o}N for standard deviation (STD)	km/h for standard deviation

7.6 Conclusions

Storm cell velocities have been computed automatically by using geometrical identities, and the time of travel between known rain gauge stations based on serial lag correlations. Procedures are demonstrated in Figures 7.1 to 7.6 for rain recorded in Edmonton using the RAINPAK utility in PCSWMMGIS. An extensive sensitivity analysis was carried out on various observed storm events with the following conclusions:

Standard data collection time steps (1 or 5 min) are sufficient for storm dynamics analysis, but finer resolution reduces problems associated with time-of-travel coincidence and increases the confidence in the velocity estimate.

The confidence in a velocity estimate may be increased by including more rain gages in the simulation.

The geometric arrangement of rain gages within a collection network with respect to expected storm tracks may be a factor when analyzing data for storm dynamics.

The information content, for storm dynamics, of the rainfall record varies from one storm event to another.

Monte Carlo simulation can be used to avoid the influence of operator bias when difficulties of three or more gages have identical times-of-travel.

Over many Monte Carlo trials, random timing errors in a rain record may not cause deviations from the "true" direction. Also, the true mean direction can be generated from a reasonably small number of trials. However, any given trial may depart significantly from the mean (by 50 or 60^0 or more in some cases, where the local hyetograph is poor, perhaps just one or two rain gage bucket tips).

Table 7.4 Estimates of cell velocity using different numbers of trials.

Δt (min)	TRIALS	DIR	VAR	STD	SKEW	SPEED	VAR	STD	SKEW
5	1 NON	295							
5	1 MOD	303				44.2			
5	10	291	85.00	9.22	-0.23	49.5	40.25	6.34	0.01
5	50	292	125.70	11.21	-0.59	47.9	57.22	7.56	1.20
5	100	290	123.58	11.12	-0.40	47.9	46.61	6.83	1.25
5	250	290	108.83	10.43	-0.17	47.7	41.33	6.43	1.11
5	500	289	107.01	10.34	-0.26	48.0	39.10	6.25	0.73
5	1000	289	117.57	10.84	-0.36	48.0	47.51	6.89	1.12
5	2000	289	114.89	10.72	-0.42	47.9	43.49	6.59	1.01
5	3000	289	111.70	10.57	-0.45	47.9	43.78	6.62	1.13
5	4000	289	112.97	10.63	-0.45	47.9	42.94	6.55	1.23
5	5000	289	112.37	10.60	-0.48	47.9	43.03	6.56	1.21
5	6000	289	113.63	10.66	-0.54	47.9	44.63	6.68	1.27
5	7000	289	113.49	10.65	-0.49	47.9	44.25	6.65	1.27
5	8000	289	115.61	10.75	-0.59	48.0	44.42	6.66	1.30
5	9000	289	116.28	10.78	-0.71	48.0	45.20	6.72	1.32
5	10000	289	116.72	10.80	-0.81	48.0	44.93	6.70	1.29

NOTE: NON refers to the reported non-modified velocity

NOTE: MOD refers to the reported modified velocity

The following units are applicable to the above table.

^{o}N for direction (DIR)	km/h for speed (SPD)
^{o}N^2 for variance (VAR)	(km/h)^2 for variance
^{o}N for standard deviation (STD)	km/h for standard deviation

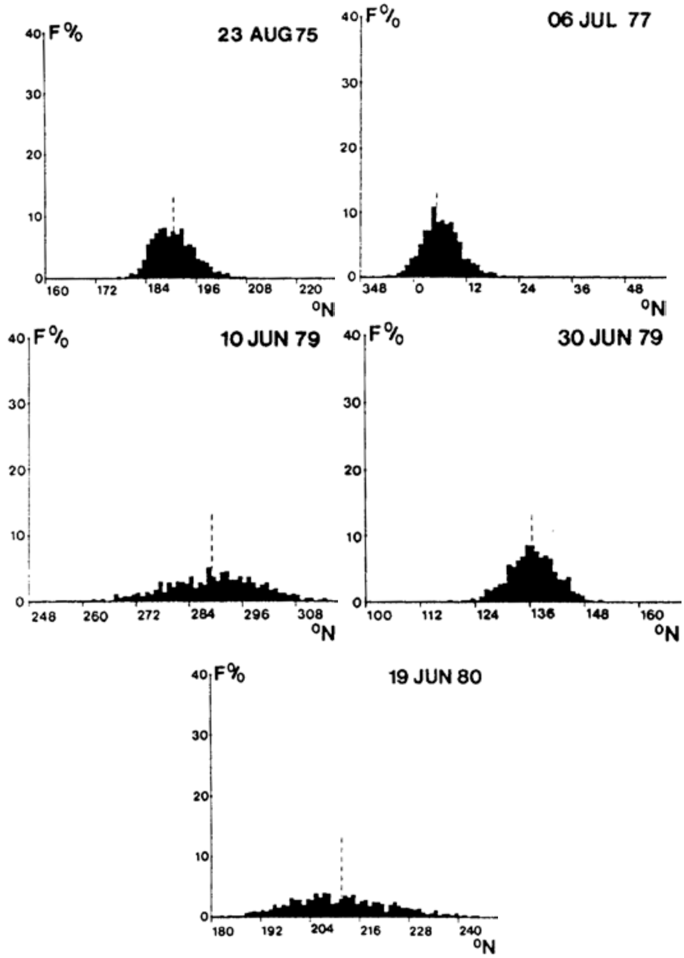


Figure 7.7 Time resolution frequency plots for storm cell direction (five Toronto storms).

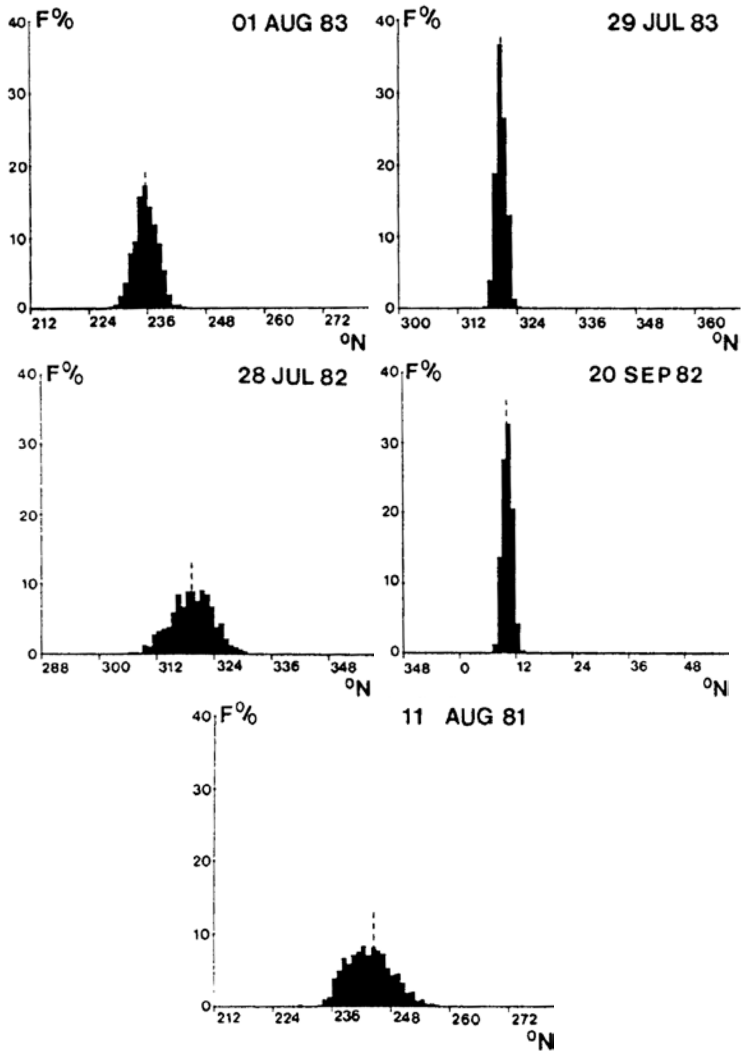


Figure 7.8 Time resolution frequency plots for storm cell direction (five Hamilton storms).

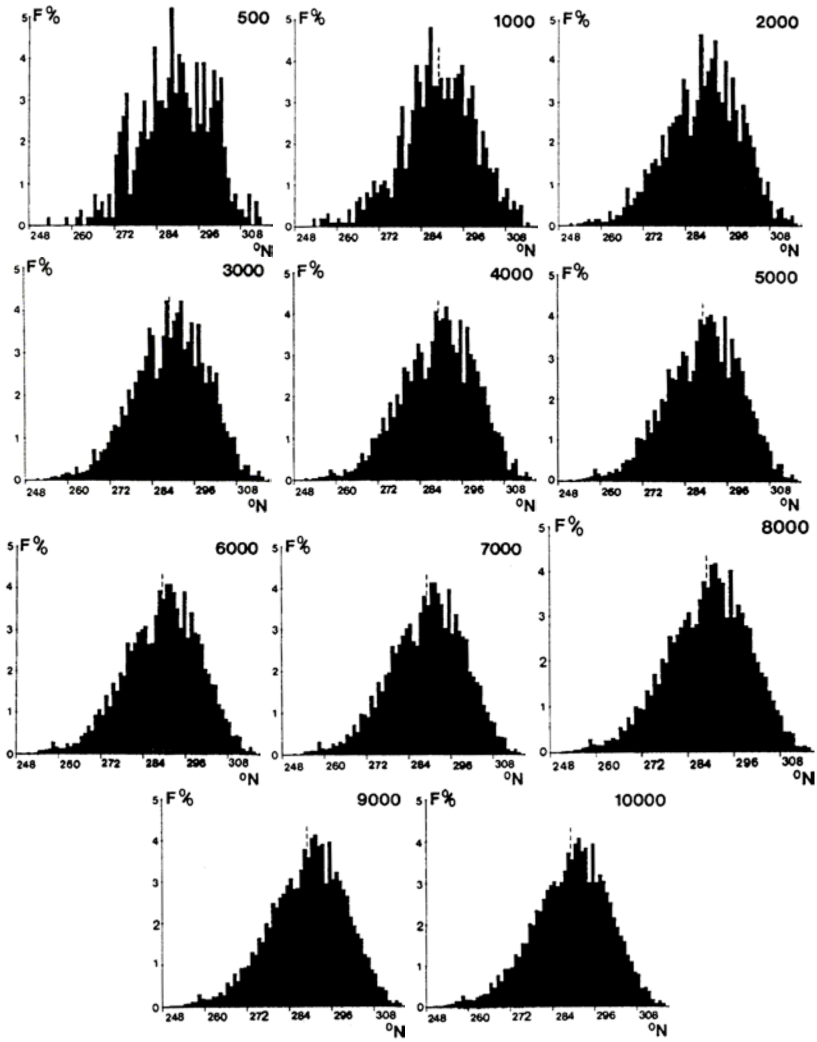


Figure 7.9 Time resolution frequency plots – effect of number of trials on storm cell direction estimates, 500 – 10,000 trials.

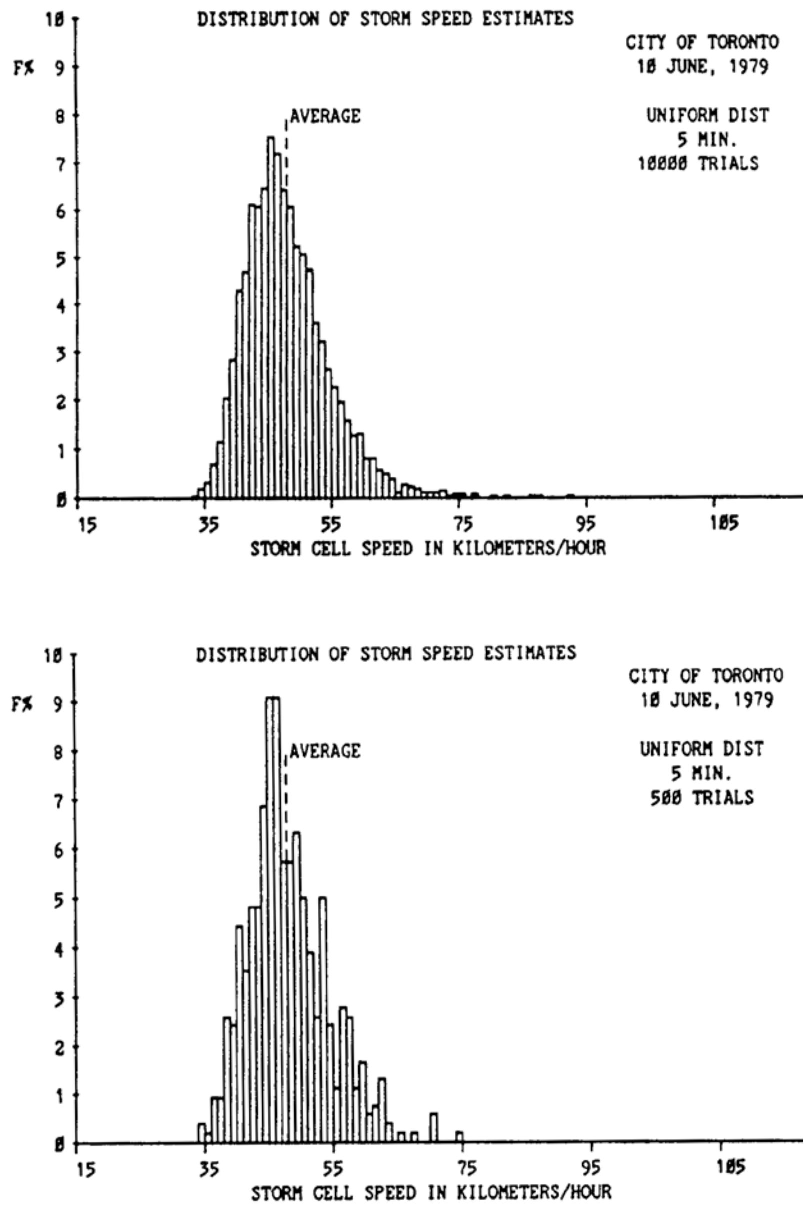


Figure 7.10 Time resolution frequency plots – effect of number of trials on storm cell speed estimates - 500 – 10,000 trials.

References

Fiering, M.B., and Jackson, B.B. 1971. Synthetic Streamflows, American Geophysical Union, Water Resources Monograph 1, Washington, 98

James, W. and Scheckenberger, R. 1983. Storm Dynamics for Urban Runoff, Proceedings of the International Symposium on Urban Hydrology, Hydraulics and Sediment Control, University of Kentucky, Lexington, Kentucky, July 25-28, pp. 11-18.

James, William. 1994. On Reasons Why Traditional Single-Valued, Single-Event Hydrology (Typical Design Storm Methodology) Has Become Simple-Minded, Dishonest and Unethical. Urban Hydrology & Hydraulics Workshop. US Army Corps of Engineers Hydrologic Engineering Center, Davis, California, September 13 to 15.

James, W.R.C. and James, W. 1996. Analysis of rain storm dynamics in the City of Edmonton. Interim report, CHI R197. 5 pp.

James, W., James, W.R.C., and Chan, A.M.C. 2000. Analysis of spatially-kinematic using a web group decision support system, and generation of long-term, fine-time-resolution rain, for urban water systems models. 5th Int. Workshop on Precip. in Urban Areas. "From precipitation measurements to design and forecasting modeling". Dec. Pontresina, Switzerland. Pp32-38.

Marshall, R.J. 1980. The Estimation and Distribution of Storm Movement and Storm Structure, Using a Correlation Analysis Technique and Rain gage Data, Journal of Hydrology, 48:19-39.

Nimmrichter, P. 1986. Dynamics of storms on the western shore of Lake Ontario. MEng. Thesis, McMaster U., 162pp.

Ormsbee, L.E. 1989. Rainfall Disaggregation Model for Continuous Hydrologic Modeling. J. of Hydraulic Engineering, 115(4):507-525.

Sargent, D.M. 1981. An Investigation Into the Effects of Storm Movement on the Design of Urban Drainage Systems, Public Health Engineering, Volume 9, pp. 201-207.

Scheckenberger, R. 1983. Dynamically Spatially Variable Rain Models for Storm water Management, M. Eng. thesis, McMaster University, Hamilton, Ontario.

Shearman, R.J. 1977. The Speed and Direction of Movement of Storm Rainfall Patterns, A Pilot Investigation with reference to Urban Storm Sewer Design, Meteorological Office, Brachnell, Berkshire, England, 53 PP.

Sobol, I.M. 1974. The Monte Carlo Method, translated and adapted from the second Russian edition by Robert Messer, John Stone and Peter Fortini, University of Chicago Press, Chicago, 63 pp.

Surkan, A.J. 1974. Simulation of Storm Velocity Effects of Flow from Distributed Channel Networks, Water Resources Research, Volume 10, Number 6, December, pp. 1149-1160.

Yen, B.C., and Chow V.T. 1969. A Laboratory Study of Surface Runoff Due to Moving Storms, Water Resources Research, Volume 5, Number 5, October, pp. 989-1006.

Bibliography

Related earlier publications by the authors

James, W. and Scheckenberger, R. 1985. RAINPAC - A Program Package for Analysis of Rainfall Inputs in Computing Storm Dynamics, Proc. of the Storm water and Water Quality Modeling Meeting, USEPA, Detroit, MI, EPA-600/9-85-003, pp. 81-100, Jan

Nimmrichter, P. and James, W. 1985 Dynamics of Storms on the Western Shore of Lake Ontario, Proc. of the Storm water and Water Quality Modeling Conf., sponsored by USEPA and Ontario Ministry of Environment, Toronto, Ontario, published by CHI Hamilton, pp. 29-60, Dec 5-6,

Nimmrichter, P. and James, W. 1986. Storm Dynamics in the Lake Ontario Region, Water Forum '86, A.S.C.E., Long Beach, California, pp. 1452-1459, Aug 4-6.

Chapter 8

DECISION SUPPORT SYSTEMS

Swift decisions are not sure. - Sophocles

But finally, with a sort of passion, as if abandoning calculation and casting himself upon the future, and uttering the phrase which men usually prelude their plunge into desperate and daring fortunes, "Let the die be cast" he hastened to cross the Rubicon. - Plutarch

8.1 Introduction

Continuous models are now reaching a level of complexity that requires that increasing attention be paid to code that manages the modeling (Orlob, 1992). Special user-friendly graphics-oriented codes are being integrated into pre- and post-processors for WQMs. Orlob calls this integrated code a decision support system (DSS). Numerous DSSs in the form of shells have been and are now being written for SWMM (James and James, 1993). Orlob (1992) does not include control of model reliability in his definition of a DSS.

But procedures to manage model error have also been aggregated under the term decision support systems, which is unfortunate in a sense, because error-management procedures relate to the control of model reliability at the model input phase, rather than output interpretation, as the phrase decision support may imply. Uber et al. (1992) strongly recommend that mathematical programming techniques be incorporated into DSSs for WQMs, especially those that have graphical user interfaces (GUIs) in a windows, interactive, menu-driven, pointing-device (WIMP) shell - in order to evaluate the larger number of alternatives that can now be generated.

Model builders and model users should quantify and present the uncertainty of their model output, to allow end-users the opportunity to evaluate the results, and the confidence that should be placed in them. In the case of SWMM and HSPF, most applications make arbitrary judgements about parameter selection and estimation, and use deficient data, and so the uncertainty and sensitivity to these assumptions are critical to end-users, but seldom reported.

Ideally, decision support systems (DSSs) should provide:

1. spatial data and facilities management systems,
2. time series management systems,
3. communications to remote, integrated, distributed databases,
4. long-term six-minutely rain generators,
5. analysis of the number and duration of exceedances and deficits in the output time series,
6. presentation tools for very long time series analysis,
7. data compression for long duration time series (eg 75 years of 6-minute data for hundreds of stations and dozens of chemicals),
8. integration with downstream, dependent models such as fluviology,
9. sensitivity analysis,
10. calibration,
11. error analysis,
12. tools to display model reliability and optimal complexity,
13. GUI WIMP presentation tools that display the uncertainty of the computed output, and
14. tools to help select the most cost-effective array of BMPs.

Such DSSs must become a normal part of the modeling procedures, so that uncertainty due to poorly defined processes and information are duly and responsibly revealed to the end-users.

There seems to be no doubt that our deterministic surface water quality models could benefit from more statistical tools. More statistical manipulations are necessary to build 75 years of data, and to present the results of 75 years of flow and many pollutants at many points. End-users of three-generation (3G) modeling will not be able to comprehend such results without a fair amount of statistical manipulation.

The melding of statistics with deterministic models will have another advantage: removal of the artificial distinction between data gatherers on the one hand and data consumers (modelers) on the other (or between monitoring people and analytical laboratory people on the one hand, and the modelers on the other). Provided such models incorporate suitable sensitivity analysis, very-long-term, complex, deterministic models with comprehensive statistical tools are useful management tools for data collection programs. They are the best means for filling in missing data. Ranking the sensitive parameters helps rank priority for selecting chemicals, sampling frequencies and locations, and accuracies of determination.

Widely distributed, integrated databases are becoming useful. Our surface water quality models should be tied in with suitable data and user networks, such as the Great Lakes Information Network (GLIN). We also need to encourage widespread use of our models, in education, decision making, engineering and research. We should write code that helps make our models available in different languages.

Continuous modeling requires a sequence of two main sets of modeling activities: i. calibration, and ii. inference.

The calibration activities involve parameter estimation and optimization against short-term, accurate, observed input functions; the inference activities involve long-term, continuous, synthetic or transposed input functions, and error analysis. Figure 8.1 shows the relationship between some of these activities: short-term calibration input functions (IFs); model; response functions (RFs); objective functions (OFs); performance evaluation functions (EFs); sensitivity analysis; parameter optimization; long-term continuous input functions; error analysis; long-term, continuous "fuzzy" response functions; and output interpretation or inference. Figure 8.1 is meant to be schematic and conceptual, and only to show the average sequence of activities in the broadest of terms.

It is clear that the modeling activities shown in Figure 8.1 will benefit from a well-written DSS. More detailed discussion on each of these activities follows.

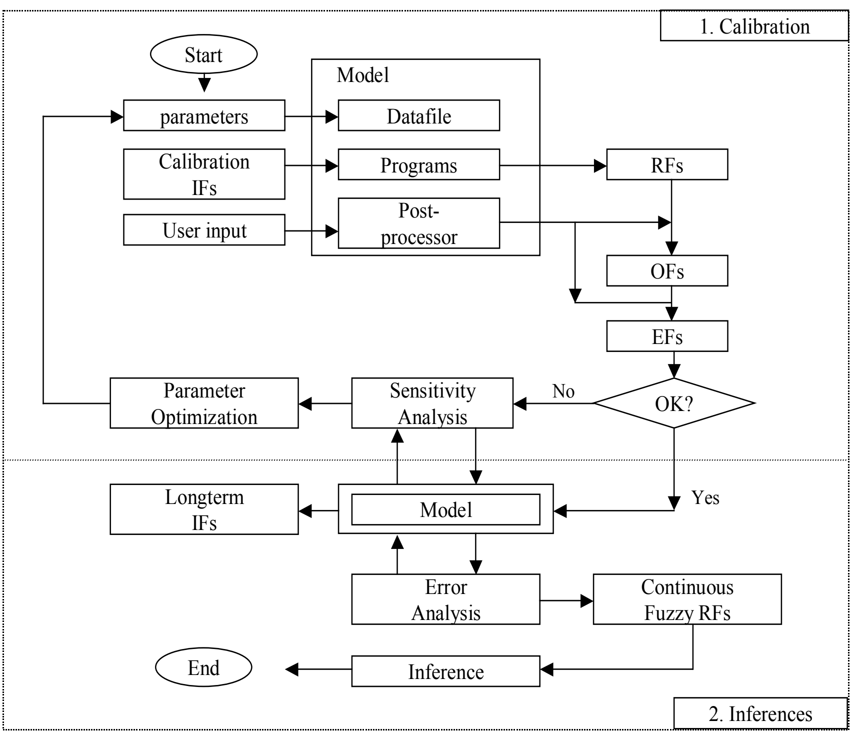


Figure 8.1: Schematic relationship between various modeling activities.

8.2 PCSWMM

PCSWMM is a web-oriented shell designed to facilitate the teaching, learning, design, and research use of the U.S. EPA Stormwater Management Model (SWMM). PCSWMM removes the error-prone steps that are a normal part of SWMM, using intelligent help, a large bibliographic database, web utilities and many other tools. From experience in international seminars and workshops, and graduate and undergraduate courses at the University of Guelph, we find using PCSWMM reduces the most prevalent run-time errors, and significantly improves the learning curve. However, there is a difficult compromise between ease-of-use and a deep comprehension.

PCSWMM (see Figure 8.2) provides a stepping stone for linking GIS and SWMM, and its features in point-form are:

- Seamless integration with PCSWMM modelling tools,
- Representation of geospatial data for Runoff, Transport and Extran in separate themes,
- Support for common map layer formats, including ArcInfo/ArcView™ shape files, Autocad™ DXF and DWG files, MapInfo™, and Microstation™, as well as raster image support for TIFF, JPG, BMP,
- Linking to ODBC databases with SQL query support,
- Importing from existing SWMM Runoff, Transport or Extran input files,
- Interactive of Runoff, Transport, or Extran input data files,
- Graphical editing of nodes/conduits in any SWMM layer (includes adding, deleting, editing and converting between SWMM layers),
- Path and element selection for display in other PCSWMM tools (e.g. Dynamic Hydraulic Gradeline),
- Automatic or manual reduction of model complexity with the Aggregation wizard,
- Intelligent connectivity checking and reporting,
- Complete customization of SWMM layers, and
- Easy learning curve with extensive online help, Getting Started video tutorials and web support.

In more detail, PCSWMM is designed as a standalone tool, and does not drive the GIS application using ESRI or MapInfo's VB/VC++ components. The model interfaces directly with the underlying database(s). ESRI, MapInfo and AutoCAD layers/themes can all be displayed. Look and feel is most similar to ArcView (although MapObjects were not employed). Data is extracted directly from the GIS's underlying database (MS Access or ODBC compliant) using SQL queries and setup in PCSWMM's internal database (MS Access) for 'data tweaking' into a suitable model (i.e. element aggregation, etc.). External access to the PCSWMM internal database is also possible.

Output is tightly linked with PCSWMM's array of tools. For example, profiles can be selected graphically within the PCSWMM view and displayed in the Dynamic HGL tool (with simulation playback). The background can be a *.bmp or meta-file. Completely accurate scaling is possible with a geo-reference file. Many tiled or layered images may be loaded. Supported raster file formats include: .TIF, .BMP, .JPG, .PCX, .DIB, and .TGA. Other (vector) layer formats supported are: MapInfo MIF, MicroStation DGN, AutoCAD DXF, AutoCAD DWG, and ESRI Shape File. However, PCSWMM does not render in 3D (no fly-thoughs or 3D points of view). The view is restricted to a plan view. Vertical detail is stored in the Extran layer, and PCSWMM supports real world coordinates (and any other coordinate system). PCSWMM can be used without a GIS/IMS as a graphical plan view editor (with or without backgrounds) for SWMM or PCSWMM.

The interface can be used to automate data input such as watershed areas, land use, and soils related parameters such as infiltration rates. Node, conduit, and subcatchment data is pulled from the underlying GIS database (ODBC) into an intermediate database for processing into a useful model. This data is then exported to a SWMM input file.

Apart from the Aggregation wizard, which calculates equivalent parameters when aggregating two or more conduits to a single conduits, PCSWMM does not calculate/estimate any parameters. To update SWMM input data files, users should associate an input file with a SWMM layer (Runoff, Transport or Extran) and click on the 'Export to Input File' menu item. The selected elements are then written to the input file.

Basically PCSWMM allows the user to pull in entities and attributes from a GIS database, edit the data (subcatchment, node, conduit) to build a model and update an existing SWMM input file with this data. All other aspects of modeling (file management, input data file development, running the model, output visualization and interpretation, sensitivity, calibration and error analysis, storm dynamics analysis, and much more) are handled by PCSWMM.

PCSWMM was written for the web. It is a decision support system for the SWMM, providing a large array of file management, data file creation, output interpretation, model calibration, and reference tools for the stormwater modeler. Users can develop their own in-house modeling environments using the extensive selection of plug-in tools and PCSWMM's seamless integration of any external processes (programs, batchfiles, macros, etc.). It is distributed with the latest version of SWMM4 from Oregon State University, as well as the official EPA SWMM5 release.

Advanced file management methodology has been enhanced for use with data distributed over local area networks (LAN's), Intranets and the Internet and fully supports long file names. Using a similar paradigm as previous versions, it simplifies the organization, connectivity and manipulation of the many files generated and used by SWMM. Groups of related files are represented by data-aware objects. The image of the object depicts the type of files it represents.

These objects can be arranged logically in folders, connected to other objects, and cut/copy and pasted between folders. Project information management is provided by means of note objects which can store text in ASCII, HTML or any document format. File extensions are handled automatically by PCSWMM, and can be specified by the user. Existing SWMM files can be easily imported into PCSWMM, and/or new data files can be created from user specified template files.

Data file structures have been simplified by eliminating the need for the Executive module command lines from all input data files. Instead, data file connectivity is displayed graphically in the PCSWMM window, and can be modified at any point with a simple mouse click. This eliminates the confusion and headaches associated with misplaced interface files, wrong scratch file designations, and long text-oriented commands and input.

Object oriented runs: SWMM runs are executed by selecting either a single or groups of objects and choosing the Run SWMM tool. The SWMM engine can be either run in a window or full-screen and changed between the two engines provided (or any other SWMM compilation) on the fly. Options are provided for the type of SWMM output generated.

Help: An extensive, online, indexed help file that provides a variety of information. Step-by-step instructions are available for all PCSWMM procedures. Known trouble spots are highlighted for quick explanations. PCSWMM also provides complete online help for all aspects of SWMM data file development for any module of SWMM. Up-to-the-minute help is available through the WWW and email.

Flexible tools registry: Tools take the form of external processes that can be plugged-in with the Tools Registry. This enables a completely flexible modelling environment created from third-party and in-house tools, as well as the tools bundled with PCSWMM. Its purpose is to encourage students and other users to contribute their own routines.

Edit is a 32bit ASCII text editor for SWMM files offers fast load times for input or output files, no matter what length. It also provides context sensitive links to our SWMM online help file using the F1 key. Its purpose is to encourage very long-term, continuous simulations.

OpenURL enables access to email, web pages, ftp, and telnet from either Internet Shortcut objects, the Tools menu, or any URL address found in any object file. The most recent version of SWMM may be downloaded with a mouse click, an important attribute for instructors who teach such courses semester after semester. The tool may include links to related information or data sources on local intranet or the internet in users' input data files or note files, and/or provide easy modeling coordination with remote teams. This tool opens up the universe of web resources to students from their executing PCSWMM desktop. It requires a WWW browser and internet access (can be dial-up). OpenURL comes with a help file for instruction on usage.

Dynamic HGL, as shown in Figure 8.3, displays the hydraulic gradeline for any sewer profile in the EXTRAN output. Both dynamic simulations and static displays can be viewed at any time step. The program provides surcharge warning, and can search for surcharge events. Simulation summary tables can be viewed for the selected profile, and profiles can be easily switched or zoomed in on, even during simulation playback. Pump stations, storage facilities, weirs and orifices are supported.

Graph creates fully customizable hydrographs and pollutographs directly from any SWMM binary interface file as well as from ASCII text plot files. Continuous simulation output is easily and quickly analyzed with unlimited zoom and pan in both the X and Y directions, and support for millions of datapoints. Quality constituents, velocity and depth values can also be graphed, with simultaneous support for up to six Y-axes. The customized, high-resolution plots can be printed, copied to the clipboard and/or saved to file in a variety of graphic and ASCII text formats.

Run SWMM facilitates the running of any SWMM4/5 or newer compilation. Run SWMM automatically creates batch runs for multiple files and takes care of the Executive module command lines for you (file connectivity, module parameters and scratch file designations), reducing the chance of user errors. All aspects of the SWMM engine are supported, including Extran hot-start files, automatic collation of two upstream interface files, and WASP output. Command line switches provide options for the generation of plot files (ASCII text versions of interface files), as well as for invoking any PCSWMM plug-in tool at the conclusion of the SWMM run(s). Run SWMM comes with a help file for instruction on usage.

RainPak can be used to:

- access user-defined datalogger files (recorded by exception - no loss of detail),
- compute speeds and directions for individual storm cells,
- generate statistics on the computed relative frequency of storm cell direction, and
- facilitate the analysis of the distribution of storm cell speeds
- generate discretized timeseries (at any time step from 1 to 60 minutes) from an array of raingauges for pasting into the Runoff module.

RainPak consists of four distinct modules: Import, Velocity, Collate and Direction. RainPak Import provides the capability for: identifying storm systems and storm cells, interpolating the various precipitation time series to a fixed, uniform time-step, and saving the interpolated precipitation time-series to a single file for use by the RainPak Velocity module. RainPak Velocity provides both visual and computational methods for determining storm cell kinematics. RainPak Collate facilitates the collation of various combinations of storm cell analysis results into a single comma-delimited file for import into spreadsheets and/or the RainPak Direction module. Finally, RainPak Direction creates rosette plots of relative frequencies of storm cell directions. Confidence in the computed storm cell velocities can be evaluated from statistical error estimates included in the software and/or by subjective analysis of the plotted vector arrays and comparison to the dynamic rain intensity bar plots generated by RainPak Velocity.

Biblio provides instant access to reference information on 4500 papers given at SWMM users meetings and conferences. The database includes over 3500 full searchable abstracts, and two powerful search engines, and an updated user interface.

Sensitivity Wizard is a powerful tool for gaining insight to the dominant processes of a particular model, and how they can change through an array of input functions (i.e. different rainfall intensities and durations). Virtually any number of Runoff parameters can be tested automatically, reducing a virtually unmanageable task to a few mouse clicks and a couple sips of coffee. Each chosen parameter is tested at five positions over user-specified ranges of uncertainty, the SWMM runs are automatically batched and results compiled from the output files. Sensitivity results are presented in a variety of ways, including both non-linear sensitivity gradient and ranked mean sensitivity gradient plots, and can be easily customized and printed or copied into reports at high resolution.

Calibration Wizard generates observed vs computed plots, complete with an array of evaluation functions, for any objective function. The program reads in a user-generated list of observed/computed data points, and the resulting plots can be customized and printed/copied into reports.

Tutorials are given on-line as videos.

Manuals that are meticulously indexed and edited are provided to students (James et al., 2005).

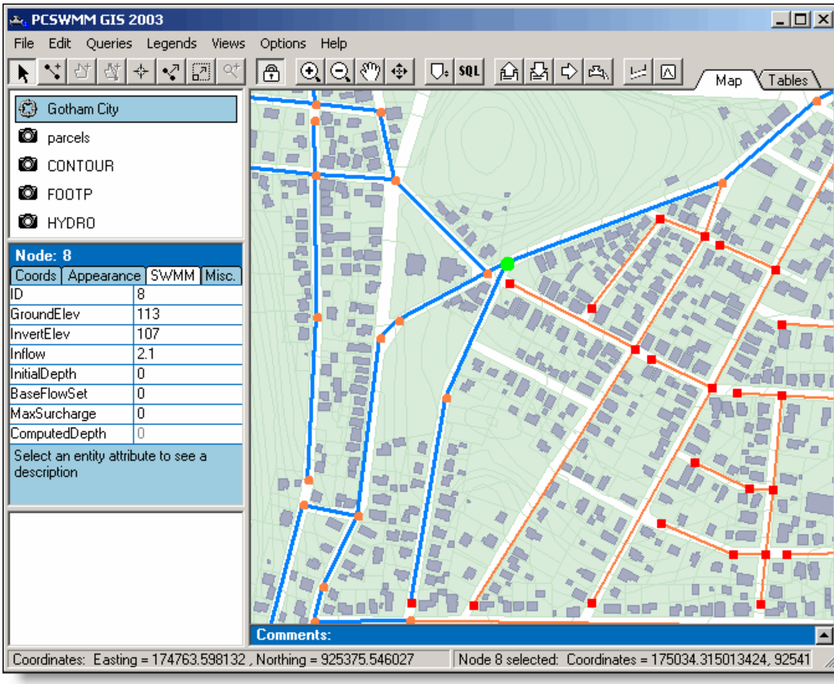


Figure 8.2: Screen dump of the GIS utility.

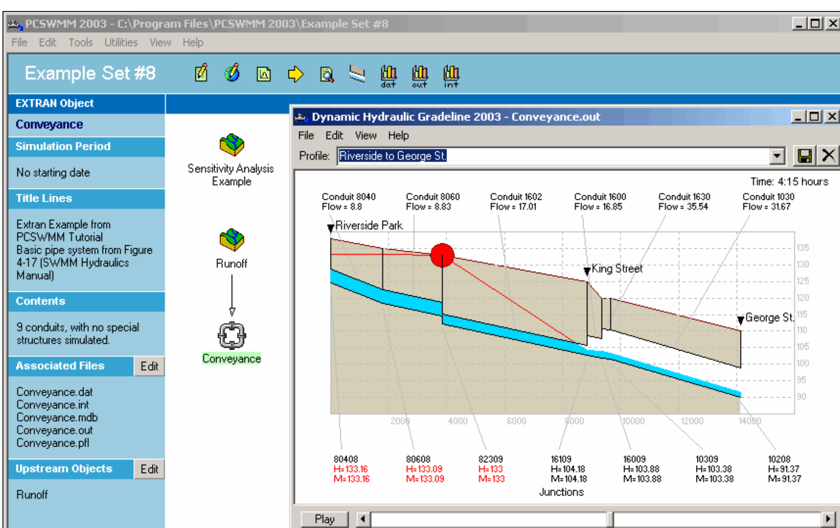


Figure 8.3: Screen dump of the dynamic hydraulic grade line animation.

8.3 Concluding remarks

Figure 8.1 shows the relationship between the various components: short-term calibration input functions (IFs); model; response functions (RFs); objective functions (OFs); performance evaluation functions (EFs); sensitivity analysis; parameter optimization; long-term continuous input functions; error analysis; long-term, continuous "fuzzy" response functions; and output interpretation or inference. Figure 8.1 is meant to be schematic and conceptual, and only to show the average sequence of activities in the broadest of terms.

It is clear that the modeling activities shown in Figure 8.1 will benefit from a well-written DSS. More detailed discussion on each of these activities is the principal subject of the rest of this book.

Based on the experience gained in a generation of development and use of earlier shells, and their application at over 50 professional workshops, as well as in graduate and undergraduate credit courses over that time, a wholly new shell for the U.S. EPA Stormwater Management Model SWMM was developed. The new shell was designed as a web-oriented tool to facilitate the teaching and learning of SWMM in a series of international seminars and workshops, as well as in graduate and undergraduate courses at the University of Guelph. Subsequent experience helped design a system that speeds classroom learning times, and maximizes use of available resources, by eliminating the most-expected errors. For example, the shell removes the error-prone executive module data lines that are normally required in the SWMM input data files, using intelligent pull-down menus instead, and provides fast graphics. Additionally, the shell uses full-page editors and a help menu to access SWMM example and .DOC files. Classroom use of PCSWMM has demonstrated significant improvement in learning times, and reductions in the most prevalent run-time errors.

Continuous modeling requires a sequence of two main sets of modeling activities:

1. *calibration or parameter optimization,* and
2. *inference or design*:

calibration activities involve parameter estimation and optimization against short-term, accurate, observed input functions;
inference activities involve long-term, continuous, synthetic or transposed input functions, and error analysis.

References

Adderley, V.C. and Liebe, M.A. 1993. Development of a SWMM graphical interface for Portland CSO models. Preprint for 1993 Runoff Quantitiy and Quality Modeling conference. ASCE Reno, Nev. Nov 1993.

Baffaut, C. 1988. Knowledge techniques for the analysis of urban runoff using SWMM. Ph.D. thesis. Purdue U. 321 pp.

Baffaut, C. and Delleur, J.W. 1989. Expert system for calibrating SWMM. J. Water Resources Planning and Management. 115(3): 278-298.

Baffaut, C. and Delleur, J.W. 1990. Calibration of SWMM runoff quality model with expert system. J. Water Resources Planning and Management. 116(2): 247-261.

Brown, L.C. 1985. Uncertainty analysis using QUAL2E. Proc. Stormwater and water quality model users group meeting, Dec 5-6 1985 pp 125-140.

Brown, L.C. and Barnwell, T.O. 1987. The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS: documentation and user manual. U.S.EPA Report EPA/600/3-87/007. 199pp.

Dennison, K.D. 1985. A Database Environment and Sensitivity Framework for Continuous Water Quality Modelling", M.Eng. Mc Master U. Oct 1985, 292 pp.

Dunn, A.R., 1986. Feasibility of a Real-Time Auto Sensitivity and Error Analysis Framework for Fuzzy Stormwater Modelling on Personal Computers, M.Eng. Thesis, McMaster U., May 1986, 130pp.

Dunn, A.R. and James, W., 1985. Extended Expert Systems. Proc. of the Stormwater and Water Quality Modelling Conf., sponsored by USEPA and Ontario Ministry of the Environment, Toronto, Canada, published by CHI, Hamilton, pp. 373-380, Dec 5-6, 1985

\Griffin, S., Bauwens, W, and Ahmad, K. 1993. UDMIA: urban drainage modelling intelligent assistant. 6th Int. Conf. Urban Storm Drainage, Niagara Falls, Ontario. Sept. 1993. pp 1314-1319.

James, W. and James, W.R.C. 1993. PCSWMM4 - a public-domain, graphics-oriented shell for easing the design-use of SWMM4. Runoff Quantity and Quality Model Group Conference, Sponsored by ASCE . Reno NV, Nov 8-9 1993.

James, W., Huber, W.C., Rossman, L.A., Dickinson, R.E., Pitt, R.E., James, W. R.C. , Roesner, L.A. and J.A. Aldrich, 2005. User's guide to SWMM. CHI pubs. ISBN 0-9736716-1-0. Guelph. 942 pp.

Kuch, A. J. 1994. A sensitivity analysis shell for complex continuous water quality management modelling. M.Sc. Thesis. Univ of Guelph. ca.120 pp. January 1994.

Kuch, A. and James, W. 1993. A sensitivity analysis tool for three generation continuous urban storm water quality management modelling. Runoff Quantity and Quality Model Group Conference, Sponsored by ASCE . Reno NV, Nov 8-9 1993. *in press.*

LaVeck, G. 1993. Development of a WINDOWS-based interface to SWMM. Preprint for Conf. on runoff quantity and quality modelling. ASCE. Reno, Nev. Nov. 1993. 7pp.

Leavesley, G.H., Lichty, R.W., Troutman, B.M. and Saindon, L.G. 1983. Precipitation-runoff modelling system user's manual. Water Resources Investigations 83-4238. USGS. Denver Colo. 80225. ca 207 pp.

Leavesley, G.H., Restrepo, P.J., Stannard, L.G., Frankoski, L.A. and Sautins, A.M 1993. The modular modelling system (MMS) - a modelling framework for multidisciplinary research and operational applications. Preprint for Conf. on GIS and Environmental Modelling. Breckenridge, Colo. Sept 1993 . 9pp.

Liong, S.Y. , Chan, W.T and Lum, L.H. 1990. Knowledge-based system for SWMM runoff component calibration. J. Water Resources Planning and Management. 117(5): 507-524.

Lumb, A.M. and Kittle, J.L. 1983. ANNIE - an interactive processor for hydrologic modeling. USGS Water resources Investigation report 83- ? 133pp.

Lundgren, P.S. and Barber, M.E. 1993. Site-specific application of SWMM using expert system. Preprint 1993 Runoff quantity and quality modeling conf. ASCE. Reno, Nev. Nov 1993. 8pp.

Orlob, G.T. 1992. Water-quality modelling for decision making. J. Water Resources Planning and Management. 118(3): 295-307.

Rauf, A. and Chaffey, D. 1993. A new interactive graphics interface for urban drainage applications. 6th Int. Conf. on Urban Stormwater Drainage. Niagara Falls. Nov. 1993. pp 1345-1350

Reboh, R.J., Reiter, R.J. and Gashnig, J. 1982. Development of a knowledge-based interface to a hydrologic simulation program. SRI International, Menlo Park, CA.

Schmidt, M.F., TenBroek, M., Aldrich, J., and Plante, S.V. 1993. CDM: Stormwater system planner. Preprint for 1993 Runoff Quantity and Quality conf. ASCE. Reno, Nev. Nov. 1993. 10pp.

Thompson, G.R., Bogiatzis, T., Phillips, B.C. and Dickinson, R.E. 1993. XP: a graphical decision support system for water resources models. 6th Int. Conf. Urban Stormwater Drainage. Niagara Falls, Sept. 1993. pp 1351-1356.

Uber, J.G., Brill, E.B., and Pfeffer, J.T. 1992. Use of mathematical programming methods for complex systems. J. Water Resources Planning and Management. 118(3): 281-294.

Unal, A., **1985.** "Centralised Time Series Management for Continuous Hydrology on Personal Microcomputer Networks", Ph.D. Thesis, McMaster U., Dec 1985, 185 pp.

Walker, W.W. 1982a. A sensitivity and error analysis framework for lake eutrophication modelling. Water Resources Bulletin. 18(1):53-60.

Walker, W.W. 1982b. MODELER a computer modeling and data management system. Ver 4.0 Mar. 1982. W.W.Walker, Concord, MA. 21pp.

Wu 1993. WINDOWS-SWMM Conf on Water quantity and quality models, ASCE. Reno, Nev.

Bibliography

Crawford, N. 1993. The next major breakthrough... Paper presented at the U.S.EPA workshop on Developing dynamic watershed modelling capabilities for the Great Lakes tributaries. Heidelberg College, Tiffin, Oh. August 4-6, 1993. 8pp.

Griffin, S., Bauwens, W, and Ahmad, K. 1993. UDMIA: urban drainage modelling intelligent assistant. 6th Int. Conf. Urban Storm Drainage, Niagara Falls, Ontario. Sept. 1993. pp 1314-1319.

Savic, D.A. 1990. An intelligent decision support system and use of fuzzy sets for reservoir analysis. Ph.D. thesis. U. of Manitoba. 253 pp.

Sircar, J.K. 1986. Computer Aided Watershed Segmentation for Spatially Distributed Hydrologic Modelling. Ph.D., Univ. of Maryland,

Stanley and Assocs, Undated. (1993?) RTSWMM 4-page pamphlet, pub by Hydrotechnical services group, Edmonton, Alberta.

Thompson, D.B. and Westphal, J.A. 1989. Tilt at the windmill of runoff simulation. Proc of the Int Conf on catchment runoff, U of Virginia, Charlottesville. 1989. pp176-185.

Von Euw, E. 1990. A comprehensive comparison of three non-point-source pollution models on an agricultural watershed in Southern Ontario. M.Sc. Thesis. U. of Guelph. 134 pp.

Editor's Note: Pages 131-138 are omitted from this edition.

Chapter 9

OBJECTIVE FUNCTIONS

In the Middle Ages, subject meant substance, and has this sense in Descartes and Spinoza: sometimes, also, in Reid. Subjective is used by William of Occam to denote that which exists independent of mind; objective, what is formed by the mind. This shows what is meant by realitas objectiva in Descartes. Kant and Fichte have inverted the meanings. Subject, with them, is the mind which knows; object, that which is known; subjective, the varying conditions of the knowing mind; objective, that which is in the constant nature of the thing known.

-Trendelenburg.

Objective means that which belongs to, or proceeds from, the object known, and not from the subject knowing, and thus denotes what is real, in opposition to that which is ideal -- what exists in nature, in contrast to what exists merely in the thought of the individual.

- W. Hamilton.

Objective has come to mean that which has independent existence or authority, apart from our experience or thought. Thus, moral law is said to have objective authority, that is, authority belonging to itself, and not drawn from anything in our nature.

-Calderwood (Fleming's Vocabulary).

9.1 Introduction

In this chapter the various driving input hydro-meteorological TS for a WQM are termed *input functions*, or input TS. Typical examples are rainfall, evapo-transpiration, wind speed, wind direction, snowfall, radiation, humidity, temperature, and some pollutant generating mechanisms such as atmospheric fallout, and traffic. Other input data that are constants and control independent component processes, are termed *parameters*. It is sometimes not clear how to distinguish between input variables and environmental parameters. Parameters are less likely to be in the form of TS. They are generally *coefficients*, and generally remain invariable through all or part of the simulation run.

The computed hydrological and water quality TS output by the WQM, is termed the *response function*. Programs such as HSPF and SWMM may be fed a large number of very long input functions (IFs), and can then provide very many response functions (RFs), e.g. pollutographs (pollutant concentration), loadographs (pollutant flux), hydrographs, water levels and velocities, cost estimates, and capacities and geometries of storages and conveyances. Response functions can be computed for many locations

An *objective function* is a statistic or a representative number derived from the response function, chosen to dovetail with the design objectives. Post-processors such as the STATISTICS module in SWMM and the DURANL module in HSPF compute several *objective functions* (OFs), e.g. event peaks, event means, numbers and durations of exceedances and deficits, etc, for most of these input and response functions, for both water quantity and quality. There are of course also similar objective functions for the equivalent measured or observed TS.

An *evaluation function* seeks to compare computed and observed objective functions. Evaluation functions are carefully chosen to evaluate the performance of the model.

9.2 Another question of terminology

Readers are warned again that our term "objective function" is used most often in the scientific literature to denote what we term the "evaluation function". Our objective function (OF) should not be confused with our *evaluation function* (EF), which is used in this book exclusively to denote the measure of agreement between the observed OFs (OOFs) and the computed OFs (COFs). Objective functions here are taken to be the primary objective of a water quality simulation run, and that is why we have taken such trouble to relate them to the principal design objectives. Calibration *per se* is not the principal objective of modeling in this book. Calibration is part of the validation process, and the functions chosen for this purpose are therefore better termed performance evaluation functions. An example of an EF is the sum of the squares of the deviations between the observed event peak flows and the computed event peak flows in a calibration plot. Efs, validation and verification are discussed later.

9.3 Response functions and statistical objective functions in SWMM

The response functions computed by SWMM include both complete and summary hydrographs and pollutographs, placed near the end of each output file. These summaries for selected nodes include certain TS statistics such as flow-weighted-average, standard deviation, maximum and minimum flow rates and total volume of runoff for the hydrographs and flow weighted average, standard deviation, maximum, minimum concentrations and total load for pollutants and constituents. Such statistical derivatives from the computed response functions are called objective functions herein, because they are objectively related to the computed response function - they do not depend on the error when compared to an observed record. Table 9.1 is a sample output file summary.

Table 9.1 Sample RUNOFF output summary

	Flow (cfs)	TOT SOL MG/L	COD MG/L	TOT NIT MG/L
Flow wtd means	0.0242	7.851E+ 02	4.906 + 01	7.633 - 01
Flow wtd std. devs	0.1715	8.757 + 02	1.223 + 01	1.447 - 01
Maximum value	2.456	1.701 + 04	6.106E + 01	9.124 - 01
Minimum value	0.000	0.0000E - 01	0.0000E - 01	0.0000E - 01
Total load	5.318+03	2.606 + 02	1.629E + 01	2.534E - 01
	Cub -ft	Pounds	Pounds	Pounds

Except for the flow-weighted standard deviation, the above output objective functions are often used, the four most common being the peak and total (flows and pollutants). These eight objective function types (four each for hydrographs and pollutographs) and many others are user-selectable in PCSWMM.

Figure 9.1 illustrates the short list of simple OFs given in Table 9.2, by depicting a typical cycle of the input and response function (IF and RF) TS. The literature stresses RFs, but similar OFs apply to all input hydrometeorological TS (e.g. temperature, wind speed).

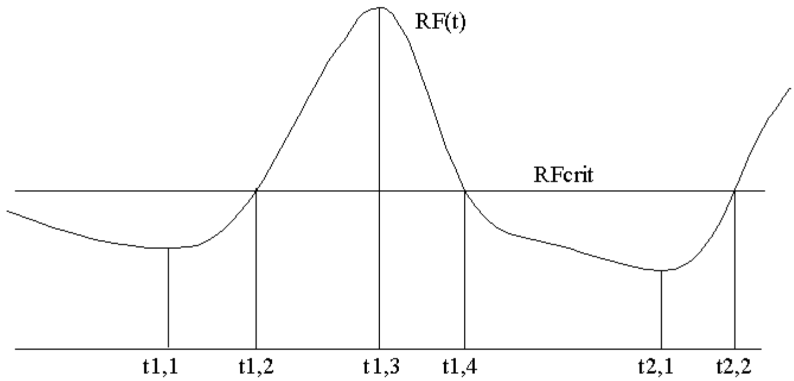


Figure 9.1 Typical cycle in a response or input function.

The functions may be observed, synthetic or computed. RFcrit and IFcrit are arbitrary.

Table 9.2 Short list of some simple response OFs.
[* the INT functions are defined below; note that the OFs and time variable are not subscripted in the table, to save space.]

OF1	(t2,1 – t1,1)	duration of wet event
OF2	(t2,2 – t1,3)	duration of dry event
OF3	RF(t1,3)	peak flow, flux, or concentration
OF4	RF(t1,1)	minimum flow, flux or concentration
OF5	*INT(t1,4 – t1,1)	total wet event flow or flux
OF6	(t1,4 – t1,2)	duration of exceedenace
OF7	(t2,2 – t1,4)	duration of deficit
OF8	N[RF>RF crit]	number of exceedances
OF9	N[RF<,RF crit]	number of deficits
OF10	*INT (t2,2 - t1,4)	volume of deficit
OF11	*INT (t1,4 – t1,2)	volume of excess
OF12	OF5/OF1	wet event mean concentration
OF13	*INT (t2,1 – t1,4)	total dry event flow or flux
OF14	OF13/OF2	dry event mean concentration

The four OFs in Table 9.2 that require integration are:

1 $$OF_5 = \int_{t_{1,1}}^{t_{1,4}} RF(t)dt$$

2 $$OF_{10} = \int_{t_{1,4}}^{t_{2,2}} \left[RF_{crit} - RF(t)\right]dt$$

3 $$OF_{11} = \int_{t_{1,2}}^{t_{1,4}} \left[RF(t) - RF_{crit}\right]dt$$

4 $$OF_{13} = \int_{t_{1,4}}^{t_{2,1}} RF(t)dt$$

In Table 9.2 only fourteen simple OFs are shown, yet even for only (say) sixteen pollutants, the total number of possible OFs for response functions alone (i.e. ignoring IFs) would be 238 per location (of which there may be several thousand, so that for 4000 locations SWMM can generate about one million RFs, a point not generally realized).

9.4 Multi-objective functions

Yan (1990) believes that the parameter set that is optimal with respect to one objective function is quite likely to be poor with respect to another OF. He therefore developed two parameter-estimation methodologies for multipurpose rainfall-runoff models, using PRMS:

1. The multi-objective function (MOF) is the product of the error sum of squares of objectives and weighting coefficients, and
2. The MOF is the Bayesian determinant with weighted errors instead of original errors.

The objective functions used were: event peak flow, event flow volume, and daily discharge. Only four parameters were estimated. Rosenbrock's optimiz-ation scheme was used for finding the parameter estimates that produced the minimum value of the MOF.

The best set of parameter estimates was the set that yielded relatively small error sum of squares over all objectives. The result is that the parameter estimates are unlikely to be both 1. very good with respect to one objective, and 2. quite poor with respect to another objective. It is important to realize that not all OFs are relevant to all parameters, or design questions. In fact the applicability is quite restricted. In matching them, Table 9.3 (refers to just the OFs listed in Table 9.2) may be used as a first guide.

9.5 Concluding remarks

Careful selection of the best OF is necessary, a point that is perhaps seldom stressed enough. Users must themselves thoughtfully choose both the RFs and the OFs that most closely relate to the design questions to be answered. Also, the number of OFs should be minimized, if the amount of computed output TS is to be kept within manageable limits. Some guidance is provided in Table 9.3 below. In the literature, some researchers have reported that it is difficult to calibrate certain WQMs to more than one OF (Yan, 1990; Seo, 1991). Note that this has not been the present writer's experience, however.

Depending on the application, the choice of OF may predicate which of the model parameters are eventually optimized, and their absolute values.

Table 9.3 Dominant processes and their potentially relevant objective functions.

Dominant process	Objective function
overland flow over impervious area	OF3
infiltration into the upper soil mantle	OF4
pollutant washoff	OF5
erosion	OF1
overland flow over pervious area	OF3
pollutant build-up	OF5
recovery of storages	OF2
recovery of loss (infiltration) rates	OF4
recession of storages	OF7
evaporation	*IF8
snowmelt	*IF11
Snow accumulation	*IF7

*Note: Some WQMs such as SWMM can produce event analysis of input TS, so that in this Table, "IF" refers to *input function* (in this case, air temperature and/or wind speed); the OFs are not subscripted to save space in the table.

Select the best objective function thoughtfully, by relating it back to the original design questions. Use the minimum acceptable number of objective functions.

References

Yan, J. 1990. Parameter estimation for multipurpose hydrologic models. Ph.D. thesis. Oklahoma State U. 184 pp.

Seo, Dong-il. 1991. Optimal complexity analysis of total phosphorus models. Ph.D. thesis. U. of Michigan Ann Arbor. 173 pp.

Bibliography

Diskin, M.H. and Simon, E. 1977 A procedure for the selection of objective functions for hydrologic simulation models. J. of Hydrology, 9:34-56.

Karlinger, M.R. and Moss, M.E., April **1973**. Regression-Analysis Simulation for Evaluation of Hydrologic-Parameter Accuracy. Unpublished manuscript, 13 pp.

Thomann, R.V., **1982.** Verification of water quality models, *Proc. Am. Soc. Civ. Eng. J. Environ. Eng. Div., 108*(EE5), 923-940,

Yen, B.C. 1993. Reliability and uncertainty analyses in hydraulic design. A sub-committee report. ASCE.

This page intentionally left blank.

Chapter 10

UNCERTAINTY ANALYSIS

> *A greater acknowledgment of model uncertainty often has the consequence of widening our uncertainty bands…. Since hedging against uncertainty is hard work, this is an unpopular turn of events, at least in the short run.* – Draper, 1995.

10.1 Introduction

In this chapter we limit discussion of uncertainty analysis to counting probable sources of error. Discussion of the various sources and kinds of model error may seem to be unnecessary, since uncertainty is so pervasive in engineering design that it is taught in the very first undergraduate course (Andrews and Ratz, 1997). Nevertheless uncertainty estimation is seldom done.

Measurement itself means a comparison against a standard and is always accompanied by error. Whereas *error* is the difference between a measured (or computed) value and the *true* value, and has a specific value for each incident, *uncertainty* is slightly different, being the *possible error*. Error is usually the difference expressed as a percentage of the expected value, (e.g.+10.9%), and uncertainty as a range of the error, e.g. –15% to +23%. It follows that the error falls within the uncertainty. Readers need to be aware that some writers report uncertainty as the computed value itself (rather than a difference) expressed as a percentage e.g. 85% to 123% in the last case.

Model uncertainty is the comparison or difference between a value taken to be true and the estimated value that could be computed based on an assessment of the integrated effects of all possible sources of error. Model uncertainty is thus partly subjective. Notice that it does not include *mistakes*, although bugs in the model code may well produce model-structure uncertainty.

Make no mistake about the importance of rainfall data in your modeling (Melching, 1987). Most forecast uncertainty is cause by uncertainties in rainfall. The second largest source is in the estimation of rainfall excess when subtracting infiltration. The reliability of those model responses that are computed by using just a few isolated rain gages should always be questioned, especially for water quality modeling (Chaubey et al., 1999). Parameters that we analyze for sensitivity and then optimize probably contribute the minor part of the uncertainty. Better models and model procedures are warranted when the input rain data is detailed and has reasonably small uncertainty.

Beck (1987) assumes that the following characterize an accurate model:

1. Errors are small in magnitude, and not attributable to any causal mechanisms of a significantly non-random character.
2. Model error variances and covariances of the model parameters are low.
3. Parameters are demonstrably invariant with time.

> Model uncertainty is the difference between a value taken to be true and the estimated value that could be computed based on an assessment of the integrated effects of all possible sources of error.

WQMs are subject to considerable uncertainty, the impacts of which are often just as significant as the model results themselves. Uncertainty should always be considered to be as important as the expected result. Rationally, when using a model in design, the design decisions should be open to evidence that the model and design were, perhaps by some evolving criteria, wrong. To be fully useful, a computed result should have an uncertainty assessment attached to it, or the design implementation may have too much or too little hedging against the actual uncertainty (Draper, 1995). Regrettably, uncertainty analysis seldom forms an important part of an engineering design that involves WQMs (Dilks,1987).

Comprehensive uncertainty analyses take effort and are costly. For one thing, data needs for characterizing input uncertainties are problematical. So it is important to be clear about the benefits of uncertainty estimation. The purposes of model uncertainty analysis are to answer the following questions:

1. What are the principal sources of uncertainty in the computed response?
2. How can the uncertainties be managed?
3. How confident may we be in the computed response for the *as-is*, *as-was*, and *to-be* scenarios?

Beck (1987), citing others, classifies sources of uncertainty into three categories:

1. Uncertainty resulting from model construction, composed of model aggregation and model structure, both of which are covered by the errors of a priori assumptions about the internal description of the model system.
2. Uncertainty resulting from parameter optimization errors.
3. Uncertainty resulting from

(a) the natural variability in the real system and to natural disturbances in the input parameters,
(b) spatial heterogeneity, which is absorbed into error associated with aggregation or lumping, and
(c) genetic variability, which is in practice indistinguishable from the errors of parameter estimation.

So far as the first class (model-structure uncertainty) is concerned, in a design-office application a model that may be said to be wrong or inherently unsuitable, should be replaced, or enhanced, or rebuilt. Recourse should be made to the best available program, and a reasonable modeler should be open-minded about model selection, using several models, so that the most appropriate model for each of the various design problems would be used. We discuss later how model uncertainty due to model structure may be uncovered.

So far as the second (parameter optimization uncertainty) and third (natural uncertainty) classes are concerned, the recourse in either case is to calibrate the model appropriately, keeping the parameter values within their reasonable ranges, and correcting the observed input and output function data whenever warranted. Calibration is covered elsewhere in this book.

Willems (2000) also classifies the sources of uncertainty into three categories: 1. input uncertainties, 2. parameter uncertainties, and 3. model-structure uncertainties. Model input is time-variable, whereas model parameters (also part of the input file) are essentially constant, being measured, estimated or calibrated. Model structure is assumed to be fixed and not time-varying. Parameters that are estimated or measured are subject to observation uncertainty. Parameters that are calibrated are also subject to observation uncertainty in the calibration datasets, and for this reason calibration is usually sub-optimal. These ideas are illustrated in Figure 10.1.

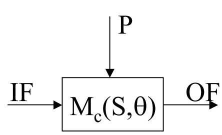


Figure 10.1 Terms used in model uncertainty.

IF = input function (time-variable), OF = output function
p = parameter set; M = model, C denotes various complexities,
S = structural-uncertainty sources, θ = parameter uncertainty sources
P + IF = input data file.

Model M has known input time series IF and input parameter set p, and produces unknown time series OF. Each parameter set is particular to the model complexity c. Both parts of the input data file, IF and p are subject to more or less uncertainty. M has two components parts that produce uncertainty, one is the model-structure S and the other is parameter uncertainty θ. (One can think of them this way: structural uncertainty is the type of uncertainty that results when one has a bad model with correct parameters, and parameter uncertainty results from a correct model with wrong parameters.) The one best parameter set is denoted p*, and the best complexity c*. We assume that the model structure for this best condition is also the best model structure S*, and the best response, OF*. It follows of course that each time the complexity of the model is changed, the model should be re-calibrated.

Generalized likelihood uncertainty estimation procedures (Beven and Binley, 1992) arose because of the observation that more than one calibration dataset p may produce equally valid responses OF*. They argue that optimal calibration p* is only possible with an optimal model structure S*. For optimal calibration p*, the derived parameter set should be unique (there should not be an equally good but significantly different parameter set), the optimal parameter values should be reasonable, and the calibrated parameters should cause their sub-models to accurately describe the component physical processes. In many cases, optimal calibration p* of an existing model with fixed structure S may not be possible. Willems (2000) cites the case of Johnston and Pilgrim (1976) who spent two years searching for but failing to find a unique parameter set for a nine-parameter rainfall-runoff model.

Model structure uncertainty has to do with the model-builders' limited ability to describe the physical processes perfectly. Willems (2000) considers that model-structure uncertainties are the uncertainties that remain after use of error-free input and measurements, and after an optimal calibration. This is an important point that we exploit in this work, and return to later in this chapter.

It is not generally necessary to decompose structural uncertainty into the structural uncertainties of the constituent sub-models. When modifications to the code are being considered, uncertainty decomposition should be carried out to determine which sub-models require improvement.

Model-structure uncertainties are the uncertainties that remain after use of error-free input and measurements, and after optimal calibration.

10.2 Previous work on model uncertainty

For two WQMs in particular, PRMS (Leavesley et al., 1983) and MODELER (Walker, 1982a&b), DSSs were developed at the outset for managing model reliability, error and uncertainty. For another, QUAL2E-UNCAS (Brown 1985; Brown and Barnwell, 1987) code was written as a later add-on shell for the steady-state river water quality model QUAL2E. MODELER was written for a steady-state lake-eutrophication model. All three use sensitivity and first-order error analysis (S&EA). No such general S&EA shells exist for the full HSPF and SWMM4 packages, to the knowledge of the writer, although some partial expert systems codes were developed (Baffaut,1988; Dunn,1986; Liong et al.,1990; Kuch, 1994). Literature searches on MODELER and HSPF were not conducted by the present writer, but for PRMS a search revealed several publications and dissertations (Rivera-Santos,1990; Yan, 1990; Thompson and Westphal, 1989; Thompson, 1989; Rivera-Santos,1988; Brabets, 1987; Reed, 1986; Bower, 1985; Brendecke et al., 1985; Carey and Simon, 1984;).

On the other hand, there is a considerable literature on uncertainty in water quality modeling (see e.g.: Qaisi, 1985;Troutman, 1985a&b; Sorooshian and Gupta, 1985; Gardner et al., 1980; Kitandas and Bras, 1980; Wood, 1976).

Qaisi (1985) estimated total uncertainty *TU* as the variance of the model output about the observed data:

$$TU = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \frac{(Y_{ij}^{\,m} - Y_i)^2}{n_1 \cdot n_2}$$

where n_1 = number of observed data values
n_2 = number of model runs
Y_{ij}^{m} = model response for the jth run at location i
Y_i = mean value of the data at location i

Parameter uncertainty *PU* was defined as the variance of the model response about the mean value of the model output:

$$PU = \sum \sum \frac{(Y_{ij}^{\,m} - Y_i^{mb})^2}{n_1 \cdot n_2}$$

where Y_i^{mb} = mean value of the model output at location i

Structural uncertainty *SU* is defined as the difference between total and parameter uncertainty and can be simplified to:

$$SU = \sum \frac{(Y_i^{mb} - Y_i)^2}{n_1}$$

10.3 Sources of error

In practical design-office applications, the most frequent errors may be traced to wrong data in the input files caused by blunders and data entry errors, user/modeler misconceptions, and bugs in the code. Such uncertainties are not explicitly counted in uncertainty estimation.

But the most serious errors are probably those that are made well after the model runs are completed: poor interpretation of the results, and of their inherent error, and reliability, by model builders, users and decision-makers alike. (Assertion of a unique, single-valued response function is typical.) We return to this topic later.

The several sources of error associated with modeling should be kept under control during design applications. Errors most commonly dealt with in the literature exclude user shortcomings, but cover those below (I have decomposed model-structure uncertainty, the relationship is also shown schematically in Figure 10.2).

1. *observation* error, related to field instrumentation, comprising two components, one random and the other systematic;
2. *sampling* error, associated with the timing and location of field equipment;
3. *numerical* error, identified with the numerical mathematics used in the code;
4. *structural* error, related to *disaggregation* (the number and resolution of the processes active during the run);
5. *structural* error, related to *discretization* (the spatial resolution, sometimes called lumping in the input datafile);
6. *structural* error, related to poor *formulation* of one or more of the component process relations and code in the program;
7. *structural* error, related to poor connectivity between the component processes;
8. *propagated* error, related to erroneously estimated values of input parameters; and
9. *start-up* error, the error in the assumed initial values of all the model storages and state-variables in the run.

Observation and sampling errors, and the structural error due to discretization, are related to the errors associated with attempts to model the inherent variability of natural hydrological and ecological systems. Models require spatial averaging, even where the spatial heterogeneity in nature may be very large. The only way to deal with this difficulty is to increase the spatial resolution or discretization, and thus the amount of data collection, preparation and the computing. Increasing model complexity in this way obviously results in increased design office costs. The issue is covered elsewhere in this book.

Structural error due to disaggregation and poor component process models, is also called *framework error*. Propagated error is also known as parameter error, and is a focus of PCSWMM. Framework error is difficult to estimate, and is therefore generally taken to be the difference between the total model error and the parameter error. It is related to the residuals between the true and computed values after the model has been carefully calibrated. Since it is possible to calibrate a model with many parameters to perfectly match a small set of objective functions, the residual must be obtained for a much larger set of observed objective functions.

Table 10.1 lists some of the sources of model and parametric uncertainty associated with the formulation and the application of transport-transformation models Isukapalli, 1999). Figure 10.2 provides a sequential framework for the systematic evaluation of model uncertainty. Figure 10.2 may be compared with Table 10.1. Model users should evaluate their model by providing a written discussion of these uncertainties.

The goal of uncertainty analysis is to rank the contributions of individual parameters and input functions to each of the above sources of uncertainty.

When interpreting the computed output from your model, all sixteen sources of error listed in the framework for uncertainty analysis should be explicitly interpreted.

	External description	Internal description
Prior knowledge	1. uncertainty due to natural variability, or unobserved input disturbances. 2. measurement and sampling errors of observed input and output.	1. aggregation error 2. numerical error 3. structural error
Calibration process identify *as-is* model	3. start-up error 4. input TS datafile error 5. model error	4. discretization error 5. input environment datafile error 6. model structure and state-parameter error 7. parameter optimization error
Design process (inference to the *to-be* and *as-was* scenarios	6. uncertainty of to-be parameters 7. user output-interpretation error	8. parameter propagation error 9. error analysis

Figure 10.2 Sequential framework for systematic model uncertainty analysis. Modelers should interpret their computed output in terms of all of these sources of error.

Table 10.1 Examples of the sources of uncertainty in the formulation and application of transport-transformation models (Source: Isukapalli, 1999)

Uncertainty in Model Formulation (Structural Uncertainty)	Uncertainty in Model Application (Data/Parametric Uncertainty)
Simplifications in conceptual formulation	Constitutive parameter selection
Simplifications in mathematical formulation	Design/structural parameter selection
Ergodic-type hypotheses	Input data development/selection
Idealizations in constitutive relations	Source information
Independence hypotheses	Meteorology (air quality models)
Spatial averaging	Hydrology (water quality models)
Temporal averaging	Initial and boundary conditions
Process decoupling	Operational model evaluation
Lumping of parameters	Uncertainty in model estimates
Discretization	Uncertainty in observations
Numerical algorithm/operator splitting	Nonexistence of observations
Approximations in computer coding	Response interpretation
	ΔE caveat and other implicit hypotheses

10.4 Discretization and observation error

Because of the cost of collecting, monitoring, analyzing and abstracting data for a large number of parameters and variables, it seems desirable to reduce both the process disaggregation and spatial discretization to the minimum needed for sound design. Also, by determining which parameters should be assigned firm values *a priori*, or *zeroed* out of the model structure for the particular study, the model may be rendered simpler, or parsimonious.

Expert storm water management modelers are characterized by their ability to conduct innovative and fast model runs that search parameter sensitivity, search for the coarsest acceptable time step, and search for minimum discretization. Rules for discretization have been set out earlier. The following strategy is recommended:

1. Establish the required accuracy for the computed response function, bearing in mind its manipulation into the objective functions, and their often tenuous association with the original design questions.
2. Match this accuracy with estimated observation errors, as discussed below.
3. Develop a series of datafiles with increasingly coarse discretization and timestep, and run these sequentially comparing

the values of the objective functions. The computed objective function will begin to deviate from the results for a fine resolution, and at a certain critical coarse description, the deviation will begin to grow rapidly.

4. The best level of discretization and the best time-step are the coarsest values that produce a deviation that meets the required design accuracy. The best value may be found by plotting the computed objective function against the number of discretized sub-spaces, or size of the time-step, and interpolating.

Observation errors:

Uncertainties arise from measurement errors; and these in turn can involve

- i) random errors in equipment,
- ii) systematic biases due to imprecise calibration,
- iii) inaccuracies in the assumptions used to infer the quantity of interest from the observed surrogate or proxy variable,
- iv) small sample size and non-representative samples.

So far as observation errors are concerned, Linsley (1973) shows that streamflow measurements within 5% of actual flows are difficult to achieve. The U.S.G.S. (1985) defines

excellent flow measurements as those for which 95% of the [daily] discharges are within 5%;

good is defined as those having 95% within 10%; and

fair as being within 15%.

Because of :

1. the short reaches common in urban conduits,
2. the discontinuities caused by lateral junctions, etc,
3. the short sampling time interval required, typically one minute, and
4. the transient or rapid flow variations,

it follows that a flow accuracy of about 10% should be considered to be excellent for urban drainage.

Due to its spatial variability, precipitation errors are believed to be worse than stream flow. Usually a few isolated rain gages are used to indicate area-mean rain. Where rain gages are separated, the error may be extremely high. In summer convective storms, gages placed 1 km apart may incur infinite error, missing the rainstorm altogether. Since convective storms are so common and significant in urban runoff design and planning, claims of measurement accuracies of 10% at high rates of rain seem implausible.

Of all quantities measured in the field, water quality indices are most problematical. Sampling error is difficult to estimate, especially for micro-biology. The obvious remark to be made is that the sampling and measurement errors are probably commensurate with the accuracy of the water quality algorithms used in surface water quality process models.

Yoon (1994) describes how a model with sensitivity and uncertainty analyses can be used to 1. reduce input data uncertainty, to achieve a desired level of model uncertainty, and 2. determine an optimum field sample size for more accurate estimates of input parameter values.

> The best level of discretization and the best time-step are the coarsest values that produce a deviation that meets the required design accuracy.

10.5 Uncertainty due to user estimates of the limiting values of parameters

A somewhat abstruse point to be emphasized relates to the estimates of the limiting values of model parameters that are made by various users. Assume that three estimates are to be made for every parameter: the expectation, a lower limiting value and an upper limiting value. Quite widely reported in the literature is the suggestion that the three values be simply gleaned from a manual, guide or textbook. For example, the three values of Manning's n for range land suggested in the SWMM manual (James et al., 2005) are 0.130, 0.010, and 0.320 respectively. Is it wise to use these values?

Perhaps not. What after all is the purpose of the exercise? It is probably not to run the model for extreme values that appear in the literature, which were derived (in this case) for range land in radically different condition. Presumably the condition of the range land is known *a priori*, and a better, narrower estimate of the expected value can be made. Admittedly estimates by experienced users may differ from those of novices, nevertheless we are interested in estimates of the expected values by a large number of reasonably well-informed users, not the widest possible published range of values for extreme conditions.

User and model constitute an entity, in which user-quality is paramount. Indeed there is evidence to show that a variety of users of a single model can produce as wide a range of model results as can a single user using a variety of models. Similar comments, if not more so, can be made about the final design emanating from the model results.

The question to be addressed is: what is the range of predictions produced by many reasonable users of this model? Estimates of the expected values by, say, 12 users should be sought, and the range for the 4^{th} and 8^{th} highest reported (assuming a normal distribution, this coincides with two standard deviations). In this case the range of three values that might be found from such an enquiry would likely be much narrower, (say) 0.130, 010, and 0.150. Model users are an integral part of the process, their intelligent use of a model is informed by their experience, the model structure itself, the human interface, the on-line help, the documentation and the available literature and observed data.

In summary, the point being made here is that we are not so much interested in what extrema might conceivably be produced by model runs for various extreme conditions, as we are in the distribution of values produced by a range of reasonable, average users.

> Do not run your model for the extreme parameter values for radically different conditions that appear in the literature – use a range of values that a reasonable person would take for the

10.6 How to carry out an uncertainty analysis

The following steps are sometimes described in the literature on uncertainty analysis:

1. characterization of input uncertainties - estimation of uncertainties in model input functions and input parameters,
2. characterization of uncertainty propagation - estimation of the uncertainty in model outputs resulting from the uncertainty in model input functions and model input parameters,
3. characterization of model uncertainty - uncertainties associated with different model complexities, and
4. characterization of the uncertainties in model predictions - uncertainties resulting from the evaluation data.

Scavia (1980) describes three methods that are commonly used for uncertainty estimation: Monte Carlo methods, first-order uncertainty analysis and Kalman filters. Kalman filters are related to first-order uncertainty analysis, which uses sensitivity analysis. Hoybye (1998) states that sensitivity analysis and Monte Carlo methods are the two most cost-effective methods for improving accuracy and analyzing error structures in hydrologic models. He gives the following overview of methods for reducing uncertainty:

A. Input data.

1. Select measurements and methods that achieve a required accuracy in terms of bias and imprecision of the data to be used for input and calibration. Obviously the user requires prior knowledge of the bias and imprecision of all relevant instruments and methods.
2. Design to the required level of detail and accuracy the sampling program in terms of size, frequency, and location.

3. Identify by sensitivity analysis those inputs that contribute most to uncertainty, and allocate priorities to improve their determination. Parameters are ranked by sensitivity gradient.
4. Identify the same by Monte Carlo simulation rather than sensitivity analysis. Parameters are ranked by correlation coefficient.

B. Prediction uncertainty.

1. Predict a range of possible outcomes (typical, minimum, maximum).
2. Use Monte Carlo simulation to predict the probability distribution of possible outcomes.
3. Use first-order analysis to predict the mean and variance of the outcome uncertainty.
4. Obtain an analytical derivation of the output probability distribution function (PDF) or moments based on the PDF and moments of the input.
5. Elicit expert opinion on the system and its uncertainty, called the "Delphic monkeys" method.

For uncertainty analysis, the Monte Carlo method has the disadvantage that the required number of simulations may be very large, a special problem for large complex models such as typical SWMM applications. First order error analysis has the advantage of generating first and second moment statistics simply, and from a cost perspective is to be preferred (Scavia, 1980). However first-order analysis does not generate frequency distributions. We use first-order analysis (method B3 above) later in presenting uncertainty, and so we recommend sensitivity analysis as a fundamental part of the procedure for uncertainty determination.

Using our earlier notation, we may write

$$RF_l = M_c\left(IF, p_1, p_2, .., p_n\right) = M_c(\mathbf{X})$$

where RF_l = a particular computed response function

M_c = model expressed a s a function of or an operator on input variables and parameters,

IF = input function,

P = input parameters,

$\mathbf{X}$ = input file expressed as a vector of uncertain variables and parameters

We take it for granted that the mean (or expected vale *E)* of the computed response of a particular model is given by the computed response of the model using the mean values of all the input parameters and the expected values of the input variables:

$$E\{RF_l\} = E\{M_c(\mathbf{X})\} = M_c(E\{\mathbf{X}\})$$

Now if we assume that the model behaves approximately linearly (meaning that the principle of superposition can be invoked as an approximation, and if the standard deviations are small, then the variance of the computed response is given by

$$Var\{RF_l\} = \sum_{i=1}^{n}\left(\frac{\partial M_c}{\partial p_i}\right)Var\{X_i\} + 2\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\frac{\partial M_c}{\partial X_j}\frac{\partial M_c}{\partial X_k}S\{X_j\}S\{X_k\}\rho_{Xj,Xk}$$

where Var = the variance operator,
S = standard deviation
$\rho_{Xj,Xk}$ = correlation coefficient between X_j and X_k
n = number of parameters and sensitivity gradients

For independent parameters,

$$Var\{RF_l\} = \sum_{i=1}^{n} Var\{p_i\}\left(\frac{\partial M_c}{\partial p_i}\right)^2$$

Note that for practical purposes:

$$\frac{\partial M_c}{\partial p_i} = \frac{\Delta RF_{li}}{\Delta p_i}$$

$$E\{RF_l\} = E\{M_c(\mathbf{X})\} + \sum_{i=1}^{n}(E\{p_i\} - p_i)\frac{\Delta RF_l}{\Delta p_i}$$

> Use first-order analysis to predict the mean and variance of the outcome uncertainty.

10.7 Uncertainty analysis for a water distribution model

Water distribution modeling is fundamentally different from runoff modeling, in that the demands at each node are uncertain, and virtually unmeasurable in the field. In a sewer network the inflows at nodes are similarly uncertain. The following example of a detailed uncertainty analysis is abstracted from: *Analysis of the 1998 water-distribution system serving the Dover Township* *area, New Jersey*. (Water Science and Technology Vol 36 No 5 pp 141-148.(Anonymous, 1997).

The American Water Works Association (AWWA) Engineering Computer Applications Committee indicate that *true model calibration is achieved by adjusting whatever parameter values need adjusting until a reasonable agreement is achieved between model-predicted behavior and actual field behavior* (AWWA Engineering Computer Applications Committee 1999). Once a model is considered to be calibrated, it can then be used, among other purposes, to estimate hydraulic characteristics of the real-world system at locations where measured data are unavailable or unknown, spatially and temporally.

In the US, definitive standards to assess the accuracy of water distribution model calibration have yet to be agreed upon or established. However, the following calibration criteria have been suggested:

1. An average pressure difference of ±2.2 psi with a maximum difference of ±7.3 psi for a *good* data set, and an average pressure difference of ±4.3 psi with a maximum difference of ±14.2 psi for a *poor* data set (Walski 1983); and
2. The difference between measured and simulated values should be ±5 psi to ±10 psi (Cesario and Davis 1984).

According to the AWWA Engineering Computer Applications Committee (1999), ten sources of possible error could cause poor agreement between water distribution model values and measured field values. These sources of error, which provide a potential list of factors that can be adjusted during the model-calibration process, are:

1. errors in input data (measured and typographic),
2. unknown pipe roughness values (i.e., Hazen-Williams C-factors),
3. effects of system demands (distributing consumption along a pipe to a single node),
4. errors in data derived from network maps,
5. node elevation errors,
6. errors introduced by time variance of parameter values such as storage tank water levels and pressures,
7. errors introduced by a skeletal representation of the network as opposed to modeling all small-diameter pipes,
8. errors introduced by geometric anomalies or partially closed valves,
9. outdated or unknown pump-characteristic curves, and
10. poorly calibrated measuring equipment including data loggers, tank water-level monitors, and SCADA systems.

Model calibration entails adjusting model parameter values until an acceptable match is achieved between measured data and model-simulated values (e.g. pressures at the test hydrants, water levels in the storage tanks, flows from booster pumps, and pumpage from groundwater wells). The ten sources of possible error that could lead to model simulated values not agreeing with measured values are listed above and these provide a list of potential model parameters that can be modified during the calibration process. To decide which parameters might require more, less, or no modification, investigators evaluated each parameter as to the qualitative magnitude of error (high, moderate, or low) that could result from uncertainty and variability of the parameter. Three of the sources of possible error were evaluated as having a qualitatively high or moderate error magnitude:

1. unknown pipe roughness (Hazen-Williams "C-factor") values,
2. effects of system demands and consumption, and
3. outdated or unknown pump-characteristic curve data.

The initial estimates for these three parameters were subjected to possible variation during the calibration process. The remaining seven sources of possible error are believed to introduce minor to insignificant errors to model simulations, and therefore, were not modified during the calibration process. Results of the calibration analysis are presented in Figure 10.3. In this case a rather large number of points are plotted, and their scatter is of interest. Wide scatter indicates high model uncertainty. After parameter uncertainty is removed by careful calibration, the residuals are due to observation and model structural uncertainty. By plotting estimated bounds of observation error, conclusions may be drawn about model structural error.

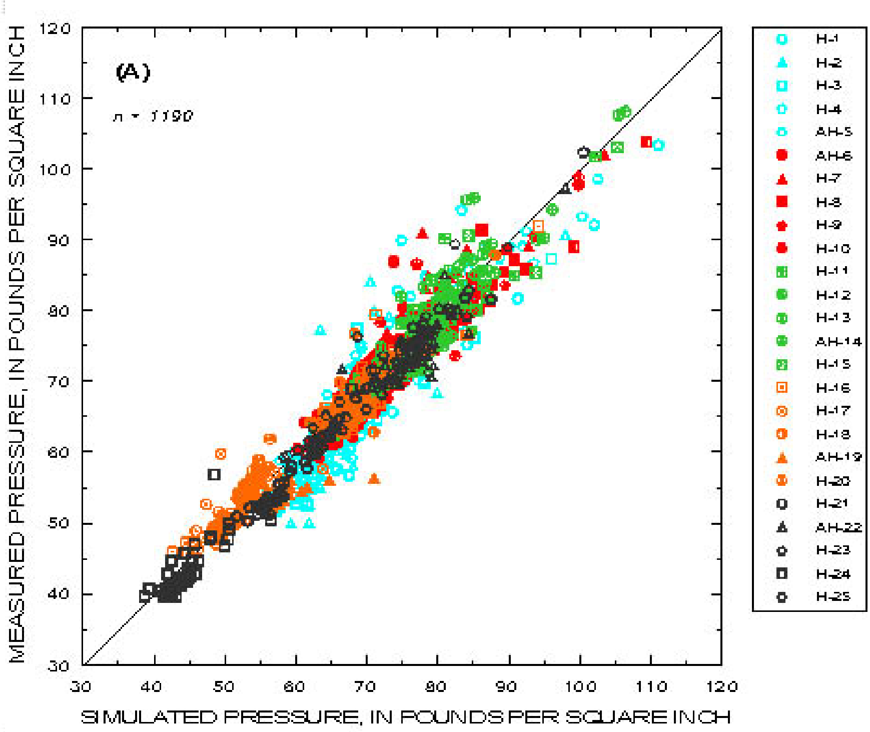


Figure 10.3 Comparison of measured and simulated hourly pressure data: (A) March 24-25, 1998; and (B) August 14-15, 1998.

For N>C, scatter of the residuals (after calibration and correcting for observation error) is a measure of the structural uncertainty (N=number of objective functions, C = complexity, or number of uncertain parameters).

10.8 Model reliability

Evaluation of the reliability of a model is recommended by several researchers (see for example Qaisi, 1985). Citing previous workers he used a reliability index RI, where RI reflects how closely the computed responses agree with the true responses. More reliable models have RI closest to 1, and if RI = 1, the model is perfect. RI is given by:

$$RI = \frac{1+S}{1-S} \geq 1,$$

where

$$S = \sqrt{\left(\frac{1}{n}\right)\sum_{i=1}^{n}\left|\frac{1-\left(OOF_i/COF_i\right)^2}{1+\left(OOF_i/COF_i\right)^2}\right|}$$

where COF = computed objective function,
OOF = observed objective function

10.9 Concluding remarks

Modelers should perform as much uncertainty analysis as they can afford, even though only a subset of it may be of interest to clients and other downstream users of the model results (Kann and Wyant, 2000). Complete uncertainty analyses is not only responsible modeling practice, it increases confidence in the model structure, parameter values and computed results.

Extending Willems (2000), a fairly simple procedure may be recommended for uncertainty analysis: 1. list and estimate all sources of uncertainty; 2. quantify the input (e.g. rain) and output (e.g. runoff) uncertainties, 3. estimate independently the measurement errors for those parameters that are measured, 4. quantify total uncertainty by sensitivity analysis for each parameter and sub-system, and 5. consider the remaining uncertainty to be model-structure uncertainty.

There is some evidence that the effect of several of the sources of error listed can be mitigated by careful parameter estimation. Certainly the perhaps unintentional effect in much surface water quality modeling, is to correct for what are sometimes gross errors in all eight categories of error listed in this chapter, by a substantial parameter optimization effort. To put it crudely: we must avoid the common trap where for instance poor rain data is corrected by a thoughtless quick fix to the parameters controlling soil surface infiltration capacities. (This is an example of *irresponsible* modeling.)

To conclude this section, some of our terms in this book have been defined as follows:

error, is the difference between a computed and an observed value, expressed as a percentage of the observed value;
uncertainty is taken to mean a range of possible values that an error may have, expressed as a percentage of the observed value;
variability, the different values that a parameter may have;
variance, the square of the standard deviation (a measure of uncertainty);
model complexity is a measure of the number of uncertain parameters in a model; and

model structure is the complex relationship of and between the component processes, most of which use uncertain parameters.

1. list and estimate all sources of uncertainty;
2. quantify the input (e.g. rain) and output (e.g. runoff) uncertainties,
3. estimate independently the measurement errors for those parameters that are measured,
4. quantify total uncertainty by sensitivity analysis for each parameter and sub-system, and
5. consider the remaining uncertainty to be model-structure uncertainty.

References

Anonymous. Analysis of the 1998 water-distribution system serving the Dover Township area, New Jersey. Water Science and Technology Vol 36 No 5 pp 141-148.

Andrews, G.C. and Ratz, H.C. 1997. Introduction to professional engineering. Faculty of Engineering, U of Waterloo. ISBN 0-9698850-1-6. 296pp

Baffaut, C. 1988. Knowledge techniques for the analysis of urban runoff using SWMM. Ph.D. thesis. Purdue U. 321 pp.

Beck, M.B. 1987. Water quality modelling: a review of the analysis of uncertainty. Water Resources Research 23(8):1393-1442. August 1987.

Beven, K. and Binley, A. 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes, 6:279-298.

Bower, D.E. 1985. Evaluation of the Precipitation-runoff modelling system, Beaver Creek, Kentucky. USGS Water Resources Investigations report 84-4316. 1985. 39pp.

Brabets, T.P. 1987. Quantity and quality of urban runoff from the Chester Creek Basin, Anchorage, Alaska. USGS Water Resources Investigations report 83-4238. 1983. 207pp.

Brendecke, C.M. Laiho, D.R., and Holden, D.C. 1985. Comparison of two daily streamflow simulation models of an alpine watershed. J. Hydrology. 77:171-186. 1985.

Brown, L.C. 1985. Uncertainty analysis using QUAL2E. Proc. Stormwater and water quality model users group meeting, Dec 5-6 1985 pp 125-140.

Brown, L.C. and Barnwell, T.O. 1987. The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS: documentation and user manual. U.S.EPA Report EPA/600/3-87/007. 199pp.

Carey, W.P. and Simon, A. 1984. Physical basis and potential estimation techniques for soil erosion parameters in the precipitation-runoff modelling system. USGS Water Resources Investigations report 84-4218. 32pp 1984.

Chaubey, I., Haan, C.T., Salisbury, J.M. and Grunwald, S. 1999. Quantifying model output uncertainty due to spatial variability of rain. J. Amer Water Resources Assoc. 35(5) 1113-1123.

Dilks, D.W. 1987. Analysis of water quality model uncertainty using a Bayesian Monte Carlo method. Ph.D. Thesis. U. of Michigan. Ann Arbor. 184 pp.

Draper, D. 1995. Assessment and propagation of model uncertainty. J.R. Statist. Soc. B 57 No.1 pp. 45-97.

Dunn, A.R., 1986. Feasibility of a Real-Time Auto Sensitivity and Error Analysis Framework for Fuzzy Stormwater Modelling on Personal Computers, M.Eng. Thesis, McMaster U., May 1986, 130pp.

Gardner, R.H., D.D. Huff, R.V. O'Neill, J.B. Mankin, J.H. Carney, and J. Jones, 1980. Application of error analysis to a marsh hydrology model. *Water Resour. Res. 16:*659-664,

Hoybye, J.A. 1998. Uncertainty modelling in hydrological applications. Doctoral dissertation, Lund University. 249pp.

James, W., Huber, W.C., Rossman, L.A., Dickinson, R.E., Pitt, R.E., James, W. R.C. , Roesner, L.A. and J.A. Aldrich, 2005. User's guide to SWMM. CHI pubs. ISBN 0-9736716-1-0. Guelph. 942 pp. page 148.

Johnston, P.R. and Pilgrim, D. 1976. Parameter optimization for watershed models. Water Resources Research, 12(3):477-486.

Kann, A. and Wyant, J.P. 2000. Approaches for performing uncertainty analysis in large-scale energy/economic policy models. Environmental modeling and assessment, 5 (2000) 29-46

Kitandis, P.K. and Bras, R.L. 1980. Real-time forecasting with a conceptual hydrologic model: 1. Analysis of uncertainty. Water Resources Research 16(6):1025-1033.

Kuch, A. J. 1994. A sensitivity analysis shell for complex continuous water quality management modelling. M.Sc. Thesis. Univ of Guelph. ca.120 pp. January 1994.

Isukapalli, Sastry S. 1999 Uncertainty Analysis of Transport-Transformation Models Doctor of Philosophy Rutgers, The State University of New Jersey

Leavesley, G.H., Lichty, R.W., Troutman, B.M. and Saindon, L.G. 1983. Precipitation-runoff modelling system user's manual. Water Resources Investigations 83-4238. USGS. Denver Colo. 80225. ca 207 pp.

Linsley, R. 1973. A manual on collection of hydrologic data for urban drainage design. Hydrocomp Inc. Palo Alto, CA.

Liong, S.Y. , Chan, W.T and Lum, L.H. 1990. Knowledge-based system for SWMM runoff component calibration. J. Water Resources Planning and Management. 117(5): 507-524.

Melching, C.S. 1987 A reliability analysis on flood event forecasting with uncertainties. PhD thesis U of Illinois at Urbanan Champan. 290 pp.

Qaisi, K.M. 1985. Uncertainty analysis of stream dissolved oxygen models. Ph.D. thesis. U. of Michigan Ann Arbor. 234 pp.

Reed, L.A. 1986. Verification of the PRMS sediment-discharge model. Proc 4th Federal Interagency sedimentation conf. Mar 24-27. Las Vegas, Nev. pp6-44 to 6-54.

Rivera Santos, J. 1990. Parameter estimation in conceptual precipitation-runoff models with emphasis on error analysis. Proc of Int. Symp. on Tropical hydrology and 4th Caribbean Islands water resources congress, San Juan, Puerto Rico, Jul 22-27, 1990. Amer. Water Resources Assoc. pp91-100.

Rivera Santos, J. 1988. A methodology for parameter estimation in conceptual precipitation-runoff models with emphasis on error analysis. Ph.D. thesis. U. of Colorado Boulder. 279 pp.

Scavia, D. 1980. Uncertainty analysis of a lake eutrophication model. Ph.D. thesis. U. of Michigan Ann Arbor. 174 pp.

Sorooshian, S., and V.K. Gupta, **1985.** The analysis of structural identifiability: Theory and application to conceptual rainfall-runoff models, *Water Resour. Res., 21*(4): 487-496, 1985.

Thompson, D.B. 1989. Determining parameters for a continuous simulation model by estimation, synthetic calibration and analytic calibration. Ph.D. thesis. U. of Missouri-Rolla. 261 pp.

Thompson, D.B. and Westphal, J.A. 1989. Tilt at the windmill of runoff simulation. Proc of the Int Conf on catchment runoff, U of Virginia, Charlottesville. 1989. pp176-185.

Troutman, B.M., **1985a**: "Errors and Parameter Estimation in Precipitation-Runoff Modeling. 1. Theory." *Water Resources Research,* 21(8):1195-1213.

Troutman, B.M. 1985b: "Errors and Parameter Estimation in Precipitation-Runoff Modeling. 2. Case Study." *Water Resources Research,* 21(8):1214-1222.

Walker, W.W. 1982a. A sensitivity and error analysis framework for lake eutrophication modelling. Water Resources Bulletin. 18(1):53-60.

Walker, W.W. 1982b. MODELER a computer modeling and data management system. Ver 4.0 Mar. 1982. W.W.Walker, Concord, MA. 21pp.

Willems, P. 2000. Probabilistic immission modelling of receiving surface waters. Doctoral dissertation, Catholic University of Leuven, Belgium. Ca. 250pp.

Wood, E.F., 1976: "An Analysis of the Effects of Parameter Uncertainty in Deterministic Hydrologic Models." *Water Resources Research,* 12(5):925-932.

Yan, J. 1990. Parameter estimation for multipurpose hydrologic models. Ph.D. thesis. Oklahoma State U. 184 pp.

Yoon, C.G. 1994 Uncertainty analysis in stream water quality modeling: reliability and data collection for variance reduction. PhD thesis, Rutgers State Univ of New York. 164pp.

Bibliography

Aitken, A.P., **1973.** Assessing Systematic Errors in Rainfall-Runoff Models. *Journal of Hydrology,* 20:121-136.

Amorocho, J. and B. Espildora, 1973: "Entropy in the Assessment of Uncertainty in Hydrologic Systems and Models." *Water Resources Research,* 9(6):1511-1522.

Andersson, Lotta. 1992. Improvements of runoff models What way to go? Nordic Hydrology, 23:315-332.

Angulo, C.E. Puente, and Bras, R.L. 1984. Non-linear filtering, parameter estimation and decomposition of large rainfall-runoff models. Report 297, (R-84-13; OSP 89310) Ralph M. Parsons laboratory, MIT. Oct 1984.

Ayyub, B.M. and McCuen, R.H. 1985. Quality and uncertainty assessment of wildlife habitat with fuzzy sets. J. Water Resources Planning and Management. 113(1):95-109.

Beck, M.B. 1983. Uncertainty, system identification, and the prediction of water quality. In: Uncertainty and forecasting of water quality, M.B. Beck and G. van Straten, eds. Springer-Verlag 1983. pp 3-68.

Beck, J.V. and K.J. Arnold, **1977**: Parameter Estimation in Engineering and Science. John Wiley and Sons, Inc. NY.

Benjamin, J.R., and C.A. Cornell, **1970.** *Probability, Statistics, and Decision for Civil Engineers,* McGraw-Hill, New York, 1970.

Bergstrom, S. 1991. Principles and confidence in hydrological modelling. Nordic Hydrology, 22: 123-136.

Blockley, D.I., Pilsworth, B.W. and Baldwin, J.F. 1983. Measures of Uncertainty. Civil Engineering Systems, Vol.1, pp.3-9, Sept. 1983.

Burn, D.H. **1985** Optimization Techniques Applied to Water Quality in an Uncertain Environment. Ph.D., Univ. of Waterloo, 1985.

Calver, A. 1988. Calibration, sensitivity and validation of a physically-based rainfall-runoff model. J. of Hydrology. 103:103-115.

Chapman, T.G. 1968. Catchment Parameters for a Deterministic Rainfall-Runoff Model. In: 'Land Evaluation', MacMillan of Australia. pp.312-323.

Crawford, N. 1993. The next major breakthrough... Paper presented at the U.S.EPA workshop on Developing dynamic watershed modelling capabilities for the Great Lakes tributaries. Heidelberg College, Tiffin, Oh. August 4-6, 1993. 8pp.

Cunge, J.A. **1974.** Evaluation Problem of Storm Water Routing Mathematical Models. Water Research, Vol.8, pp.1083-1087.

Duckstein, L., Monarchi, D., and Kisiel, C.C. 1973. Interactive Multi-Objective Decision Making Under Uncertainty. A paper prepared for the 1973 NATO Conference on "The Role and Effectiveness of Decision Theories in Practice:, Luxembourg, August 1973. 39 pp.

Edwards, D.R. 1988. Incorporating Parametric Uncertainty into Flood Estimation Methodologies for Ungaged Watersheds and for Watersheds with Short Records. Ph.D., Oklahoma State Univ.,

El-Hosseiny, T. and James, W.,1991. Urban Drainage Simulation with Real-Time Error Analysis. Annual Conf. of CSCE, Vancouver, B.C., pp 306-314, May 29-31, 1991.

Goldberg, M.A. **(undated)**. The Last of the Dinosaurs or the First of the Primates: Yet Another Simulation Study. Unpublished manuscript. 32 pp.

Helton, J.C., Iman, R.L., Johnson, J.D. and Leigh, C.D. 1986. Uncertainty and sensitivity analysis of a model for muticomponent aerosol dynamics. Nuclear tech. 73:320-342.

Hromadka, T.V. 1987. Uncertainty in hydrologic models: a review of the literature. Proc. Stormwater and water quality model users group meeting, Mar 23-24 1987. Denver Colo. USEPA pp 217-227.

Hwang, J.H. 1985. Uncertainty analysis in mathematical water quality modelling. Proc.Stormwater and water quality model users group meeting, Dec 5-6 1985 pp 81-100.

Ibbitt, R.P and Hutchinson, P.D. 1984 Model parameter consistency and fitting criteria. 9th IFAC Triennial World Congress, Budapest, Hungary.

Larson, L.W. and E.L. Peck, 1974: "Accuracy of Precipitation Measurements For Hydrologic Modeling." *Water Resources Research,* 10(4):857-863.

Lei, J. and Schilling, W. 1993. Propagation of model uncertainty. 6th Int. Conf. Urban Stormwater Drainage. Niagara Falls, Ont. Sept 1993. pp465-470.

Litchfield, J.B. and Warwick, J.J. 1993. The effect of spatial data aggregation on urban stormwater quality modeling. Preprint for 1993 Runoff quantity and quality modelling conf. ASCE. Reno, Nev. Nov.1993.5pp.

Melching, C.S. and Anmangandla, S. 1992. Improved first-order uncertainty method for water-quality modelling. J. Hydraulics Div ASCE. 118(5):791-805.

Moore, S.F. 1973. Estimation theory applications to design of water quality monitoring systems. J. of the Hydraulics Div. ASCE. 99(5):815-831.

Nnaji, S., Davis, D.R. and Fogel, M.M., April **1974.** Uncertainty in Predictions from a Deterministic Watershed Model. Paper presented at the 55th Annual Meeting American Geophysical Union, Washington DC 20036, 21 pp.

Rogers, C.C.M., Beven, K.J., Morris, E.M. and Anderson, M.G. 1985. Sensitivity analysis, calibration and predictive uncertainty of the Institute of Hydrology Distributed Model. J. of Hydrology, 18:179-191.1985.

Sircar, J.K. 1986. Computer Aided Watershed Segmentation for Spatially Distributed Hydrologic Modelling. Ph.D., Univ. of Maryland,

Stephenson, D., 1989. Selection of stormwater model parameters. J. of Environmental Engrg. 115(1): 210-220.

Troutman, B.M., **1982:** "An Analysis of Input Errors in Precipitation-Runoff Modeling Using Regression with Errors in the Independent Variables." *Water Resources Research,* 18(4):947-964.

Troutman, B.M., **1983:** "Runoff Prediction Errors and Bias in Parameter Estimation Induced by Spatial Variability of Precipitation." *Water Resources Res.* 19(3):791-810.

Troutman, B.M. 1981. The Effect of Input Errors in Using Precipitation-Runoff Models for Runoff Prediction. Reprinted from the book: Statistical Analysis of Rainfall and Runoff. Proc. of the International Symposium on Rainfall-Runoff Modeling held May 18-21, 1981 at Mississippi State University, Mississippi State, Mississippi, USA. pp.305-314.

Troutman, B.M. 1986. Reducing Bias in Parameter Estimation Caused by Model Input Errors. Presented at AGU-AGM, Baltimore, May 18, 1986. 16 pp.

Van-Thanh-Van Nguyen and Wang, H. 1993. An assessment of uncertainty in the modelling of rainfall runoff process. 6th Int. Conf. Urban Stormwater Drainage. Niagara Falls, Sept 1993. pp 453-458.

Warwick, J.J., 1990. Estimating uncertainty of stormwater runoff computations. J. Water Resources Planning and Management. 116(2):187-204.

-----, and Cale, W.G., 1987. Determining the likelihood of obtaining a reliable model. J. Environmental Eng'rg. 113(5):1102-1119.

-----, and Cale, W.G., 1986. Effects of parameter uncertainty in stream modelling. J.Environmental Engrg. 112(3):479-489.

Yen, B.C. 1993. Reliability and uncertainty analyses in hydraulic design. A sub-committee report. ASCE.

Young, P. 1983. The validity and credibility of models for badly defined systems. In: Uncertainty and forecasting of water quality, M.B. Beck and G. van Straten, eds. Springer-Verlag 1983. pp 69-98.

Zaghloul, N.A. 1981. SWMM model and level of discretization. J. of the Hydraulics Div. ASCE. 107(11):1535-1545.

Chapter 11

SENSITIVITY ANALYSIS

A little inaccuracy sometimes saves tons of explanation – Moung Ka.

11.1 Introduction

Interestingly, only three types of modeling experiments are possible:

1. to *run* the model for a certain duration of input time series (*inference*);
2. to run the model several times for different values of an input environmental parameter, comparing the output (*sensitivity analysis*); and
3. to run the model repetitively to find the optimum value of an environmental parameter, i.e., one that provides a computed response that best-fits an equivalent function that is deemed to be correct (*parameter optimization*).

Methods for model calibration and error analysis (e.g., least squares) have as a basic goal, to correlate the computed and observed objective functions (*COF*s and *OOF*s). This goal is achieved in two phases: first, the independent parameters p_i that most affect the behavior of the model are analyzed; and second, an error analysis is used to determine the effect of the uncertainty of these parameters on the *COF*. This chapter covers the first phase.

Sensitivity analysis is only carried out on models, their input and responses. We are unable in his book to say anything definite about the sensitivities of real processes, other than by inferring them from model performance. Sensitivity illustrates the impact of input parameter uncertainties (e.g. sub-catchment width, slopes, infiltration parameters) on model responses (e.g. peak runoff).

In fact, and the importance of this observation cannot be overstated, the set of dominant parameters is not constant with time, but varies according to model state variables and the input driving forces (e.g., soil water content and water levels, rainfall intensity and duration).

The aim of sensitivity analysis is to estimate the rate of change in the response of a model with respect to changes in model input parameters. Such knowledge is important for

1. evaluating the applicability of the model,

2. determining parameters for which it is important to have more accurate values, and
3. understanding the behavior of the system being modeled.

Generally speaking, sensitivity is defined as the rate of change of a computed objective function (a *COF* is one particular statistic or property of a computed response function, or *CRF*) with respect to change of a selected model input or environmental parameter *p*. We may write

$$COF_{c,j} = M_c\left(p_1, p_2, ..., p_n\right)$$

where M denotes model (operator),
c denotes model complexity
j denotes a particular objective function.

An expression for sensitivity can be written using the Taylor series:

$$M_c\left(p_i + \Delta p_i, p_{j|j\neq i}\right) = COF_0 + \frac{\partial COF_0}{\partial p_i}\Delta p_i + \frac{1}{2!}\frac{\partial^2 COF_0}{\partial p_i{}^2}\Delta p_i{}^2 + ...$$

where COF_0 = computed objective function at value of 0 at some specified level of each p_i. The equation can be reduced if the higher order terms are considered to be negligible:

$$M_c\left(p_i + \Delta p_i, p_{j|j\neq i}\right) = COF_0 + \frac{\partial COF_0}{\partial p_i}\Delta p_i$$

$$\Delta COF_0 = M_c\left(p_i + \Delta p_i, p_{j|j\neq i}\right) - COF_0 = \frac{\partial COF_0}{\partial p_i}\Delta p_i$$

The latter equation is called the linearized sensitivity equation. The choice of a sensitivity analysis method depends to a great extent on

1. the sensitivity measure employed,
2. the desired accuracy in the estimates of the sensitivity measure, and
3. the computational cost involved.

Table 11.1 presents some of the sensitivity measures that are often employed in sensitivity analysis.

Table 11.1 Summary of sensitivity measures used in sensitivity analysis (Isukapalli, 1999).

Sensitivity Measure	Definition
Response from arbitrary parameter variation	$COF = M_c\left(\overline{p} + \delta p\right) - M_c\left(p\right)$
Normalized Response	$D_i = \dfrac{\delta p_i}{M_i\left(\partial p\right)}$
Average Response	$\overline{COF_i\left(\overline{p}\right)} = \dfrac{\int_n \cdots \int M_i\left(\overline{p}\right)dp}{\int_n \cdots \int dp}$
Expected Value	$\left\langle COF_i\left(p\right)\right\rangle = \int_n \cdots \int M_i\left(pP\left(p\right)\right)dp$
Variance	$\delta_i^2\left(p\right) = \left\langle M_i\left(p\right)^2\right\rangle - \left\langle M_i\left(p\right)\right\rangle^2$
Extrema	$\text{Max}\left[M_i\left(p\right)\right],\ \text{Min}\left[M_i\left(p\right)\right]$
Local Gradient Approximation	$\delta COF \approx \left[SG\right]\delta p; SG_{ij} = \dfrac{\partial COF_i}{\partial p_j}$
Normalized Gradient	$SG_{ij}^n = \dfrac{\overline{p}_j}{M_i\left(\overline{p}\right)} \dfrac{\partial COF_i}{\partial p_j}$

Sensitivity analysis methods can be broadly classified into the following categories:

***Variation of parameters or model formulation*:** In this approach, the model is run at a set of sample points (different combinations of active parameters) or with straightforward changes in model structure (e.g., in model comlexity). Sensitivity measures that are appropriate for this type of analysis include the response from arbitrary parameter variation, normalized response and extrema. Extreme values are often critical in environmental applications.

***Domain-wide sensitivity analysis*:** Here sensitivity involves the study of the system behavior over the entire range of parameter variation, often taking the uncertainty in the parameter estimates into account.

***Local sensitivity analysis*:** Here the focus is on estimates of model sensitivity to input and parameter variation in the vicinity of a sample point. This sensitivity is often characterized through gradients or partial derivatives at the sample point.

In the next section different methods for sensitivity analysis are discussed, to help us establish the extent and limitations of using first order analysis to calibrate model parameters. It draws from the work by El Hossieny (1996).

Factor Perturbation Method:

The factor perturbation method is commonly used for hydrologic modeling due to its simplicity (McCuen, 1973). Sensitivity coefficients or gradients (SG_i) for a particular parameter can be defined by the partial derivative of an objective function *COF* with respect to a parameter p_i.

$$SG_i = \frac{\partial COF_0}{\partial p_i}$$

The procedure allows exploration of model responses to small changes in input variables, parameters, initial conditions, functional relationships, and coefficients associated with numerical techniques. The limitations of this method are:

1. The sensitivity gradient is assumed to be linear, an assumption that may be valid only over a limited range of the parameter (most hydrologic models are thought to be nonlinear).
2. The sensitivity is estimated by holding the other parameters constant at some expected value, which may not be valid for real-world processes.
3. The method gives a single-valued indication of the effect on the objective function. In most cases, the decision maker would prefer to have an idea of the distribution of the design parameter.

Monte Carlo Method:

The Monte Carlo method treats parameter variability by combining the estimates of errors from all sources. A particular set of parameters is selected in a random generation procedure. Each error source is characterized by its frequency distribution rather than by a single value. For each simulation, a set of parameters **P** is selected, and an objective function (*COF*) is calculated. If a large number of vectors **P** is generated randomly the probability of a particular set being optimum tends to unity as the number of generated vectors tends to infinity.

The Monte Carlo method has many applications for assessing the degree of confidence, and is thought by many to provide a better representation of the reliability and the precision of the computed results than do linear techniques (e.g., first order analysis), when dealing with nonlinear or discontinuous systems.

Despite its flexibility the Monte Carlo method was excluded from PCSWMM because of the large number of trials needed. Availability of reliable parameter observations in enough quantity to define the probability distribution is a demanding requirement of this method (e.g., of infiltration parameters, watershed width). Also, with the Monte Carlo method, the lumped uncertainty of all model parameters keeps the individual contribution unknown, which is considered to be a disadvantage (El Hosseiny, 1996).

11.2 Sensitivity Analysis and Hydrological Models:

Several researchers have applied sensitivity analysis to hydrological models. Table 11.2 lists three investigations of the sensitivity of different input parameters for the SWMM program (Graham et al., 1974, Huber et al., 1975, Jewell et al., 1978). All of these investigations kept every parameter except one constant, and the sensitivity gradients were developed through a range of the parameter values around its mean or expectation. Table 11.2 lists the wide range of their various ratios to the mean values, and the corresponding ratios for computed runoff volumes. Evidently sensitivity analysis (SA) yields different results for different applications. Specific SA should be carried out for every application.

Table 11.2: Sensitivity Analysis Results

Parameter	Col. 1		Col. 2		Col .3	
	A	B	A	B	A	B
% of impervious area	1:7	1:2.01	-	-	1:4	1:3.58
Pervious area minimum infiltration	1:10	1.82:1	1:150	1.75:1	-	-
Characteristic width	1:79	2.00:1	-	-	2:1	1.03:1
Pervious area Manning's n	1:4	1.36:1	1:150	1.00:1	-	-
Impervious area depression storage	1:4	1.24:1	1:200	1.22:1	1:7	1.28:1
Pervious area depression storage	1:4	1.02:1	1:50	1.00:1	-	-
Impervious area Manning's n	1:3	1.02:1	1:100	1.02:1	1:2.3	1. 021: 1

Column 1 = Graham, et al. (1974)
Column 2 = Huber, et al. (1975)
Column 3 = Jewell, et al.(1978)
A = Range = Original parameter value/maximum parameter value
B = tot flow vol using initial parameter value/tot flow vol using max parameter value

McCuen (1973) used sensitivity and error analysis to examine the structure of a commonly used evaporation model. Three models were investigated, the Fractional-Evaporation Equivalent (FEE), the Penman model and the Weather Bureau (WB) model. He recommended a relative sensitivity:

$$DSG_r = \frac{\partial COF/COF}{\partial p/p}$$

where: DSG_r = relative sensitivity
COF = objective function
p = input parameter

After applying the sensitivity analysis for the three models McCuen (1973) determined that the FEE is an inadequate method to simulate evaporation because it assumes that the effect of change in radiation or evaporation rate does not change with time, while the Penman and the WB model are more accurate due to higher sensitivity to the effects of the meteorological variables.

Mein and Brown (1978) applied a sensitivity analysis procedure to the Australian Boughton Model. The model has thirteen parameters, all of which were evaluated by curve fitting. The model was applied to four catchments, and the watershed parameters were changed by increasing and decreasing their values by 5% to produced a matrix of sensitivity gradients **SG**. To determine the best estimate of the true parameter, the least squares functions were calculated and minimized. The covariance matrix was used to calculate the coefficient of variation (standard deviation/parameter value). It was concluded that, for the Boughton Model, relationships between parameter values and model prediction are imprecise, due to the wide range of uncertainty in the parameters.

James and Robinson (1981) applied sensitivity analysis using the perturbation method on the RUNOFF block of SWMM4. The analysis was applied to a system of eleven subcatchments for quantity and quality simulations. A total of twelve parameters were tested (four in quantity and eight for quality). A series of procedures to ensure credibility and confidence in model results was recommended. These procedures are a series of verification, validation, and sensitivity tests that produce information about model performance.

Rogers, et al. (1985) applied sensitivity analysis to the U.K. Institute of Hydrology distributed Model (IHOM). The analysis examined parameter sensitivity and its influence on computed flow using the perturbation method. The sum of the squared errors between observed and computed discharge was used as a measure of the effect of changes in the model parameters. The analysis was limited to surface flow roughness (Chezy C) and soil properties. Initial estimates were determined from previous work. It was concluded that the parameter sensitivity is storm-intensity dependent. In other words, a parameter that may be sensitive for one specific storm intensity may not be sensitive for a different storm (which should come as no surprise to readers of this book). The most sensitive input parameters were the Chezy roughness coefficient C and the saturated hydraulic conductivity K_{sat} for a hill-slope surface. These results agree with other studies conducted by Yen and Akan (1983) and Freeze (1972).

In PCSWMM, sensitivity analysis consists of :

1. varying model coefficients or parameters one at a time, with the amount varied being representative of the uncertainty in the parameter being analyzed,
2. dividing the resulting normalised change in computed response by the normalised parameter variation,
3. ranking the resulting modulus of the sensitivity gradients, highest to zero, and then
4. making inferences about the model uncertainty.

PCSWMM uses state variable (SV) sub-spaces to calibrate model parameters. Each process is active only during a limited state of the model. This state can be predetermined and is related to an input function. The SV space is the area where the process is active. The approach is described elsewhere.

A general sensitivity matrix of sensitivity coefficients must be developed for the entire range of rainfalls:

$$\mathbf{X}(p) \equiv \left[\nabla_p COF(p)\right]$$

$$\nabla_p \equiv \begin{vmatrix} \dfrac{\partial}{\partial p_1} \\ \dots \\ \dfrac{\partial}{\partial p_n} \end{vmatrix}$$

$$\mathbf{X}(p) = \begin{vmatrix} \dfrac{\partial COF_1}{\partial p_1} & \dots & \dfrac{\partial COF_1}{\partial p_n} \\ \dots & \dots & \dots \\ \dfrac{\partial COF_m}{\partial p_1} & \dots & \dfrac{\partial COF_m}{\partial p_n} \end{vmatrix}$$

where n = number of active parameters
m = number pf rainfall events
p = active parameters
COF = computed objective function

However, sometimes the available input functions (e.g., rainfall) and the corresponding observed objective functions do not cover the entire range of intensities for the modeled space. The user must ensure the availability of data before generating the sensitivity matrix.

Beck and Arnold (1977) and Mein and Brown (1978) applied the least squares technique to estimate error in model results. A sensitivity matrix was used to produce a covariance matrix. From the covariance matrix an objective function for a set of parameters, P could be estimated. The covariance matrix is defined:

$$\mathrm{var}\left(\beta_{ls}\right) = \mathbf{Z}^{-1}\sigma^2$$

where $\mathbf{Z} = \mathbf{A}^{\mathbf{T}}\mathbf{A}$

$\mathbf{A}$ = sensitivity matrix for m events and n parameters

$\mathbf{A}^{\mathbf{T}}$ = transpose of sensitivity matrix

β_{ls} = mean least squares of active parameters

$$\sigma^2 = \frac{\sum_{i=1}^{n}\left[OOF_i - COF_i\right]^2}{n}$$

In linear models, the sensitivity gradient is constant and does not depend on the values of model parameters. However, the above analysis fails when the determinant of ($\mathbf{A}^{\mathbf{T}}\mathbf{A}$) equals zero, as the minimum of the sum of squares does not provide a unique solution.

Some researchers caution against sensitivity analysis (e.g. Dilks, 1987) on the grounds that parameter uncertainty is so large that non-linear interactions become unavoidable. His argument is that interactions between uncertain parameters magnifies model uncertainty. This has not been the writer's experience, given the approaches advocated in this book.

Each parameter is associated with a process (or several), and each process is active only under certain model states. To estimate the best value of a parameter, the only states that need to be examined are the states or events when the related processes are active. The causative events need to be established, and selected from the record to be used for calibration; the specific, observed events are then used to calibrate the specific active processes and parameters.

In order to account for the expected range and distribution of estimates of parameters, which clearly vary widely in their certainty, it is useful to request the user to input three values for each parameter (Dunn, 1986; Dunn and James, 1985) based on the user's estimate of fellow users' best estimates. We use 12 estimates by way of illustration; because it helps convert into standard deviations, if the probability distribution of these estimates is known. The distributions by the way are likely to be normal as in percent imperviousness, or skewed as in Manning's n, and unlikely to be uniform. These three estimates are denoted:

1. E_{med}, the mean (or median for skewed distributions) expected value (the mean of 12 hypothetical users' best estimates),
2. E_{max}, a value near the top of the range, say the 8th highest of the 12 users' estimates of the expected value, and
3. E_{min}, a value near the bottom of the range, say the 4th highest of the same 12 user estimates of the expected value.

Note that the range of 12 users' best estimates of the expected value is not the same thing as one user's estimate of the feasible range – the latter will be much smaller than the former. The point is stressed elsewhere in this book as well. Values used in the perturbation for the sensitivity analysis should be a reasonably small fraction *s* of this range:

$$[E_{med}(1\text{-}s) - s.E_{min}], \quad E_{med}; \quad [E_{med}(1\text{-}s) + s.E_{max}]$$

The three values are shown schematically in Figure 11.1.

```
     <-------------------published parameter range----------->
                        <--SA range---->
     Emin               -0.25  Emed   +0.25                    Emax
computer run:             3     1       2
```

Figure 11.1: Schematic of the range of perturbed values for each parameter analyzed (s = 0.25; the origin may be far to the left).

> Analyze for sensitivity for at least three values of each parameter. Estimate three likely values nearer the expected value than the published ranges.

PCSWMM uses five parameter estimates so that the total number of runs required for *n* parameters is *4n + 1* For (say) 49 active parameters in a SWMM-RUNOFF data file, 197 event runs are required. Monte Carlo analysis on the other hand, would require perhaps hundreds of thousands of runs of the full continuous dataset.

11.3 Sensitivity gradients

Sensitivity analysis reveals the inner structure of the model, beyond those invented by the model builder. Sensitivity gradients can be used to estimate the propagated error, and optimize input parameters. In each limited SV sub-space use at least as many observed OFs as there are parameters to be optimized. Excess observations will improve the reliability of the parameter estimates. PCSWMM integrates the 197 input data files into one run, and accumulates a large sensitivity output file, by appending results from subsequent runs. Dimensionless sensitivity gradients (DSGs) are presented in a family of plots, one curve for each parameter, one family for each process per screen, as shown in Figure 11.2. DSGs are ranked as shown in Figure 11.3.

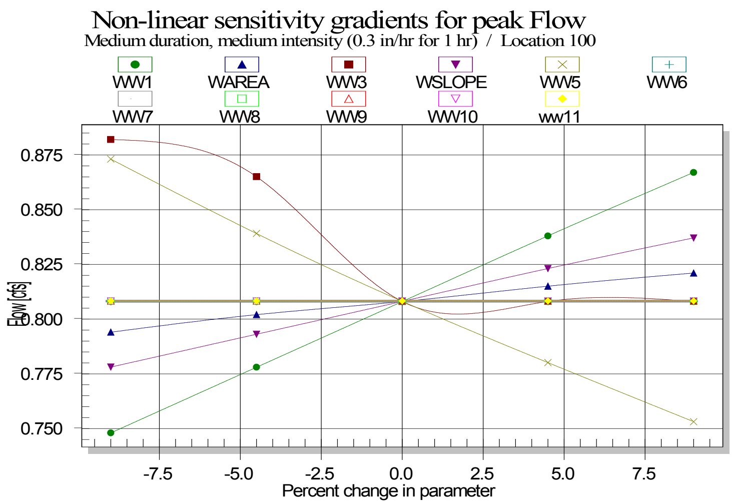


Figure 11.2 Non-linear sensitivity gradients.

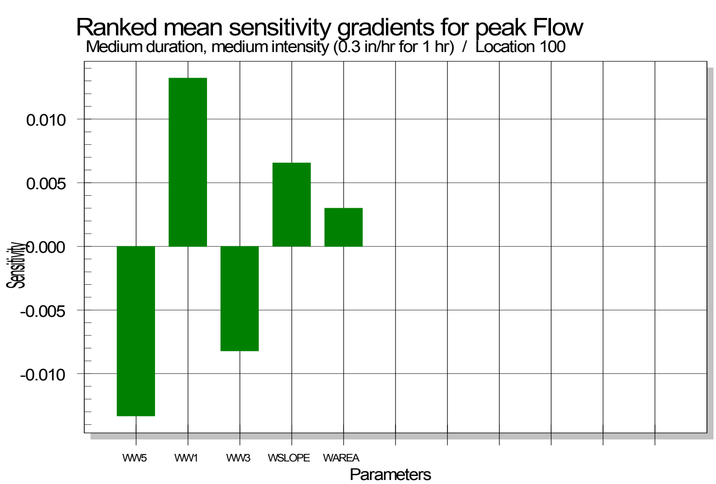


Figure 11.3 Ranked sensitivity.

Certain urban drainage processes are clearly non-linear. For instance, doubling the rate of rain does not double the runoff at a point if it flows through a surcharged sewer system, especially one that incorporates active or remotely and arbitrarily controlled diversions. The widespread practice of simply assigning an arbitrary perturbation of each input parameter based on its expected value, e.g. 25% of its *absolute* value, is probably not appropriate, as it could carry the sensitivity analysis into non-linear situations, which in turn will be difficult to manage. Due account should be taken of the model state. The sensitivity of some parameters may change with large perturbations - they may be curved as shown in Figure 11.2.

At the end of the sensitivity analysis, PCSWMM displays the ranked sensitivity gradients as shown in Figure 11.3 where the selected objective functions are i) peak flow, and ii) peak concentration of total solids. The colors selected by the program are used consistently for the same variables in all displays for all selected objective functions.

The trick to the success of this overall methodology is the structure of the logic underlying the limitation of the analyses to just the minimum number of runs absolutely essential for the analysis. Insensitive processes and parameters are eliminated at the outset, and only the most sensitive parameters are subject to further parameter optimization, certainty and error analysis. The next section uses fuzzy logic to explain the approach.

It is not necessary to calibrate all processes simultaneously for the entire duration of the continuous model study. Calibrate only against many, short causative events.

REFERENCES

Baffaut, C., 1989. Expert System for Calibrating SWMM, Journal of Water Research Planning and Management, Vol. 115, No. 3.

Beck, J.V. and K.J. Arnold, 1977, "Parameter Estimation in Engineering and Science", John Wiley, Inc., New York, New York.

Dawdy, D.R. and T. O'Donnell, 1965, 'Mathematical Models of Catchment Behaviour", Journal of the Hydraulics division, ASCE, Vol. 91, No. HY4, pp. 123-137.

Dilks, D.W. 1987. Analysis of water quality model uncertainty using a Bayesian Monte Carlo method. Ph.D. Thesis. U. of Michigan. Ann Arbor. 184 pp.

Dunn, A. 1986 Feasibility of a Real-Time Auto Sensitivity and Error Analysis Framework for Fuzzy Stormwater Modelling on Personal Computers, M.Eng. Thesis, McMaster U., May 1986, 130pp.

Dunn, A. and James, W. 1985. . Extended Expert Systems. Proc. of the Stormwater and Water Quality Modelling Conf., sponsored by USEPA and Ontario Ministry of the Environment, Toronto, Canada, published by CHI, Hamilton, pp. 373-380, Dec 5-6.

El Hosseiny, T. 1996. Uncertainty optimization for urban drainage systems. Sept. 1996. PhD thesis, Univ. of Guelph. Ca 250 pp.

Freeze, R.A., 1972, "The Role of Subsurface Flow in the Generation of Surface Runoff Upstream Source Areas", Water Resources Research, Vol, 8, No. 5, pp. 1272-12183.

Graham, P.H., L.S. Costello, and H.J. Mallon, 1974, "Estimation of Imperviousness and Specific Curb Length for Forecasting Stormwater Quality and Quantity", Journal of Water Pollution Control Federation, Vol. 46, No. 4, pp. 717-725.

Gupta, V.K and S. Sorooshian, 1983, "Uniqueness and Observability of Conceptual Rainfall - Runoff Model Parameters: The Percolation Process Examined", Water Resources Research, Vol. 19, No. 1, pp. 269-276.

Heaney, J.P., W.C. Huber, MA Medina, Jr., M.P. Murphy, S. J. Nix, and S. M. Hasan, 1977, "Nationwide Evaluation of Combined Sewer Overflows and Urban Stormwater Discharges Volume 11: Cost Assessment and Impacts", EPA-600-2-77-064, U.S. Environmental Protection Agency, Cincinnati, OH.

Huber, W.C. and R. Dickinson; 1988,"Storm Water Management Model Version 4, User's Manual, U.S. Environmental Protection Agency, Athens, Georgia.

Isukapalli, 1999.

James, W. And M. Robinson, 1981, "Standard Terms of Reference to Ensure Satisfactory Computer-based Urban Drainage Design Studies", Canadian Journal of Civil Engineering, Vol. 8, pp. 294-303.

James, W. 1993 Stormwater Management Modelling Conceptual Workbook. Pub. by CHI Guelph. 200 pp.

Jewell, TX, T.J. Nunno and D.D. Adrian, 1978, "Methodology for Calibrating Stormwater Models", Journal of the Environmental Engineering, ASCE, 104 (3): 485-501.

Leavesley, G.H., R.W. Lichty, B.M. Troutman and L.S. Saindon, 1983,'Precipitation - Runoff Modeling System" User's Manual, U.S. Geological Survey, Water Resources Investigations Rep., 83-4238.

Lee, H.L. and L.W. Mays, 1986, "Hydraulic Uncertainties in Flood Levee Capacity", Journal of Hydraulic Engineering, Vol. 112, No. 10, pp. 928-934.

McCuen, R.H., 1973, 'The Role of Sensitivity Analysis in Hydrologic Modeling", Journal of Hydrology, 18: pp. 37-53.

Mein, R.G. and B.M.Brown, 1978, "Sensitivity of Optimized Parameters in Watershed Models", Water Resources Research, Vol. 14, No. 2, pp. 299-303.

Melching, C.S. and S. Anmangandla, 1992, 'Improved First-Order Uncertainty Method for Water-Quality Modeling", Journal of Environmental Engineering, 188(5):791-805.

Rogers, C.C.M., 1CJ. Bewen, E.M. Morris and M.G. Anderson, 1985, "Sensitivity Analysis, Calibration and Predictive Uncertainty of the Institute of Hydrology Distributed Model", Journal of Hydrology, Vol. 81, pp. 179-191.

Troutman, B.M., 1985, 'Errors and Parameter Estimation in Precipitation-Runoff Modeling, Theory", Water Resources Research, Vol. 21, No. 8, pp. 1195-1213

Tung, B.Y. and L.W. Mays, 1980, 'Risk Analysis for Hydraulic Design", Journal of the Hydraulics Division, ASCE, Vol. 106, No. HY5, pp. 893-913.

Yeri, B.C. and A.0. Akan, 1983, 'Effects of Soil Properties on Overland Flow and Infiltration", Journal of Hydraulic Research, Vol. 21 (2), pp. 153-174.

Chapter 12

STATE VARIABLE SPACE

One of the fundamental tasks of engineering and science, and indeed of mankind in general, is the extraction of information from data. – J.V. Beck and K.J Arnold.

The things of the universe are not sliced off one from another with a hatchet, neither the hot from the cold, nor the cold from the hot. – Anaxagoras

12.1 Introduction

The inherent structure of a WQM is very much dependent on the model builder's notion of the processes involved, their relationships and interactions, and the meaning of the technical words and phrases used to describe them. The model builder uses ideas that are not always obvious to the user from the documentation for the WQM. Indeed several expressions may appear to a reader to be used ambiguously. Nevertheless, one of the inherent model building rules rules seems to be that model builders use basic schematics as building blocks, such that processes are separated and connected logically, and become active under previously conceived conditions. Programming languages and structured programs facilitate the procedure. In other words model builders deliberately define their processes so that each process is active only in limited states of the model, and so that the states can be predetermined and related to an input variable. Of course many processes may be simultaneously active, but the processes are each described by independent algorithms. Dependent variables that are computed within the algorithm and related to the response, are called state variables here. Water level in a pond or the groundwater, or on the overland flow surface, is an example. There is a clear and distinct boundary to the state-variable (SV) space in which each process is active, although of course more than one process may be active simultaneously {their state variable spaces overlap). The limited SV space is called the SV *sub-space* herein (Figure 12.1).

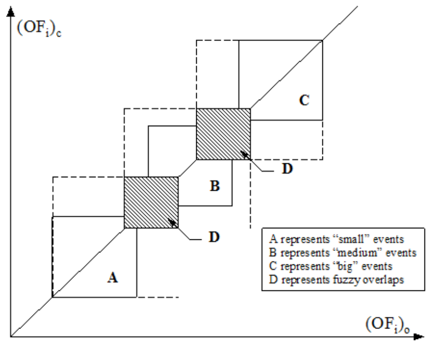


Figure 12.1 Calibration plot translating SV sub-spaces

These concepts allow model users to:

1. isolate the important empirical parameters that require refining (calibration),
2. associate these parameters with their correct processes (may be more than one),
3. isolate the conditions under which the processes are active (again may be more than one), and then
4. select state-variable events (SV sub-spaces) for sensitivity, and
5. select events from the observed record for calibration.

However, the effort is complicated by the fact that it is the relative contribution of each parameter that is important - users need to isolate the events in which the contribution from particular processes are dominant, rather than merely active. Each of these steps is covered by menus and displays in PCSWMM.

12.2 State variable sub-spaces

To clarify the foregoing discussion, consider for example overland flow from impervious surfaces, which becomes active as soon as light rain exceeds the impervious area depression storage. It then dominates the runoff process until the rain exceeds infiltration capacities and runoff is contributed from adjacent pervious areas (if any - the relative sensitivity is clearly affected by the percent imperviousness). The state variables here are both rate-of-rain (must not exceed infiltration capacities of pervious areas) and duration of rain (total rain must exceed impervious area depression storage). Since two dimensions of rain are being used, we refer to state-variable space. Relevant SVs are rain, snow, evapo-transpiration, wind, temperature, radiation, etc. Figure 12.1 illustrates the point. Sub-spatial effects, where certain processes dominate, are readily translated to the calibration plot.

In Figure 12.1 c denotes computed; o denotes observed; and OFi denotes the objective function chosen. Square sub-spaces have been chosen for mathematical simplicity. They can be explained simply as follows: consider peak flows as the objective function, in which case the state-variable is rainfall (product of rate and duration). *A* then denotes light rains, when infiltration capacities are not exceeded, *C* denotes heavy rains when all surfaces contribute and infiltration capacities have reached their asymptotic lowest values, *B* denotes intermediate events, when infiltration capacities are of the same order of magnitude as rains, and *D* denotes the fuzzy overlapping zones. Similar illustrations can be provided for processes such as, e.g. pollutant washoff, erosion, or snowmelt.

Fortunately it is not necessary to precisely determine entire boundaries of SV sub-spaces manually, because the sensitivity analysis routines developed herein are useful for this purpose. Also, it seems likely that fuzzy logic could be helpful in dealing with the fringes of the applicable SV sub-spaces (fuzzy logic is also discussed further later).

Table 12.1 Runoff parameters and their SV spaces (Kuch, 1997)

Parameter	Group	Parameter Space for Maximum Sensitivity Gradients			
		Duration	Rain Intensity	Temp.	Wind
RAINIT	Erosion	Short to Long	Med. to High	NA	NA
ERODAR	Erosion	Short to Long	Med. to High	High	Med. to High
ERLEN	Erosion	Short to Long	Med. to High	High	Med. to High
SOILF	Erosion	Short to Long	Med. to High	NA	Med. to High
CROPMF	Erosion	Short to Long	Med. to High	High	Med. to High
CONTPF	Erosion	Short to Long	Med. to High	High	Med. to High
DRYDAY	Quality	Short	Zero	NA	NA
DRYBSN	Quality	Short	Zero	NA	NA
DDLIM	Quality	Long	Zero	NA	NA
DDPOW	Quality	Long	Zero	NA	NA
DDFACT	Quality	Short	Zero	NA	NA
WASHPO	Quality	Short	Low	High	Med. to High
RCOEFF	Quality	Short	Low	High	Med. to High
CBFACT	Quality	Short	Low	Medium	Low
CONCRN	Quality	Short to Long	Med. to High	High	Med. to High

PCTZER	Quantity	Short	Low	High	Low
WIDTH	Quantity	Short to Med.	Low	Medium	Med. to High
AREA	Quantity	Short to Long	Low to High	Medium	Med. to High
%IMPER	Quantity	Short to Long	Low to Med.	Medium	Low
WSLOPE	Quantity	Short	Low	Medium	Med. to High
IMPERN	Quantity	Short to Med.	Low	Medium	Med. to High
PERVN	Quantity	Short	High	High	Med. to High
IMPDEP	Quantity	Medium	Low	Medium	Med. to High
PERDEP	Quantity	Short	High	High	Med. to High
WW9	Quantity	Short	High	High	Med. to High
WW10	Quantity	Medium	Medium	High	Med. to High
FWFRC1	Snowmelt	Long	Na	All	Med. to High
FWFRC2	Snowmelt	Long	Na	All	Med. to High
TIPM	Snowmelt	Long	Na	All	Med. to High
RNM	Snowmelt	Long	Na	All	Med. to High
WIND	Snowmelt	Long	Na	NA	All
GWIDTH	Pipe/Channel	Short to Long	Med. to High	High	High
LENGTH	Pipe/Channel	Short to Long	Med. to High	High	High
ISLOPE	Pipe/Channel	Short to Long	Med. to High	High	High
ROUGH	Pipe/Channel	Short to Long	Med. to High	High	High
SDEPTH	Pipe/Channel	Short to Long	Med. to High	High	High
A1	Groundwater	Long	Low	NA	NA
B1	Groundwater	Long	Low	NA	NA
A2	Groundwater	Long	Low	NA	NA
B2	Groundwater	Long	Low	NA	NA
A3	Groundwater	Long	Low	NA	NA
POR	Groundwater	Long	Low	NA	NA
WP	Groundwater	Long	Low	NA	NA
FC	Groundwater	Long	Low	NA	NA
HKSAT	Groundwater	Long	Low	NA	NA
TH1	Groundwater	Long	Low	NA	NA
HCO	Groundwater	Long	Low	NA	NA
PCO	Groundwater	Long	Low	NA	NA
REGEN	Not included	Medium	Zero	Low	Low
WW11	Not included	Short	High	High	Med. to High

Table 12.1 lists SWMM-RUNOFF input parameters showing the parameter names as they appear in the SWMM documentation. The location of each variable in the SWMM input datafiles is shown by the two-character line identifier (e.g.K1) and the location on the line (e.g. K1.8 for the eighth variable). Also shown are the process groups and the parameter space where each parameter is expected to be most sensitive.

For the sensitivity analysis, artificial time series (TS) are used, with constant intensities. The intensities and durations are chosen so that they relate in a fuzzy way to the scale of the model problem. These artificial state variable spaces are like simple design storms or design droughts; they are named according to their fuzzy zones as listed in Table 12.2 below.

Table 12.2: State-variable sub-spaces for auto-sensitivity analysis

The values shown as typical are suggested values for the mid-points of the fuzzy ranges. Users are required to select their own fuzzy zones.

description	acronym	typical values	
Rain:			
short-duration-high-intensity	SDHI	20 mins;	3 in/hr
medium-duration-high-intensity	MDHI	60 mins;	1.0 in/hr
long-duration-high-intensity	LDHI	600 min;	0.2 in/hr
short-duration-medium-intensity	SDMI	20 min;	0.4 in/hr
medium-duration-medium-intensity	MDMI	60 min;	0.3 in/hr
long-duration-medium-intensity	LDMI	600 min;	0.1 in/hr
short-duration-low-intensity	SDLI	20 min;	0.1 in/hr
medium-duration-low-intensity	MDLI	60 min;	0.1 in/hr
long-duration-low-intensity	LDLI	600 min;	0.1 in/hr
Evapo-transpiration:			
short-duration-high-intensity	SDHI	1 day;	0.5 in/day
long-duration-high-intensity	LDHI	10 days;	0.3 in/day
short-duration-low-intensity	SDHI	1 day;	0.05 in/day
long-duration-low-intensity	LDLI	10 days;	0.05 in/day

As a guide, to start the discussion, it may be helpful to relate dominant processes to SV sub-spaces as shown in Table 12.3.

> Associate parameters with processes, and processes with causative events, and causative events with limited state-variable sub-spaces.

Table 12.3 Process likely to dominate in various situations.

State variable sub-space	dominant processes
rain:	
light rate of rain	overland flow over impervious areas
medium rate of rain	infiltration into upper soil mantle; pollutant washoff
heavy rate of rain	erosion; pollutant washoff; flow over pervious area
long duration rain	overland flow over pervious areas
no rain:	
long duration drought	pollutant build-up; groundwater depletion
short duration drought	storage recessions
temperature:	
high temperatures	evapo-transpiration; snowmelt
low temperatures	snow accumulation and ripening
wind:	
high wind	snowmelt

Two dimensions of a driving SV are used to plot the SV space shown in Figure 12.2, with rate of rain or ET plotted vertically, and the duration of rain or ET horizontally. From the center of the SV space, the catchment becomes increasingly wet as we move to the right and up, and increasingly dry to the left and down. This tendency is similar to the calibration plots shown elsewhere in this book. The SV space can be divided up into sub-spaces, chosen to suit the model concepts. Since we have not (yet) placed numerical values on the input function, we retain flexibility to choose the ranges to suit the parameters and circumstances of the application (a classic case of circular thinking). As shown in Figure 12.2, each sub-space is shown with a fuzzy overlap shaded gray. For simplicity only eight processes are represented, and the sub-spaces are drawn as regular rectangles.

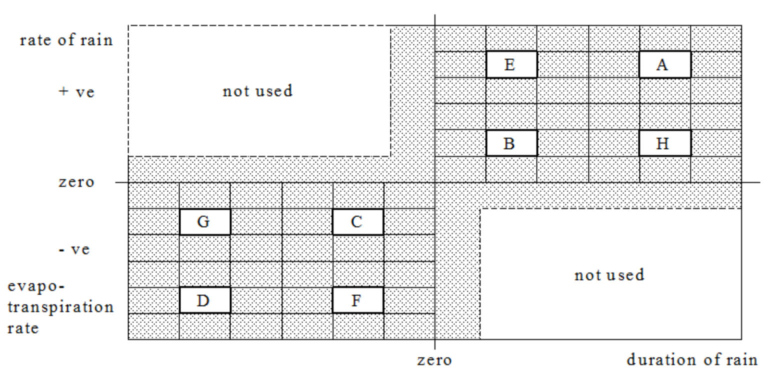


Figure 12.2 Fuzzy state-variable sub-spaces and their associated processes

A: flow over pervious areas; B: flow over impervious areas; C: recovery of storages and loss rates; D: pollutant accumulation; E: pollutant washoff; F: recovery of depression storage; G: groundwater depletion; H: rainout.

Some of the processes and the spaces where they dominate warrant further explanation. For light rains of short duration, runoff will occur from impervious but not pervious surfaces (sup-space B). This is because of our circular definitions – it will take heavier rain than this to satisfy infiltration capacities, and cause runoff from pervious surfaces. Thus infiltration processes dominate in sub-space B, but not A. On the other hand in sub-space A infiltration is satisfied and no longer dominates – the entire urban surface contributes runoff.

Definition of the Meteorological Types

This section describes the selection of different storm types (Kuch, 1997). Processes in SWMM (of which there are dozens) become dominant when their contribution to a selected objective function is larger than that of other processes. Consider for example the process of depression storage for an impervious area, which acts as an initial abstraction from rainfall (a depth). This process will be dominant when the amount of rainfall is less than, say, twice this depth, and its dominance is related to the rainfall input function as well as other physically-based parameters within the model.

Dominance is measured relative to a specific objective function, e.g. the total volume of runoff from pervious and impervious areas. The dominance will change based on environmental conditions. Its relative importance changes with every application.

Isolation and use of specific events within the continuous time series of precipitation is important because it is not possible to calibrate a continuous model without calibrating all the component processes individually. If all the processes are calibrated the model will produce an acceptable response throughout the simulation period. On the other hand, there is little confidence in a model calibrated only to a few high-intensity storms since the continuous input rainfall TS includes events of different intensity or duration, and will not produce responses that match observations if no effort was made to investigate or calibrate the model for such conditions.

For analysing parameter sensitivity at various state variable subspaces, 30 different synthetic storms were developed by Kuch (1997) and used for each of eleven overland flow parameters, two washoff parameters and the PCTZER parameter used in his Redhill data file. Synthetic storms of constant intensity were derived for six durations (0.25, 0.50, 1.0, 3.0, 6.0, 12.0 h) and five intensities (0.1, 0.25, 0.50, 1.0, 3.5 inch/h).

Additionally, the sensitivity of the REGEN parameter and both buildup variables was analysed using 30 different dry events that were synthesized using various periods of no rain and constant evaporation rates. These events were keyed into SWMM by bounding these time periods with fixed rainfall events so that the resulting flows and pollutant concentrations could be used for the objective functions. The synthetic time series in this case used five constant evaporation rates (0.05, 0.10, 0.2, 0.3, 0.5 inch/day) and six durations (0.25, 0.50, 1.0, 3.0, 6.0, 12.0 d).

In total seventeen parameters were changed by plus and minus 10% of their value using 30 different input time series. The total number of runs was 1080 including the 60 base runs that were performed to obtain the objective function values with no parameter adjustments (seventeen parameters times 30 events times two perturbations). The average sensitivity for each parameter was calculated using the relative change in the objective function for the 20% range. Peak flows and total runoff volumes were used for the objective functions for water quantity, and peak suspended solids (SS) concentration and SS load were used as the objective functions for water quality. Kuch (1997) calculated and plotted the relative sensitivity of each parameter was plotted by dividing the average parameter sensitivity for each storm by the largest average parameter sensitivity of any storm. Thus the subspace with the largest change in objective function plots as 1. If the parameter is insensitive for a given storm event the average sensitivity and the relative sensitivity plots as a zero. The relative sensitivity of each parameter was plotted against total storm volumes or evaporation totals, and separate graphs were prepared for peak flows or SS concentrations and total volume and SS load. These plots facilitate identification of regions where parameters are most sensitive, in other words, the domain of sensitivity is plotted using fuzzy logic concepts.

Storm intensities less than the minimum infiltration rate of 0.30 inch/h were designated as low intensity. This includes all of the 0.1 and the 0.25 inch/h storms. No overland flow from pervious areas was expected or found in any of these runs since the infiltration rate was not exceeded by the rainfall rate at any time in the simulations. Storm intensities greater than the minimum infiltration rate and less than the maximum of 3.0 inch/h were designated as medium intensity for this study. This includes both sets of the 0.50 and the 1.0 inch/h storms. The synthetic storms of 3.5 inch/h were designated as high intensity because this intensity generates runoff on the pervious areas for any event duration.

The synthetic storms of durations of 0.25 and 0.50 h were grouped as short duration. Medium duration storms were storms of 1 and 3 h duration and the storms of 6 and 12 h durations were considered to be long.

Similarly, evaporation rates of 0.05 and 0. 10 inch/d are denoted low, 0.2 and 0.3 are medium and 0.50 are high. Dry periods of 0.25 and 0.50 d were considered short, 1 and 3 d medium and 6 or 12 d long.

Selected Events for Calibration

A total of fifteen events were selected for calibration from the continuous input TS comprising three events of each of the following: SDLI for impervious area and washoff parameters, SDMI for routing parameters, LDMI for pervious area parameters, LDZI for buildup parameters, and MDZI for the REGEN parameter. Table 12.4 a and b summarize the events.

Table 12.4a: Events selected for calibration.

Date	Duration (h)	Average Intensity (in/h)	Depth (inches)	Group Label
9/5/56	1	0.150	0.150	SDLI
6/11/55	1	0.070	0.070	SDLI
7/8/55	1	0.060	0.060	SDLI
7/27/55	1	0.480	0.480	SDMI
5/17/56	1	0.350	0.350	SDMI
6/23/56	1	0.220	0.220	SDMI
4/28/56	4	0.588	2.350	LDMI
6/17/56	6	0.270	1.620	LDMI
8/13/56	14	0.133	1.860	LDMI
4/28/56	36+4	NA	2.350	MDZI
6/17/56	24+6	NA	1.620	MDZI
8/13/56	36+14	NA	1.860	MDZI
7/15/56	80.5	NA	0.590	LDZI
7/27/55	75	NA	0.480	LDZI
8/29/56	282	NA	0.240	LDZI

Table 12.4b contd.: Events selected for calibration.

Date	Observed Peak Flow, (cfs) SS Conc,(mg/1)	Observed Volume, (ft) Load, (lbs)
9/5/56	25.257	4.535E+5
(SS)	558.2	1.318E+4
6/11/55	28.773	9.416E+5
(SS)	305.7	8.878E+3
7/8/55	2.930	6.082E+4
(SS)	263.2	7.513E+2
7/27/55	110.766	1.886E+6
5/17/56	62.743	1. 18E3+6
6/23/56	43.684	7.627E+5
4/28/56	1680.589	2.940E+7
6/17/56	705.151	1.829E+7
8/13/56	804.228	1.462E+7
4/28/56	1680.589	2.940E+7
6/17/56	705.151	1.829E+7
8/13/56	804.228	1.462E+7
7/15/56		
(SS)	398.900	7.461E+4
7/27/55		
(SS)	191.000	4.634E+4
8/29/56		
(SS)	783.600	5.541E+4

A sample sensitivity family of plots is shown in Figure 12.3. Each parameter was adjusted plus and minus 10% of its value and the dataset run in the model for 30 different wet or dry events.

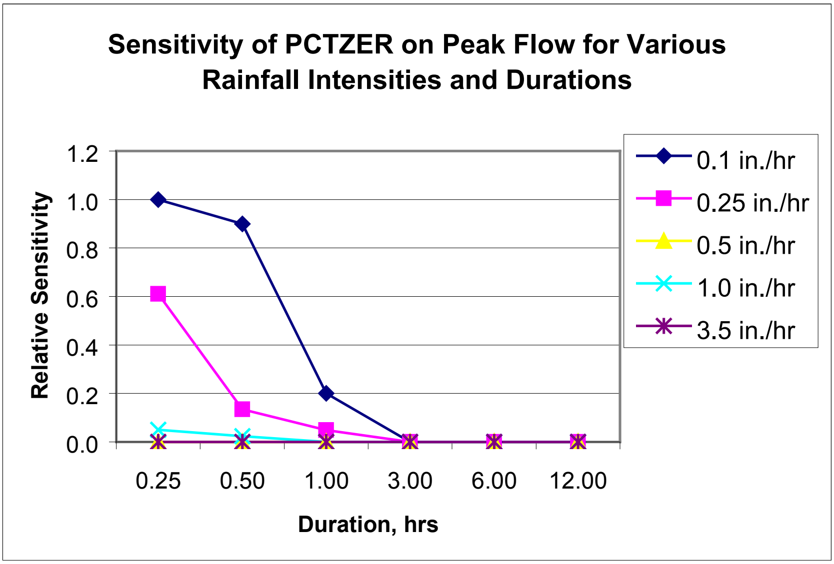


Figure 12.3 Typical sensitivity plot.

References

Kuch, A.W. 1997. Sensitivity abnalysis and calibration decision-support tools for continuous modelling with the storm water management model. MSc Thesis Univ of Guelph, ca 300pp.

Chapter 13

PERFORMANCE EVALUATION FUNCTIONS

Is it not strange that desire should so many years outlive performance? - William Shakespeare
All promise outruns performance. - Ralph Waldo Emerson
True genius resides in the capacity for evaluation of uncertain, hazardous, and conflicting information. - Winston Churchill
The best of them is but an approximation, while the worst bears nor relation whatever to the truth. – E.C.K Gonner
Sir, I would rather be right than be President. - Henry Clay

13.1 Introduction

In the literature, the tests for goodness-of-fit are usually called *objective functions*, and normally operate on the computed and observed hydrographs. In this Chapter we list some common functions that have been used to evaluate the goodness of fit between a set of computed and observed functions (Rivera-Santos, 1988).

In the methodology of this book, we introduce the idea of working with various simple statistics of the response functions (RFs). These measures are closely identified with the original design objectives, and so we call them objective functions (OFs). Now since we need to distinguish these measures from those that we use to determine the goodness-of-fit, we call the latter *performance evaluation functions* EFs. Our usage in this book differs from conventional - the distinctions are:

example	**conventional term**	**our usage here**
rain	input	input function, IF
runoff	output	response function, RF
peak flow	(no one term)	objective function, OF
standard error	objective function	evaluation function, EF

Few writers have reported on the problem of the best evaluation function for the calibration of long-term continuous water quality models. Han and Rao (1980) in their examination of event modeling using ILLUDAS found that the sum of the squared deviations between observed and computed flows was the most frequently used, and also gave the best overall performance.

Deyda (1993) states that the parameter optimization method should be chosen to suit the properties of the chosen evaluation function. For linear EFs, the simplex method is the most efficient. For non-linear EFs, one of many different search techniques can be used. Deyda optimized four parameters of an extremely simple quantity-only, event-hydrologic model, against three events, one fewer than the number of parameters sought.

In the list of EFs below, the computed objective function is *COF*, the observed objective function *OOF*, and the difference between them is the error:

$$e_i = OOF_i - COF_i, \quad 1 \le i \le n$$

where n is the number of points compared.

13.2 Survey of evaluation functions

In his study on the calibration and error analysis of PRMS, Rivera Santos (1988) reviewed many of the objective functions for calibration of precipitation-runoff models reported in the literature up till that time. Many in the following list were taken from the appendix to his thesis, but the notation and the emphasis has been changed here to suit our purposes. Some of the functions have been omitted, and others have been added from other sources.

1. Simple least squares (tends to favor large errors and large flows):

$$EF_1 = \sum_{i=1}^{n} (OOF_i - COF_i)^2 \tag{13.1}$$

2. First dimensionless form of simple least squares (tends to be independent of long records and also to favor large flows):

$$EF_2 = \sum_{i=1}^{n} \left(\frac{OOF_i - COF_i}{OOF_i} \right)^2 \tag{13.2}$$

3. Second dimensionless form of simple least squares:

$$EF_3 = \frac{\sum_{i=1}^{n}(OOF_i - \overline{OOF})^2 - \sum_{i=1}^{n}(OOF_i - COF_i)^2}{\sum_{i=1}^{n}(OOF_i - \overline{OOF})^2} \tag{13.3}$$

where $\overline{OOF}$ is the mean observed flow.

4. The logarithmic simple least squares (tends to overcome the bias of the previous EFs to large flows):

$$EF_4 = \sum_{i=1}^{n}(\ln OOF_i - \ln COF_i)^2 \tag{13.4}$$

5. Simple least squares of *m*-powered flows (tends to put more emphasis on lower flows than larger flows for m=1/2, 1/3; and more weight on larger flows for m>1):

$$EF_5 = \sum_{i=1}^{n}(OOF_i^m - COF_i^m)^2 \tag{13.5}$$

6. Simple least squares of the reciprocal flows (same as EF_5 where m = -1; tends to favor small flows):

$$EF_6 = \sum_{i=1}^{n}\left(\frac{1}{OOF_i} - \frac{1}{COF_i}\right)^2 \tag{13.6}$$

7. Weighted least squares:

$$EF_7 = \sum_{i=1}^{n} W_i (OOF_i - COF_i)^2 \tag{13.7}$$

The problem here is to select an appropriate weight. The US Army Corps of Engineers use the following weight in HEC-1:

$$W_i = \frac{OOF_i + \overline{OOF}}{2 \cdot \overline{OOF}} \tag{13.7a}$$

8. Mean weighted least squares (used in HEC-1):

$$EF_8 = \sum_{i=1}^{n} \frac{W_i (OOF_i - COF_i)^2}{n} \tag{13.8}$$

where W_i has the same meaning as above.

9. Root mean square errors:

$$EF_9 = \sum_{i=1}^{n} \left[\frac{(OOF_i - COF_i)^2}{n} \right]^{\frac{1}{2}} \tag{13.9}$$

10. Dimensionless form of root mean square error:

$$EF_{10} = \frac{\sum_{i=1}^{n} \left[\frac{(OOF_i - COF_i)^2}{n} \right]^{\frac{1}{2}}}{\overline{OOF}} \tag{13.10}$$

11. Standard error of estimate (Jewell, 1978, uses this criterion for the quantity portion of SWMM, a small value indicating a good calibration):

$$EF_{11} = \sqrt{\frac{\sum_{i=1}^{n} (OOF_i - COF_i)^2}{n - 2}} \tag{13.11}$$

12. Generalized least squares:

$$EF_{12} = \frac{1}{n} \mathrm{E}^{\mathrm{T}} \Omega^{-1} \mathrm{E} \tag{13.12}$$

where:

$$\mathrm{E} = \left[\mathrm{e_i} - \rho \mathrm{e_{i-1}} \right]^{\mathrm{T}} \tag{13.12a}$$

and the covariance matrix Ω is estimated from the sample covariance:

$$Cov\ (e_i, e_{i-1}) = \frac{1}{n}\sum_{i=2}^{n}([e_i - re_{i-1}][e_i - re_{i-1}]^{\mathrm{T}}) \tag{13.12b}$$

where r is the sample estimate of the lag-1 correlation matrix ρ:

$$r = \left[\frac{1}{n}\sum_{i=2}^{n}[e_i][e_{i-1}]^{\mathrm{T}}\right]\left[\frac{1}{n}\sum_{i=2}^{n}[e_i][e_{i-1}]^{\mathrm{T}}\right]^{-1} \tag{13.12c}$$

and e_i is the error OOF_i - COF_i

13. Normalized objective function:

$$EF_{13} = \frac{\sqrt{n\ \sum_{i=1}^{n}(OOF_i - COF_i)^2}}{\sum_{i=1}^{n}OOF_i} \tag{13.13}$$

14. Sum of the m-powered absolute errors:

$$EF_{14} = \sum_{i=1}^{n}|OOF_i - COF_i|^m \tag{13.14}$$

15. Sum of the m-powered absolute log errors:

$$EF_{15} = \sum_{i=1}^{n}|\ln OOF_i - \ln COF_i|^m \tag{13.15}$$

16. Dimensionless form of the sum of the absolute error:

$$EF_{16} = \frac{\sum_{i=1}^{N}|OOF_i - COF_i|}{\sum_{i=1}^{n}OOF_i} \tag{13.16}$$

17. Reduced error estimate:

$$EF_{17} = \sqrt{\frac{\sum_{i=1}^{n}(OOF_i - COF_i)^2}{\sum_{i=1}^{n}(OOF_i - \overline{OOF})^2}} \qquad (13.17)$$

18. Proportional error of estimates:

$$EF_{18} = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{OOF_i - COF_i}{\overline{OOF}}\right)^2}{n}} \qquad (13.18)$$

19. Composite absolute log error:

$$EF_{19} = \sum_{i=1}^{n} |\ln(OPF_i) - \ln(CPF_i)|^2 + \frac{\sum_{i=1}^{n} |\ln(OVF_i) - \ln(CVF_i)|^2}{2} \qquad (13.19)$$

where: *OPF, CPF, OVF* and *CVF* are the observed and computed peak and volumes of flow respectively.

20. Composite weighted sum of least squares (used by Rao and Han, 1981, to calibrate urban storm water models):

$$EF_{20} = \sum_{i=1}^{n} W(OPF_i - CPF_i)^2 + (1 - W)(OVF_i - CVF_i)^2 \qquad (13.20)$$

21. Composite least squares:

$$EF_{21} = \sum_{i=1}^{n}[e_i]^{\mathrm{T}}[e_i] + \lambda \sum_{i=1}^{n}[\Theta_i - \overline{\Theta}]^{\mathrm{T}} \rho_e^{-1}[\Theta_i - \overline{\Theta}] \qquad (13.21)$$

where λ is a weighting factor, $\overline{\Theta}$ is the mean value of the parameters Θ, and $\boldsymbol{\rho_e}$ is the covariance matrix of the parameters.

22. An *ad hoc* function proposed by Ibbitt and Hutchinson (1984) to accurately estimate 30-day and total water balances:

$$EF_{22} = \sum_{j=0}^{\frac{n-30}{30}} \left[\frac{\left| \sum_{i=30j+1}^{30J+30} OOF_i - \sum_{i=30j+1}^{30j+30} COF_i \right|}{\sum_{i=30j+1}^{30j+30} OOF_i + \sum_{i=30j+1}^{30j+30} COF_i} \right] + \frac{\sum_{i=1}^{n} OOF_i - \sum_{i=1}^{n} COF_i}{\sum_{i=1}^{n} OOF_i}$$

(13.22a)

23. First lag auto correlated maximum likelihood estimate:

$$EF_{23} = \frac{1}{2}\left[n \ln(2\pi) + \ln\left(\frac{\sigma_v^{2n}}{1-\rho^2}\right) - \rho^2 \sigma_v^2 e_1^2 - \sigma_v^{-2} \sum_{i=2}^{n} (e_i - \rho e_{i-1})^2 \right]$$

(13.22b)

where σ_v^2 is the variance of the random portion of the errors and is given by:

$$\sigma_v^2 = \frac{1}{n}\left[-\rho^2 e_i^2 + \sum_{i=2}^{n} (e_i - \rho e_{i-1})^2 \right] \quad (13.23a)$$

an d ρ is the first lag correlation coefficient, which can be estimated from the implicit equation:

$$A\rho^3 + B\rho^2 + C\rho + D = 0 \quad (13.23b)$$

where:

$$A = e_i^2 - \sum_{1=2}^{n} e_{i-1}^2 \quad (13.23c)$$

$$B = \sum_{i=2}^{n} e_i \, e_{i-1} \quad (13.23d)$$

$$C = \sigma_v^2 - e_1^2 + \sum_{i=2}^{n} e_{i-1}^2 \quad (13.23e)$$

$$D = -\sum_{i=2}^{n} e_i \; e_{i-1} \quad (13.23f)$$

24. Heteroscedastic maximum likelihood estimate:

$$EF_{24} = \frac{\sum_{i=1}^{n} W_i e_i^2}{n \left[\prod_{i=1}^{n} W_i \right]^{\frac{1}{n}}} \tag{13.23g}$$

where W_i is the weight at the *i*th time interval given by:

$$W_i = flow_i^{2(\lambda-1)} \tag{13.24a}$$

where λ is the unknown variance stabilizing transformation parameter and estimated from the implicit equation

$$\left[\sum_{i=1}^{n} \ln(f_i) \right] \left[\sum_{i=1}^{n} W_i e_i^2 \right] - n \left[\sum_{i=1}^{n} W_i \ln(f_i) e_i^2 \right] = 0 \tag{13.24b}$$

where $f_i = E\ (flow_i)$ is the expectation of the "true" flow.

For more details on EF_{23} and EF_{24}, readers are referred to the publications by Sorooshian (1978) and Sorooshian et al. (1981). Other reviews of objective functions are provided by Han and Rao (1980), Green and Stephenson (1986) and Diskin and Simon (1977).

25. Baffaut (1988) sequentially ranked the objectives of her modeling effort, considering that: firstly, the runoff volumes should be correct, on the average; secondly, the peak discharges should match, on the average; thirdly, the times-to-peak should be correct on the average; and lastly, the root mean square of the hydrograph shape should match on the average. To avoid the problem of an evidently poor evaluation being computed for a relatively small error in timing, but for an otherwise good fit, the standard square error of the shape was weighted 0.7, and the volume 0.3.

The equations used are:

$$volume\ difference = \frac{1}{N} \sum \frac{V_m - V_p}{V_m}$$

$$peak\ difference = \frac{1}{N} \sum \frac{P_m - P_p}{P_m}$$

$$time\ difference = \frac{1}{N}\sum\left(T_m - T_p\right)$$

$$weighted\ error = \frac{1}{N}\sum a\frac{\sqrt{\sum_{j=1}^{N}\left(m_j - p_j\right)^2}}{m_i} + b\left|\frac{V_m - V_p}{V_m}\right|$$

where:

V is the volume;
P is the peak flow;
T is the time-to-peak;
m and p refer to observed and computed;
N is the number of storms;
N_i is the number of data for event i;
m_j and p_j ($j = 1,.., N_i$) are the observed and computed flows of the ith hydrograph, and
m_i is the average flow for event i.

26. Warwick and Wilson (1990) used a total error statistic (EF_t) similar to EF_{20} to quantify overall goodness of fit:

$$EF_t = (1.0 - W)\left[\sum_{i=1}^{n}\frac{(COF_i - OOF_i)^2}{n}\right]^{\frac{1}{2}} + (W|OPF_p - CPF_p|)$$

where:

EF_t	=	total error statistic (m^3/s);
W	=	weighting factor;
n	=	number of measured hourly flows;
OOF	=	measured flow (m^3/s);
COF	=	computed flow (m^3/s);
OPF	=	measured peak flow (m^3/s); and
CPF	=	computed peak flow (m^3/s).

Assigning a value of 1.0 to W results in a complete focusing upon matching the peak flow. Reasonable balances between hydrograph shape and peak were obtained with $W = 0.2$.

27. The so-called Nash coefficient:

$$EF_{27} = 1 - \frac{\sum_{i=1}^{n}\left(COF_i - OOF_i\right)^2}{\sum_{i=1}^{n}\left(OOF_i - \overline{OOF}\right)^2}$$

28. Another measure of goodness of fit is the correlation coefficient which is independent of the number of observations and their uncertainty, and allows results of different parameter exercises to be directly compared (Doherty, 2001).

$$EF_{28} = \frac{\sum_{i=1}^{n}\left(w_i OOF_i - \overline{wOOF}\right)\left(w_i COF_i - \overline{wCOF}\right)}{\left[\sum_{i=1}^{n}\left(w_i OOF_i - \overline{wOOF}\right)^2 \sum_{i=1}^{n}\left(w_i COF_i - \overline{wCOF}\right)^2\right]^{1/2}}$$

Generally for the fit to be considered acceptable, we should have

$$EF_{28} > 0.9$$

(see Cooley and Naff, 1990, and Hill, 1998.)

13.3 Concluding remarks

Gan et al. (1997) and Lamb (1999) found that the evaluation function is an important element when automatically calibrating models. Gan (1997) found, where high flows are important, a least squares procedure is preferable to a maximum likelihood method, and Rivera-Santos (1990), using PRMS to compare various calibration procedures, found that the least squares procedure yielded better estimates when the serial correlation of the errors is taken into account. He used sensitivity analysis to select parameters and to estimate errors, and recommended that the error structure be taken into account when choosing the calibration procedure. Redundant parameters and non-linear parameter interaction caused poor performance of the optimization algorithm. Lamb (1997) recommends careful consideration of the evaluation function, using one that favors high flows if these are important.

Han and Rao (1980) showed that, for SWMM model calibration purposes, of a total of seventeen objective functions reviewed, the sum of the squared deviations between the observed and computed flows EF_1 gave the best overall performance. Cooper et al. (1997) like others before them studied several evaluation functions and found simple least squares to be best but opted for the equally good, so-called Nash coefficient, EF_{27} . Javaheri (1998) also used EF_{27} (but did not caution that his results might be influenced by this choice).

In general, an evaluation function incorporating a weight factor might make a useful algorithm if the weight factor can be chosen to suit the design problem being solved - for storage elements, it should favor flow volume, for conveyance structures, it should favor peak flow rates.

Using multiple objectives, the problem of having many equally valid but quite different optimal parameter sets may be solved (Lamb, 1999). This being so, and there being no clear rules for selecting the best evaluation function, we have provided in PCSWMM a wide selection of evaluation functions from the above list.

> Use several different evaluation functions, especially where several design objectives are involved. For one of them, consider using simple least squares, because of its historical associations.

References

Baffaut, C. 1988. Knowledge techniques for the analysis of urban runoff using SWMM. Ph.D. thesis. Purdue U. 321 pp.

Cooley, R.L. and Naff, R.L. 1990. Regression modeling of ground-water flow. USGS Techniques in Water Resources Investigations, Book 3, Chap. B4, 232p.

Cooper, V. A., Nguyen, V. T. V., and Nicell, J. A. 1997. Evaluation of global optimization methods for conceptual rainfall-runoff model calibration. Water Science and Tech., 36(5):53-60.

Deyda, S. 1993. Automatic calibration of runoff models with the help of mathematical optimization methods. 6th Int. Conf. Urban Storm Drainage, Niagara Falls, Ontario. Sept. 1993. pp 838-843.

Diskin, M.H. and Simon, E. 1977 A procedure for the selection of objective functions for hydrologic simulation models. J. of Hydrology, 9:34-56.

Gan, T.Y., Dlamini, E.M., and Biftu, G.F. 1997 Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. J. Hydrology, 192:81-103.

Green, I.R.A. and Stephenson, D. 1986 Criteria for comparison of single event models. Hydrologic Sciences J., 31(3):395-411.

Han, J. and Rao, A. Ramachandra. 1980. Optimal parameter estimation of objective functions of urban runoff models. Tech. report No 135. Water Resources Research Center, Purdue University. Sept. 1980

Hill, M.C. 1998 Methods and guidelines for effective model calibration. USGS Water Resources Investigations Report 98-4005.

Javaheri, J. 1998. Automatic calibration of urban runoff models using global optimization techniques. M.Eng, thesis, McGill U., Canada. 119p.

Jewell, T.K., Nunno, T.J., and Adrian, D.D. 1978. Methodology for calibrating stormwater models. Proc. Stormwater and water quality model users group meeting, May 4-5 1978. USEPA 600/9-78-019 pp 125-169.

Lamb, R. 1999. Calibration of a conceptual rainfall-runoff model for flood frequency estimation by continuous simulation. Water Rsources Research 35(10) 3103-3114.

Doherty, J. 2001. PEST documentation, 2001**,** Published by Watermark Numerical Computing, downloaded from the PEST website at www.ozemail.au.

Qaisi, K.M. 1985. Uncertainty analysis of stream dissolved oxygen models. Ph.D. thesis. U. of Michigan Ann Arbor. 234 pp.

Rivera Santos, J. 1990. Parameter estimation in conceptual precipitation-runoff models with emphasis on error analysis. Proc of Int. Symp. on Tropical hydrology and 4th Caribbean Islands water resources congress, San Juan, Puerto Rico, Jul 22-27, 1990. Amer. Water Resources Assoc. pp91-100.

Rivera Santos, J. 1988. A methodology for parameter estimation in conceptual precipitation-runoff models with emphasis on error analysis. Ph.D. diss. U of Colorado Boulder. 279p

Sorooshian, S., **1978**: Considerations of Stochastic Properties in Parameter Estimation of Hydrologic Rainfall-Runoff Models. Ph.D. Dissertation, University of California, Los Angeles, CA, pp.248.

Thompson, D.B. 1989. Determining parameters for a continuous simulation model by estimation, synthetic calibration and analytic calibration. Ph.D. thesis. U. of Missouri-Rolla. 261 pp.

Warwick and Wilson 1990 source misplaced.

Yan, J. 1990. Parameter estimation for multipurpose hydrologic models. Ph.D. thesis. Oklahoma State U. 184 pp.

Chapter 14

PARAMETER OPTIMIZATION AND PROCESS CALIBRATION

You see, one thing is, I can live with doubt and uncertainty and not knowing. I think it's much more interesting to live not knowing than to have answers which might be wrong. I have approximate answers and possible beliefs and different degrees of uncertainty about different things, but I am not absolutely sure of anything and there are many things I don't know anything about, such as whether it means anything to ask why we're here. – Richard P. Feynman

14.1 Introduction

Model calibration is important. Only about ten years ago, some consultants argued that calibration constituted *research* whereas *design* merely required intelligent parameterization ("guestimation"). But today calibration is an expected part of engineering modeling. Calibration is more compute-intensive than sensitivity analysis. The purpose of the procedure is quite simple: find "best" values for the difficult parameters. Start calibration with the most sensitive parameters, proceeding towards the less sensitive. Insensitive parameters and their associated processes should be zeroed out of the data file and the study – one should not waste time and money collecting field data for relatively dormant processes. Procedures are indicated in Figure 14.1

In the Figure the upper part relates to event analysis, and the lower to continuous modeling. Parameters are optimized on the basis of events, do not attempt to calibrate the whole dataset for a long continuous model. Calibrate just the dominant processes: snowmelt during snowmelt, infiltration during medium rains, recovery of infiltration capacity during dry events. Only calibrate processes for and during the events when and where those processes are dominant.

Collect up all rainfall and dry events in your dataset and calibrate parameters that are shown to be sensitive for each and every type of event. Collect up all the light rainfall events and calibrate the parameters for impervious areas. Collect up all the dry weather events and calibrate recovery parameters. Collect up all the snowmelt events and calibrate snowmelt parameters.

Much time can be spent on a calibration effort. Only when all pertinent processes are calibrated, can they be run in a continuous model as shown in Figure 14.1. Disregard the EPA manual for HSPF where it recommends calibrating against long-term records, because it is impossible, unnecessary and a waste of time trying to obtain long-term continuous records for all quantities of interest (e.g. 100 y of 1-minute data for microbiology, to belabor the point).

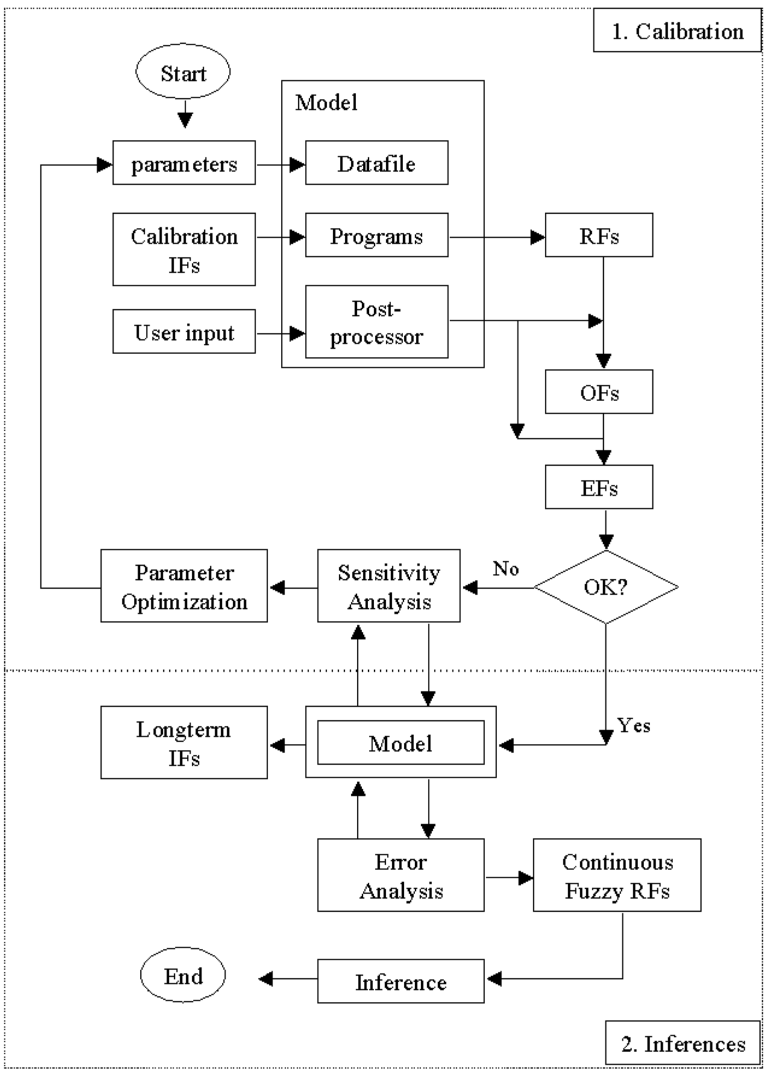


Figure 14.1 Modeling calibration procedures.

In urban runoff quantity modeling, calibration can be comparatively simple: e.g. computed peak flow is plotted against observed peak flow for a number of different rainfall events, typically about twenty. Examples are presented later in this Chapter. If the model is good, the points should fall on a 45-degree line with little scatter. If the plot shows a systematic variation in the output, in other words if the curve fitting the plotted points deviates from the line of equality, the data is said to be biased. The process of calibration is to best-fit the line of equality, and to reduce the scatter (improve the model precision).

The points nearer the origin are for low rainfall rates where the impervious parameters are important; the events around the middle relate to intermediate rainfalls, where infiltration may be significant; and the heavy rains relate to the parameters for overall slopes, areas and so on.

The next question that arises is: why calibrate peak flows instead of, say, the entire hydrograph shape? The answer to this question depends on the original design problem to be resolved. If the volume of a detention basin is to be optimized, the volume of runoff may be important. Short, sharp peak flows may not be important. Calibrating the parameters that govern low flows may be a waste of time. Consequently, fitting at the outset of the calibration exercise the entire long-term hydrograph seems unnecessarily demanding. Evidently effort can be reduced by attending more to those parts of the response hydrograph that relate directly to the design problems being resolved.

Another important question that must be addressed is: where we have many lines of input data all with similar parameters that have been derived systematically, is it essential that each and every parameter be separately and independently optimized? For instance, sub-catchment data contain the same geophysical parameters on each line, and there may be 1000 lines of such data. Once again the answer depends on the problem, for this example it could depend on the spatial distribution of the subcatchments and their relative dominance of the computed response. Sensitivity analysis is the key. A large, well-drained impervious sub-catchment near the problem outlet might merit special calibration, as a sensitivity analysis would show.

But in general it would be impractical to separately calibrate each parameter in a very large array, because, in the first place, the relative sensitivity of parameters in a finely discretized data file is small, and secondly, it would be too onerous to separately optimize hundreds of thousands of parameters. We consider instead that the method of parameter estimation initially used is itself the subject of optimization, by optimizing the entire array of (say) 1000 similar parameters, e.g. widths of sub-catchments. An optimum correction factor is applied to all 1000 widths.

Points made in the last three paragraphs make it possible to calibrate very large and complex urban runoff models.

14.2 Model validation

Model validation is an extension of model calibration, supposedly testing the optimum parameter set against an independent set of field observations not used in the calibration procedure. If the parameters are then further modified using the additional information, the procedure becomes essentially the same as calibration. If the parameters are not modified, the procedure becomes tantamount to deliberately ignoring the additional information. In practice, modelers either use the complete set of event data for calibration in the first case, or use part of the observed data set for calibration and the rest for validation. For example, use 1990 storms and runoff events for calibration and then run the 1991 dataset for validation. Personally I usually simply calibrate against the entire dataset, providing of course that the landscape itself has not undergone any interventions that are not reflected in the input data file.

14.3 When is a model "sufficiently calibrated"?

Whether a model may or may not be considered sufficiently calibrated depends in part on the intended use of the model (e.g., planning or operation). There is no standard for the adequacy of calibration. For some users a 25% accuracy is adequate while others insist on 10%. The Integral Square Error (ISE) is one measure of goodness-of-fit between observed and modeled responses. Calibration and verification costs money, and is the most painful phase of any modeling project. Also, the cost of collecting calibration data is substantial, perhaps more than the modeling per se.

Three response hydrograph statistics are normally analyzed: 1. volume, 2. peak flow and 3. time-to-peak. One of these statistics could be more important in a given specific situation:

Volume: becomes predominant when retention is considered.

Peak discharge: For conduit sizing and estimating the remaining available capacity in existing systems.

Time-to-peak: Could be important in a network when large branches combine together, such as in combined sewer overflow regulators. If the times to peak are not evaluated, it could have an important impact on the total peak discharge in the downstream branch.

If the end result is to determine the diameter of a pipe, what is the required accuracy? Using standard pipe sizes, the incremental difference in capacity may be as low as 10% or as great as 200% depending on the diameter that is being used as the basis (assuming of course that slopes are equal).

In model calibration contracts, there is often a stated accuracy level of calibration. Typically:

A. for Dry Weather: +/- 5% of volume and +/-10% of peak, and

B. for Wet Weather: +/- 10% of volume and +/- 15% of peak

A group in the UK (WaPug) recommends that the number of events be three significant events - meaning big enough to cause surcharge. The comparison should then be on four criteria.

1. Model peak flow +25% to -15%
2. Model volume of flow +20% to -10%.
3. Model depth of surcharge +0.5m to -0.1m
4. General shape of the hydrographs should be similar.

Calibration tolerances need to consider inherent inaccuracies in the observed data, as discussed earlier. Typical portable depth and velocity sewer flow meters have flow measurement errors that can range to greater than 20%. Careful field work can minimize these errors. Stream flow measurements can be more problematic. Only the reliable events should be used for calibration, e.g. no more than the top 5% of the total. Also the modeler should set upper and lower bounds on parameters during the calibration process, to ensure that estimated values are in range.

PCSWMM uses the following graphssome of which are illustrated later to indicate the success of the calibration:

calculated vs. observed values (line graph and scatter graph)
parameter values (line graph)
parameter sensitivities (line graph and bar chart)
objective function (line graph)
calibration residuals (bar chart)

In addition PEST (Kuch, 1994) supplies:

parameter correlation coefficient matrix
normalized eigenvectors of the covariance matrix

14.4 Sensitivity-based calibration

This part of the book draws from the work by Benny wan (2001). Trial and error has been the conventional way to calibrate models. However, accurate and satisfactory results were rarely obtained. So various calibration methods and algorithms have been developed to improve the efficiency of determining the "best" parameter sets. For automating the calibration of the runoff model, a sensitivity-based system called the "knowledge-based calibration system" was developed by Liong et al., (1991) for a watershed in Singapore. It comprised three key components:

1. A sensitivity study of calibration parameters.
2. A strategy for parameter selection and assigning values so as to achieve a match between simulated and measured hydrographs.
3. Knowledge representation scheme used to implement the calibration.

Wan (2001) developed a similar sensitivity-based calibration procedure based on the synthesis of four components: (1) sensitivity analysis, (2) setting calibration parameter limits, (3) genetic algorithm calibration strategy and (4) error analysis. We have adopted Wan's methodology in PCSWMM and in this book. Figure 14.2 outlines the steps in the method.

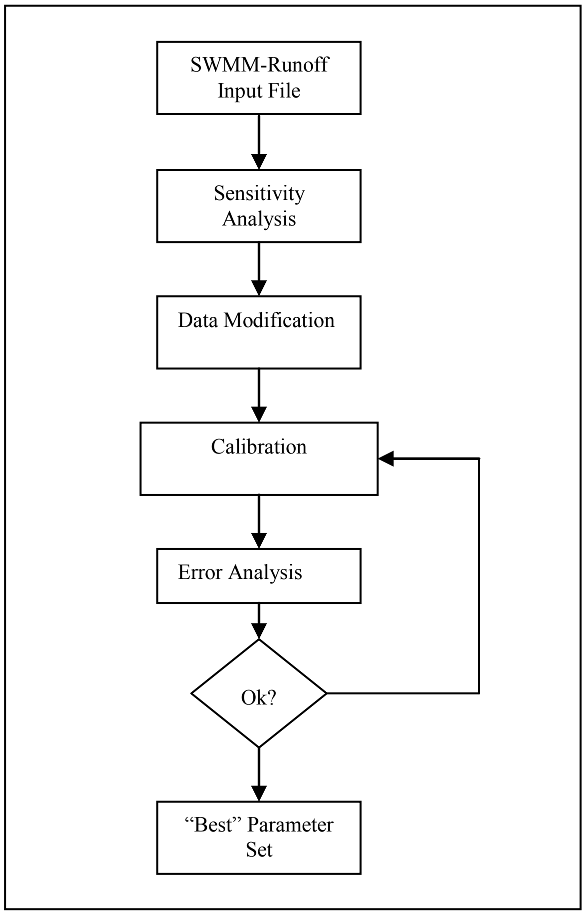


Figure 14.2: Sensitivity based Genetic Algorithm Calibration

14.5 Robustness and Efficiency

According to Thyer et al., (1999) robustness is related to the probability of finding the same global optimum from a series of independent trials. Efficiency is the number of function evaluations (model runs per optimization trial) required by the algorithm to satisfy prescribed convergence criteria. To evaluate the calibration algorithm the robustness and efficiency should be identified during the analysis.

14.6 Limiting Values of Parameters

For a given computed response, an infinite number of parameter sets may be found for the causative input data, especially where parameters are correlated, as indeed many are. Thus errors in determining best parameter sets are potentially enormous. In order to confine these errors, determination of the limiting values of calibration parameters is critical. The calibration method should not allow parameters to exceed the maximum or be less than their minimum physically-meaningful limits. Careful determination of the limits of each calibration parameter increases the accuracy and efficiency of the calibration process. Liong and Ibrahim (1995) limit their calibration parameters to ± 50% of the default value, which, as a blanket value for every calibration parameter, seems to be arbitrary and unnecessarily wide.

In general, the input data required in SWMM is quite substantial and detailed. SWMM RUNOFF parameters can be variously categorized, for example PCSWMM uses the categories: hydrologic, hydraulic, water quality, snow and erosion. We take the hydrologic parameters to include overland flow, infiltration, evaporation and snowmelt. Hydraulic parameters include physical properties of different channels and pipes. For efficient calibration, first estimates of the model parameters should be as close as possible to the true value. Thus a realistic and physically meaningful range of the model parameters should be identified. Limiting values of a few SWMM RUNOFF parameters are suggested below.

A. Hydrologic Parameters

Subcatchment Area (WAREA) Areas of catchments for drainage are derived from topographic maps. In practice, lack of detailed contour information and the unknown inflows and outflows may cause this parameter to become more difficult to obtain. However, improved GIS technology will eventually minimize this margin of error. Hence, ± 2% of the measured data is thought to be an acceptable range. Suggested limit: ± 2% of the measured data.

Subcatchment Width (WW1) Width of the subcatchment from which overland flow is derived. Overland flow on the subcatchment is assumed to derive from an idealized rectangular subcatchment. In practice, the subcatchment will not be rectangular with properties of symmetry and uniformity (James & James, 1999). In order to systematically estimate the width of the subcatchment, the measured area of the subcatchment is divided by the length of the longest flow path. The error in measuring the area of the subcatchment is approximately ±2%. Due to the high uncertainty of fixing the length of the flow path from upstream to downstream, uncertainty of estimating the length is perhaps ±15%. Accordingly, a range of -15% to 41% is calculated. Suggested limit: -15% to 41% of the measured data.

Percent Imperviousness (WW3) Imperviousness is defined as the percentage of the area which is impervious and hydraulically directly connected to the outlet, such as driveways and rooftops with downspouts (James & James, 1999a). Imperviousness can be estimated by dividing the directly connected impervious area by the total subcatchment area. In reality, runoff that flows over the driveway and rooftops will not completely drain into the sewer due to the retention of rain water on the road surface and the rooftops. As a result, an error of 12% might be appropriate. Suggested limit: ± 12% of the measured data.

Manning's Roughness Coefficient (WW5) Manning's roughness coefficient, n, is one of the parameters used to calculate overland flow. Suggested limits are shown in Table 14.1. Due to the high variability of Manning's roughness coefficient for different types of ground cover, the limit and range of this parameter is very difficult to estimate. The values shown in Table 14.1 are taken to be the upper and lower limits of this parameter.

Impervious Area Depression Storage (WW7) Impervious area depression storage is defined as water that is stored in depressions on impervious areas, depleted only by evaporation. Depression storage varies according to soil type, subcatchment slope and pavement. Many empirical equations have been developed to model depression storage. The upper and lower limit of impervious area depression storage is taken from the PCSWMM user's manual, which was based on experimental data obtained by Kidd (1978) in Europe. Suggested limit: 0.13 mm to 1.50 mm

Pervious Area Depression Storage (WW8) Pervious area depression storage is defined as water stored in depressions on pervious areas and is subject to infiltration and evaporation. Pervious area depression storage has the same properties as impervious area depression storage, which is mainly dependent on soil type and slope. The upper and lower limit of this parameter is taken from different experimental results cited in the PCSWMM user's manual. Suggested limit: 1.5 mm to 6.5 mm

Subcatchment Slope (WSLOPE) Slope of each individual subcatchment. Subcatchment slope is the average slope along the pathway of overland flow to the inlet (node). Subcatchment slope is estimated by dividing the change in elevation by the length of the flow path. Typical change in elevation is approximately 1 to 10m. In addition, the maximum error that would occur in estimating the elevation on a contour map is about ±1m (assuming 1 m contour intervals) (James and James, 1999a). The error in change in elevation of the subcatchment is assumed to be ±10%. By using the aforementioned error for the length of the flow path in the previous section, the upper and lower limit of this parameter is calculated. Suggested limit: -12% to 29% of measured data.

B. Hydraulic Parameters

Channel / Pipe Length (GLEN) Length of different types of open channel or pipe in either meters or feet. Error might occur in field measurement, however, the accuracy of measuring this parameter would not be very low. Suggested limit: ±10 cm of the measured data

Width/Diameter of Channel/Pipe (GWIDTH) Width of an open channel or diameter of the pipe. Errors might occur in measuring this parameter in the field, and would not be large. ±0.25cm for open channel and ±5mm for pipe is suggested. Suggested limit: ±0.25m (open channel) and ±5mm (pipe)

Invert Slope (G3) Inverted slope of channel or pipe. Explanation of this parameter is same as that for subcatchment slope. Suggested limit: -12% to 29% of measured data.

Manning's Roughness Coefficient for Channel / Pipe (G6) Manning's roughness coefficient, n, is one of the parameters used to calculate flow in a pipe or open channel. Limiting values of this parameter were taken from a water resources handbook (Mays, 1996). Covered earlier in this book is a lengthy argument on why the range used in your model should be narrower than the limite suggested below.

Suggested limit for open channel: much less than 0.009 to 0.200

Suggested limit for pipe: much less than 0.011 to 0.026

Table 14.1: Range of Manning's n

Ground Cover	n	Range	% Range
Concrete or asphalt	0.011	0.01 – 0.013	-9 to 18
Bare sand	0.01	0.01 – 0.016	0 to 60
Graveled surface	0.02	0.012 - 0.03	-40 to 50
Bare clay-loam (eroded)	0.02	0.012 - 0.033	-40 to 65
Range (natural)	0.13	0.01 - 0.32	-0.92 to 146
Bluegrass sod	0.45	0.39 - 0.63	-13 to 40
Short grass prairie	0.15	0.10 - 0.20	-33 to 33
Bermuda grass	0.41	0.30 - 0.48	-27 to 17

14.7 Automatic Calibration

Automatic calibration is becoming increasingly used for routine calibration of conceptual rainfall-runoff (CRR) models. The goal is to determine a feasible and unique global set of parameter estimates which produce computed responses that best fit the observed runoff response. Most if not all optimisation algorithms have difficulty in locating the global optimum due to response surfaces that contain multiple local optima with regions of attraction of differing size, discontinuities, and long ridges and valleys (Thyer et al., 1999). Over the last ten years extensive research has been undertaken to develop efficient and robust global optimisation algorithms. Multi-Extremum (Global) Optimization, Down-Hill Simplex (DHS), Shuffled Complex Evolution (SCE), and Simulated-

Annealing Algorithms (SimA) are reviewed here. GA methods are presented in more detail later.

Solomatine (1995 – see URL3) presented a multi-extremum (global) optimization method to calibrate the evaluation function (model error). In his optimization method, nine global optimization (GO) methods were compared in terms of effectiveness (accuracy), efficiency (number of needed function evaluations) and reliability. The nine GO methods that he used are:

1. Set (space) covering methods.
2. Pure direct random search (uniform sampling).
3. Controlled random search (CRS).
4. Evolutionary strategies and GAs (GA).
5. Multistart and clustering.
6. Adaptive cluster covering (ACCO).
7. ACCOL strategy - combination of ACCO with the multiple *local searches.*
8. ACD algorithm - a random search algorithm that combines with ACCO and downhill simplex descents (DSD).
9. ACDL algorithm - combination of ACD and multiple local searches.

A PC-based optimization tool *Globe* incorporating all nine GO algorithms was used to perform the optimization. Results were:

1. ACCOL and CRS4 have the highest effectiveness, efficiency and reliability.
2. ACDL is efficient and effective on some of the runs with functions of higher dimensions.
3. M-simplex (multi-start algorithm) works well with functions of low dimension.
4. Other algorithms provide reasonable solutions

In general, the reliability and accuracy assessment of each algorithm is strongly dependent on the type of problem, e.g., GO is better for calibration of pipe network problems. The assessments obtained by Solomatine are based on the overall performance of the calibration (effectiveness and efficiency) of each algorithm.

Down-Hill Simplex Method. The Down-Hill Simplex (DHS) Method was first developed by Nelder and Mead in 1965. It is a local and direct search algorithm that has long been used for calibration of conceptual rainfall-runoff models (Javaheri, 1998). The basic concept of DHS is that it searches the search space repeatedly and replaces the point that has the highest function value in a simplex with another point. When combined with other operations, the simplex under consideration adapts itself to the local landscape, elongating down long inclined planes, changing direction on encountering a valley at an angle, and contracting in the neighborhood of a minimum. There are four important steps involved in DHS method: *Reflection, Contraction, Expansion,* and *Shrinkage*. Figure 14.3 shows an example of DHS optimization (see http://neural.cs.nthu.edu.tw/jang/courses/cs4601/simplex.htm, 2001).

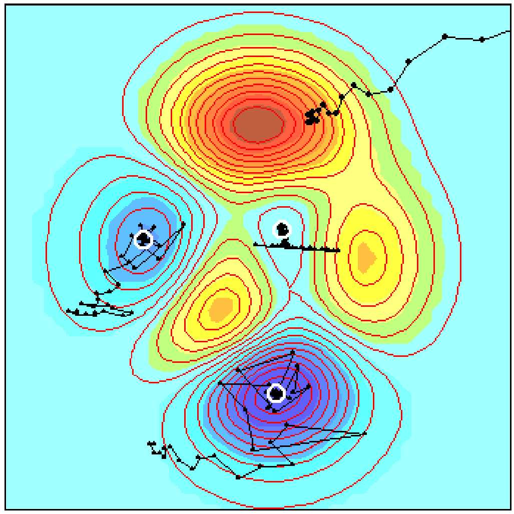


Figure 14.3: Example of Simplex Search

The advantages of using the Simplex Search are:

1. It is derivative-free.
2. It has an interesting geometric interpretation.

The disadvantages are:

1. It is slow.
2. It is not stochastic.
3. The underlying heuristics are only good for continuous optimization.

Shuffled Complex Evolution. The general Shuffled Complex Evolution (SCE) optimization technique was developed at the University of Arizona by Duan et al., (1992). The technique locates a unique global optimum within the random set of points, called the "population". SCE is designed to handle the various response surface problems encountered in the calibration of non-linear simulation models, particularly the multi-level problem encountered with urban runoff models (Javaheri, 1998). SCE is based on four main concepts:

1. combination of probabilistic and deterministic approaches,
2. systematic evolution of a "complex" of points spanning the space in the direction of global improvement,
3. competitive evolution, and
4. complex shuffling (Thyer et al., 1999).

Initially a population is sampled from the parameter space and partitioned into a number of complexes. Each of these complexes is allowed to evolve by using competitive evolution techniques, which are mainly based on the simplex method. In order to ensure information sharing in the evolution process, the population is completely shuffled and reassigned to new complexes during periodic stages. It is important to ensure that the population size is sufficiently large in the search operation that the entire population will move toward the neighborhood of the global optimum. SCE uses the strength of the local optimization simplex procedure to locate the global optimum. With the idea of competitive evolution and complex shuffling, SCE can ensure that the information of the parameter space gained by each of the individual complexes is shared throughout the entire population. Thyer et al., (1999) found that SCE conducts a far more effective search of the entire parameter space. Implementation of SCE is much simpler and more applicable under certain conditions. The overall performance of this optimization technique is good in term of robustness and efficiency.

Simulated-Annealing Algorithm. The Simulated-annealing (SimA) algorithm (Thyer et al., 1999) is a three-phase algorithm that combines simulated annealing with the simplex method of Nelder and Mead (1965). It was first proposed by Kirkpatrick et al. (1983) and originally developed for minimizing multivariate functions. The general idea of SimA is to explore the parameter space by randomly perturbing the value of the evaluation function at the simplex vertices. In the perturbation of the objective function, the acceptance of uphill as well as downhill steps on the response surface can be enabled. The magnitude of the random perturbation is determined by one control parameter, T. T is analogous to temperature in the physical annealing processes. It is decremented from its original value to zero in order to determine the optimum point. The rate of decrement of T is controlled by an annealing schedule. During the process of searching for a candidate global optimum, it is possible to escape from local minima by using stochastic step acceptance criterion (Thyer et al., 1999). As T is decreased slowly, a gradual reduction in the probability of accepting an uphill move will eventually occur. More detailed explanations of the SimA algorithm are given by Bates (1994) and Sumner et al. (1997). Thyer et al., (1999) concluded that the SimA algorithm must be tuned for different conditions and characteristics of catchments and the parameterization level at hand. However it can still efficiently determine the optimum value.

14.8 Genetic Algorithm

Genetic algorithms (GAs) were formally introduced in the United States in the 1970s by John Holland at the University of Michigan. The continuing price/performance improvement of computer systems has made some types of optimization more attractive. To use a GA, the solution of a problem must be represented as a *genome* (or *chromosome*). The GA then creates a population of solutions and applies genetic operators such as mutation and crossover to evolve the solutions in order to find the best one(s). The three most important aspects of using GAs are: (1) definition of the objective function, (2) definition and implementation of the genetic representation, and (3) definition and implementation of the genetic operators.

To use a genetic algorithm, the biological background should be recognized. Biological terms used to describe the optimization process are given in Table 14.2.

Table 14.2 GA terms

Terminology	Microbiology Definitions	Implication
Chromosomes	Structure that carries genes and is responsible for heredity	Possible Solutions
Genes	Segment of a DNA molecule that contains information directing a cell to synthesize a specific protein product	Information of solutions, eg. Significant Digit
Alleles	One or more alternative forms of a given gene concerned with the same trait or characteristic; one of a pair or multiple forms of a gene located at the same locus of homologous chromosomes	Characteristic of solutions, eg. The smaller the solution, the larger the error
Genome	A complete set of genetic information contained in a set of chromosomes	Possible solution set
Genotype	The genetic information contained in the entire complement of alleles	Particular type of solution set
Phenotype	The totality of observable structural and functional characteristics of an individual organism, determined jointly by combination of its genotype and the environment	Criteria of optimizing solution, the particular solution (Genotype) set is developed based on the criteria (Phenotype)
Reproduction	A fundamental property of living systems by which organisms give rise to other organisms of the same kind	Two solutions are taken as one of the objective to be changed.
Recombination (crossover)	The exchange and incorporation of genetic information into a single genome, resulting in the formation of new combinations of alleles	Two solutions will be changed by exchanging some of the number within the solution, eg. 12.345 & *11*.234 => *11*.345 & 12.234
Mutation	A stable condition in which two organisms of different species live in close physical association; symbiosis	Solutions will be randomly changed, eg. 12.234 => 12.133

The GA is based on the precept that: "*Good parents have better children*" (even though this is not true in reality). According to this principle, the process starts with a set of **chromosomes** (solutions), called a **population**. Two chromosomes are selected from the population and represent the **parents**. Each chromosome in the population is evaluated by their degree of **fitness**, $f(x)$, which is used to decide whether the process should be continued. The second step is to select two chromosomes from a new population as parents and perform a process called **crossover**. According to the user's assigned **crossover probability**, **children** (offspring) might be generated by crossing over the parent's chromosomes. If no crossover is performed, the offspring are an exact copy of parents. After the process of crossover, **mutation** takes place. Mutation is a random perturbation process of the **genes** within the chromosomes according to the **mutation probability** assigned by the user. After the process of mutation, the new chromosomes are born and are evaluated by their *f(x)*. In the performance evaluation process, if the *f(x)* of the children are better than the parents, this pair of chromosomes is placed into a new population. On the other hand, if the parents are better than the children, the parents' chromosomes are inserted into the next generation. The best *f(x)* of children or parent will be inserted into the new population for the next iteration. Finally, the whole process is repeated again until the criteria are satisfied. Figure 14.4 illustrates the process.

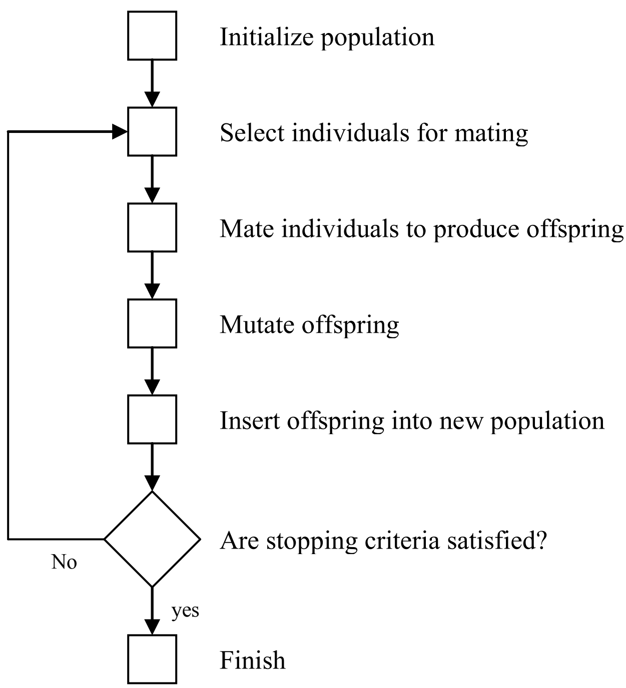


Figure 14.4 GA procedure

GA Parameters and Applications

Two GA parameters that play an important role in the optimization process are crossover probability and mutation probability. Other issues are the population size and the encoding.

Crossover Probability. Crossover probability is the frequency at which the two chromosomes are crossed over. If no crossovers were performed, the new offspring would be an exact copy. The difference between the offspring and the original solution is caused by the crossover probability. For example, if the crossover probability is *100%*, then all offspring chromosomes are crossed over. If the crossover probability is *0%*, then the offspring are an exact copy of the original solution.

Mutation Probability. Mutation probability is the degree to which the offspring are mutated. If there is no mutation (mutation probability = *0%*), offspring will be placed into the new population without any changes. Similarly, if the mutation probability is *100%*, the offspring will be changed completely.

Population Size. Another parameter that should also be noted in a GA is the population size. The population size is related to the number of solutions that were taken into the solution set. On the one hand, a small population size will provide few solutions to crossover. As a result, a small part of the search space is explored. On the other hand, a large population will slow down the optimization process and the process will be considered inefficient. However, with present computer technology, the optimization speed might not be affected greatly.

Encoding. Encoding identifies the representation of chromosomes for different applications. Various encoding methods are selected depending on the conditions. Generally, encoding is categorized into four types: (1) binary, (2) permutation, (3) value and (4) tree encoding. Value encoding is used in PCSWMM.

14.9 Application of the Genetic Algorithm to Calibration

The GA consists of two main operations: crossover and mutation. Two runoff data input files represent the parents of the chromosome. After completing the crossover operation, several children are generated and some parameters within the children are mutated randomly. The performance of these children is evaluated by their degree of fitness (*f(x)*). Any two of the children or parents may be selected to populate the next generation if their performance is the best among these children and parents. GA calibration is sequential. The process is stopped when an acceptable performance of any member of the population is achieved. The procedures are discussed in the following sections.

Two problem sets (chromosome) are required to populate the next generation by creating different solution sets (children). In SWMM modeling, two runoff data files (parents) are required initially. The first data file is the "best estimate" input file that was initially built by the modeler and it is manually corrupted. The second data file is automatically generated by randomly corrupting the manually-corrupted data file. This random corruption procedure is only applied to sensitive parameters. Generally the target data set (TD) will be the observed field data.

Crossover Operation

In microbiological terminology, crossover is an exchange of genes between two chromosomes. There are two different types of crossover: single and multi crossover. Figures 14.5 and 14.6 demonstrate this operation.

	WW1	WAREA	WW3	WSLOPE	WW5
Parent 1	95	480	20	0.0545	0.0432
Parent 2	86	484	19.8	0.041	0.0344
Child 1	95	480	19.8	0.041	0.0344
Child 2	86	484	20	0.0545	0.0432

Figure 14.5: Single Genetic Crossover Operation

	WW1	WAREA	WW3	WSLOPE	WW5
Parent 1	95	480	20	0.0545	0.0432
Parent 2	86	484	19.8	0.041	0.0344
Child 1	95	480	19.8	0.041	0.0432
Child 2	86	484	20	0.0545	0.0344

Figure 14.6: Multi Crossover Operation

In Figure 14.6, WW1, WAREA, WW3, WSLOPE and WW5 are the sensitive SWMM runoff parameters and are selected as the calibration parameters in this example. Initially, the calibration parameters are separated into several different strata. Each strata of the parent that is crossed over is determined by the crossover probability. In Figure 14.6 the crossover probability is 40%, 40% and 20%, so WW1 and WAREA are in the 1st strata, WW3 and WSLOPE are in the 2nd strata, and WW5 is in the 3rd strata. The number of parameters in each stratum can be determined by the following equation.

No. parameters in each strata = Crossover Probability × Total No. Parameters

According to Figure 14.6, two parents generate two children after the crossover operation. This kind of crossover is the most typical GA procedure and is commonly used. However, it is possible to generate more than two children if multi crossover is used. In reality parents can have many children. The maximum possible combination of crossing over the multiple strata of the parent solution sets is examined below.

	WW1	WAREA	WW3	WSLOPE	WW5
Parent 1	95	480	20	0.0545	0.0432
Parent 2	86	484	19.8	0.041	0.0344
Child 1	95	480	19.8	0.041	0.0432
Child 2	86	484	20	0.0545	0.0344
Child 3	95	480	20	0.0545	0.0344
Child 4	86	484	19.8	0.041	0.0432
Child 5	95	480	19.8	0.041	0.0344
Child 6	86	484	20	0.0545	0.0432

Figure 14.7: Maximum Possible Number of Children

In Figure 14.7, six possible children are born for from both parents with three strata. The total number of children depends on the number of strata and is decided by the user. Total maximum possible children is determined by the following equation.

$$\textit{Total Maximum Possible Children} = 2^n - N_{parent}$$

where n = total number of strata

N_{parent} = total number of parents

After the crossover operation, the next procedure is mutation.

Mutation Operation

Mutation is a part of the genetic reproduction process. The purpose of this operation is to perturb the genetic information within a chromosome. In practice, each parameter that was contained in a solution set of each child will be perturbed by adding or subtracting a small number. The parameter perturbation is decided by the user-defined mutation probability. Figure 14.8 shows the procedure.

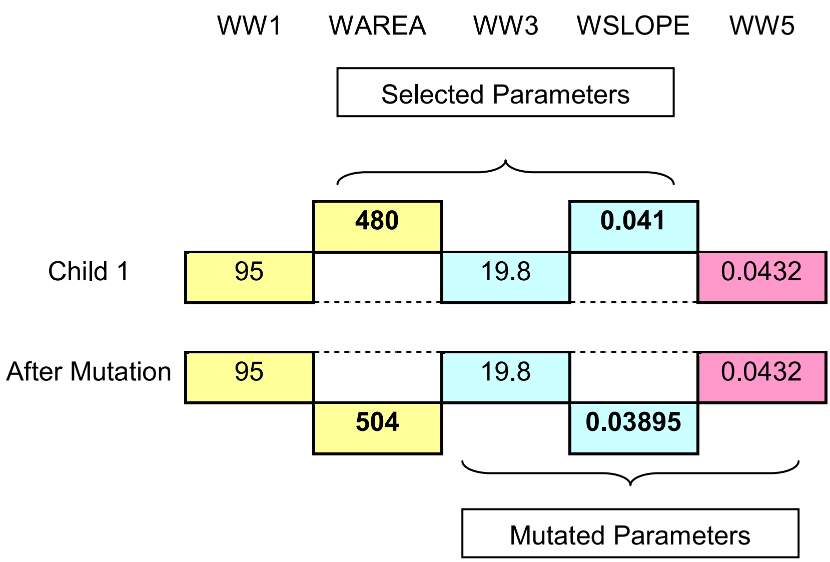


Figure 14.8: Mutation Operation

In Figure 14.8, the subcatchment area (WAREA) and subcatchment slope (WSLOPE) are chosen to be the parameters to mutate and the mutation values are +5% and –5%, respectively. The user defines this mutation value; and the mutated parameters are randomly selected. In this example, the mutation value is fixed at ±5%. However, this value is further investigated in this study in order to improve the calibration process. Finally, the order of mutation is calculated by the following equation:

$$M_{order} = (-1)^{n-1}$$

where n = parameter number
M_{order} = mutation order

Once the crossover operation and mutation operation are completed, the next step is to input the new SWMM runoff parameter sets into PCSWMM and run the model. The objective function (peak runoff flow rate) is computed in PCSWMM.

Performance Evaluation

In order to determine the performance of the GA calibration, the computed objective function of each child is compared with the TD's objective function. The performance of the calibration process is evaluated by the degree of fitness (*f(x)*). In practice, *f(x)* is calculated by using an evaluation function (EF). Various EFs are discussed elsewhere. In this chapter, the simple least square approximation (1 dimension) is selected:

$$EF = f(x) = \sum \left(\frac{OOF_i - COF_i}{OOF_i} \right)^2$$

where i = 1, … , n
OOF = observed objective function
COF = computed objective function

As discussed in Chapter two, there is no standard method of determining the best EF and the selection is totally subjective. Simple least squares was used because of its long historical associations.

Toleration of *f(x)*

Tolerance is assigned to pause the automatic GA calibration method. In the previous chapter, 5% of true value is considered to be excellent for stream flow measurement. The tolerance is calculated based on this accuracy:

$$\begin{aligned} COF &= (100\% - 5\%) \times OOF \\ &= 0.95\ OOF \end{aligned}$$

Substitute into the selected EF

$$E.F. = f(x) = \sum \left(\frac{OOF_i - COF_i}{OOF_i} \right)^2 = \sum \left(\frac{OOF_i - 0.95\,OOF_i}{OOF_i} \right)^2 = \sum (0.05)^2$$

$$\therefore \quad Tolerance = (0.05)^2 \cdot n$$

where n = number of calibration parameters

The main purpose of the sensitivity-based GA calibration method is to minimize the error between the computed objective function and observed objective function for all rainfall events. Figures 14.9 and 14.10 show the calibration plot before and after using the preliminary GA calibration method, based on work done by Wan (2001).

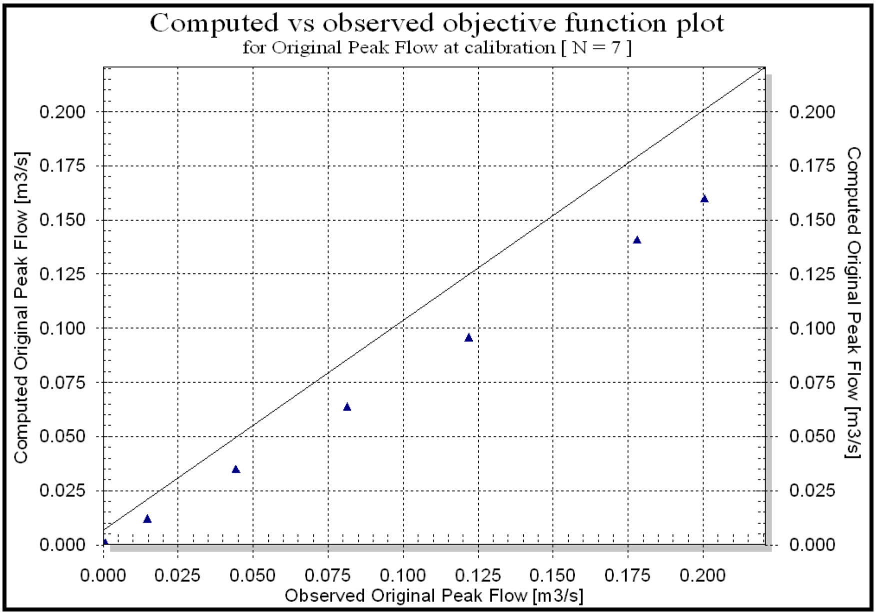


Figure 14.9: Before Applying GA Calibration Method

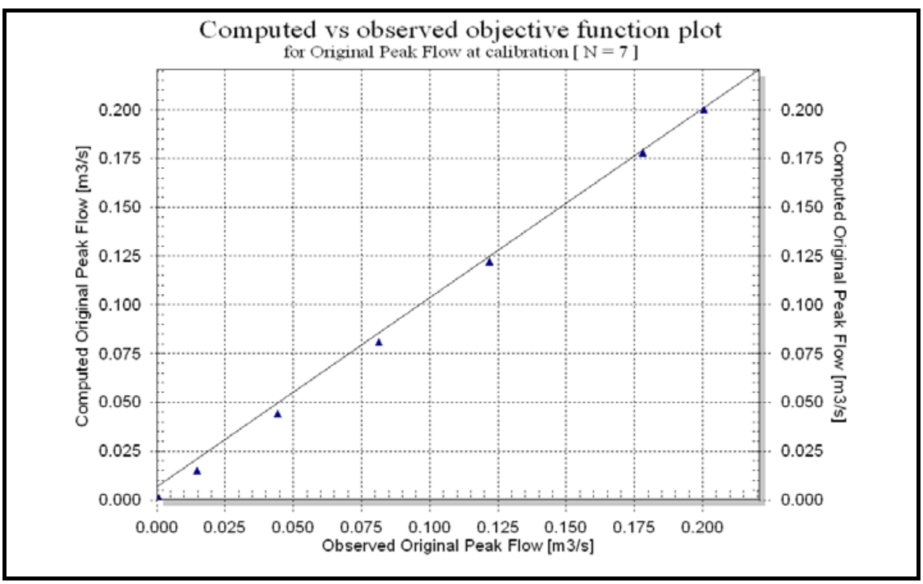


Figure 14.10: After Applying GA Calibration Method

After only 7 cycles the calibrated parameters provide nearly the same result as the TD. Hence, the GA calibration method is evidently efficient. In order to optimize the GA calibration method, the following GA components were analyzed:

Number of Parents
Number of Crossovers
Crossover Probability
Mutation Probability
Evaluation Functions

Number of Parents

Wan's analysis determined the number of parents that produced the minimum iteration when *f(x)* reaches a constant value (no further changes of *f(x)*). In addition, the minimum number of parents was 2 in order to access the GA calibration method. In this analysis, several initial inputs were assigned and the results are:

Number of Crossovers = 3 Crossover probability = 40%, 40%, 20%
Mutation Range = 5% Mutation Probability = 0.4

Table 14.3: Number of Parents

	Number of Parents			
Cycle	2	3	4	5
1	2.18E-01	2.18E-01	2.18E-01	2.18E-01
2	5.84E-02	**4.92E-02**	**4.92E-02**	**4.92E-02**
3	5.09E-02	4.92E-02	4.92E-02	4.92E-02
4	4.99E-02	4.92E-02	4.92E-02	4.92E-02
5	**4.92E-02**	4.92E-02	4.92E-02	4.92E-02
6	4.92E-02	4.92E-02	4.92E-02	4.92E-02
7	4.92E-02	4.92E-02	4.92E-02	4.92E-02
8	4.92E-02	4.92E-02	4.92E-02	4.92E-02
9	4.92E-02	4.92E-02	4.92E-02	4.92E-02
10	4.92E-02	4.92E-02	4.92E-02	4.92E-02
# of Children:	**6**	**24**	**60**	**120**

Table 14.3 shows the *f(x)* for various number of parents. The best *f(x)* was achieved in 5th cycle and 2nd cycle for the number of parents equal to 2 and 3, respectively. The GA calibration method with 3 parents required more runs of SWMM than did 2 parents due to the fact that more children are generated.

Computational time was relatively longer for 3 parents. Hence the GA with 2 parents is a better choice.

Crossover Parameters

Wan varied the number of crossovers and the crossover probability in order to analyze the GA calibration program. Initial conditions were the same as those given in the previous section, except that the number of parents was equal to 2 and the tolerance for the program was 4.92E-02.

Table: 14.4: Results of Different Crossover

Crossovers	%Crossover Probability	Av. Cycles	Exceed-ance	Children
2	0, 100 (No crossover)	24.5	0	2
	20, 80	16.5	1	
	40, 60	9.0	2	
	60, 40	9.3	1	
	80, 20	6.2	4	
3	20, 20, 60	5.7	3	6
	20, 60, 20	4.2	3	
	60, 20, 20	4.1	1	
	20, 40, 40	**4.4**	**4**	
	40, 20, 40	3.6	1	
	40, 40, 20	3.9	3	
4	20, 20, 20, 40	3.1	3	14
	20, 20, 40, 20	4.5	1	
	20, 40, 20, 20	**2.1**	**3**	
	40, 20, 20, 20	3.2	1	
5	20, 20, 20, 20, 20	**2.1**	**8**	30

Table 14.4 shows the results of the GA-CP with the varied number of crossovers and crossover probability. This analysis was done by running the GA-CP with the same input ten times in order to obtain an approximate average number of runs for each of the different entities. The exceedance of the average number of computations was counted in order to determine the reliability of the approximation. The highlighted cells indicate that he GA's parameters provide

an efficient computation in terms of number of runs and their exceedance. Three crossovers with crossover probability of 20%, 40% and 40% was the best selection for the GA-CP in this study due to the shorter computational time.

Note that the number of cycles fluctuates even for the same conditions (e.g. 2 crossovers with 60% & 20%). This phenomenon was caused by the randomization in GA-CP. The number of runs is small when the *f(x)* of the second parent is lower than the first parent, and vice versa. Hence, it is necessary to ensure the *f(x)* of the second parent is lower than the first parent in order to minimize the computational time.

Mutation Parameters

Wan analyzed the two mutation parameters: mutation range and mutation probability, in order to increase the efficiency of the GA-CP. Initial conditions were:

Number of Parents = 2
Number of Crossovers = 3
Crossover Probability = 20%, 40%, 40%
Evaluation Functions = Simple least squares (first dimensionless form)

Tables 14.5 and 14.6 show the results.

Table 14.5: Mutation Range

Mutation Range	*f(x)*	# of Runs
10%	0.04908222	15
5%	0.04899342	5
2%	0.04899342	8
1%	0.04903386	20

* Program discontinued when *f(x)* = 0.04899342

Table 14.6: Mutation Probability

Mutation Probability	# of Runs
100%	6
80%	6
60%	7
40%	5
20%	7

Ideally, a small mutation range should bring the calibration parameters closer to the true value. However, Table 14.5 shows that a smaller mutation range will not achieve the ideal solution. In fact, a smaller mutation range was even worse. The mutation range of 5% will efficiently determine the best parameter and this was set as a default in the GA program.

In Table 14.6, the mutation probability of 100% shows that all calibration parameters have a chance to be mutated. The results show that 40% was an appropriate mutation probability and can determine the best parameter sets (lowest *f(x)*) within 5 runs. Thus 40% was set as a default mutation probability.

Table 14.7 summarizes the optimized parameters, all of which were then set as default values for the program.

Table 14.7: Default Value for GA-CP

Number of Parents	2
Number of Crossovers	3
Crossover Probability	20%, 40%, 40%
Mutation Range	5%
Mutation Probability	40%
Evaluation Function	Simple least squares (1st dimensionless form)

In the GA calibration program, users can set their own preferred parameters.

Validation of a Genetic Algorithm Calibration Program (GACP)

A different SWMM runoff data file (TDIF2.dat) was used to verify the consistency of GA-CP.

Sensitivity Analysis

Figure 14.11 shows the results obtained in the sensitivity analysis of TDIF2.dat

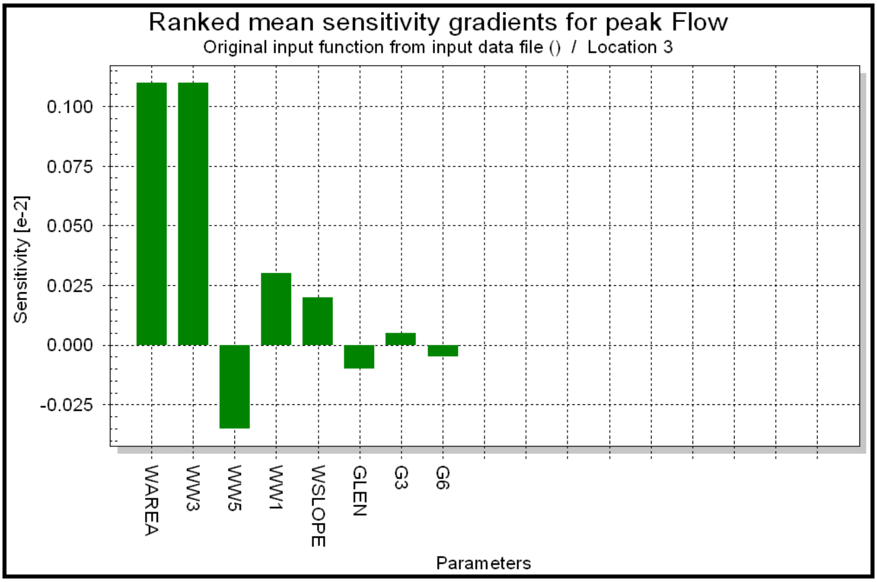


Figure 14.11: Sensitivity Gradient for TDIF2.dat

Comparing all sensitive parameters, the length of channel/pipe (GLEN), the invert slope (G3) and the Manning's n (G6) were relatively insensitive. Thus only five sensitive parameters were selected for calibration.

As mentioned earlier, the GA's parameters listed in Table 14.7 were used as default values in the GA program. The best parameter set was obtained after 49 cycles and the *f(x)* was **4.892218E-03**. Figures 14.12 and 14.13 show the calibration plots before and after using the GA program and Figure 14.14 compares the target and computed hydrographs.

Discussion

Since 5 calibration parameters were used, the tolerance of this analysis was **0.0125**. After the first six runs, the *f(x)* of **5.99459E-03** was achieved and the tolerance was met.

In order to complete the second task, the GA calibration program continued with the same mutation range of 5% for another 30 runs and the *f(x)* of **4.984318E-03** was estimated in the 36th run. The mutation range was changed to 3% after the 36th run in order to determine a better answer. A lower value of *f(x)*, **4.918529E-03**, was achieved in the 44th run. Subsequently, the mutation range of 1% and 0.5% was used to determine an even lower *f(x)*. Finally, **4.892218E-03** was determined in the 49th run. Figure 14.15 shows the highest peak runoff rate of Figure 14.14.

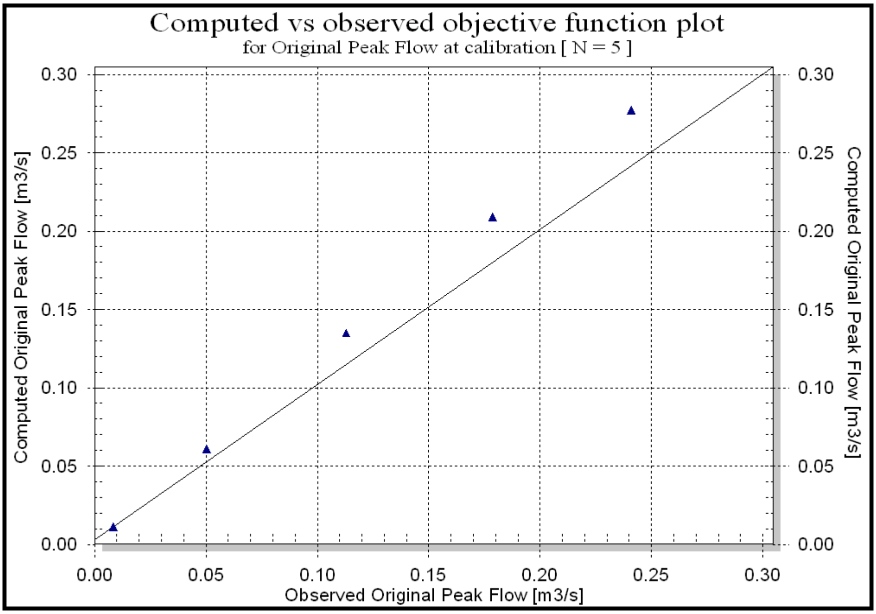


Figure 14.12: Calibration Plot of TDIF2.dat (Before Calibration)

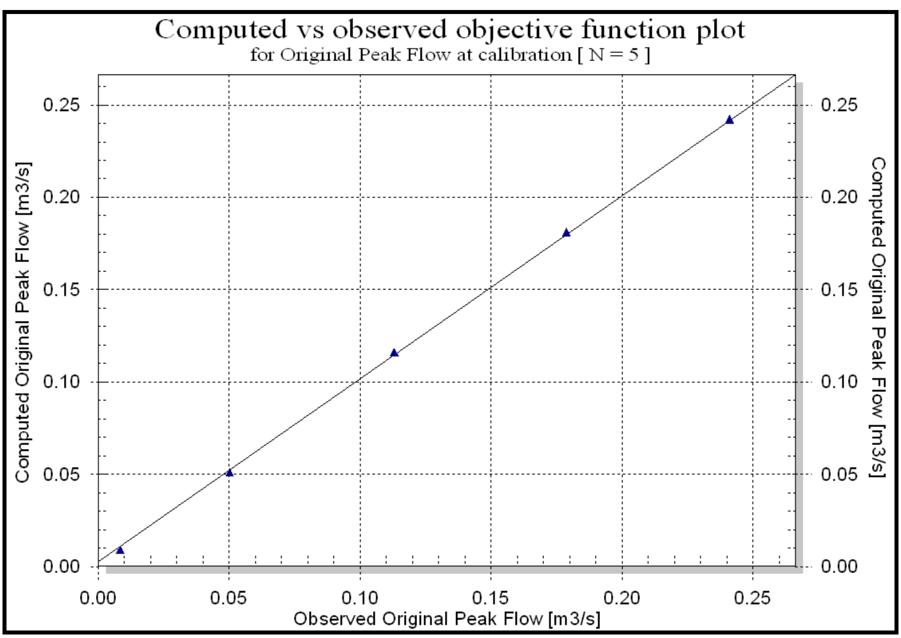


Figure 14.13: Calibration Plot of Best Parameter Set (After Calibration)

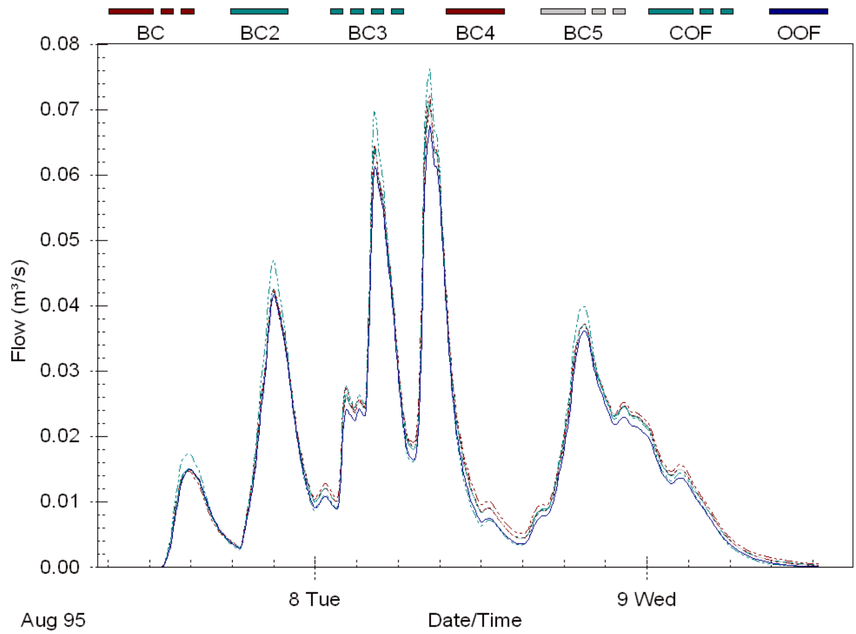


Figure 14.14: Hydrograph of Observed Data and Computed Data

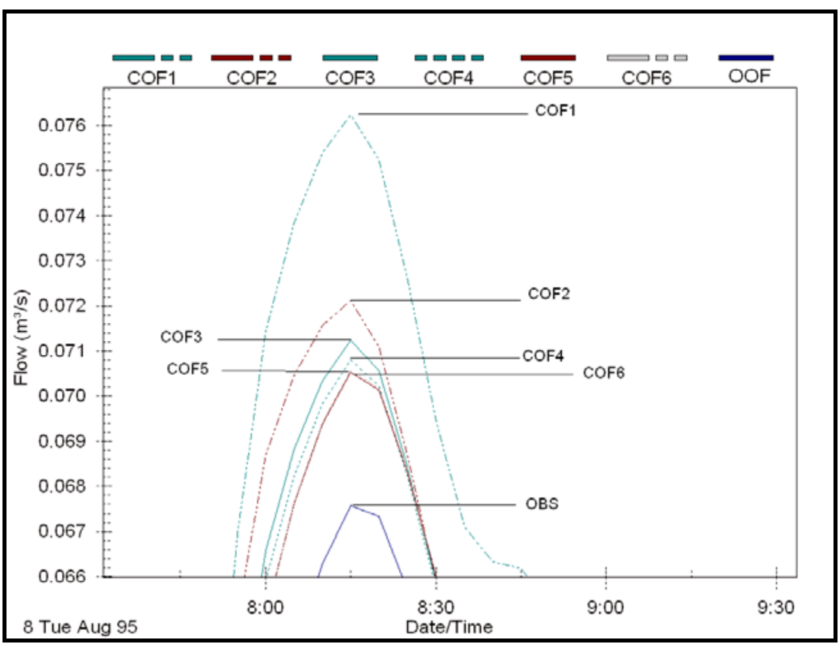


Figure 14.15: Highest Peak Runoff Rate

Table 14.8 explains the labels in Figure 14.15 and summarizes the results.

Table 14.8: Labels and Results of GA-CP for TDIF2.dat

Labels	Definition	f(x)	Max. flow	Cycle
OOF	Observed OF	N/A	0.06756 m^3/s	N/A
COF1	Computed OF	N/A	0.07623 m^3/s	N/A
COF2	2^{nd} COF	5.99459E-03	0.07211 m^3/s	6
COF3	3^{rd} COF	4.984318E-03	0.07124 m^3/s	36
COF4	4^{th} COF	4.918529E-03	0.07081 m^3/s	44
COF5	5^{th} COF	4.892218E-03	0.07057 m^3/s	56
COF6	6^{th} COF	4.892218E-03	0.07057 m^3/s	61

The % difference for each calibration parameter is shown in Table 14.9.

Table 14.9: % difference in Calibration Parameters for TDIF2.dat

TD-IF:

Subcatchment	WW1	WAREA	WW3	WSLOPE	WW5
100	25	4	20	0.015	0.05
200	36	12	18	0.009	0.03
300	97	8.5	27	0.01	0.04

Computed:

Subcatchment	WW1	WAREA	WW3	WSLOPE	WW5
100	27.11	4.04	20.8	0.0248	0.0423
200	42.17	13.09	20.63	0.009	0.0428
300	119.1	7.748	26.87	0.0172	0.0615

Percent Errors:

Subcatchment	WW1	WAREA	WW3	WSLOPE	WW5
100	8%	1%	4%	39%	18%
200	15%	8%	13%	0%	30%
300	19%	10%	0%	42%	35%

In Table 14.9, only two calibration parameters were the same as those in the TD-IF. Large % differences occur in WSLOPE and WW5 due to the their low sensitivity gradient. For WW1, WAREA and WW3, relatively small errors were found due to their high sensitive gradient. Overall, the performance of the calibration method is considered to be acceptable. In fact, the percent differences between the OOF and COF6 was only 4.5%, better than excellent stream flow measurements.

Four screens from the calibration wizard in PCSWMM are shown in the calibration workspace in Figure 14.16, and this is followed by several more.

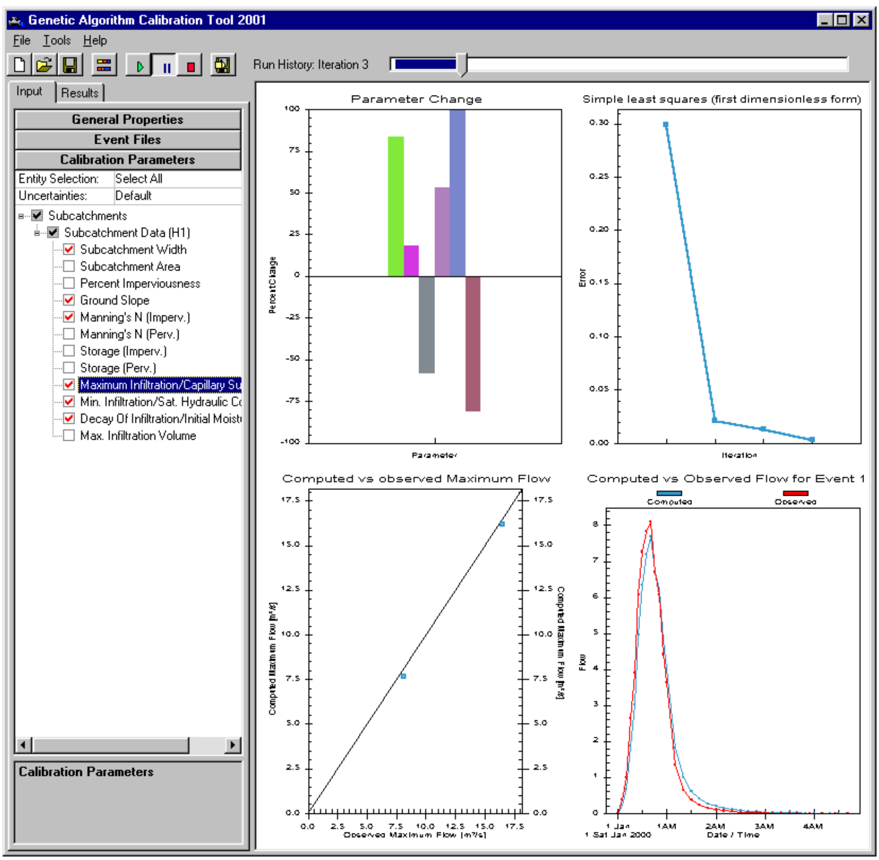
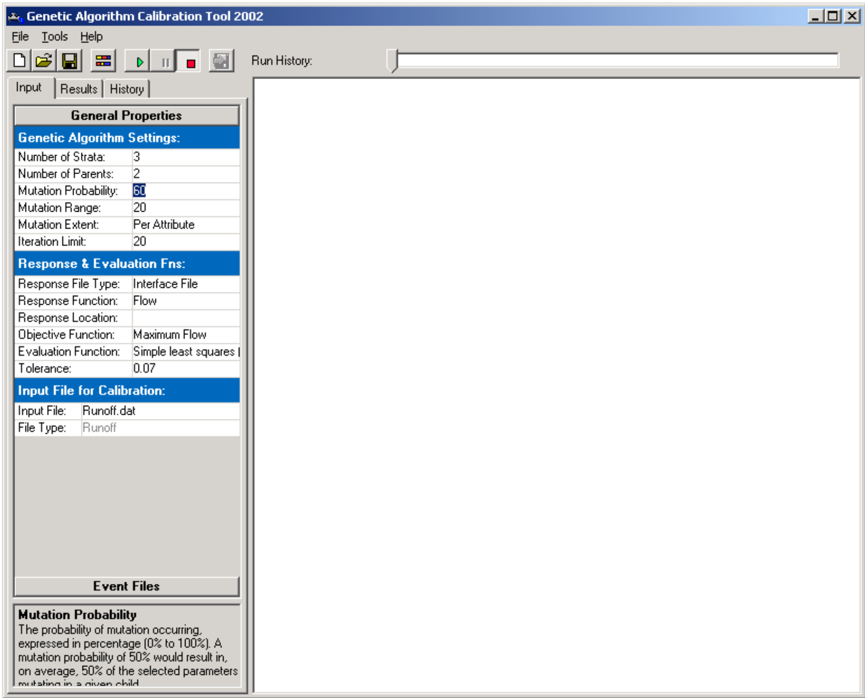
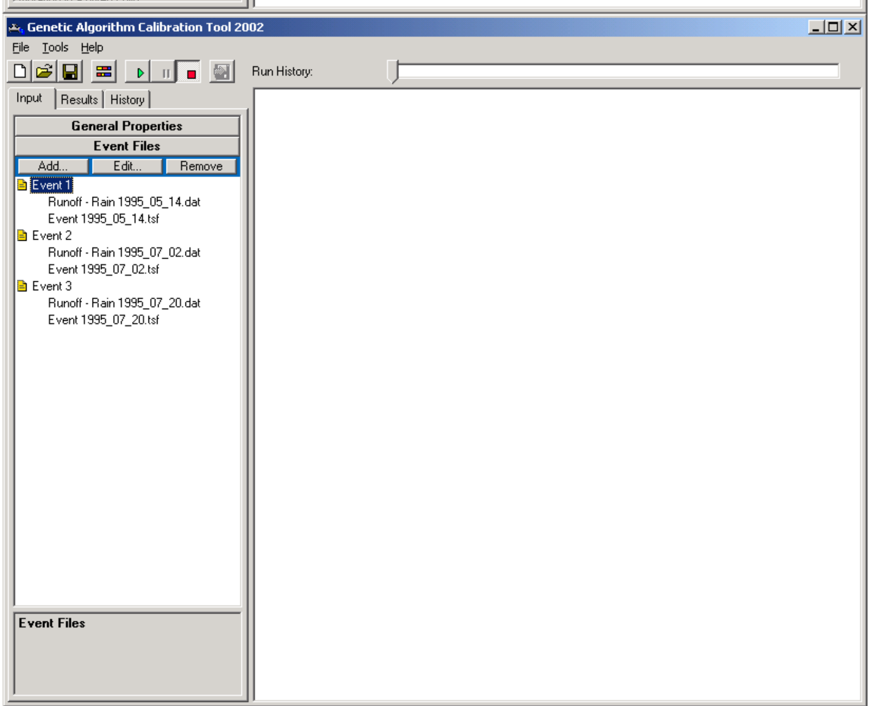
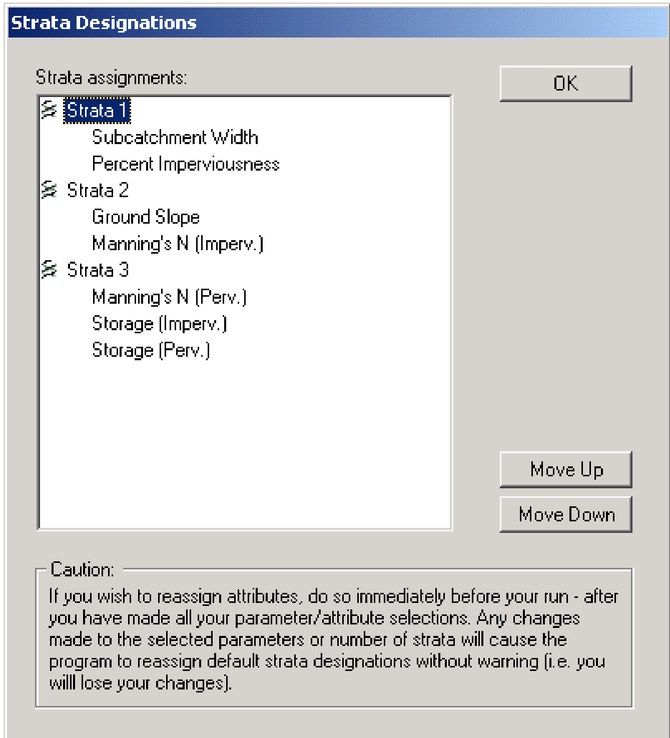
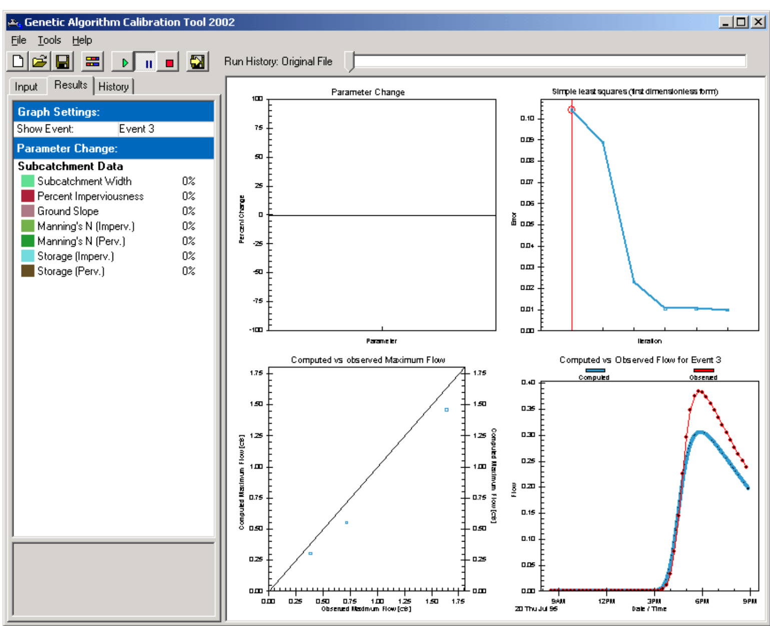


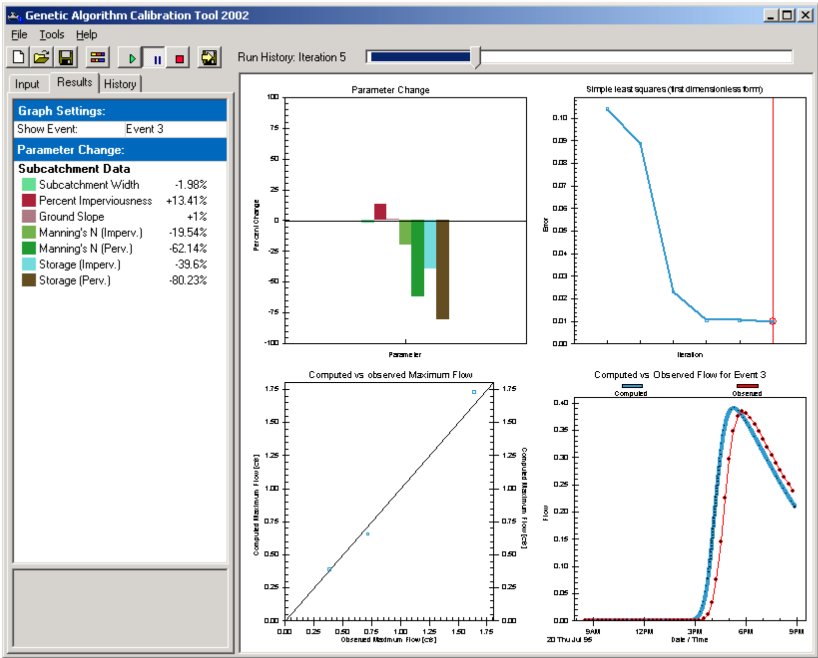
Figure 14.16: The PCSWMM calibration wizard workspace.


Genetic Algorithm Calibration Tool 2002
File Tools Help
Run History:
Input Results History
General Properties
Genetic Algorithm Settings:
Number of Strata: 3
Number of Parents: 2
Mutation Probability: 60
Mutation Range: 20
Mutation Extent: Per Attribute
Iteration Limit: 20
Response & Evaluation Fns:
Response File Type: Interface File
Response Function: Flow
Response Location:
Objective Function: Maximum Flow
Evaluation Function: Simple least squares
Tolerance: 0.07
Input File for Calibration:
Input File: Runoff.dat
File Type: Runoff
Event Files
Mutation Probability
The probability of mutation occurring, expressed in percentage (0% to 100%). A mutation probability of 50% would result in, on average, 50% of the selected parameters


Genetic Algorithm Calibration Tool 2002
File Tools Help
Run History:
Input Results History
General Properties
Event Files
Add... Edit... Remove
Event 1
Runoff - Rain 1995_05_14.dat
Event 1995_05_14.tsf
Event 2
Runoff - Rain 1995_07_02.dat
Event 1995_07_02.tsf
Event 3
Runoff - Rain 1995_07_20.dat
Event 1995_07_20.tsf
Event Files


Strata Designations
Strata assignments:
OK
Strata 1
Subcatchment Width
Percent Imperviousness
Strata 2
Ground Slope
Manning's N (Imperv.)
Strata 3
Manning's N (Perv.)
Storage (Imperv.)
Storage (Perv.)
Move Up
Move Down
Caution:
If you wish to reassign attributes, do so immediately before your run - after you have made all your parameter/attribute selections. Any changes made to the selected parameters or number of strata will cause the program to reassign default strata designations without warning (i.e. you willl lose your changes).


Genetic Algorithm Calibration Tool 2002
File Tools Help
Run History: Original File
Input
Results
History
Graph Settings:
Show Event: Event 3
Parameter Change:
Subcatchment Data
Subcatchment Width 0%
Percent Imperviousness 0%
Ground Slope 0%
Manning's N (Imperv.) 0%
Manning's N (Perv.) 0%
Storage (Imperv.) 0%
Storage (Perv.) 0%
Parameter Change
Simple least squares (first dimensionless form)
Computed vs observed Maximum Flow
Computed vs Observed Flow for Event 3
Computed
Observed



REFERENCES

Duan, Q., Sorooshian, S., Gupta, V., (1992), *Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models*, Water Resources Research, 28(4).1015-1031.

Holland, J. H., (1975), *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence,* Ann Arbor, MI, University of Michigan Press. ISBN 0-262-58111-6

James, W., James, R. C., (1999), *Water Systems Models – Hydrology* 4th Printing, CHI Inc., Guelph, Ontario

Javaheri, H. (1998), *Automatic Calibration of Urban Runoff Models using Global Optimization Techniques*, Master Thesis, Department of Civil Engineering and Applied Mechanics, McGill University, Montréal, Canada.

Kuch, A. J. (1994), *A Sensitivity Analysis Shell for Complex Continuous Water Quality Management Modelling.* M. Sc. Thesis, University of Guelph. Ca., 120 pp.

Liong, S.Y., Ibrahim, Y. (1995), *Catchment Calibration Using Fractional-Factorial and Central-Composite-Designs-Based Response Surface*, Journal of Hydraulic Engineering, Vol. 121 (6), p. 507-617.

Liong, S.Y., Chan, W.T., Lum, L.H. (1991), *Knowledge-Based System for SWMM Runoff Component Calibration*, Journal of Water Resources Planning and Management, Vol. 117 (5), p. 507-524.

Mays, L.W., (1996), *Water Resources Handbook* Internation Edition, McGraw-Hill Co. Inc., New York ISBN 0-07-114517-6

Rechenberg, I., (1973), *Evolution Strategies*, Stuttgart-Bad Cannstatt : Frommann-Holzboog, Germany, ISBN 377280374

Thyer, M., Kuczera, G., and Bates, C. (1999), *Probabilistic Optimization for Conceptual Rainfall-Runoff Models: A Comparison of the Shuffled Complex Evolution and Simulated Annealing Algorithms*, Water Resources Research, Vol. 35, p. 767-773

Wan, B., (2001), Auto-calibration of SWMM-RUNOFF using sensitivity-based genetic algorithms, MSc thesis, U of Guelph, Ontario. 159pp.

Yan, J. (1990), *Parameter Estimation for Multipurpose Hydrologic Models*, Ph.D. Thesis, Department of Agricultural Engineering, Oklahoma State University, Oklahoma.

Web References:

Annoymous URL – 1. USGS 1999 California Hydrologic Report – Accuracy of the Records, California, http:// www.water.wr.usgs.gov/data/text/accuracy.html

URL – 2. Introduction to Intelligent Computing, http://neural.cs.nthu.edu.tw/jang/courses/cs4601/simplex.htm

URL – 3, Solomatine, D. P., (1995), *Experience with Nine Automatic Calibration Algorithms: Global Optimization with GLOBE System*, http://www.dhi.dk/softcon/papers/003/003.htm

BIBLIOGRAPHY

Bäck, T. (1996), Evolutionary Algorithms in Theory and Practice, Oxford University Press, New York.

Baffaut, C. (1988), Knowledge Techniques for the Analysis of Urban Runoff using SWMM, Ph.D. Thesis, School of Civil Engineering, University of Purdue, Purdue.

Bos, M. G., Replogle, J. A., Clemmens, A. J. (1943), Flow Measuring Flumes for Open Channel Systems, John Wiley & Sons, New York. ISBN 0-471-80637-4

Calver, A. (1988), Calibration, Sensitivity and Validation of a Physically-Based Rainfall-Runoff Model, Journal of Hydrology, Vol. 103, p. 103-115.

Chambers, L. (1995), Practical Handbook of Genetic Algorithm, CRC Press, Inc., New York. ISBN 0-8493-2519-6

Chong, E. K. P., Zak, S. H., (1996), An Introduction to Optimization, John Wiley & Sons, Inc., New York, ISBN 0417-08949-4

Duan, Q., Sorooshian, S., Gupta, V., (1995), Optimal use of the SCE-UA Global Optimization Method for Calibrating Watershed Models, Journal of Hydrology, 158: 241-264.

Franchini, M. (1996), Use of a Genetic Algorihtm Combined with a Local Search Method for the Automatic Calibration of Conceptual Rainfall-Runoff Models, Journal of Hydrological Sciences, Vol. 41 (1), p. 21-39.

Franchini, M., Galeati, G., (1997), Comparing Several Genetic Algorithm Schemes for the Calibration of Conceptual Rainfall-Runoff Models, Journal of Hydrological Sciences, Vol. 43, p.357-379

Gaume, E., Villeneuve, J. P., Desbordes, M., (1998), Uncertainty Assessment and Analysis of the Calibrated Parameter Values of an Urban Storm water Quality Model, Journal of Hydrology, Vol 210, p.38-50

Gan, T.Y., Biftu, G.F. (1996), Automatic Calibration of Conceptual rainfall-Runoff Models: Optimization Algorithms, Catchment Conditions, and Model Structure, Water Resources Research, Vol. 32 (12), p. 3513-3524.

Gupta, V.K., Sorroshian, S. (1985), the Automatic Calibration of Conceptual Catchment Models Using Derivative Based Optimization Algorithm, Water Resources Research, Vol. 21 (4), p.473-485.
Ibrahim, Y. (1993), A Method of Estimating Optimal Catchment Model Parameter, Water Resources Research, Vol. 29 (9), p. 3049-3058.
Ibrahim, Y., Liong, S.Y. (1992), Calibration Strategy for Urban Catchment Parameters, Journal of Hydraulic Engineering, Vol. 118 (11), p. 1550-1570.
James, W. (2000), Applied Modelling of Urban Water Systems - Monograph 8, CHI, Guelph, Ontario.
Johnston, P.R., Pilgrim, D.H. (1976), Parameter Optimization for Watershed Models, Water Resources Research, Vol. 12 (3), p. 477-486.
Lamb, R., (1999), Calibration of a Conceptual Rainfall-Runoff Model for Flood Frequency Estimation by Continuous Simulation, Water Resources Research, Vol. 35, p.3103-3114.
Liong, S.Y., Ibrahim, Y. (1994), Estimation of Peak Flow and Runoff Volume with Response Surface Method, Journal of Water Resources Planning and Management, Vol. 120 (2), p. 161-173.
López, A., (1999), Stochastic Water Quality Models: Solution, Calibration and Application, Ph. D. Thesis, Faculty of the Graduate School, University of Cornell, Cornell, USA
Pham, D.T., Karaboga, D. (1988), Intelligent Optimisation Techniques: Genetic Algorithms, Tabu Search, Simulated Annealing and Neural Networks, Springer c2000, New York. ISBN 1-85233-028-7
Rivera Santos, J. (1988), A Methodology for Parameter Estimation in Conceptual Precipitation-Runoff Models with Emphasis on Error Analysis, Ph.D. Thesis, Department of Civil, Environmental and Architectural Engineering, University of Colorado, CO.
Rogers, C.C.M., Beven, K.J., Morris, E.M., Anderson, M.G. (1985), Sensitivity Analysis, Calibration and Predictive Uncertainty of the Institute of Hydrology Distributed Model, Journal of Hydrology, Vol. 28, p. 179-191.
Sulistiyono, H., (1999), Rainfall-Runoff Model Calibration Using Experimental Designs and Response surface Methodology, M.Eng. Thesis, Faculty of Engineering and Applied Science, University of Newfoundland, Newfoundland, Canada.
Thompson, D. (1989), Determining Parameter for a Continuous Simulation Model by Estimation, Synthetic Calibration and Analytic Calibration, Ph.D. Thesis, School of Civil Engineering, University of Missouri-Rolla, Missouri.
Vose, M. D., (1999), The Simple Genetic Algorithm – Foundations and Theory, The MIT Press, Cambridge, Massachusetts, ISBN 0-262-22058-X (hc: alk. Paper)
Wang, Q. J., (1991), the Genetic Algorithm and Its Application to Calibrating Conceptual Rainfall-Runoff Models, Water Resources Research, Vol. 27, p.2467-2471

Web References:

URL 1. – Solomatine, D. P., (2000), Application of Global Optimization to Models Calibration, http://www.ihe.nl/hi/sol/p_blob1/onglob06.htm
URL 2. – Hsul, K. L., Gupta, H. V., Sorooshian, S., (2000), Artificial Neural Network Modelling of the Rainfall-Runoff Process, http://www.hwr.arizona.edu/~hoshin/papers/wrr31_10_1995/wrr_ann.htm
URL 3. – Schwar, M., (1996), Rainfall-Runoff Model Calibration, http://ftp.engr.wisc.edu/~schwar/rrmodelcalib.html
URL 4. – Parker, M. A., Savic, D. A., Walters, G. A., Kappelan, Z., (2000) SEWERNET: A Genetic Algorithm Application for Optimising Urban Drainage Systems, http://www.hydroinform.com/ICUDI/pages/papers/parker.htm

URL 5. – Obitko, M., (1998), Introduction to Genetic Algorithms with Java Applets, http://cs.felk.cvut.cz/~xobitko/ga/
URL 6. – Pohlheim, H., (1997), GEATbx: Genetic and Evolutionary Algorithm Toolbox for use with MATLAB, http://www.systemtechnik.tu-ilmenau.de/~pohlheim/GA_Toolbox/index.html
URL 7. – Annoymous, (undated), Global Optimization, http://www.fi.uib.no/~antonych/glob.html
URL 8. – Wall, M., (undated), Introduction to Genetic Algorithms, http://lancet.mit.edu/~mbwall/presentations/IntroToGAs/index.html
URL 9. - Annoymous, (undated), The Genetic Algorithms Archive, http://www.aic.nrl.navy.mil/galist/
URL 10. - Walters, G. A., Savic, D. A., (undated), Recent Applications of Genetic Algorithms to Water System Design, http://www.hydroinform.com/ICUDI/pages/papers/parker.htm
URL 11. - Walters, G. A., Savic, D. A., (1995), Genetic Algorithm Techniques for Calibrating Network Models, http://www.hydroinform.com/ICUDI/pages/papers/parker.htm
URL 12. - Walters, G. A., Savic, D. A., Morley, M., Schaetzen, W., Atkinson, R., (undated), Calibration of Water Distribution Network Models using Genetic Algorithms, http://www.hydroinform.com/ICUDI/pages/papers/parker.htm

This page intentionally left blank.

Chapter 15

ON USING A LITTLE FUZZY LOGIC

Words [resemble] those nebulous masses familiar to the astronomer, in which a clear and unmistakable nucleus shades off on all sides, through zones of decreasing brightness, to a dim marginal film that seems to end nowhere, but to lose itself imperceptibly in the surrounding darkness. - James A. H. Murray

Arthur: You speak with the tongue of snakes, I will no more of this.
Soldier: Not at all. It's just that ants are my special subject. Ants, bees, wasps, all the hymenoptera, and you often get people who just bandy the word "ant" around as if it meant something. It's like saying "I am a human". It's so unspecific. – Monty Python and the Holy Grail.

15.1 Introduction

In this chapter we use some elements of fuzzy logic to develop arguments for defining relevant state variable (SV) sub-spaces. We use the information later to compute and represent model reliability in real time (model reliability is presented in Chapter 16). Fuzzy logic is *not* used to re-develop the basic algorithms used in the SWMM or HSPF engines, as you might otherwise have expected. Instead we fuzzify the dominance of the processes, and then defuzzify for selecting rules for real-time error estimation. Defuzzification procedures used here are not the same as those frequently seen in the literature.

Many fundamental design problems can be formulated in fuzzy terms, for example, whether a culvert may be "too big", or the pollutant may have "somewhat too high a concentration", or the detention pond is "too small", or the stream flow depths and velocities are "possibly a little low". More particularly, we use the approach to determine which processes are more or less "dominant" or "dormant".

Fuzzy procedures are based on concepts that are inexact quantitatively, yet important in engineering design. Examples of such design questions and their associated OFs from earlier are:

problem expressed fuzzily	**objective function**
Is the diameter of the conduit *too big*?	OF_3
Is the detention basin *small*?	OF_5
Is there *too much* pollutant?	OF_5
Will the creek erode *a lot*?	OF_8, OF_6
Will the riparian ecosystem *suffer*?	OF_3, OF_4,OF_7

We strive to associate processes with various types of SV sub-spaces, such as the wetness or water storage in the catchment. Accumulated moisture depends on antecedent meteorology such as rate and duration of precipitation, ET, temperature, and wind.

This complicated task can be reduced somewhat by considering the concept of water vapor moving in the direction of the vapor gradient, from higher concentration to lower. For evapo-transpiration (ET) the water vapor gradient is negative and water transport is away from the surface. So ET can be conceptualized as a negative form of precipitation. Some terms for the fuzzy sub-ranges of rates-of-precipitation and its duration are shown schematically in Figure 15.1. Note again that while rain and ET are input functions, their cumulative value correlates with the moisture state of the catchment, which in turn is a factor in the response. Another factor is the runoff state, also related to the input functions.

<-----rate of evapo-transpiration----rate of rain------------------------>

very high medium low zero light medium very heavy

<----duration of evapo-transpiration-------duration of rain----------->

long medium short zero short medium long

<--------------------------temperature------------------------------------->

very low low freezing moderate high

I-----------------------------wind--->

zero light moderate high

Figure 15.1 Fuzzy ranges of input functions that indicate state-variables

15.2 Fuzzification

SV space is plotted in two dimensions in Figure 12.2, rate of rain or ET plotted vertically, and duration horizontally. Infiltration processes dominate in sub-space B, but not A. On the other hand in sub-space A infiltration is satisfied and no longer dominates – the entire urban surface contributes runoff.

Thus it is in the definition of these sub-spaces that fuzzification is used. Typical questions that must be addressed are: What is a *medium* duration of precipitation? What is a *short* duration of evapo-transpiration? What is a *very heavy* rate-of-rain? What is a *medium* evapo-transpiration rate? While it is possible to derive order-of-magnitude estimates for some of these dimensions, e.g. relating the rain duration to the characteristic response time of the drainage system (analogous to the time-of-concentration), and rain rate to infiltration capacity, the effort to draw up a general rule is likely to be frustrating, because these characteristics are dependent on individual applications, and not simple constants. Further details are provided in Chapter 12.

Kuch (1997) tested all parameters for sensitivity for all rains reported in Chapter 12, and a sample sensitivity plot is shown in Figure 12.3. The discussion in the next few pages explains the results of the fuzzy sensitivity analysis in excruciating detail. Inclusion of the discussion is warranted, however, because it exemplifies how knowledge may be gained from sensitivity analysis.

Sensitivity of PCTZER

The PCTZER parameter designates the percentage of the impervious area having zero depression storage. This allows the generation of an overland flow hydrograph for all rainfall events regardless of depth, duration or intensity. The balance of the impervious area must first receive a rainfall depth that exceeds the impervious depression storage before overland flow is generated.

The value for this parameter in the Redhill dataset was the program default value of 25%. This parameter displayed its maximum sensitivity at the smallest rainfall depths, a relative sensitivity of 1.0 (maximum) for a storm depth of 0.025 inches. The parameter became less sensitive as the total volume of rainfall increased and at a depth of more than 0.50 inches was not sensitive. This parameter was most sensitive during short duration low intensity events since the product of intensity and duration yields the depth of the storm. The sensitivity of the parameter was not influenced by the duration or the intensity so much as by the storm volume.

Sensitivity of REGEN

The REGEN parameter is only active when the Horton infiltration method has been selected. It describes the rate at which the infiltration rate recovers after the cessation of rainfall. The sensitivity of this parameter requires a range of durations and evaporation rates to describe the rainfall inter-event. The REGEN parameter is only active during dry periods and its effect on the dynamic infiltration rate is noticed when subsequent rainfall occurs after the dry period.

The value for this parameter in the Redhill dataset was 0.01 s. This parameter displayed its maximum sensitivity for medium inter-event durations of 1 and 3 d. The parameter became insensitive after a duration of 6 d of no rain. The evaporation rates had no impact on the sensitivity of this parameter using peak flow as the objective function. Minimal sensitivity was seen for different evaporation rates using runoff volume as the objective function. This parameter never displayed more than a 0.7% change in any objective function for a 10% change in the variable. This parameter was directly influenced by the duration of no rain.

Sensitivity of WIDTH

SWMM assumes sheet flow on the subcatchment and non-linear reservoir routing. WIDTH is a measure of overland flow width for both the pervious and impervious areas. Larger subcatchment widths result in faster catchment response and are representative of highly drained areas. Smaller subcatchment widths mean longer overland flow lengths, longer response time and are representative of open areas without well defined surface and engineered drainage systems. WIDTH can be assumed to be a tuning parameter, since it controls attenuation.

The value for this parameter in the Redhill dataset ranged from 2700 to 39300 feet. The two corresponding subcatchments for this low and high WIDTH value had catchment areas of 595 and 2516 acres and % impervious of 17.0 and 0.0 respectively. This parameter displayed its maximum sensitivity at the smallest rainfall totals and some sensitivity was seen for all rainfall depths. The parameter became less sensitive as the total volume of rainfall increased and at a depth of more than 0.50 was markedly less sensitive. This parameter was most sensitive during short duration low intensity events and short duration high intensity events. A second region of sensitivity for high intensity short events was also discovered. This is when pervious area runoff contributed to the overall peak and runoff volume totals. For the two storms of 3.5 inch/h and 0.25 and 0.50 h duration the WIDTH parameter was as sensitive as it was for most intensities at the short durations. WIDTH was not sensitive for durations in excess of 3 h except for medium intensity storms with respect to the volume of runoff.

Generally the parameter was sensitive when a new source of overland flow contributed to the total runoff, such as the pervious area with depression storage. At the smallest of rainfall intensities and volumes overland flow consists of only overland flow on the impervious area without any depression storage. At a rainfall depth beyond the impervious area depression storage the impervious area contributed to the total runoff and the WIDTH parameter was sensitive. Finally, at high intensities above the infiltration rate the WIDTH parameter was again sensitive as pervious area overland flow became a significant portion of the total flow. Some sensitivity although less than the others was also noticed at long durations of medium intensity as the infiltration rate was reduced to a rate less than the rainfall intensity and overland flow occurred on the pervious area.

Sensitivity of WAREA

WAREA is used in SWMM to designate the total area of each subcatchment in the watershed. Impervious areas are described as a % of the total area and thus increasing or decreasing WAREA will adjust the total area of the subcatchment and not affect the ratio of impervious and pervious areas. WAREA is not often considered to be a candidate for calibration because of the accuracy used in survey and GIS techniques. However, the WAREA parameter is a measure of the hydraulically effective area. If a portion of the subcatchment although clearly within a defined subcatchment watershed boundary does not directly connect to the drainage network, only the area contributing to the overall measured runoff should be counted as the area of the subcatchment. For example, a sag in the landscape, that drains only by infiltration and no groundwater flow, should be excluded from the area. In addition, the sensitivity of the subcatchment area can show the location of hydraulic limitation in a sewer shed when a proportional amount of flow is not seen for an increase or decrease in the WAREA parameter. Theoretically there should be a 10% increase in runoff for a 10% increase in area.

The values for the subcatchment areas in the Redhill dataset ranged from 355 to 3463 acres. This parameter displayed its maximum sensitivity at the smallest rainfall totals where there was an even larger increase in peak flow than parameter change. At most other rainfall storms the sensitivity was nearly 1: 1 with a 10% increase in peak flows and runoff volume for a 10% increase in the WAREA parameter. Correspondingly, the peak flow and runoff volumes decreased by 10% for a decrease in the WAREA parameter. The parameter was equally sensitive for all ranges of storms except at very high intensity and high runoff volumes. At these times the model was hydraulically controlled and the runoff rates were reduced at the outfall by limited hydraulic capacity of conduits causing excess flows to be held in artificial storage at inlets. This was found at rain depths of more than 1.5 inches and rainfall rates of 3.5 inch/h. During simulations when more flow was generated than the hydraulic system can handle, RUNOFF stores water at the upstream inlets or junctions until there is capacity later in the simulation for these flooded volumes to be entered into the system. This method may or may not mimic the physical system being modeled. EXTRAN offers alternatives to modeling hydraulics and surcharge.

Sensitivity of Percent Impervious PCTZER

PCTZER is used to assign a portion of the total subcatchment area as impervious area and hence not subjected to infiltration losses, This area allows the complete generation of all rainfall to overland flow for rainfall amounts greater than the depression storage. Additionally, some of this impervious area can have no depression storage and is designated with the PCTZER parameter discussed earlier. Hence, a larger percent impervious or total area would also increase the total area for the impervious area without depression storage.

The value for this parameter among all subcatchments ranged from 0% to 24%. This parameter displayed equal sensitivity of nearly a 1:1 correlation while there was only overland flow on the impervious areas. At high rainfall intensities and long duration medium intensities where overland flow was generated on the pervious area, the sensitivity was significantly diminished. At these times the impervious area runoff became less than 100% of the total runoff. The domain of sensitivity for this parameter was all low intensity and short duration medium intensity storms. The parameter did not become less sensitive as the total volume of rainfall increased except when total runoff was hydraulically limited by the sewer network. This parameter was most sensitive during short duration low intensity events. The sensitivity of the parameter was influenced by increased duration for all storms with intensities greater than the minimum infiltration rate and duration long enough to cause overland flow on the pervious area. Similar sensitivity was displayed using peak flows or runoff volumes as the objective function.

Sensitivity of WSLOPE

WSLOPE is used in SWMM to assign an average slope for the idealized subcatchment. The same slope is common to all surfaces on the subcatchment (pervious and impervious). If the physical slopes on each surface are significantly different, two separate subcatchments should be considered, one subcatchment of a given slope for the entire pervious area and another for the entire impervious area. Alternatively, some small differences in overland flow caused by having different slopes for each surface can be incorporated by using suitable values of roughness for each surface. Larger WSLOPE resulted in a faster response to rainfall, larger peak flows and runoff volumes in this analysis.

The value for this parameter ranged from 0.01 to 0.24 feet/feet. This parameter displayed its maximum sensitivity for short duration low intensity. However, the parameter was sensitive for all rainfall intensities of short duration. The parameter became less sensitive as the duration increased and at a duration of more than 1.0 h was markedly less sensitive. Some sensitivity was found for all events with runoff volume as the objective function, but no sensitivity was discovered for durations of more than 3 h with peak flow as the objective function. This parameter was not highly sensitive even at short duration low intensity. In this case a 10% change in the parameter resulted in a 2.7% change in the peak flow and a 2.0% change in the runoff volume. Generally the parameter was sensitive for short durations. For these storms a higher slope resulted in larger peak flows and volumes since the depth of water on the idealized catchment builds up over time.

Sensitivity of IMPERN

IMPERN is used to assign a Manning's roughness to the impervious area of the subcatchment. The subcatchment width and the slope are common to both the pervious and impervious area. Allowing different roughness values for each surface permits the runoff hydrographs for each surface to have different resulting time-to-peak and hydrograph shapes. Impervious areas respond to all rainfall intensities and volumes if some of the impervious area has no depression storage. Correspondingly, the domain for the sensitivity of the IMPERN parameter is when the overland flow on the impervious area is the dominant process. In addition, the sensitivity of the IMPERN parameter should decrease as overland flow on the pervious area becomes a larger portion of the total subcatchment response.

The value of IMPERN for each subcatchment in the Redhill dataset was 0.017 although the model allows different IMPERN for each subcatchment. This parameter displayed its maximum sensitivity at the smallest rainfall totals where only overland flow was occurring from the impervious area without depression storage. Unlike all of the previously discussed parameters, when this parameter was increased the peak flow and runoff volume decreased, because a rougher surface impedes the overland flow. At its most sensitive state a 10% change in the IMPERN parameter resulted in a 5.5% change in peak flow and a 4.0% change in the runoff volume. The parameter sensitivity was very uniform with respect to duration and intensity. As the intensity increased and the duration increased the sensitivity reduced for both peak flow and runoff volume. Generally the parameter was sensitive for short and medium durations and low and medium intensities. The sensitivity dropped off sharply above 1 h durations and 0.5 inch/h. The diminished sensitivity for this parameter at longer durations and intensities is attributed to the increase of the overland flow on the pervious area as a portion of the total flow.

Sensitivity of PERVN

PERVN is used to assign Manning's roughness to the pervious area. Just as was the case for the impervious area roughness an increase in this parameter allows a longer time to peak and a flatter hydrographs for the pervious area. Pervious areas only generate runoff in SWMM when the depression storage has been exceeded and the rainfall rate is greater than the infiltration rate.

Correspondingly, the PERVN parameter will be most sensitive when the overland flow on the pervious area is a dominant process. Additionally, the domain should not include short-duration low and medium-intensity events. Sensitivity of the PERVN parameter should be at its highest at the onset of overland flow on the pervious area.

The value of the PERVN parameter for each subcatchment in the Redhill dataset was 0.025 although the model allows different PERVN for each subcatchment as was the case for IMPERN.

This parameter displayed its maximum sensitivity for high intensity short duration rainfall. Unlike the IMPERN parameter which displayed sensitivity for all storms this parameter did not display any sensitivity for all storm intensities below 0.5 inch/h (the minimum infiltration rate) and short duration medium intensity events. This was expected because no overland flow was generated for pervious areas during these events i.e. the process was not active. At its most sensitive state a 10% change in the PERVN parameter resulted in a 2.5% change in peak flow and a 3.7% change in the runoff volume. Similar parameter sensitivity was observed for both these objective functions. Generally, the parameter was sensitive for events that activated the overland flow process for the pervious area and when the hydraulic routing of the sewer system did not limit the flow to the outfall, since the flows, volumes, SS concentration and load at the outfall were used as the objective functions.

Sensitivity of IMPDEP

IMPDEP is used to assign a depression storage to the impervious area. No runoff will be generated from the impervious area with depression storage until a rainfall depth in excess of IMPDEP occurs. Correspondingly, the IMPDEP parameter will be most sensitive when the overland flow on the impervious area without depression storage is the dominant process. In addition, the sensitivity of the IMPDEP parameter should decrease as overland flow on the pervious area becomes a larger portion of the total subcatchment response.

The value of the IMPDEP parameter for each subcatchment in the Redhill dataset was 0.050 inches although the model allows different IMPDEP for each subcatchment. This parameter displayed its maximum sensitivity at a rainfall depth just exceeding the value of IMPDEP. The sensitivity of the parameter was independent of rainfall intensity and duration but was dependent on the depth of rainfall. Increasing the value of the parameter decreased the computed peak flows and runoff volumes. This parameter was hypersensitive for rainfall depths just beyond the value of the parameter. At a rainfall depth of 0.0625 inches a 10% change in the IMPDEP parameter resulted in a 24.5% change in peak flow and a 17.7% change in the runoff volume. The parameter sensitivity was equally distributed for peak flows and runoff volumes. The sensitivity reduced for both peak flow and runoff volume from this depth until a depth of 0.750 inches after which the parameter was insensitive. Generally, the domain of parameter sensitivity was short duration low intensity storms with rainfall depths above 0.050 inches. The total range of the domain was very small (event volumes from the IMPDEP to about 5 times the IMPDEP). The diminished sensitivity for this parameter at medium and longer durations and medium intensities results from the reduction of the relative value of this initial abstraction to the overall storm depth.

Sensitivity of PERDEP

PERDEP is used to assign a depression storage to the pervious area. The rainfall rate must exceed the infiltration rate for a period of time long enough to buildup a depth greater than PERDEP before overland flow occurs on the pervious area. The PERDEP parameter is most sensitive when the overland flow on the pervious area becomes a dominant process. In addition, the sensitivity of the PERDEP parameter should decrease as overland flow on the pervious area increases as a portion of the total subcatchment response since the PERDEP parameter acts as a fixed initial abstraction of rainfall. However, in the case of a rainfall storm the depth on the watershed is dynamic with the depth increasing when rainfall exceeds infiltration and falling when the inverse is true.

The value of the PERDEP parameter for each subcatchment in the Redhill dataset was 0.20 inches. This parameter displayed its maximum sensitivity for high intensity short duration and long duration medium intensity events. These two regions of input variable subspaces activated overland flow over the pervious areas. There was no sensitivity for events with less than 0.875 inches of rain or events with less than 0.50 inch/h rainfall intensity. The sensitivity was dependent on rainfall intensity and duration because all intensities had to be greater than the minimum infiltration rate and have a duration long enough to develop a depth on the watershed more than PERDEP. In a similar fashion to IMPDEP increasing the value of the parameter decreased the peak flow and runoff volumes. The parameter was most sensitive for the highest intensity shortest duration event with a 10% change in the PERDEP parameter resulted in a 4.0% change in peak flow and a 6.6% change in the runoff volume.

Sensitivity of WLMAX

WLMAX is used to assign an initial infiltration rate for the pervious area of the subcatchment when the Horton infiltration option is used. No runoff is generated from the pervious area unless the rainfall rate is more than the infiltration rate and the depression storage on the pervious area is filled. The infiltration rate in SWMM is dynamic, using the Horton method has the infiltration rate decay exponentially throughout the storm from the WLMAX rate to the WLMIN rate using the DECAY parameter to describe the rate of infiltration decay. The sensitivity of WLMIN and DECAY is described below. High WLMAX rates lower the peak flows and volumes of runoff as more rainfall is allowed to infiltrate. Correspondingly, the WLMAX parameter will be most sensitive when the overland flow on the pervious area is dominant. The value of the WLMAX was 3.00 inch/h for each subcatchment in the Redhill dataset. Individual infiltration parameters can be used for each subcatchment, each is a lumped effective rate for an individual subcatchment. This parameter shared the same domain of sensitivity as the PERDEP parameter with the sensitivity being about twice as great. This parameter displayed its maximum sensitivity for high intensity short duration and long duration medium intensity events. In these two regions the rainfall activated the overland flow over the pervious areas. There was no sensitivity of the parameter for events with less than 0.750 inches of rain or events with less than 0.50 inch/h. The sensitivity was dependent on rainfall intensity and duration because all intensities had to be greater than the minimum infiltration rate and have a duration long enough to develop a depth on the watershed more than PERDEP. The parameter was most sensitive for the highest intensity shortest duration event with a 10% change in the PERDEP parameter resulting in a 9.3% change in peak flow and a 12.7% change in the runoff volume.

Sensitivity of WLMIN

The WLMIN parameter is used to assign a minimum infiltration rate when the Horton infiltration option is used. High WLMIN rates lower the peak flows and volumes of runoff since more rainfall is allowed to infiltrate. For long durations and large storm volumes the infiltration rate will reduce down to or near the minimum infiltration rate described by WLMIN. Correspondingly, the WLMIN parameter will be most sensitive when the overland flow on the pervious area is the dominant process and infiltration has reached this minimum.

The value of the WLMIN was 0.30 inch/h for each subcatchment in the Redhill dataset. This parameter was sensitive for very large storm volumes. It was most sensitive for storms of 1.5 to 6 inches in depth, and for the medium intensity long duration events. Although some sensitivity was seen for high intensity long duration events these generated extremely high flows and the runoff obtained at the outfall was limited by the conveyance system. As the intensity increased above the minimum infiltration rate and the duration increased the parameter sensitivity decreased. This showed that the sensitivity was at a maximum with storm intensities just above the minimum infiltration rate. There was no sensitivity of the parameter for events with less than 0.750 inches of rain or events with an intensity of less than 0.50 inch/h. The sensitivity was dependent on rainfall intensity and duration because all intensities had to be greater than the minimum infiltration rate and have a duration long enough to saturate the soil and reduce the infiltration rate to less than the rainfall rate. The parameter was most sensitive for the lowest medium intensity with the shortest duration that lowered the infiltration rate to or near WLMIN. For the case of the maximum sensitivity using peak flow and runoff volume as objective functions a 10% change in the WLMIN parameter resulted in a 13.9% change in peak flow and a 18.4% change in the runoff volume.

Sensitivity of DECAY

DECAY is used to describe the rate at which the Horton infiltration rate will decrease from WLMAX to WLMIN. The infiltration rate will decay over time for all cases where the rainfall rate is greater than WLMIN. Thus no sensitivity will be seen for this parameter for events with the rainfall rate less than WLMIN. Additionally, for events when the rainfall rate exceeds the dynamically changing infiltration rate and the depth of water on the subcatchment equals or exceeds PERDEP there will be flow on the pervious area and sensitivity gradients can be calculated using the peak flow and runoff volume for this parameter. High DECAY rates raise the peak flows and volumes of runoff as less rainfall is allowed to infiltrate with the infiltration rate decreasing faster. Correspondingly the DECAY parameter will be most sensitive when the overland flow on the pervious area is the dominant process.

The value of the DECAY was 0.00115 s^{-1} for each subcatchment in the Redhill dataset. This parameter was sensitive in the same identical input parameter subspaces as the WLMAX and WLMIN parameters with the sensitivity gradients in the case of DECAY being opposite in sign. This parameter displayed its maximum sensitivity for high intensity short duration and long duration medium intensity events. These two regions of input variable subspace activated overland flow over the pervious areas. There was no sensitivity of the parameter for events with less than 0.750 inches of rain or events with intensities less than 0.50 inch/h. The sensitivity was dependent on rainfall intensity and duration because all intensities had to be greater than the minimum infiltration rate and be of sufficient duration to develop a depth on the watershed greater than PERDEP. The parameter was most sensitive for the highest intensity shortest duration event with a 10% change in the PERDEP parameter resulting in a 4.2% change in peak flow and a 7.6% change in the runoff volume.

Sensitivity of WASHPO

SWMM can use an exponential washoff equation for computing dynamic runoff concentration. Two variables are used to describe the washoff of available pollutants from the catchment. WASHPO is the exponent in the washoff equation. For the sensitivity analysis of this study only the SS concentration and load were used as objective functions, all other water quality parameters would have similar sensitivity. A washoff exponent of more than 1.0 is generally used to model sediment source constituents and a value of less than 1.0 is often used to model dissolved constituents whose concentration diminishes strongly with increasing flow rate. A value of 1.0 (used in the Redhill Creek example) causes the equation to be linear. The same set of storm events that were used for all of the hydrology parameters were used for the analysis of WASHPO and RCOEFF (the coefficient in the washoff equation). The difference in the analysis is that the peak SS concentration and load are used as objective functions to measure sensitivity instead of the peak flow and runoff volume. For all cases the number of dry days before the start of each event used in the sensitivity analysis was 2.0 d. This DRYDAY parameter initializes the model with a fixed amount of pollutants on the subcatchment described by the buildup equation.

The value of the WASHPO for SS was 1.0 in the Redhill dataset. A deviation from 1.0 results in a nonlinear equation and correspondingly the parameter is very sensitive. Individual WASHPO parameters are used for each pollutant. In the case of the Redhill Creek dataset the three other pollutants also used a value of 1.0. The washoff parameter was most sensitive for the short duration low intensity storms. Decreasing sensitivity was found with increased storm depth and the sensitivity of the parameter to load was very small beyond 0.750 inches of rainfall. However, using peak concentration as the objective function the parameter was sensitive for all storm events. For all events an increase in the WASHPO parameter resulted in a decrease of the objective function. The parameter displayed a near linear sensitivity for each rainfall intensity regardless of duration. The same result was not found for the load as the longer durations resulted in lower sensitivity for each intensity. This result can be explained by the continual depletion of available pollutant by washoff. Longer durations approached complete depletion of pollutant load. With an infinite supply of pollutants a similar graph of sensitivity would be seen for concentration and load. Additionally, for each increased rainfall intensity and increased duration the sensitivity decreased. The parameter was very sensitive. The largest sensitivity displayed was for the short duration low intensity event with a 79.4% change in peak SS concentration and a 79.1 % change in SS load for a 10% change in the WASHPO parameter for SS.

Sensitivity of RCOEFF

The second washoff parameter is RCOEFF, or the coefficient in the washoff equation. SS concentration and load were used as objective functions. Both the WASHPO and RCOEFF parameters were analyzed using the same storm events. The Redhill dataset uses linear time dependent buildup and exponential washoff. For the RCOEFF parameter to be sensitive with respect to concentrations and loads there must be pollutant washoff occurring and the process must be dominant. The washoff algorithm removes pollutants on the surface that have been built-up over a dry period. For the sensitivity analysis of the Redhill data file this was a 2 day period for all events. With a finite amount of pollutants built up and available to be washed off the washoff parameters will be most sensitive when most of the pollutant is still available.

The value of the RCOEFF parameter for SS was 40.0 in the Redhill dataset. The washoff coefficient parameter was most sensitive for the short duration low intensity storms. The parameter displayed a near linear sensitivity for each rainfall intensity independent of duration. The same result was not found for the load, longer durations resulted in lower sensitivity for each intensity. This result can be explained by the continual depletion of available SS to washoff. Additionally, for each increased rainfall intensity and increased duration sensitivity decreased. The largest sensitivity displayed was for the short duration low intensity event with a 10% change in peak SS concentration and a 9.9% change in SS load for a 10% change in the RCOEFF parameter for SS .

Sensitivity of DDLIM

SWMM incorporates a time dependent buildup of each pollutant or the buildup of dust and dirt with pollutants being defined as factors of dust and dirt buildup. The time dependent buildup can be estimated as linear, exponential or have the form of the Michaelis-Menton equation (Huber and Dickinson, 1988). In the case of the Redhill dataset linear buildup of dust and dirt was estimated and a potency factor for SS was used to relate it to the buildup of dust and dirt. The buildup can be based on area, or total catchment curb length and the buildup may be limited by the incorporation of a buildup limit. For the case of SS in the Redhill model, the SS buildup was modeled using a factor of 360 times the buildup of DD with the buildup of DD being described linearly by total curb length in each subcatchment using the DDLIM parameter. Situations arise where the DDLIM parameter is not at all sensitive. If the buildup rate is small and the interevent time is not long enough (or dry days before storm for single events) DDLIM will not be sensitive since the pollutants have not reached the limit. No buildup occurs in the SWMM model during rainfall.

To measure the sensitivity of DDLIM and DDPOW (exponent of pollutant buildup, 30 different dry events were used followed by a fixed rainfall to wash off the pollutants. This allows the peak concentration and total load to be used as objective functions. The dry periods ranged from a short 0.25 d to a long 12 d. For each duration varying evaporation rates were also used.

The value of the DDLIM in the Redhill data was 1.0E6 pounds, in other words no buildup is accumulated beyond a total mass of 1.0E6 pounds of SS. For all cases analysed no sensitivity was found for concentration or load. The concentration and load were not influenced by the duration, or evaporation rate. Closer examination of the output files revealed that even after 12 d the total SS buildup did not attain 1.0E6 pounds. With the buildup limit set so high, it was not found to be sensitive. It would however be sensitive at a much lower value and for much longer durations. Generally, in arid climates the DDLIM parameter would be sensitive and the DDPOW less sensitive because of the long durations of no rainfall. On the other hand, in regions which receive frequent storms, the SWMM models have sensitive buildup rates and insensitive buildup limits - the limit is never reached. This is the case for the Redhill model: computed pollutant buildup is controlled only by the DDPOW parameter.

Sensitivity of DDPOW

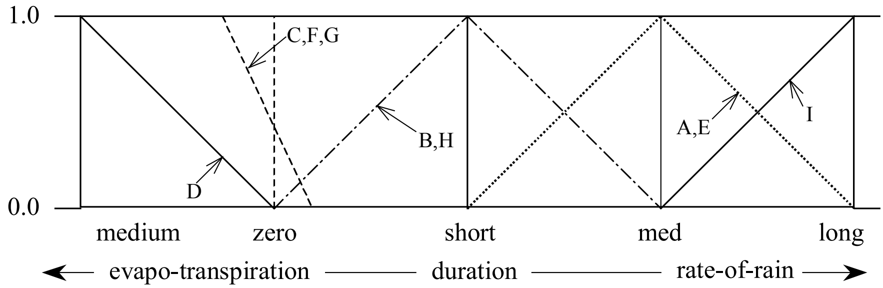
SWMM computes time-dependent buildup of dust and dirt with linear, exponential or Michaelis-Menton equations. For the Redhill dataset, linear buildup was chosen (a value of 1.0 for DDPOW). Values other than 1.0 will cause the equation to become non-linear and the parameter can be very sensitive. The sensitivity of DDPOW was measured using the same events as DDLIM and REGEN. Dry periods ranged from 0.25 d to 12 d and the DDPOW parameter was adjusted ±10%. Increasing the DDPOW parameter increases the computed peak SS concentration and load.

No buildup beyond a total mass of 1.0E6 pounds of SS was permitted as discussed earlier with the DDLIM parameter. For all events the parameter was sensitive for both computed peak concentration and load as objective functions. The peak sensitivity was found for the longest dry durations and largest evaporation totals. DDPOW was the most sensitive parameter of all examined in the Redhill Creek dataset. For twelve day dry periods a 10% parameter change resulted in a 187.3% change in the computed peak concentration of SS and a 159.% change in the computed SS load. Evaporation rates had only a very small influence in the concentrations and loads; greater input rates of evaporation caused higher computed concentrations and loads. The sensitivity of the parameter increased dramatically for durations beyond 1 day. The buildup limit was set so high in this dataset that there was no limiting effect on the sensitivity of the DDPOW parameter.

Graphical representation of the fuzzy relations

SV subspaces may be simply fuzzified as shown hypothetically in Figure 15.4. In the figure it is assumed that the sensitivity weight factor *w* will be evaluated only for a limited number of typical input functions representative of the fuzzy model states, and then *w* is then simply drawn in linearly.

Figure 15.4 Fuzzification of state-variable sub-spaces



A: flow over pervious areas; B: flow over impervious areas; C: recovery of storages and loss rates;
D: pollutant accumulation; E: pollutant washoff; F: recovery of depression storage;
G: groundwater depletion; H: rainout; I: erosion.

Figure 15.4 shows how the relative importance of nine arbitrarily chosen dominant processes might vary. In this case dominance is plotted as a weight factor *w* between 0 and 1. In this case for SV, we use duration of rain or ET. The applicability is 1 at the center of the applicable sub-range, and 0 at the center of adjacent sub-ranges. Exceptions are the processes of recovery of storages and loss rates, including groundwater depletion, which change more steeply with the onset or cessation of rain (in conventional hydrology, the change is assumed to be discontinuously abrupt). As is widely reported in the fuzzy logic literature, it is sufficient for the moment to assume that the relationship is linear. Detailed sensitivity analysis for every application will reveal the exact relation, and show that the relations are unique to the application, and also to the selected values of the range center points.

When the duration of rain or drought has a value that is not at the center of the sub-ranges, then the relative importance of the dominant processes changes, and adjacent dominant processes may become equally important.

Scores of processes are modeled in the entire SWMM suite. However, Kuch (1997) limited his work to a discussion of the RUNOFF processes that are modeled in his Redhill dataset, a total of seventeen parameters. The parameters can be grouped by macroscopic processes such as overland flow over the pervious area or pollutant buildup. The absolute parameter sensitivity is compared for parameters sensitive to a group of meteorological events. The sensitivity is ranked in Table 15.2 a and b, from most sensitive to least, for storms of short, medium, or long duration and low, medium, and high rainfall intensity.

Table 15.2a: Ranked peak flow and runoff volume parameter sensitivity for low events.

Note: The five most sensitive parameters are shown for peak flow and runoff volume, WAREA and %IMP are excluded from the table because they were always very sensitive.

	0. 10 in/h		0.25 in/h	
	Flow	Volume	Flow	Volume
Short				
.25 h	PCTZER	PCTZER	IMPDEP	IMPDEP
	IMPERN	IMPERN	PCTZER	PCTZER
	WIDTH	WIDTH	IMPERN	IMPERN
	WSLOPE	WSLOPE	WIDTH	WIDTH
			WSLOPE	WSLOPE
0 .5 h	PCTZER	PCTZER	IMPDEP	IMPDEP
	IMPERN	WIDTH	IMPERN	WSLOPE
	WIDTH	IMPERN	WIDTH	PCTZER
	IMPDEP		WSLOPE	PCTZER
Medium				
1.0 h	IMPDEP	IMPDEP	WIDTH	IMPDEP
	IMPERN	PCTZER	IMPERN	WSLOPE
	WIDTH	IMPERN	IMPDEP	IMPERN
	PCTZER	WIDTH	WSLOPE	WIDTH
	WSLOPE	WSLOPE	PCTZER	WSLOPE
3.0 h	IMPERN	IMPDEP	NONE	IMPDEP
	IMPDEP	PCTZER		PCTZER
	WIDTH	IMPERN		IMPERN
	WSLOPE	WIDTH		WIDTH
		WSLOPE		WSLOPE
Long				
6.0 h	NONE	IMPDEP	NONE	IMPDEP
		WIDTH		PCTZER
		PCTZER		IMPERN
		IMPERN		WIDTH
		WSLOPE		
12.h	NONE	IMPDEP	NONE	IMPERN
		IMPERN		WIDTH
		WIDTH		IMPDEP
		PCTZER		WSLOPE
		WSLOPE		

Table 15.2b: Ranked peak flow and runoff volume parameter sensitivity for medium and long events.

Note: The five most sensitive parameters are shown for peak flow and runoff volume, WAREA and %IMP are excluded from the table because they were always very sensitive.

	0.50 in/h		1.00 in/h		3.50 in/h	
	Flow	Vol	Flow	Vol	Flow	Vol
short						
.25 h	IMPDEP	IMPDEP	WIDTH	IMPDEP	WLMAX	WLMAX
	IMPERN	PCTZER	IMPERN	PCTZER	DECAY	DECAY
	WIDTH	IMPERN	IMPDEP	WIDTH	PERDEP	PERDEP
	PCTZER	WIDTH	PCTZER	IMPERN	WIDTH	PERVN
	WSLOPE	WSLOPE	WSLOPE	PERVN	PERVN	WIDTH
0 .5 h	WIDTH	IMPDEP	WIDTH	IMPDEP	WLMAX	WLMAX
	IMPERN	PCTZER	IMPERN	PCTZER	DECAY	DECAY
	IMPDEP	IMPERN	WSLOPE	WIDTH	WIDTH	WIDTH
	PCTZER	WIDTH	IMPDEP	IMPERN	PERVN	WLMIN
	WSLOPE	WSLOPE	PCTZER	PERVN	PERDEP	PERDEP
medium						
1.0 h	WIDTH	IMPDEP	WIDTH	IMPDEP	WLMAX	WLMAX
	IMPERN	PCTZER	IMPERN	PCTZER	DECAY	DECAY
	WSLOPE	IMPERN	WSLOPE	WIDTH	WIDTH	WLMIN
	IMPDEP	WIDTH	IMPDEP	IMPERN	PERVN	WIDTH
	PCTZER	WSLOPE	PERVN		PERDEP	PERDEP
3.0 h	WLMIN	WLMIN	WLMIN	WLMIN	NONE	WLMAX
	DECAY	DECAY	DECAY			WLMIN
	WLMAX	WLMAX	WLMAX			DECAY
	PERDEP	PERDEP	PERDEP			WIDTH
	PERVN	PERVN	WIDTH			PERDEP
long						
6.0 h	WLMIN	WLMIN	WLMIN	WLMIN	NONE	WLMAX
	DECAY	DECAY		DECAY		WLMIN
	WLMAX	WLMAX		WLMAX		DECAY
	PERVN	PERDEP		PERDEP		WIDTH
	WIDTH	WIDTH		WIDTH		PERDEP
12.h	WLMIN	WLMIN	WLMIN	WLNMIN	NONE	WLMAX
		DECAY		DECAY		WLMIN
		WLMAX		WLMAX		DECAY
		PERDEP		PERDEP		WIDTH
		WIDTH		WIDTH		PERDEP

Using the sensitivity results obtained by Kuch, the activity of the process and the sensitivity of the following groups of parameters are shown in fuzzy terms in Figure 15.5

Note: Separate plots are used for each of the following rainfall descriptions; depth, duration, and intensity.

A = impervious area parameters overland flow on the impervious area
B = pervious area parameters overland flow on the pervious area
C = routing parameters overland flow on all surfaces
D = Pollutant washoff parameters pollutant washoff from impervious area

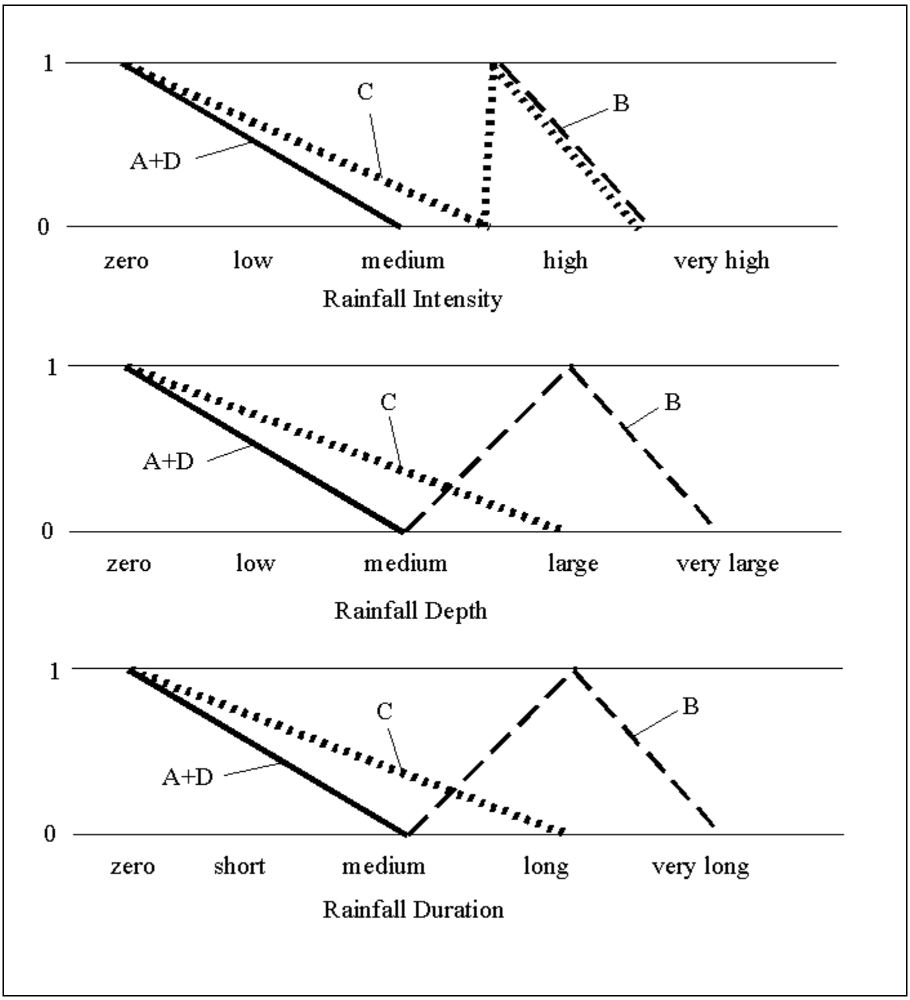


Figure 15.5 Fuzzification of the model processes (Kuch, 1997).

Calibration Procedure

Continuous SWMM models are not calibrated as continuous simulations because of the high parameter correlation and significantly long model execution times. Rather, as has been discussed earlier, SWMM models can be calibrated to a wide range of events that occur in the continuous TS. The following general procedure is suggested:

1. Identify
   a. the major modeled processes chosen in the data file,
   b. suitable objective functions,
   c. evaluation functions and
   d. parameters for sensitivity analysis and calibration.
2. Construct a series of synthetic storm events or extract suitable real storm events of varying duration, depths, dry periods and intensity to test the sensitivity of the selected parameters.
3. Perform a comprehensive sensitivity analysis with each parameter for all event types.
4. Plot or tabulate the parameter sensitivity gradients in a form that will allow the modeler to extract the input-variable state that causes the parameter to be most sensitive, and the gradient.
5. Within each meteorological type rank the parameters from most to least sensitive.
6. Select comparable measured rainfall-runoff-water quality sets from existing TS or discrete sampling.
7. Calibrate each of the following groups of parameters in the order presented starting with the most sensitive parameter in each group.
   A) Impervious Parameters (PCTZER, IMPERN, IMPDEP)
   B) Routing Parameters (WIDTH, SLOPE) - then recheck storms in A and redo B
   C) Pervious Parameters (Infiltration Parameters, PERVN, PERDEP)
   D) Recovery of Infiltration (REGEN)
   E) Pollutant Buildup (DDLIM, DDPOW, QFACT)
   F) Pollutant Washoff (RCOEF, WASHPO) - then recheck events in E and redo F
8. Verify all objective functions with other events if available and validate the continuous TS computed using the new optimized parameter set.
9. Restart from step three if the calibration is unsatisfactory.

15.3 De-fuzzification

Considering only the duration of rain or drought as the SV for now, the fuzzy rules may be derived from Figure 15.3 as follows:

general form:

If ***X*** period is ***Y***, analyze ***Z*** parameters.

where the ***X, Y, Z*** have the following set of nine meanings:

	X	***Y***	***Z***
1.	rain	long	erosion
2.	rain	medium	pervious area flow
3.	rain	medium	pollutant washoff
4.	rain	short	impervious area flow
5.	rain	short	rain-out
6.	evapo-transpiration	short	recovery of storages
7.	evapo-transpiration	medium	recovery of loss rates
8.	evapo-transpiration	long	groundwater depletion
9.	evapo-transpiration	medium	pollutant build-up

For this SV, extended rules can be written for the other intermediate durations, interpolating the weight factor for each process according to whatever relation has been found to be appropriate.

And then, of course, similar sets of rules, and extended rules, can be derived for all appropriate processes and SVs.

> Analyse only sensitive parameters, and then only against relevant events.

Additional rules are then also developed:

1. If the product $DSG.(E_{max} - E_{min})$ for any parameter is less than the required accuracy, eliminate that parameter from further analysis (note: you should previously establish the requisite accuracy - see James and Robinson, 1985). Note that the entire process can also be zeroed (meaning deleted) out of the model in many cases.

2. Perform the rest of the calibration and error analysis in descending order of DSG, starting with the most sensitive parameter. When the requisite model reliability is achieved, or when the incremental improvements are less than the required accuracy, curtail the analyses (again, your study group must previously have set acceptable model reliability limits). In the overall method shown

schematically in Figure 15.6, fuzzy methods for calibration take place in the validation box, and fuzzy methods for presentation of model reliability in the inference box.

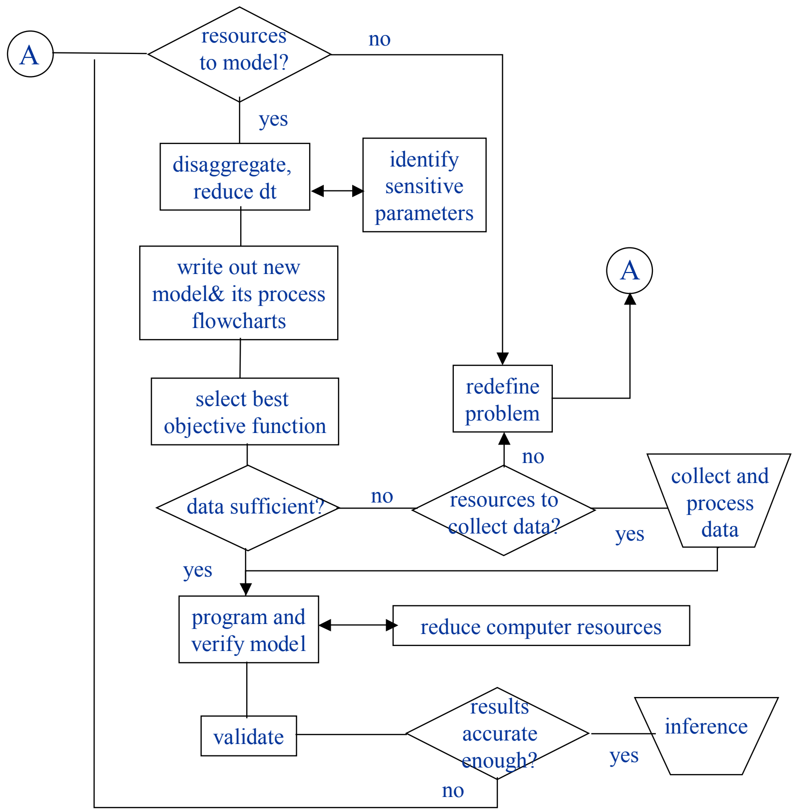


Figure 15.6 Schematic showing the sequence of modeling activities

15.4 Conclusion

The fuzzy rules that were developed in this chapter are programmed into the PCSWMM. The shell suggests various OFs based on the reasoning deduced in this book. The shell defines various SV spaces, and suggests appropriate sub-spaces for various processes and parameters. Some of the processes were pre-selected, based on their inherent uncertainty, and vice versa. None of the rules are rigidly enforced in the shell.

References

Kuch, A. J. 1994. A sensitivity analysis shell for complex continuous water quality management modelling. M.Sc. Thesis. Univ of Guelph. ca.120 pp. January 1994.

James, W. and Robinson, M.A., 1987. "Hydrologic Modelling of Continuous Urbanisation and Consequent Effects Downstream", 8th Can. Hydrotech. Conf., Montreal, Quebec, pp. 524-531, May 17-22, 1987.

Bibliography

Ayyub, B.M. and McCuen, R.H. 1985. Quality and uncertainty assessment of wildlife habitat with fuzzy sets. J. Water Resources Planning and Management. 113(1):95-109.

Bardossy, A. and Disse, M. 1993. Fuzzy rule-based models for infiltration. Water Resources Research 29(2)373-382.

Dunn, A.R., 1986. Feasibility of a Real-Time Auto Sensitivity and Error Analysis Framework for Fuzzy Stormwater Modelling on Personal Computers, M.Eng. Thesis, McMaster U., May 1986, 130pp.

Dunn, A.R., and James, W., 1985. Extended Expert Systems. Proc. of the Stormwater and Water Quality Modelling Conf., sponsored by USEPA and Ontario Ministry of the Environment, Toronto, Canada, published by CHI, Hamilton, pp. 373-380, Dec 5-6, 1985

Dunn, A.R., & James, W. 1986 "Real-Time Error and Sensitivity Analysis to Help Engineers Use Complex Programs Expertly", Int. Symp. on Water and Data Processing, Ecole Nationale Des Ponts Et Chaussees, Paris, France, pp. 196-205, May 28-30, 1986.

James, W. and Dunn, A., **1986** "An Integrated Conceptual Expert System for Flood and Water Pollution Management", Water Forum '86, A.S.C.E., Long Beach, California, pp. 158-165, Aug 4-6, 1986.

James, W. and Dunn, A., **1985**. "Integrated Knowledge-based Software for Flood and Water Pollution Management", ASCE Water Resources Planning and Management Specialty Conf. on Computer Applications in Water Resources, Buffalo, New York, pp. 611-619, Jun 10-12, 1985.

Jowitt, P.W., **1984**. Risk analysis, fuzzy logic, and river basin management, *Water Sci. Technol., 16*(5-7), 579-585, 1984.

Kindler, J. 1992. Rationalizing water requirements with aid of fuzzy allocation model. J. Water Resources Planning and Management. 118(3):308-323.

Kuch, A.J. and James, W. 1993. A sensitivity analysis tool for three generation continuous urban storm water quality management modelling. Runoff Quantity and Quality Model Group Conference, Sponsored by ASCE . Reno NV, Nov 8-9 1993. *in press.*

McNeil, D. and Freiberger, P. 1993. Fuzzy Logic. Simon and Schuster, New York, NY, 1993 319 pp.

Morita, M., Jinda, S. and Izumi, K. 1993. Application of fuzzy theory to urban runoff prediction. 6th Int. Conf. on urban storm drainage. ASCE. Vol 1: pp 820-825.

Savic, D.A. 1990. An intelligent decision support system and use of fuzzy sets for reservoir analysis. Ph.D. thesis. U. of Manitoba. 253 pp.

Chapter 16

PRESENTING CONTINUOUS UNCERTAINTY AND MODEL RELIABILITY IN REAL TIME

The next great awakening of human intellect may well produce a method of understanding the qualitative content of equations. Today we cannot see that the water flow equations contain such things as the barber pole structure of turbulence that one sees between rotating cylinders. Today we cannot see whether Schrodinger's equation contains frogs, musical composers, or morality – or whether it does not.

- Richard P. Feynman.

16.1 Introduction

In this chapter we present proposals for computing and presenting various measures of model performance. By *real time* we mean that the measures are presented while the model response is being computed, as part of the same plot – the computation is not delayed by separately requesting, computing and plotting the additional information (normally models plot only the computed response function). Besides the expected response plotted as a time series, the additional measures of interest include: the range of expected response (model uncertainty); observed response function also as a time series; model reliability as a continuous time series (indicating the degree to which the model is reproducing the observed functions); model state as a discrete time series (indicating which processes are dominant, active or dormant); observed rainfall and other input functions as time series; and parameter sensitivity as a discrete time series. Figure 16.1 (Wan, 2001) is an example of such a plot.

Dunn (1986) asked how such a plot might be generated in real time. The following steps are proposed:

1. Scan the input functions for rain and dry weather types
2. Classify the wetness/dryness types
3. Correlate the wetness/dryness types with runoff
4. Perform sensitivity analysis
5. Optimize the uncertain parameters
6. Determine from the SA the dominant, active and dormant processes

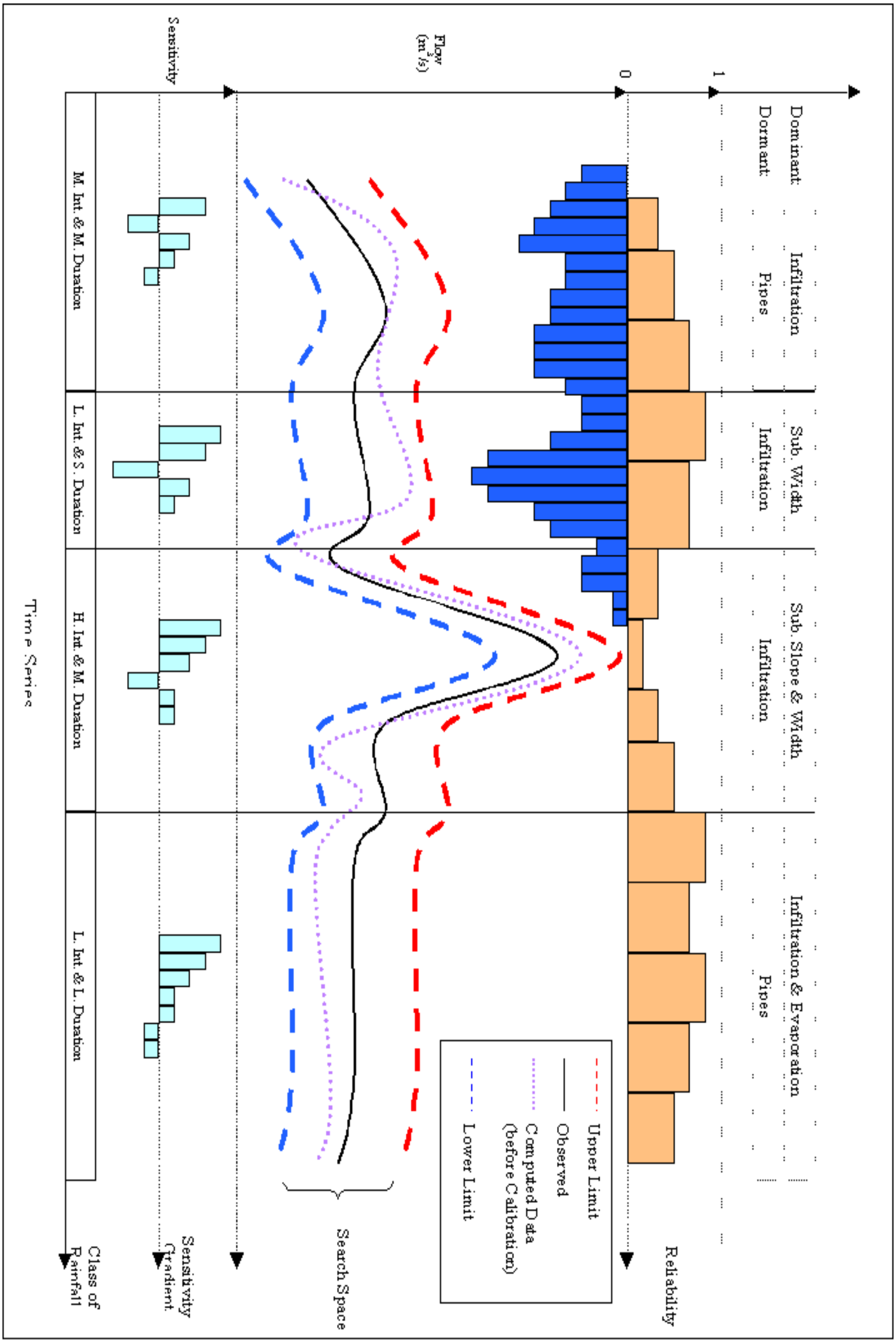


Figure 16.1: Presentation of model states, reliability and uncertainty.

7. Determine the fuzzy relations for relative parameter sensitivity for various wetness/dryness states
8. Use defuzzification logic and linear error to compute uncertainty
9. Compute model reliability for each state
10. Plot them.

Sensitivity and error analysis are the key procedures used.

16.2 Parameter uncertainty

In PCSWMM, parameter uncertainty, called "input doubt" by Dunn (1986), is defined as the range between a user's lowest and highest estimates for the true value of a parameter. Parameter uncertainty is thus:

$$\Delta p = \left| p_{\max} - p_{\min} \right|$$

where p_{max} = user's maximum estimate
p_{min} = user's minimum estimate.

For computing sensitivity and model error, PCSWMM uses only part of the parameter uncertainty as follows:

Uncertainty half ranges:

$$\Delta p_{1/2} = E[p] - p_{\min}; \quad p_{\max} - E[p]$$

where $E[p]$ = expected value of the parameter also estimated by the user.

Uncertainty quarter-range: half ranges plus:

$$\Delta p_{1/4} = E[p] - 0.5\, p_{\min}; \quad 0.5\, p_{\max} - E[p]$$

Parameter uncertainty can be elicited from users by requesting estimates of the approximate upper and lower bounds p_{max} , p_{min} of each parameter. The estimate for which the user is 95% certain the true value is not above can be converted into a statistically relevant term. Specifically, the difference between the user's best estimate and an estimate which s/he is 95% certain the true value is below, represents a range of two standard deviations in the estimate for normal distributions. Similarly, the difference between the best estimate and a value which the true value is 95% likely to be above also represents two standard deviations for normal distributions. Figure 16.2 shows the statistical relationship between confidence levels and standard deviations.

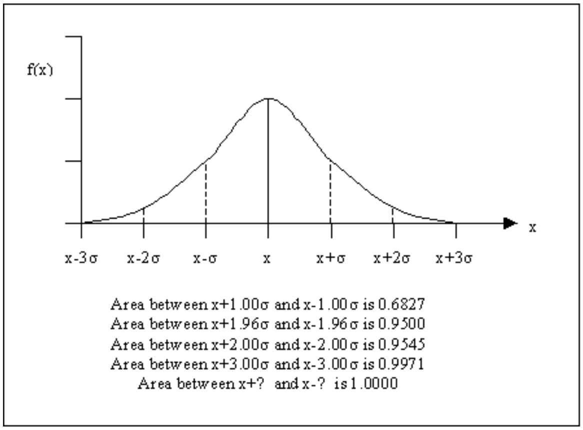


Figure 16.2: normal distribution

Both upper and lower bounds are required since the distribution may be skewed, in which case, the standard deviation is calculated for an approximate skewed distribution. For example, Figure 16.3 shows that for Manning's roughness a best estimate coefficient is 0.015, that the true value will most likely be above 0.010 but will also probably be below 0.025 at the 95% confidence level.

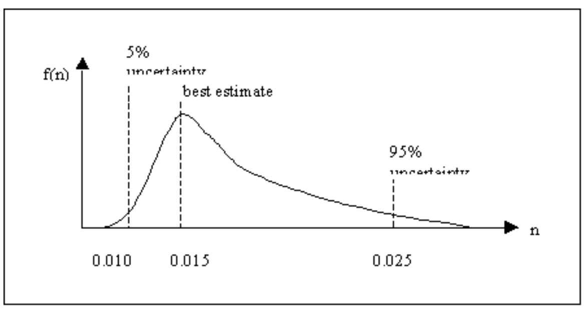


Figure 16.3: Typical estimates for Manning’s n, a skewed parameter.

By prompting for these three values for each of the eleven parameters PCSWMM collects data regarding statistical moments for the parameter estimates (the mean value and upper and lower standard deviations). This data represents parameter uncertainty.

16.3 Sensitivity analysis

If the objective function, OF_j , j = 1,2 .... m, is dependent on a set of constant parameters, p_i, i = 1,2,...n, then the governing relation can be differentiated with respect to the constant parameters, p_i, to create a set of sensitivity gradients,

$$SG_{ij} = \frac{\partial OFj}{\partial p_i}.$$

While most sensitivity analyses have been connected with sensitivity gradients, other methods have been developed (Koda et al., 1978). Parameters which are related to spatial or temporal variables create greater complexity in sensitivity analysis than do parameters which are constants throughout a simulation. This was investigated by Porter (1967) who proposed that parameters could be represented by expansions in the orthogonal basis functions. Three methods of sensitivity analysis include: 1. the direct method, 2. the variational method, and 3. the Fourier Amplitude Sensitivity Test (FAST) (Koda et al., 1978). While direct and variational methods are linearized theories, the FAST method is valid for non-linear situations. Direct and variational methods are applicable when the parameter uncertainties are relatively small. The FAST method will work well when large uncertainties exist. Koda et al. (1978) found that any of these three methods is capable of providing similar information and that the major consideration in choosing one is computational efficiency.

1. The direct method requires all parameters to be constant throughout the simulation. The method requires the solution of two sets of ordinary differential equations, the system and the sensitivity equations. Three dimensional problems could create large storage (memory) requirements.
2. The variational method provides a good sensitivity measure but requires the solution of two sets of partial differential equations which may be lengthy.The method calculates sensitivity gradients with the use of adjoint variables. The distributed nature of the variables is maintained and the final answer requires the solving of the state and adjoint equations (Koda et al., 1978).
3. The FAST method requires auxillary calculations, such as the Fourier amplitude. Additionally, the solution of the system may require repeated iterations. The computation of the amplitudes of an expansion of the system state forms the basis of this method. Cukier, et al. (1973) devised a system which assigns the periodic function of a new variable s to each parameter $p_i(s)$. Distributions are assigned to the parameters and correspond to the periodic functions. The model produces a quasi-periodic function $u(s)$. The Fourier amplitudes of

u(s) are computed. The dependence of the solution on p_i is indicated by the generation of the frequency *w* for the parameter under consideration.

A statistical approach may be used for simple hydrologic models. The central limit theorem states that the sum of a large number of individual random components will tend toward a normal distribution as the number of components increases (Ang and Tang, 1975). In a hydrological situation, the probability distribution for outflows will tend to a normal distribution since the models are usually multivariate.

A derivation from first order error analysis is the normalized or dimensionless sensitivity coefficient matrix (Walker, 1981, 1982):

$$DSG_{ij} = \frac{\partial OF_j}{\partial p_i} \bigg/ \frac{p_i}{OF_j}$$

The sensitivity of a parameter relates the amount of change in the objective function caused by a given perturbation to one of the input parameters. Specifically, the matrix can be described by the percentage change in a selected output variable caused by a 1% change in the selected input parameter (Walker, 1981, 1982).

16.4 Error analysis

First order analysis requires the use of a Taylor series expansion. If the flow, OF_t, is a function of the driving force precipitation, *IF*, where variations are not large, using the expected value and variance:

$$E[OF_t] = E[M_c(p)] = M_c(E[p]) + \frac{1}{2}Var[p]\frac{\partial^2 M_c}{\partial p^2}$$

and

$$Var[M_c(p)] = \left[\frac{\partial M_c}{\partial p}\right]^2 Var[p] - \frac{1}{4}Var^2[p]\left[\frac{\partial^2 M_c}{\partial p^2}\right] +$$

$$E[p - E[p]]\frac{\partial M_c}{\partial p}\frac{\partial^2 M_c}{\partial p^2} + \frac{1}{4}E[p - E[p]]^4\left[\frac{\partial^2 M_c}{\partial p^2}\right]^2$$

Usually truncation eliminates the third and fourth moments leaving:

$$Var[OF_t] = \left[\frac{\partial^2 M_c}{\partial p^2}\right] Var[p]$$

Numerical differentiation may be required for complex systems. First order analysis is most easily adapted to complex models.

Burnash and Ferral (1982) studied the hierarchy of parameters in the Sacramento Model, a time dependent water balance system. They noted that although a set of parameters that represents the best fit for an observed data set was unlikely, a reasonable parameter set is usually attainable. In the Sacramento Model there were seventeen parameters. After elimination of insignificant parameters and parameters which could be analytically obtained, only six were found to be valid candidates for an optimization search. The sensitivity of these six was not however significant when compared to the driving force uncertainties. The authors noted that a 1% change in the precipitation input would cause the same distortion of the computed output as a 10% change in the most sensitive parameter. A 10% change in rainfall generated the same error as a 59% change in the most sensitive parameter. Their conclusion was that the concern of modelers over parameter sensitivity and uncertainty is uncalled for, considering the sensitivity of the model to the input function.

First-order error analysis procedures provide the mean and an approximate standard error for the objective functions. The first order approximation for the variance of the response, RF, can be expressed (Benjamin and Cornell, 1970):

$$Var[RF] = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial M_c}{\partial p_i}|_m \frac{\partial M_c}{\partial p_j}|_m Cov[p_i, p_j]$$

If the p are not correlated then the expression can be simplified to:

$$Var[OF] \cong \sum_{i=1}^{n}\left(\left(\frac{\partial M_c}{\partial p_i}\right)|_m\right)^2 Var[p_i]$$

where, $\frac{\partial M_c}{\partial p_i}$ is the sensitivity of the computed objective function to input parameters p_i. Each of the input variables contributes an amount proportional both to its own variance and to the square of its sensitivity.

Three methods of error analysis are widely reported in the literature: Kalman filters, first-order sensitivity-error-analysis, and Monte Carlo. No use of Kalman and/or Monte Carlo methods on long-term, continuous applications of surface WQMs has been reported, so far as this literature search has revealed. On the other hand, it seems that for most applications, first-order approximations are adequate, if not preferable, because:

1. the input variable error and model error estimates are themselves approximate, and
2. the first-order procedure readily ranks parameter sensitivity and uncertainty (Walker, 1982), which are useful for aiding in understanding the model performance.

The output function OF_i is a function of the n model parameters $p_1, p_2, , , p_n$:

$$OF_j = M_{cj}\left(p_1, p_2, \ldots p_n\right)$$

where M_{jc} denotes a function dependent on the state j of a model M of complexity c.

In general, the absolute error in the computed objective function caused by parameter variation is given by the first-order error equation:

$$\varepsilon\left\lfloor OF_{ij} \right\rfloor = SG_1 \Delta p_1 + SG_2 \Delta p_2 + \ldots + SG_n \Delta p_n$$

where ε is the error, SG_i is the dimensional sensitivity gradient for parameter p_i, and Δp_i is the perturbation in the input parameter. PCSWMM uses the half-range and quarter-range in the estimates, for both positive error and for negative error as indicated above.

As expressed here, the error arises by virtue of reasonable estimates by different users, and, if both the negative and positive errors are computed, will represent a range of about four standard deviations for each term (assuming the parameter estimates were normally distributed; this may also apply to other distributions).

A remaining conundrum is the question of the direction of the errors summed in the last equation. Since uncertainty is the theoretical range of the error, the error should be maximized, so that directions are chosen to increase the uncertainty. This seems not to take account of the small likelihood of maximizing all errors. PCSWMM therefore allows users to select a plotted range less than the maximum (e.g. one half).

There is another way to deal with the difficulty of the low probability that 1. a user may have estimated every one of the parameters involved to compound the effect of the sign of the SG, overestimating positive SGs and underestimating negative SGs, or 2. have estimated half of them up and half down, thus producing no difference ultimately between the computed and observed OF. A simple Monte Carlo technique may be used: generate a random sequence of binary positive/negative decisions, apply them serially to all of the parameter half-ranges and then substitute them into the first-order error equation. This will be relatively fast, because the error equation is explicit and simple, and the sensitivity gradients will be known, so that a large number of runs will present no difficulty. This modified Monte-Carlo/first-order-error procedure, and the requisite routines, are currently under development in PCSWMM.

Use first-order error analysis to report the estimated propagated error in your recommended design solution.

References

Ang, Alfredo H-S. and Tang, Wilson H., (1975). "Probability Concepts i Engineering Planning and Design, Volume I", John Wiley and Sons, Ne York.

Benjamin, J.R. and Cornell, C.A., (1970). "Probability, Statistics and Decision for Civil Engineers", MoOraw Hill, New York.

Burnash, Robert J.C. and Ferral, R.L., (1982). "The Significance ol Various Parameters Estimated in Fitting the Sacramento Model", Proceedings of the Fall Meeting of the American Geophysical Union, Sat Francisco, Califonia.

Cukier, R.I. et al., (1973). "Study of the Sensitivity of Couplec Reaction Systems to Uncertainties in Rate Coefficients", Journal of Chemical Physics, Vol. 59, No. 8, pp. 3873 - 3878.

Dunn, A.R., 1986. Feasibility of a Real-Time Auto Sensitivity and Error Analysis Framework for Fuzzy Stormwater Modelling on Personal Computers, M.Eng. Thesis, McMaster U., May 1986, 130pp.

Dunn, A.R. and James, W., (1985). "Extended Expert Systems", Proceedings of the Stormwater and Water Quality Modelling Conference, sponsored by the U.S. EPA and Ontario Ministry of the Environment, Toronto, Canada, published by CHI, Hamilton, December 5 - 6, pp. 373 - 380.

Koda, M., Dogru, A.H. and Seinfeld, J.H., (1979). "Sensitivity Analysis of Partial Differential Equations with Application to Reaction and Diffusion Processes", Journal of Computational Physics, Vol 30. pp. 259 -282.

Porter, W.A., (1967). "Sensitivity Problems in Distributive Systems", International Journal of Control, Vol. 5, No. 5., pp. 393 - 412.

Walker, W.W. 1982a. A sensitivity and error analysis framework for lake eutrophication modelling. Water Resources Bulletin. 18(1):53-60.

Walker, W.W. 1982b. MODELER a computer modeling and data management system. Ver 4.0 Mar. 1982. W.W.Walker, Concord, MA. 21pp.

Wan, Benny. 2001. Auto-calibration of SWMM-RUNOFF using sensitivity-based genetic algorithms. MSc dissertation, U of Guelph, Canada. Ca 150pp.

Bibliography

Andersson, Lotta. 1992. Improvements of runoff models What way to go? Nordic Hydrology, 23:315-332.

Beck, J.V. and K.J. Arnold, **1977**: Parameter Estimation in Engineering and Science. John Wiley and Sons, Inc. NY.

Benjamin, J.R., and C.A. Cornell, **1970.** *Probability, Statistics, and Decision for Civil Engineers,* McGraw-Hill, New York, 1970.

Bergstrom, S. 1991. Principles and confidence in hydrological modelling. Nordic Hydrology, 22: 123-136.

Clarke, R.T. 1971. The Use of the Term 'Linearity' as Applied to Hydrological Models. J. of Hydrology, 13, pp.91-95.

Crawford, N. 1993. The next major breakthrough... Paper presented at the U.S.EPA workshop on Developing dynamic watershed modelling capabilities for the Great Lakes tributaries. Heidelberg College, Tiffin, Oh. August 4-6, 1993. 8pp.

Diskin, M.H. and Simon, E. 1977 A procedure for the selection of objective functions for hydrologic simulation models. J. of Hydrology, 9:34-56.

Garrick, M., C. Cunnane, and J.E. Nash, 1978: "A Criterion of Efficiency for Rainfall-Runoff Models." *Journal of Hydrology,* 36:375-381.

Griffin, S., Bauwens, W, and Ahmad, K. 1993. UDMIA: urban drainage modelling intelligent assistant. 6th Int. Conf. Urban Storm Drainage, Niagara Falls, Ontario. Sept. 1993. pp 1314-1319.

James, W. and Dunn, A., **1986** "An Integrated Conceptual Expert System for Flood and Water Pollution Management", Water Forum '86, A.S.C.E., Long Beach, California, pp. 158-165, Aug 4-6, 1986.

James, W. and Dunn, A., **1985**. "Integrated Knowledge-based Software for Flood and Water Pollution Management", ASCE Water Resources Planning and Management Specialty Conf. on Computer Applications in Water Resources, Buffalo, New York, pp. 611-619, Jun 10-12, 1985.

Rivera Santos, J. 1988. A methodology for parameter estimation in conceptual precipitation-runoff models with emphasis on error analysis. Ph.D. thesis. U. of Colorado Boulder. 279 pp.

Chapter 17

CONCLUSIONS AND RECOMMENDATIONS

It has been said that the computer is incredibly fast, accurate and stupid; users are unbelievably slow, inaccurate and brilliant - the combination is an opportunity beyond imagination. – source unknown

17.1 Framework for continuous modeling:

Using the logical approach described in this book, the number of runs for sensitivity, parameter and complexity optimization can be reduced to manageable amounts. The methodology has been found to be useful for honest modeling. It is summarized here as a set of rules. The method is based on the not unreasonable premises that the WQM has been structured such that component processes do not interfere with one another unpredictably, and that the overall model behaves as if it were largely linear.

At your desk:

1. Make a list of simplified design questions, and postulate the relationship between your list and your proposed objective functions.
2. Select the best objective functions and response functions for your study problem. Minimize the computed output and computer execution times. Allocate computer storage for computed time series management.
3. Obtain or generate a credible, very-long-term time series to drive your model for design inference.
4. Obtain a short but sufficient record of good, observed events to calibrate your as-is model.

Using the PCSWMM shell:

A. Calibrate:

5. Estimate the uncertainty for all input parameters.
6. List all parameters that need to be optimized, and their associated processes.
7 Associate all processes with the limited state-variable sub-spaces where they dominate.
8. Search the good observed record for a sufficient number of appropriate events.
9. For each of all input parameters, estimate:
   1. the expected (mean or most likely) value,
   2. a higher most likely value,
   3. a lower most likely value, and
   4. choose the sensitivity test range, but keep it small.
10. Rank all parameters, in terms of their dimensionless sensitivity gradients.
11. Optimize the parameters to give the smallest error.

B. Design:

12. Run the calibrated model for the long term time series for the as-was condition and each to-be array of BMPs.
12. Infer which is the best design, or array of BMPs.
13. Rerun the model for this array estimating the error in the computed response functions.
14. Study all the input and output information again; make certain that it is logical, and gain knowledge about the performance of the drainage system.
15. Interpret the impact of the errors.

At your client's office:

16. Report your recommendations, and,
17. If you follow the logic, become rich and famous.

17.2 Recommendations:

The following rules, taken from a personal catechism for honest, very-long term, continuous surface water quality modeling, have been espoused in this book.

Precepts

***Rule 1**:*
It does not make sense to test a WQM *per se* for sensitivity, parameter optimization, or error, because individual applications are likely to be radically different. It is fundamentally important to realise that the values of the parameters in the input datafile influence which processes are relatively active - from dominant to dormant - and that their relative values change both the model sensitivity and the model error.

***Rule 2**:*
A model is used to help select the best among competing proposals. It is fundamentally irresponsible and unethical for modelers not to interpret the inherent uncertainty of their model output, since this information is at least as important to decision makers as the computed expected response.

***Rule 3**:*
The basic design questions must be simplified, and directly related to the model objectives. The model must include code that adequately describes the important processes.

***Rule 4**:*
Carefully choose the best objective functions that represent the design questions and the model variability. Get the advisory committee to justify the selections in writing.

Rule 5:

The essence of good design is to know what lies behind the impressive computed output. Precisely what mathematics, algorithmic structure and code allow us to infer that an aquatic ecosystem will or will not be remediated as a result of the *to-be* arrays of BMPs?

Rule 6:
Computed and observed time series are more ethically represented as smudges than single-valued lines.

Rule 7:

Because of:

1. the modern wide distribution of inexpensive computers;
2. freely available knowledge and information on continuous modeling;
3. the urgent need to develop ecosystem sensitive methods; and

4. the informed engineering community, itself the product of excellent higher educational institutions, and an informed society; - there can no longer be any case for event-hydrology methods.

Rule 8:
Never use event hydrology for design. For continuous modeling, a short observed input and equivalent response time series, containing only sufficient events to calibrate all parameters, is required at or near the study site; the inferential long-term input time series may be transposed or generated from external sources.

Rule 9:
Continuous modeling requires a sequence of two main sets of modeling activities:

1. calibration or parameter optimization, and
2. inference or design:

calibration activities involve parameter estimation and optimization against short-term, accurate, observed input functions;
inference activities involve long-term, continuous, synthetic or transposed input functions, and error analysis.

Calibration:

Rule 10:
Do not calibrate all parameters simultaneously against a long-term, continuous, observed record, notwithstanding any early advice to the contrary in the literature.

Rule 11:
Transpose or synthesize a long-term, hydro-meteorologic input time-series from the same hydrologic region, and use this for inferring comparative performance of various arrays of BMPs. Many records of 50 years duration or longer are available.

Rule 12:
In order to control the amount of computing, associate the input parameters with processes, and processes with causative events, and causative events with limited state-variable sub-spaces. For this activity, sensitivity analysis code in PCSWMM is helpful. Do not analyze parameters outside these spaces.

Rule 13:
In determining the best level of complexity, test simple models first, proceeding to more complex, until the required accuracy of the computed response function

is achieved. Use the least number of processes, discretized spaces, and the biggest time step that delivers the required accuracy.

Rule 14:
Use three estimates of the most likely parameter values. It is more meaningful to compare the computed response from several reasonable estimates, rather than responses computed using plausible extreme values.

Rule 15.
Assume that the WQM is approximately linear, for the purposes of optimizing parameters, and estimating the propagated error. Then analyze for sensitivity near the mean expected values of all input parameters.

Rule 16:
Calibrate only sensitive parameters, and then only against relevant events for which you have good, short-term observed data.

Rule 17:
Make certain that you have good rate-of-rain observations with adequate coverage and spatial resolution.

Design

Rule 18:
Sixteen sources of error are listed in the framework for uncertainty analysis presented here. When interpreting the computed output from your model, all sixteen sources should be explicitly interpreted.

Rule 19:
Use first-order linear error analysis, and report the estimated propagated error in your recommended design solution. Plot results depicting their expected range.

The last word

Models are made by humans for other humans to enable understanding. Any model that improves understanding is a good model. - Anon.

The purpose of models is not to fit the data but to sharpen the questions. - Samuel Karlin

The sciences do not try to explain, they hardly even try to interpret, they mainly make models. By a model is meant a mathematical construct which, with the addition of certain verbal interpretations, describes observed phenomena. The justification of such a mathematical construct is solely and precisely that it is expected to work. - John von Neumann

Progress does not consist of replacing a theory that is wrong with one that is right. It consists of replacing a theory that is wrong with one that is more subtlety wrong.. - Hawking's Theory of Progress

As complexity rises, precise statements lose meaning, and meaningful statements lose precision. - Lotfi Zadeh

What is wanted is not the will to believe, but the will to find out, which is the exact opposite. - Bertrand Russell "Skeptical Essays" 1928

Obesa cantavit

References

Author? 1997. Analysis of the 1998 water-distribution system serving the Dover Township area, New Jersey. Water Science and Technology Vol 36 No 5 pp 141-148.

Abbott, M.B. 1991. Hydroinformatics: information technology and the aquatic environment. Avebury Technical. ISBN 1-85628-832-3. 157pp

Adderley, V.C. and Liebe, M.A. 1993. Development of a SWMM graphical interface for Portland CSO models. Preprint for 1993 Runoff Quantitiy and Quality Modeling conference. ASCE Reno, Nev. Nov 1993.

Aitken, A.P., 1973. Assessing Systematic Errors in Rainfall-Runoff Models. *Journal of Hydrology,* 20:121-136.

Alley, W.M., and Smith, P.E. 1981. Estimation of Accumulation Parameters for Urban Runoff Quality Modeling. Water Resources Research, Vol.17, No.6, pp.1657-1664. Dec. 1981.

Alvarado, L.G. 1982. Calibration strategies for SWMM modelling. Proc. Stormwater and water quality model users group meeting, Oct 18-19, 1982. Ottawa, Ontario. pp 36-50.

Amorocho, J. and B. Espildora, 1973: "Entropy in the Assessment of Uncertainty in Hydrologic Systems and Models." *Water Resources Research,* 9(6):1511-1522.

Andersson, Lotta. 1992. Improvements of runoff models What way to go? Nordic Hydrology, 23:315-332.

Angulo, C.E. Puente, and Bras, R.L. 1984. Non-linear filtering, parameter estimation and decomposition of large rainfall-runoff models. Report 297, (R-84-13; OSP 89310) Ralph M. Parsons laboratory, MIT. Oct 1984.

Anthony, A.L. and Watson, H.K. 1972. Techniques for Developing Analytic Models. IBM Systems Journal, Vol.11, No.4, pp.316-328.

Ayyub, B.M. and McCuen, R.H. 1985. Quality and uncertainty assessment of wildlife habitat with fuzzy sets. J. Water Resources Planning and Management. 113(1):95-109.

Baffaut, C. 1988. Knowledge techniques for the analysis of urban runoff using SWMM. Ph.D. thesis. Purdue U. 321 pp.

-----, and Delleur, J.W. 1989. Expert system for calibrating SWMM. J. Water Resources Planning and Management. 115(3): 278-298.

-----, &-----, 1990. Calibration of SWMM runoff quality model with expert system. J. Water Resources Planning and Management. 116(2): 247-261.

Bao, Y-B, and Thunvik, R. 1991. Sensitivity analysis of groundwater flow. Nordic Hydrology, 22:175-192.

Bardossy, A. and Disse, M. 1993. Fuzzy rule-based models for infiltration. Water Resources Research 29(2)373-382.

Baumgardner, J., Ruby, A., Walker, M. , and Brent, D. 1993. Discharge characterization of urban stormwater runoff using continuous simulation. 1993 Runoff quantity and quality modelling conf., ASCE, Reno, Nev. Nov 1993. Preprint. 15pp.

Baumgardner, J., Ruby, A. and Walker, M. 1993. Discharge characterization of urban storm water runoff using continuous simulation. Chapter 15 in: Current practices in modelling the management of stormwater impacts. Pub. by Lewis publishers. ISBN 1-56670-052-3. pp243- 256.

Bayes, A.J. 1972. A Minimum Variance Sampling Technique for Simulation Models. J. of the Assoc. for Computing Machinery, Vol.19, No.4, pp.734-741, Oct. 1972.

Bayes, A.J. **1970.** Statistical Techniques for Simulation Models. The Australian Computer Journal, Vol.2, No.4, pp.180-184. Nov. 1970.

Beck, M.B., **1980**: "Application of System Identification and Parameter Estimation in Water Quality Modelling." In: Hydrological Forecasting, Proceedings of the Oxford Symposium, April, 1980, *International Association of Hydrological Sciences,* IAHS, 129:123-131.

Beck, M.B., **1985.** Structures, failure, inference, and prediction, in *Identification and System Parameter Estimation,* edited by H.A. Barker and P.C. Young, pp.1443-1448, Pergamon, New York, 1985d.

Beck, M.B. 1987. Water quality modelling: a review of the analysis of uncertainty. Water Resources Research 23(8):1393-1442. August 1987.

Beck, M.B. 1976. Identification and parameter estimation of biological process models. In:Systems simulation in water resources, G.C.Vansteenkiste, ed. North-Holland. pp 19-43.

-----, 1983. Uncertainty, system identification, and the prediction of water quality. In: Uncertainty and forecasting of water quality, M.B. Beck and G. van Straten, eds. Springer-Verlag 1983. pp 3-68.

Beck, J.V. and K.J. Arnold, **1977**: Parameter Estimation in Engineering and Science. John Wiley and Sons, Inc. NY.

Benjamin, J.R., and C.A. Cornell, **1970.** *Probability, Statistics, and Decision for Civil Engineers,* McGraw-Hill, New York, 1970.

Bergstrom, S. 1991. Principles and confidence in hydrological modelling. Nordic Hydrology, 22: 123-136.

Berry, W. cited by Orr (1992)

Bhaskar, N.R., James, W.P. and Devulapalli, R.S. 1992. Hydrologic parameter estimation using geographic information system. J. Water Resources Planning and Management. 118(5):492-512.

Biswas, A.K. 1976 Mathematical modelling and water resources decision making. In: Biswas (Ed), Systems approach to water management. McGraw-Hill NY pp398-414.

Blockley, D.I., Pilsworth, B.W. and Baldwin, J.F. 1983. Measures of Uncertainty. Civil Engineering Systems, Vol.1, pp.3-9, Sept. 1983.

Bower, D.E. 1985. Evaluation of the Precipitation-runoff modelling system, Beaver Creek, Kentucky. USGS Water Resources Investigations report 84-4316. 1985. 39pp.

Brabets, T.P. 1987. Quantity and quality of urban runoff from the Chester Creek Basin, Anchorage, Alaska. USGS Water Resources Investigations report 83-4238. 1983. 207pp.

Brendecke, C.M. Laiho, D.R., and Holden, D.C. 1985. Comparison of two daily streamflow simulation models of an alpine watershed. J. Hydrology. 77:171-186. 1985.

Brown, L.C. 1985. Uncertainty analysis using QUAL2E. Proc. Stormwater and water quality model users group meeting, Dec 5-6 1985 pp 125-140.

Brown, L.C. and Barnwell, T.O. 1987. The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS: documentation and user manual. U.S.EPA Report EPA/600/3-87/007. 199pp.

Brundtland, H., 1987. Our Common Future - World Commission on Environment and Development, United Nations Brundtland Commission 1987. pp. ?.

Burn, D.H. **1985** Optimization Techniques Applied to Water Quality in an Uncertain Environment. Ph.D., Univ. of Waterloo, 1985.

Calver, A. 1988. Calibration, sensitivity and validation of a physically-based rainfall-runoff model. J. of Hydrology. 103:103-115.

Carey, W.P. and Simon, A. 1984. Physical basis and potential estimation techniques for soil erosion parameters in the precipitation-runoff modelling system. USGS Water Resources Investigations report 84-4218. 32pp 1984.

Chapman, T.G. 1968. Catchment Parameters for a Deterministic Rainfall-Runoff Model. In: 'Land Evaluation', MacMillan of Australia. pp.312-323.

Chaudhury, R.R. 1993. Stormwater modelling of a combined sewer system. Ph.D. thesis. U. of Rhode Island. 707 pp.

Chen, H-Y, and Adams, B.J. undated. Stochastic optimization and reliability analysis for optimal seasonal control of a reservoir. Unpub. manuscript. 24pp.

Chu, C.S., and Bowers, C.E. 1978. An optimization technique for a mathematical urban runoff model. Int. Symp. on urban stormwater management. U. of Kentucky. July 24-27, 1978. pp95-102.

Clarke, R.T. 1971. The Use of the Term 'Linearity' as Applied to Hydrological Models. J. of Hydrology, 13, pp.91-95.

Crawford, N. 1993. The next major breakthrough... Paper presented at the U.S.EPA workshop on Developing dynamic watershed modelling capabilities for the Great Lakes tributaries. Heidelberg College, Tiffin, Oh. August 4-6, 1993. 8pp.

Cunge, J.A. 1974. Evaluation Problem of Storm Water Routing Mathematical Models. Water Research, Vol.8, pp.1083-1087.

Cunningham, B.A., and Huber, W.C. 1987. Economic and predictive reliability implications of stormwater design methodologies. Pub No 98, U of Florida Water Resources Research Center. Gainesville, FL.

Davis, D.R., Duckstein, L., Kisiel, C.C., 1973. Model Choice and Evaluation from the Decision Viewpoint. A paper prepared for the 1973 NATO Conference on "The Role and Effectiveness of Decision Theories in Practice", Luxembourg, August 1973. 18 pp.

DeCoursey, D.G. and W.M. Snyder, 1969: "Computer-Oriented Method of Optimizing Hydrologic Model Parameters." *Journal of Hydrology,* 9:34-56.

Dennison, K.D. 1985. A Database Environment and Sensitivity Framework for Continuous Water Quality Modelling", M.Eng. Mc Master U. Oct 1985, 292 pp.

Dennison, K.D. and James, W. 1985. "A Database Environment and Sensitivity Framework for Continuous Water Quality Modelling", Proc. of the Stormwater and Water Quality Conf., sponsored by USEPA and Ontario Ministry of Environment, Toronto, Ont. published by CHI Hamilton, pp. 101-124, Dec 5-6, 1985.

Deyda, S. 1993. Automatic calibration of runoff models with the help of mathematical optimization methods. 6th Int. Conf. Urban Storm Drainage, Niagara Falls, Ontario. Sept. 1993. pp 838-843.

Dilks, D.W. 1987. Analysis of water quality model uncertainty using a Bayesian Monte Carlo method. Ph.D. Thesis. U. of Michigan. Ann Arbor. 184 pp.

Diskin, M.H. and Simon, E. 1977 A procedure for the selection of objective functions for hydrologic simulation models. J. of Hydrology, 9:34-56.

Donigian, A.S. and Linsley, R.K. 1979. Continuous simulation for water quality planning. Water Resources Bulletin, 15(1): 1-16

Duckstein, L., Monarchi, D., and Kisiel, C.C. 1973. Interactive Multi-Objective Decision Making Under Uncertainty. A paper prepared for the 1973 NATO Conference on "The Role and Effectiveness of Decision Theories in Practice:, Luxembourg, August 1973. 39 pp.

Dunn, A.R., 1986. Feasibility of a Real-Time Auto Sensitivity and Error Analysis Framework for Fuzzy Stormwater Modelling on Personal Computers, M.Eng. Thesis, McMaster U., May 1986, 130pp.

-----, and James, W., 1985. Extended Expert Systems. Proc. of the Stormwater and Water Quality Modelling Conf., sponsored by USEPA and Ontario Ministry of the Environment, Toronto, Canada, published by CHI, Hamilton, pp. 373-380, Dec 5-6, 1985.

-----, & -----, 1986 "Real-Time Error and Sensitivity Analysis to Help Engineers Use Complex Programs Expertly", Int. Symp. on Water and Data Processing, Ecole Nationale Des Ponts Et Chaussees, Paris, France, pp. 196-205, May 28-30, 1986.

Edwards, D.R. 1988. Incorporating Parametric Uncertainty into Flood Estimation Methodologies for Ungaged Watersheds and for Watersheds with Short Records. Ph.D., Oklahoma State Univ.,

El-Hosseiny, T. and James, W.,1991. Urban Drainage Simulation with Real-Time Error Analysis. Annual Conf. of CSCE, Vancouver, B.C., pp 306-314, May 29-31, 1991.

El-Sharkawy, A and Kummler, R.H. 1984. Sensitivity analysis of SWMM predictions of water quality in the Detroit River. Proc. Stormwater and water quality model users group meeting, Apr 12-13, 1984. USEPA 600/9-85-003. pp 67-80.

Eykhoff, P.,1974. *System Identification - Parameter and State Estimation,* John Wiley, New York, 1974.

Field, R. and Lager, J.A. 1975. Urban runoff pollution control - state of the art
J. Environmental Eng'rg, ASCE EE1 pp107-125.

Fleming, G. 1975. Computer simulation techniques in hydrology. Elsevier. ISBN 0-444-00157-3. 333pp.

Frank, P.M. 1978. Introduction to system sensitivity theory. Academic Press. 1978. ISBN 0-12-265650-4

Gallagher, D.L. 1986. The Application of Data Based Transformations to Parameter Estimation in Water Quality Models. Ph.D., Univ. of North Carolina, .

Gardner, R.H., D.D. Huff, R.V. O'Neill, J.B. Mankin, J.H. Carney, and J. Jones,1980. Application of error analysis to a marsh hydrology model. *Water Resour. Res., 16,* 659-664,

Garnatz, D.K. and Hunt, E. 1973. Eyeball Parameter Estimation with a Computer. IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC-3, No.1, Jan. 1973. pp.45-51.

Garrick, M., C. Cunnane, and J.E. Nash, 1978: "A Criterion of Efficiency for Rainfall-Runoff Models." *Journal of Hydrology,* 36:375-381.

Goldberg, M.A. (undated). The Last of the Dinosaurs or the First of the Primates: Yet Another Simulation Study. Unpublished manuscript. 32 pp.

Gore, A., 1992. Earth in the Balance - Ecology and the Human Spirit. Houghton Mifflin Company, New York, New York, 1992, 408 pp.

Goyen, A.G. 1981. Determination of Rainfall/Runoff Model Parameters. M.E. Thesis, New South Wales Inst. of Technology,

Green, I.R.A. and Stephenson, D. 1986 Criteria for comparison of single event models. Hydrologic Sciences J., 31(3):395-411.

Griffin, S., Bauwens, W, and Ahmad, K. 1993. UDMIA: urban drainage modelling intelligent assistant. 6th Int. Conf. Urban Storm Drainage, Niagara Falls, Ontario. Sept. 1993. pp 1314-1319.

Gupta, V.K. and S. Sorooshian, 1983: Uniqueness and observability of conceptual rainfall-runoff model parameters: the percolation process examined. *Water Resources Research,* 19(1):269-276.

Gupta, V.K. and S. Sorooshian, 1985: The Automatic Calibration of Conceptual Catchment Models Using Derivative-Based Optimization Algorithms *Water Resources Research,* 21(4):473-485.

Han, J. 1981. Investigation and development of urban hydrologic models. Ph.D. thesis Purdue U. 308pp.

Han, J. and Rao, A. Ramachandra. 1980. Optimal parameter estimation of objective functions of urban runoff models. Tech. report No 135. Water Resources Research Center, Purdue University. Sept. 1980

Harlin, J. 1991. Development of a process-oriented calibration scheme for the HBV hydrological model. Nordic Hydrology, 22: 15-36.

Helton, J.C., Iman, R.L., Johnson, J.D. and Leigh, C.D. 1986. Uncertainty and sensitivity analysis of a model for muticomponent aerosol dynamics. Nuclear technology 73:320-342.

Hendrickson, J.D. and Sorooshian, S. 1988. Comparison of Newton-type and direct search algorithms for calibration of conceptual rainfall-runoff models. Water Resources Research. 24(5):691-700.

Hornberger, G.M., K.J. Beven, B.J. Cosby, and D.E. Sappington,1985. Shenandoah Watershed study: Calibration of a topography-based, variable contributing area hydrological model to a small forested catchment, *Water Resour. Res. 21*(12):1841-1850, 1985.

Houck, M.H., Bell, J.M. and Wood, D.M. 1990. Automated calibration and use of stream quality simulation model. J. Env. Eng'rg. 116:236-49.

Hromadka, T.V. 1987. Uncertainty in hydrologic models: a review of the literature. Proc. Stormwater and water quality model users group meeting, Mar 23-24 1987. Denver Colo. USEPA pp 217-227.

Hromadka, T.V., McCuen, R.H. and Yen, C.C. 1987. Comparison of overland flow hydrograph models. J. Hydraulic Eng'rg. 113(11):1422-1440.

Huber, W.C. 1993. Methods vs. parameters: the importance of parameter estimation and sensitivity analysis in hydrology. In: Water resources education: a lifetime of learning. AWRA. 1993 pp151-158.

Huber, W.C., 1992. Experience with the U.S. EPA SWMM Model for Analysis and Solution of Urban Drainage Problems. Inundaciones Y Redes de Drenaje Urbano, Dolz J., Gomez, M., and J.P. Martin eds., Colegio de Ingenieros de Caminos, Canales y Puertos, May 1992. pp. 199-220.

Huber, W.C., 1992. Prediction of Urban Nonpoint Source Water Quality: Methods and Models. Paper presented at the International Symposium on Urban Stormwater Management, Sydney 4-7 February 1992. pp.?.

Huber, W.C. and Dickinson, R.E 1988. Storm water management model user's manual, version 4. EPA/600/3-88/001a. (NTIS PB88-236641/AS) U.S.EPA, Athens, Ga.

Huet, 1954. Cited but not referenced by Lazaro, 1990.

Hwang, J.H. 1985. Uncertainty analysis in mathematical water quality modelling. Proc.Stormwater and water quality model users group meeting, Dec 5-6 1985 pp 81-100.

Ibbitt, R.P and Hutchinson, P.D. 1984 Model parameter consistency and fitting criteria. 9th IFAC Triennial World Congress, Budapest, Hungary.

Ibrahim, Y and Liong, S-Y. 1993. A method of estimating optimal catchment model parameters. Water Resources Research 29(9):3049-3058.

Irvine, K.N., Loganathan, B.G., Pratt, E.J., and Sikka, H.C. 1993. Calibration of PCSWMM to estimate metals, PCBs and HCB in CSOs from an industrial sewershed. In: New techniques for modelling the management of stormwater quality impacts. Ed. W. James. Lewis Publishers, ISBN 0-87371-898-4. pp 215-242.

Isabel, D. and J.-P. Villeneuve, 1986: "Importance of the Convergence Criterion in the Automatic Calibration of Hydrological Models." *Water Resources Research,* 22(10):1367-1370.

Isukapalli, Sastry S. 1999 Uncertainty Analysis of Transport-Transformation Models Doctor of Philosophy Rutgers, The State University of New Jersey

Jackson, D.R. and G. Aron, 1971: "Parameter Estimation in Hydrology: The State of the Art." *Water Resources Bulletin,* 7(3):457-472.

Jaffe, P.R., Paniconi,C. and Wood, E.F. 1988. Model calibration based on random environmental fluctuations. J. Environmental Engrg. 114(5): 1136-1145.

James, L.D. and Burges, S.J. 1982. Selection, calibration and testing of hydrologic models. In: Hydrologic modelling of small watersheds, Ed by Haan, C.T., Johnson, H.P. and Brakensiek, D.L. ASAE monograph no. 5. pp 436-472.

James, W. 1995. Channel and Habitat Change Downstream of Urbanization, Chap. 7 in: Herricks, E. (Ed.), Stormwater Runoff and Receiving Systems - Impact, Monitoring and Assessments. Lewis Pubs. ISBN 1566701597 (Invited paper - Engineering Foundation Conf. on Urban Runoff and Receiving Streams, Crested Butte, Colorado, Aug 4-9, 1991.) pp 103-112.

-----, 1993a Need for 3-generation stormwater quality impacts models. USEPA Invited Conference on Model-building for Great Lakes Water Quality problems of the 2000's. Heidelberg College, Tiffin, OH. Aug 4-6. *unpublished; copy available from author*

-----, 1993b Stormwater Management Modelling Conceptual Workbook. Pub. by CHI Guelph. 200 pp.

-----, and James, W.R.C. 1993. PCSWMM4 - a public-domain, graphics-oriented shell for easing the design-use of SWMM4. Runoff Quantity and Quality Model Group Conference, Sponsored by ASCE . Reno NV, Nov 8-9 1993

James, W. and Dunn, A., **1986** "An Integrated Conceptual Expert System for Flood and Water Pollution Management", Water Forum '86, A.S.C.E., Long Beach, California, pp. 158-165, Aug 4-6, 1986.

----, & ----, 1985. "Integrated Knowledge-based Software for Flood and Water Pollution Management", ASCE Water Resources Planning and Management Specialty Conf. on Computer Applications in Water Resources, Buffalo, New York, pp. 611-619, Jun 10-12, 1985.

---, and Robinson, M.A. 1981a. Standard terms of reference to ensure satisfactory computer-based urban drainage design studies. Can J Civ. Engrg 8:294-303.

-----, ----- 1981b. Standards for Computer-based Design Studies, J. of the Hydraulics Division, ASCE, 107(HY7) Proc. Paper 16385: 919-930, July.

-----, ----- 1985. Required Time and Space Resolution for Continuous Dynamic Storm and Runoff Modelling and for Associated Field Data. ASCE Hydraulics Specialty Conf. on Hydraulics and Hydrology in the Small Computer Age, Orlando, pp. 1160-1165, Aug 1985.

-----, ----- 1986a. Continuous Deterministic Urban Runoff Modelling. Int. Symp. in Urban Drainage Modelling, Dubrovnik, Yugoslavia, published by Pergamon Press, London, pp. 347-378, Apr 1986.

-----, ----- 1986b. Continuous Modelling of Downstream River Erosion caused by 40 Years of Urbanization. Water Forum '86, A.S.C.E., Long Beach, California, pp. 1710-1717, Aug 4-6, 1986.

------, & -----, 1987. "Hydrologic Modelling of Continuous Urbanisation and Consequent Effects Downstream", 8th Can. Hydrotech. Conf., Montreal, Quebec, pp. 524-531, May 17-22, 1987.

----, and Shivalingaiah, Boregowda, 1986. Continuous Mass Balance of Pollutant Build-up Processes, Proc. NATO Advanced Research Workshop on Urban Runoff Pollution, Montpellier, France, Aug 25-30, 1985, published by Springer-Verlag in **Urban Runoff Pollution** (Series G. Vol. 10), pp. 243-271, 1986.

Jewell, T.K. 1974. Application and testing of the U.S.EPA SWMM to Greenfield, Mass. MSc project report. Univ of Massachusetts.

Jewell, T.K. 1980. Urban stormwater pollutant loadings. PhD Thesis. U. Mass. Feb 1980. 280pp.

Jewell, T.K. and Adrian, D.D. **1982.** Statistical Analysis to Derive Improved Stormwater Quality Models. J. WPCF, Vol.54, No.5, pp.489-497.

Jewell, T.K., Nunno, T.J., and Adrian, D.D. 1978. Methodology for calibrating stormwater models. Proc. Stormwater and water quality model users group meeting, May 4-5 1978. USEPA 600/9-78-019 pp 125-169.

Johanson, R.C., Imhoff, J.C., Kittle, J.L., and Donigian, A.S. 1984. Hydrologic simulation program - FORTRAN (HSPF): User manual. U.S.EPA. Report EPA/600/9-80-015.

Johnston, P.R. and D.H. Pilgrim, 1976: "Parameter Optimization for Watershed Models." *Water Resources Research,* 12(3):477-486.

Jowitt, P.W., **1984**. Risk analysis, fuzzy logic, and river basin management, *Water Sci. Technol., 16*(5-7), 579-585, 1984.

Jowitt, P.W. and J.P. Lumbers, 1982. Water quality objectives, discharge standards, and fuzzy logic, Optimal Allocation of Water Resources, *IAHS-AISH Publ., 135,* 241-250, 1982.

Karlinger, M.R. and Moss, M.E., April **1973**. Regression-Analysis Simulation for Evaluation of Hydrologic-Parameter Accuracy. Unpublished manuscript, 13 pp.

Kelton, W.D. IV. 1980. The Startup Problem in Discrete-Event Simulation. Ph.D., Univ. of Wisconsin-Madison

Kennedy, P. 1993 Preparing for the twenty-first century. Random House. ISBN 0-394-58443-0. 445pp.

Kindler, J. 1992. Rationalizing water requirements with aid of fuzzy allocation model. J. Water Resources Planning and Management. 118(3):308-323.

King, R.P. 1967. Estimation of parameters in systems defined by differential equations. S. Afr. J. of Sci. 63(3):91-96.

Kisiel, Chester C. 1971. Efficiency of Parameter and State Estimation Methods in Relation to Models of Lumped and Distributed Hydrologic Systems. Proc. of the First Bilateral U.S.-Japan Seminar in Hydrology Jan. 1971. Chapt.12, pp.12.1-12.27.

Kitandis, P.K. and Bras, R.L. 1980. Real-time forecasting with a conceptual hydrologic model: 1. Analysis of uncertainty. Water Resources Research 16(6):1025-1033.

Klemes, V.1987. Empirical and causal models in hydrologic reliability analysis. Procs NATO Institute on Engineering reliability and risk in water resources, Tucson, Ariz. May19-Jun2. 1985. ISBN 90-247-3492-4. pp391-403.

Kohberger, R., Scavia, D and Wilkinson, J.W. 1978. A method for parameter sensitivity analysis in differential equation models. Water Resources Research 14(1):25-29.

Kouwen, N. and Harrington, R.A. 1986. Automatic Estimation of Model Parameters. Proc. of the First Canadian Conference on Computer Applications in Civil Engineering: Micro-Computers. pp.443-457.

Kuch, A. J. 1994. A sensitivity analysis shell for complex continuous water quality management modelling. M.Sc. Thesis. Univ of Guelph. ca.120 pp. January 1994.

-----, and James, W. 1993. A sensitivity analysis tool for three generation continuous urban storm water quality management modelling. Runoff Quantity and Quality Model Group Conference, Sponsored by ASCE . Reno NV, Nov 8-9 1993. *in press.*

Larson, L.W. and E.L. Peck, **1974:** "Accuracy of Precipitation Measurements For Hydrologic Modeling." *Water Resources Research,* 10(4):857-863.

LaVeck, G. 1993. Development of a WINDOWS-based interface to SWMM. Preprint for Conf. on runoff quantity and quality modelling. ASCE. Reno, Nev. Nov. 1993. 7pp.

Lazaro, T.R. 1990 Urban hydrology - revised Technomic. ISBN 87762-547-6. ca 250pp.

Leavesley, G.H., and Stannard, L.G. 1989. A modular watershed modelling and data management system. Flood forecasting Symposium and Workshop. Mar-Apr 1989 NWS. pp71-95

Leavesley, G.H., Restrepo, P.J., Stannard, L.G., Frankoski, L.A. and Sautins, A.M 1993. The modular modelling system (MMS) - a modelling framework for multidisciplinary research and operational applications. Preprint for Conf. on GIS and Environmental Modelling. Breckenridge, Colo. Sept 1993 . 9pp.

Leavesley, G.H., Lichty, R.W., Troutman, B.M. and Saindon, L.G. 1983. Precipitation-runoff modelling system user's manual. Water Resources Investigations 83-4238. USGS. Denver Colo. 80225. ca 207 pp.

Lei, J. and Schilling, W. 1993. Propagation of model uncertainty. 6th Int. Conf. Urban Stormwater Drainage. Niagara Falls, Ont. Sept 1993. pp465-470.

Leopold, L.B. 1968. Hydrology for urban land planning - a guidebook on the hydrologic effects of land-use. USGS. Circ. no. 554. 18pp.

Lettenmaier, D.P., and S.J. Burges, 1976. Use of state estimation techniques in water resource modelling. *Water Resour. Bull., 12*(1):83-99,

Linsley, R. 1973. A manual on collection of hydrologic data for urban drainage design. Hydrocomp Inc. Palo Alto, CA.

Liong, S.Y. , Chan, W.T and Lum, L.H. 1990. Knowledge-based system for SWMM runoff component calibration. J. Water Resources Planning and Management. 117(5): 507-524.

Litchfield, J.B. and Warwick, J.J. 1993. The effect of spatial data aggregation on urban stormwater quality modeling. Preprint for 1993 Runoff quantity and quality modelling conf. ASCE. Reno, Nev. Nov.1993.5pp.

Lumb, A.M. and Kittle, J.L. 1983. ANNIE - an interactive processor for hydrologic modeling. USGS Water resources Investigation report 83- ? 133pp.

Lundgren, P.S. and Barber, M.E. 1993. Site-specific application of SWMM using expert system. Preprint 1993 Runoff quantity and quality modeling conf. ASCE. Reno, Nev. Nov 1993. 8pp.

Manley, R.E., 1978a: "Calibration of Hydrological Model Using Optimization Techniques." *Journal of the Hydraulics Division,* Proceedings of the American Society of Civil Engineers (ASCE), 104(HY2):189-202.

Marsalek, J. 1982 Discussion of: SWMM model and level of discretization. J. Hydraulics Div ASCE. 108:1773-1776. *See Zaghloul, below.*

McCuen, R.H. 1973a. The role of sensitivity analysis in hydrologic modelling. J of Hydrology, 18:37-53.

-----, 1973b. Component sensitivity: a tool for the analysis of complex water resources systems. Water Resources Research. 9(1):243-247.

McLaughlin, D.B., **1978a.** Parameter estimation problems in water resources modelling, in *Modelling, Identification and Control in Environmental Systems,* edited by G.C. Vansteenhiste, pp.137-151, North-Holland, Amsterdam,

McNeil, D. and Freiberger, P. 1993. Fuzzy Logic. Simon and Schuster, New York, NY, 1993 319 pp.

Medina, M. 1992. Paper on Surface water quality modelling. In: Pollution Control Planning pub by CHI Guelph and MOEE Toronto.

Mein, R.G. and Brown, B.M. 1978. Sensitivity of optimized parameters in watershed models. Water Resources Research, 14(2):299-303.

Melching, C.S. and Anmangandla, S. 1992. Improved first-order uncertainty method for water-quality modelling. J. Hydraulics Div ASCE. 118(5):791-805.

Metropolis, N. and Ulam, S. 1949. The Monte Carlo method. J. Amer. Statistical Assoc. 44(247):335-341.

Mihram, G.Arthur. (undated) Some Practical Aspects of the Verification and Validation of Simulation Models. Operational Research Quarterly, Vol.23(1) pp.17-29.

Mihram, G.A. **1972.** The Modelling Process. IEEE Trans. on Systems, Man, and Cybernetics, Vol.SMC-2, No.5, pp. 621-629. Nov. 1972.

Mineart, P. nd Lobascio, M. 1993. Sensitivity analysis of non-point loads assessment using Monte Carlo simulation. Preprint. 1993 Runoff quantity and quality modeling conf. ASCE. Reno, Nev. Nov 1993. 6pp.

Moore, S.F. 1973. Estimation theory applications to design of water quality monitoring systems. J. of the Hydraulics Div. ASCE. 99(5):815-831.

Morita, M., Jinda, S. and Izumi, K. 1993. Application of fuzzy theory to urban runoff prediction. 6th Int. Conf. on urban storm drainage. ASCE. Vol 1: pp 820-825.

Nash, J.E and Sutcliffe, J.V. 1970. River flow forecasting through conceptual models Part 1 - a discussion of principles. J. of Hydrology, 10:282-290.

Navarro, J.E., McCauley, D.J. and Blystra, A.R. 1994. Instream flow incremental methodology (IFIM) for modelling fish habitat. In: Current practices in modelling the management of stormwater impacts. Lewis Pub. ISBN 1-56670-052-3 pp1-22.

Nihoul, J.C. 1976. Applied mathematical modelling in marine science. Applied Math. Modelling Vol 1:3-8.

Nix, S.J., Moffa, P.E. and Davis, D.P. 1991. The practice of combined sewer system modelling. Water Resources Bulletin. 27(2):189-200.

Nnaji, S., Davis, D.R. and Fogel, M.M., April **1974.** Uncertainty in Predictions from a Deterministic Watershed Model. Paper presented at the 55th Annual Meeting American Geophysical Union, Washington DC 20036, 21 pp.

Norris, J.M. 1986. Application of the precipitation runoff model system to small basins in the Parachute Creek Basin, Colorado. USGS. Water Resources Investigations report 86-4115. 43pp

Orlob, G.T. 1992. Water-quality modelling for decision making. J. Water Resources Planning and Management. 118(3): 295-307.

Orr, D.W. 1992. Ecological literacy. State Univ of NY Press. ISBN 0-7914-0874-4 224pp.

Patry, G.G. 1983 Real-Time Forecasting of Water Quantity and Quality in Urban Catchments. Ph.D., Univ. of California Davis,.

Patry, G.G. and M.A. Marino, 1983: "Nonlinear Runoff Modeling: Parameter Identification." *Journal of Hydraulic Engineering,* Proceedings of the American Society of Civil Engineers (ASCE), 109(6):865-880.

Philip, J.R. **1975.** Some Remarks on Science and Catchment Prediction. Prediction in Catchment Hydrology, Australian Academy of Science, 1975. pp.23-30.

Pickup, G., 1977: "Testing the Efficiency of Algorithms and Strategies for Automatic Calibration of Rainfall-Runoff Models." *Hydrological Sciences Bulletin,* 22(2):257-274.

Pitman, W.V. **1978.** Flow Generation by Catchment Models of Differing Complexity - A Comparison of Performance. J. of Hydrology, 38, pp.59-70.

Plinston, D.T., **1972:** "Parameter Sensitivity and Interdependence in Hydrological Models." In: Mathematical Models in Ecology, edited by J.N.R. Jeffers, Blackwell, Oxford, pp.237-247.

Proctor Redfern and MacLaren, J.F. 1976. Storm water management model study, Vol 1 Final report. Ontario Ministry of the Environment, Toronto.

Posada, P.J Restrepo, and Bras, R.L. 1982. Automatic parameter estimation of a large conceptual rainfall-runoff model: a maximum liklihood approach. Report 267. Ralph M Parsons Labortory, MIT. (R82-5; OSP 89310) ca 260pp.

Postma, L. 1981. Modelling as a Tool. Waterloopkundig Laboratorium, Delft Hydraulics Laboratory, Pub. No. 252, 12 pp. April 1981.

Qaisi, K.M. 1985. Uncertainty analysis of stream dissolved oxygen models. Ph.D. thesis. U. of Michigan Ann Arbor. 234 pp.

Ramachandra Rao, A. and J. Han, 1981: "Optimal Parameter Estimation in Urban Runoff Models." Paper presented at the 2nd International Conference on Urban Storm Drainage, Urbana, IL.

Rauf, A. and Chaffey, D. 1993. A new interactive graphics interface for urban drainage applications. 6th Int. Conf. on Urban Stormwater Drainage. Niagara Falls. Nov. 1993. pp 1345-1350.

Reboh, R.J., Reiter, R.J. and Gashnig, J. 1982. Development of a knowledge-based interface to a hydrologic simulation program. SRI International, Menlo Park, CA.

Reckow, 1979. source misplaced.

Reckhow, K.H. 1981. Confirmation of Water Quality Simulation Models. Presented at the Fall 1981 Meeting of the American Geophpysical Union, San Francisco, California. 36 pp.

Reed, L.A. 1986. Verification of the PRMS sediment-discharge model. Proc 4th Federal Interagency sedimentation conf. Mar 24-27. Las Vegas, Nev. pp6-44 to 6-54.

Rivera Santos, J. 1990. Parameter estimation in conceptual precipitation-runoff models with emphasis on error analysis. Proc of Int. Symp. on Tropical hydrology and 4th Caribbean Islands water resources congress, San Juan, Puerto Rico, Jul 22-27, 1990. Amer. Water Resources Assoc. pp91-100.

Rivera Santos, J. 1988. A methodology for parameter estimation in conceptual precipitation-runoff models with emphasis on error analysis. Ph.D. thesis. U. of Colorado Boulder. 279 pp.

Robinson, M. and James, W., 1984. “Continuous Variable Resolution Stormwater Modelling on a Microcomputer”, Proc. of the Stormwater and Water Quality Modelling Conf., Sponsored by USEPA and Ontario Ministry of Environment, Burlington, Ontario, published by CHI Hamilton, pp. 59-74, Sept 6-7, 1984.

Rogers, C.C.M., Beven, K.J., Morris, E.M. and Anderson, M.G. 1985. Sensitivity analysis, calibration and predictive uncertainty of the Institute of Hydrology Distributed Model. J. of Hydrology, 18:179-191.1985.

Sangal, S.K. and Bonema, S.R. 1994. A methodology for calibrating SWMM models. In: Current Practices in modelling the management of stormwater impacts. Lewis Publishers. ISBN 1-56670-052-3. 1994 pp375-388.

Sautier, J.L and Delleur, J.W. 1978. Calibration and sensitivity analysis of the continuous simulation model STORM. Int. Symp. on Urban Stormwater Management. July 24-27, 1978. pp73-79.

Savic, D.A. 1990. An intelligent decision support system and use of fuzzy sets for reservoir analysis. Ph.D. thesis. U. of Manitoba. 253 pp.

Scheckenberger, R., 1983. Dynamically Spatially Variable Rain Models for Stormwater Management. Master in Engineering Thesis, McMaster University, May 1983, 146 pp.

Schmidt, J.W., Taylor, R.E., and Bennett, G.K. 1973. A Heuristic Elimination Method for Optimum Selection of Alternatives Through Simulation. Simulation, pp.129-134.

Schmidt, M.F., TenBroek, M., Aldrich, J., and Plante, S.V. 1993. CDM: Stormwater system planner. Preprint for 1993 Runoff Quantity and Quality conf. ASCE. Reno, Nev. Nov. 1993. 10pp.

Scavia, D. 1980. Uncertainty analysis of a lake eutrophication model. Ph.D. thesis. U. of Michigan Ann Arbor. 174 pp.

Seo, Dong-il. 1991. Optimal complexity analysis of total phosphorus models. Ph.D. thesis. U. of Michigan Ann Arbor. 173 pp.

Shastry, J.S., Fan, L.T. and Erickson, L.E. 1973. Nonlinear Parameter Estimation in Water Quality Modeling. J. of The Environmental Engineering Div., Vol.99,No.EE3, June 1973, pp.315-331.

Sircar, J.K. 1986. Computer Aided Watershed Segmentation for Spatially Distributed Hydrologic Modelling. Ph.D., Univ. of Maryland,

Smith, J.A. and Karr, A.F. 1985. Parameter estimation for model of space-time rainfall. Water Resources Research. 21(8):1251-1257.

Snyder, W.M. **1955**. Hydrograph Analysis by the Method of Least Squares. Proc.ASCE Hydraulics Div., Vol.81 Paper No.793, Sept. 1955. 25 pp.

Sorooshian, S. 1987. The impact of catchment modelling on hydrologic reliability. Procs NATO Institute on Engineering reliability and risk in water resources, Tucson, Ariz. May19-Jun2. 1985. ISBN 90-247-3492-4. pp 365-389.

Sorooshian, S. (undated). Automatic Calibration of Conceptual Rainfall-Runoff Models: The Question of Parameter Observability and Uniqueness. Submitted to Water Resources Research. 27 pp.

Sorooshian, S., **1978**: Considerations of Stochastic Properties in Parameter Estimation of Hydrologic Rainfall-Runoff Models. Ph.D. Dissertation, University of California, Los Angeles, CA, pp.248.

Sorooshian, S., **1980**: "Comparison of Two Direct Search Algorithms Used in Calibration of Rainfall-Runoff Models." In: Water and Related Land Resources Systems, Symposium of the International Federation of Automatic Control, Cleveland, Ohio, May 28-31, 1980, pp.477-485.

Sorooshian, S. and F. Arfi, 1982: "Response Surface Parameter Sensitivity Analysis Methods For Postcalibration Studies." *Water Resources Research,* 18(5):1531-1538.

Sorooshian, S., and Arfi, F. 1982. Parameter Sensitivity Analysis for Conceptual Hydrologic Rainfall-Runoff Models. Water Resources Research 1982. 35 pp.

Sorooshian, S., and V.K. Gupta, **1985.** The analysis of structural identifiability: Theory and application to conceptual rainfall-runoff models, *Water Resour. Res., 21*(4): 487-496, 1985.

Sorooshian, S. and V.K. Gupta, 1983: "Automatic Calibration of Conceptual Rainfall-Runoff Models: The Question of Parameter Observability and Uniqueness." *Water Resources Research,* 19(1):260-268.

Sorooshian, S., V.K. Gupta, and J.L. Fulton, **1983**. Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall-runoff models: Influence of calibration data variability and length on model credibility, *Water Resour. Res., 19*(1):251-259,

Sorooshian, S., Gupta, V.K., and Fulton, J.L. **undated**. Parameter Estimation of Conceptual Rainfall-Runoff Models Assuming Autocorrelated Streamflow Data Errors - A Case Study. Unpublished manuscript, 14 pp. (undated).

Sorooshian, S., Gupta, V.K., Fulton, J.L. and Arfi, F. 1981. Advanced calibration methodology for hydrologic rainfall-runoff simulation models. Tech. rep. LSSPA-81-6. Case Inst. of Tech., Case Western Reserve University. Dec 1981. ca 160pp.

Southerland, E. 1982. A continuous simulation modelling approach to nonpoint pollution management. Ph.D. thesis. Virginia Polytechnic Institute and State U. 165 pp.

Srinanthakumar, S.K. and Codner, G.P. 1993. Selection of calibration events for urban runoff quality modelling. 6th Int. Conf. Urban Stormwater Drainage. Niagara Falls, Ont. Sept. 1993. pp 909-914.

Srikanthan, R. and McMahon, T.A., 1985. Stochastic Generation of Rainfall and Evaporation Data. Australian Water Resources Council Technical Paper No. 84, 1985. 301 pp.

Stanley and Assocs, Undated. (1993?) RTSWMM 4-page pamphlet, pub by Hydrotechnical services group, Edmonton, Alberta.

Steinhorst, R.K., **1979**. Parameter identifiability, validation, and sensitivity analysis of large system models, in *Systems Analysis of Ecosystems,* edited by G.S. Innis and R.V. O'Neill, pp.33-58, International Co-operative Publishing House, Fairland, Md.,

Stephenson, D., 1989. Selection of stormwater model parameters. J. of Environmental Engrg. 115(1): 210-220.

Stevens, E.W and Haan, C.T. 1993. Effect of parameter variation and model nonlinearity on first order analysis. 6th Int. Conf. Urban Stormwater drainage. Niagara Falls, Sept. 1993. pp 441-446.

Szollosi-Nagy, A. 1981. State Space Models of the Nash-Cascade - Kinematic and Diffusion Waves. Research Report TULEA 1981:14, Serie A No. 68, 24 pp. April 1981.

Thomann, R.V., **1982.** Verification of water quality models, *Proc. Am. Soc. Civ. Eng. J. Environ. Eng. Div., 108*(EE5), 923-940,

Thompson, D.B. 1989. Determining parameters for a continuous simulation model by estimation, synthetic calibration and analytic calibration. Ph.D. thesis. U. of Missouri-Rolla. 261 pp.

Thompson, D.B. and Westphal, J.A. 1989. Tilt at the windmill of runoff simulation. Proc of the Int Conf on catchment runoff, U of Virginia, Charlottesville. 1989. pp176-185.

Thompson, G.R., Bogiatzis, T., Phillips, B.C. and Dickinson, R.E. 1993. XP: a graphical decision support system for water resources models. 6th Int. Conf. Urban Stormwater Drainage. Niagara Falls, Sept. 1993. pp 1351-1356.

Tomovic, R. 1963. Sensitivity analysis of dynamic systems. McGraw-Hill Book co.1963. ca 160 pp.

Troutman, B.M., **1982:** "An Analysis of Input Errors in Precipitation-Runoff Modeling Using Regression with Errors in the Independent Variables." *Water Resources Research,* 18(4):947-964.

Troutman, B.M., **1983:** "Runoff Prediction Errors and Bias in Parameter Estimation Induced by Spatial Variability of Precipitation." *Water Resources Research,* 19(3):791-810.

Troutman, B.M., **1985a**: "Errors and Parameter Estimation in Precipitation-Runoff Modeling. 1. Theory." *Water Resources Research,* 21(8):1195-1213.

Troutman, B.M. 1985b: "Errors and Parameter Estimation in Precipitation-Runoff Modeling. 2. Case Study." *Water Resources Research,* 21(8):1214-1222.

Troutman, B.M. 1981. The Effect of Input Errors in Using Precipitation-Runoff Models for Runoff Prediction. Reprinted from the book: Statistical Analysis of Rainfall and Runoff. Proc. of the International Symposium on Rainfall-Runoff Modeling held May 18-21, 1981 at Mississippi State University, Mississippi State, Mississippi, USA. pp.305-314.

Troutman, B.M. 1986. Reducing Bias in Parameter Estimation Caused by Model Input Errors. Presented at AGU-AGM, Baltimore, May 18, 1986. 16 pp.

Uber, J.G., Brill, E.B., and Pfeffer, J.T. 1992. Use of mathematical programming methods for complex systems. J. Water Resources Planning and Management. 118(3): 281-294.

Udhiri, S. 1984. The need to validate hydrologic models. Proc. Stormwater and water quality model users group meeting, Apr 12-13, 1984. USEPA 600/9-85-003. pp 175-188.

Unal, A., **1985.** "Centralised Time Series Management for Continuous Hydrology on Personal Microcomputer Networks", Ph.D. Thesis, McMaster U., Dec 1985, 185 pp.

Unal, A. and James, W., 1984. "Centralized Time Series Management Programs for Continuous SWMM3 on Personal Microcomputers", Proc. of the Stormwater and Water Quality Modelling Conf., Sponsored by USEPA and Ontario Ministry of Environment, Burlington, Ontario, published by CHI Hamilton, pp. 91-100, Sept 6-7, 1984.

Vajda, S., P. Valko, and A. Yermakova, **1985**. A multistage parameter estimation procedure for continuous systems, in *Identification and System Parameter Estimation,* edited by H.A. Barker and P.C. Young, pp.1579-1584, Pergamon, New York,

Van-Thanh-Van Nguyen and Wang, H. 1993. An assessment of uncertainty in the modelling of rainfall runoff process. 6th Int. Conf. Urban Stormwater Drainage. Niagrara Falls, Sept 1993. pp 453-458.

Voit, G., 1993. Fuzzy Logic: Concepts to Constructs. AI Expert. November 1993 pp. 26-33.

Von Euw, E. 1990. A comprehensive comparison of three non-point-source pollution models on an agricultural watershed in Southern Ontario. M.Sc. Thesis. U. of Guelph. 134 pp.

Walker, W.W. 1982a. A sensitivity and error analysis framework for lake eutrophication modelling. Water Resources Bulletin. 18(1):53-60.

Walker, W.W. 1982b. MODELER a computer modeling and data management system. Ver 4.0 Mar. 1982. W.W.Walker, Concord, MA. 21pp.

Wang, Q.J. 1991. The genetic algorithm and its application to calibrating conceptual rainfall-runoff models. Water Resources Research. 27(9): 2467-2471.

Warwick, J.J., 1990. Estimating uncertainty of stormwater runoff computations. J. Water Resources Planning and Management. 116(2):187-204.

-----, and Cale, W.G., 1987. Determining the likelihood of obtaining a reliable model. J. Environmental Eng'rg. 113(5):1102-1119.

-----, and Cale, W.G., 1986. Effects of parameter uncertainty in stream modelling. J.Environmental Engrg. 112(3):479-489.

-----, and Tadepalli, P., 1991. Efficacy of SWMM application. J. Water Resources Planning and Management. 117(3): 352-366.

Warwick and Wilson 1990 source misplaced.

Wigan, M.R. 1972. The Fitting, Calibration, and Validation of Simulation Models. Simulation, Vol.18, No.5, pp.188-192. May 1972.

Wildermuth, M.J. and Yeh, W.W-G. 1979. Calibration of rainfall-runoff models: a case study demonstrating a unified approach to calibration . Int Symp. on Urban Storm Runoff. U. of Kentucky. July 23-26, 1979. pp 175-181.

Willgoose, G.R. and R.L. Bras, 1987: "Automatic Calibration of Conceptual Rainfall-Runoff Models." In: Engineering Hydrology, Proceedings of the Symposium Sponsored by the Hydraulics Division of the American Society of Civil Engineers (ASCE), August 3-7, 1987, pp.732-737.

Williams, K.K. 1979. Oklahoma City Urban Storm Runoff Quantity: Comparison and Calibration of Predictive Methods. Ph.D., Univ. of Oklahoma,

Williams, B.J. and W.W-G. Yeh, **1983:** "Parameter Estimation in Rainfall-Runoff Models." *Journal of Hydrology,* 63:373-393.

Wood, E.F., 1976: "An Analysis of the Effects of Parameter Uncertainty in Deterministic Hydrologic Models." *Water Resources Research,* 12(5):925-932.

Wu 1993. WINDOWS-SWMM Conf on Water quantity and quality models, ASCE. Reno, Nev.

Xie J.D.M 1993 Modelling solar thermal enrichment of streams due to urban storm wter, MSc Thesis, U of Guelph. 130 pp.

Xie, J.D.M. and James, W. 1994. Modelling solar thermal enrichment of urban stormwater. In: Current practices in modelling the management of stormwater impacts. Lewis Publishers, ISBN 1-56670-052-3 pp 205-220.

Yan, J. 1990. Parameter estimation for multipurpose hydrologic models. Ph.D. thesis. Oklahoma State U. 184 pp.

Yeh, W.W-G., 1982: "Parameter Estimation in Rainfall-Runoff Modeling." In: Statistical Analysis of Rainfall and Runoff, Edited by V.P. Singh, Water Resources Publishers, Littleton, CO., pp.481-490.

Yen, B.C. 1993. Reliability and uncertainty analyses in hydraulic design. A sub-committee report. ASCE.

Yih, S.-M., and B. Davidson, **1975**. Identification in nonlinear, distributed parameter water quality models, *Water Resour. Res. 11*(5):693-704,

Young, P.C. 1981., Parameter estimation for continuous-time models - A survey, *Automatica, 17*(1):23-39, 1981.

Young, P. 1983. The validity and credibility of models for badly defined systems. In: Uncertainty and forecasting of water quality, M.B. Beck and G. van Straten, eds. Springer-Verlag 1983. pp 69-98.

Zaghloul, N.A. (undated). Sensitivity Analysis of the SWMM Runoff-Transport Parameters and the Effects of Catchment Discretization. 40 pp. Unpublished manuscript.

-----, 1981. Same title. Int Symp. on Urban Hydrology. U. of Kentucky. July 27-30 1981. pp 25-34.

Zaghloul, N.A. 1981. SWMM model and level of discretization. J. of the Hydraulics Div. ASCE. 107(11):1535-1545. *see also Marsalek above.*

Zaghloul, N.A and Al-Shurbaji, A.M. 1990. A storm water management model for urban areas in Kuwait. Water Resources Bulletin, 26(4):563-575.

Glossary

best management practice (BMP) Structural devices that temporarily store or treat urban stormwater runoff to reduce flooding, remove pollutants, and provide other amenities.
calibration *See parameter estimation.*
catchment That area determined by topographic features within which falling rain will contribute to runoff at a particular point under consideration.
channel A natural stream that conveys water; a ditch or drain excavated for the flow of water.
channel erosion The widening, deepening, and headward cutting of small channels and waterways, due to erosion caused by moderate to large floods.
cold water fishery A fresh water, mixed fish population, including some salmonids.
combined sewer A sewer intended to carry surface runoff, sewage and industrial wastes allowed by sewer by-laws.
combined sewer overflow Flow from a combined sewer, in excess of the sewer capacity, that is discharged into a receiving water.
computable A simulation that can be performed in a working day (eight hours) and the output of the simulation to require normal hard drive working space of a typical engineering office work station (in the order of 200 Mb).
continuous modelling A simulation that models both the dry and wet processes of hydrology with a continuous record of atmospheric data. In contrast event modelling is a simulation of short defined storm events with subjective startup conditions.
decision support system Code that manages the simulation system, as poosed to the internal process code (the engine).
detention The slowing, dampening, or attenuating of flows either entering the sewer system or within the sewer system, by temporarily holding the water on a surface area, in a storage basin, or within the sewer itself.
detention time The amount of time a parcel of water actually is present in a BMP. Theoretical detention time for a runoff event is the average time parcels of water reside in the basin over the period of release from the BMP.
dissaggregation The degree to which the components of a physical system are modelled by increasing the number of defined processes.
discretization The number of components selected to represent the physical system that has been dissaggregated into processes on those components and the degree to which the physical parameters are lumped as spatial and temporal averages.
dominant processes The hydrological processes coded in the program and coupled with the input file (forming a model of the physical system) that are active and represent large percentages of contribution to the selected objective function. For

example infiltration is a process coded in SWMM and during *low* rainfall intensities this process would be dominant in extracting water from the surface and reducing runoff volumes and peak flows.

drainage 1. To provide channels, such as open ditches or dosed drains, so that excess water can be removed by surface flow or internal flow. 2. To lose water (from the soil) by percolation.

dry weather flow Combination of domestic, industrial and commercial wastes found in sanitary sewers during dry weather not affected by recent or current rain.

engine The internal hydrologic process code in a model.

erodibility (of soil) The susceptibility of soil material to detachment and transportation by wind or water.

erosion 1. The wearing away of the land surface by running water, wind, ice or other geological agents, including such processes as gravitational creep. 2. Detachment and movement of soil or rock fragments by water, wind, ice or gravity.

error The difference between a computed and an observed value.

See also: observation error, sampling error, numerical error, structural error, propagated error, framework error and parameter error.

error analysis The computation of the likely error that a computed response may have.

event mean concentration (EMC) The average concentration of an urban pollutant measured during a storm runoff event. The EMC is calculated by flow-weighting each pollutant sample measured during a storm event.

event modelling - *see continuous modelling*

expert system Code that follows a certain logic involving rules, inference and on-line infromation.

facilities management system Code for keeping track of inventory and costs of urban infrastructure systems.

first flush The condition, often occurring in storm sewer discharges and combined sewer overflows, in which an unusually high pollution load is carried in the first portion of the discharge or overflow.

flood frequency A measure of how often a flood of given magnitude should, on an average, be equalled or exceeded.

framework error The structural error due to disaggregation and poor component process models.

fuzzy process A process that has different levels of dominance which is dependent on the state or input variable. For example the modelling of pollutant buildup and washoff is a fuzzy process when the rainfall intensity oscillates from zero to very low levels; is pollutant buildup or pollutant washoff occuring?

heuristic Problem-solving techniques that use self-educating ideas.

hydrograph A graph showing variation in stage (depth) or discharge of a stream of water over a period of time.

impervious area Impermeable surfaces, such as pavement or rooftops, which prevent the infiltration of water into the soil.

infiltration The seepage in dry or wet weather or both of groundwater or vadose water into any sewer (storm, sanitary, combined). Generally, infiltration enters

through cracked pipes, poor pipe joints or cracked or poorly jointed manholes; also the loss of surface runoff into pervious ground.
infiltration (of soils) Movement of water from the ground surface into a soil.
input function, or input variable The driving input hydro-meteorological time series for a surface WQM. Typical examples are rainfall, evapo-transpiration, wind speed, wind direction, snowfall, radiation, humidity, temperature, and some pollutant generating mechanisms such as traffic.
input variable space/state variable space The combinations of input time series that alter the dominance of and trigger the processes coded in the program.
loadograph A plot of pollutant load flux rate against time.
model A program that has been attached to a specific, hydro-topographic input-data-file and calibrated.
model complexity A measure of the number of uncertain parameters in a model.
model reliability An overall measure of the model's performance, especially its accuracy.
model uncertainty The degree to which the output of a simulation represents the observed outcome of the physical system. The model uncertainty comprises the uncertainty of many sources including the parameter estimation uncertainty and the degree to which the code of the program models the physical system.
Monte Carlo method Uses probability distributions of input parameters and random number generators to develop large numbers of computed responses, and draw statistical inferences from them.
non-point source An area from which pollutants are exported in a manner not compatible with practical means of pollutant removal (e.g. crop lands).
numerical error Identified with the numerical mathematics used in the code.
objective function A statistic or a representative number derived from the response function.
objective, water quality A designated concentration of a constituent, based on scientific judgements, that, when not exceeded will protect an organism, a community of organisms, or a prescribed water use with an adequate degree of safety.
observation error Related to field instrumentation, comprising two components, one random and the other systematic;
optimal complexity The level of discretization and dissagregation that yields the minimum modelling cost for a given level of model accuracy.
parameters. Coefficients that are also input, but less likely to form very long times series at a fine time resolution, and generally control independent component processes.
parameter error *same as propagated error.*
parameter estimation/calibration A procedure to discover the global optimum of an array of modelling parameters that can only be discovered with estimation and are not directly measurable in the field. Calibration is completed when an objective function defined as a degree of fit between measured and computed output is minimized.
parameter sensitivity The influence of a parameter's value on the model output. Parameters are very sensitive when small changes in the value of a parameter have a

significant effect on an output objective function such as peak concentration or total constituent load.

peak discharge (flow) The maximum instantaneous flow at a specific location resulting from a given storm condition.

pollutant Dredged soil, solid waste, incinerator residue, sewage, garbage, sludge, chemical wastes, biological materials, radioactive materials, heat, wrecked or discarded equipment, rock, sand, dirt and industrial, municipal and agricultural waste discharged into water.

pollutograph Plot of pollutant concentration against time.

propagated error Related to erroneously estimated values of input parameters.

recurrence interval (return period) The average interval of time within which the magnitude of a particular event (e.g. storm or flood) will be equalled or exceeded. e.g 1 in 5 year frequency of 1:5 AEP.

resolution 1. The scale of spatial and temporal disscretization. 2. The size of the time step in a continuous simulation.

response function The computed hydrological and water quality time series, output by the WQM,

return period *See recurrence interval.*

riparian A relatively narrow strip of land that borders a stream or river, often coincides with the maximum water surface elevation of the 100 year storm.

runoff That portion of the precipitation on a drainage area that is discharged from the area into stream channels.

sampling error Associated with the timing and location of the field equipment.

sanitary sewer A sewer that carries liquid and water-borne wastes from residences, commercial buildings, industrial plants, and institutions, together with relatively low quantities of ground, storm, and surface waters that are not admitted intentionally.

sediment Solid material, both mineral and organic, that is in suspension, is being transported, or has been moved from its site of origin by air, water, gravity, or ice, and has come to rest on the earth's surface either above or below sea level.

sedimentation The process of subsidence and deposition of suspended matter carried by water, sewage, or other liquids, by gravity.

sensitivity analysis Consists of 1. varying model coefficients or parameters one at a time, with the amount varied being representative of the uncertainty in the parameter being analyzed, 2. dividing the resulting dimensionless change in computed response by this dimensionless parameter variation, giving the sensitivity gradient, and 3. ranking the resulting modulus of the sensitivity gradients, highest to zero.

sensitivity gradients The ratio of the relative computed response to the relative perturbation in a causative parameter.

sewershed The area of a municipality served by a given sewer network. For example, the area tributary to a given combined sewer overflow or a given WPCP would be termed the sewershed tributory to the overflow or WPCP.

shell Code that is written around existing code, usually to provide better human interaction.

simulation Representation of physical systems and phenomena by mathematical models.

state variable A variable that indicates the various states of the model, where one group of processes may dominate over another.
stormflow The portion of flow which reaches the stream shortly after a storm event.
storm sewer A sewer that carries storm water and surface water, street wash and other wash waters or drainage, but excludes sewage and industrial wastes.
stormwater Water resulting from precipitation which either percolates into the soil, runs off freely from the surface, or is captured by storm sewer, combined sewer, and to a limited degree, sanitary sewer facilities.
streamflow Water flowing in a natural channel, above ground.
structural error 1. related to *disaggregation* (the number and resolution of the processes active); 2. related to *discretization* (the spatial resolution); 3. related to poor *formulation* of one or more of the component process relations and code.
subcatchment Tributary area, part of a catchment.
surcharge The flow condition occurring in closed conduits when the sewer is pressurized or the hydraulic grade line is above the crown of the sewer.
time of concentration (hydraulics) The shortest time necessary for all points in a catchment area to contribute simultaneously to flow past a specified point.
uncertainty A possible value an error may have.
urban runoff Surface runoff from an urban drainage area that reaches a stream or other body of water or a sewer.
urbanization Transformation of a landscape from forest and/or prairie to rural, or from rural to urban.
urbanized area Central city, or cities, and surrounding closely settled territory. Central city (cities) have populations of 50,000 or more. Peripheral areas with a population density of one person per acre or more are included (United States city definition).
variability The different values that a parameter may have.
variance The square of the standard deviation (a measure of uncertainty).
watercourse A natural or constructed channel for the flow of water.
watershed The region drained by or contributing water to a stream, lake, or other body or water. *See also catchment.*
waterway A natural or man-made drainage way. Commonly used to refer to a channel which has been shaped to a parabolic or trapezoidal cross-section and stabilized with grasses (and sometimes legumes), and which is designed to carry flows at a velocity that will not induce scouring.
wet weather flow A combination of dry weather flows, infiltration and inflow which occurs as a result of rain and storms.

This page intentionally left blank.

Converting SI units to U.S. Customary Units

To convert from (SI)	Conversion factor	To get U.S. Customary
Length		
Meters (m)	Multiply by 3.28	Feet (ft)
Meters	Multiply by 1.094	Yards
Meters per second (m/s)	Multiply by 2.237	Miles per hour (mph)
Centimeters (cm)	Multiply by 0.39	Inches (in.)
Milimeters (mm)	Divide by 25.4	Inches
Kilometers (km)	Divide by 1.608	Miles (mi)
$m^3/(m^2\text{-day}) = m/\text{day}$	Multiply by 24.6	gpd/ft^2
Area		
cm^2	Divide by 6.45	Square inch
m^2	Multiply by 10.76	Square foot (ft^2)
Hectare (ha = 10 000 m^2)	Multiply by 2.46	Acre
km^2 (= 100 ha = 106 m^2)	Multiply by 0.387	Square mile
Volume		
Liters (= 1 dm^3)	Divide by 28.3	Cubic foot (ft^3)
m^3	Multiply by 35.4	ft^3
Liters	Divide by 3.78	U.S. gal
Liters	Divide by 4.54	Imp. gal
Liters/(s-ha)	Multiply by 0.014	Inches/hour
m^3	Divide by 3 780	Million U.S. gal (mg)
m^3	Multiply by 8.54	Barrel
m^3	Divide by 1 230	Acre-ft
Liters/s	Divide by 43.75	Million U.S. gal/day (mgd)
Liters/s	Multiply by 15.87	gal/min (gpm)
Liters/m^2	Divide by 40.76	gal/ft^2
Mass		
Gram (g)	Divide by 454	Pounds (lb)
Gram	Multiply by 15.43	Grain
Kilogram[a](kg = 1 000 g)	Multiply by 2.2	Pounds
Newton (= 0.1 kg[b])	Multiply by 0.225	Pounds
Metric ton (= 1 000 kg)	Multiply by 1.1	U.S. ton
Metric ton	Multiply by 0.98	English ton

[a]Metric kilograms in this table are weight kilograms which equal 9.81 (m/s^2) x kg (mass) = 9.81 Newtons.

To convert from (SI)	Conversion factor	To get U.S. Customary
Concentration		
Milligram per liter (mg/liter = g/m^3)	Multiply by 1.0	Parts per million (ppm)
mg/liter	Divide by 2.29	Grain/ft^3
Microgram per liter (ug/liter = 10^{-3} g/m^3)	Multiply by 1.0	Parts per billion (ppb)

To convert from (SI)	Conversion factor	To get U.S. Customary
Density		
kg/m^3	Divide by 16	lb/ft^3
g/m^3	Multiply by 6.24×10^{-5}	lb/ft^3
m^3/kg	Multiply by 16.03	ft^3/lb
Pressure		
g/m^2	Divide by 4 885	lb/ft^2
Bar (= 10^5 N/m^2)	Multiply by 14.2	psi (= lb/in.^{2})
kg/m^2	Divide by 4.89	lb/ft^2
kg/cm^2	Multiply by 14.49	psi
Energy etc.		
Watt (W = N x m/s)	Multiply by 3.41	Btu/hr
Kilowatts (kW = 1 000 W)	Multiply by 1.34	Horsepower (hp)
Kilowatt-hours (kW-hr)	Multiply by 3 409.5	Btu
W/m^3	Multiply by 5	hp/mg
kW-hr/(m^2x^{o}C)	Multiply by 176	Btu/ft^2/^{o}F
kW-hr/(m^3x^{o}C)	Multiply by 53.6	Btu/ft^3/^{o}F
Calories (gram)	Divide by 252	Btu
1 calorie	Multiply by 1.16×10^{-6}	kW-hr
Degree celsius (^{o}C)	Multiply ^{o}C by 1.8 and add 32	Degrees Fahrenheit (^{o}F)

Some constants

1 m^3 of water weighs 1 000 kg.
1 ft^3 of water weighs 62.4 lb.
1 U.S. gal of water weighs 8.34 lb.
1 Imp. (English) gal of water weighs 10 lb.
1 day has 1 440 minutes and 86 400 seconds.

Some magnitude prefixes for SI units

G	giga	10^9
M	mega	10^6
k	kilo	10^3
m	milli	10^{-3}
μ	micro	10^{-6}

S.I. standard time abbreviations
(WARNING – these may not be the convention adopted by SWMM).

Second	s
Minute	min
Hour	h
Day	d
Week	week
Month	month
Year	y

Programs and Models

Access	MS Database
ADAPT	*see GIS ADAPT*
AGNPS	Agricultural Non-Point Source
AML	Arc Macro Language
ANNIE	a USGS hydrologic timeseries data management system
ArcFM	AM/FM/GIS application for the utility industry
ARCINFO	a GIS program (ESRI *(q.v.)*)
ArcView	a GIS program (ESRI *(q.v.)*)
ASDM	Atmospheric and Sediment Deposition Model
AutoCAD(r)	an automated computer-aided drafting package
Avenue	a scripting language bundled with Arcview (*q.v.*)
BASIC	a programming language
BASINS	USEPA Better Assessment Science Integrating Point and Nonpoint Sources program
CASE	Computer Aided Software Engineering
CFA	Consolidated Frequency Analysis Package
CMP	Computer Mapping Program
COGO	a CAE application module
COMBINE	a time-series module of the SWMM (*q.v.*) program
CORMIX	Cornell Mixing Zone Model (USEPA)
CUF	California Urban Futures Model (to evaluate spatial consequence of loccal land use alternatives)
DBMS	Database Management System
DEM	Digital Elevation Model
DOMECOL	a water pollution model
DOMOD7	Dissolved Oxygen Model Version 7
DOS	a computer operating system
DSPLAY	a graphical display module of HECDSS
DSS	Data Storage System, a type of file format
DSSUTL	a data storage system utility module of HECDSS
DTM	Digital Terrain Model
DUH	digital elevation models
DWOPER	Dynamic Wave Operational Model
EPICWQ	Erosion/Productivity Impact Calculator -Water Quality
ERDAS	a raster based GIS (Earth Resource Data Analysis System)
EXAMSII	Exposure Analysis Modeling Systems II
EXSUDS	extended SUDS program
EXTRAN	extended transport, a hydraulics computational module of SWMM (*q.v.*)
FDAM	flood damage analysis model
FORTRAN	a high-level programming language
GCM	General Circulation Models
GeoGUIDE	an ARC/INFO GUI
Geo/SQL(r)	a geographical program
GeoSelect	GIS interface to facilitate raingage selection and data retrieval
GeoSTORM	a river program (Innovative Systems Developers Inc)
GH	GIS-Hydrology interface (Univ. of Texas CRWR)
GISFPM	GIS floodplain management
GPS-X	a dynamic model for the design, operation and control of wastewater treatment plants
GRASS	Geographical Resources Analysis Support System (US Army COE)

GUI	Graphical User Interface
HEC1	hydraulic modeling program to determine discharges (US Army Corps of Engineers)
HEC2	hydraulic modeling program to determine water surfaceelevations (do)
HECAD	interface linking HEC1 with ADAPT (do)
HECDSS	HEC hydrologic time-series data management system (do)
HECRAS	HEC river analysis program (do)
HSPF/HSP-F	Hydrologic Simulation Program-Fortran (USEPA)
HWY DSS	Highway Decision Support System
HYDRA	Hydrologic Data Retrieval and Alarm system
HYDRO	a steady state model for WWTP (*q.v.*) hydraulic processes. (CH2MHill).
HYMO	a hydrologic model (USDA)
IDRISI	GIS program (Clark Univ.)
IFG4	hydraulic/velocity simulation program
IFORM	a rainfall file format defined by RAIN module of SWMM (*q.v.*)
ILLUDAS	Illinois Urban Drainage Area Simulator
IOWDM	a file input/output utility for ANNIE (*q.v.*)
LFA	Low Flow Frequency Analysis
MATHPK	a math and statistics module of HECDSS
MODFLOW	Modular 3-Dimensional Finite-Difference Ground-water Flow Model
MTOPOND	Ontario Ministry of Trans-portation quality control performance model for stormwater ponds
MTV	Model Turbo View
MTVE	Model Turbo View -EXTRAN (a post-processor for EXTRAN)
NT	Microsoft Windows NT operating system
OASIS	On-line Access and Service Information System
PCSWMM	a decision support system for the SWMM (*q.v.*) program (CHI (*q.v.*))
PCSWMM GIS	a graphical model development environment which provides links to GIS/AM/FM databases (CHI (*q.v.*))
PCSWMMPP	A tool to aid the hydrologic design of permeable pavement
PDF	Probability Density Function
PDM	Probabilistic Dilution Model
PLAN	a CAE applications module
QUAL2E	Enhanced Stream Water Quality Model
QUALHYMO	a version of HYMO (*q.v.*)
RAIN	a time-series module of the SWMM (*q.v.*) program
RAINEV	a rainfall analysis computer module of SIMPTM (*q.v.*)
RAP	Rainfall Analysis Program
REPGEN	a report presentation module of HECDSS
RUNOFF	a hydrology computational module of the SWMM (*q.v.*) program
SAMS	Sewerage Analysis and Management System
SDE	Spatial Database Engine
S/T; S-T	Storage/Treatment module of SWMM (*q.v.*)
SIMPLE	a geo-referenced hydrologic model for remotely sensed data
SIMPTM	Simplified Particulate Transport Model
SLAMM	Source Loading and Management Model (Pitt, Univ. of Alabama)
STATS	a time-series module of the SWMM *(q.v.)* program (STATISTICS)
STORAGE	a computational module of the SWMM (*q.v.*) program
STORET	Storage and Retrieval System
STORM	Storage Treatment Overflow Runoff Model (US Army COE (*q.v.*))
SWAT	Soil and Water Assessment Tool
SWMM	US EPA Stormwater Management Model
SWRRB-WQ	Simulator Water Resources In Rural Basins - Water Quality
SWSTAT	a statistics module of ANNIE (*q.v.*)
SYNOP	statistical rainfall analysis program
TEMP	a time-series module of the SWMM (*q.v.*) program
TIGER	Topographically Integrated Geographic Encoding and Referencing
TOXIWASP	See WASP4

TR-20	USDA (*q.v.*) SCS (*q.v.*) Technical Release 20 - a watershed hydrology model to route a design storm hydrograph through a pond
TR55	USDA SCS Technical Release 55 -
TRANSPORT	a computational module of the SWMM (*q.v.*) program
WASP	Water (Quality) Analysis Simulation Program (USEPA)
WATRSHD	Watershed Analysis Tools for Reporting, Statistics, Hypothesis-testing and Display
WATSTORE	Water data storage and retrieval system
WDM	Watershed Data Manage-ment, a type of file format
XP-SWMM	An expert systems version of SWMM (XP Software)

This page intentionally left blank.