uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Igor Iskra

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Civil Engineering)

GRADE / DEGREE

Department of Civil Engineering

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Propagation of Uncertainty in a Watershed Model

TITRE DE LA THÈSE / TITLE OF THESIS

Ronald L. Droste

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Konrad Gajewski

Colin Rennie

Chandra Madramootoo

Paul Van Geel

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

PROPAGATION OF UNCERTAINTY IN A WATERSHED MODEL

by

Igor Iskra

Ph. D. Thesis

Submitted to the School of Graduate Studies and Research

Under the supervision of Prof. Ronald L. Droste

In partial fulfillment of the requirements for the degree of Ph. D. in Civil Engineering

The Ottawa-Carleton Institute for Environmental Engineering
Department of Civil Engineering
University of Ottawa, Ottawa, ON
Canada, K1N 6N5

April, 2007

Your file Votre référence
ISBN: 978-0-494-49361-8
Our file Notre référence
ISBN: 978-0-494-49361-8

ABSTRACT

The Hydrological Simulation Program FORTRAN (HSPF) model was calibrated for the South Nation watershed located in Eastern Ontario. Three nonlinear automatic optimization techniques were applied and compared: Gauss-Marquardt-Levenberg (GML) method, Random Multiple Search Method (RSM), and Shuffled Complex Evolution method developed at the University of Arizona (SCE-UA). The best GML, RSM, and SCE-UA variable values beyond which objective function improvement is insignificant were suggested.

It was found that more than one parameter set is able to maintain the model in a calibrated state which reflects correlation among model parameters and equations. The lowest value of the objective function (OF) does not necessarily correspond to the optimum solution. Comparison of scatter plots, graphs of residuals, and plots of cumulative differences are required to determine the best model parameter set. Combination of Nash and Sutcliffe (NS) model fit, coefficient of efficiency, and index of agreement proved to be the best statistics for model comparison.

Proper definition of the OF is crucial to successful model calibration. Over 60 single and compound OFs were compared. The OF expressed as a log of observed and simulated flows was found to be the most appropriate single OF. A compound OF expressed as a sum of squared residuals of equally weighed log-transformed maximum (top 1% of flows), minimum (bottom 20% of flows), and middle flows was found to be best for most general model applications.

Uncertainty of HSPF parameters was explored by the method of moments (MM), Monte Carlo (MC) with Latin hypercube and induced correlation, and response surface (RS) methods. Typically, the MM results in the most conservative uncertainty. The 95% confidence intervals of parameter uncertainty correspond to up to 10% variations in spring maximum flows.

The predictive confidence interval and predictive noise for spring maximum and autumn minimum flows using single and compound OFs were computed. The predictive intervals were computed from the 95% confidence limits of the OF.

It was found that HSPF can be an efficient tool to predict flows in ungaged watersheds with parameter transfer from calibrated neighbouring watersheds. The impact of DEM resolution on HSPF topographical parameters was studied.

ABSTRACT

Watershed models are an essential tool for quantitative and qualitative assessment, planning, development, and management of water resources. The Hydrological Simulation Program FORTRAN (HSPF) developed by the US Environmental Protection Agency (USEPA) is one of the most comprehensive and widely used watershed models in North America. It was the model used in this study.

The model was calibrated for the South Nation watershed located in Eastern Ontario. Calibration of HSPF is a very tedious and time consuming task involving more than 100 optimization parameters. Three nonlinear automatic optimization techniques were applied and compared: Gauss-Marquardt-Levenberg (GML) method, Random Multiple Search Method (RSM), and Shuffled Complex Evolution method developed at the University of Arizona (SCE-UA). RSM is a significant improvement to GML. All three automatic methods produced good results and proved to be more efficient than manual HSPF calibration. The best GML, RSM, and SCE-UA variable values beyond which objective function improvement is insignificant were suggested.

More than one parameter set is able to maintain the model in a calibrated state which reflects correlation among model parameters and equations. The lowest value of the objective function (OF) does not necessarily correspond to the optimum solution. Comparison of scatter plots, graphs of residuals, and plots of cumulative differences are required to determine the best model parameter set. The statistical parameters of coefficient of determination, mean error (ME), and daily root mean square (DRMS) error should not be used for comparison of observed and simulated data because they are insensitive to changes in the OF. Combination of Nash and Sutcliffe (NS) model fit, coefficient of efficiency, and index of agreement proved to be the best statistics for model comparison.

Proper definition of the OF is crucial to successful model calibration. Given the various response characteristics of the observed data, it is important to include all relevant information in the OF for a particular problem. Also, it is important to assign appropriate weighting factors to different OFs. Weighting factors will depend on the study objectives. Typically, weights are assigned proportional to their perceived significance in the overall optimization. For example, if low flows are of importance for model calibration the appropriate OF can be given higher weight.

Over 60 single and compound OFs were compared. The OF expressed as a log of observed and simulated flows was found to be the most appropriate single OF. A compound OF expressed as a sum of squared residuals of equally weighed log-transformed maximum (top 1% of flows), minimum (bottom 20% of flows), and middle flows was found to be best for most general model applications.

The Pareto front concept was used in selecting the optimal parameter set. Several single OFs were minimized individually over the feasible parameter space. Due to various trade-offs among the different OFs, the solution, in general, is not a single unique set of parameters but a Pareto set of solutions which visualizes and quantifies tradeoffs between individual OFs in a multi-component OF. Possibly due to structural and data errors, irregular Pareto fronts were observed.

Sensitivity analysis was conducted and the most and the least sensitive parameters were identified. Sensitivity of a parameter depends on the number of parameters adjusted simultaneously. As more parameters are optimized simultaneously, a wider range of parameter values can maintain the model in the calibrated state. The parameters most and least sensitive to high and low flows were identified.

Uncertainty of HSPF parameters was explored. Three uncertainty methods were used: the method of moments (MM), Monte Carlo (MC) with Latin hypercube and induced correlation, and response surface (RS) methods. As more parameters are optimized simultaneously, typically higher uncertainty for the value of any given parameter is observed while maintaining the same error in the OF. Typically, the most sensitive parameters have the lowest parameter uncertainty and vice-versa.

Different methods provide different estimations of uncertainty for each parameter. Typically, the MM results in the most conservative uncertainty. The largest uncertainty was estimated by the MC method. The 95% confidence intervals of parameter uncertainty correspond to up to 10% variations in spring maximum flows. Although RS methods are fast, they only provide a rough estimation of uncertainty and are unlikely to describe the complex relationship between input and output variables in a nonlinear model. Taking into account disadvantages of the MM and RS methods, it is beneficial to use MC, ensuring that the proper sampling technique is used.

The predictive confidence interval and predictive noise for spring maximum and autumn minimum flows using single and compound OFs were computed. The predictive intervals were computed from the 95% confidence limits of the OF. The random predictive error comprises the

major part of uncertainty in prediction of high and low flows. Higher predictive noise is associated with the compound OF. Higher predictive confidence intervals are typical for a single OF.

Spatial scaling of 14 watersheds in Eastern Ontario was assessed using the product of moments. The logarithms among the kth moments of average annual runoff, river length, peak flows, and watershed area were found to be linear, which indicates the applicability of simple scaling laws for the these watersheds. HSPF parameters were analysed at different spatial scales. HSPF parameters were not scale dependent and can be transferred. It was found that HSPF can be an efficient tool to predict flows in ungaged watersheds with parameter transfer from calibrated neighbouring watersheds and digital elevation model (DEM) data. The impact of DEM resolution on HSPF topographical parameters (area, cross-section geometry, length and slope of overland plane, stream length, and elevation differences) was studied. Small watersheds are more sensitive to DEM resolution. Area is the most sensitive parameter for simulation of minimum and maximum flows. Assuming that the error in maximum flows should not exceed 6% and the error in minimum flows should be within 12%, it is critical to have DEM at 35 m resolution for small watersheds (<200 km^2) and at 100 m resolution for larger watersheds.

Availability of detailed meteorological data is important. For daily flow prediction at least three hourly time series are required: precipitation, potential evaporation, and air temperature. Slight improvement in calibration is observed if wind speed, solar radiation, cloud cover, and dewpoint temperature data are available. The most important meteorological data are precipitation. For a reasonable flow prediction the missing precipitation data should not exceed 20 - 30%, and they should be substituted with the data from a nearby meteorological station.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor Dr. Ronald L. Droste for continuous support, encouragement, and guidance throughout the project. I feel truly privileged and honoured to be his student.

I would like to thank my committee members Dr. Paul Van Geel, Dr. Colin Rennie, Dr. Konrad Gajewski, and Dr. Pedro Lomónaco Tonda for their time and efforts in reviewing this work at different stages. Their suggestions were constructive and beneficial for the research.

During the years of my Ph.D. study, I have been blessed with friends who became family to me. My officemate Cigdem Eskicioglu deserves special thanks for her courage and patience in tolerating me for the last four years.

I would like to extend my thanks to Frank Aposaga for our invaluable discussions on various research topics. I am very grateful to my friends Muna, Isil, and Nuno for creating a friendly and productive environment.

I would like to thank the International Development Research Centre (Ottawa) for awarding me a scholarship which supported me during the first three years of my Ph.D. research.

I would also like to acknowledge the love and encouragement of my mother, Zoia, father, Victor, mother-in-law, Tatiana, and father-in-law, Vitaliy. They relieved me of considerable burden by baby-sitting my children during this seemingly endless journey.

I dedicate this thesis to my loving wife, Yuliya, and my children, Alice and Edwin. Their love and support made this Ph.D. thesis possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

AAFC	Agriculture and Agri-Food Canada
AGWETP	Fraction of remaining evapotranspiration from active groundwater (HSPF parameter)
AGWRC	Base groundwater recession (HSPF parameter)
ARS	Adaptive random search
ASTM	American Society of Testing Materials
BASETP	Fraction of remaining evapotranspiration from baseflow (HSPF parameter)
CCFACT	Condensation and convection melt factor (HSPF parameter)
CE	Coefficient of efficiency
CEPSC	Interception storage capacity (HSPF parameter)
Cfs	Cubic feet per second
DEEPFR	Fraction of groundwater inflow to deep recharge (HSPF parameter)
DEM	Digital elevation model
DLG	Digital line graph
DRMS	Daily root mean square error
EA	Evolutionary algorithm
ET	Evapotranspiration
FOVE	First-order variance estimation
GA	Genetic algorithm
GIS	Geographic information system
GML	Gauss-Marquardt-Levenberg method
GLOBE	Global optimization tool
GLUE	Generalized likelihood uncertainty estimation
GOM	Global optimization method
HBV	Swedish hydrological model
HEC	Hydrologic engineering center
HSPF	Hydrological simulation program FORTRAN
HSPFEXP	Hydrological simulation program FORTRAN expert system
IA	Index of agreement
IMPLAN	Impervious land subroutine in HSPF
INFILT	Index to infiltration capacity (HSPF parameter)
IRC	Interflow recession parameter (HSPF parameter)
LHS	Latin hypercube sampling
LOESS	Locally weighted scatterplot smoothing
LOM	Local optimization method
LZETP	Lower zone evapotranspiration (HSPF parameter)
LZSN	Lower zone nominal soil moisture storage (HSPF parameter)
MAE	Mean absolute error
MAP	Mean areal precipitation
MAPE	Mean areal potential evaporation
MAT	Mean areal temperature
MC	Monte-Carlo
MCAT	Monte-Carlo analysis toolbox
ME	Mean error
MM	Method of moments
MOCOM-UA	Multi-objective complex evolution

MOE	Ministry of Environment
NRVIS	Natural resource and values information system
NSUR	Manning's n for overland flow (HSPF parameter)
NS	Nash and Sutcliffe model fit
NWSRFS	National weather service river forecast system
OF	Objective function
OMNR	Ontario Ministry of Natural Recourses
PDF	Probability distribution function
PERLND	Pervious land subroutine in HSPF
PEST	Parameter estimation tool
PMAE	Percent mean absolute error
PME	Percent mean error
RS	Response surface
PSRM	Pennsylvania state runoff model
PWATER	HSPF subroutine which calculates runoff from a pervious land segment
PWM	Probability-weighted moment
PWQMN	Provincial water quality monitoring network
RCHRES	Water body and stream reach subroutine in HSPF
RSM	Random multiple search method
SA	Simulated annealing
SCE-UA	Shuffled complex evolution method – University of Arizona
SNOWCF	Snow gage correction factor (HSPF parameter)
SNR	South Nation river
SVD	Saturated vapor density
SWAT	Soil and water assessment tool
SWG	Stochastic weather generator
SWM	Stanford watershed model
TIN	Triangulated irregular network
TMDL	Total maximum daily load
USEPA	United States Environmental Protection Agency
UZSN	Upper zone nominal soil moisture storage (HSPF parameter)
WMO	World Meteorological Organization
WSC	Water Survey of Canada

CHAPTER 1

INTRODUCTION

Watershed models are an assembly of concepts in the form of mathematical equations that portray watershed phenomena. An increasing demand for effective water resources management, wide availability of satellite imagery, geographical information system (GIS), and computing technology has led to rapid development of watershed modeling. The principal question to be answered by the models is how human society can be better prepared for major environmental challenges and better manage and protect the environment. In particular, watershed models are an essential tool for quantitative and qualitative assessment of streamflow, groundwater, and water resources management. It is believed that in the near future, the use of models will become routine for water resources planning, development, and management.

Watershed models are nonlinear and complex. Interactions among watershed components are essentially nonlinear and involve physical, chemical, and biological processes. Furthermore, watershed processes take place over a wide range of temporal and spatial scales.

Watershed modeling allows exploration from an integrated perspective. In contrast, field observations are inevitably limited to certain aspects of a complex watershed system. A well-tested model can be a good representation of the watershed as a whole, its dynamics, and its responses to possible external changes and therefore can provide a foundation for prediction.

Models can be used to analyze a problem, to organize our understanding, and to formulate critical experiments to test different hypotheses. However, models are always approximations of reality and they cannot precisely represent the whole system. Moreover, there is no single accepted method or test that determines whether or not a model is valid *(Beven, 2000)*. A famous statistician, George Box, once said "all models are wrong, some of them are useful." *(Clauson, 1997)*. This classical saying should be kept in mind when dealing with models.

Model parameters are the key elements of any model. Model parameters are often conceptually related to physical processes. The better natural process are understood – the higher degree of precision that can be achieved for model parameters. In watershed hydrology the physical processes are complex and highly spatially and temporally variable. As a result they are not well understood and naturally variable from watershed to watershed. Therefore, model parameters have to be estimated and calibrated for each particular application.

Model equations and parameters are idealized representations of a real watershed, and often they are not directly related to observable watershed properties. There are many errors and uncertainties in model structure, input data, and in parameter formulation. As a result, it is difficult to estimate all uncertainties and develop reliable procedures for parameter estimation.

In recent years the demand for water resource studies has increasingly shifted to ungaged areas where meteorological and hydrological data are not available, or only partly available. In such studies hydrological models are the only means to provide a basis for development and management decisions. The models are often blindly used with different resolutions of input data and on different temporal and spatial scales without evaluating their uncertainty. At some point the model may become a useless tool, because using these data may lead to uncertainty similar to guessing.

1.1 Objectives

The current research has four main objectives.

- Improve efficiency of HSPF by investigating data requirements (amount and resolution), robustness, and limitations of the model. Examine ways of overcoming overparameterization in order to enhance the model performance. Study the impact of spatial scaling and data resolution on sensitivity and uncertainty of HSPF.
- Enhance effectiveness of some nonlinear optimization methods through detailed study of their variables. The problems of parameter intercorrelation, insensitivity, and matrix singularity are to be addressed.
- Study the role and ways of defining the objective function (OF) for watershed model calibration. Determine the optimum formulation of single and compound OFs in multi-objective optimization.
- Examine different types and sources of uncertainties in the watershed model. Investigate the propagation of uncertainty in the watershed model: input data uncertainty, parameter uncertainty, and uncertainty of predictions.

1.2 Scope of research

The overall flow chart of the research is presented in Figure 1.1. Numbers on the figure indicate the sequence of steps that were fulfilled.

The thesis is organized in a paper-format presented in 8 chapters. The main results are presented in Chapters 3, 4, 5, 6, and 7.

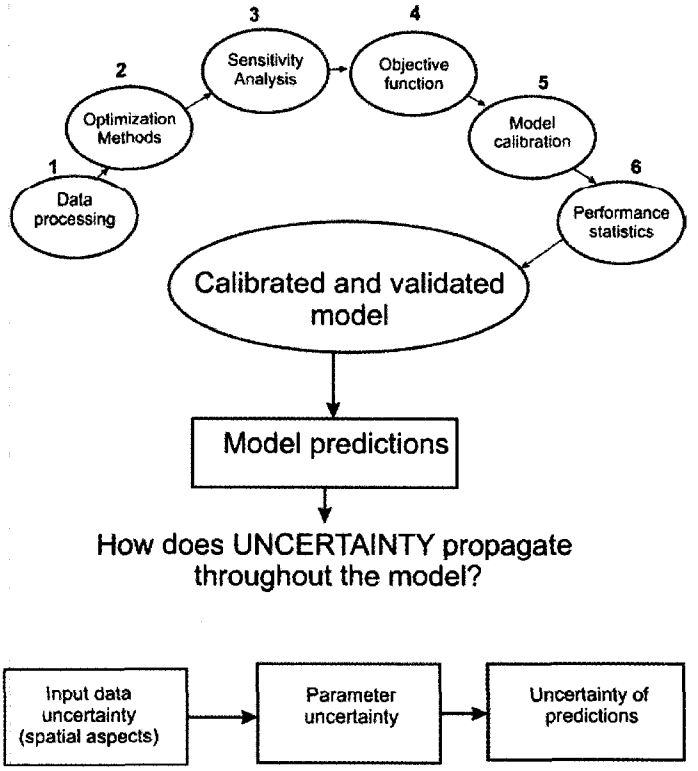


Figure 1.1 Flow chart of the research

General background of the research and extensive literature review on the topic are provided in Chapter 2. Its subsections deal with the model description, ways of calibration and validation, advantages and disadvantages of various local and global optimization techniques, fundamentals of uncertainty analysis, and peculiarities of spatial scaling and temporal resolution in watershed modeling. Application of nonlinear automatic optimization techniques for calibration of HSPF are investigated in Chapter 3. Three different methods are compared and efficient mechanisms of their realizations are suggested. This manuscript was accepted for publication in the journal *Water Environment Research*. Parameter uncertainty of HSPF is discussed in Chapter 4. Propagation of parameter uncertainty to the simulated results is discussed. The manuscript is currently under review in the *Canadian Water Resources Journal*. Predictive uncertainty of maximum spring flows and minimum flows is discussed in Chapter 5. The manuscript is under review in the *Journal of Hydrological Processes*. Impact of spatial scaling and resolution on sensitivity and uncertainty of a watershed model's is presented in Chapter 6. The manuscript was

submitted to the *Journal of Hydrology*. Ways of defining the objective function for model calibration are discussed in Chapter 7. This chapter will be finalized as a paper at a later date. Conclusions of the research are presented in Chapter 8.

Results of this research were presented at the following conferences and workshops.

- 41th Central Canadian Symposium on Water Quality Research, February 13-14, 2006, Burlington, Canada.
- 78th Water Environment Federation Conference and Exhibition, WEFTEC, October 29-November 2, 2005, Washington, USA.
- Workshop on Distributed Hydrological Modeling for Water Resources Management at River Basin Scale, International Research School of Water Resources, August 22-27, 2005, Copenhagen, Denmark.
- 21st Eastern Regional Conference of Canadian Association on Water Quality, November 4, 2005, Quebec City, Canada.
- 40th Central Canadian Symposium on Water Quality Research, February 14-15, 2005, Burlington, Canada.
- USEPA workshop, Application of BASINS-HSPF Watershed Modeling System, September 20-24, 2004, Santa Clara, CA, USA.
- 39th Central Canadian Symposium on Water Quality Research, February 9-10, 2004, Burlington, Canada.
- 19ème Congrès de l'Est du Canada de l'ACQE. 24 October 2003, Universitè de Sherbrooke, QC.

1.3 References

Beven, K.J. (2000). Rainfall-runoff modeling: the primer, Wiley, Chichester.

Clauson, J. (1997). Cyberquality resources: tools and tutorials, Quality Progress, 11, 112-114.

CHAPTER 2

BACKGROUND AND LITERATURE REVIEW

This chapter describes the background of the problem and provides a literature review of existing approaches and practices in watershed modeling.

2.1 Watershed models

2.1.1 Overview

The integration of simple hydrologic models and simulation of virtually entire watersheds began in the 1960s. The most comprehensive model, the Stanford watershed model (SWM), was developed by Crawford and Linsley (1966). Later on, SWM was updated and several new subroutines were added. Today SWM is known as HSPF. Simultaneously with SWM, several other models were developed, for example, the watershed models of *Dawdy and O'Donnell (1965)*, *Kuchment (1971)*, and the Hydrologic Engineering Centre (*Feldman, 1981*).

A number of watershed models are used today. They differ in complexity and internal structure. Some models have mainly regional application. For example, the University of British Columbia model (UBC) and WATFLOOD are popular for hydrological simulation in Canada. The RORB and WBN models are commonly used for flood forecasting and landuse effects in Australia *(Singh and Woolhiser, 2002)*. TOPMODEL and SHE are standard hydrologic simulation models in many European countries. The HBS model is the standard model for watershed simulation and flow forecasting in Scandinavian countries. The ARNO, LCS, and TOPIKAPI models are popular in Italy. The Xinanjiang model is commonly used in China *(Singh and Woolhiser, 2002)*. HEC-HMS is considered a standard model for the private sector in the US. An overview of the most widely used watershed models is presented in Appendix A.

Several attempts have been made to compare different watershed models. For example, an evaluation of the capabilities of 28 hydrological models was conducted in 1985 by the US Task Committee on Quantifying Land-Use Change Effects. The commission concluded that confidence in the capabilities of surface water models "appear(s) to be based upon personal experience, possibly tempered by belief in the model originators" *(US Committee, 1985)*.

The World Meteorological Organization (WMO) sponsored three comprehensive studies on comparison of watershed models. The first study was published in 1975 and it was titled "Intercomparison of conceptual models used in operational hydrological forecasting" *(WMO,*

1975). The second study was "Intercomparison of models of snowmelt runoff" *(WMO, 1986)*. The third study was "Simulated realtime intercomparison of hydrological models" *(WMO, 1992)*. These studies have limited value today, since models have changed dramatically over the last 15 years. Besides these WMO studies, no comprehensive efforts have been made to compare major watershed models *(Singh and Woolhiser, 2002)*. However, some work has been done by researchers to compare separate components of watershed models.

It is important to remember that different models were developed for different purposes, and often different simplifications and assumptions underlie model structure. The search for the best model should typically begin with the objectives of the study. In watershed modeling the interest is often in overall performance of the system (e.g., daily flow, snow depth) rather than the details (e.g., moisture content in the lower zone). In using a watershed model one is often looking for a reliable predictor, rather than a fundamental explanation, or a detailed reproduction of the system.

The HSPF model was employed in the current research. It is the most widely used watershed model in North America. It was developed by USEPA as a standard tool for many environmental procedures including total maximum daily load (TMDL) development. The model is semi-distributed, open-source, and in the public domain.

2.1.2 Classification of watershed models

All watershed models are approximations of reality. Output of the natural system can hardly be simulated with certainty. The main reason is that variability of the hydrological phenomena exists at least in five axes (three space dimensions, time, and randomness). There is no known practical way to take into account all sources of variability in a single model. Although a number of model classifications exist, probably the most comprehensive classification was proposed by *Chow et al. (1988)*. It is based on how a model treats the variability in space, time, and randomness. A slightly modified version of *Chow et al.'s (1988)* classification is presented in Figure 2.1.

The first level is based on how a model treats the randomness of hydrological processes. Models can be deterministic or stochastic. Stochastic models allow some randomness or uncertainty in the model outcomes due to input variables, boundary conditions, or model parameters (*Beven, 2000*). As a result, the output of a stochastic model contains a random component; it has its own distribution and can be presented as a range of values with confidence limits. Stochastic models are typically used when the number of random variables in the simulated process is large.

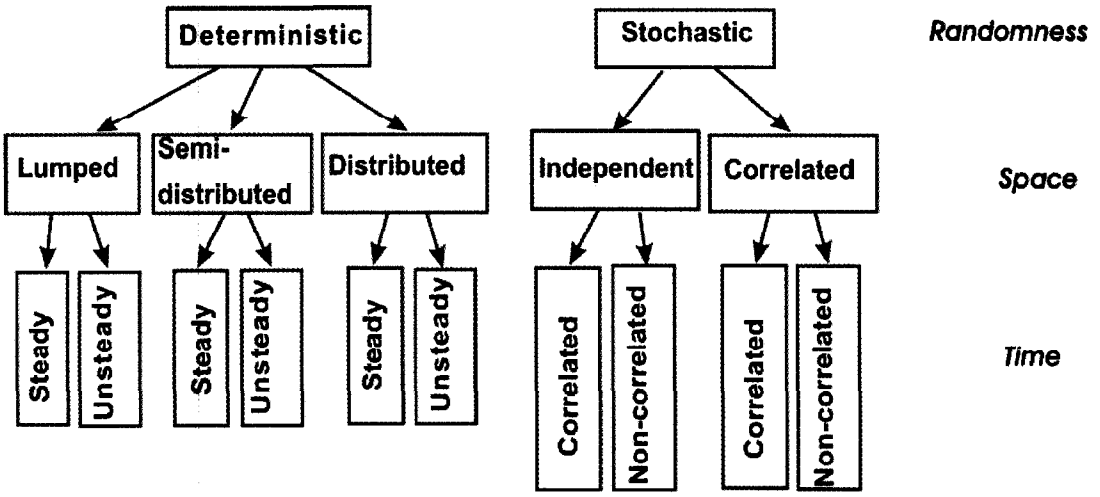


Figure 2.1 Classification of watershed models according to the way they treat randomness, time, and space variability (adapted from Chow et al., 1988)

Deterministic models allow only one outcome for a simulation with one set of input and parameter values, i.e., for each unique set of input data the model computes fixed and repeatable results. Most watershed models are deterministic because the variability of unknown parameters is relatively small in comparison to variability in known factors. However, nowadays, some models add stochastic components, such as probability distribution function (PDF), stochastic error, etc.

The second level is based on how a model treats spatial variation. Two groups are distinguished in stochastic models: space-independent and space-correlated. Deterministic models are subdivided into lumped, semi-distributed, and distributed models.

Lumped models deal with an entire watershed as one unit and do not respect spatial variability of parameters, boundary conditions, or properties within a watershed. Spatial variability of model parameters usually is evaluated by an area-weighted average procedure. The structure of lumped models is simple, data requirements are minimal, and usually they are easy to understand, and fast to set up and calibrate. Lumped models are not a good choice for event predictions. If the primary interest is discharge simulation, these models, once calibrated, can be as good as very complex physically-based models (*Beven, 2000*).

Semi-distributed models or simplified distributed models divide the basin into smaller subwatersheds. Model parameters are partially allowed to vary spatially. There are two main types of semi-distributed models. The first is kinematic wave theory models (e.g., HEC-HMS, HSPF), which are simplified versions of surface/subsurface flow equations of physically-based

hydrologic models (*Beven, 2000*). The second group is probability distributed models (e.g., TOPMODEL) where spatial resolution is accounted for by using probability distributions of input parameters across the basin (*Cunderlik, 2003*). An advantage of semi-distributed models is that their structure is more physically based than lumped models and they are less data demanding than fully distributed models.

Distributed models allow parameters to vary in space and in resolution chosen by the modeller (e.g. MIKE-SHE). Equations for state variables are solved for every grid cell and parameters have to be specified for every element of the model. Distributed models require a large amount of data for parameterization of each grid cell. Physical processes in distributed models are modeled in detail and, if properly applied, they can provide the highest degree of accuracy *(Cunderlik, 2003)*. In strict terms, in a fully distributed model all its aspects must be distributed: parameters, initial and boundary conditions, sources, and sinks. In reality, due to practical limitations of data and discrete watershed geometry, fully distributed models do not exist, only partially distributed models are achieved. Some examples of lumped, semi-distributed, and distributed watershed models are presented in Appendix B.

At the third level the model accounts for time variations. Deterministic models can be steady-state and unsteady state. Watershed models can also be divided according to the type of hydrologic event into two categories: event driven and continuous process. The latter is capable of simulating both short term and continuous events. Event-driven models simulate short individual precipitation-runoff events with emphasis on infiltration and runoff. Continuous process models take into account all runoff components, including rate of moisture recovery during dry seasons.

Given such a variety of models, the question arises as to what model should be chosen for solving a particular problem. The interesting graph (Figure 2.2) illustrating the conceptual relationship between model complexity, data availability, and performance of the model was suggested by *Grayson and Bloschl (2001)*. "Data availability" implies both the amount and quality of data. Having detailed soil, landuse, hydrogeological, and meteorological data is equivalent to "large" availability while having only runoff and precipitation data would mean "small" availability. Model complexity ranges from small (lumped models) to large (physically-based, fully distributed models). The graph suggests that if a certain amount of data is available, an optimal model exists beyond which results deteriorate. Beyond the optimum, the model becomes too complex with too many parameters, and there are insufficient data to support the results. The

general approach for solving "what model complexity is needed" is, probably, not to add complexity when there is no means to test whether this improves a model or makes it worse.

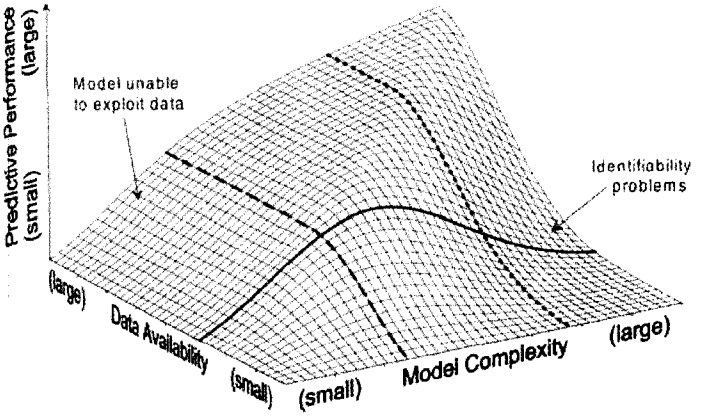


Figure 2.2 Schematic diagram of the relationship between model complexity, data availability, and predictive performance (adapted from *Grayson and Bloschl, 2001*)

Although a certain consensus exists about theoretical superiority of distributed physically-based watershed models, there is a wide range of opinions whether they offer a significant improvement in solving practical watershed problems compared to the well-proven lumped conceptual models. *Beven (1989)* argues that theoretical consideration of scale problems in the current generation of distributed models is the same as in lumped conceptual models and even further, they "are not well suited to applications to real catchments". *Grayson et al. (1992)* support his view and claims that distributed models have been oversold by their developers.

2.1.3 HSPF (critical analysis)

HSPF is a set of computer codes that can simulate hydrologic processes and water quality in water bodies. It is a comprehensive watershed model that allows the integrated simulation of land and soil contaminant runoff processes with in-stream hydraulic and sediment-chemical interaction. HSPF is a fully dynamic model with continuous simulation of runoff and sediment mobilization. It models generic water-quality parameters (pH, DO, etc.) and more complex ones BOD, P- and N-nutrients, pesticides, and phytoplankton. The chemical processes modeled include hydrolysis, oxidation, photolysis, biodegradation, volatilization, and sorption, all of which require coefficients and rate constants. A number of options are available for depicting various agricultural land practices. One-dimensional lakes can be included in the stream segmentation. The subsurface budget is modeled in two vadose layers, which can interact with the surface

through plants or interflow. Percolation to a deep aquifer is included. The aquifer is not modeled, but treated as a sink of water.

In HSPF the watershed is subdivided into smaller subwatersheds based on location of meteorological and hydrological monitoring stations, elevation, stream connectivity, and soil types. Each subwatershed is further subdivided into land uses segments (forest, pasture, urban, etc.). The receiving watercourse for each subwatershed is modeled as completely (longitudinally, laterally, and vertically) mixed. It is assumed that streams have a trapezoidal profile. For modeling purposes, water, sediment, and water quality constituents move laterally to a downslope segment or to a reach. Each subwatershed is modeled with data from one representative meteorological station. The hydrological component of HSPF computes surface storage, infiltration flux, and storage in the two vadose zones and two groundwater zones (one drives baseflow and the other represents percolation to a deep aquifer).

Literally all processes identified in the surface water budget correspond to an equation in HSPF. Sometimes it is difficult to evaluate relations of these equations to standard processes discussed in the literature *(Ward and Benaman, 1999)*. Model documentation *(Bicknell et al., 2001)* gives limited descriptions of the mechanisms of key modeling processes such as overland flow, sediment detachment, and surface erosion.

The user of HSPF must be aware of a number of model assumptions. Some model inputs (e.g., topography, cross-sections, etc.) are assumed to be constant over a whole period of modeling. Some inputs vary in time but are assumed to be uniform in space (e.g., air temperature and other meteorological characteristics).

Among other disadvantages, HSPF requires a large amount of data, which includes watershed data, soil data, vegetative cover, and detailed meteorological data sets. Acquisition, reformatting, and management of input time-series consume much time in model application. In addition, the model has a wealth of empirical calibration parameters (more than 100) that must be determined from handbooks, by calibration, or from field measurements.

The last available HSPF version 12.0, released in 2001, was used in this research. It operates under Windows XP. The code and manual were prepared by AQUA TERRA Consultants of Mountain View, CA under sponsorship of the USEPA *(Bicknell et al., 2001)*.

2.1.3.1 Model structure

HSPF is a complex model with a modular or compartmental structure. The modeller can isolate and focus on modules of interest, bypassing others. Hydrology is simulated in three main modules: PERLND (pervious land), IMPLAN (impervious land), and RCHRES (water body or a stream reach). Descriptions of HSPF modules and their required data are given in Appendix C.

The key section of PERLND is PWATER which calculates runoff from a pervious land segment. The pervious module has a number of zones where water can be stored or diverted to other zones based on a series of equations. There are 18 subroutines to control the water budget distribution in a pervious zone. The list and relationships of the subroutines are shown in Appendix D. The pervious land segment is represented by four conceptual layers: surface, upper vadose, lower vadose, and groundwater. Different subroutines simulate the processes inside each layer and calculate the water budget in a user-specified time step. Water in the reach comes from precipitation falling on the reach surface and upstream reach, as well as from surface runoff, interflow, and baseflow.

HSPF parameters fall into three categories:

- Prescriptive parameters that set flags and specify algorithms to use.
- Measured or estimated parameters, such as basin area, overland slope, reach length, width, and others.
- Calibrated parameters that may be estimated by measurement, but likely must be adjusted during model calibration. Examples of calibrated parameters are interception storage, upper and lower zone storage, infiltration coefficient, Manning's n, among others.

2.1.3.2 Pervious land modeling principles

The hydrological processes in HSPF are almost identical to the subroutine LANDS in SWM and Watershed Model IV *(Crawford and Linsley, 1966)*. Conceptual layers and movements of water inside the pervious layer are shown in Figure 2.3. Moisture supplied to the land segment includes precipitation (rain, snow) and water from snowpack. Interception storage is formed by water retained above the overland flow plain. Overflow from interception may be added to the total inflow into the surface detention storage. Water from surface detention storage is available for infiltration and runoff. Depression storage can be modeled when necessary. Water from the depression storage does not infiltrate, it can only evaporate. The subsurface budget is modeled in

two vadose layers, which can interact with the surface through plants or interflow. Penetrated moisture may infiltrate to the lower zone, active groundwater storage, or may be lost by deep percolation. Active groundwater eventually reappears as baseflow, and is subject to evapotranspiration (ET). ET is modeled from reach surface, depression storage, interception storage, upper and lower vadose zone, active groundwater, and baseflow. Deep percolation water is considered permanently lost from the simulated system. The aquifer is not modeled, but treated as a sink of water. Lateral external flows to upper zone, lower zone, and active groundwater storage are also possible to model.

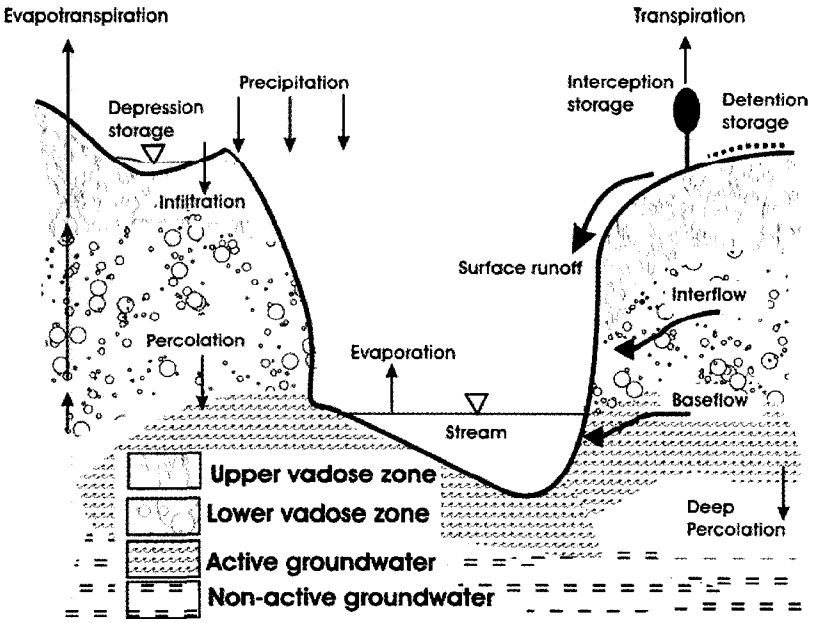


Figure 2.3 Schematic cross section showing conceptual layers and movement of water in HSPF

Water storages of each zone are also simulated. Water holding capacities of upper and lower zones are defined in terms of nominal (not absolute) capacities in order to smooth any abrupt changes if an absolute capacity is reached *(Bicknell et al., 2001)*. Storages also affect ET loss.

A simplified flowchart of HSPF storages (bold font) and fluxes (bold and italic font) is presented in Figure 2.4. The processes in each zone are simulated subsequently and independently. At any time step it is possible to check the volume of water stored in each zone. Circles with numbers indicate location of main subroutines: 1 – ICEPT, 2 – EVICEP, 3 – DIVIZN, 4 – UZINF, 5 – INTFFLW, 6 – PROUTE, 7 – UZONE, 8 – ETUZON, 9 – LZONE, 10 – ETLZONE, 11 – GWATER, 12 – ETAGW, and 13 – ETBASE. Structure and explanation of these subroutines are presented in Appendix D. Some key water budget parameters are shown in Figure 2.4 as well.

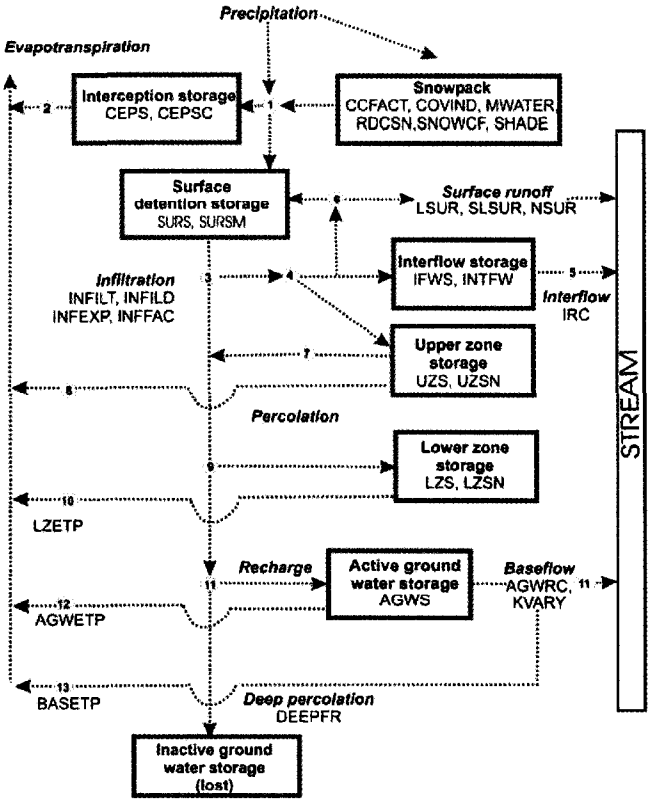


Figure 2.4 Storages and fluxes in HSPF

2.2 Calibration and validation

Two main approaches exist for estimation of model parameters: *a priori* and model calibration. The *a priori* approach estimates model parameters based on measurable characteristics of a watershed using different theoretical or empirical relationships. The calibration approach adjusts parameter values, in order to achieve a close match of model output response and observed data for a historical period of time. The model calibration approach is discussed below.

Model calibration is a critical step in any model application. The American Society of Testing Materials (ASTM) defined calibration as a test of a model with known input and output information that is used to adjust or estimate factors for which data are not available *(ASTM, 1984)*. A more specific definition is proposed by *Refsgaard and Henriksen, (2004)*: calibration is a process of adjustment of parameter values of a model to reproduce the response within the range of accuracy specified in the performance criteria. The definition of calibration can be formulated in different ways, but its essence remains the same – it is solving the problem of model error minimization. For most watershed models, calibration is an iterative procedure of parameter evaluation and refinement. During calibration the observed and simulated values of

interest are compared. Usually calibration enhances understanding of the physical processes occurring in the watershed. Three types of calibration are usually applied in watershed modeling: manual-expert, automatic, and multiple-criteria calibration.

2.2.1 Manual-expert calibration

Manual calibration is still widely used in watershed modeling. Similarity between model outputs and measurement data is evaluated by an expert using objective (different mathematical criteria) and subjective (visual comparison, experience, and intuition) methods. Parameter adjustments are done by a semi-intuitive trial and error process. Although most models have documented guidelines for parameter adjustments, in reality the sequence is varied from expert to expert based on their experience, model understanding, available data, and watershed properties *(Boyle et al., 2000)*. The manual process of calibration typically proceeds in three steps.

Step 1. Expert examines available data, watershed characteristics, model structure, and develops rough *a priori* estimates of a range for each model parameter. The manual for the model and literature review can be used to obtain the first estimation. Sometimes physical reasoning plays an important role.

Step 2. Ranges of parameters become more refined. Specific segments of the hydrograph are analyzed and detailed topography data are considered. A feasible parameter space is the final output of this step.

Step 3. The model is used for output simulation and comparison with the observed data. Due to parameter intercorrelation, adjustment of the model may be a difficult process. Some parameters influence overall model behavior and some parameters are essential only for specific hydrological events. The expert must simultaneously evaluate several subjective and objective criteria while iteratively adjusting model parameters.

Consistent and reliable results in manual calibration are only possible having a broad understanding of the model and the realities of the watershed. Such knowledge is not easily transferable and a great deal of practice and experience with the specific model is needed.

For lumped watershed models with 15 or less parameters the manual approach can give excellent results, but at the expense of considerable time and energy *(Gupta et al., 2003)*. Manual calibration procedures are model-specific and have to be developed for every model. For example, the Interactive Calibration Program (ICP) of the National Weather Service River

Forecast System (NWSRFS) was designed specifically for the Sacramento Soil Moisture Accounting (SAC-SMA) model and HSPFEXP was compiled for HSPF. These calibration procedures are not easily transferable to other models.

In general, manual calibration is strongly dependent on the modeller's knowledge and expertise. Invention of the graphic-user interface has facilitated manual calibration. Still its execution is complicated, labour intensive, and time consuming. However, the overall results of manual calibration may be better.

2.2.2 Automatic calibration

In contrast to manual calibration, automatic calibration is faster and simpler to apply. However, it may give less consistent model performance. Methods of automated calibration are continuously improving usually in parallel with increasing computer power. The goal is to develop calibration procedures that result in the same (or better) model performance compared to performance predictions from a highly trained expert. Subjective human judgments are eliminated. Parameter evaluation and adjustments are objective and explicit rules are employed to calibrate the model. The modeller typically specifies upper and lower limits for each parameter, as well as the initial parameter value. Simple automatic calibration employs a single-criterion approach, which searches in a feasible parameter space to optimize the selected mathematical criterion according to the selected optimization algorithm. Optimization algorithms can vary according to the model and expert preferences. Different types of optimization algorithms are discussed in Section 2.3.

The broad introduction of automatic calibration is promoted by availability of sophisticated distributed and semi-distributed watershed models with dozens of interrelated parameters. An advantage of automatic calibration is that its procedures can be generalized for different models.

2.2.3 Multiple-criteria calibration

Multiple-criteria calibration is an advanced automatic procedure which began to be used in watershed modeling about a decade ago. In multiple-criteria calibration the problem is formulated as an optimization problem and a myriad of optimization methods can be employed to minimize the OF and locate optimal model parameters. Multiple-criteria methods came from the field of economic analysis (*Gupta et al., 2003). Gupta et al. (1998)* suggested using a Pareto-set of trade off solutions instead of a single-set solution. An automatic search in a feasible parameter space finds a set of solutions (the so-called "Pareto optimal") which simultaneously optimizes several user-defined criteria. This results in several viable solutions with different kinds of minimal

errors. The tools which are used for the multiple-criteria calibration are essentially the same as for model optimization: genetic algorithm (GA), SCE-UA, MOCOM-UA, and others.

2.2.4 Validation

Validation is an extension of a calibration procedure. It describes the level of confidence one has in a particular model. The purpose of validation is to ensure that the calibrated model correctly assesses all the variables and conditions that can affect model results. As well, the validation should demonstrate the ability of the model to predict field observations for periods of time separate from the time period used in calibration efforts *(Donigian, 2002)*. Several approaches exist to validate the model. In the present research, the so-called split-sample calibration/validation approach was used. In this approach only a portion of the available data is used for calibration. Once the final parameter values are found through calibration, validation is performed for the remaining period of observed data and goodness-of-fit between recorded and simulated values is reassessed. The model is considered validated when a single parameter set can reasonably represent a wide range of observed data.

Every model has its calibration and validation targets. They are often expressed as the percent mean square error between simulated and observed values. The target values for HSPF suggested by *Donigian (2000)* are presented in Table 2.1. It is worth noting that the presented targets are relevant for monthly and annual mean values. Storm peaks may show larger differences and still be acceptable. Moreover, the level of agreement between simulated and observed values depends on data quality, purpose of the study, available resources, and accessible alternative assessment procedures.

Table 2.1 General calibration and validation targets for HSPF application (adapted from *Donigian, 2000*)

	% difference between simulated and observed values		
	Very Good	Good	Fair
Hydrology/Flow	<10	10 - 15	15 - 25
Sediment	<20	20 - 30	30 - 45
Water Temperature	<7	8 - 12	13 - 18
Water Quality/Nutrients	<15	15 - 25	25 - 35
Pesticides/Metals	<20	20 - 30	30 - 40

It is important not to mix two related terms: validation and verification. Verification is examination of the numerical technique in the computer code to ascertain that it truly represents the conceptual model and there are no inherent numerical problems associated with obtaining a

solution. Verification ensures that computer algorithms and code accurately solve the equations upon which the model is based. In the current research verifying models was not envisaged, assuming this was already done by the model developers.

2.2.5 Calibration of a watershed model

The schematic approach for HSPF calibration which was used in this research is shown in Figure 2.5. Basically, the calibration consists of three main steps: to choose the optimization method, properly define the OF, and select the appropriate evaluation criteria for the results. Each of these steps is discussed in the following chapters.

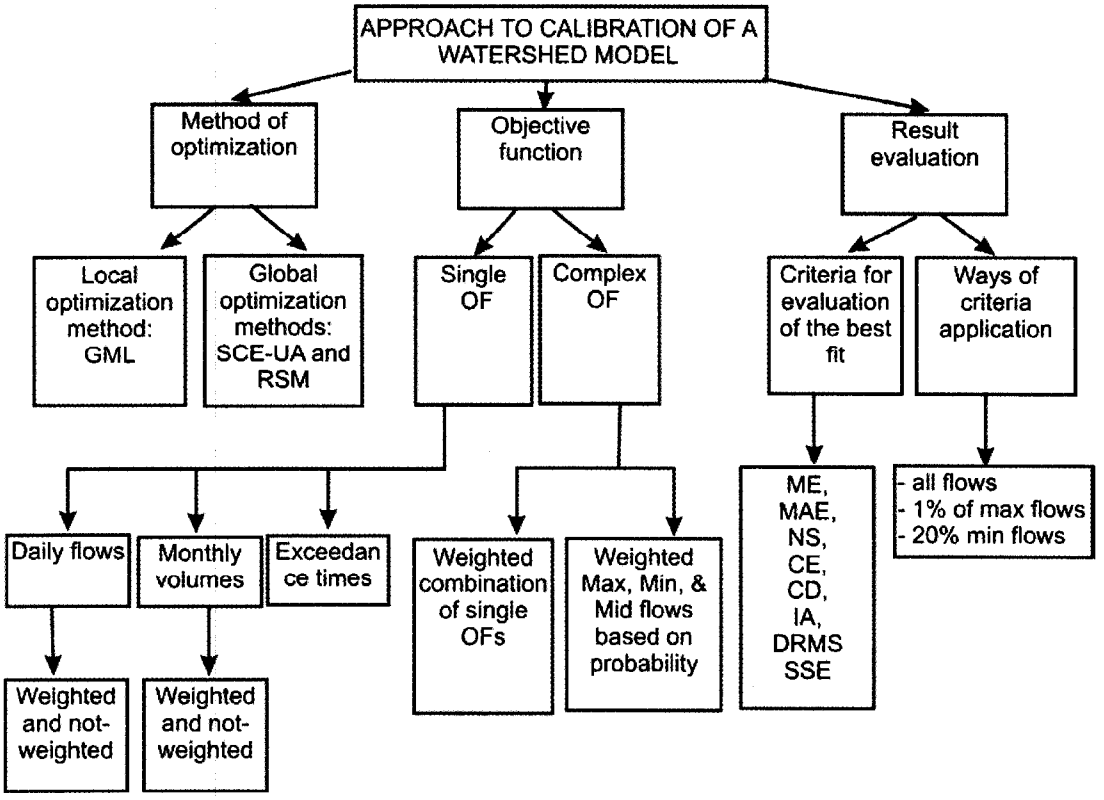


Figure 2.5 Model calibration procedures in the current research

2.3 Optimization methods

The success of model calibration depends heavily on four factors: model structure, calibration data, calibration criteria, and optimization methods *(Duan, 2003)*. The review below focuses on existing and applicable optimization methods for watershed modeling.

2.3.1 Local optimization methods

Local (single extreme) optimization methods, often called deterministic methods, appeared in the late 1960s almost simultaneously with the first watershed model. These methods do not require huge computer power which made them quite popular *(Dawdy and O'Donnell, 1965)*. Historically there are two popular ways of solving a nonlinear unconstrained minimization problem: direct optimization and gradient type methods. Direct optimization methods are derivative-free. They are used when a function is not differentiable or too complicated to compute. Among the popular direct optimization methods are the Axis-Rotating method of *Rosenbrock (1960)*, the Pattern Search method of *Hooke and Jeeves (1961)*, and probably the most widely used is the Simplex Nelder and Mead method *(Nelder and Mead, 1965)*.

Gradient methods are used when an analytical expression of the OF is known and derivatives can be computed. A large family of Newton and Quasi-Newton methods belongs to gradient methods. They require that model equations be continuous to second-order and function values and gradients to be known. Gradient methods usually converge faster than direct methods. However, they may perform poorly when derivative continuity is violated *(Duan, 2003)*. Some applications of gradient type methods for watershed models were not successful *(Hendrickson and Sorooshian, 1988; Pickup, 1977; Duan, 2003)*. This is due to difficulties involved in evaluating derivatives in model equations.

Local optimization methods (LOMs) are very effective for simple, unimodal spaces. In a real-life watershed problem they can be trapped in non-optimum areas, due to the presence of many optima, flat surfaces, ridges, etc. *(Efstratiadis and Koutsoyiannis, 2002)*.

Duan et al. (1992) used two watershed models, SWM and SAC-SMA, for comprehensive analysis of LOMs. They concluded that LOMs are not the best choice for watershed models due to the following reasons: a) the model parameter space contains several regions of attraction into which a search strategy may converge; b) each attraction region has many local minima; c) the OF in a multi-parameter space is not smooth and sometimes not continuous, and d) near the region of a global optimum the parameters often interact nonlinearly. These deficiencies directed much research in the past few years to further study of global optimization methods (GOMs) for watershed modeling. Advances in computer technologies greatly facilitate this process.

Given the arbitrary behavior of the optimized function, there is no general way to verify whether a minimum point is the global minimum or not. One of the proposed strategies *(Gill et al., 1981)* is to run the search algorithm starting from different points as many times as possible.

2.3.2 Global optimization methods

Shortcomings of deterministic local optimization algorithms especially in the presence of multiple regions of attraction, multi-local optima, discontinuous derivatives, insensitivities, and parameter interdependency led to development of probabilistic schemes that use stochastic transition rules and do not impose any special requirement on the nature of the OF. These methods are called global (multi-extreme) optimization methods; they guarantee asymptotic convergence to the global optimum of the function *(Efstratiadis and Koutsoyiannis, 2002)*. There are several GOMs that are applied for watershed model calibration. A review of the most widely used methods is presented in the following sections.

2.3.2.1 Monte-Carlo methods

Probably the simplest GOM is pure random sampling. The parameter space is randomly sampled, choosing the best point as the estimate of the global solution. Several early applications of this method were not efficient and effective *(Ibbitt, 1970; Karnopp, 1963)*. Random sampling methods evolved to the Adaptive Random Search (ARS) method, which focused attention on promising regions *(Masri et al., 1980)*. Via perturbation it generates the next point around the initial one and accepts it only if it improves the OF. Some researchers *(Brasil, 1989)* reported that the ARS method together with proper initial parameter estimates can produce acceptable results. However, some other applications for watershed modeling were not successful *(Armour, 1990; Weinig, 1991; Duan et al., 1992)*. Most studies show that for a successful ARS application the search space should be confined to a narrow range around the optimum.

In this research the random multiple search method (RSM) is applied. RSM performs a random parameter sampling in a user-defined space, then it runs a LOM such as Gauss-Marquardt-Levenberg method (GML), records the value of the OF for each parameter set, and ranks them. The parameter set corresponding to the lowest OF is used for model optimization. After its completion RSM picks another random parameter set which is as far as possible from the trajectory of the previous parameters and performs optimization again. This is done to minimize the probability of finding the same minimum. In the next step RSM tests the parameter set which is maximally distant from all previous parameter trajectories *(Doherty, 2004)*. Termination

criteria such as maximum number of inversions runs, number of runs with no improvements, and the OF improvement judged to be negligible are supplied by the modeller. The strength of this method is in introduction of a certain degree of randomness into the optimization. Randomness ensures that all "regions of attractions" in a parameter space are found and checked during the optimization process. Although RSM needs more computer power and is significantly lengthier than any LOM, it often yields superior calibration.

2.3.2.2 Evolutionary algorithms

Evolutionary algorithms (EAs) brought some important modifications to RSMs. The mechanism of EA was inspired from natural evolution and based on principles of genetics. Searching procedures in EAs are implemented at stages called generations. At each stage the population of randomly generated points evolves by applying selection, recombination, and mutation operators. The most common EA is Genetic algotithm (GA) (*Schwefel, 1981*). GA was invented in the late 1970s by *Holland (1975)* and significantly improved by *Goldberg (1989)*. GA was first applied for watershed calibration by *Wang (1991)*. He applied GA to the Xinganjiang watershed model and reported consistent excellent results in 10 random trials for all 7 parameters. Many other scientists reported that GA is an efficient and robust method for calibrating watershed models *(Duan et al., 1992; Franchini and Galeati, 1997; Savik et al., 1999; Agrawal and Singh, 2003).*

The main operators of elementary GA are reproduction, selection, crossover, and mutation. The key algorithmic parameters are population size, crossover, mutation probability and number of generations. An important step in GA is the coding of the model variables. Usually the variables are transformed to a binary string of specific length where each string is called a chromosome. Then variables must be discretized and approximated by integers. The precision of the optimal solution depends on the encoded bit length of integers *(Duan, 2003)*.

2.3.2.3 Simulated annealing

The simulated annealing (SA) method was first proposed by *Metropolis et al. (1953)*. The method is based on thermodynamic principles employed in metallurgy. Molten metal is cooled slowly with irregular reheating to allow a stable crystal system to develop. The system sometimes goes uphill as well as downhill to provide a chance to escape a local energy minimum in favour of finding the global minimum. Thus, SA is a kind of random walk that migrates through a number of local minima and eventually converges to the global minimum *(Duan, 2003)*. As other random

search algorithms, SA is criticized for being slow. SA becomes most efficient when coupled with local search methods *(Desai and Patil, 1996; Thyer et al., 1999)*.

2.3.2.4 Shuffled complex evolution method

The SCE-UA was developed at the University of Arizona as part of a PhD thesis by Duan *(Duan et al., 1992)*. SCE-UA was specially designed for watershed models and it quickly became popular. The method is based on synthesis of GA, the simplex downhill method, and the control random search method plus a newly introduced concept – partition of complex shuffling.

The procedure for SCE-UA consists of several steps.

- Sample s parameter sets with random values and compute the OF for each parameter set. Rank the sets in order of increasing OF.
- Partition s parameter sets into p complexes, with m points in each. The first complex contains $p \times (k-1) +1$ ranked parameter sets, the second complex contains $p \times (k-1) +2$ sets and so on, where $k = 1, 2, 3\ldots, m$.
- Each complex evolves according to the *Nelder and Mead (1965)* simplex downhill search method. The worst parameter sets are improved by reflection, contraction, and expansion rules.
- Shuffling stage. The evolved parameter sets in each complex are combined into a single sample population. It is sorted according to the value of the OF and re-partitioned into p complexes according to the rule in Step 2. Then, the convergence criteria are checked. If they are not met, the process is repeated.

Some researchers have applied SCE-UA for watershed calibration. In many cases this method is superior to other GOMs *(Gan and Biftu, 1996; Kuczera, 1997; Franchini et al., 1998)*. SCE-UA is a robust, effective, and efficient global optimization algorithm combining the power of deterministic and probabilistic methods.

2.3.2.5 MOCOM-UA

Among the methods that employ multi-criteria optimization is the recently developed multi-objective complex evolution (MOCOM-UA) algorithm *(Yapo et al., 1998)* which is a general purpose GOM capable of optimizing a model simultaneously with respect to different OFs in a single run. MOCOM-UA has evolved from SCE-UA and its procedures are based on the

following three concepts: (1) population, (2) rank-based selection, and (3) competitive evolution. A population of possible solutions is evolved in the feasible space to search for the Pareto set. Ranking of the population is accomplished by successively placing all points on different Pareto fronts. Competitive evolution consists of selecting points based on their rank and moving the worst points toward the Pareto set using the newly developed multi-objective simplex procedures. More details on the MOCOM-UA algorithm can be found in *Gupta et al. (1999)* and *Bastidas et al. (1999)*. MOCOM-UA application for real watersheds is limited.

2.3.2.6 Bayesian

Conventional batch calibration methods assume time-invariant model parameter values and require a considerable amount of data, typically covering 8-10 years. Bayesian methods are often used when historical data are insufficient. The procedure was proposed by *Moskus (1989)*. Bayesian methods allow forecasts to be generated soon after the first observation values are available. Among the simple recursive methods developed by *Beven and Binley (1992)* is the generalized likelihood uncertainty estimation (GLUE) method. GLUE is a simple Bayesian estimation that uses MC simulation based on sampling parameter sets from assumed prior probability distributions. The method is based on estimation of the probabilistic weights associated with different parameter sets using user-selected goodness of fit criteria and the derivation of a probability distribution. Then, the obtained distribution is used to derive the predictive probability of the output variables *(Baysal, 2003)*. Although Bayesian methods are attractive theoretically, they are not widely used in watershed modeling due to complicated algorithmic realization *(Törn and Zilinskas, 1987)* and subjectivity in evaluating parameter probabilities *(Baysal, 2003)*.

2.3.3 Optimization software

There are a number of optimization software packages used for calibration and parameter estimation in watershed modeling. Among the most popular are PEST, MCAT, GLOBE, MODINV, MODFLOP, AutoCal, and MOUSE. The discussion below highlights several of them. Information about software developers and market availability is presented in Table 2.2.

MCAT (Monte-Carlo analysis toolbox) is a collection of MATLAB analysis and visualization functions integrated through a graphical user interface. The toolbox can be used to analyze the results from Monte-Carlo methods or from model optimization methods (e.g., GA). A number of powerful techniques are included in the toolbox to investigate the structure, parameter sensitivity,

and output uncertainty of hydrological models. They include an extension of the regional sensitivity analysis method, various components of the GLUE method, options for the use of multiple-objective model assessment, and response surface plots *(Wagener et al., 2001)*.

Table 2.2 Optimization software for watershed models

Model name	Developer	Code	Availability
MCAT – Monte-Carlo analysis toolbox (2001)	T. Wagener, Imperial College of Science, Technology and Medicine London, UK	Matlab	Free download at http://ewre-www.cv.ic.ac.uk/software/toolkit.htm
PEST – Parameter estimation tool, Version 10.0 (2006)	J. Doherty, Watermark Computing, Corinda, Australia	Fortran	Free download at http://www.sspa.com/PEST/download.html, Visual Pest – 590 USD at http://www.scientificsoftwaregroup.com
GLOBE – Global optimization tool, version 1.5 (1999)	D.P. Solomatine, IHE Delft, Delft, The Netherlands	Fortran	499 USD, available at http://www.xs4all.nl/~dpsol/data-machine/order.htm

PEST (Parameter estimation tool) is a nonlinear parameter estimation package for data interpretation and model calibration. PEST communicates with a model through the model's own input and output files. PEST implements a particularly robust variant of the GML method of nonlinear parameter estimation. Individual model parameters can be designated as adjustable, fixed, or linked to other parameters. Adjustable parameters can be log-transformed to increase optimization efficiency. Optimum parameter values can be constrained to lie between individually-specified upper and lower bounds. PEST calculates statistical data for optimized parameter values including 95% confidence intervals together with the parameter covariance and correlation-coefficient matrices and many other statistics. PEST includes the SCE-UA method for parameter estimation *(Doherty and Johnston, 2003)*.

GLOBE (Global Optimization Tool) is a global optimization program. The modeller supplies files with the error value generated by the hydrological model and GLOBE using different optimization techniques calculates optimized parameter values. In the last version of GLOBE the following algorithms are included: controlled random search, GA, multistart algorithm, and adaptive cluster covering with local search *(Solomatine et al., 1999)*.

The available optimization software was compared for its functionality, effectiveness, and cost. Based on research conducted and review of the recently published papers, it was concluded that PEST, at present, is a superior optimization tool. It is free, regularly updated, and supported by

USEPA. PEST was specially designed for environmental problems. Some of its components are specifically made for watershed models.

2.4 Mathematics of parameter optimization

The relationship between parameters and model-simulated observations in a watershed model can be described as

$$C^o = X \times b^o \tag{2.1}$$

where C^o – model simulated output, b^o – vector which holds system parameters, and X – matrix which describes the system, it is constant and independent of b^o.

In matrix form Eq. 2.1 can be written as

$$\begin{Bmatrix} C_1^o \\ C_2^o \\ C_3^o \\ C_4^o \\ \dots \\ C_n^o \end{Bmatrix} = \begin{bmatrix} f_1(x_1) & f_1(x_2) & f_1(x_3) & \dots & f_1(x_m) \\ f_2(x_1) & f_2(x_2) & f_2(x_3) & \dots & f_2(x_m) \\ f_3(x_1) & f_3(x_2) & f_3(x_3) & \dots & f_3(x_m) \\ f_4(x_1) & f_4(x_2) & f_4(x_3) & \dots & f_4(x_m) \\ \dots & \dots & \dots & \dots & \dots \\ f_n(x_1) & f_n(x_2) & f_n(x_3) & \dots & f_n(x_m) \end{bmatrix} \times \begin{Bmatrix} b_1^o & b_2^o & b_3^o & b_4^o & \dots & b_m^o \end{Bmatrix} \tag{2.2}$$

where $C^o_1 \dots C^o_n$ are model generated outputs, $b^o_1 \dots b^o_m$ is the parameter vector of order m, n is the number of observations, m – number of parameters, and $f_1(x_1) \dots f_n(x_m)$ is the matrix which describes the system.

Because the model is nonlinear, the relationship 2.2 has to be linearised. Often Taylor's approximation is used. A new set of observations, C, is generated when parameter vector b is slightly different from b^o.

$$C = C^0 + J(b - b^o) \tag{2.3}$$

where J is a Jacobian matrix of X, which is comprised of n rows and m columns. Each element of the matrix is a derivative of one particular observation with respect to each of the m parameters.

The proximity of simulated and observed values is measured by their squared residuals. Weighted square residuals are often called the OF (Φ). During calibration the OF should be minimized to the lowest possible value. From Eq. 2.3 the OF is defined as

$$\Phi = \left[C - C^0 - J \cdot \left(b - b^0\right)\right]^t \left[C - C^0 - J \cdot \left(b - b^0\right)\right] \tag{2.4}$$

where t – indicates matrix transpose operation.

From Eq. 2.4 it is possible to find a parameter upgrade vector that minimizes the residual sum of squares (Φ). For that purpose the derivative is set to 0. Denoting upgrade vector ($b - b^o$) as u and the vector of residuals ($C - C^o$) as r, Eq. 2.4 becomes

$$\begin{aligned}
&\Phi = (r - Ju)^t \cdot (r - Ju) \\
&\Phi = r^t r - r^t Ju - J^t u^t r + J^t u^t Ju \\
&0 = -\frac{d}{du} r^t Ju - \frac{d}{du} J^t u^t r + Ju \frac{d}{du} J^t u^t + J^t u^t \frac{d}{du} Ju \\
&0 = -r^t J - Jr^t + JJ^t u^t + JJ^t u^t \\
&u = J^t r \cdot (J^t J)^{-1}
\end{aligned} \tag{2.5}$$

Eq. 2.5 is the classical formula for the least squares estimator in matrix notation.

Not all observations have equal importance for the parameter estimation process. Some measurements may be more prone to errors than others or the magnitude of different observations can vary. In order to minimize this problem a weighting factor has to be applied to each observation. The weighting factors are represented by a square, diagonal matrix W. Squared weight w_i is attached to the ith observation.

$$W = \begin{Bmatrix} w_1^2 & 0 & 0 & 0 & 0 \\ 0 & w_2^2 & 0 & 0 & 0 \\ 0 & 0 & w_3^2 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & w_m^2 \end{Bmatrix} \tag{2.6}$$

The modified Eq. 2.4 is now

$$\Phi = [C - C^o - J \cdot (b - b^o)]^t \cdot W \cdot [C - C^o - J \cdot (b - b^o)] \tag{2.7}$$

and the modified expression for the parameter upgrade vector is

$$u = J^t \cdot W \cdot r \cdot (J^t \cdot W \cdot J)^{-1} \tag{2.8}$$

where W – is a weighting factor matrix.

Another, simpler way of expressing the OF is

$$\Phi = \sum_{i=1}^{m} (W_i \cdot r_i)^2 \tag{2.9}$$

where r_i is the i^{th} residual expressed as the difference between simulated and observed values.

Calibration is started with the initial parameter set b^o and then the update vector is calculated and hopefully the global minimum of the OF will be found iteratively. A good initial parameter set improves stability of the optimization and reduces the number of iterations required to determine the optimal parameter set. Iterative improvement of initial parameter values towards the minimum of the OF is shown in Figure 2.6.

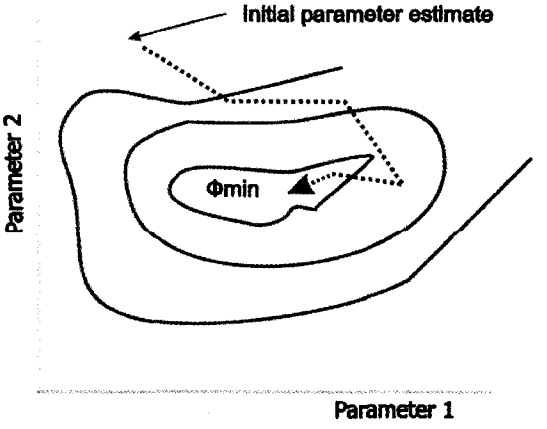


Figure 2.6 Contours of the OF and path of iterative improvement of initial parameter value towards the global minimum

Equation 2.3 is a classical simple gradient descent method. It does not perform well when the function is complex, especially when it has a long narrow valley. The Gauss-Newton method added a second derivative to the denominator to make finer steps on approaching the minimum. Improvements to the Gauss-Newton method were suggested by *Levenberg (1944)*, who proposed the following update rule

$$b = b^o + (H + \lambda \cdot I)^{-1} \cdot J \cdot (C - C^o) \tag{2.10}$$

where H – Hessian matrix evaluated at b^o, λ – damping factor which is adjusted during each iteration, I – identity matrix.

The initial λ is chosen somewhere between 1 and 10. If the error decreases, the iteration is accepted and λ is reduced for the next iteration by a user-supplied factor. This action reduces the

influence of gradient descent. On other hand, if error increases, the step is retracted and λ is increased.

The disadvantage of the Levenberg method is that if λ is large the Hessian matrix is not used. *Marquardt (1963)* replaced the identity matrix in Eq. 2.10 with the diagonal of the Hessian resulting in the Marquardt-Levenberg update rule.

$$b = b^o + (H + \lambda \cdot diogonal[H])^{-1} \cdot J^T (C - C^o) \tag{2.11}$$

Since the Hessian matrix is proportional to the curvature of the function, a larger step is taken when the curvature is low (function is flat), and oppositely, when the function has a steep incline, the step for iteration is small. Often, for computational simplification there is no need to compute the full Hessian. The approximation $H \approx J^T \cdot J$ works well, especially when the residuals are small *(Doherty, 2004)*. Denoting u as the parameter upgrade vector and r as the vector of residuals and introducing weighting matrix W, Eq. 2.11 becomes

$$u = (J^T \cdot W \cdot J + \lambda \cdot diogonal[H])^{-1} \cdot J^T \cdot W \cdot r \tag{2.12}$$

During model calibration the parameter upgrade vector u is calculated for each iteration with the aim to minimize the vector of residuals r. Equation 2.12 represents the GML method which was used for parameter optimization in this research. As seen from the equation, the GML algorithm requires computation of the Jacobian matrix and matrix inversion at each iteration step as many times as there are adjustable parameters. Matrix inversion is a significant disadvantage. When the matrix is singular, GML still performs due to the Marquardt lambda; however, the risk that optimization does not reach the global minimum is high *(Doherty, 2004)*. A matrix becomes singular when columns (or rows) are linearly dependent.

2.5 Uncertainty analysis

Uncertainty refers to lack of knowledge about specific factors, parameters, or models. Uncertainty analysis aims to quantify the level of confidence the modeller has in inputs, parameters, model structure, and as a result, in model outcomes. A number of uncertainty classifications are proposed in the literature *(USEPA, 1999; Yen, 2002; Willems, 2000)*. Although they have the same components, their groupings are different. In this study three main types of uncertainty are differentiated: model uncertainty, input data uncertainty, and parameter uncertainty. They are based on the different physical natures of the uncertainties.

2.5.1 Model uncertainty

Model uncertainty represents model structure with equations, assumptions, and the modeller's vision of reality. Since any model is a simplification of the physical world, uncertainty associated with the model structure is unavoidable. This uncertainty does not vary in time since the model equations do not change. An approach to estimate model uncertainty may consist of multiple model applications for various watersheds and comparisons with other watershed models. Model uncertainty can be considered for an individual subroutine or it can be described in a lumped approach as a total model uncertainty of the output variables. An interesting example of model uncertainty decomposition is discussed by *Timbe and Willems (2004)*.

2.5.2 Input data uncertainty

Differences between estimated and actual input data values are usually described as uncertainty. Model input contains directly measured or estimated data which vary for different watersheds. Input data uncertainty can be caused by temporal and spatial variability in natural processes or by measurement, sampling, and estimation errors.

Some input data are temporally constant but spatially variable (topography, landuse, river bed geometry, etc.). During the period of modelling their values are constant from year to year for a particular cell grid. Some input data can be temporally variable but spatially constant (precipitation, solar radiation, and pan evaporation). Other input data are spatially and temporally variable (infiltration, upper and lower zone storage moisture, interception storage, etc.).

Almost all factors and processes affecting flow routing in a watershed model are subject to natural stochastic variability. There is always uncertainty in the description of channel geometry, roughness, specification of sources and sinks, and initial and boundary conditions *(Singh, 2004)*. Stochastic variability is caused by heterogeneities, nonuniformities, random irregularities, and different types of errors in input data. Input data uncertainty can be quite high. For example, *Willems (2000)* reported that spatial variation in rainfall accounts for approximately 30% of the overall variability in model results. Natural variability of data is a stochastic component and cannot be reduced but should be accounted for. Measurement and sampling errors exist in every watershed model and depend on quality of the data. For a realistic estimate of the input uncertainty the modeller should know the instruments used for the measurements, the sampling methodology, sampling frequency, how and where samples were stored before analysis, and other similar information.

2.5.3 Parameter uncertainty

Different coefficients and exponents which need adjustment during model calibration are often referred to as model parameters. Model performance is highly sensitive to some parameters and relatively insensitive to others. Although a margin between model input and model parameter is not always visible, in general, model parameters are related to model characteristics that are adjustable by the modeller during calibration *(Duan et al., 2003)*. Estimates of parameters are generally uncertain because data used for calibration are uncertain and the model never exactly represents the natural system. Parameter uncertainty can be caused by correlated and insensitive parameters, poor initial and boundary parameter values, and inadequate optimization methods.

2.5.4 Expression of uncertainty

There are three common ways to express uncertainty: a single value, an interval, or a probability distribution. When a single value is used the second statistical moment (variance or standard deviation) as a measure of dispersion of the variable is often a choice. Variance is easy to manipulate mathematically; however, comparison of it with the observations is difficult due to its squared units. Standard deviation is expressed in the same units as the studied characteristic. The coefficient of variation is a convenient normalized measure for comparison of uncertainties of different variables. By themselves, point estimates do not fully portray the variability of the system and usually more sophisticated expressions are needed.

Confidence and prediction intervals are convenient measures to convey uncertainty. They show an explicit degree of confidence in a specified interval of variability. Confidence intervals are usually associated with a mean value and prediction intervals with an individual value *(Kirchner, 2001)*. The method of intervals has potential problems which should be accounted for. The first problem is most of the conventional procedures for determination of confidence intervals assume a normal distribution which is not always the case, especially when sample size is small. The second problem is confidence intervals of individual random components cannot be directly and easily combined to estimate the confidence interval of the system as a whole *(Yen, 2002)*.

The PDF is the most complete and ideal description of uncertainty. The PDF is described by skewness, shape of the distribution, characteristics of tails, and some other characteristics. Unfortunately, in most practical model applications the PDF is difficult to find precisely *(Yen, 2002)*. A large number of measurements are necessary to approximate any one of the well-known distributions.

2.5.5 Distributions in uncertainty analysis

The basic assumption of uncertainty analysis is that the uncertainty in the value of interest is equal to the uncertainty of error or deviation in its value. Another assumption is that measurement error and parameter deviation in a calibration process are random variables that follow a statistical distribution *(Castrup, 2002)*.

The distribution of measurement errors and deviation of parameter values during the calibration can be assumed to be uniform, log-normal, triangular, normal, etc. (Figure 2.7). In a uniform distribution, the probability of finding a value between the minimum and maximum boundaries is unchanged and equal to unity. A uniform distribution is easy to use and analyze and it may be appropriate for some limited applications. The log-normal distribution can be used to estimate uncertainty in parameters with asymmetric limits or with asymmetric tolerance, for example, when one of the physical limits lies close to the nominal value. Uncertainty estimates for the log-normal distribution are obtained by numerical iterations *(Castrup, 2002)*. The triangular distribution can be helpful when a central tendency exists and the containment probability is 100%. However, abrupt transitions to zero at the boundary limits are often physically unrealistic. Moreover, the linear increase and decrease in behaviour is not suitable for the majority of parameters. In some cases, the quadratic distribution that exhibits a central tendency without discontinuities is used. The Student's t normal distribution, which is based on degrees of freedom, may be used as well.

The most widely used distribution for uncertainty analysis is the normal distribution. Reasons for this are firstly, this distribution is the best approximation of many physical processes in the universe *(Castrup, 2002)* and secondly, the central limit theorem proves that even though individual random errors or deviations may not be normally distributed, the combined error is approximately normal *(Clarke, 1994)*. The normal distribution is a quite common assumption for uncertainty analysis *(Morgan, 1999; Chang et al., 1994; Castrup, 2002)*.

2.5.6 Methods for uncertainty estimation

Methods for uncertainty estimation vary by level of sophistication, computational complexity, and data requirements. The choice of method is dictated by available information, required precision, and accessible computer resources. In the ideal case, uncertainty can be characterized by an exact PDF of model outcomes as a function of the stochastic input parameters. However, in practical watershed applications due to nonlinearity and complexity, the analytical derivation of a

PDF is not often feasible. As an alternative, a number of methods attempt to approximate the statistical properties of uncertain model output. Detailed description of methods for uncertainty analysis is presented in Section 4.2.4.

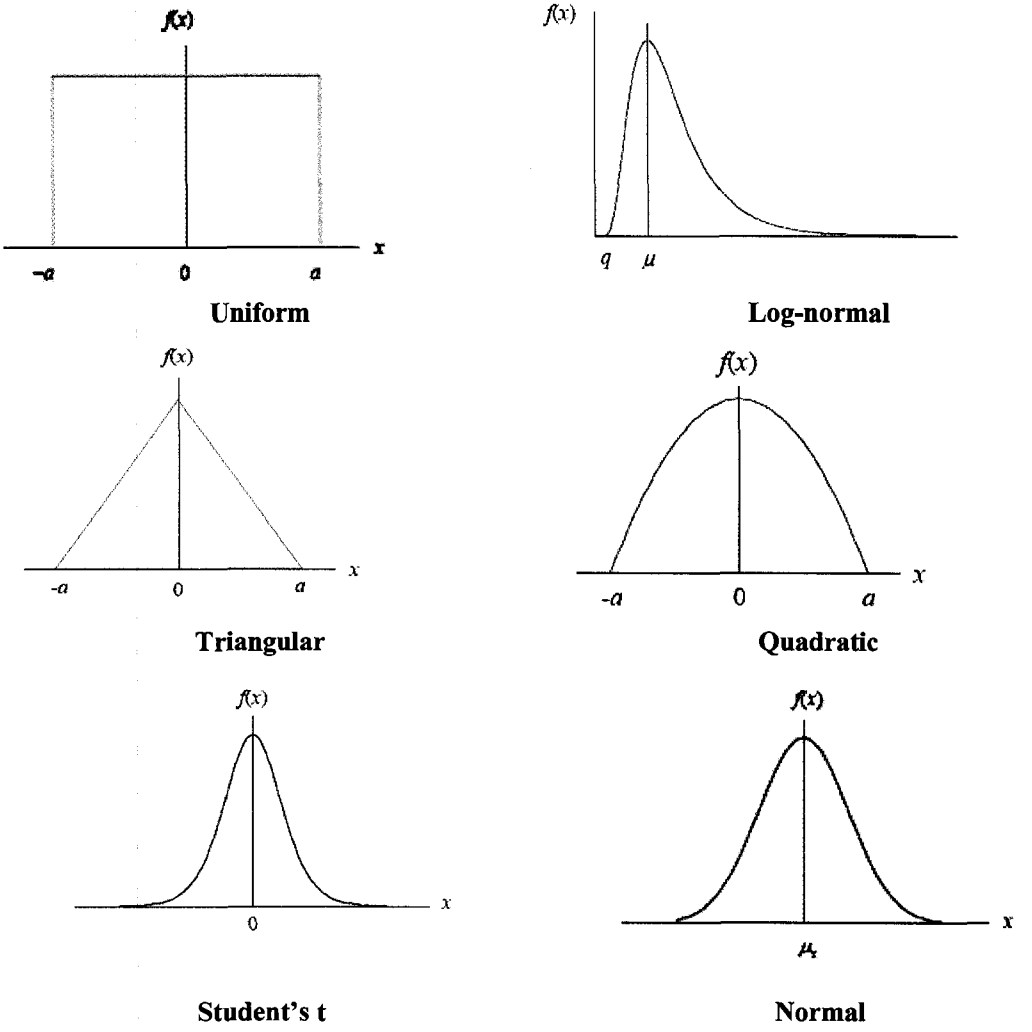


Figure 2.7 Typical probability distributions for uncertainty analysis

2.6 Spatial scaling and temporal resolution

2.6.1 Scaling in nature

It was noticed long ago that processes behave differently at different scales. However, recently it was found that there are orderly ways, scaling laws, which connect various phenomena in biology, ecology, hydrology, and other sciences. The scaling exponents in biology and ecology are often simple fractions based on underlying mathematical and physical principles. One example is the well-known quarter-power scaling law in biology. Comparing the mass of a

mouse, a squirrel, and an elephant (30 grams, 1 kilogram, and 5 tonnes, respectively) with their pulse rate (600, 250, and 30 beats per minute, respectively), one can notice that the pulse rate scales inversely as the quarter power of the animal's mass *(Klarreich, 2005)*. An animal's metabolic rate appears to be proportional to its mass raised to the 3/4 power, which holds from microbes to whales, creatures ranging over 21 orders of magnitude in size *(Savage et al., 2004)*. In a mature forest, the average distance between trees of the same mass as well as trunk diameter follow a quarter-power scaling law *(West et al., 1997)*. Recent advances in physics showed that all transportation networks have the same general behaviour if they have achieved maximum efficiency through natural selection. Circulatory and respiratory networks in plants and animals, stream networks in a watershed, and networks of pipes transporting water to homes have been under close investigation and it appears that they exhibit the same general behaviour. It was proved that allometric scaling exists. The stationary solutions of the derived dynamic equation for the erosion in river networks have allometric scaling behaviour that was confirmed with observational data *(Banavar et al., 1997)*.

Based on the research results in biology and physics, it is logical to hypothesize that some scaling equations should exist in physically-based model parameters when upscaling or downscaling a well-calibrated model.

2.6.2 Scaling approaches in watershed modeling

Although hydrological processes at larger scales are the resultant of smaller scale processes, models representing these processes vary considerably. One of the major objectives in watershed modeling is to understand the impact of changing scales on hydrological processes and model parameters. The importance of scale has long been recognized by water resources engineers; however, many questions related to modeling scale and model parameters remain unresolved. Do the mathematical equations developed and valididated for laboratory scale apply at the watershed scale? Where are the upper and lower limits of applicability for a watershed model? What approaches does the model use for spatial and temporal scaling? Obviously, processes interact at different scales; however, "no one discipline has completely resolved the problem of translation of knowledge from lower scales to higher, or the reverse" *(Wagenet, 1998)*.

The suggestion to test similarity of hydrological parameters on different scales came from soil physics *(Miller, 1980)* and now it is enthusiastically being explored by some watershed modellers. Two fundamental hypotheses underlie this research. The first is that larger basins are more linear (governed by simpler laws) in their response than smaller basins *(Amorocho, 1963)*.

Another vantage point on this stems from input data often being absent or of a much lower quality at a larger scale, which results in a tendency to use simpler, empirical models at the larger scale *(Heuvelink, 1998)*. The second hypothesis is that different processes dominate at different scales and as a result, different questions arise at different scales. Also, it is important to remember that different processes are ignored in the simplification steps of model development.

Scaling is a huge problem in hydrologic modeling and it will likely remain a problem for a long time. The main reasons for this are:

- the governing equations in the code are difficult to solve effectively for different scales,
- temporal and spatial lumping cause uncertainty which is difficult to estimate,
- resolution of input data often needs to be different at various scales, and
- scaling issues are likely to influence the search for the global optimum in model calibration.

As a final point, watershed models are often calibrated and validated where observed data are available. However, the models are also needed in ungaged basins where calibration is impossible. Understanding spatial and temporal scaling properties of the studied characteristic (flow, snow depth, water temperature, etc.) and model parameters is needed if such models are to be used for ungaged areas.

2.6.3 Size of watershed

Hydrological modeling can be performed on a variety of scales ranging from point scale (<10 m^2) to global scale ($>1{,}000{,}000$ km^2). Hydrological processes are often studied and modeled at a very low scale. As a result, many hydrological models were originated and tested at the hill-slope scale. Examples are TOPMODEL *(Beven, 1995)* and MIKE SHE *(Bathurst, 1986)*. The problem with these models is the methodology to transfer a field measurement or governing parameter to a watershed scale. Various aggregating and upscaling approaches exist, the main principle of which is to perform another calibration when the scale is changed *(Vazquez et al., 2002)*. Some other physically-based models such as HSPF and SWAT are developed and tested explicitly for the watershed scale. Thus, the equations and parameters generally are representative at watershed scale and not necessarily at point scale. For example, the simple conceptual Swedish hydrological model (HBV) is reported to be stable over a wide range of scale from several tens of thousands of kilometers to the continental scale of the Baltic Sea *(Bergstrom and Graham, 1998)*.

Intuitively, physically-based models suggest that finer grid (landuse segment in HSPF, hydrological response unit in SWAT) size and time steps result in better representation of a watershed. However, owing to computation limitations, data availability, and model structure these models still can only be applied using effective grid (or segment) parameters. *Vazquez et al. (2002)* found that the effective grid resolution for achieving reasonably accurate prediction in MIKE-SHE is 600 m. Effective grid resolution for HSPF is not yet studied.

More research is needed to examine interrelation (from a modeling perspective) between large and small watersheds. Understanding model behaviour at different scales opens the possibility of transferring model parameters to neighbouring ungaged watersheds. The scale-dependency of the physically-based parameters was explored in this study.

2.6.4 Digital elevation model

Topography effectively controls the distribution of hydrological processes over the whole range of scales from hill slope to continental basins. The topographic configuration of the landscape is one of several controlling boundary conditions for watershed modeling. Recent development of remote sensing provides accurate topographic description of practically any area of the world. Topographic products currently are available at various levels of accuracy and resolution.

One of the ways to represent topography is a digitized array of elevation values called a digital elevation model (DEM). DEMs are usually visualized by means of GIS. DEM is a viable alternative to traditional field surveys and manual analysis of topographic maps *(Moore et al., 1991)*. DEMs are usually stored in one of three data structures: grid data, triangulated irregular network (TIN), and contour-based structure *(DeBarry et al., 1999)*. Grid structure consists of a square matrix (called a pixel) with the assigned elevation. TIN is a continuous surface generated from variable size interconnected triangles with known elevation. Smaller triangles are used in areas with rapidly changing topography and larger triangles are used in relatively smooth topography. The contour-based or digital line graph (DLG) consists of digitized contour lines of specified elevation *(DeBarry et al., 1999)*. Each DEM structure has its advantages and disadvantages *(Moore et al., 1991)*. In this research square-grid DEMs were used due to their simplicity, processing ease, and computational efficiency.

For watershed modeling two aspects are important for DEM selection: quality and resolution of DEM data. Quality refers to the accuracy of elevation data, and resolution refers to the horizontal and vertical grid spacing. Quality and resolution must be consistent with the study objective, scale

of studied processes, and with model limitations. Research is needed to assess the impact of accuracy limitations and low resolution of DEM data on modeling results. Although some work has been done *(Wolock and Price, 1994; Zhang and Montgomery, 1994)* many issues are currently unresolved.

Often the simulation is done with the most detailed DEM available; the question of how it impacts the slope, aspect, flow length, drainage divide, and channel network in a watershed model remains open. In the current research the question is answered. Furthermore, the impact of parameters individually and altogether on model simulations, in particular, maximum and minimum flows, is addressed.

One approach to study impact of DEM on model response is to explore the effect of spatial aggregation in a watershed with detailed data. DEM aggregation obviously leads to coarsening of terrain attributes (elevation, slope, and curvature). *Kuo et al. (1999)* highlighted the variation of curvature with the grid size as an important control on hydrological prediction. As grid size increases, details of lateral flow pathways are lost, with subsequent effect on the transport and accumulation of water across the watershed. Since watersheds vary in topography, it is expected that the aggregation effect is site specific and related to terrain roughness and drainage density *(Wolock and McCabe, 2000)*.

Interesting results for the Penn state runoff model (PSRM) were obtained by *Seybert (1996)*. He concluded that volume estimates in the PSRM model are less sensitive to spatial data resolution than peak flow estimates. Also, a ratio of average watershed area to grid cell area of about 100 was found to be an acceptable threshold of spatial resolution for reasonable model results.

Vivoni et al. (2005) studied the impact of DEM on the hydrograph prediction and spatial distribution of surface-subsurface watershed response in MIKE-SHE. He concluded that over a broad range of DEM resolution (30 to 90 m) many processes exhibit weak resolution dependence; total runoff and ET are approximately constant. Their results further suggested that the critical DEM size is about 100 m and it is related to the mean hill slope length of the studied watershed. An analogous study for HSPF has not yet been conducted.

2.6.5 Spatial variability in HSPF

Spatial variability in a watershed is treated differently in various watershed models. In some large-scale watershed models, such as HSPF, SWAT, etc., the concept of effective parameters is used. A uniform value of effective parameter is assigned to a heterogeneous area *(Bloschl and*

Sivapalan, 1995). Effective parameters are usually estimated from inverse modeling *(Schoups et al., 2005)*. The value of the effective parameter may not exist at point scale; nevertheless, such an approach may still produce good results for some applications.

An example of an effective parameter is infiltration in HSPF. Spatially averaged infiltration capacity over the land segment (Figure 2.8) is calculated according to the soil moisture ratio in the lower zone *(Bicknell et al., 2001)*. Point **A** corresponds to the spatial mean infiltration capacity and point **B** to the spatial mean infiltration and interflow capacity of the land segment.

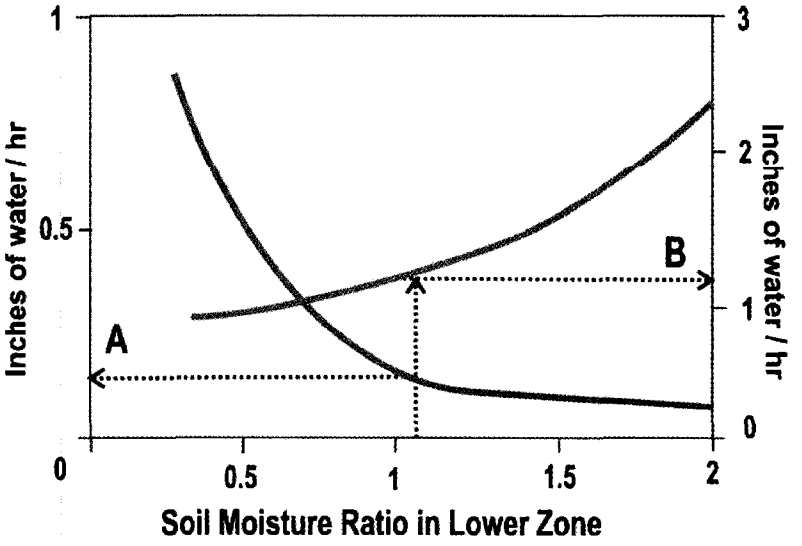


Figure 2.8 Spatial mean infiltration capacity (A) and spatial mean infiltration and interflow capacity (B) of the land segment as function of a soil moisture ratio in the lower zone *(adapted from Bicknell et al., 2001)*

Then, in HSPF, infiltration capacity is assumed to vary linearly based on a PDF. Infiltration capacity varies in time to represent changes in soil moisture in the unsaturated zone (Figure 2.9). LINE I separates the water available for potential direct runoff and infiltration. LINE I is parallel to the x-axis when there is no spatial variation over the subwatershed, i.e., the infiltration capacity of soil is the same at any point along the land segment. The larger the angle of LINE I, the higher spatial variability exists in the subwatershed. LINE II divides potential direct runoff into interflow and surface runoff. The proposed scheme was suggested in order to take into consideration that some parts of the subwatershed have very low infiltration capacity (point A_{MIN}) and some very high (point A_{MAX}). The area of the shaded trapezoid is equal to total volume of infiltrated water from the land segment in a specified time step, and it is basically a function of current soil-moisture ratio in the lower zone and the moisture supply rate *(Bicknell et al., 2001)*.

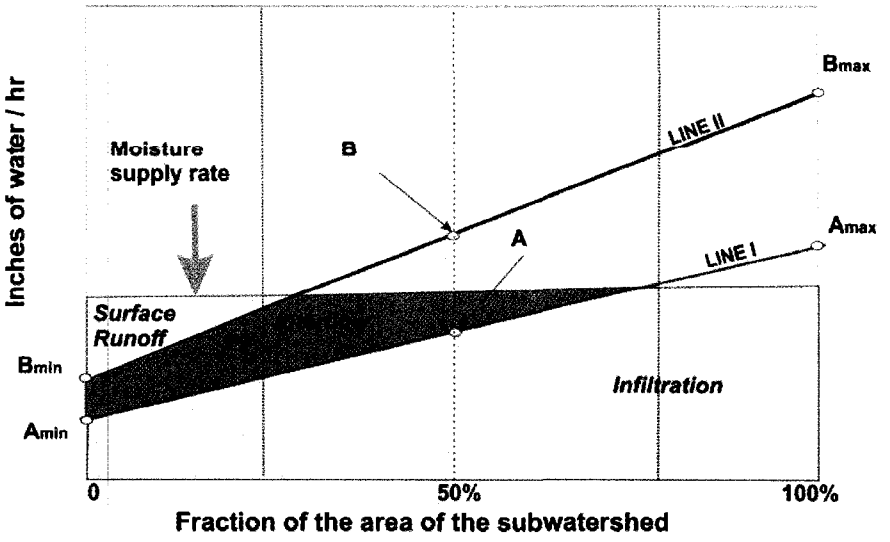


Figure 2.9 Conception of aerial variation of infiltration, interflow, and surface runoff in HSPF ***(adapted from Bicknell et al., 2001)***

Maximum infiltration capacity of the land segment (A_{MAX}) is determined by multiplication of A and a user-defined parameter. A_{MIN} is calculated by doubling A and subtracting A_{MAX}. A similar approach is used for calculating B_{MIN} and B_{MAX}.

2.6.6 Missing data

Temporal resolution is an important factor in hydrological modeling. Some watershed models (HSPF, MIKE-SHE, etc.) allow virtually any temporal scale (down to several minutes), others (SWAT) are restricted only to a daily time step. Given almost unlimited computer resources nowadays, generally it is more appropriate to use fine time scales with fine spatial scales *(Croley and He, 2005)*. However, due to inadequate availability of meteorological and hydrological data, it is not always possible. In practical applications the temporal resolution of a model is controlled by resolution of available meteorological data.

Watershed models require continuous meteorological and hydrological time-series. If a value is missing, which happens quite often, the modeller should devise a method to substitute this value. If a single hourly or daily observation were missing, it would not likely affect model calibration; however, situations with months or even years of missing data may happen.

Three meteorological characteristics play dominant roles in watershed modeling: precipitation, temperature, and evaporation. Several approaches exist to deal with the problem of missing meteorological data.

2.6.6.1 Neighbouring station approach

Meteorological processes occur on a much larger scale than the scale of a typical watershed; therefore, precipitation, temperature, and evaporation data should be analyzed on a regional scale. Usually several meteorological stations exist in a watershed and some hourly or daily missing values can be substituted from a nearby station. A number of software tools have been developed for deriving missing meteorological data based on neighboring stations and historical data. The most widely used software are the mean areal precipitation (MAP) preprocessor, the mean areal temperature (MAT) preprocessor, and the mean aeral potential evaporation (MAPE) preprocessor. They are used for non-mountainous areas and assume weights of neighboring stations based solely on their location in the x, y plain *(Smith et al., 2003)*. Since precipitation is highly spatially variable, usually all available stations inside and outside the watershed are included in the analysis. Given that temperature is less spatially variable, usually fewer temperature stations are required.

The basic equation for missing data estimation in MAP is expressed as

$$P_x = \frac{\sum_{i=1}^{n} \frac{\overline{P_x}}{\overline{P_i}} \cdot P_i \cdot w_{i,x}}{\sum_{i=1}^{n} w_{i,x}} \tag{2.13}$$

where P_x – precipitation at the station with missing data, P_i – precipitation at the estimator station, n – number of estimating stations, i – station being used as an estimator, $\overline{P_x}$ and $\overline{P_i}$ mean monthly precipitation for station x and i, $w_{i,x}$ – station weight, computed according to

$$w_{i,x} = \frac{1}{d_{i,x}^2} \tag{2.14}$$

where $d_{i,x}$ – distance from station x to i.

Missing temperature data in MAT are computed according to

$$T_x = \frac{\sum_{i=1}^{n} \left(\overline{T}_x - \overline{T}_i + T\right) \cdot w_{x,i}}{\sum_{i=1}^{n} w_{x,i}} \tag{2.15}$$

where T – computed temperature and $w_{x,i}$ is estimated by

$$w_{x,i} = \frac{1.0}{d_{i,x} + \Delta h_{x,i}} \tag{2.16}$$

where $d_{i,x}$ – distance, and $\Delta h_{x,i}$ – elevation difference.

Potential evaporation (PE), in mm/d, for stations with missing data can be estimated from energy budget methods such as those from *Kristensen and Jensen* (1975) (Eq. 2.17) or *Hamon* (1961) (Eq. 2.18).

$$PE = (0.016T + 0.186)(Q_s \cdot 0.000673) \cdot 25.4 \tag{2.17}$$

$$PE = 0.656\left(\frac{D}{12}\right)^2 \cdot \frac{SVD}{100} \cdot 25.4 \tag{2.18}$$

where T – air temperature in °C, Q_s – solar radiation in cal/cm^2-d, D – hours of daylight, and SVD – saturated vapor density at mean temperature in g/m^3.

2.6.6.2 Weather generator

Another approach to fill missing data gaps is to use stochastic weather generators (SWGs). SWG simulates climate variables such as temperature and precipitation based on the statistical distribution of historical data. Two types of SWGs are distinguished: parametric and non-parametric. Parametric SWGs need to consider the correlation and dependence of the weather variables with each other on the same day as well as over time both at a station (temporally) and between stations (spatially). Many weather variables are correlated with precipitation, thus meteorological variables are often based on stochastically generated precipitations *(Rajagopalan and Lall, 1999; Yates et al., 2003)*.

Non-parametric SWGs resample data from historical records assuming that past meteorological conditions are representative of what may occur in the future *(Yates et al., 2003; Clark et al., 2004)*. These methods compare a vector of all weather variables for a given day with a vector containing the same variables from similar dates in the historic record. A weighted probability function is assigned to each day with the nearest neighbours having higher probabilities. A random selection is based on a number of similar days. Unique spatial and temporal dependencies can be preserved by resampling from the historic record. It is believed that non-parametric methods allow simple and effective multivariate, multi-site weather generation *(Clark et al., 2004)*.

2.6.7 Disaggregation of input data

Sometimes meteorological data exist at a temporal scale that is too coarse to run the model. For example, instead of hourly precipitation, only daily precipitation exists, or weather service reports a multi-day total instead of individual daily values. A number of software algorithms have been developed to deal with these problems. WDMUtil is most widely used by watershed modelers. It allows analysis of missing or faulty data, generating new time-series using existing data and disaggregating time series from daily to hourly time steps *(USEPA, 2001)*. For example, daily solar radiation can be computed from cloud cover and latitude; potential evaporation may be computed using the Hamon or Jensen method *(Hamon, 1961)*. Time-series can be disaggregated based on data from other stations, latitude, minimum, and maximum values, etc.

An alternative software routine for data disaggregation is the U.S. Army Corps of Engineers program NetStorm that uses an algorithm described in *Socolofsky et al. (2001)*. It can also fill in missing data using a stochastic analyzer.

2.7 Study area and data collection

2.7.1 The South Nation watershed

The research was conducted on the South Nation River (SNR) which is located in south-eastern Ontario and flows in a northeastern direction from headwater near Brockville into the Ottawa River near Plantagenet. The length of the SNR is about 175 km. Elevation difference between its source and mouth is only 84 m due to a very flat landscape. SNR drains about 3810 km^2 (Figure 2.10). Glacially derived marine clays were laid down in the Champlain Sea which covers the study area. The underlaying bedrock is characterized by limestone deposited nearly 10,000 years ago as the Champlain Sea receded *(Haughton, 2002)*. These limestone and sensitive marine clays formed river banks that are subject to extensive erosion which gives the water a brownish color. The clay soils along the SNR often contribute high levels of turbidity. There is a 20 km stretch of highly sensitive Leda clay which is prone to landslides between Casselman and Lemieux. The most recent landslide occurred in July 1993 involving from 2.5 to 3.5 million cubic meters of sand, silt, and clay *(Brooks et al., 1994)*.

The South Nation watershed began to suffer from environmental degradation from the early 1800s as forest was cleared for lumbering. In the late 1800s agriculture moved in and many wetlands were converted to farmlands. The SNR at Casselman was dammed in the early 1900s in order to produce hydroelectric energy. At the present time about 57% of the watershed is used for

farming. Due to reduction of wetlands and forests the SNR has become highly vulnerable to droughts in summer and flooding in spring *(Haughton, 2002)*.

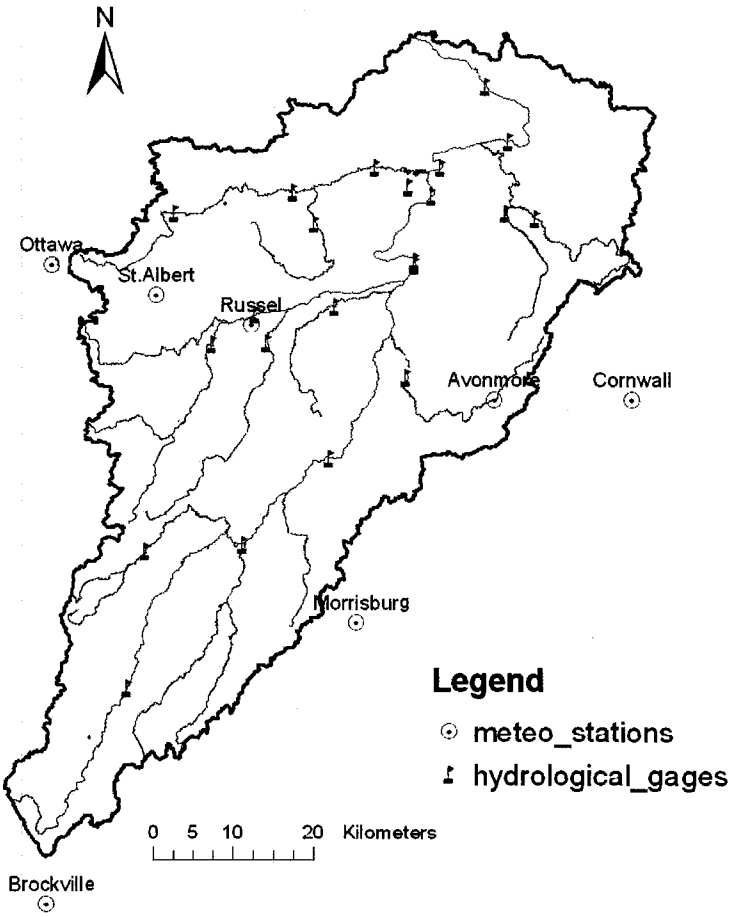


Figure 2.10 Map of the South Nation watershed

The SNR has a seasonal nature of flows. On average, 52 to 62 percent of annual streamflow volume is delivered from local watersheds during the months of March and April *(EOWRMS, 2002)*. The spring high-flow period is followed by relatively rapid recession and a low flow period from late spring to early autumn. The flow increases in October-November due to rainfall. Such seasonality underlines the need for careful management of wastewater discharges and surface water withdrawals (irrigation, water supply) in order to protect aquatic habitats and ensure that downstream water users have adequate water supply.

The watershed receives on average about 930 mm of precipitation annually, of which approximately 420 mm is lost through evaporation and transpiration. The remaining 510 mm of water is partitioned between surface water runoff and deep groundwater reserves *(EOWRMS, 2001)*.

Approximately 57% of the South Nation watershed area is in agricultural use, primarily hay, corn, and soybean crops. Forest is the next major category (about 35%). Urban territories are less than 5% of the total area *(EOWRMS, 2001)*.

2.7.2 Data acquisition and processing

Collection of data is an important step of the research. The observed data are a limiting factor for the model, because the model cannot be expected to be more accurate than the data errors (confidence intervals).

Initially the information needs of the model were defined. Then an inventory of existing information and its sources was made. Information gaps, if any, and possible ways of filling these gaps were identified. The last step of data processing was to obtain, format, and input data into the model. All available data at the maximum spatial and temporal resolution were employed.

The study period covered 11 years from 1990 to 2000. Extensive meteorological data were collected to set up the model. Hourly data include: precipitation, potential evaporation, air temperature, wind speed, solar radiation, and dewpoint temperature. Landuse data were taken from 1998 Landsat imagery with resolution of 30 m. Several landuse categories were identified: agricultural, forest, pasture, bareland, and urban territories. GIS topographical maps, DEM with 25 m resolution, and available hydrographical data were used. Each dataset was examined with respect to homogeneity, completeness, and accuracy. Type and sources of information that were used in this research are presented in Table 2.3.

For modeling purposes, the watershed was divided into subwatersheds. Each segment was assumed to be homogeneous in hydrologic, hydraulic, and water quality response. Watershed segmentation is based on individual characteristics of the watershed, including topography, channel network, geography, land use, meteorology, types of soils, etc. Essentially the segmentation is an iterative procedure of overlaying different data layers and selecting the territories with similar characteristics. The key element of segmentation is to divide the channel network. The following spatially dependent characteristics play the most important role: location of observation gages, channel junctions, characteristics of channel and floodplain (slope,

roughness, depths, etc.), hydrotechnical structures (dams, etc.), and location of point sources of pollution. A stream segment should be short enough to reduce numerical dispersion and accurately represent point and nonpoint sources of pollution. On the other hand, the stream segments should be long enough such that the residence time in every reach exceeds the simulation time step (*Patwardhan and Kittle, 2002*). In this research, GIS (ArcGIS 8.2) tools were used for landuse segmentation and watershed delineation.

Table 2.3 Type of data and its sources

Type of Data	Type of information	Source
Watershed topography	Topographic maps of watershed area, digital elevation data, river networks, drainage areas, roads, water bodies	NRVIS, OMNR, Landsat imagery
Hydrometeorological	Rainfall, snowfall, air temperature, solar radiation, sunshine hours, wind velocity, evaporation	Environment Canada
Hydrologic	Flow depth, streamflow discharge, stream-aquifer interaction, water table	WSC
Agricultural	Vegetative cover, landuse, treatment, fertilizer, and pesticide application	AAFC
Pedologic	Soil type, texture and structure, soil conditions, soil particle size diameter	AAFC
Water supply and wastewater	All available and relevant data	Municipalities
Surface Water Quality	DO, water temperature, pH, BOD, phosphate, nitrate, etc.	PWQMN
Groundwater quality and level monitoring	All available and relevant data	MOE water well database

2.8 References

Agrawal, R.K. and Singh, J.K. (2003). Application of a genetic algorithm in the development and optimisation of a non-linear dynamic runoff model, Biosystems Engineering, 86, 87-95.

Amorocho, J. (1963). Measures of the linearity of hydrologic system, J. Geophys. Res., 68, 8, 2237-2249.

Armour, A. (1990). Adaptive random search evaluated as a method for calibration of the sma-nwsrfs model, M.S. Thesis, Dept of Hydrology and Water Resources, University of Arizona, Tucson.

ASTM. (1984). Standard practice for evaluating environmental fate models of chemicals, designation E978-84. American Society of Testing Materials. Philadelphia, PA.

Banavar, J.R., Colaiori, F., Flammini, A., Giacometti, A., Maritan, A., and Rinaldo, A. (1997). Sculpting of a fractal river basin, Phys. Rev. Lett., 78, 4522-4525.

Bastidas, L.A., Gupta, H.V., Sorooshian, S., Shuttleworth, W.J., and Yang, Z.L. (1999). Sensitivity analysis of a land surface scheme using multicriteria methods, J. Geophys. Res., 104, 19481-19490.

Bathurst, J.C. (1986). Physically-based distributed modeling of an upland catchment using the Systeme Hydrologique Europeen, J. of Hydrology, 87, 79-102.

Baysal, F.M. (2003). Improving efficiency and effectiveness of Bayesian recursive parameter estimation for hydrologic models, PhD dissertation, Department of Hydrology and Water Resources, The University of Arizona, Tucson.

Bergstrom, S. and Graham, L.P. (1998). On the scale problem in hydrological modeling, J. of Hydrology, 211, 1-4, 253-265.

Beven, K.J. (1995). Chapter 18: TOPMODEL, Computer models of watershed hydrology, V.P. Singh, ed., Water Resources Publications, Littleton, Colo.

Beven, K.J. (1989). Changing ideas in hydrology - the case of physically-based models, J. of Hydrology, 105, 157-172.

Beven, K.J. (2000). Rainfall-runoff modeling: the primer, Wiley, Chichester.

Beven, K.J. and Binley, A. (1992). The future of distributed models: model calibration and uncertainty prediction, Hydrol. Process., 6, 279-298.

Bicknell, B.R., Imhoff, J.C., Kittle, J.L., Jobes, T.H., and Donigian, A.S. (2001). Hydrological simulation program HSPF. Version 12. user's manual, U.S. Environmental Protection Agency, Athens, Georgia.

Bloschl, G. and Sivapalan, M. (1995). Scale issues in hydrological modelling: a review, Hydrol. Process., 9, 251-290.

Boyle, D.P., Gupta, H.V., and Sorooshian, S. (2000). Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, Water Resour. Res., 36, 12, 3663-3674.

Brazil, L.E. (1989). Multilevel calibration strategy for complex hydrologic simulation models, PhD Dissertation, Dept. of Civil Engineering, Colorado State University, Fort Collins, Colorado.

Brooks, G.R., Aylsworth, J.M., Evans, S.G., and Lawrence, D.E. (1994). The Lemieux landslide of June 20, 1993, South Nation valley, southeastern Ontario - a photographic record, Geological Survey of Canada, Miscellaneous Report 56.

Burnash, R.J.C. (1995). The NWS river forecast system – catchment modeling, in computer models of watershed hydrology, Singh, V.P., ed., 311-366.

Castrup, H. (2002). Distributions for uncertainty analysis, IDW conference, May 6-9, Knoxville, TN.

Chang, C.H., Tung, Y.K., and Yang, J.C. (1994). Monte Carlo simulation for correlated variables with marginal distribution, J. Hydr. Eng., 120, 313-331.

Chow, V. T., Maidment, D.R., and Mays, L.W. (1988). Applied hydrology. McGraw-Hill.

Clark, M.P., Gangopadhyay, S., Brandon, D., Werner, K., Hay, L., Rajagopalan, B., and Yates, D. (2004). A resampling procedure for generating daily weather sequences, Water Resour. Res., 40, 4, 1-15.

Clarke, R.T. (1994). Statistical modeling in hydrology, John Wiley and Sons, UK.

Crawford, N.H. and Linsley, R.K. (1966). Digital simulation in hydrology: Stanford watershed model IV, Technical Report No. 39, Dept. of Civil Engineering, Stanford University, Stanford, CA.

Croley, T.E. and He, C. (2005). Distributed-parameter large basin runoff model I: model development, J. of Hydrol. Engineering, 10, 3, 173-181.

Cunderlik, J. (2003). Hydrologic model selection for the CFCAS project: assessment of water resources risk and vulnerability to changing climate conditions, Project Report I, Report submitted to Canadian Foundation for Climatic and Atmospheric Sciences (CFCAS).

Dawdy, D.R. and O'Donnell, T. (1965). Mathematical models of catchment behavior, J. Hydraul., 91, 123-127.

DeBarry, P.A., Quimpo, R.G., Garbrecht, J., Evans, T.A., Garcia, L., Johnson, L.E., Jorgeson, J., Krysanova, V., Leavesley, G., Maidment, D.R., Nelson, E.J., Ogden, F.L., Olivera, F., Seybert, T.A., Sloan, W.T., Burrows, D., Engman, E.T., Binger, R., Evans, B.M., and Theurer, F. (1999). GIS modules and distributed models of the watershed, ASCE Task Committee on GIS Modules and Distributed Models of the Watershed, Special Report, ASCE, Reston, VA.

Desai, R. and Patil, R. (1996). SALO: Combining simulated annealing and local optimization for efficient global optimization, Proceedings of the 9th Florida AI Research Symposium, Key West, FL, 233-237.

Doherty, J. (2004). PEST – model-independent parameter estimation, User manual: 5th edition, Watermark Computing, Corinda, Australia.

Doherty, J. and Johnston J.M., (2003). Methodologies for calibration and predictive analysis of a watershed model, J. American Water Res. Association, 39, 2, 251-265.

Donigian, A.S. (2000). HSPF training workshop handbook and CD. Lecture #19. Calibration and Verification Issues, EPA Headquarters, Washington Information Center, Office of Water, Office of Science and Technology, Washington, D.C.

Donigian, A.S. (2002). Watershed model calibration and validation: The HSPF Experience, WEF Specialty Conference Proceedings, WEF National TMDL Science and Policy, Phoenix, AZ.

Duan, Q.Y., Gupta, V., Sorooshian, S., Rousseau, A.N., and Turcotte, R. (2003). Calibration of watershed models, American Geophysical Union, Washington DC.

Duan, Q.Y., Sorooshian, S., and Gupta, V. (1992). Effective and efficient global optimisation for conceptual rainfall-runoff models, Water Resour. Res., 28, 1015-1031.

Duan, Q.Y. (2003). Global optimization for watershed model calibration, in calibration of watershed models, Water Science and Application, American Geophysical Union, Editors: Duan Q., Gupta H., Sorooshian S., Rousseau A., and Turcotte R., 89-104.

Efstratiadis, A. and Koutsoyiannis, D. (2002). An evolutionary annealing-simplex algorithm for global optimisation of water resource systems, Proceedings of the Fifth International Conference on Hydroinformatics, Cardiff, UK, 1423-1428.

EOWRMS (2001). Eastern Ontario water resources management study. Final report is prepared by CH2MHILL, Ottawa, Ontario.

Feldman, A.D. (1981). HEC models for water resources system simulation: theory and experience, Advances Hydrosciences, 12, 297-423.

Franchini, M. and Galeati, G. (1997). Comparing several genetic algorithm schemes for the calibration of conceptual rainfall-runoff models, Hydrol. Sci. J., 42, 3, 357-380.

Franchini, M., Galeati, G., and Berra S. (1998). Global optimization techniques for the calibration of conceptual rainfall-runoff models, Hydrolog. Sciences J., 43, 3, 443-458.

Gan, T.Y. and Biftu, G.F. (1996). Automatic calibration of conceptual rainfall-runoff models: Optimization algorithms, catchment conditions, and model structure, Water Resour. Res., 32, 12, 3513-3524.

Gill, P.E., Murray, W., and Wright, M.H. (1981). Practical optimization, Academic Press, London.

Goldberg, D. E. (1989). Genetic algorithms in search, optimization and machine learning, Addison-Wesley, Reading, Mass.

Grayson, R.B. and Bloschl, G. (2001). Spatial modelling of catchment dynamics, spatial patterns in Catchment Hydrology, 51-81.

Grayson, R.B., Moore, I.D., and McMahon, T.A. (1992). Physically-based hydrologic modeling II: Is the concept realistic? Water Resour. Res., 28, 2659-2666.

Gupta, H.V., Bastidas, L.A., Sorooshian, S., Shuttleworth, W.J., and Yang, Z.L. (1999). Parameter estimation of a land surface scheme using multi-criteria methods, J. Geophys. Res., 104, 19491-19504.

Gupta, H.V., Sorooshian, S., and Yapo, P.O. (1998). Towards improved calibration of hydrologic models: Multiple and noncommensurable measures of information, Water Resour. Res., 34, 4, 751-763.

Gupta, H.V., Sorooshian, S., Hogue, T., and Boyle, D.P. (2003). Advances in automatic calibration of watershed models. Chapter 1 of Advances in Calibration of Watershed Models, Edited by Duan, Q., Sorooshian, S., Gupta, H.V., Rousseau, A.N. and Turcotte, R. Monograph Series on Water Resources, American Geophysical Union, Washington DC.

Hamon, W.R. (1961). Estimating potential evapotranspiration, Proceedings of American Society of Civil Engineers, 87, 107-120.

Haughton, J. (2002). A state of the environment report for the South Nation River watershed, environmental Science Program, University of Ottawa.

Hendrickson, J.D. and Sorooshian, S. (1988). Comparison of Newton-type and direct search algorithms for calibration of conceptual rainfall-runoff models, Water Resour. Res., 24, 5, 691-700.

Heuvelink, G.B.M. (1998). Uncertainty analysis in environmental modelling under a change of spatial scale, Nutrient Cycling in Agroecosystems, 50, 255-264.

Holland, J. (1975). Adaptation in natural and artificial systems, University of Michigan Press, Ann Arbor, Michigan.

Hooke, R. and Jeeves, T. (1961). Direct search solutions of numerical and statistical problems, J. of the Association for Computing Machinery, 8, 2, 212-229.

Ibbitt, R.P. (1970). Systematic parameter fitting for conceptual models of catchment hydrology, Ph.D. thesis, University of London, UK.

Karno, D.C. (1963). Random search techniques for optimization problems, Automatica, 1, 111-121.

Kirchner, J. (2001). Data analysis toolkit: uncertainty analysis and error propagation, Course notes, University of California, Berkeley.

Khuri, A.I. and Cornell, J.A. (1987). Response surfaces: design and analyses, Marcel Dekker, NY.

Klarreich, E. (2005). Life on the Scales, Science News, 167, 7, 106-107.

Kristensen, K.J. and Jensen, S.E. (1975). A model for estimating actual evapotranspiration from potential transpiration, Nordic Hydrology, 6, 70-88.

Kuchment, L.S. (1971). Mathematical modeling of river runoff, Gidrometeoizdat, Leningrad.

Kuczera, G. (1997). Efficient subspace probabilistic parameter optimization for catchment models, Water Resour. Res., 33, 1, 177-185.

Kuo, W.L., Steenhuis, T.S., McCulloch, C.E., Mohler, C.L., Weinstein, D.A., DeGloria, S.D., and Swaney, D.P. (1999). Effect of grid size on runoff and soil moisture for a variable-source-area hydrology model, Water Resour. Res., 35, 3419-3428.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares, Quart. Appl. Math., 2, 164-168.

Marquardt, D.W. (1963). An algorithm for least-squares estimation of nonlinear parameters, J. of the Society of Industrial and Applied Mathematics, 11, 2, 431-441.

Masri, S.F., Bekey, G.A., and Safford, F.B. (1980). A global optimization algorithm using adaptive random search, Appl. Math. and Comput., 7, 353-375.

Metropolis, N.A., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953). Equation of state calculation by fast computing machines, J. Chem. Phys., 21, 1087.

Miller, E.E. (1980). Similitude and scaling of soil water phenomena, Applications of Soil Physics, 300-318.

Mockus, J. (1989). Bayesian approach to global optimization, Kluwer Academic Publishers, Dordrecht-Boston-London.

Moore, I. D., Grayson, R.B., and Ladson, A.R. (1991). Digital terrain modeling: A review of hydrological, geomorphological and biological applications, Hydrological Processes, 5, 1, 3-30.

Morgan, L. (1999). Pesticides and groundwater in the state of Washington, Environmental Management and Health, 10, 7-17.

Nelder, J.A. and Mead, R. (1965). A simplex method for function minimization, Comput. Journal, 7, 308-313.

Patwardhan, A. and Kittle, J. (2002). Technical memorandum #6. Modeling results for exciting conditions, Prepared by CH2MHILL for Metropolitan North Georgia Water Planning District.

Pickup, G. (1977). Testing the Efficiency of algorithms and strategies for automatic calibration of rainfall-runoff models, Hydrological Sciences Bulletin, 22, 2, 257-274.

Rajagopalan, B. and Lall, U. (1999). A k-nearest-neighbor simulator for daily precipitation and other weather variables, Water Resour. Res., 35, 3089-3101.

Refsgaard, J.C. and Henriksen, H.J. (2004). Modelling guidelines – terminology and guiding principles, Advances in Water Resour., 27, 1, 71-82.

Rosenbrock, H.H. (1960). An automatic method for finding the greatest or least value of a function, Comput. J., 3, 175-184.

Savage, V.M., Gillooly, J.F., Woodruff, W.H., West, G.B., Allen, A.P., Enquist, B.J., and Brown, J.H. (2004). The predominance of quarter-power scaling in biology, Functional Ecology, 18, 257-282.

Savic, D.A., Walters, G.A., and Davidson, J.W. (1999). A genetic programming a roach to rainfall-runoff modelling, Water Resour. Management, 13, 219-231.

Schoups, G., Hopmans, J.W., Young, C.A., Vrugt, J.A., and Wallender, W.W. (2005). Multi-criteria optimization of a regional spatially-distributed subsurface flow model, J. of Hydrology, 311, 20-48.

Seybert, T.A. (1996). Effective partitioning of spatial data for use in a distributed runoff model, Doctor of Philosophy Dissertation, Department of Civil and Environmental Engineering, The Pennsylvania State University.

Schwefel, H.P. (1981). Numerical optimisation of computer models, John Wiley, Chichester.

Singh, V.P. and Woolhiser, D.A. (2002). Mathematical modeling of watershed hydrology, J. of Hydrologic Engineering, 7, 4, 270-292.

Singh, V.P. (2004). Flow routing in open channels: some recent advances, Second International Conference on Fluvial Hydraulics, June 23-25, 1-23.

Smith, M., Laurine, D., Koren, V., Reed, S., and Zhang, Z. (2003). Hydrologic model calibration in the national weather service, water science and application, American Geophysical Union, 133-152.

Socolofsky, S., Adams, E.E., and Entekhabi, D. (2001). Disaggregation of daily rainfall for continuous watershed modeling, Journal of Hydrologic Engineering, 6, 4, 300-309.

Solomatine, D.P., Dibike, Y.B., and Kukuric, N. (1999). Automatic calibration of groundwater models using global optimization techniques, Hydrological Sciences Journal, 44, 6, 879-894.

Thyer, M., Kuczera, G., and Bates, B.C. (1999). Probabilistic optimization for conceptual rainfall-runoff models: A comparison of the shuffled complex evolution and simulated annealing algorithms, Water Resour. Res., 35, 3, 767-773.

Timbe, L. and Willems, P. (2004). Uncertainties in hydrodynamic flood simulation, ACTIF Workshop on Quantification, reduction and dissemination of uncertainty, Delft, November 22-23, 2004, 1-15.

Törn, A. and Zilinskas, A. (1987). Global optimization, Springer, Verlag, Berlin.

USEPA. (1997). Technical Panel: Guiding principles for Monte Carlo analysis, EPA /630/R-97/001.

USEPA. (1999). Treatment of uncertainty and variability, in report Total risk integrated methodology, EPA-453/R-99-010.

USEPA. (2001). WDMUtil version 2.0: A tool for managing watershed modeling time-series, Data User's Manual, U.S. Environmental Protection Agency.

US Committee. (1985). Evaluation of hydrologic models used to quantify major land-use change effects, J. of Irrigation and Drainage Engineering, 11, 1, 1-15.

Vazquez, R., Feyen, J., Feyen, L., and Refsgaard, J. (2002). Effect of grid size on effective parameters and model performance of the MIKE-SHE code, Hydrological Proces., 16, 355-372.

Vivoni, E.R., Ivanov, V.Y., Bras, R.L., and Entekhabi, D. (2005). On the effects of triangulated terrain resolution on distributed hydrologic model response, Hydrological Processes, 19, 11, 2101-2122.

Wagener, T., Lees, M.J., and Wheater, H.S. (2001). A Matlab toolkit for conceptual rainfall-runoff modelling and the analysis of Monte-Carlo simulations, BHS Circulations PLUS, Su lement to BHS Circulations, 68, 11-12.

Wagenet, R.J. (1998). Scale issues in agroecological research chains, Nutrient Cycling in Agroecosystems, 50, 23-34.

Wang, Q.J. (1991). The genetic algorithm and its application to calibrating conceptual rainfall-runoff models, Water Resour. Res., 27, 9, 2467-2471.

Ward, G.H. and Benaman, J. (1999). A survey and review of modeling for TMDL application in Texas watercourses, report for the Texas Natural Recourses Conservation Commission CRWR-99-8, Austin, Texas.

Weinig, W.T. (1991). Calibration of the soil moisture accounting model using the adaptive random search algorithm, M.S. Thesis, Dept of Hydrology and Water Resources, University of Arizona, Tucson.

West, G.B., Brown, J.H., and Enquist, B.J. (1997). A general model for the origin of allometric scaling models in biology, Science, 276, 122-126.

Willems, P. (2000). Probabilistic emission modelling of receiving surface waters, PhD thesis, Katholieke Universiteit Leuven, Leuven.

Wolock, D.M. and McCabe, G.J. (2000). Differences in topographic characteristics computed from 100 and 1000 meter resolution digital elevation model data, Hydrological Processes, 14, 987-1002.

Wolock, D.M., and Price, C.V. (1994). Effects of digital elevation model map scale and data resolution on a topography-based watershed model, Water Resour. Res., 30, 11, 3041-3052.

World Meteorological Organization (WMO). (1975). Intercomparison of conceptual models used in operational hydrological forecasting, Operational Hydrology Paper, 429, Geneva.

World Meteorological Organization (WMO). (1986). Intercomparison of models of snowmelt runoff, Operational Hydrology Paper, 646, Geneva.

World Meteorological Organization (WMO). (1992). Simulated realtime intercomparison of hydrological models, Operational Hydrology Paper, 779, Geneva.

Yapo, P.O., Gupta, H.V., and Sorooshian, S. (1998). Multi-objective global optimization for hydrologic models, J. Hydrol., 204, 83-97.

Yates, D., Gangopadhyay, S., Rajagopalan, B., and Strzepek, K. (2003). A technique for generating regional climate scenarios using a nearest-neighbor algorithm, Water Resour. Res., 39, 7, 1199-1213.

Yen, B.C. (2002). System and component uncertainties in water recourses, in book: Risk, reliability, uncertainty and robustness of water recourses systems, ed. Bogari J. and Kundzewicz, Z.W., Cambridge university press.

Zhang, W. and Montgomery, D. (1994). Digital elevation model grid size, landscape representation, and hydrologic simulations, Water Resour. Res., 30, 4, 1019-1028.

CHAPTER 3

APPLICATION OF NONLINEAR AUTOMATIC OPTIMIZATION TECHNIQUES FOR CALIBRATION OF HYDROLOGICAL SIMULATION PROGRAM FORTRAN

Igor Iskra and Ronald Droste
Department of Civil Engineering, University of Ottawa,
161, Louis Pasteur, Ottawa, Ontario, K1N 6N5, Canada

The novel contribution of the presented paper is in studying the impact of GML, RSM, and SCE-UA variables on the methods' ability to find the global minimum of the OF. The best variable values for HSPF calibration were suggested.

3.1 Abstract

Development of total maximum daily loads is often facilitated by using the Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) software system. One of the key elements of BASINS is the watershed model Hydrological Simulation Program-Fortran (HSPF), developed by the U.S. Environmental Protection Agency. Calibration of HSPF is a tedious and time-consuming task in that more than 100 parameters are involved in the calibration process. In this current research, three nonlinear, automatic optimization techniques are applied and compared and an efficient way to calibrate HSPF is suggested. Parameter optimization using local and global optimization techniques for the watershed model is discussed. Approaches to automatic calibration of HSPF using the nonlinear parameter estimation tool (PEST) with its Gauss-Marquardt-Levenberg (GML) method, random multiple search method (RSM), and Shuffled Complex Evolution method University of Arizona (SCE-UA) are presented.

A sensitivity analysis was conducted and the most- and least-sensitive parameters were identified. It was noted that sensitivity depends on a number of adjustable parameters. As more parameters were optimized simultaneously, a wider range of parameter values maintained the model in the calibrated state. The impact of GML, RSM, and SCE-UA variables on the ability to find the global minimum of the objective function was studied and the best variables are suggested. All three methods proved to be more efficient than manual HSPF calibration. Optimization results obtained by these methods are similar, although, in most cases, RSM outperforms the GML method and the SCE-UA method outperforms RSM. The GML method is very fast and can perform as well as the SCE-UA method when the variables are properly adjusted, the initial guess is good, and insensitive parameters are eliminated from the optimization process. The SCE-UA method is very robust and convenient to use. In most cases, logical definition of key variables leads to the global minimum.

Keywords: calibration, Hydrological Simulation Program Fortran, parameter estimation tool, Gauss-Marquardt-Levenberg method, shuffled complex evolution-University of Arizona.

3.2 Introduction

Watershed models are becoming increasingly widely used by engineers, ecologists, and scientists. Almost every project that deals with watershed planning, development, or management involves modeling. A number of watershed models exist today ranging from simple steady-state models to

complex distributed models with hundreds of parameters. One of the most widely used watershed models in North America is Hydrological Simulation Program Fortran (HSPF), which can simulate hydrological processes and water quality in water bodies. HSPF is a semi-distributed model in which a watershed is subdivided into subwatersheds with homogeneous properties and model parameters are allowed to vary spatially.

HSPF is a part of the software system Better Assessment Science Integrating Point and Nonpoint Sources, which is commonly used to develop total maximum daily loads. The last available HSPF version (12.0) was used in the current research. Detailed documentation and model descriptions are available from *Bicknell et al. (2001).* To run the model properly, HSPF requires a large amount of data including watershed data, topography, land use, soil data, hydrological characteristics, and detailed meteorological datasets. Acquisition, reformatting, and management of input time series consume most of the time in model application. In addition, the model has a wealth of empirical calibration parameters that must be determined from handbooks, field measurements, and by calibration.

Model calibration is a critical step in model application. Parameter values need to be adjusted to reproduce the response as closely as possible to the observed data. For most watershed models, calibration is a difficult and time-consuming iterative procedure of parameter evaluation and refinement. Presently, models are more sophisticated with hundreds of interrelated parameters. Semidistributed and distributed models are quickly replacing lumped parameter models. This factor has motivated the development of many automatic techniques in watershed modeling during the last decade that are constantly improving parallel to increasing computer power. All of these automatic techniques have to overcome the highly nonlinear nature of the watershed models and the existence of multilocal minima in the pbjective function (OF). The nonlinear parameter estimation tool (PEST), with its Gauss-Marquardt-Levenberg (GML) method, random multiple search method (RSM), and shuffled complex evolution method developed at the University of Arizon (SCE-UA), was used for HSPF calibration the present research. Much useful information on HSPF calibration using PEST is presented in *Doherty and Johnston (2003).*

Research was conducted in the South Nation watershed located in southeastern Ontario, Canada. The South Nation River (SNR) flows in a northeastern direction from headwaters near Brockville, Ontario, into the Ottawa River near the community of Plantagenet, Ontario. The SNR is approximately 175-km long, and the elevation difference between its source and mouth is 84 m; the landscape of the river is very flat. The SNR drains an area of approximately 3810 km^2.

Glacially derived marine clays were laid down in the Champlain Sea, which covered the area. The underlaying bedrock is characterized by limestone deposited nearly 10 000 years ago as the Champlain Sea receded *(Haughton, 2002)*. Presently, approximately 57% of the watershed is used for farming. Due to wetland and forest reduction, SNR has become vulnerable to droughts in the summer and flooding in the spring *(Haughton, 2002)*. For this study, the watershed was subdivided into 34 subwatersheds. The subdivisions were based on the distribution of meteorological and hydrological stations as well as relief and soil types. Each subwatershed was further subdivided into land use categories.

The study period covers 8 years, from 1990 to 1998. Extensive meteorological data from Environment Canada were used to set up the model. Hourly data include precipitation, potential evaporation, air temperature, wind speed, solar radiation, cloud cover, and dewpoint temperature. Daily data include snow depths and minimum and maximum temperature. Land use data were taken from Landsat imagery with a resolution of 30 m. The following five land use categories were identified: agriculture, forest, pasture, bare land, and urban territories. Geographical Information System topographical maps, a digital elevation model with 25-m resolution, and hydrographical data were used. Flow and water quality data were obtained from the South Nation Conservation Authority.

3.3 Methodology

3.3.1 Optimizations methods

□□□□ □ollo□i□□□ two broad groups of nonlinear optimization methods are widely used in watershed model calibration: single extreme (local) and multi-extreme (global) optimization methods. In the present research, one local method (GML) and two global optimization methods (RSM and SCE-UA) were compared and evaluated. The nonlinear estimation package PEST, version 9, was used throughout the research.

PEST implements a particularly robust variant of the GML method, which is actually a blend of gradient-descent and Gauss-Newton methods. When a function is far from its minimum, the GML method resembles the steepest decent method; around the minimum, the GML method works as a Gauss-Newton method. The Marquardt lambda (Eq. 3.1) is a blending factor that determines the mix between gradient-descent and Gauss-Newton methods. If the error is increasing, the function is far from its minimum and lambda should be increased approaching the

gradient-descent method. If the error is decreasing, the approximation is working well so lambda is decreased. The parameter upgrade vector (u) is calculated according to the following equation:

$$u = (J^T \cdot W \cdot J + \lambda \cdot I)^{-1} \cdot J^T \cdot W \cdot r \tag{3.1}$$

where J – a Jacobean matrix, each element is a derivative of one particular observation with respect to each parameter, W – square diagonal matrix with squared weights attached to each observation, λ – Marquardt lambda or damping factor, which is adjusted during each iteration, r – vector of residuals which is minimized during calibration, I – an Identity matrix, and T – matrix transpose operator.

As seen from Eq. 3.1, the GML algorithm requires computation of the Jacobian matrix and matrix inversion at each iteration step as many times as there are adjustable parameters. Matrix inversion is a significant disadvantage. When the matrix is singular, GML still performs due to the Marquardt lambda; however, the risk that optimization does not reach the global minimum is too high. A matrix often becomes singular when columns (or rows) are linearly dependent.

Besides the aforementioned local optimization method (LOM), the following two global optimization methods were used: RSM and SCE-UA. Subroutines developed by Watermark Numerical Computing, Brisbane, Australia and the University of Idaho, Moscow (*pd_ms2* and *sceua_p*, respectively) were used *(Doherty, 2004)*.

The RSM does a random parameter sampling. When running PEST from different starting locations in a user-specified parameter space, every next starting location is based on "previous mistakes." The RSM can be seen as an improved version of the GML method. After random sampling in a parameter space, RSM records values of the objective function for each parameter set and ranks them. Then, the parameter set corresponding to the lowest objective function is used for PEST optimization. After its completion, RSM picks another random parameter set that is as far as possible from the trajectory of the previous parameters and performs optimization again. This is done to minimize the probability of finding the same minimum. In the next step, RSM tests the parameter set that is maximally distant from all previous parameter trajectories. Termination criteria such as maximum number of inversions runs, number of runs with no improvements, and objective function improvement judged to be negligible are supplied by the user. The strength of this method is in its introduction of a certain degree of randomness into the optimization process. Randomness ensures that all "regions of attractions" in a parameter space are found and checked during the optimization process. Although RSM requires more computer

power and is significantly lengthier than the GML method, it often yields superior results of calibration.

The SCE-UA method is a global optimization method developed by *Duan et al. (1992)* at the University of Arizona explicitly for calibration of watershed models. The SCE-UA method is based on synthesis of genetic algorithm, simplex downhill search, control random search method, and complex shuffling. The procedure for the SCE-UA method consists of the following steps.

- Sample s parameter sets with random values and compute the objective function for each parameter set. Rank the sets in order of increasing objective function.
- Partition s parameter sets into p complexes, with m points in each. The first complex contains $p \times (k - 1) + 1$ ranked parameter sets, the second complex contains $p \times (k - 1) + 2$ sets and so on, where $k = 1, 2, 3\ldots, m$.
- Each complex evolves according to the *Nelder and Mead (1965)* simplex downhill search method. The worst parameter sets are improved by reflection, contraction, and expansion rules.
- This is the shuffling stage. The evolved parameter sets in each complex are combined into a single sample population. The population is sorted according to the objective function and repartitioned into p complexes according to the rule in step 2. Then, the convergence criteria are checked. If they are not met, the process is repeated.

Thus, the SCE-UA method is a robust, effective, and efficient global optimization algorithm combining the powers of deterministic and probabilistic methods. *Duan et al. (1994)* and *Hashino et al. (2002)* documented several successful applications of the SCE-UA method for calibrating watershed models.

3.3.2 Sensitivity analysis

An integral part of the calibration process, a sensitivity analysis of HSPF parameters was performed. Parameter sensitivity reflects the effect of parameter changes on the output function. The sensitivity of each parameter with respect to all weighted observations was computed. That composite sensitivity (S) of parameter i is defined as (Doherty, 2004)

$$S_i = \frac{(J^t W J)_{ii}^{1/2}}{m} \tag{3.2}$$

As such, the composite sensitivity of parameter i is the square root of the dot product of the column i of the Jacobian matrix with itself, normalized with the respect to the number of observations m.

Additionally, having the Jacobian matrix calculated allows analysis and comparison of observation sensitivity (i.e., how sensitive is observation j with respect to a particular parameter in column i). Often, the composite sensitivity of a particular observation to all adjustable parameters is used. Composite observation sensitivity is defined as the magnitude of the "ith" row of the Jacobian matrix multiplied by the observation weight and divided by the number of adjustable parameters. Although the observation sensitivity was examined, its analysis is beyond the scope of this paper.

3.3.3 Objective function

Calibration is an iterative procedure of parameter evaluation and adjustment, with the objective to reproduce a modeled response identical to the observed reality. The essence of calibration is to minimize the error between observed and model data. The error is often expressed in weighted squared residuals (of the objective function). In the present research, the classical approach to assign different weights to several single objective functions was used. The compound objective function (ϕ) minimized during optimization is a combination of the following three single objective functions: weighted squared differences between simulated and observed daily flows (objective function 1 [OF1]), weighted squared differences between simulated and observed monthly volumes (objective function 2 [OF2]), and weighted squared exceedance times (objective function 3 [OF3]) for the chosen flow thresholds. The exceedance times are calculated on the basis of the daily flow. Φ is defined as:

$$\phi = \sum_{i=1}^{n}[w_1^i \times (q_{obsr}^i - q_{sim}^i)]^2 + \sum_{i=1}^{k}[w_2^i \times (V_{obsr}^i - V_{sim}^i)]^2 + \sum_{i=1}^{j}[w_3^i \times (E_{obsr}^i - E_{sim}^i)]^2 \quad (3.3)$$

where n – number of daily flow observations, k – number of months, j – number of flow thresholds, q_{obsr} and q_{sim} – the values of daily observed and simulated flows, V_{obsr} and V_{sim} – the values of monthly observed and simulated volumes, E_{obsr} and E_{sim} – the exceedance times for the defined thresholds, and w_1, w_2, w_3 – the weighting factors.

Weights were assigned to achieve approximately equal contributions from all three groups to the compound objective function. Multiple trial and error simulations were performed to obtain the

weighting coefficients for each OF which will contribute approximately 33.3 % into the total OF. This necessitated the following weight factors: the assigned weight for OF1 was 0.1 and the assigned weight for OF2 was $0.008/\sqrt{V_{obsr}}$. The exceedance times in days were calculated for the following flow thresholds: 5, 7, 10, 100, 1000, 2000, 3000, 4000 ft^3/s. The weighting factor for OF3 was assigned a value of 15 000. Depending on the objectives of the study, the weighting factors as well as the objective function itself can vary. In general, the proper formulation of objective functions and their weights is key to proper automatic optimization.

It was crucial to obtain proper snow calibration for predicting daily flows. Although snow calibration was performed using the same GML, RSM, and SCE-UA methods, the objective function was defined as a simple sum of squared residuals of snow depths, as follows:

$$OF_{snow} = \sum_{i=1}^{n} (S_{obsr}^{i} - S_{sim}^{i})^2 \tag{3.4}$$

where S_{obsr} and S_{sim} are the values of daily observed and simulated snow depths.

3.3.4 Performance comparison

Although correlation and correlation-based measures have been widely used for model calibration, it was observed that they are oversensitive to extreme values and insensitive to additive and proportional differences between observed and simulated values *(Legates and McCabe, 1999)*. As a result, alternative relative error measures such as coefficient of efficiency and index of agreement were used in this research alongside the traditional coefficient of determination. During calibration, the best parameter sets were selected based on the following statistical characteristics:

a) Coefficient of determination (r^2). This is the square of the Pearson's product-moment correlation coefficient. It describes the proportion of the total variance in the observed data that can be explained by the model. The coefficient of determination varies from 0 to 1, with higher values indicating better agreement between observed and modeled data.

$$r^2 = \frac{\sum_{i=1}^{n} (q_{obsr}^{i} - \bar{q}_{obsr}) \times (q_{sim}^{i} - \bar{q}_{sim})}{\left[\sum_{i=1}^{n} (q_{obsr}^{i} - \bar{q}_{obsr})^2\right]^{0.5} \left[\sum_{i=1}^{n} (q_{sim}^{i} - \bar{q}_{sim})^2\right]^{0.5}} \tag{3.5}$$

$\bar{q}_{obsr}$ *and* $\bar{q}_{sim}$ are mean values of observed and simulated daily flows

b) Nash and Sutcliffe (NS) model fit. Suggested by *Nash and Sutcliffe (1970),* this is a unitless measure of the ratio of the residual variance ("noise") to the variance of observed flow ("information"). Values may vary from - ∞ to 1.0, with higher values indicating better agreement.

$$NS = 1 - \sum_{i=1}^{n} (q_{obsr}^{i} - q_{sim}^{i})^2 \Big/ \sum_{i=1}^{n} (q_{obsr}^{i} - \bar{q}_{obsr})^2 \tag{3.6}$$

c) Coefficient of efficiency (CE). This characteristic is similar to the Nash and Sutcliffe model fit in that higher values indicate better agreement.

$$CE = 1 - \sum_{i=1}^{n} \left| q_{obsr}^{i} - q_{sim}^{i} \right| \Big/ \sum_{i=1}^{n} \left| q_{obsr}^{i} - \bar{q}_{obsr} \right| \tag{3.7}$$

d) The index of agreement (IA). This characteristic was proposed by *Willmott (1981)* to overcome the insensitivity of correlation-based measures to differences in the observed and modeled means and variances.

$$d = 1 - \sum_{i=1}^{n} \left| q_{obsr}^{i} - q_{sim}^{i} \right| \Big/ \sum_{i=1}^{n} \left(\left| q_{sim}^{i} - \bar{q}_{obsr} \right| + \left| q_{obsr}^{i} - \bar{q}_{obsr} \right| \right) \tag{3.8}$$

e) The daily root mean square error (DRMS). This characteristic simply computes the standard deviation of the model prediction error; a smaller value indicates better agreement between observed and modeled data.

$$DRMS = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (q_{sim}^{i} - q_{obsr}^{i})^2} \tag{3.9}$$

It was sometimes valuable to consider mean error or bias, which measures the average tendency of the simulated value to be larger or smaller than its observed counterparts. The optimal value is zero. Positive values indicate a model bias toward overestimation and vice-versa. Occasionally, mean absolute error (MAE), percent mean error (PME), and percent mean absolute error (PMAE) were considered.

3.4 Results

3.4.1 Sensitivity analysis

Parameter sensitivity is important in distinguishing parameters that are insensitive to model outputs and that contribute little to model improvement. Table 3.1 shows composite parameter sensitivity (computed using Eq. 3.2) of the calibrated model with respect to the following three objective functions: daily flows (3287 observations), monthly volumes (108 observations), and exceedance times (8 thresholds). Table 3.1 shows that daily flows are most sensitive to the index of infiltration capacity of soil (INFILT). This was not surprising because INFILT relates to mean soil infiltration rates and effectively controls the division of rain water to surface and subsurface storages. High values of INFILT result in increased base flow, while low values produce more overland flow and interflow. Monthly volumes are most sensitive to base groundwater recession (AGWRC), one day's groundwater discharge to previous day's groundwater discharge. The importance of AGWRC for monthly flows probably stems from the fact that AGWRC mostly impacts the timing of seasonal and low flows. It is a complex function of watershed land use and soils. As Table 3.1 shows, exceedance times are most sensitive to the fraction of evapotranspiration from active groundwater (AGWETP) as well as AGWRC. The importance of AGWETP for exceedance times is a little surprising because AGWETP mostly controls evapotranspiration losses from groundwater storage; moreover, its weight should be visible mostly in low-flow simulations and affect the annual water balance. Parameter interception storage capacity (CEPSC) (in inches) is insensitive to all objective functions and probably can be eliminated from the optimization process. CEPSC estimates the amount of rainfall retained by vegetation that never reaches the land surface, and plays an important role for watersheds with extensive dense forests. Because the South Nation is mostly an agricultural watershed, insensitivity of CEPSC was expected.

It is worth noting, however, that sensitivity of parameters changes during the optimization process. Analyzing the Jacobian matrix after the first several iterations, it was noted that sensitivity of CEPSC is relatively high. Therefore, CEPSC is an important parameter when the objective function is far from its minimum and, essentially, insensitive when the objective function is around its global minimum. Analogous behavior was noted for the fraction of evapotranspiration from base flow (BASETP) in that its sensitivity between the first and the last iterations declines by a factor of two. INFILT and lower zone evapotranspiration (LZETP), conversely, are very insensitive at the beginning of optimization and become three times more sensitive around the global minimum.

Among the snow parameters, the highest sensitivity to OF1 (daily flows) was the snow gage catch correction factor (SNOWCF) (Table 3.1). This parameter affects snow accumulation in the watershed by adjusting the catch factor in the precipitation gage. SNOWCF depends on the type of precipitation gage, how it is shielded, and wind speed during snowfall. SNOWCF is also a powerful tool to adjust snow precipitation in the watershed. However, other authors have indicated high sensitivity of SNOWCF *(Zarriello, 1999)*. *Crawford (1999)* pointed out that SNOWCF values between 1 and 1.5 are physically realistic.

Table 3.1 Composite parameter sensitivity with respect to OF1, OF2, and OF3

Parameter	Definition	OF1	OF2	OF3
Hydrological parameters				
LZSN	Lower zone nominal storage	0.04	0.91	27.7
INFILT	Index of infiltration capacity of soil	2.86	36.9	1733
AGWRC	Base groundwater (GW) recession	1.40	53.9	3450
DEEPFR	Fraction of ground water inflow to deep recharge	0.29	13.4	688.7
BASETP	Fraction of evapotranspiration (ET) from baseflow	0.13	3.64	92.2
AGWETP	Fraction of ET from active GW	0.16	26.6	3461
CEPSC	Interception storage capacity (in.)	3.8×10^{-16}	$9.1\times10^{-\square\square}$	6.0×10^{-13}
UZSN	Upper zone nominal storage (in.)	0.11	1.81	24.1
NSUR	Roughness of overland flow	0.05	0.23	2.61
IRC	Interflow recession parameter	0.54	3.70	70.8
LZETP	Lower zone ET	0.31	11.7	579
Snow parameters				
SHADE	Fraction of land segment shaded from solar radiation	0.36	4.46	68.2
SNOWCF	Snow gage catch correction factor	0.96	13.4	109
COVIND	Snowfall required to fully cover surface	7.3×10^{-2}	0.84	1.90
RDCSN	Density of new snow	1.1×10^{-2}	0.20	2.29
CCFACT	Condensation/convection melt factor	0.87	15.1	160
MWATER	Liquid water storage capacity in snowpack	0.40	7.73	127

The condensation/convection melt factor (CCFACT), which controls the rate of convective or condensation heat transfer to the surface, is sensitive for OF2 (monthly volumes) and OF3 (exceedance times). In a sense, CCFACT controls heat transfer by leveling possible uncertainties in wind velocities and air temperature.

The least-sensitive parameters are the snowfall required to fully cover the surface (COVIND) (mostly due to the flat topography of the South Nation watershed) and the density of new snow

(RDCSN) (due to the nature of the parameter). Fortunately, however, their low sensitivity does not cause problems for matrix inversion and they were optimized simultaneously with other parameters.

The aforementioned 11 hydrological parameters and six snow parameters were optimized simultaneously. During the course of the sensitivity analysis, it was noted that parameter sensitivity depends on the number of parameters being optimized. The classical approach to study sensitivity is to fix all parameters but one and to use Monte Carlo or other analysis to perturb the parameter and record the outcome. The results of this classical approach are shown in Figure 3.1(a). The lower zone nominal storage (LZSN) and INFILT parameters were allowed to vary and the objective function was recorded for each iteration. In both cases, parabolas with clear minima around 2.9 and 0.08, respectively, were observed. Then, three parameters (Figure 3.1[b]) and 11 parameters (Figure 3.1[c]) were simultaneously optimized and the objective function response was recorded. It can be seen that variability of LZSN and INFILT is dramatically increased and more parameter values correspond to the minimum in the observed function.

Figures 3.1(b) and 3.1(c) were generated using the SCE-UA method with 2000 and 15 000 model runs, respectively. Computational time soars from several hours for one parameter (Figure 3.1[a]) to 3 days for 11 parameters (Figure 3.1[c]). Table 3.2 presents potential parameter variability depending on the number of adjustable parameters. Assuming that the model is calibrated if the objective function is less than 1.9×10^6 (slightly above the global minimum), it can be seen that LZSN values between 2.803 and 3.126 preserve the model with one adjustable parameter in the calibrated state. The range of "suitable" LZSN values increases two-fold when 11 parameters are optimized simultaneously. This may be explained by parameter intercorrelation. The parameters can somewhat compensate upward or downward movement of each other. Generally, as more parameters are optimized simultaneously; a wider range of parameter values can maintain the model in the calibrated state.

3.4.2 Optimization of GML variables

During HSPF calibration, it was noted that GML variables can significantly impact how fast and precise the minimum value of the objective function is found. During any optimization iteration, GML tries several lambdas. Lambda is lowered if the objective function decreases, and lambda is increased if the objective function increases. With higher lambda, as previously noted, the optimization resembles the steepest descent method, which works very well during initial iterations, especially when the initial guess is poor.

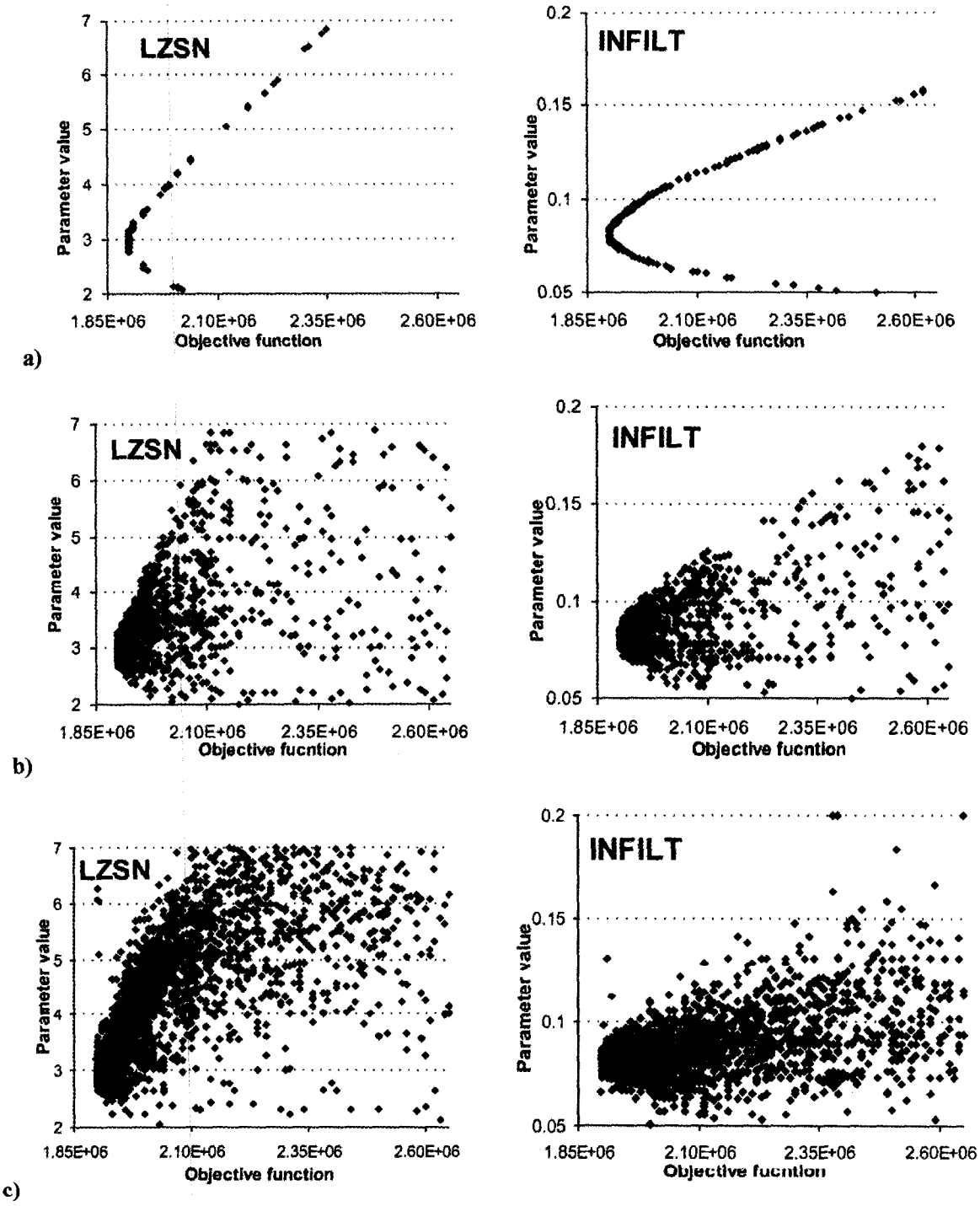


Figure 3.1 **Sensitivity of parameters when (a) one parameter is optimized independently, (b) three parameters are optimized simultaneously, and (c) 11 parameters are optimized simultaneously**

On approaching the minimum value of the objective function, lambda becomes smaller. Two variables, initial Marquardt lambda (RLAMBDA1) and the maximum factor change of parameter that is allowed in one optimization iteration (FACPARMAX), were found to be sensitive. In the current application, all adjustable parameters were stated as factor-limited. If any parameter is chosen to be relative-limited, the maximum relative change of parameter that is allowed in one optimization iteration (RELPARMAX) variable becomes important.

Table 3.2 Range of parameter values that maintain the model in the calibrated state (objective function < 1.9×10^6)

Parameter	No. of Perturbing Parameters	Min.	Max.	Difference
LZSN	1	2.803	3.126	0.323
	3	2.764	3.264	0.500
	11	2.643	3.290	0.647
INFILT	1	0.077	0.084	0.008
	3	0.076	0.086	0.010
	11	0.077	0.087	0.010
UZSN	1	1.886	2.000	0.114
	3	1.874	2.000	0.126
	11	1.849	2.000	0.151

Figure 3.2 shows the surface of the objective function as a function of RLAMBDA1 and FACPARMAX. Locally weighted scatterplot smoothing (LOESS) with tricube weighting and third-order polynomial regression was used to build the error surface. The figure is based on results of 180 runs.

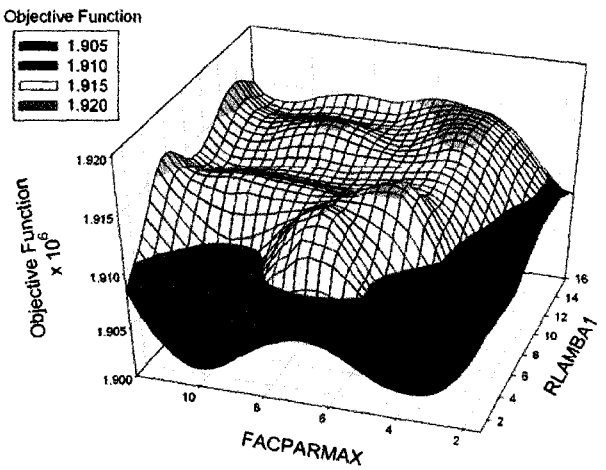


Figure 3.2 Impact of RLAMBA1 and FACPARMAX on the objective function

The darkest color in Figure 3.2 indicates areas where the objective function is at its minimum. Two minima were found at RLAMBA1 =3, FACPARMAX =4 and at RLAMBA1 =3, and FACPARMAX =9. Based on the experiments conducted, it is best to begin optimization with the initial lambda =3 and relative allowable parameter change =4. It was concluded that due to the sensitive nature of the aforementioned variables in HSPF calibration, it is necessary to try several initial lambdas and maximum factor change to ensure that the objective function is at its minimum. The default PEST variables setting with RLAMBA1 =5 and RELPARMAX =5 do not perform well.

Another important variable is the factor by which lambda is adjusted (RLAMFAC) during each iteration. Lambda is divided or multiplied by RLAMFAC to improve the objective function. Table 3.3 shows the impact of RLAMFAC on the objective function and performance statistics. Although the impact of RLAMFAC on statistics is hardly visible, the minimum in objective function suggests using a value of 2. This is in accord with the recommendations of *Doherty (2004).*

Table 3.3 Impact of the factor by which the lambda is adjusted on the objective function and model statistics

Value	OF	DRMS	NS	CE	IA
1.1	1.9358×10^{6}	205.0	0.7557	0.6144	0.8008
1.5	1.8989×10^{6}	204.9	0.7560	0.6188	0.8030
2	1.8978×10^{6}	205.1	0.7557	0.6189	0.8032
2.5	1.9033×10^{6}	205.2	0.7554	0.6185	0.8027
3	1.9120×10^{6}	205.1	0.7556	0.6172	0.8015
3.5	1.9159×10^{6}	206.9	0.7520	0.6185	0.8034

If PHIRARSUF cannot be satisfied, PEST uses another criterion to move to the next iteration. If a relative reduction between two consecutive iterations is less than the maximum relative reduction of the objective function between two consecutive lambdas (PHIREDLAM), it is more efficient to begin a new iteration instead of testing new lambdas. PHIREDLAM should not be too high giving PEST the opportunity to try several lambdas and not too low wasting time trying many lambdas. Figure 3.3 shows the impact of different values of PHIREDLAM on the objective function. Values of PHIREDLAM between 0.0003 and 0.001 were found to be appropriate.

The variable NUMLAM controls the number of lambdas PEST can try during each optimization iteration. A value of 10 was found to be suitable for the current application. The maximum number of PEST iterations for a particular optimization run (NORTMAX) was set at 50.

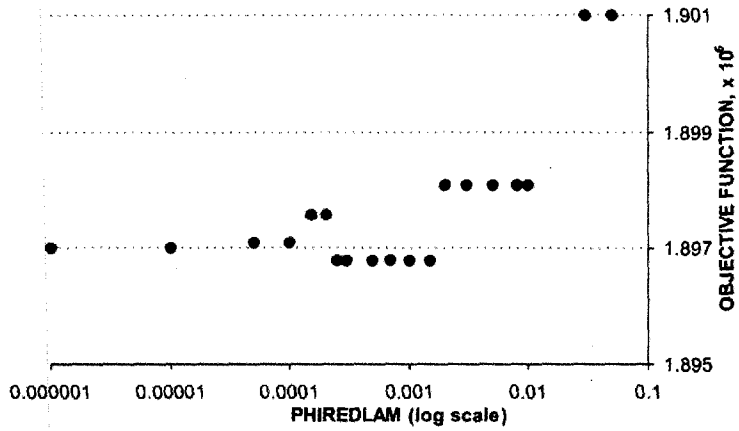


Figure 3.3 Impact of variable PHIREDLAM on the objective function

Table 3.4 lists values of GML variables that have been used in the current research. Other GML variables were found to not play a significant role in the optimization process and their values were taken as suggested in the PEST manual *(Doherty, 2004).*

Table 3.4 Best GML variable values for HSPF calibration

PEST variable	Description	Best value
RLAMBDA1	Initial Marquardt lambda	3
RLAMFAC	Factor by which the lambda is adjusted	2
PHIRATSUF	Sufficient ratio of the OF	0.3
FACPARMAX	Factor change of parameter that is allowed in one optimization iteration	4
PHIREDLAM	Maximum relative reduction of the OF between two consecutive lambdas	0.001
NUMLAM	Number of lambdas that PEST can test during any one optimization	10
NORTMAX	Maximum number of PEST iterations for a particular optimization run	50

3.4.3 Optimization of RSM variables

The RSM method adds certain randomness to the GML procedure. Although the aforementioned statements about GML variables are valid for RSM, their impact is significantly less. The GML method initiates a single optimization run from the location prespecified by the user. If the initial

parameter values are poorly chosen, the GML method can be trapped in local minima and stop performing. Contrary to the GML method, RSM generates many parameter sets, the best of which are used as starting points for new optimization runs. The number of pre-inversion random runs (NPIR) is supplied by the user. Although it may appear that increasing the parameter sets from which to choose will lead to better optimization, Figure 3.4 shows that no obvious correlation between NPIR and the objective function exists. The explanation lies in the fact that RSM chooses the next parameter set as maximally distant from the parameter trajectory of the previous parameter set, trying all "regions of attraction" in the parameter space. That is why NPIR does not significantly impact the optimization process. However, the experiments suggest NPIR should not be less than 20. *Doherty (2004)* suggests NPIR to be 4 to 10 times the number of adjustable parameters.

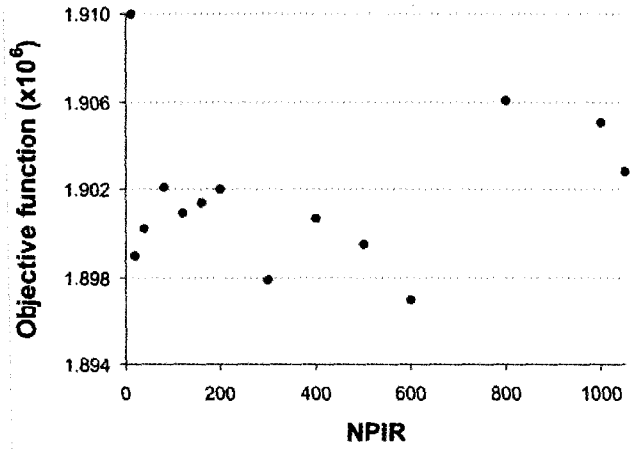


Figure 3.4 Impact of the number of RSM pre-inversion runs on the objective function

It was noted that approximately 4-8 PEST inversion runs are required to minimize the objective function from a preselected parameter set. This implies that the number of PEST inversion runs should be at least 10 to allow for completion of the search to the possible minimum. The number of runs with no improvement in objective function was set to 5. The improvement fraction in the objective function to be judged negligible depends on the value of the objective function. Improvement less than 0.1% was judged to be negligible. Key variable values used for RSM optimization are listed in Table 3.5.

Table 3.5 RSM variable values used for HSPF calibration

Description	Optimum value
Number of pre-inversion random runs (NPIR)	20
Maximum PEST inversion runs to be performed (MINR)	10
Number of runs with no improvements in the objective function (NRNI)	5
Improvement fraction in OF judged to be negligible (NOF)	0.001

3.4.4 Optimization of SCE-UA variables

Although the SCE-UA method has fewer variables than other methods, experience with the method has indicated that the effectiveness and efficiency of the algorithm are influenced by the choice of variables *(Duan et al., 1994)*. The variables that control the execution of the SCE-UA method are specified in Table 3.6. The maximum number of runs before optimization is terminated (MAXN) should be set high enough so that optimization is completed before MAXN runs are performed. Because 15 000 – 20 000 runs are, in most cases, performed for optimizing 11 parameters, MAXN was chosen to be 50 000. Termination criteria KSTOP and PCENTO are equivalent to the number of runs with no improvements in the objective function (NRNI) and the improvement fraction in the objective function judged to be negligible (NOF). In most cases, KSTOP should be set to 5. However, in some cases slight improvement in the objective function with KSTOP =10 was observed at a cost of almost double computational time.

Table 3.6 SCE-UA variables for optimization of 12 HSPF parameters

Variable	Description	Value
MAXN	Maximum number of runs before optimization is terminated	50,000
KSTOP	Number of shuffling loops in which OF should improve by the specified PCENTO	5
PCENTO	Percentage by which OF must change in KSTOP loops. Otherwise the optimization is terminated	0.001
NGS	Number of complexes used for optimization search	20
NPG	Number of points in each subcomplex	13
NSPL	Number of evolution steps taken by each complex before shuffling	13

Table 3.7 shows the impact of PCENTO on the objective function. Improvement in the objective function is barely visible, which determines that it is not practical to set PCENTO below 0.001; computational time increases significantly. The number of complexes used for optimization search (NGS) depends on the number of parameters to be optimized. *Doherty (2004)* suggests NGS can vary from 2 to 20. It was found that a value of 20 performs well. The number of points

in each subcomplex (NPG) is chosen to be the number of adjustable parameters +1 as suggested in *Doherty (2004)*. Default values of the number of evolution steps taken by each complex before shuffling (NSPL), which are equal to NPG, work well.

Table 3.7 Impact of termination criterion PCENTO on the objective function

PCENTO	Objective function
1	$1.9500\times10^{\square}$
0.1	$1.9300\times10^{\square}$
0.01	$1.8990\times10^{\square}$
0.001	$1.8965\times10^{\square}$
0.0001	$1.8959\times10^{\square}$
0.00001	$1.8958\times10^{\square}$

Advantages of the SCE-UA method are that it requires minimum user intervention and logically selected variables perform well in most cases. As a direct search method that relies only on function evaluation, the SCE-UA method does not require computation of derivatives and matrix inversion. Although it is significantly lengthier (an optimization run can last 1-2 days on an Intel's Pentium IV) than the GML method and RSM, the SCE-UA method saves time in the long run by not attempting numerous variables in the GML method; additionally, frustration with matrix inversion is avoided.

3.4.5 Results of hydrologic al calibration

Calibration of HSPF was started by first adjusting hydrological parameters, then snow parameters, and finally sediment and water quality parameters. The calibration procedures followed for HSPF were those described in *Donigian (2002)*. This paper focuses on hydrological and snow calibration. Daily flows and snow depths from January 1990 to December 1998 were calibrated for SNR.

More than 30 parameters are used in HSPF for water budget calibration in a pervious land segment. Some of these parameters are fixed for a period of modeling and some vary monthly. A number of HSPF parameters can be obtained from land use maps, high resolution topographical maps, and geographical information system or field surveys. Other parameters should be estimated from climatological conditions, geology, soil types, and literature. The lower and upper limits for the calibrated parameters were taken from literature and from *U.S. EPA (2000)*. Table 3.8 shows the optimized water budget parameters estimated using different methods, the values of

which are similar. Moreover, the estimated parameter values are similar to the range that other researchers such as *Zarriello (1999)*, *Senior and Koerkle (2003)*, and *Hoyer and Larson (2005)* obtained during HSPF calibration.

Table 3.8 Water budget parameters estimated by different methods

Parameter	Parameter range	GML	RSM	SCE-UA
LZSN (in.)	2.00-10.00	3.042	3.084	2.927
INFILT (in.)	0.01-0.50	0.079	0.079	0.081
AGWRC	0.85-0.99	0.937	0.936	0.937
DEEPFR	0.01-0.50	0.287	0.274	0.286
BASETP	0.01-0.20	0.198	0.200	0.999
AGWETP	0.01-0.20	0.010	0.010	0.010
CEPSC (in.)	0.01-0.40	0.010	0.050	0.097
UZSN (in.)	0.05-2.00	2.000	2.000	1.998
NSUR	0.01-0.50	0.500	0.500	0.499
IRC	0.30-0.85	0.567	0.569	0.572
LZETP	0.10-0.90	0.382	0.398	0.382

Performance comparisons and values of different objective functions are shown in Table 3.9. It is evident that the objective functions are very close to each other, although the SCE-UA method achieved better error minimization mostly attributed to better simulation of the exceedance times (OF3). The differences in statistics describing the accuracy of the simulation are minimal.

Table 3.9 Performance comparison of different optimization methods

Methods	CE	NS	DRMS	IA	OF1	OF2	OF3	OF1+OF2+OF3
GML	0.6193	0.755	205.2	0.803	1.38×10^{6}	4.38×10^{5}	7.63×10^{4}	1.8980×10^{6}
RSM	0.6188	0.755	205.21	0.803	1.38×10^{6}	4.37×10^{5}	7.59×10^{4}	1.8970×10^{6}
SCE-UA	0.6192	0.756	205.22	0.804	1.38×10^{6}	4.39×10^{5}	7.31×10^{4}	1.8962×10^{6}

Daily flows and monthly volumes were of primary interest in hydrological calibration. Table 3.10 shows performance comparison statistics between the observed and modeled data for the GML method. The values are satisfactory. Obviously, the performance statistics for monthly volumes are better than those for daily flows

Table 3.10 Performance comparison statistics for daily flows and monthly volumes from 1990 to 1998 using GML

	r^2	ME	PME	MAE	PMAE	DRMS	NS	CE	IA
Daily flows	0.76	11.8	6.1	82.5	42.43	205.2	0.755	0.619	0.803
Monthly volumes	0.83	3.1×10^7	6.05	1.3×10^8	25.07	2.2×10^8	0.905	0.729	0.860

In addition to the statistical analysis, graphical comparison of the percent exceedance flow was performed (Figure 3.5). High- and middle-range flows are reasonably well-calibrated; however, extremely low flows (<10 ft^3/s) are underpredicted. It is worth noting that it is almost impossible to obtain equally good fit of observed and simulated flows throughout the entire range of observations. In defining the objective function, one usually has to sacrifice low or high flows.

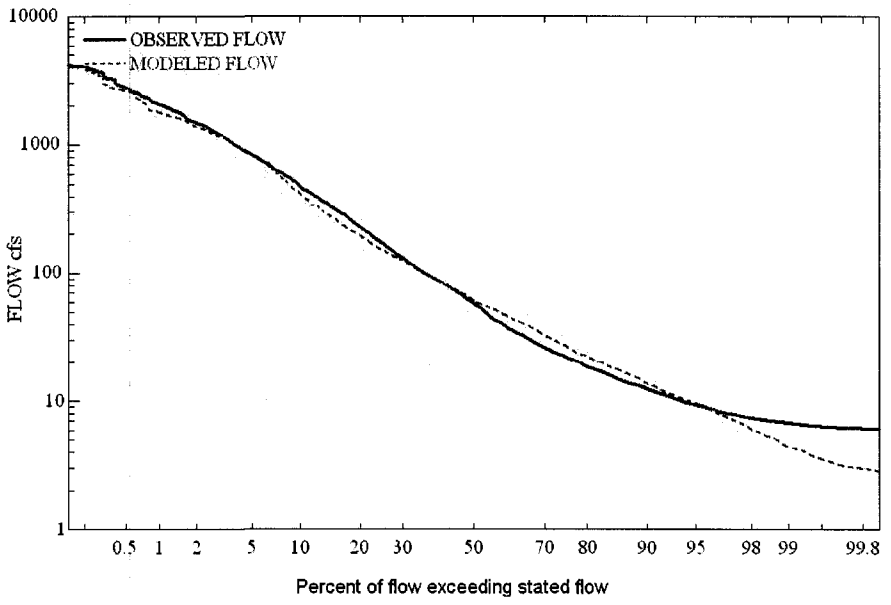


Figure 3.5 Observed and modeled exceedance curve (cumulative frequency distribution) of daily flow for SNR from 1990 to 1998 (Reach #8)

The parameter correlation coefficient matrix is presented in Table 3.11. Most parameters are poorly correlated. The highest negative correlation is between LZETP and the fraction of groundwater inflow to the deep recharge (DEEPFR). This can be explained by the fact that, in a way, LZETP and DEEPFR are competing for the moisture in the lower zone; the moisture either goes to evapotranspiration or it is lost to a deep aquifer, which is why an increase in value of one parameter leads to a decrease in another.

Some positive correlation was observed between the groundwater parameters AGWETP and AGWRC. This is quite natural because higher AGWRC values indicate slow groundwater recession rates and, subsequently, more potential for direct evaporation from groundwater storage.

Table 3.11 Parameter correlation coefficient matrix

	INFILT	AGWRC	DEEPFR	BASETP	AGWETP	CEPSC	UZSN	NSUR	IRC	LZETP
LZSN	−0.29	−0.43	−0.73	0.09	−0.28	−0.29	−0.61	−0.02	−0.35	0.57
INFILT	1	−0.25	0.24	0.01	0.01	0.17	−0.31	−0.06	−0.39	−0.15
AGWRC		1	0.59	−0.23	0.65	−0.02	0.24	0.21	0.50	−0.39
DEEPFR			1	−0.49	0.57	0.18	0.37	0.11	0.46	−0.82
BASETP				1	0.01	0.05	0.04	−0.08	−0.11	0.17
AGWETP					1	−0.13	0.33	0.05	0.29	−0.78
CEPSC						1	0.01	0.02	−0.01	0.02
UZSN							1	−0.15	0.2	−0.56
NSUR								1	0.11	0.01
IRC									1	−0.41

Ideally, all parameters should be independent and not correlated with each other. Generally, as the number of correlated parameters rises, model calibration becomes more difficult. Correlated parameters can move upward or downward in unison having almost no impact on the objective function.

During calibration, OF1, OF2, and OF3 were minimized simultaneously and the compound objective function (OF1 + OF2 + OF3) was recorded. Examples of shapes of the compound objective function for different pairs of hydrological parameters are presented in Figures 3.6 and 3.7. Twelve thousand model runs and LOESS with tricube weighting and polynomial regression were used to generate a surface. It can be seen that the surface of the objective function is complex with many local minima. The GML technique, as a LOM, can be trapped in these minima. To ensure that the global minimum was achieved, many optimization runs with different initial parameter values, various lambdas, and different factors of changing lambda were used. Six hydrological parameters in HSPF are allowed to vary monthly. Two main approaches for estimation of monthly values were explored. The simple and faster approach is to assume that monthly values have a sinusoidal form and to ensure that the peak occurs during the appropriate season.

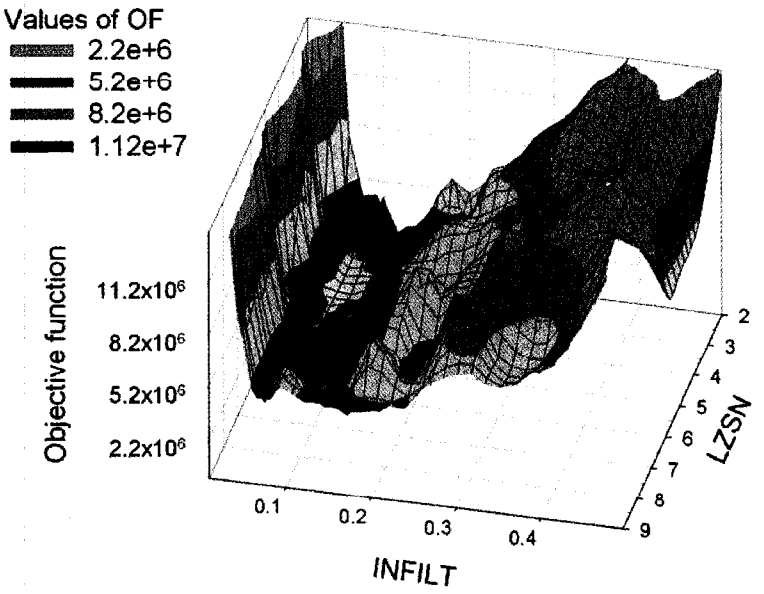


Figure 3.6 **Surface of compound objective function in the parameter space of INFILT and LZSN**

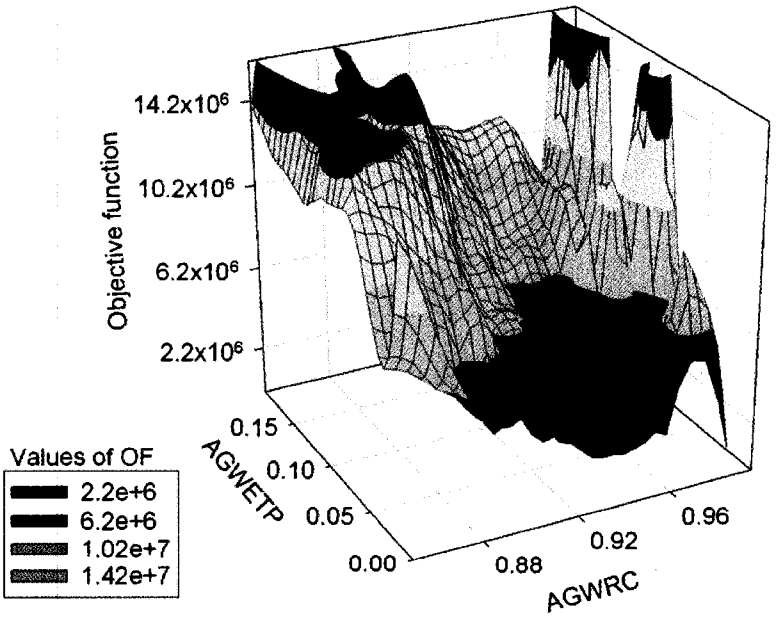


Figure 3.7 **Surface of the compound objective function in the parameter space of AGWETP and AGWRC**

The nonlinear parameter estimation tool was used to estimate the mean value and amplitude of variation. Then, monthly values were written to HSPF by parameter preprocessor PAR2PAR (Parameter Estimation Tool, 2003). The second approach is to estimate each monthly value independently. Obviously, the latter approach requires more computations and time. Both approaches were tried and the second was found to be superior in terms of minimizing the objective function and improving performance indicators. All three optimization methods achieved good results in minimizing the objective function. Slightly better results were obtained with the SCE-UA method. Table 3.12 shows lower and upper bounds of the monthly varied HSPF parameters. Interflow inflow parameter INTFW is a monthly varied parameter that divides water between interflow and direct surface runoff. It affects the timing of runoff, reducing or decreasing peaks maintaining the same volume.

Table 3.12 Lower and upper bounds of monthly-varied water budget parameters

Parameter	Lower bound	Upper bound
CEPSC	0.01	0.1
UZSN	0.1	2.0
NSUR	0.3	0.5
INTFW	3.7	10.0
IRC	0.58	0.66
LZETP	0.1	0.9

3.4.6 Snow calibration

Hydrological Simulation Program Fortran has more than 20 parameters in its snow modeling subroutine that simulates accumulation and melting of snow and ice. Some parameters, such as latitude of watershed segment, mean elevation, temperature at which precipitation becomes snow, and so on, are straightforward and do not require calibration. Based on a review of the literature and some sensitivity tests, six parameters were singled out to be adjusted during calibration. These parameters, as well as initial, minimum, maximum, and optimized values, are shown in Table 3.13. All calibrated snow parameter values are close to those found in literature by Crawford (1999), Hoyer and Larson (2005), Zarriello (1999), and Senior and Koerkle (2003).

Table 3.13 Snow parameters estimated by different methods

Parameter	Initial value	Possible range of values	GML	RSM	SCE-UA
SHADE	0.4	0.0-0.8	0.727	0.705	0.662
SNOWCF	1.2	1.0-2.0	1.078	1.085	1.096
COVIND (in)	3	0.1-10	4.094	4.125	4.139

RDCSN	0.1	0.05-0.3	0.062	0.3	0.124
CCFACT	1	0.5-2.0	0.582	0.597	0.584
MWATER (in/in)	0.1	0.005-0.2	0.2	0.2	0.196

Table 3.14 provides correlation coefficients for snow parameters. Condensation/convection melt factor (CCFACT) and liquid water storage capacity (MWATER) show the highest correlation, which is logical. No difficulties were observed with matrix inversion in the GML method for snow parameters.

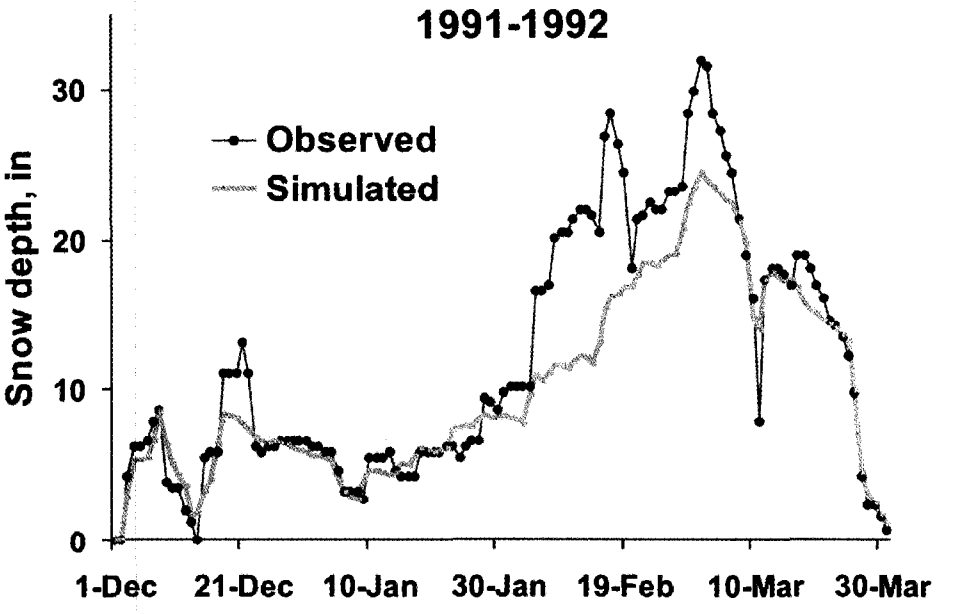
Table 3.14 Correlation coefficient matrix of snow parameters

Parameter	SNOWCF	COVIND	RDCSN	CCFACT	MWATER
SHADE	−0.37	−0.49	−0.09	−0.47	0.03
SNOWCF	1	0.54	−0.02	0.49	0.04
COVIND		1	−0.03	0.53	0.14
RDCSN			1	0.05	0.23
CCFACT				1	0.59

Results of snow depth simulation for the Castor River subwatershed taken upstream of the South Nation are presented in Figure 3.8. Observed snow depths were obtained from a nearby airport meteorological station in Ottawa. Although there are some discrepancies on a rising limb of the graphs, in general, the predictions are close enough to the observed data. Comparative statistics of observed and simulated snow depths for winter seasons from 1990 to 1997 are presented in Table 3.15. For this particular climate zone, the winter season was assumed to commence on November 1 and to end on March 31. On average, the model underpredicts snow depth in five out of seven seasons, although the total simulated snow storage is close to that observed. Comparison statistics indicate that the model predicts relatively well the snow depths throughout the period of investigation.

Based on performance statistics and a graphical comparison of modeled and observed daily flows as well as snow depth calibration, it was concluded that sufficient precision of hydrological calibration was achieved.

A manual process of calibration in watershed modeling is still widely used today. Similarity between model outputs and measurement data is evaluated by an expert using objective (different mathematical criteria) and subjective (visual comparison, experience, and intuition) methods.



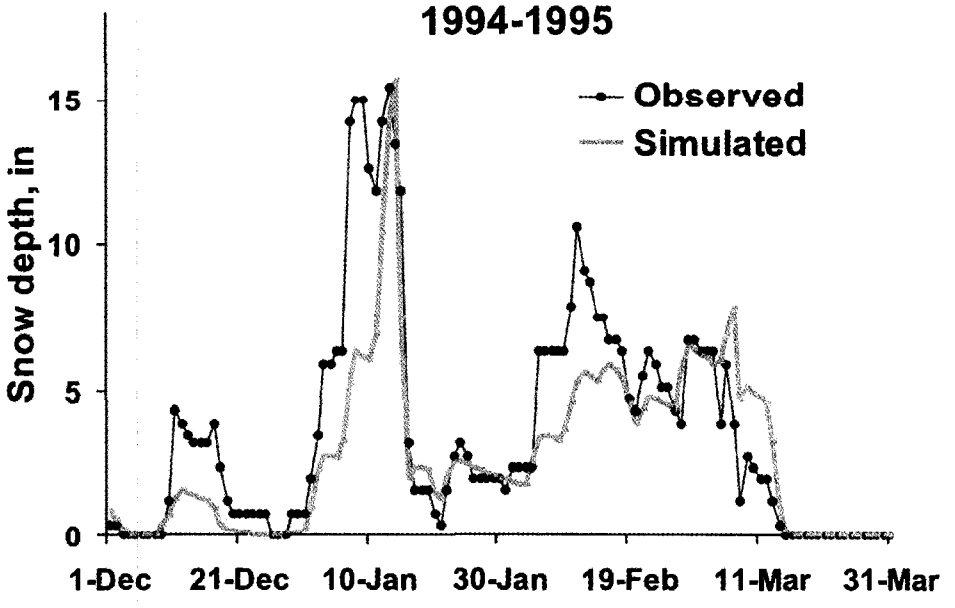


Figure 3.8 Observed and calibrated snow depths of two winter seasons

Table 3.15 Performance comparison statistics for daily snow depths

Winter season	r^2	ME	MAE	DRMS	NS	CE	IA
1990-1991	0.763	2.014	2.362	3.310	0.475	0.366	0.716
1991-1992	0.910	−1.525	1.952	3.399	0.849	0.737	0.855
1992-1993	0.923	−2.290	3.601	5.448	0.867	0.708	0.842
1993-1994	0.943	−1.521	1.818	2.611	0.901	0.755	0.872
1994-1995	0.743	−0.660	1.147	2.044	0.699	0.604	0.784
1995-1996	0.907	−2.168	2.550	3.711	0.814	0.667	0.821
1996-1997	0.668	1.766	2.554	4.608	0.483	0.546	0.795

Parameter adjustments are done by semi-intuitive trial and error processes. Although most models have documented guidelines for parameter adjustments, in reality the sequence is varied from expert to expert based on their experience, model understanding, available data, and watershed properties *(Boyle et al., 2000)*. Because of parameter intercorrelation, adjustment of the model may be a difficult process. Some parameters influence overall model behavior and some parameters are essential only for specific hydrological events. The expert must simultaneously evaluate several subjective and objective criteria while iteratively adjusting model parameters. For lumped watershed models with 15 or less parameters, the manual approach can give excellent results, albeit at the expense of considerable time and energy *(Gupta et al., 2003)*. Manual calibration procedures are model-specific and not easily transferable to other models. For example, HSPFEXP was compiled for HSPF *(Lumb et al., 1994)*. While use of HSPFEXP was attempted in the current research, it turned out to be very simplistic and time-consuming.

3.5 Discussion

In general, manual calibration is strongly dependent on user knowledge and expertise. Although invention of the graphic user interface has made manual calibration much easier, its execution is complicated, labor-intensive, and time-consuming. Still, the overall results of manual calibration may be good.

In contrast to manual calibration, automatic calibration is faster and simpler to apply. Methods of automated calibration are improving constantly, usually parallel to increasing computer power. The goal is to develop calibration procedures that result in the same (or better) model performance compared to that of a highly trained expert. Subjective human judgments are eliminated. Parameter evaluation and adjustments are objective and explicit rules are used to calibrate the model. The user typically specifies upper and lower limits for each parameter as well

as the initial parameter value. Another advantage of automatic calibration is that its procedures can be generalized for different models.

A hybrid approach that combines the strengths of manual and automatic calibration was proposed by *Boyle et al. (2000)*. A multicriteria formulation was used to imitate the evaluation techniques and strategies used in manual calibration, and the resulting optimization problem is solved using a computerized algorithm. The Sacramento Soil Moisture Accounting model was used as an example.

Although manual calibration helps to understand model equations and interparametric relations, it was not possible to achieve satisfactory results of HSPF manual calibration for SNR. As a result, various local and global optimization techniques were used and compared.

Local optimization methods (often called deterministic), which are widely used because of their speed and low requirement of computer power, are effective for simple, unimodal spaces. When there is only one minimum, the LOM is usually the fastest way to find it. However, in a real-life watershed problem LOMs can be trapped in non-optimum areas due to the presence of many local minima, flat surfaces, and so on. In this research, the GML method as a popular LOM was used.

After extensive application of the GML method for HSPF calibration, several observations were made. First, insensitive parameters can significantly deteriorate the optimization process. An insensitive parameter can dominate the parameter upgrade vector (whose magnitude is limited by FACPARMAX) resulting in much smaller changes for other, more sensitive parameters. As a result, the objective function changes little and is often not minimized at all. The problem is easily recognized by analyzing the PEST record file, which lists the parameter with maximum factor change per iteration. If its change is equal to FACPARMAX, this parameter most likely is insensitive and causes problems for optimization. Solutions to the problem can be to increase the FACPARMAX, change the parameter space, or, more preferably, fix the troublesome parameter. In the current application, CEPSC is a troublesome parameter because of its low sensitivity (Table 3.1). As a result, PEST constantly reports that the matrix is singular and inversion is not possible. Additionally, the objective function was not lowered efficiently. As soon as CEPSC was eliminated from the optimization process, the problem disappeared.

It was noted that the GML method is sensitive to the initial parameter values supplied by the user. If the initial value is located in the catchment area of a local minimum, then the GML method

finds a local minimum and efforts to find a global minimum will not be made. To avoid this error, it is a good practice to initiate several runs from different starting locations and analyze their performance.

An additional significant disadvantage of the GML method is matrix inversion. When the matrix is singular, the GML method still performs because of the Marquardt lambda; however, the risk that optimization will not reach the global minimum is high. A matrix becomes singular when columns (or rows) are linearly dependent. The easiest remedy is to fix one of the correlated parameters or to optimize them separately.

Shortcomings of deterministic local optimization algorithms, which are a result of the presence of multiple regions of attraction, multilocal optima, discontinuous derivatives, insensitivities, and parameter interdependency, led to wide application of global optimization methods. The global optimization methods applied in this work are RSM and SCE-UA.

The random multiple search method conducts a random parameter sampling, running PEST from different starting locations in a parameter space; additionally, RSM "learns from previous mistakes". Although the problem of matrix inversion is not avoided, RSM is superior to the GML method in that it tries many regions of attraction in the parameter space, thereby increasing chances of finding the global minimum. Ideally, RSM finds the global minimum passing through many local minima.

SCE-UA method has already been used for successful watershed model calibration *(Duan et al., 1994)*. It is characterized by increased robustness at a cost of significantly higher computational time. Noticeably more model runs are required. The GML optimization run lasts from half an hour to two hours on a Pentium IV at 2GHz. However, it is important to try several initial lambdas, different factors of parameter adjustments, and different initial parameter values. On average, 30-50 GML optimizations are usually sufficient for calibration of HSPF with 11 parameters taken at a time. Optimization runs for RSM can take several hours to a day; a SCE-UA optimization run may last 1-2 days.

As shown in Table 3.9, optimization results conducted by different methods are very similar, although, in most cases, RSM outperforms the GML method and the SCE-UA method outperforms RSM. However, the advantage of the SCE-UA method is not better performance. Rather, it is convenience for the user that makes RSM superior to the GML method and the SCE-UA method superior to RSM. The GML method can perform as well as the SCE-UA method

when the user takes time to manually adjust numerous variables, tries many different initial parameter values, and eliminates problematic parameters. On the contrary, the SCE-UA method has a minimum number of simple variables and a user with little knowledge can obtain good calibration.

In addition to optimization techniques, there are many factors beyond the scope of this work that contribute to parameter sensitivity and, ultimately, to the calibration results. Climate, geology, topography, and land cover of the watershed are represented in the model in a simplified fashion. Changing the "fixed" parameters such as slopes, vegetation, roughness, and groundwater depth obviously would lead to an altering of the interaction between the parameters and, as a result, to different parameter values. Some parameter uncertainty stems from the fact that several meteorological time series used in HSPF cannot fully describe the climate conditions in the basin. Time step of meteorological observations often is not sufficient. Moreover, the meteorological stations often are several dozen kilometers away from the subwatershed where the modeling is being conducted. Although it is important to acknowledge these factors impacting the calibration process, the primary objective of the present work was to compare different optimization techniques assuming the climate and watershed characteristics are properly defined and fixed for a research period.

3.6 Conclusions

Sensitivity analysis has shown that INFILT and AGWRC are the most sensitive parameters and that CEPSC is the least sensitive. High sensitivity of INFILT stems from its physical nature of controlling the division of precipitation to surface and subsurface storages. CEPSC was practically insensitive due to the low forest ratio in the watershed. Some parameters are sensitive to one objective function and insensitive to another. For example, AGWETP is very sensitive to exceedance times, but relatively insensitive to daily flows. AGWRC is mostly sensitive to monthly volumes and exceedance times due to its regulation role of seasonal and low-flow timing. It was observed that parameters can increase or decrease their sensitivity by several times during the optimization run. This is explained by the complex surface of the objective function in a parameter space.

The range of parameter values that can maintain the model in a calibrated state depends on the number of adjustable parameters. When more parameters are optimized simultaneously, a wider range exists for a "suitable" parameter. For example, the range of suitable LZSN values increases two-fold when 11 parameters are optimized simultaneously instead of just one. This may be

explained by increased parameter intercorrelation. The correlated parameters can move upward or downward in unison, having almost no impact on the objective function.

The impact of some GML, RSM, and SCE-UA variables on the ability to find the global minimum was studied. The best GML, RSM, and SCE-UA variable values beyond which objective function improvement is insignificant were suggested. It was noted that the GML method is very sensitive to initial Marquardt lambda, factor of lambda adjustment, and initial parameter guess. Many GML optimization runs with different variables should be conducted to ensure that the global minimum is found. Problems with parameter intercorrelation, insensitivity, and matrix singularity are common.

The random multiple search method represents a significant improvement to the GML method in that it adds randomness into initial parameter selection as well as a mechanism for making a smart choice about the starting locations for optimization runs. Moreover, RSM guarantees that all regions of attraction within the parameter space are tested at various stages of the optimization process, although problems with matrix inversion still occur.

Contrary to the GML method and RSM, SCE-UA is a very robust method and convenient to use. The genetic algorithm and shuffling of the SCE-UA method help to avoid trapping in local minima. In most cases, logical definition of key variables leads to the global minimum. Additionally, user intervention is limited and problems with matrix inversion and insensitivity do not occur.

All three methods produced good results and proved to be more efficient than manual HSPF calibration. The GML method is very fast; it optimizes 11 parameters in less than 2 hours on a Pentium IV with 2.0 Ghz. The GML method can perform as well as the SCE-UA method if the GML variables are properly adjusted, the initial guess is good, and insensitive parameters are eliminated from the optimization process. An optimization run for RSM can take several hours to a day, while an SCE-UA optimization run may last 1-2 days.

3.7 Acknowledgments

Extensive, valuable comments by anonymous reviewers contributed substantially to the quality of the research presented herein.

3.8 References

Bicknell, B. R.; Imhoff, J. C.; Kittle, J. L., Jr.; Jobes, T. H.; Donigian, A. S., Jr. (2001) *Hydrological Simulation Program Fortran. Version 12. User's Manual;* U.S. Environmental Protection Agency: Athens, Georgia.

Boyle, D. P.; Gupta, H. V.; Sorooshian, S. (2000) Toward Improved Calibration of Hydrologic Models: Combining the Strengths of Manual and Automatic Methods. *Water Resour. Res.,* **36** (12), 3663–3674.

Crawford, N. H. (1999) *Hydrologic Journal—Snowmelt Calibration*; Hydrocomp, Inc., www.hydrocomp.com .

Doherty, J. (2004) *PEST—Model Independent Parameter Estimation. User Manual,* 5th ed.; Watermark Computing: Corinda, Australia.

Doherty, J.; Johnston, J. M. (2003) Methodologies for Calibration and Predictive Analysis of a Watershed Model. *J. Am. Water Resour. Assoc.,* **39** (2), 251–265.

Donigian, A. (2002) Watershed Model Calibration and Validation: The HSPF Experience. *Proceedings of the Water Environment Federation National Total Maximum Daily Load Science and Policy conference* [CD-ROM]; Phoenix, Arizona, Nov. 13–16; Water Environment Federation: Alexandria, Virginia.

Duan, Q. Y.; Sorooshian, S.; Gupta, V. K (1992) Effective and Efficient Global Optimization Method for Calibrating Watershed Models. *J. Hydrol.,* 158, 265–284.

Duan, Q. Y.; Sorooshian, S.; Gupta, V. K. (1994) Optimal Use of the SCE-UA Global Effective and Efficient Global Optimisation [sic.] for Conceptual Rainfall–Runoff Models. *Water Resour. Res.,* 28, 1015–1031.

Gupta, H. V.; Sorooshian, S.; Hogue, T.; Boyle, D. P. (2003) Advances in Automatic Calibration of Watershed Models. In *Calibration of Watershed Models;* Duan, Q., Sorooshian, S., Gupta, H. V., Rousseau, A. N., Turcotte, R., Eds.; American Geophysical Union: Washington, D.C.

Hashino, T.; Bradley, A. A.; Schwartz, S. S. (2002) *Verification of Probabilistic Streamflow Forecasts.* Report No. 427; Hydroscience & Engineering and the University of Iowa: Iowa City, Iowa.

Haughton, J. (2002) *A State of the Environment Report for the South Nation River Watershed.* Environmental Science Program, University of Ottawa, Ottawa, Ontario, Canada.

Hoyer, D. P.; Larson, A. M. (2005) *Belle Fourche River Watershed Assessment*. Final Report and Total Maximum Daily Load for Butte, Lawrence, and Meade Counties, South Dakota; Rapid City, South Dakota.

Legates, D. R; McCabe, G. J., Jr. (1999) Evaluating the Use of "Goodness-of-Fit" Measures in Hydrologic and Hydroclimatic Model Validation. *Water Resources Research,* **35**, 233–241.

Lumb, A. M.; McCammon, R. B.; Kittle, J. L. (1994) *User's Manual for an Expert System (HSPFEXP) for Calibration of the Hydrologic Simulation Program-Fortran.* Water-Resources Investigations Report 94-4168; U.S. Geological Survey: Reston, VA; p 102.

Nash, J. E.; Sutcliffe, J. V. (1970) River Flow Forecasting through Conceptual Model, Part I, A Discussion of Principles. *J. Hydrol.*, **10** (282–290), 1070.

Nelder, J. A.; Mead, R. (1965) A Simplex Method for Function Minimization. *Comput. J.,* **7**, 308–313.

Parameter Estimation Tool (2003) *Surface Water Utilities.* Prepared by Watermark Numerical Computing: Brisbane, Australia and the University of Idaho: Moscow; p 122.

Senior, L.A.; Koerkle, E.H.(2003) *Simulation of Streamflow and Water Quality in the Red Clay Creek Subbasin of the Christina River Basin, Pennsylvania and Delaware, 1994-98*. Water-Resources Investigations Report 03-4138; U.S. Department of the Interior and U.S. Geological Survey: New Cumberland, Pennsylvania.

U.S. Environmental Protection Agency (2000) *Estimating Hydrology and Hydraulic Parameters for HSPF.* BASINS Technical Note 6; U.S. Environmental Protection Agency: Washington, DC.

Willmott, C. J. (1981) On the Validation of Models. Phys. Geog., **2**, 184–194.

Zarriello, P. (1999) *A Precipitation-Runoff Model for Part of the Ninemile Creek Watershed near Camillus, Onondaga County, New York*. Water-Resources Investigations Report 98-4201; U.S. Geological Survey: Ithaca, New York

CHAPTER 4

PARAMETER UNCERTAINTY OF A WATERSHED MODEL

Igor Iskra and Ronald Droste
Department of Civil Engineering, University of Ottawa
161, Louis Pasteur, Ottawa, K1N 6N5, Canada

The novel contribution of this paper is in estimation and comparison of parameter uncertainty of the HSPF using the method of moments, Monte Carlo with LHS and induced correlation, and response surface methods. The parameter uncertainty was expressed as confidence intervals and a probability distribution function. Impact of parameter uncertainty on maximum spring flows was studied.

4.1 Abstract

Parameter uncertainty of the Hydrologic Simulation Program Fortran (HSPF) was the focus of this work. The model was calibrated using data from the South Nation watershed located in Eastern Ontario. Up to 18 parameters were used in independent model calibrations. Three uncertainty methods were used: the method of moments, Monte Carlo (MC) with LHS and induced correlation, and response surface methods. It was observed that as more parameters are optimized simultaneously typically higher uncertainty is observed. The method of moments shows the most conservative uncertainty. The largest uncertainty typically was estimated by the MC method. The 95% confidence intervals of allowable parameter uncertainty correspond to up to 10% variations in spring maximum flows.

Résumé

L'incertitude de paramètre du programme hydrologique de simulation (HSPF) était l'objectif principal de ce travail. Le modèle a été calibré en utilisant des données du bassin de South Nation situé dans l'est de l'Ontario. Jusqu'à 18 paramètres ont été employés indépendamment dans les calibrages modèles. Trois méthodes d'incertitude ont été employées: la méthode de moments, Monte Carlo (MC) avec le LHS et corrélation induite, et des méthodes surfaces de réponse. Lorsque plus de paramètres sont optimisés simultanément, on observe généralement une incertitude plus élevée. La méthode de moments montre l'incertitude la plus conservatrice. La plus grande incertitude a typiquement été estimée par la méthode de MC. Les intervalles de confiance de 95% du paramètre d'incertitude correspondent aux variations de 10% des écoulements maximums printemps.

4.2 Introduction

Probably the earliest justification of uncertainty analysis was made by *Aristotle (350 BC)* in his Nicomachian Ethics. He wrote "...it is the mark of an educated mind to rest satisfied with the degree of precision that the nature of the subject admits, and not to seek exactness when only an approximation is possible". Processes in nature are stochastic by their nature and therefore never perfectly certain. Any model is based on many assumptions and simplifications, including parameterization of different physical, chemical, and biological processes. The current research aims to look further into various aspects of uncertainty in a watershed model.

Watershed models are becoming an everyday routine for watershed planning, development, and management. Most watershed models are fairly complex; interactions among watershed components are essentially nonlinear, involve a range of physico-chemical processes, and take place on a wide range of temporal and spatial scales. Each of these factors individually and the system as a whole are subject to uncertainties. Uncertainty of the system is a probabilistic combination of the component uncertainties. Individual uncertainties (due to data errors, model assumptions, natural variability, etc.) must be studied in order to identify their relative contribution. Strengthening of weak components could be the most effective means of uncertainty reduction of the system.

4.2.1 Types of uncertainties

Uncertainty refers to lack of knowledge about specific factors, parameters, or models. Uncertainty analysis is aimed to quantify the level of confidence the modeler has in the inputs, the parameters, the model structure, and as a result in the model outcomes. A number of uncertainty classifications are proposed in the literature (*USEPA, 2002; Yen, 2002; Willems, 2000*). Although they have the same components, their groupings are different. In this paper three main types of uncertainty are differentiated: model uncertainty, input data uncertainty, and parameter uncertainty (Figure 4.1). They are based on the different physical natures of the uncertainties.

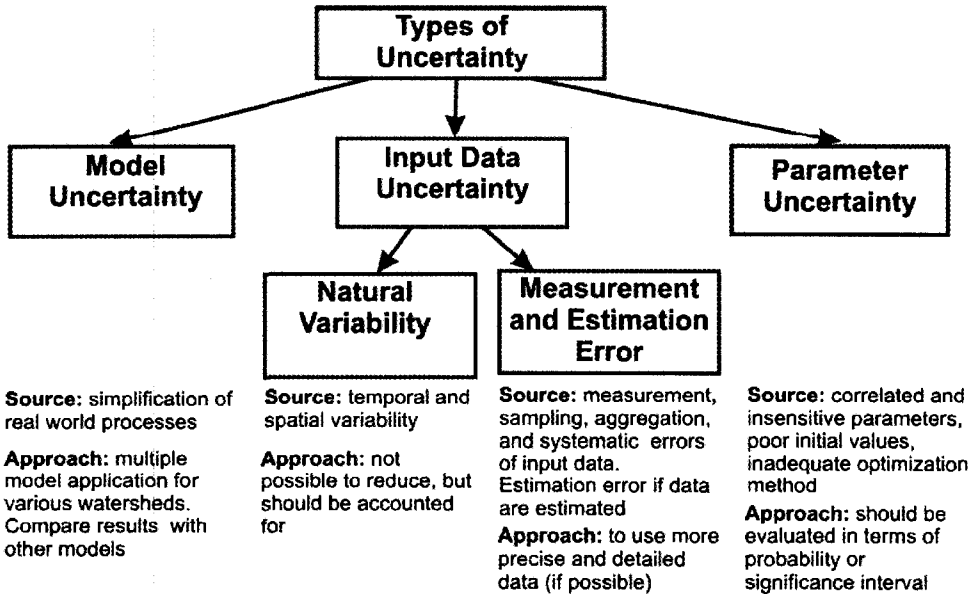


Figure 4.1 Types of uncertainty

Model uncertainty represents model structure with equations, assumptions, and the modeler's vision of the reality. Since any model is a simplification of the real physical world, uncertainty associated with the model structure is unavoidable. This uncertainty does not vary in time since the model equations do not change. An approach to estimate model uncertainty may consist of multiple model applications for various watersheds and comparisons with other watershed models. Model uncertainty can be considered for an individual subroutine or can be described in a lumped approach as a total model uncertainty of the output variables. An interesting example of model uncertainty decomposition is discussed by *Timbe and Willems (2004).*

Model input contains directly measured or estimated data which vary for different watersheds. Input data uncertainty can be caused by temporal and spatial variability in natural processes or by measurement, sampling, and estimation errors. Model input can be variable in time (e.g., meteorological time series) or relatively time invariant (e.g., river bed geometry, landuse, DEM).

Actually almost all factors and processes affecting flow routing in a watershed model are subject to natural stochastic variability. There is always uncertainty in the description of channel geometry, roughness, specification of sources and sinks, and initial and boundary conditions *(Singh, 2004)*. The stochastic variability is caused by heterogeneities, nonuniformities, random irregularities, and different types of errors in input data. Input data uncertainty can be quite high. For example, *Willems (2000)* reported that spatial variation in rainfall alone input accounts for approximately 30% of overall variability in model results. Natural variability of data is a stochastic component and cannot be reduced but should be accounted for. Measurement and sampling errors exist in every watershed model and depend on quality of data. For a realistic estimate of the input uncertainty a modeler should know instruments used for the measurements, the sampling methodology, sampling frequency, how and where samples are stored before analysis, and other similar information.

Model parameters are different coefficients and exponents which often need adjustment by model calibration. Model performance is highly sensitive to some parameters and relatively insensitive to others, although a margin between model input and model parameter is not always visible, in general, model parameters are referred to model characteristics which are adjustable by a modeler during calibration *(Duan et al., 2003)*. Estimates of parameters are generally uncertain because data used for calibration are uncertain and the model never exactly represents the natural system. Parameter uncertainty can be caused by correlated and insensitive parameters, poor initial and

boundary parameter values, and inadequate optimization methods. The present paper only deals with parameter uncertainty.

4.2.2 Measures of uncertainties

Uncertainty can be expressed as a single value, an interval, or probability distribution (Figure 4.2). When a single value is used the second statistical moment (variance or standard deviation) as a measure of the dispersion of the variable is often a choice. Variance is easy to manipulate mathematically; however, due to its squared units it is difficult to compare it directly with the observations. Standard deviation is expressed in the same units as the studied characteristics. The coefficient of variation is a convenient normalized measure for comparison uncertainties of different variables. By themselves, point estimates do not fully portray the variability of the system and usually more sophisticated expressions are needed.

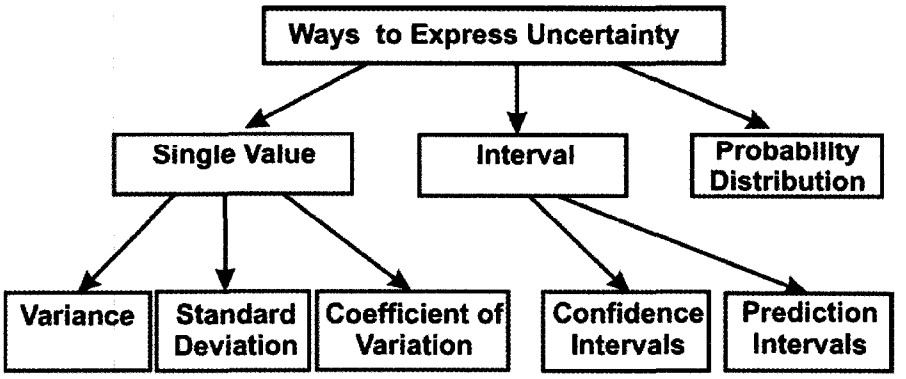


Figure 4.2 Ways of expressing the uncertainty

Confidence and prediction intervals are a convenient way to convey uncertainty. They show an explicit degree of confidence on a specified interval of variability. Confidence intervals are usually associated with a mean value and prediction intervals with an individual value *(Kirchner, 2001)*. The method of intervals has potential problems which should be accounted for. Firstly, most of the conventional procedures for determination of confidence intervals assume a normal distribution which is not always the case, especially when sample size is small. Secondly, confidence intervals of individual random components cannot be directly and easily combined to estimate the confidence interval of the system as a whole *(Yen, 2002)*.

The most complete and ideal description of uncertainty is the probability distribution function (PDF). The probability distribution is described by skewness, shape of distribution, characteristics of tails, and some other characteristics. Unfortunately, in most practical model applications the

PDF is difficult to find precisely *(Yen, 2002)*. A large number of measurements are necessary to approximate any one of the well-known distributions.

4.2.3 Distributions for parameter uncertainty

The basic assumption of uncertainty analysis is that the uncertainty in the value of interest is equal to the uncertainty of error or deviation in the value. Another assumption is that measurement error and parameter deviation in a calibration process are random variables that follow a statistical distribution *(Castrup, 2002)*.

The distribution of measurement errors and deviation of parameter values during the calibration can be assumed to be uniform, log-normal, triangular, normal, etc. In a uniform distribution, the probability of finding a value between the minimum and maximum boundaries is unchanged and equal to unity. The uniform distribution is easy to use and analyze and it may be appropriate for some limited applications. The log-normal distribution can be used to estimate uncertainty in parameters with asymmetric limits, for example, when one of the physical limits lies close enough to the nominal value. Uncertainty estimates for the log-normal distribution are obtained by numerical iterations *(Castrup, 2002)*. The triangular distribution can be of help when the central tendency exists and the containment probability is 100%. However, abrupt transitions at the boundary limits to zero are physically unrealistic in most cases. Moreover, the linear increase and decrease in behavior is not suitable for the majority of parameters.

The most widely used distribution for uncertainty analysis is the normal distribution. Firstly, because this distribution is the best approximation of many physical processes in the universe *(Castrup, 2002)* and secondly, the central limit theorem proves that even though individual random errors or deviations may not be normally distributed, the combined error is approximately normal *(Clarke, 1994)*. The normal distribution is a quite common assumption for uncertainty analysis *(Morgan, 1999; Chang et al., 1994; Castrup, 2002)*. In the present research the distributions of HSPF parameters are assumed to be normal.

4.2.4 Methods for uncertainty estimation

A number of methods exist for uncertainty estimation (Figure 4.3). They vary by the level of sophistication, computational complexity, and data requirements. The choice of method is dictated by the available information, needed precision, and available computer resources. In the ideal case the uncertainty can be characterized by an exact PDF of model outcomes as a function of the stochastic input parameters. However, in practical watershed applications due to

nonlinearity and complexity, the analytical derivation of a PDF is not often feasible. As an alternative, a number of methods attempt to approximate the statistical properties of uncertain model output.

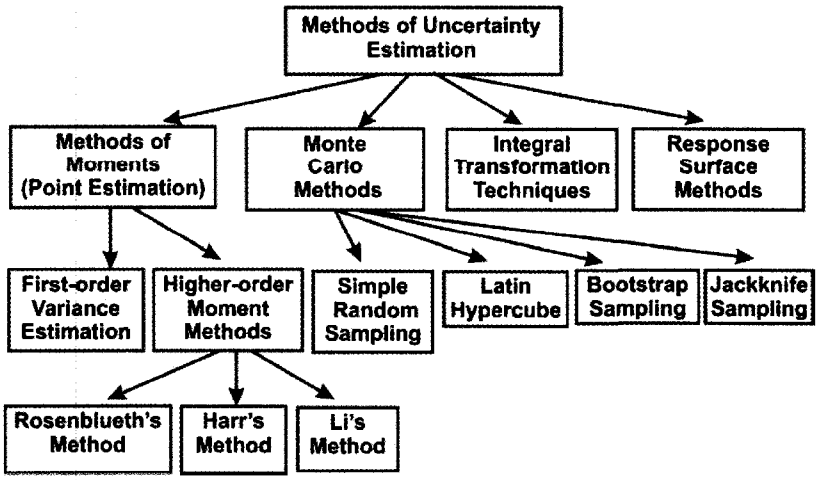


Figure 4.3 Methods of uncertainty estimation

4.2.4.1 Method of moments

The method of moments is a general technique for estimation of the first (mean), the second (variance or standard deviation), the third (skewness), and sometimes the forth (kurtosis) moments of a variable, based on various approximations to the function. The most common form of the method of moments is the first-order variance estimation (FOVE) that approximates the function by a flat plane tangent to the curved surface of the function at the mean value *(Kirchner, 2001)*. In the simplest case of a single-variable function $z = f(x)$, the variance of z (model outcome) is approximated by a product of squared slope (dz/dx) of the tangent line and a variance of x (model parameter). In a case with two variables, z is approximated by the tangent planes with slope $\partial z/\partial x$ in the x dimension and $\partial z/\partial y$ in the y dimension. The variance of z is calculated according to

$$Var(z) \approx \left(\frac{dz}{dx}\right)^2 \cdot Var(x) + \left(\frac{dz}{dy}\right)^2 \cdot Var(y) + 2\frac{\partial z}{\partial x}\frac{\partial z}{\partial y} Cov(x,y) \qquad (4.1)$$

where $Cov(x,y)$ is the covariance of x and y. If x and y are uncorrelated, the third term in Eq. 4.1 disappears and the formula becomes the Gaussian error propagation formula *(Kirchner, 2001)*. In a multivariable case with the assumption that all variables are uncorrelated the variance of z is equal to

$$Var(z) \approx \left(\frac{dz}{dx}\right)^2 \cdot Var(x) + \left(\frac{dz}{dy}\right)^2 \cdot Var(y) + ... + \left(\frac{dz}{dw}\right)^2 \cdot Var(w) \qquad (4.2)$$

The FOVE is a straightforward method which does not require knowledge of the PDF of the input parameters. The only requirement is that the function should be differentiable.

The higher order moment methods are capable of taking into account the higher moments of the involved random variables. Among the most popular is *Rosenblueth's (1975)* method which can be used to estimate statistical moments of any order with the correlated or uncorrelated input variables *(Harr, 1987)*. The number of point estimates of Rosenblueth's method is 2^N, where N is the number of random variables. The difficulty of obtaining point estimates in Rosenblueth's method increases exponentially with N. To address this issue, *Lind (I983)* and *Harr (1989)* developed alternative probabilistic methods which require only $2 \times N$ model estimates. *Li (1992)* has proposed a computational point estimate method that allows incorporation of the first four moments of individual random variable and the correlation among the variables. The number of model runs required is expressed as $(N^2 + 3 \times N + 2)/2$. *Singh (2004)* looked at application of Li's method for a flow routing model and found it to be efficient and accurate.

The general shortcoming of the point estimate methods is that the number of points depends on the number of variables.

4.2.4.2 Monte-Carlo (MC) methods

Monte-Carlo (or simulation) methods are popular among watershed modellers. These methods rely on direct model simulation and they do not require access to model equations *(Thomas, 1971)*. Given or assuming the PDF of input parameters, random values for these parameters can be generated and the output variable calculated. When this is repeated a number of times, a PDF for the dependent variables can be developed. The MC implementation may vary in terms of the sampling, termination criteria, and the way the results are presented.

Among the advantages of MC methods are possibilities to use them with many parameters and easy handling of different parameter distributions even in complex models. The precision of MC methods can be improved by performing additional model runs. MC methods are simple, informative, and widely used in the scientific community; however, they are computationally expensive. The main disadvantage of MC is that the role of each variable in contributing to total uncertainty is often difficult to infer *(Kirchner, 2001)*. Another limitation of MC analysis is that it requires estimates of the parameter distribution that reflect both uncertainty and variability

individually, and information of the parameter distribution is not available for most parameters *(USEPA, 1999)*.

The parameter distribution for MC methods can be approximated from the random pick using simple random sampling, Latin hypercube, bootstrap, or any other sampling technique.

4.2.4.3 MC with simple random sampling

Simple random sampling (SRS) produces a random vector of parameters from a prescribed probability distribution. Depending on type of data various PDFs can be assigned: normal, uniform, log-normal, triangular, exponential, etc. A normally distributed random variable x with mean μ and standard deviation σ can be repeatedly generated by

$$x_n = \sigma \cdot r_n + \mu \tag{4.3}$$

where r_n are normally distributed random numbers with mean 0 and variance 1. For a uniform distribution r is a uniformly distributed random number ranging from 0 to 1. The same approach is applied for other distributions. The main disadvantage of SRS is that a representative distribution of model outputs requires a large number of parameter vectors and hence a large number of computationally expensive model runs.

4.2.4.4 MC with Latin hypercube sampling

Latin hypercube sampling (LHS) was introduced by *McKay et al. (1979)* with the idea to achieve similar confidence in statistics of model outputs as with SRS using a smaller number of realizations. LHS is a stratified random procedure of sampling variables from their distribution *(Iman and Conover, 1982)*. Unlike SRS, this method ensures a full coverage of the range of each variable. Additionally, *Stein (1987)* shows that LHS yields an asymptotic variance smaller than that for SRS. Procedures for LHS include dividing the cumulative distribution of each variable into N equiprobable intervals. Then, from each interval a value is randomly selected. The sampled cumulative probability for the i^{th} interval can be written as *(Wyss and Jorgensen, 1998)*

$$\text{Prob}_i = (1/N)\, r_n + (\text{i} - 1)/\text{N} \tag{4.4}$$

Then, using the inverse of the distribution function, the sampled probability values are transformed into the real values. A total of N values are obtained for each variable. At the end, random values of each variable are randomly paired.

LHS is based on the assumption that the variables are independent of each other which is not true for most hydrological models. Random pairing may result in impossible combinations; furthermore, independent variables tend to bias the uncertainty *(Minasny and Bratney, 2002)*. *Iman and Conover (1982)* proposed a method for inducing correlation among the variables with the aim to restrict the way the variables are paired. Their method is based on the Cholesky decomposition of the correlation matrix. The proposed algorithm is limited by small matrices because the Cholesky decomposition of very large matrices is difficult to obtain *(Zhang and Pinder, 2003)*. *Stein (1987)* proposed a method for sampling dependent variables which does not have the stated limitation. Stein's LHS is based on the rank of a target multivariate distribution. The first step is simulation of a set of simple random realizations using any standard random generator. The second step is to generate LHS realizations using the rank of the simple random realization from the first step. The lowest i value is assigned rank 1 and the highest rank is assigned N. Duplicate or multiplicate i-values are assigned different ranks The Latin hypercube sample x_{ij} ($i = 1,\ldots, N; j = 1, \ldots k$) is obtained *(Minasny, 2000)* by

$$x_{ij} = F_j^{-1}\left(\frac{u_{ij} - r_u}{N}\right) \tag{4.5}$$

where F^{-1} – inverse of distribution function, u_{ij} – rank correlation matrix with k columns and N rows, k – number of variables (parameters), N – number of equiprobable intervals.

With this transformation the sampled values yields an approximately joint distribution. But these methods do not necessarily produce the right correlation matrix. Correlations among the variables make some hypercubes more probable than the others *(Minasny, 2000)*.

LHS with the induced correlation as proposed by *Stein (1987)* was used along with others in the current research.

4.2.4.5 Integral transformation techniques

Integral transformation techniques can be used as a problem solving tool in uncertainty analysis. The most popular transformations in water resource modeling are Fourier, Laplace, and exponential transforms *(Yen, 2002)*. In all cases the PDF of the random variable X should be known. The Fourier transform is a reversible integral transform of one function into another. For a continuous aperiodic function the transformation is

$$\Im(\omega) = \int_{-\infty}^{\infty} f(x) \cdot e^{i \cdot \omega \cdot x} dx \tag{4.6}$$

where $\Im(\omega)$ – Fourier transform of function f(x) and $i = \sqrt{-1}$. If X is a random variable, the resulting Fourier transform $\Im(\omega)$ is called the characteristic function and the r-th moment of the random variable X can be obtained *(Yen, 2002)* as

$$E[X^r] = (-1^r)\left[\frac{d^r \Im(\omega)}{d\omega^r}\right]_{\omega=0} \tag{4.7}$$

The Fourier transform is practically useful when random variables are independent and linearly related. In such cases, the convolution property of the Fourier transform can be applied to derive the characteristic function of the resulting stochastic variable *(Yen, 2002)*. Laplace and exponential transforms use moment generating functions which are similar to the characteristic function. Tables on characteristic functions and moment generating functions can be found in mathematical handbooks *(Abramowitz and Stegun, 1972)*. More details on integral transformation techniques and their application can be found in *Duffy (2004)* and *Tung (1996)*.

4.2.4.6 Response surface methods

Response surface (RS) methods were developed to overcome the disadvantages of the MC methods. The idea is to replace the complicated output function by a simple analytical approximation which is termed a RS *(Cox and Baybutt, 1981)*. The procedures are the following: (i) determination of important input parameters, (ii) making multiple model runs using specific combinations of these input parameters, and (iii) fitting a polynomial to the model data. Often the method of least squares is used. The fitted RS is used as a replacement of the computer model and uncertainty is approximated from this fitted polynomial. *Box and Draper (1987)* describe the adaptive nature of RS methods. The methodology for identifying the sampling points and subsequent analysis is presented in the literature *(Draper and Smith, 1981; Khuri and Cornel, 1987)*. The principal advantage of the RS is economy in model runs. Even for very complicated output functions the RS surface can be fit using only a few tens of runs. The major disadvantage is the difficulty of estimating the accuracy with which the RS approximates the original output *(Cox and Baybutt, 1981)*. Using the RS method it is difficult to detect model thresholds, discontinuities, nonlinearities, and include correlations and restrictions of input parameters. RS methods typically work only when there are less than 10 input parameters, a limited number of

output variables, and the relationship between input and output variables is basically linear or quadratic or involves a few cross-products *(Helton, 1993)*. Otherwise, the structure of the input-output relationships is too complicated to be captured in an efficient manner.

4.3 Materials and methods

4.3.1 Study area and collected data

The research was conducted on the South Nation watershed located in southeastern Ontario, Canada. The South Nation River (SNR) flows in a northeastern direction from headwaters near Brockville into the Ottawa River near the community of Plantagenet. The length of the SNR is about 175 km; the elevation difference between its source and mouth is only 84 m due to a very flat landscape. SNR drains about 3810 km^2.

For the purpose of this study the South Nation watershed was subdivided into 34 subwatersheds. Subdivision was based on distribution of meteorological and hydrological stations as well as relief and soil types. Each subwatershed was further subdivided into landuse segments.

The study period covers 11 years from 1990 to 2000. Extensive meteorological data were used to set up the model. Hourly data included: precipitation, potential evaporation, air temperature, wind speed, solar radiation and dewpoint temperature. Landuse data were taken from the Landsat imagery with a resolution of 30 m. Five landuse categories were identified: agriculture, forest, pasture, bareland, and urban territories. Geographical information system (GIS) topographical maps, digital elevation model with 25 m resolution and hydrography data were used. Flow and water quality data were obtained from the South Nation Conservation Authority.

4.3.2 Watershed model and optimization techniques

HSPF, one of the most popular watershed models, was used in the current research. HSPF is a part of the software system BASINS (Better Assessment Science Integrating Point and Nonpoint Sources) which is developed and supported by USEPA. It is a comprehensive continuous simulation watershed model that simulates flows, point and non-point pollution, and performs its routing in stream reaches. Detailed description of the model is available in *Bicknell et al. (2001)*.

Two nonlinear methods were used for model calibration: Gauss-Marquardt-Levenberg (GML) developed by *Marquardt (1963)* and Shuffled Complex Evolution (SCE-UA) developed by *Duan et al. (1992)*. GML is implemented in the Parameter estimation tool (PEST) which was used for calibration and sensitivity analysis *(Doherty, 2004)*. SCE-UA was employed to ensure that the

global minimum was found. Although SCE-UA is superior in many aspects, GML can produce good calibration results once GML variables are adjusted and initial parameter values are chosen properly *(Iskra and Droste, 2005)*.

Sampling for MC and RS methods was done according to Stein's LHS technique *(Stein, 1987)* using code written in Matlab 7.1.

4.3.3 Sensitivity analysis

In a broad sense sensitivity refers to the rate of change in one factor with respect to change in another. Basically, sensitivity is a ratio of changes. In the current research finite difference or perturbation sensitivities, obtained by perturbing each parameter and calculating the sensitivity of each observation are used. This requires $(N + 1)$ runs to obtain the forward or backward derivatives and $(N \times 2) + 1$ runs for more accurate central derivatives *(Doherty, 2004)*. N is number of parameters.

Two types of sensitivity were evaluated: parameter sensitivity and observation sensitivity. Composite parameter sensitivity, i.e., the effect of parameter changes with respect to all weighted observations was estimated according to

$$S_i = \frac{(J^t W J)_{ii}^{1/2}}{m} \tag{4.8}$$

where m is number of observations. Simply, the composite sensitivity of parameter i is a weighted magnitude of the column i of the Jacobian matrix.

Having calculated the Jacobian matrix, analysis and comparison of observation sensitivity becomes easy. Observation sensitivity shows how sensitive observation j is with respect to a particular parameter in column i. Often the composite sensitivity of a particular observation to all adjustable parameters is calculated. Composite observation sensitivity is defined as the magnitude of the ith row of the Jacobian multiplied by the observation weight and divided by the number of adjustable parameters *(Doherty, 2004)*. Observation sensitivity characterizes the importance of a particular observation for the optimization process. Composite observation sensitivity (S_j) is defined as

$$S_j = \frac{(J^t W J)_{jj}^{1/2}}{n} \tag{4.9}$$

According to *Baker and Daley (2000)* the observation sensitivity is the largest when the observations are relatively isolated and assumed to be more accurate than the background. Also, large observation sensitivities suggest that these data have a greater potential to impact the forecast aspects of the model *(Baker and Daley, 2000)*.

4.3.4 Objective function

Calibration of the model is an iterative procedure of parameter evaluation and adjustment. The essence of calibration is to minimize differences between observed and simulated values. In this research, the differences or errors were chosen to be expressed in weighted squared residuals. The sum of squared weighted residuals which is minimized during the calibration is called an objective function (OF). In this research it is assumed that the optimal parameter set corresponds to the minimum value of the OF. The general expression of the OF (Φ) is

$$\Phi = \sum_{i=1}^{n} (w_i r_i)^2 \tag{4.10}$$

where r – residuals expressed as the difference between simulated and observed values for the i^{th} observation, and w – weight attached to the i^{th} observation.

In the current paper the best fit of the observed and simulated daily flows for a period of 11 years was required. The OF was chosen to be expressed as a sum of log transformed daily flows.

$$\Phi = \sum_{i=1}^{n} (\log Q^i - \log q^i)^2 \tag{4.11}$$

where Q – observed and q – simulated daily flows, respectively.

Log transformation keeps more equal contributions of high and low flows into the OF. In a non-log OF, high values dominate and the fit of low flows is poor.

4.3.5 Model calibration

The hydrological calibration consists of adjusting daily flows, monthly volumes, and snow depths in the watershed. About 30 parameters impact water budget calibration in a pervious land segment. Some of them are fixed for a period of modeling, some vary monthly. A number of parameters were obtained from landuse maps, high resolution topographical maps, GIS or field surveys. Other parameters were estimated from climatological conditions, geology, soil types, and literature. The HSPF snow modeling subroutine with more than 20 parameters simulates

accumulation and melting of snow and ice. Some snow parameters, such as latitude of watershed segment, mean elevation, temperature at which precipitation becomes snow, etc. are straight forward with no need for further correction. Based on literature review and sensitivity analysis two snow parameters were adjusted during calibration: snow gage catch correction factor (SNOWCF) and condensation/convection melt factor (CCFACT).

Calibration consists of defining a physically possible range of parameters and selecting parameters that produce the best agreement between observed and simulated values throughout the calibration period. Model performance comparison was based on the following statistical characteristics: coefficient of determination (r^2), mean error or bias (ME), mean absolute error (MAE), percent mean error (PME), percent mean absolute error (PMAE), Nash and Sutcliffe (NS) model fit, coefficient of efficiency (CE), index of agreement (IA) proposed by *Willmott (1981)*, and daily root mean square error (DRME). Description of the statistics and the equations are presented in *Iskra and Droste (2005)*.

4.3.6 Allowable parameter uncertainty

Several guidelines exist for performing the uncertainty analysis *(IAEA, 1989; Helton, 1993; Hammonds et al., 1994)*. The generalized procedures implemented in the research include several steps.

- The assessment endpoint is defined.
- All uncertain parameters and their upper and lower bounds are identified.
- A subjective PDF for each parameter is specified.
- The correlation among the parameters is determined and possibly taken into account.
- Using either analytical or numerical procedures the uncertainty in model parameters is propagated to produce a PDF of the model outcome.
- The quantitative estimation of uncertainty in terms of a subjective confidence interval for the unknown value is derived.
- The relationship between input parameters and output variables is examined. The parameters contributing most to uncertainty in the model outcome are ranked.
- The results of the analysis are presented and interpreted.

The assessment endpoint is the set of best parameter values which is found during the calibration process using the global optimization technique. In this research the best parameter set was chosen to correspond to the lowest achievable value of the OF. In reality, there is more than one parameter set that can maintain the model in a calibrated state. This is partly due to the fact that the lowest OF has its own confidence limits. Moreover, each parameter has its own probability distribution around the best value. The shape of this distribution is determined by the nature of the parameter, its sensitivity, and degree of correlation to other parameters.

Boundaries of OF confidence regions were computed as suggested in *Seber and Wild (1989)* and *Vecchia and Cooley (1987)*. The assumption of linearised confidence interval around Φ_{min} was made because the increment, δ, above the minimum OF is very small. δ was found according to

$$\delta = m\sigma^2 F_a(m, n-m) \tag{4.12}$$

where $F_\alpha(m, n-m)$ denotes the upper α probability point of a F distribution with m and $n-m$ degrees of freedom, m – number of parameters and n – number of observations, σ^2 – a reference variance estimated as $\Phi_{min}/(n-m)$.

Physically realistic boundaries for all uncertain parameters were identified. The range of parameter values was taken from *USEPA (2000)* and adjusted to SNR climatic and geologic conditions.

After the analysis of possible probability distributions for uncertain parameters it was assumed that all parameters were normally distributed around the calibrated value.

Three methods where used to propagate the allowable parameter uncertainty to the model outcomes: methods of moments, MC with LHC sampling, and RS.

4.4 Results

4.4.1 Sensitivity analysis

Parameter sensitivity is important for distinguishing parameters that are insensitive to model outputs and contribute little to model improvement. Composite parameter sensitivity of the calibrated model with respect to daily flows is shown in Table 4.1. The daily flows are most sensitive to INFILT, AGWETP, and CEPSC. This was not surprising since INFILT is related to mean soil infiltration capacity and effectively controls the division of rain water to surface und subsurface storages. High values of INFILT result in increased baseflow, low values produce

more overland flow and interflow *(USEPA, 2000).* AGWETP characterizes the direct evaporation from the groundwater storage. Typically this parameter is sensitive in areas with shallow groundwater, such as the SNR. The effect of AGWETP is especially visible on the low-flow parts of the hydrograph.

Although the observation sensitivity conveys less useful information than the parameter sensitivity, sometimes it may be helpful. High observation sensitivity suggests that a particular observation is important for calibration and sensitive to the estimated parameters. *Doherty (2004)* noted that a high value of observation sensitivity does indeed mean that the observation is possibly sensitive to many parameters; however, it does not indicate that the observation is particularly indispensable to the parameter estimation process. This can only be determined in the context of the presence or absence of other observations with similar sensitivities *(Doherty, 2004).*

Table 4.1 Results of sensitivity analysis for some HSPF parameters

Parameter	Description	Value	Sensitivity
INFILT	Index to infiltration capacity of soil (in.)	0.0351	0.0546
AGWETP	Fraction of ET from active groundwater (GW)	0.0342	0.0363
CEPSC	Interception storage capacity (in.)	0.0170	0.0190
IRC	Interflow recession parameter	0.7534	0.0129
LZETP	Lower zone evapotranspiration (ET)	0.3385	0.0109
DEEPFR	Fraction of GW inflow to deep recharge	0.5000	0.0093
CCFACT	Condensation/convection melt factor	0.9528	0.0051
SNOWCF	Snow gage correction factor	1.2813	0.0036
UZSN	Upper zone nominal storage (in.)	1.9871	0.0014
BASETP	Fraction of ET from baseflow	0.2000	0.0011
KVARY	Variable GW recession (1/in.)	1.4035	0.0009
LZSN	Lower zone nominal storage (in.)	2.2915	0.0006

Simulated daily flows and their sensitivities are shown in Figure 4.4. Twelve parameters were used for the calibration. It can be seen that the sensitivity varies insignificantly, typically less than one order of the magnitude. An inverse relationship between flow and observation sensitivity was observed. Observations with high flows tend to have smaller sensitivity. That was somewhat unexpected.

The model was calibrated using from 1 to 18 hydrologic and snow parameters. In total, about 25 calibrations were built. Parameters were added in order of decreasing sensitivity. Calibration was tedious and involved many optimization runs. Daily flows and monthly volumes were of primary interest in the hydrological calibration. Although snow calibration was conducted the results are

not presented here. As an example, the statistical performance of simultaneous calibration of 12 parameters is presented in Table 4.2.

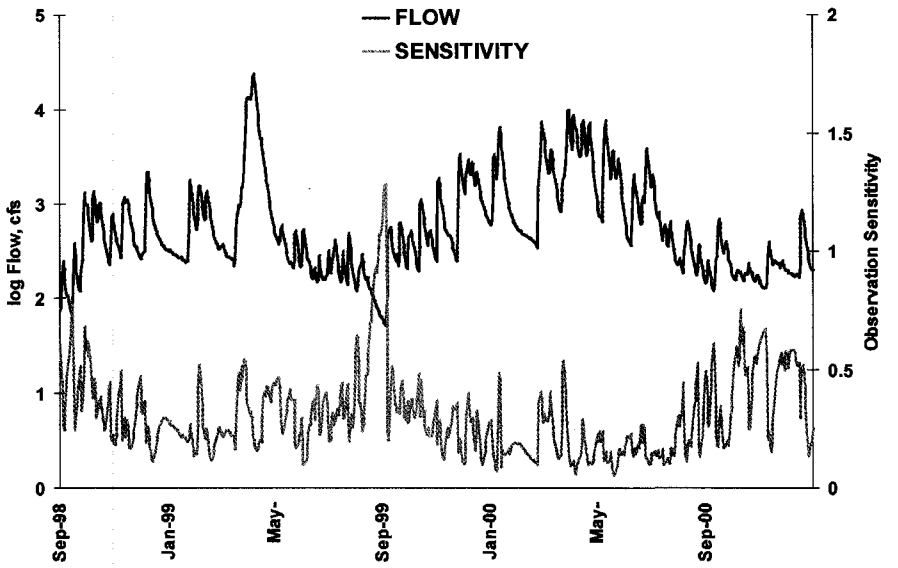


Figure 4.4 Daily flows and their correspondent observation sensitivity

Table 4.2 Performance statistics for observed and simulated daily flows

r^2	ME	MAE	PME	PMAE	NS	CE	IA	DRMS
0.808	198.4	695.7	11.2	39.29	0.773	0.649	0.815	1646

4.4.2 Calibration

In addition to statistical analysis, graphical comparisons of observed and simulated flows were made including percent exceedance flow curves, scatter plots, graphs of residuals, and plots of cumulative differences. Based on the statistics and graphical comparison of modeled and observed daily flows it was concluded that sufficient precision of each calibration was achieved.

4.4.3 Parameter uncertainty

4.4.3.1 First-order variance estimation

The least sophisticated estimation of uncertainty is the FOVE. It estimates variance by considering first-order Taylor expansion at a selected point in the model parameter space. The confidence intervals are calculated based on the covariance matrix which was generated during the GML parameter update. The covariance matrix is directly proportional to the OF and thus varies inversely with the goodness of fit between observed and simulated data. The computed variance of parameters, coefficient of variation, and 95% confidence intervals for 12

possibly sensitive to many parameters; however, it does not indicate that the observation is particularly indispensable to the parameter estimation process. This can only be determined in the context of the presence or absence of other observations with similar sensitivities *(Doherty, 2004)*.

Table 4.1 Results of sensitivity analysis for some HSPF parameters

Parameter	Description	Value	Sensitivity
INFILT	Index to infiltration capacity of soil (in.)	0.0351	0.0546
AGWETP	Fraction of ET from active groundwater (GW)	0.0342	0.0363
CEPSC	Interception storage capacity (in.)	0.0170	0.0190
IRC	Interflow recession parameter	0.7534	0.0129
LZETP	Lower zone evapotranspiration (ET)	0.3385	0.0109
DEEPFR	Fraction of GW inflow to deep recharge	0.5000	0.0093
CCFACT	Condensation/convection melt factor	0.9528	0.0051
SNOWCF	Snow gage correction factor	1.2813	0.0036
UZSN	Upper zone nominal storage (in.)	1.9871	0.0014
BASETP	Fraction of ET from baseflow	0.2000	0.0011
KVARY	Variable GW recession (1/in.)	1.4035	0.0009
LZSN	Lower zone nominal storage (in.)	2.2915	0.0006

Simulated daily flows and their sensitivities are shown in Figure 4.4. Twelve parameters were used for the calibration. It can be seen that the sensitivity varies insignificantly, typically less than one order of the magnitude. An inverse relationship between flow and observation sensitivity was observed. Observations with high flows tend to have smaller sensitivity. That was somewhat unexpected.

The model was calibrated using from 1 to 18 hydrologic and snow parameters. In total, about 25 calibrations were built. Parameters were added in order of decreasing sensitivity. Calibration was tedious and involved many optimization runs. Daily flows and monthly volumes were of primary interest in the hydrological calibration. Although snow calibration was conducted the results are not presented here. As an example, the statistical performance of simultaneous calibration of 12 parameters is presented in Table 4.2.

simultaneously optimized parameters are presented in Table 4.3. It is important to note that 95% confidence intervals rely on a linearity assumption that allows the limits to extend beyond the parameter space (e.g., LZSN and USZN).

Table 4.3 Variance and 95% confidence intervals for some HSPF parameters

Parameter	Current value	Variance	Coefficient of variation	95% confidence interval
INFILT	0.0351	8.09×10^{-6}	0.081	± 0.0056
AGWETP	0.0342	3.31×10^{-5}	0.168	± 0.0113
CEPSC	0.0170	6.44×10^{-5}	0.472	± 0.0157
IRC	0.7534	4.80×10^{-5}	0.009	± 0.0136
LZETP	0.3385	3.78×10^{-4}	0.057	± 0.0381
DEEPFR	0.5000	3.42×10^{-4}	0.037	± 0.0363
CCFACT	0.9528	3.93×10^{-4}	0.021	± 0.0389
SNOWCF	1.2813	7.12×10^{-4}	0.021	± 0.0523
UZSN	1.9871	6.21×10^{-3}	0.040	± 0.1545
BASETP	0.2000	6.49×10^{-3}	0.403	± 0.1579
KVARY	1.4035	3.74×10^{-2}	0.138	± 0.3794
LZSN	2.2915	4.15×10^{-2}	0.089	± 0.3991

It can be seen that the most sensitive parameters, INFILT and AGWRC, have the smallest variance and as a result the narrowest confidence intervals. The least sensitive parameters, KVARY and LZSN, are the most uncertain. Coefficient of variation compares the degree of variation of each parameter from its optimal value. It is seen that the highest coefficient of variation is for interception storage capacity (CEPSC) and ET from the groundwater (BASETP), due to their small parameter values and high variance.

Calibrations with different numbers of parameters were built and compared (Figure 4.5). It was noted that the optimized parameter value is stochastically changing as more parameters are added to the calibration processes. This is likely a result of parameter intercorrelation and the complex shape of the resulting OF. Also, it was observed that the confidence intervals depend on number of simultaneously optimized parameters. Typically, increasing the number of parameters leads to greater allowable parameter uncertainty.

The MC method with LHS and induced correlation as proposed by *Stein (1987)* was used in the current research. Based on the optimal parameter set, assumed normal distribution, and correlation matrix which was obtained during GML estimation, a number of parameter vectors were generated. Then, the model was run for every parameter set and values of the model output (OF in our case) were recorded. From 1,000 to 50,000 parameter sets and corresponding model

runs were performed during the experiment. The PDF for output function was built based on the model realizations.

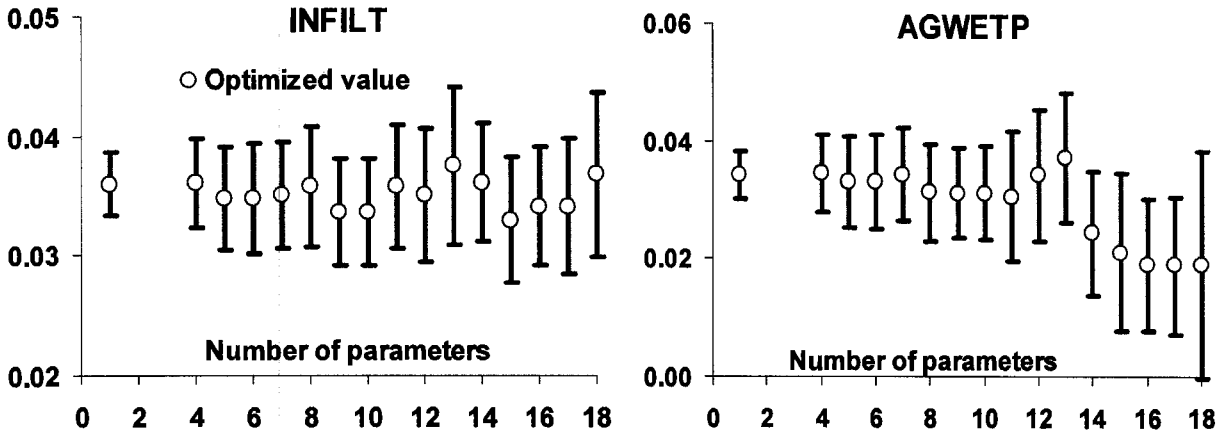


Figure 4.5 Optimized parameter values and their 95% confidence intervals using the method of moments

4.4.3.2 Monte-Carlo uncertainty

No significant changes in the PDF were observed if more than 10,000 model runs were performed, which is why in all experiments the sample size was 10,000. The confidence limits for several HSPF parameters are presented in Figure 4.6. It can be seen that uncertainty increases as number of the calibrated parameters grows. The confidence intervals are asymmetrical, especially when less than 7 parameters are used.

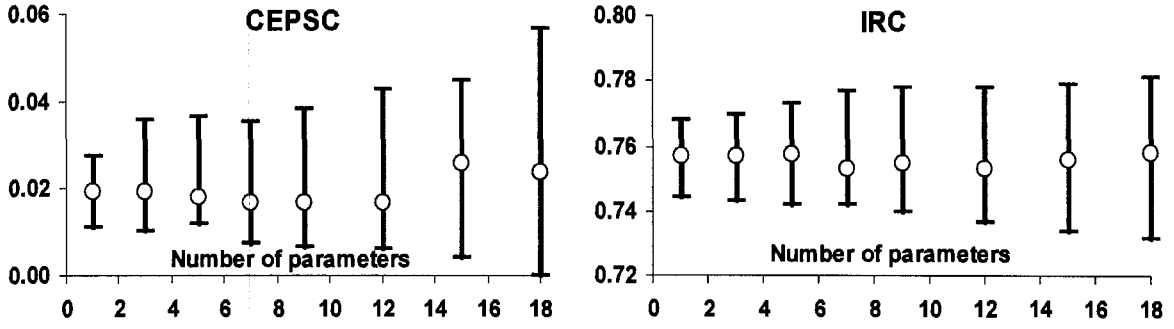
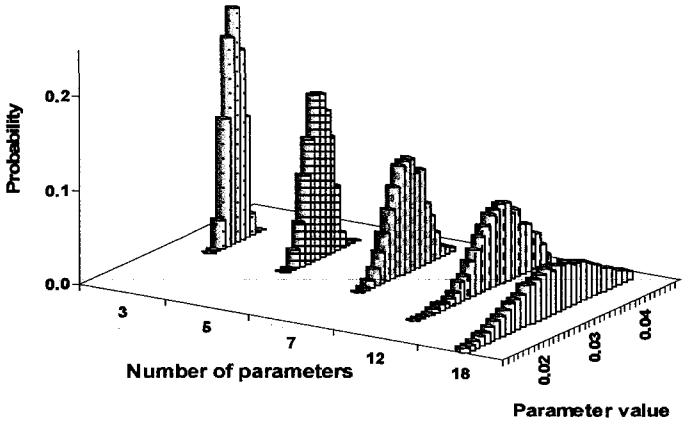


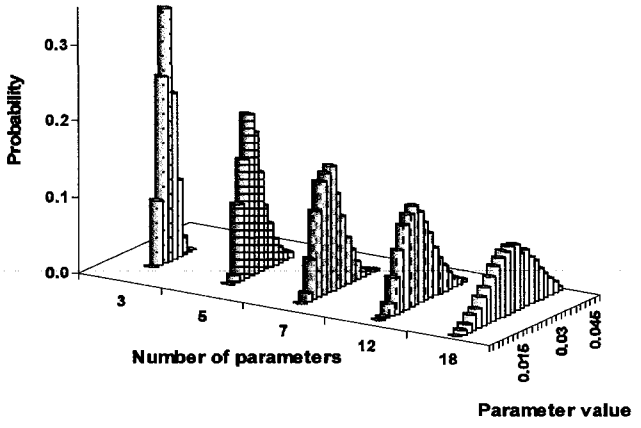
Figure 4.6 Optimized parameter values and their 95% confidence intervals using MC method

A more interesting picture is seen when the probability histograms are plotted (Figure 4.7).

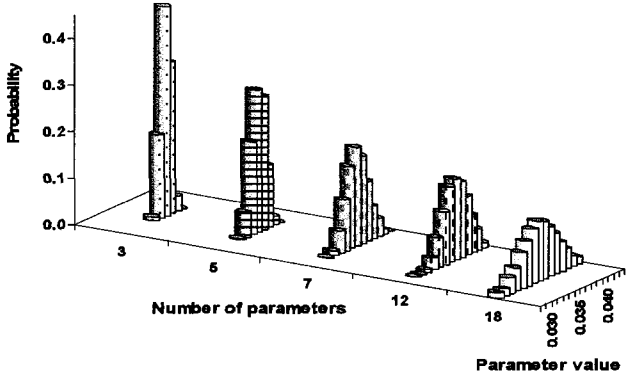
AGWETP



CEPSC



INFILT



IRC

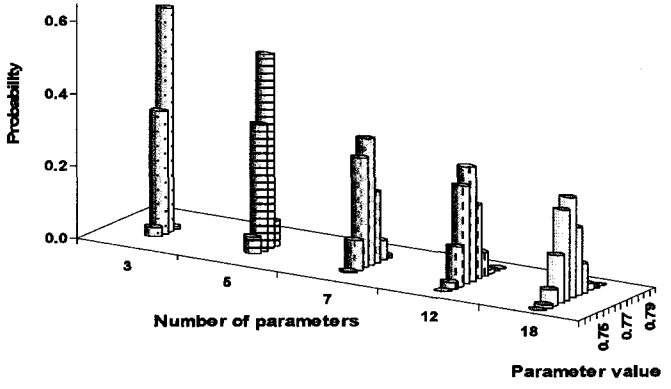


Figure 4.7 Histograms obtained by MC method with LHS based on 10,000 sampled parameter sets for the model with different number of the simultaneously optimized parameters

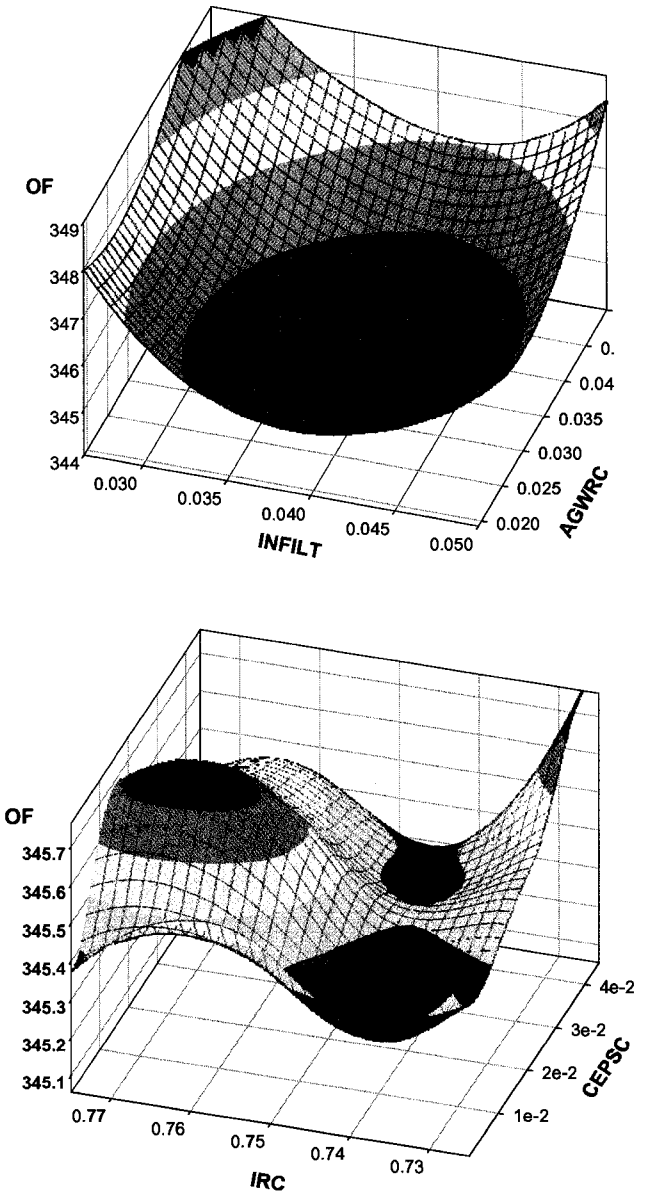


Figure 4.8 Response surfaces

It can be seen that as more parameters are optimized simultaneously, the deviation from the mean value increases. Probability histograms carry important information about allowable parameter uncertainty: confidence intervals, shape, tails, skewness, etc.

4.4.3.3 Response surfaces

Two aspects are crucial for performance of the RS: the sampling technique and the smoothing methods which are used to build the surface. In the current research LHS was used as the

sampling technique. Locally weighted scatterplot smoothing (LOESS) with tricube weighting and 3rd order polynomial regression was used to build the surface (Figure 4.8); 1,000 model runs were used to build the surface.

Once the RS is built the uncertainty can be estimated from the graph or more preferably from the smoothed data matrix. The concern with the RS method is difficulty to analyze several parameters simultaneously and to find the true shape of the nonlinear function (thresholds and discontinuities). Should a different sampling or smoothing technique be used, the results may be completely different. Although the RS method is fast, it is unlikely to determine the complex relationship between input and output variables in a nonlinear model.

4.4.3.4 Comparison of methods

Comparison of results of the methods of moments, MC, and RS methods is shown in Table 4.4. The 95% confidence intervals were approximated for different parameters. It can be seen that typically the most conservative method is the method of moments. The largest uncertainty was estimated by MC. Taking into account disadvantages of the methods of moment and RS; it is beneficial to use MC, ensuring the proper sampling technique is used. MC methods normally have fewer assumptions and therefore the results are more reliable.

Table 4.4 Comparison of parameter uncertainty estimated by different methods

Parameter	Calibrated	Method of moments		Monte Carlo		Response surface	
	value	min	max	min	max	min	max
INFILT	0.0351	0.0295	0.0407	0.0295	0.0471	0.0350	0.0461
AGWETP	0.0342	0.0229	0.0455	0.0176	0.0531	0.0203	0.0377
CEPSC	0.0170	0.0013	0.0327	0.0047	0.0456	0.0051	0.0500
IRC	0.7534	0.7398	0.7670	0.7357	0.7784	0.7374	0.7787
LZETP	0.3385	0.3004	0.3766	0.2562	0.3708	0.2561	0.3465
DEEPFR	0.5000	0.4637	0.5363	0.4778	0.5000	0.4913	0.5809
CCFACT	0.9528	0.9165	0.9917	0.8799	1.0284	0.8921	0.9845
SNOWCF	1.2813	1.2290	1.3336	1.2045	1.3898	1.2412	1.3625
UZSN	1.9871	1.8326	2.1416	1.8245	2.0000	1.9221	2.2365
BASETP	0.2000	0.0000	0.5794	0.0658	0.2000	0.0810	0.4210
KVARY	1.4035	1.0241	1.7829	0.6815	2.0552	0.8010	1.9820
LZSN	2.2915	1.8924	2.6906	2.0000	2.7866	1.7803	2.5715

After the uncertainty results are obtained the next important question is how allowable parameter uncertainty translates into the model outcomes. The parameter sets which correspond to 95% minimum and maximum bounds of the confidence intervals were used to generate maximum flows in SNR (Figure 4.9). It can be seen that the predictive confidence intervals of the maximum flows are not high. Typically the parameter uncertainty is translated into a flow error not higher

than 10%. Typically, the wider predictive confidence intervals correspond to higher values of the flow.

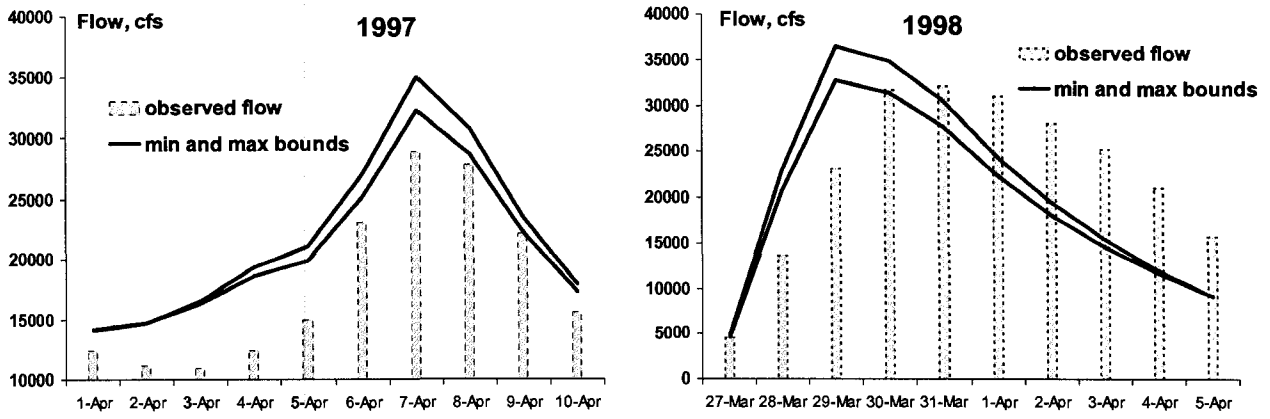


Figure 4.9 Minimum and maximum bounds of allowable parameter uncertainty for spring maximum flows at Plantagenet

4.5 Conclusions

Sensitivity analysis is an important step in estimation of parameter uncertainty. As a result of sensitivity analysis the parameters were ranked and the most sensitive ones were INFILT, AGWETP, and CEPSC. It was noted that the most sensitive parameters have the smallest variance and as a result narrow confidence intervals. The least sensitive parameters (KVARY and LZSN) are the most uncertain. Analysis of observation sensitivity revealed an unexpected inverse relationship between daily flows and individual observation sensitivities.

A number of model calibrations with 1 to 18 estimated parameters were performed. It was observed that as more parameters are optimized simultaneously, typically higher uncertainty is observed. This likely happens due to intensified correlation among the parameters and increased number of relatively insensitive parameters. Also, the optimized parameter value is stochastically changing as more parameters are added to the calibration process.

Among the tested methods the most conservative uncertainty was shown by the method of moments. FOVE relies on a linearity assumption which causes the limits of some parameters to extend beyond the parameter space. The confidence intervals are symmetrical.

The MC method with induced correlation shows the highest parameter uncertainty among the tested methods. From 1,000 to 50,000 parameter sets and corresponding model runs were performed during the experiment. The PDF for output function was built based on the model

realizations. No significant changes in the PDF were observed if more than 10,000 model runs are performed; thus 10,000 model realizations were chosen as optimal. The confidence intervals in MC are asymmetrical, especially when less than 7 parameters are used. Due to fewer assumptions MC methods are more reliable compared with the methods of moments and RS.

To build the RS for 12 parameters, 1,000 model runs were performed. The uncertainty was then estimated from the smoothed data matrix. On average, the values of parameter uncertainty were between FOVE and MC methods. Although the RS methods are fast and can provide a rough estimation of uncertainty, they are unlikely to catch complex relation between input and output variables in a nonlinear model.

The impact of parameter uncertainty on peak spring flows was studied. It was found that 95% confidence intervals of parameter uncertainty correspond to up to 10% variations in maximum spring flows. Typically, the wider predictive confidence intervals correspond to higher values of the flows.

4.6 References

Abramowitz, M. and Stegun, I.A. (1972) Handbook of mathematical functions: with formulas, graphs, and mathematical tables, 9th edition, Dover publications, NY.

Aristotle (350 BC) *Nicomachean ethics*, translated in 1962 by Martin Ostwald, Library of Liberal Arts, p. 352.

Baker, N.L. and Daley, R. (2000) Observation and background adjoint sensitivity in the adaptive observation-targeting problem. *Quart. J. Roy. Meteor. Soc.*, 126, 1431-1454.

Bicknell, B.R., Imhoff, J.C., Kittle, Jr. J.L., Jobes, T.H., and Donigian, Jr. A.S. (2001) *Hydrological simulation program HSPF. Version 12. User's manual.* U.S. Environmental Protection Agency, Athens, Georgia.

Box, G.E.P. and Draper, N.R. (1987) *Empirical model-building and response surfaces.* John Wiley & Sons, NY.

Castrup, H. (2002) Distributions for uncertainty analysis. *IDW conference*, May 6-9, Knoxville, TN.

Chang, C.H., Tung, Y.K., and Yang, J.C. (1994) Monte Carlo simulation for correlated variables with marginal distribution. *J. Hydr. Eng.,* 120, 313–331.

Clarke, R.T. (1994) *Statistical modeling in hydrology*, John Wiley and Sons, UK, p. 412.

Cox, D.C. and Baybutt, P. (1981) Methods for uncertainty analysis: a comparative survey. *Risk Analysis,* 14, 251-258.

Doherty, J. (2004) *PEST – model-independent parameter estimation.* User manual: 5th edition, Watermark Computing, Corinda, Australia.

Draper, N.R. and Smith, H. (1981) *Applied regression analysis*, John Wiley & Sons, NY.

Duffy, D.G. (2004) Transform methods for solving partial differential equations, 2nd Edition Springer, p.728.

Duan, Q.Y., Gupta, V., Sorooshian, S., Rousseau, A.N., and Turcotte, R. (2003) *Calibration of watershed models.* American Geophysical Union, Washington DC, p 345.

Duan, Q.Y., Sorooshian, S, and Gupta, V. (1992) Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.* 28, 1015-1031.

Hammonds, J.S., Hoffman, F.O. and Bartell, S.M. (1994) *An introductory guide to uncertainty analysis in environmental and health risk assessment.* Technical Rep. No. ES/ER/TM-35/R1, Oak Ridge National Laboratory, Oak Ridge, Tenn.

Harr, M.E. (1987) Reliability-based design in civil engineering, McGraw-Hill, New Your.

Harr, M. E. (1989) Probabilistic estimates for multivariate Analyses. *Applied Mathematical Modelling*, 13, 313-318.

Helton, J.C. (1993) Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering and System Safety*, 42, 327-367.

Iman, R.L. and Conover, W.J. (1982) A distribution-free approach to inducing rank correlation among input variables. *Communication in statistics: simulation and computation* 11, 311-34.

IAEA (1989). Evaluating the reliability of predictions made using environmental transfer models. IAEA Safety Series 100, Vienna, Austria.

Iskra, I. and Droste, R. (2005) Application of non-linear automatic optimization techniques for calibration HSPF. *WEFTEC conference*, October 29 - November 2, Washington, 8778-8798.

Kirchner, J. (2001) *Data analysis toolkit: uncertainty analysis and error propagation.* Course notes, University of California, Berkeley.

Khuri, A.I. and Cornell, J.A. (1987) *Response surfaces: design and analyses.* Marcel Dekker, NY.

Li, K.S. (1992) Point estimate method for calculating statistical moments. *Journal of Engineering Mechanics*, ASCE, 118, 7, 1506-1511.

Lind, N.C. (1983) Modeling of uncertainty in discrete dynamical systems. *Applied Mathematical Modelling*, 7, 146-152.

Marquardt, D. (1963) An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, 11, 431-441.

McKay, M.D., Conover, W.J., and Beckman, R.J. (1979) A comparison of three methods for selection values of input variables in the analysis of output from a computer code. *Technometrics*, 22, 239-245.

Minasny, B., and Bratney, A. (2002) Uncertainty analysis for pedotransfer functions. *European Journal of Soil Science*, 53, 3, 417-429.

Minasny, B. (2000) *Efficient methods predicting soil hydraulic properties*, PhD thesis, University of Sydney, Australia.

Morgan, L. (1999) Pesticides and groundwater in the state of Washington. *Environmental Management and Health*, 10, 7-17.

Rosenblueth, E. (1975) Point estimates for probability moments. *Proceedings of the National Academy of Sciences*, 72, 10.

Seber, G.A.F. and Wild, C.F. (1989) *Nonlinear regression*. John Wiley and Sons, NY.

Singh, V. (2004) Flow routing in open channels: some recent advances. *Second International Conference on Fluvial Hydraulics*, June 23-25, 1-23.

Stein, M. (1987) Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, 29, 143-152.

Timbe, L. and Willems, P. (2004) Uncertainties in hydrodynamic flood simulation. *ACTIF Workshop on Quantification, reduction and dissemination of uncertainty*, Delft, November 22-23 2004, 1-15.

Thomas, J.B. (1971) An introduction to applied probability and random processes, Wiley, Washington.

Tung, Y.K. (1996) Uncertainty and reliability analysis in water recourses engineering. *Water Resources Update*, 103, 13-21

USEPA (1999) Treatment of uncertainty and variability, in report *Total risk integrated methodology*. EPA-453/R-99-010.

USEPA (2000) BASINS Technical Note 6: Estimating hydrology and hydraulic parameters for HSPF, http://www.epa.gov/waterscience/basins/tecnote6.pdf.

Vecchia, A.V. and Cooley, R.L. (1987) Simultaneous confidence and prediction intervals for nonlinear regression models with application to a groundwater flow model. *Water Resources Research*, 23, 7, 1237-1250.

Willmott, C.J. (1981) On the validation of models, *Phys. Geogr.* 2, 184-194.

Willems, P. (2000) *Probabilistic immission modelling of receiving surface waters.* PhD thesis, Katholieke Universiteit Leuven, Leuven.

Wyss, G.D. and Jorgensen, K.H. (1998) *A user's guide to LHS: Sandia's Latin hypercube sampling software*, SAND98-0210, Sandia National Laboratories, Albuquerque, NM.

Yen, B.C. (2002) System and component uncertainties in water recourses, in book: *Risk, reliability, uncertainty and robustness of water recourses systems*. Ed. J. Bogari and Z.W. Kundzewicz. Cambridge university press.

Zhang, Y. and Pinder, G.F. (2003) Latin-Hypercube lattice-sample selection strategies for correlated random hydraulic-conductivity fields. *Water Resour. Res.*, 39, 1226.

CHAPTER 5

UNCERTAINTY OF PREDICTIONS MADE BY HSPF

Igor Iskra and Ronald Droste
Department of Civil Engineering, University of Ottawa
161 Louis Pasteur St., Ottawa, ON K1N 6N5, Canada

The novel contribution of the presented paper is in estimation of predictive confidence interval and predictive noise in the dependent variable for spring maximum and autumn minimum flows using single and compound OFs. The predictive intervals were computed based on the 95% confidence limits of the OF.

5.1 Abstract

The Hydrologic Simulation Program Fortran (HSPF) was calibrated for an Ontario watershed to simulate daily flows. The Parameter Estimation Tool (PEST) was used for calibration, sensitivity analysis, and estimation the uncertainty which is associated with a particular prediction. It was noted that proper definition of the objective function (OF) and weighting factors is a key to robust automatic calibration. Single (sum of log-transformed residuals) and compound (sum of squared residuals of equally weighed log-transformed maximum, minimum, and middle flows) OFs were used throughout the research. The statistical parameters of coefficient of determination, mean error, and daily root mean square error should not be used for comparison of observed and simulated data. It was found that more than one parameter set is able to maintain the model in a calibrated state. The lowest value of the OF does not necessarily correspond to the optimum solution. The most and least sensitive parameters are found from simulations at high and low flows. The predictive confidence interval and predictive noise in the dependent variable were computed for spring maximum and autumn minimum flows using single and compound OFs. The predictive intervals were computed from the 95% confidence limits of the OF. For both high and low flows the random predictive error is the major part of the uncertainty for a particular prediction. Higher predictive noise pertains to the compound OF. Higher predictive confidence intervals are typical for the single OF.

Keywords: HSPF, PEST, Uncertainty, Prediction, Model calibration

5.2 Introduction

Watershed models are becoming an everyday routine for watershed planning, development, and management. Most watershed models are fairly complex with nonlinear interactions among watershed components and involve physical, chemical, and biological processes. Furthermore, watershed processes take place over a wide range of temporal and spatial scales *(Chow, 1988)*. As any model, a watershed model is a simplified representation of a real physical system intended to improve our ability to understand, control, and predict its behavior.

HSPF is one of the most widely used watershed models. It is a comprehensive semi-distributed model that can simulate hydrologic processes and water quality in water bodies. HSPF is a part of the software system BASINS (Better Assessment Science Integrating Point and Nonpoint

Sources) developed and supported by USEPA *(Donigian and Imhoff, 2006)*. HSPF is often used to facilitate the development of total maximum daily loads (TMDLs) in the US.

The most important and difficult step in application of any watershed model is its calibration. Calibration is often referred to as a test of a model with known input and output information that is used to adjust or estimate factors for which data are not available *(ASTM, 1984)*. In a more specific watershed model context, calibration is the process of adjusting numerical or physical modeling parameters for the purpose of improving agreement with experimental data *(AIAA, 1998)*. Model calibration is quite complex due to limitations of the model and input data, and often insufficient knowledge of watershed properties *(Schaake, 2003)*. Up to 100 parameters need to be adjusted for hydrological calibration in HSPF; therefore, manual calibration of HSPF is lengthy and often inaccurate. Recent advances in computer technology and development of new nonlinear optimization techniques allow application of many nonlinear automatic methods for HSPF calibration. In the present research PEST was applied for model calibration. PEST is a nonlinear estimator which employs the GML method to minimize the differences between observed and simulated data *(Doherty, 2004)*. The Shuffled Complex Evolution method developed at the University of Arizona (SCE-UA) is a robust and effective global optimization method *(Duan et al., 1992)*. It was also used in the current research.

The current research is devoted to HSPF application for prediction of daily flows. Prediction is a model application to foretell the response variable under the conditions for which the model has not been validated *(AIAA, 1998)*. All model predictions will inevitably include a broad range of uncertainties. One of the key elements of predictive analysis is to understand and estimate the uncertainty which is associated with a particular prediction. Estimation of predictive uncertainty of the maximum spring flows and minimum autumn flows was a primary objective of the present work.

5.3 Materials and methods

5.3.1 Study area and data

The research was conducted on the South Nation watershed located in southeastern Ontario, Canada. It flows in a northeastern direction from headwaters near Brockville into the Ottawa River near the community of Plantagenet. The length of the South Nation River (SNR) is about 175 km; it drains about 3810 km^2.

For the purpose of this study the South Nation watershed was subdivided into 34 subwatersheds. Several subwatersheds with available detailed flow observations were calibrated individually. Then, the model was applied to the whole South Nation watershed and parameters for the remaining subwatersheds were obtained through the calibration. The Plantagenet reach located close to the river mouth was the focus for predictive and uncertainty analysis in the current research.

The study period consists of 8 calibration years (1990-1997) and a three year predictive period (1998-2000). Hourly meteorological data were obtained from Environment Canada and include: precipitation, potential evaporation, air temperature, wind speed, solar radiation, and dewpoint temperature. Landuse data were taken from Landsat imagery with a resolution of 30 m. The satellite image was made in 1998. It was assumed that the landuse segments in 1998 were the same during the whole research period. Five landuse categories were identified: agriculture, forest, pasture, bareland, and urban territories. Geographical information system (GIS) topographical maps, digital elevation model with 25 m resolution and hydrography data were used. Flow and water quality data were obtained from the South Nation Conservation Authority and Environment Canada.

5.3.2 Method

The latest available version 12.0 of HSPF was used as the watershed model. It is a continuous simulation watershed model that simulates flows, point and non-point pollution, and performs its routing in stream reaches. Detailed description of the model is presented in *Bicknell et al. (2001)*.

Two nonlinear methods were used for model calibration: Gauss-Marquardt-Levenberg (GML) *(Marquardt, 1963)* and SCE-UA *(Duan et al., 1992)*. Although SCE-UA is superior in many aspects, GML can produce good calibration results once GML parameters are adjusted and initial parameter values are chosen properly *(Iskra and Droste, 2005)*. GML is implemented in the PEST optimization package *(Doherty, 2004)*, which was used for calibration, sensitivity, and predictive analysis. Taking into account the drawbacks of GML method *(Doherty, 2004)*, SCE-UA was also used to ensure that the global minimum was found.

Once the model was calibrated the predictive analysis function of PEST was used to evaluate predictive uncertainty. The procedure for estimating predictive uncertainty involves many model runs and selecting the parameter sets that generate maximum and minimum outputs while

keeping the model in a calibrated state. An example of a similar approach for assessing predictive uncertainty is presented in *Doherty and Johnston (2003)*.

5.3.3 Sensitivity analysis

In a broad sense sensitivity refers to the rate of change in one factor with respect to change in another *(McCuen and Snyder, 1986)*. Basically sensitivity is the ratio of two changes. Parameter sensitivity, i.e., the effect of parameter changes on the output function, was estimated. Composite sensitivity or sensitivity of each parameter with respect to all weighted observations was computed. The composite sensitivity of parameter i is defined according to *Doherty (2004)* as

$$S_i = \frac{(J^t W J)_{ii}^{1/2}}{n} \tag{5.1}$$

where n is number of observations.

So, the composite sensitivity of parameter i is a magnitude of the column i of the Jacobian matrix, with each element multiplied by the weight of the corresponding observation.

In some cases it was convenient to compare the relative sensitivity of the parameter which was computed by multiplying its composite sensitivity by the magnitude of the parameter value (Doherty, 2004). The relative composite sensitivity gives an indication of how fractional change in parameter values affects model output.

5.3.4 Objective function

As mentioned above, calibration is an iterative procedure of parameter evaluation and adjustment. The essence of calibration is to minimize differences between observed and simulated values. The differences or errors were chosen to be expressed in weighted squared residuals. This error is called an objective function (OF). The optimal parameter set typically corresponds to the minimum value of the OF.

The general expression of the OF is

$$\Phi = \sum_{i=1}^{n} (w_i r_i)^2 \tag{5.2}$$

where r – residuals expressed as the difference between simulated and observed values for the ith observation, and w – weight attached to the ith observation.

The objective of model calibration was to obtain the best fit of the observed and simulated daily flows for a period of 8 years. Two different OFs were used: single and compound. Several different single OFs were tested and experimentally it was found that the OF expressed as a sum of weighted log transformed daily flows performs well.

$$\Phi_1 = \sum_{i=1}^{n} (\log Q^i - \log q^i)^2 \tag{5.3}$$

where $\Phi_{\square}$ – single OF, Q – observed daily flows, and q – simulated daily flows.

Log transformation keeps more equal contributions of high and low flows into the OF. In a non-log OF high values dominate and the fit of low flows is poor.

The compound OF was expressed as a sum of squared differences between observed and simulated maximum, minimum, and middle daily flows. Maximum flows were assumed to be the highest 1% of flows. Minimum flows were defined as the bottom 20% of flows. Middle flows were flows between the minimum and maximum flows. This OF was determined to be good by asssment of many other compound OFs. All flow units are cubic feet per second (cfs). The compound OF was defined as:

$$\Phi_2 = \sum_{i=1}^{j} (w_1^i [\log Q_{\max}^i - \log q_{\max}^i])^2 + \sum_{i=1}^{k} (w_2^i [Q_{\min}^i - q_{\min}^i])^2 + \sum_{i=1}^{p} (w_3^i [\log Q_{mid}^i - \log q_{mid}^i])^2 \tag{5.4}$$

where Φ_2 – compound OF, j – number of maximum daily flows, k – number of minimum daily flows, p – number of middle daily flows, Q_{max}, Q_{min}, and Q_{mid} – observed maximum, minimum, and middle daily flows, $q_{\max}$, $q_{\min}$, and q_{mid} – simulated maximum, minimum, and middle daily flows, w_1, w_2, and w_3 –weighting factors.

Weights were assigned using a trial and error method to achieve approximately equal contributions from all three groups to the compound OF.

5.3.5 Model calibration

Hydrological calibration of daily flows and monthly volumes is the first step. About 30 parameters are adjusted during water budget calibration in a pervious land segment. Some of them are fixed for the period of modeling, some vary monthly. A number of HSPF parameters can be obtained from landuse maps, high resolution topographical maps, GIS or field surveys. Other parameters should be estimated from climatological conditions, geology, soil types, and the literature. During estimation, a physically possible range of parameters (parameter space) is

specified and then parameters are changed to find their optimum values. Some hydrological parameters vary from land segment to land segment (INFILT, LZSN, UZSN), and some are assumed to be unchanged for all segments for the whole calibration period (DEEPFR, KVARY, BASETP). A different set of parameters corresponds to different subwatersheds.

Snow calibration is another important step to reliable hydrological description of the basin. The HSPF snow modeling subroutine has more than 20 parameters. This subroutine simulates accumulation and melting of snow and ice. Some parameters, such as latitude of watershed segment, mean elevation, temperature at which precipitation becomes snow are straight forward with no need for further correction. Based on a literature review *(Crawford, 1999; USEPA, 2000; Bicknell et al., 2001; Hoyer and Larson, 2005)* and some sensitivity tests the following parameters were adjusted during calibration: fraction of land segment shaded from solar radiation (SHADE), snow gage catch correction factor (SNOWCF), snowfall required to fully cover surface (COVIND), density of new snow (RDCSN), condensation/convection melt factor (CCFACT), and liquid water storage capacity in snowpack (MWATER). Observed snow depths were a reference characteristic for snow calibration.

5.3.6 Performance comparison

During the calibration the overall performance of the model was evaluated based on a number of goodness-of-fit measures that quantify the error of simulation taking into account different aspects of the fit.

One of the most commonly used statistics is the coefficient of determination (r^2) which is the square of the Pearson product-moment correlation coefficient. It describes the proportion of the total variance in the observed data that can be explained by the model. For example, $r^2 = 0.8$ means that 80% of total variation in observed flow can be explained by the calibrated model. The coefficient of determination varies from 0 to 1. Higher values indicate better agreement between the observed and simulated data.

Mean error (ME) or bias measures the average tendency of the simulated value to be larger or smaller than its observed counterparts. Positive values indicate a model bias towards overestimation and vice versa. Mean absolute error (MAE) is a weighted average of the absolute errors. MAE emphasizes large outliers, negative and positive results are not able to cancel each other as is the case with ME. ME and MAE are measured in the same units as the original data. A zero value represents a perfect fit.

Percent mean absolute error (PMAE) is the average of differences between the simulated and observed values expressed as a percentage of the observed value. Percent mean absolute error (PMAE) is similar to PME, but the absolute values are used. In general, percent error measures are often preferred to their absolute error counterparts because they provide information on the magnitude of the errors relative to the average measurement.

Nash and Sutcliffe (NS) *(Nash and Sutcliffe, 1970)* model fit is a unitless measure of the ratio of residual variance ("noise") to variance of flow ("information"). NS values may vary from $-\infty$ to 1.0. Coefficient of efficiency (CE) is similar to NS, but absolute values are used. Higher values of NS and CE indicate better agreement between observed and simulated data.

Index of agreement (IA) was recommended by *Willmott (1981)* to overcome the insensitivity of correlation based measures to differences in the observed and simulated means and variances. In this research a slightly modified version of agreement was used, with the exponent equal to 1 rather than 2 as originally proposed by Willmott. The exponent equil to 1 was slightly more sensitive to small changes in the OF than exponent 2. IA varies from zero (a theoretical minimum) to 1.0 which is perfect agreement between the observed and simulated values.

Daily root mean square error (DRMS) simply computes standard deviation of the model simulated error. A smaller value indicates better agreement between observed and simulated data. DRMS gives a measure of the model error in the same units as the data, rather than in squared units or unitless.

5.4 Uncertainty analysis

Uncertainty refers to lack of knowledge about specific factors, parameters, or models *(USEPA, 1997)*. Uncertainties in a watershed model can be split into four groups: parameter uncertainty (measurement, sampling, and systematic errors), model uncertainty (due to simplification of real-world processes, assumptions, and mathematical equations), scenario uncertainty (descriptive errors, aggregation errors, missing or incomplete data, errors in professional judgment), and natural stochasticity (temporal and spatial variability) *(USEPA, 1997)*. When a model is used for predictions a combination of the above shapes the predictive uncertainty.

The approach for quantifying predictive uncertainty is based on the assumption that there is more than one parameter set that can maintain the model in a calibrated state. This happens because parameters are not completely independent. Several different sets of parameters that yield a

calibrated model can be produced by allowing the OF to rise slightly above its minimum. The "rise" is related to the probability of having the true value within the specified confidence interval. All parameter sets within the specified confidence interval of the OF generate a calibrated model. Using PEST two parameter sets were singled out, one which produced the maximum response and another which produced the minimum response. The difference between maximum and minimum predicted values is the uncertainty of prediction.

Figure 5.1 graphically illustrates the approach to quantify predictive uncertainty. The best parameter set was determined when the OF reached its possible minimum. Out of many parameter sets obtained when the OF is perturbed within the confidence region it is possible to single out the parameter set which generates the minimum possible prediction and another which generates the maximum possible prediction. The difference between the maximum and minimum possible predictions on a given day while still maintaining the model in a calibrated state is the confidence interval for a particular prediction or predictive uncertainty.

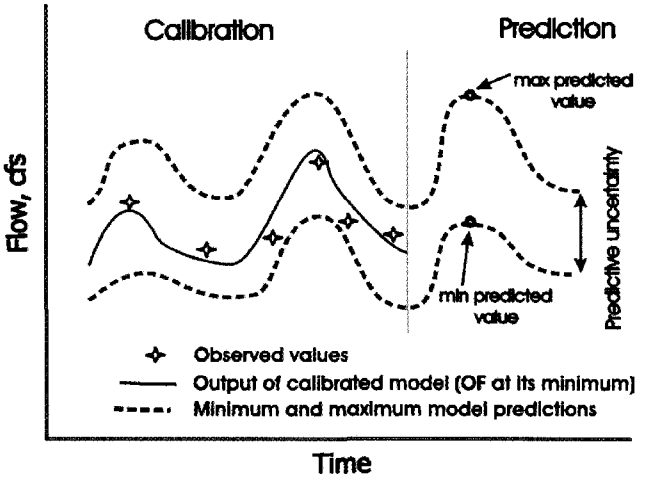


Figure 5.1 Schematic representation of predictive uncertainty

Predictive uncertainty can be shown in a parameter space. Figure 5.2 shows contour lines of the OF in parameter space of two parameters: infiltration index (INFILT) and lower zone moisture storage (LZSN). The model is deemed to be calibrated as long as the OF is less than 194.5, i.e., inside the inner contour line. Minimum (A) and maximum (B) critical points in Figure 5.2 are the two parameter sets that generate respectively minimum and maximum prediction outputs. The difference between the most widely separated points is the predictive uncertainty.

The paper's objective is to attempt to estimate uncertainty in prediction of specific future observations. Estimation of uncertainty is essential as a measure of the confidence an investigator has in the model simulation results.

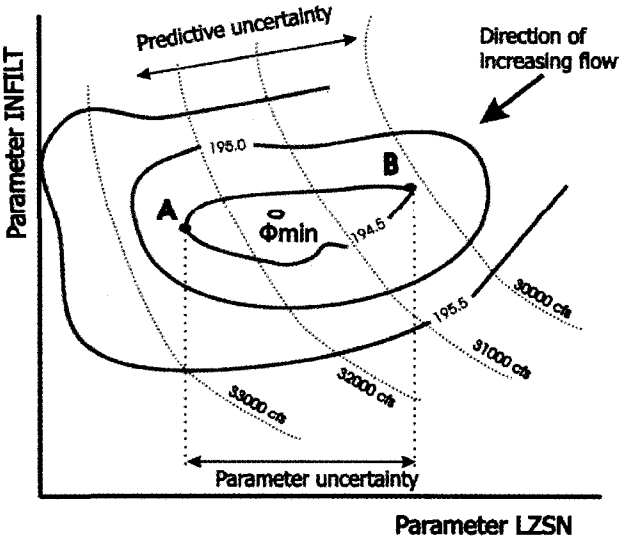


Figure 5.2 OF and critical points in parameter space

The confidence interval for a particular prediction or predictive interval is an interval between the maximum and minimum predictions that can be made by the model that is maintained in a "calibrated state" at a certain level of confidence. Predictive intervals typically are based on a number of assumptions. Firstly, the future behaves like the past. In a watershed context it is assumed that landuse, topography, and hydrography are unchanged for the calibration and predictive periods. The second assumption is that the model is well calibrated and that accurate calibration data are used. Because of the assumptions the predictive intervals tend to be optimistic.

The length of the calibration period and predictive periods are strongly dependent *(Kabaila, 1999)*. In groundwater modeling, *Bredehoeft (2003)* suggests that predications can be made with some confidence for a period equal to the period of the calibrated match. The confidence of the predictions made beyond the length of the calibration period will diminish rapidly *(Bredehoeft, 2003)*. Our experience indicates that the calibration period should be at least three times longer than the predictive period.

The predictive interval is typically defined as the confidence interval of an estimate of an individual response variable. For example, the 95% predictive interval indicates that 95% of the

time the predicted value will be within the interval. It was found that 1,000 runs were required to establish the confidence interval. According to a number of authors *(Vecchia and Cooley, 1987; Doherty, 2005)* the predictive interval consists of two parts: predictive confidence interval and predictive noise or error.

The predictive confidence interval is based on covariance of parameters and it reflects the uncertainty associated with the parameters. Uncertainty in parameter estimates is the main contributing factor to the predictive confidence intervals.

Predictive noise stems from the fact that a model cannot replicate nature in all details. An inherent random noise component exists in all observed data used to calibrate the model. Every prediction which a model makes has a certain degree of predictive noise.

Predictive intervals in the current research were estimated using the theoretical background from *Vecchia and Cooley (1987)* and practical implementation by *Doherty (2004)* in PEST. Version 10.0 of PEST is capable of evaluating not only predictive confidence intervals but noise which is added to the selected predictive output. The parameter upgrade vector for predictive analysis is calculated according to Eq. 5.5.

$$u = (J^T W J)^{-1} \cdot \left\{ J^T W h - \frac{Z}{2\lambda} \right\} \tag{5.5}$$

$$\left(\frac{1}{2\lambda}\right)^2 = \pm \frac{\Phi - h^T W h + h^T W J \left(J^T W J\right)^{-1} J^T W h}{Z^T \left(J^T W J\right)^{-1} Z + w_e^{-2}} \tag{5.6}$$

where Z – a vector of sensitivities of the model prediction to each parameter, h – vector comprises of observations employed in the calibration process, w_e – a weight associated with predictive error, Φ – OF comprised of lowest achievable Φ_{min} plus a modeler specified increment δ.

The positive sign in Eq. 5.6 is used when prediction in maximized and negative sign is used when the lower boundary of prediction is required. Predictive error is computed according to Eq. 5.7.

$$e = -\frac{w_e^{-2}}{2\lambda} \tag{5.7}$$

The objective of predictive analysis is to find minimum and maximum model predictions while maintaining the OF within (Φ_{min} +δ). Φ_{min} is the global minimum value of OF after calibration. PEST iteratively solves Eqs. 5.5-5.7 adjusting J, h, and Z during each iteration. The Marquardt lambda is adjusted correspondingly to keep the optimization stable.

PEST undertakes a line search in the direction of the parameter upgrade vector to determine the point when this vector crosses the ($\Phi_{min} + \delta$) contour. The line search algorithm monitors the value of prediction as well as value of the OF *(Doherty, 2004)*. Although most PEST variables for predictive analysis were taken as suggested in *Doherty (2005)*□ better results were obtained when the initial line search factor (INITSCHFAC) was set as low as 0.1 so that the line search begins close to current parameter values. For the first iteration the current parameter set is the calibrated set which generates the minimum possible OF.

Boundaries of OF confidence regions were computed as suggested in *Seber and Wild (1989)* and *Vecchia and Cooley (1987)*. Because δ is very small the assumption of linearised confidence interval around Φ_{min} was made. δ was found according to Eq. 5.8.

$$\delta = m\sigma^2 F_a(m, n-m) \tag{5.8}$$

where $F_\alpha(m, n-m)$ denotes the upper α probability point of a F distribution with m and $n-m$ degrees of freedom, m – number of parameters and n – number of observations, σ^2 – a reference variance estimated as $\Phi_{min}/(n-m)$.

It is also possible to obtain the critical values of prediction using a Monte Carlo scheme. With only several parameters up to 9, Monte Carlo may work well; however, with 10 or more parameters the number of optimization runs becomes enormous.

5.5 Results

5.5.1 Objective function

Single and compound OFs, as defined in Eqs. 5.3 and 5.4, were used. The compound OF consists of equally weighted 29 maximum, 584 minimum, and 2309 middle flow observations. The total calibration period was 9 years.

In the case of the compound OF it was noted that assigning different weights to the OF components leads to a different optimized parameter set. Some or all statistics can improve or deteriorate by changing the weighting factors. Increasing the value of the OF slightly above the minimum, it was observed that the parameter set which corresponds to the global minimum does not necessary represent the optimum solution. Some parameter sets with slightly higher OF can exhibit the same or better statistics and graphical plots (exceedance curves, residuals plot, etc.). This observation leads to the conclusion that the global minimum (assuming the proper global optimization technique is used) of the OF is not necessarily the best solution from a statistical or

graphical point of view. There are many other solutions close by which produce the same or even better fit.

A number of different statistics were used for comparison of observed and simulated data. It was noted that close to the global minimum of the OF, the statistics did not necessarily improve or degrade simultaneously. For example, it was observed many times that when CE improved, IA deteriorated.

Another observation is that good values of statistical coefficients do not necessarily represent the best fit. Statistics may be good but the comparison plots are poor when viewing the hydrograph of predicted and observed flows. Some statistics (based on squared values) favour high flows, some favour fit of mean and variance values (IA). This leads to the conclusion that there is no one single statistic which can be used to judge the best fit; a number of statistical coefficients should be compared simultaneously.

From this exercise it is recommended that at least three sets of statistics be used to obtain a representative picture of the fit between observed and simulated flows. One set is for high flows, one for middle flows, and one for minimum flows. Analyzing these statistics for several parameter sets around the global minimum one can usually find the parameter set which can be considered as the best fit. The concept of Pareto optimality may be used for these purposes *(Kasprzak and Lewis, 2000; Mattson and Messac, 2003)*.

In general, proper formulation of the OF based on objectives and the weight factors is the key to proper automatic calibration.

5.5.2 Sensitivity analysis

Parameter sensitivity is important for distinguishing parameters that are insensitive to model outputs and contribute little to model improvement. Composite parameter sensitivity, i.e., parameter change with respect to all weighted observations (Eq. 5.1) was a focus in the current research. The range of parameter values was taken from *USEPA (2000)* and adjusted to SNR climatic and geologic conditions. Calibrated parameter values and their sensitivities are presented in Table 5.1. SCE-UA and then GML methods with single (Eq. 5.3) and compound (Eq. 5.4) OF were used to calibrate the model.

Parameter values obtained by single and compound OFs are very close. However, their sensitivities in many cases are different. The most sensitive parameters are INFILT and AGWRC.

This was not surprising since INFILT is related to mean soil infiltration capacity and effectively controls the division of rain water to surface und subsurface storages. High values of INFILT result in increased baseflow, low values produce more overland flow and interflow. AGWRC is a groundwater recession rate, a ratio of one day's groundwater discharge to the previous day's discharge *(USEPA, 2000)*. Daily flows are sensitive to AGWRC since it impacts timing of the flow.

Table 5.1 Calibrated parameter values and their composite sensitivity

			Single OF		Compound OF	
Parameter	Definition	Range	Value	Sensi-tivity	Value	Sensi-tivity
SNOWCF	Snow gage correction factor	1.000-2.00	1.193	0.004	1.216	0.073
CCFACT	Condensation/convection melt factor	0.500-2.00	0.769	0.007	0.620	0.098
LZSN	Lower zone nominal storage (in.)	2.000-9.00	2.000	0.001	2.000	0.005
INFILT	Index to infiltration capacity of soil (in.)	0.010-0.20	0.035	0.072	0.023	0.560
AGWRC	Groundwater (GW) recession rate	0.850-0.99	0.976	0.086	0.990	1.562
KVARY	Variable GW recession (1/in.)	0.001-2.00	0.001	0.002	1.357	0.006
DEEPFR	Fraction of GW inflow to deep recharge	0.010-0.50	0.500	0.011	0.460	0.047
LZETP	Lower zone evapotranspiration (ET)	0.100-0.90	0.449	0.012	0.828	0.025
BASETP	Fraction of ET from baseflow	0.010-0.20	0.128	0.001	0.200	0.007
AGWETP	Fraction of ET from active GW	0.001-0.20	0.002	0.068	0.001	0.478
CEPSC	Interception storage capacity (in.)	0.010-0.40	0.010	0.032	0.010	0.124
UZSN	Upper zone nominal storage (in.)	0.050-2.00	2.000	0.002	1.122	0.014
NSUR	Roughness of overland flow	0.010-0.35	0.350	0.001	0.350	0.003
INTFW	Interflow inflow parameter	1.000-9.90	9.900	0.001	9.900	0.001
IRC	Interflow recession parameter	0.300-0.85	0.749	0.015	0.696	0.091

Parameters INTFW and NSUR have low sensitivity and consistently reached the limits of the parameter space. INTFW determines the portion of overland flow which becomes interflow. INTWF can be important when vertical percolation is retarded by a shallow, less permeable soil layer *(USEPA, 2000)*. It does not affect the volume of the hydrograph but may affect the timing of the peak. Due to its low sensitivity INTFW can be eliminated from future investigations.

The parameter correlation matrix was generated and analyzed. Two parameters KVARY and AGWRC show very high correlation. Because of its lower sensitivity, KVARY was fixed and excluded from further optimization.

Sensitivity of any particular parameter varies in a parameter space and as a result varies during the optimization process. Normally, the modeller is interested in a very local estimate of sensitivity evaluated immediately close to the global minimum. However for research purposes the Jacobian matrix during each iteration was analyzed and it was noted that sensitivity of some

parameters (IRC, INFILT, CCFACT, and KVARY) increases towards the optimum parameter set; other parameters (AGWETP, CEPSC, LZETP, and NSUR) become less sensitive. Sensitivity of some parameters (AGWRC, SNOWCF, LZSN, UZSN, and INTFW) changes little. Since sensitivity relates to curvature, the variations of parameter sensitivities during an GML optimization run gives an idea of the complexity of the OF surface.

Sensitivity analysis for the compound OF as defined in Eq. 5.1 was undertaken to test which parameters are the most sensitive to the maximum, minimum, and middle flows. In this particular case the relative composite parameter sensitivity is more convenient to use (Table 5.2). The relative sensitivity is obtained by multiplying the composite sensitivity and the value of the parameter *(Doherty, 2004)*. The fractional change in SNOWCF produces the largest impact on maximum flows. That was not surprising because the snow parameters regulate snow depth and rate and time of snow melt, which is the main factor contributing to the spring maximum flows *(USEPA, 2000)*. AGWRC is the most sensitive for simulating middle and low flows. AGWRC is a function of baseflow recession, and its importance probably is explained by generally shallow groundwater in the South Nation watershed. IRC is quite sensitive as well. It affects the rate at which interflow is discharged from the storage, mostly affecting the falling limb of the hydrograph. Lower values of IRC lead to the steeper falling slope of the hydrograph and interflow behaves like overland flow *(Bicknell et al., 2001)*.

Table 5.2 Relative composite sensitivity of HSPF parameters with respect to maximum, minimum, and middle flows

Parameter	OF1, max	OF2, min	OF3, mid
SNOWCF	8.4788	0.0214	0.0233
CCFACT	3.2457	0.0717	0.0625
LZSN	0.4057	0.0348	0.0079
INFILT	0.6034	0.0511	0.0064
KVARY	0.0535	0.0436	0.0019
AGWRC	6.9598	7.6849	0.3057
DEEPFR	0.2567	0.0984	0.0116
BASETP	0.0186	0.0066	0.0007
AGWETP	0.0004	0.0024	0.0001
CEPSC	0.0199	0.0057	0.0006
UZSN	0.5928	0.0572	0.0117
NSUR	0.0925	0.0008	0.0009
INTFW	0.4871	0.0167	0.0023
IRC	5.0532	0.1151	0.0350
LZETP	0.0749	0.0887	0.0138

Parameters AGWETP and CEPSC have the lowest relative sensitivity to all OFs and probably can be eliminated from the optimization process.

5.5.3 HSPF calibration

Calibration of HSFP was comprised of adjusting hydrological and snow parameters to generate daily flows as close as possible to the observed ones. Snow depths were calibrated; however, results are not presented here. The calibration period extended from Jan., 1990 to Dec., 1998. Lower and upper limits for the calibrated parameters (limits of parameter space) were taken from *USEPA (1997)* and adjusted for particular soil, hydrological, and geological conditions. After conducting sensitivity analysis and eliminating parameters whose values can be taken from maps and reviews, twelve parameters remained for detailed study. The list of calibration parameters and their best values are presented in Table 5.1.

Calibration was tedious and involved many optimization runs. Results of performance statistics for all daily, minimum, and maximum flows are presented in Table 5.3. It is seen that both single and compound OFs demonstrate satisfactory performance statistics. The compound OF had superior results for simulation of minimum and maximum flows; however, total daily flows are simulated better with the single OF.

Table 5.3 Performance statistics for observed and simulated daily flows computed using single and compound OF for calibration period 1990-1997

	r^2	ME	MAE	PME	PMAE	NS	CE	IA	DRMS
Single OF									
Daily flows	0.808	−179.2	638.3	−20.8	53.1	0.805	0.682	0.835	1484
Max. flows	0.674	−3797	4270	−34.4	30.5	−0.148	0.003	0.610	5211
Min. flows	0.275	63.3	72.5	23.2	36.6	−10.9	−1.41	0.347	120.1
Compound OF									
Daily flows	0.811	−54.9	680.9	−47.9	72.2	0.779	0.660	0.834	1581
Max. flows	0.690	33.3	4187	−6.9	20.6	−0.135	0.023	0.617	5183
Min. flows	0.148	28.7	46.2	8.5	31.5	−7.82	−0.53	0.428	103.6

In addition to statistical analysis, graphical comparisons of observed and simulated flows were made including percent exceedance flow curves, scatter plots, graphs of residuals, and plots of cumulative differences. Differences between simulated and observed flows are plotted in Figure 5.3. Small residuals indicate good fit. Large residuals, quite naturally, are typical for spring maximum flows. Taking into account that maximum flows varied from 20,000 to 32,000 cfs, the difference between observed and simulated flows can be up to 30-50%. Almost equal distribution

of positive and negative residuals indicates that the model does not have a serious bias towards over- or underprediction.

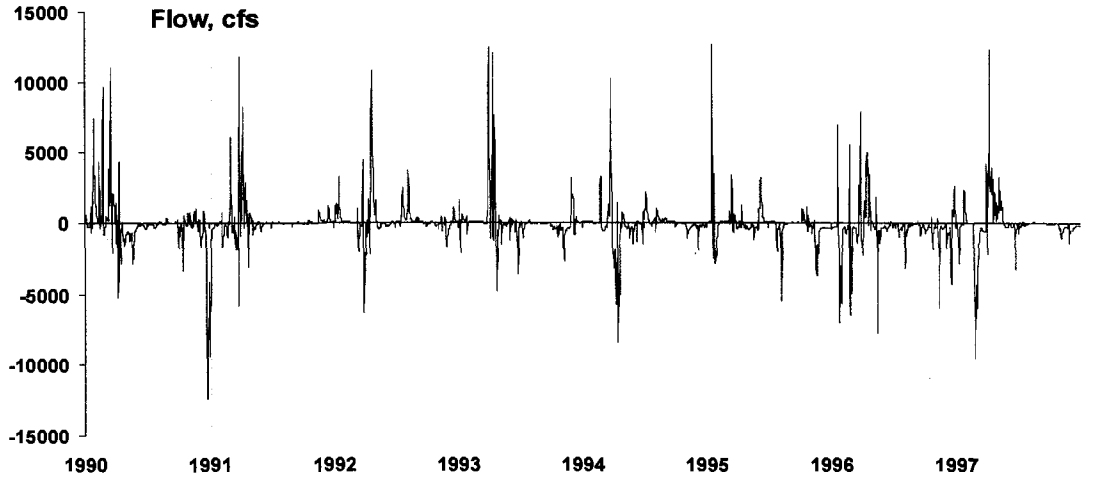


Figure 5.3 Residual plot of simulated and observed daily flows using single OF

Long term trends in simulation are evaluated using a cumulative difference plot which illustrates the periodicity of variations in the simulated and observed time series. In a good model, residuals should be centered around zero with no positive or negative cumulative trend and the sum of residuals should be zero. Application of cumulative residuals for interpretation of model fit is described well in *Lin et al. (2002)*. The plot, not presented here, had no positive or negative cumulative trend.

Based on statistical and graphical comparisons the calibration of daily flows using single and compound OFs was considered satisfactory.

5.5.4 Predictive analysis of peak flows

After model calibration the predictive uncertainty of the daily flows was investigated. The primary interest was to examine the predictive uncertainty of high spring and low autumn flows. Using PEST the OF was allowed to vary at the specified confidence level (95%) and parameter sets which generate maximum and minimum flows at a specific time point were recorded.

Two parallel approaches were used, calibration with single and compound OFs. Each approach produces the "best" parameter set which was then used to predict maximum flows in 1998-2000. It should be noted that the model was not specifically calibrated for maximum flows. The intention was to have a fairly good model for extreme events, monthly volumes, and water

balance in general. Using the observation weighting scheme as well as changing weights of individual OFs, it is possible to obtain better calibration of maximum flows at the expense of minimum flows or volumes and vice versa.

Observed and simulated maximum daily flows for calibration and predictive periods are presented in Table 5.4. Calibration performed with the single OF tends to underpredict maximum flows; at the same time the compound OF often overpredicts peak events. In general, the compound OF produces better fit for maximum flows for both calibration and predictive periods.

Table 5.4 Observed and simulated maximum daily flows (cfs) for calibration and predictive period

	Calibration period						Predictive period		
Year	1992	1993	1994	1995	1996	1997	1998	1999	2000
Observed flow	16,280	31,289	24,614	12,078	15,750	28,781	31,819	23,731	17,763
Single OF	12,198	32,460	17,713	10,536	11,124	31,428	28,690	22,948	10,284
% error	−25.1	3.7	−28.0	−12.8	−29.4	9.2	−9.8	−3.3	−42.1
Compound OF	15,288	43,383	21,021	13,338	17,149	41,041	40,874	28,326	16,376
% error	−6.1	38.6	−14.6	10.4	8.9	42.6	28.5	19.4	−7.8

Then a study of predictive interval for the spring maximum flows was conducted. The OF was allowed to vary between 0.05 and 7% above the global minimum and the parameter sets which produce the maximum and minimum values of flow prediction were recorded (Figure 5.4).

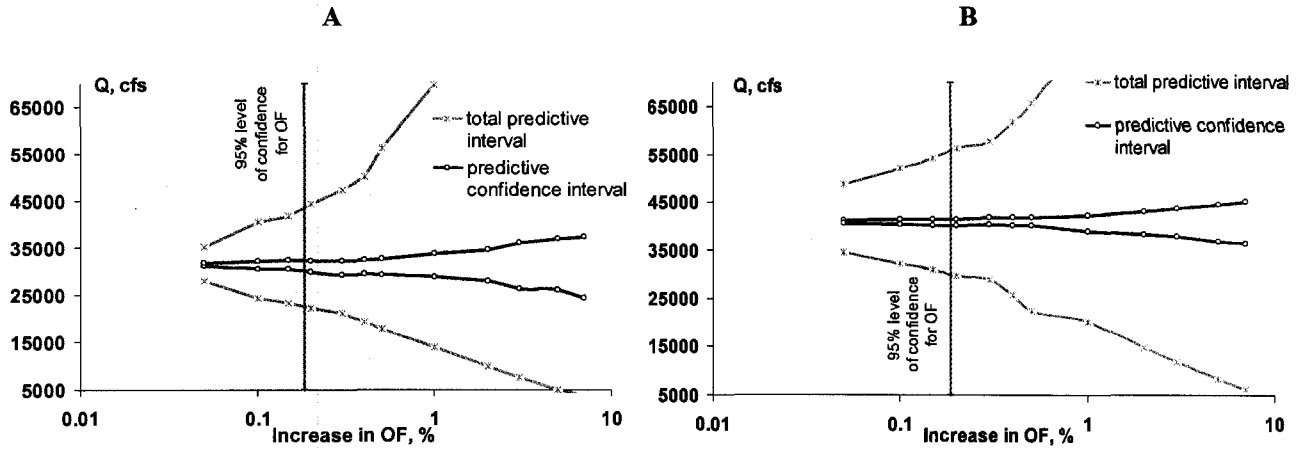


Figure 5.4 Predictive uncertainty of spring maximum flow in 1997. A – single OF, B – compound OF

Two types of predictive intervals are recorded. The smaller interval between dark lines with the circles represents predictive confidence interval which is based on covariance of parameters, i.e.,

parameter uncertainty. The larger interval represents the total predictive interval which includes the predictive noise. Predictive noise was determined using Eqs. 5.5-5.7. Predictive noise is intended to account for random noise in the observed and calibrated data. The vertical line on the graph indicates 95% confidence level of the OF computed according to Eq. 5.8. From the graph, the largest portion of the predictive interval consists of the predictive noise. Predictive noise in the case of the compound OF is greater than in the case with the single OF. Predictive confidence interval, which is a function of the parameter uncertainty, is slightly wider for the compound OF.

Numeric values of total predictive intervals are presented in Table 5.5. It is seen that increasing the OF by 0.2%, which is approximately equal to 95% confidence limits of the OF, leads to predictive uncertainty of 22,319 cfs with the single OF and 26,614 cfs with the compound OF.

Table 5.5 Total predictive intervals for 1997 at different values of the OF

		Single OF			Compound OF		
Increase in OF (%)		Flow, cfs	% deviation from observed flow	Difference, cfs	Flow, cfs	% deviation from observed flow	Difference, cfs
Observed max flow		28,781			28,781		
Simulated max flow		31,428			31,428		
	Prediction						
0.05	Max	35,843	24.5		48,887	69.8	
	Min	26,172	−9.1	9,671	34,470	19.7	14,417
0.2	Max	44,339	54.0		56,439	96.1	
	Min	22,020	−23.5	22,319	29,825	3.6	26,614
0.5	Max	58,405	102.9		65,730	128.4	
	Min	17,241	−40.1	41,164	20,893	−27.4	44,836
1.0	Max	73,365	154.9		82,203	185.6	
	Min	13,597	−52.7	59,769	20,179	−29.8	62,024
5.0	Max	196,95	584.3		204,09	609.1	
	Min	4,972	−82.7	191,984	8,258	−71.3	195,836

Predictive confidence intervals for one calibration year (1997) and three predictive years (1998-2000) are shown in Figure 5.5. The intervals were defined as a difference between maximum and minimum predictions. It is seen that the intervals are relatively small. For 95% confidence limits of the OF (increase of 0.2% above Φ_{min}) the predictive confidence intervals vary between 400 and 2,200 cfs or 1-8% deviation from the observed peak flow. The compound OF produces significantly smaller predictive confidence intervals than the single OF meaning that parameter uncertainty is smaller when the compound OF is used.

In the next exercise, the predictive confidence interval and predictive noise were computed at a daily time step during the spring flood (Figure 5.6). It is seen that the predictive confidence

intervals closely follow the calibrated flow. The total predictive intervals are significantly wider for higher flows mainly due to high predictive noise.

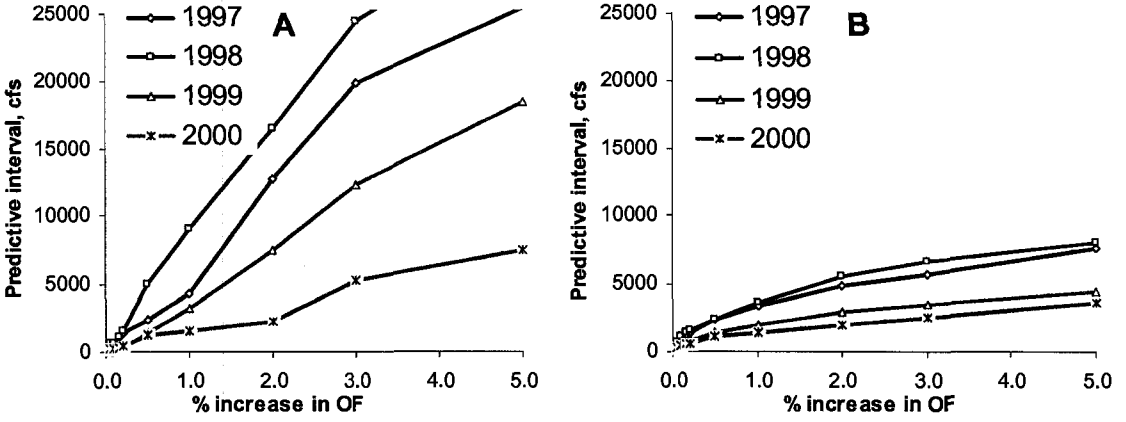


Figure 5.5 Confidence intervals for spring maximum flows at different increase of the OF above the global minimum. A – single OF, B – compound OF

Analyzing the statistics for maximum and minimum predicted flood flows in Table 5.6, it can be concluded that the OF increase leads to an insignificant degradation of statistics. Some statistics are more and others are less sensitive to increasing the OF. Several statistics (r^2, ME, and DRMS) should be discouraged for use for comparison of observed and simulated data due to their insensitivity to significant changes in the OF. It was noted that based on statistics alone it is difficult to choose the best fit and as a result the best parameter set.

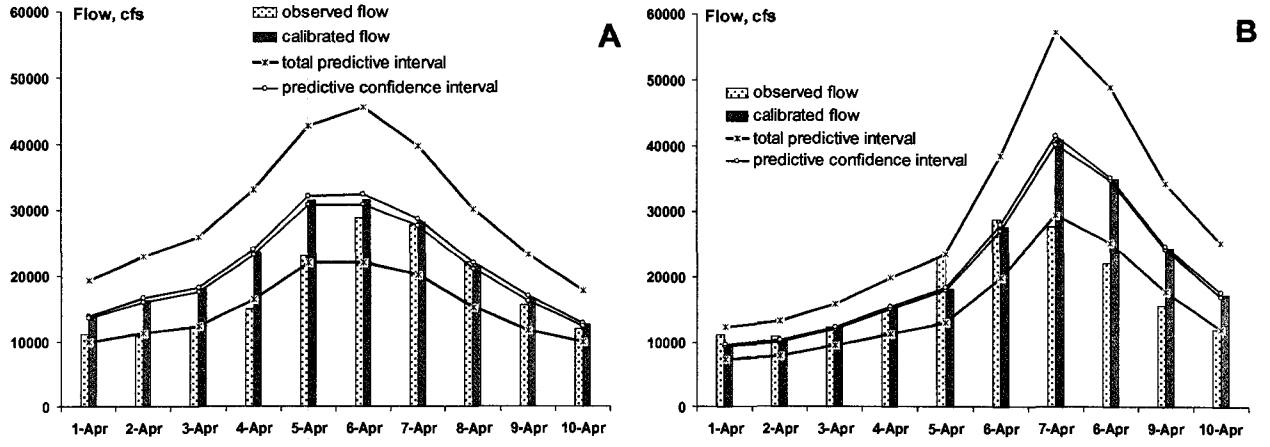


Figure 5.6 Predictive and confidence intervals for spring maximum flow in 1997. A – single OF. B – compound OF

HSPF parameter values that correspond to minimum and maximum predicted spring flows are presented in Table 5.7. The single OF was used to generate the table. Snow parameters (SNOWCF and CCFACT) are the most variable, ET and groundwater related parameters (LZETP, AGWETP, and AGWRC) are the least variable.

Table 5.6 Statistics of maximum and minimum flood predictions under different values of the OF (predictive period, 1998-2000)

Increase in OF, %		r^2	ME	MAE	PME	PMAE	NS	CE	IA	DRMS
0.05	max	0.747	−442.9	831.1	−68.18	104.7	0.732	0.588	0.777	1914
	min	0.765	−486.1	825.8	−66.81	103.3	0.739	0.591	0.776	1889
0.10	max	0.741	−431.1	838.4	−68.35	105.1	0.727	0.584	0.776	1934
	min	0.762	−495.9	833.8	−69.62	105.8	0.735	0.587	0.774	1905
0.20	max	0.734	−416.8	843.1	−68.48	105.4	0.720	0.582	0.776	1958
	min	0.786	−537.3	833.9	−72.49	107.3	0.737	0.587	0.772	1896
0.50	max	0.704	−341.3	876.1	−70.18	107.7	0.661	0.566	0.772	2156
	min	0.789	−550.1	833.0	−75.20	109.3	0.735	0.587	0.772	1904
1.00	max	0.696	−324.6	886.0	−70.40	108.2	0.641	0.561	0.771	2218
	min	0.789	−529.6	832.7	−70.96	106.5	0.739	0.587	0.772	1890
3.00	max	0.675	−266.4	920.1	−70.68	109.1	0.561	0.544	0.765	2452
	min	0.790	−538.9	839.4	−69.79	105.6	0.733	0.584	0.770	1912
calibrated		0.757	−464.1	826.7	−68.26	104.5	0.738	0.590	0.777	1892

Table 5.7 Parameter values corresponding to maximum and minimum flood predictions under different values of the OF

Increase in OF, %		SNOWCF	CCFACT	INFILT	AGWRC	DEEPFR	BASETP	AGWETP	UZSN	IRC	LZETP
0.05	max	1.197	0.771	0.037	0.975	0.499	0.128	0.003	1.992	0.744	0.447
	min	1.193	0.776	0.038	0.974	0.500	0.129	0.003	1.993	0.749	0.447
0.20	max	1.199	0.771	0.037	0.974	0.497	0.129	0.003	1.992	0.742	0.448
	min	1.186	0.775	0.038	0.974	0.500	0.129	0.003	1.993	0.750	0.447
0.50	max	1.231	0.788	0.036	0.975	0.500	0.127	0.003	2.000	0.738	0.446
	min	1.156	0.769	0.038	0.974	0.500	0.136	0.004	1.995	0.758	0.442
1.00	max	1.213	0.772	0.035	0.974	0.487	0.124	0.002	2.000	0.732	0.452
	min	1.166	0.787	0.037	0.974	0.500	0.133	0.003	2.000	0.759	0.446
3.00	max	1.432	0.812	0.039	0.974	0.455	0.200	0.002	2.000	0.722	0.474
	min	1.166	0.829	0.043	0.973	0.500	0.153	0.005	2.000	0.790	0.447
7.00	max	1.550	0.745	0.037	0.976	0.418	0.171	0.004	1.691	0.666	0.497
	min	1.004	0.832	0.042	0.972	0.500	0.174	0.005	2.000	0.795	0.435

5.5.5 Predictive analysis of low flows

Predictive uncertainty of low flows was studied using the same approach as described above. However, together with the absolute minimum flow, cumulative values of daily flow for 5, 10, 15, 20, 25, and 30 day periods were used. Historical data for the South Nation watershed indicate that September is the month when low flows are most likely to be observed; therefore, September was chosen as the reference month to study low flows. Comparison of observed and simulated cumulative low flows for September 1997 in ft^3/s is presented in Table 5.8. The model underpredicts and overpredicts minimum flows in various years. Although Table 5.3 suggests that the compound OF should simulate minimum flows more accurately than the single OF, this was not evident from the cumulative minimum flows in Table 5.8. In some case the single OF is better, in others the compound OF generates better results.

Table 5.8 Observed and simulated cumulative low flows* ($\times 10^9$) for calibration and predictive periods

	Calibration period							Predictive period		
	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Observed minimum flows										
5 days, cumulative	0.20	0.29	0.37	0.29	0.15	0.18	0.16	1.50	0.09	0.16
10 days, cumulative	0.58	1.15	1.56	1.01	0.75	0.81	0.58	5.40	1.03	0.51
15 days, cumulative	1.16	2.58	3.51	2.24	1.85	3.23	1.41	10.62	3.42	2.96
20 days, cumulative	2.00	4.93	5.70	4.17	3.25	10.54	2.76	16.83	5.76	9.57
25 days, cumulative	3.05	11.46	8.41	6.08	4.76	16.66	4.97	23.30	9.48	15.49
30 days, cumulative	4.83	19.73	13.86	10.95	6.54	25.77	7.86	40.26	13.41	20.92
Calibrated using single OF										
5 days, cumulative	0.10	1.21	0.30	1.20	0.20	0.12	0.20	0.30	0.12	0.40
10 days, cumulative	0.35	4.49	1.16	3.61	0.96	0.88	0.70	1.06	1.82	1.38
15 days, cumulative	0.87	8.75	2.56	6.98	2.23	3.16	1.76	2.08	4.62	3.66
20 days, cumulative	1.83	14.36	4.35	11.70	3.93	6.32	3.29	4.67	8.25	7.87
25 days, cumulative	2.82	25.30	6.74	16.76	6.19	9.46	5.93	7.33	14.91	12.57
30 days, cumulative	4.78	36.84	11.06	25.00	8.76	14.81	9.02	17.49	20.54	16.99
Calibrated using compound OF										
5 days, cumulative	0.11	0.77	0.29	0.78	0.16	0.09	0.17	0.26	0.10	0.28
10 days, cumulative	0.38	2.83	1.09	2.28	0.79	0.71	0.61	0.86	1.74	0.96
15 days, cumulative	0.94	5.47	2.34	4.40	1.81	2.59	1.54	1.66	4.65	2.61
20 days, cumulative	1.90	9.05	3.92	7.43	3.13	5.07	2.82	3.74	8.17	5.36
25 days, cumulative	2.87	17.03	6.01	10.65	4.91	7.31	4.94	5.71	16.47	8.38
30 days, cumulative	4.65	24.74	9.59	15.96	6.91	11.56	7.29	20.13	22.60	11.62

* All flows are in ft^3/s

Simulated minimum flows can differ more than two-fold compared to the observed values. However, when it comes to daily flow values, absolute differences are not high.

Year 1997 was taken as a reference and the OF was allowed to vary up to 7% above the global minimum. Maximum and minimum predictive values of the low flows were recorded (Figure

5.7). A 0.2% increase of the OF above the global minimum corresponds to 95% confidence level in the OF. Logically predictive uncertainty increases with raising the OF and increases its confidence level. It can be seen that predictive confidence intervals are smaller with the compound OF, which suggests the smaller parameter uncertainty using the compound OF. However, total predictive interval due to high predictive noise is smaller for the single OF.

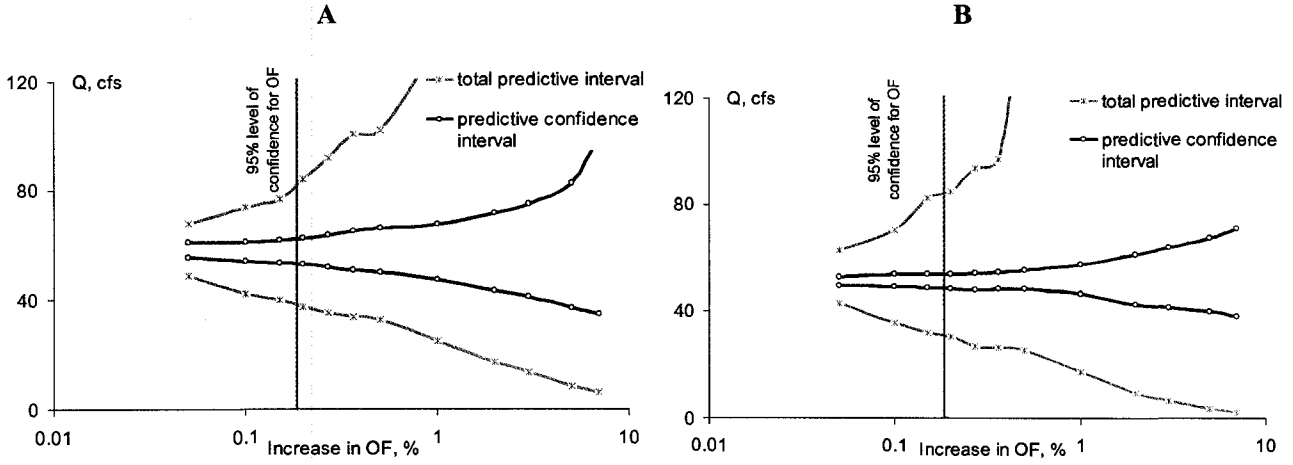


Figure 5.7 Predictive uncertainty of autumn minimum flow in 1997. A – single OF, B – compound OF

Predictive confidence intervals for one calibration year (1997) and three predictive years (1998-2000) are shown in Figure 5.8. It can be seen that using compound OF is more beneficial because the predictive confidence intervals tend to be smaller.

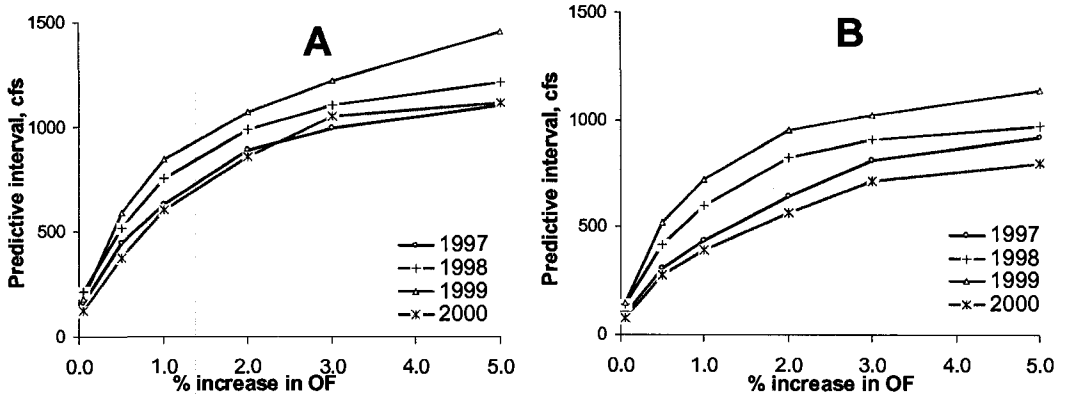


Figure 5.8 Predictive confidence interval for minimum 30 days cumulative flow at different increases of the OF above the global minimum. A – single OF, B – compound OF

Predictive confidence interval and predictive noise were computed at every time step during autumn low flows in 1997. An example with the compound OF is shown in Figure 5.9. Observed and calibrated flows are shown in circles. It was noted that observed flows with both the single and compound OFs are fairly well simulated. In general, the compound OF is preferable when minimum flows are the primary focus, due to smaller predictive confidence interval and predictive noise.

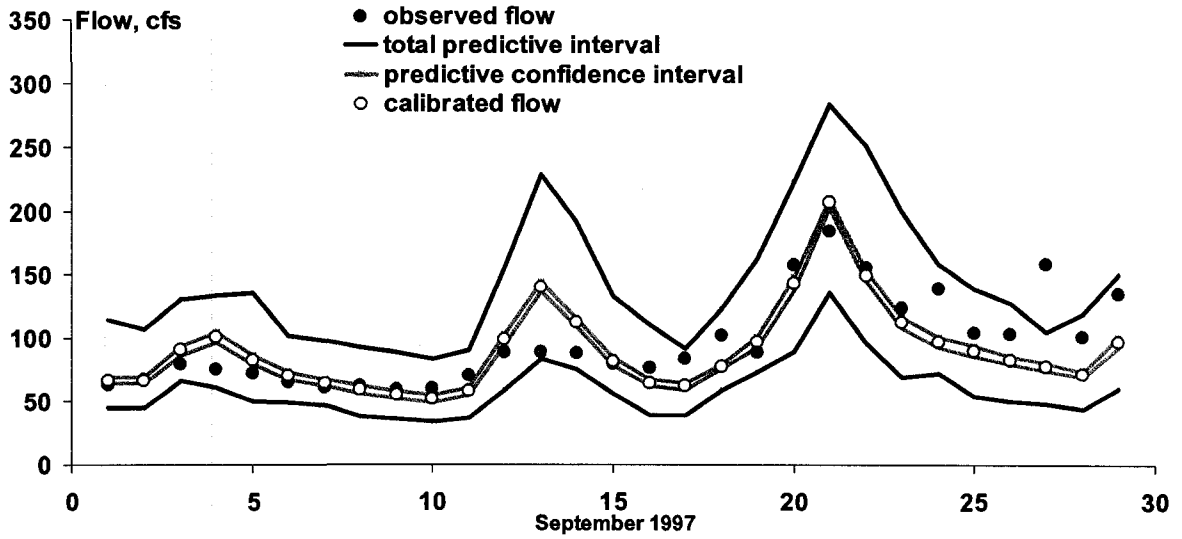


Figure 5.9 Predictive and confidence intervals for autumn minimum flow using the compound OF

5.6 Conclusions

PEST was found to be an efficient tool for sensitivity analysis, calibration, and predictive analysis of HSPF.

The lowest OF value does not necessarily correspond to the optimum solution. A number of parameter sets close to the global minimum have the same or even better performance statistics and graphical fit. Proper OF formulation including weights is the key to accurate automatic calibration.

Parameter sensitivity is an important step in parameter optimization. Sensitivity was estimated close to the achievable minimum value of the OF. When using the single OF the most sensitive parameters to daily flows are AGWRC and INFILT, the least sensitive are INTFW and NSUR. With the compound OF the most sensitive parameters to maximum flows are SNOWCF and AGWRC; to minimum flows are AGWRC and IRC. The least sensitive parameters to maximum

and minimum flows are AGWETP and CEPSC. Analyzing the Jacobian after each iteration run it was noted that parameter sensitivity changes during the optimization. Sensitivity of some parameters (IRC, INFILT, CCFACT, and KVARY) increases as the optimum parameter set is reached. On the contrary, some parameters (AGWETP, CEPSC, LZETP, and NSUR) become less sensitive around the global minimum.

Total predictive interval was computed as a sum of predictive confidence interval (uncertainty in parameter estimates) and random errors in the dependent variable (daily flows). The predictive intervals were computed for maximum spring flows and minimum autumn flows using single and compound OFs.

Based on predictive analysis of maximum spring flows it was concluded that predictive confidence interval, and as a result parameter uncertainty, is relatively small. At 95% confidence limits of the OF the uncertainty of prediction in maximum flows varies between 400 and 2,200 cfs or up to 8% deviation from the observed maximum flows. The smaller uncertainty pertains to the compound OF. Total predictive noise for maximum flows can be very high. At 95% confidence limits of the OF the total uncertainty in prediction can reach 22,000-26,000 cfs or 75-95% deviation from the observed peak flow, mostly due to the predictive noise in the data. Higher predictive noise pertains to the compound OF.

Changes in performance statistics and in parameter values around 95% confidence limits of the OF are not significant. The statistical parameters of coefficient of determination, ME, and DRMS should be discouraged to be used for comparison of the observed and simulated data due to their insensitivity to significant changes in the OF.

Total predictive interval for minimum flows is not high, at 95% confidence limits of the OF it can reach about 50 cfs for an observed flow of 60 cfs. The confidence interval due to parameter uncertainty does not exceed 10-15% of the total predictive uncertainty interval. The compound OF is more desirable to use for prediction of low flows due to smaller predictive confidence intervals.

5.7 References

AIAA.1998. Guide for the verification and validation of computational fluid dynamics simulations. American Institute of Aeronautics and Astronautics: Reston, VA.

ASTM. 1984. Standard practice for evaluating environmental fate models of chemicals, designation E978-84. American Society of Testing Materials: Philadelphia, PA.

Bicknell B R, Imhoff J C, Kittle Jr J L, Jobes T H, Donigian Jr A S. 2001. Hydrological simulation program HSPF. Version 12. User's Manual. U.S. Environmental Protection Agency: Athens, GA.

Bredehoeft J D. 2003. From models to performance assessment: The conceptualization problem. Ground Water 41: 571-577.

Chow V T, Maidment D R, Mays L W. 1988. Applied hydrology, McGraw-Hill Inc.: New York, NY.

Crawford N H. 1999. Snowmelt calibration. Hydrologic Journal, Hydrocomp, Inc., □□□.□□□□o□o□□.□o□ [accessed July 20th, 2006].

Doherty J. 2004. PEST – model-independent parameter estimation. User manual: 5th edition. Watermark Computing: Corinda, Australia.

Doherty J, Johnston J M. 2003. Methodologies for calibration and predictive analysis of a watershed model. J. American Water Resources Association 39: 251-265.

Doherty J. 2005. Addendum to the PEST manual. Watermark Numerical Computing: Corinda, Australia.

Donigian A S Jr, Imhoff J. 2006. History and evolution of watershed modeling derived from Stanford Watershed Model (SWM). In Watershed Models, Singh V, Frevert D (eds.). Taylor & Francis/CRC Press: Boca Raton; FL, 680.

Duan Q Y, Sorooshian S, Gupta V. 1992. Effective and efficient global optimisation for conceptual rainfall-runoff models. Water Resour. Res. 28: 1015-1031.

Hoyer D P, Larson A M. 2005. *Belle Fourche River watershed assessment.* Final report and TMDL for Butte, Lawrence, and Meade Counties, South Dakota: Rapid City, SD.

Iskra I, Droste R. 2005. Application of non-linear automatic optimization techniques for calibration HSPF. WEFTEC conference, October 29 - November 2, 2005: Washington; 8778-8798.

Kabaila P. 1999. The relevance property for prediction intervals. Journal of Time Series Analysis T20: 655-662.

Kasprzak E, Lewis K. 2000. An approach to facilitate decision tradeoffs in Pareto solution sets. Journal of Engineering Valuation and Cost Analysis 1: 173-187

Lin E Y, Wei L J, Ying Z. 2002. Model-checking techniques based on cumulative residuals. Biometrics 58: 1-12.

Marquardt D. 1963. An algorithm for least-squares estimation of nonlinear parameters. SIAM J. Appl. Math. 11: 431-441.

Mattson C A, Messac A. 2003. Concept selection using s-Pareto frontiers. AIAA Journal 41: 1190–1204.

McCuen R, Snyder W. 1986. Hydrologic modeling: statistical methods and applications. Prentice-Hall: Englewood Cliffs, NJ.

Nash J E, Sutcliffe J V. 1970. River flow forecasting through conceptual models, Part-I, A discussion of principles, Journal of Hydrology 10: 282–290.

Schaake J. 2003. Introduction. In calibration of watershed models, Duan Q, Gupta HV, Sorooshian A N, Turcotte R (eds), American Geophysical Union: Washington, DC; 345.

Seber G A F, Wild C F. 1989. Nonlinear regression. John Wiley and Sons: New York.

USEPA. 2000. BASINS Technical Note 6: Estimating hydrology and hydraulic parameters for HSPF, http://www.epa.gov/waterscience/basins/tecnote6.pdf [accessed July 20th, 2006].

USEPA. 1997. Technical Panel: Guiding principles for Monte Carlo analysis. EPA /630/R-97/001.

Vecchia A V, Cooley R L. 1987. Simultaneous confidence and prediction intervals for nonlinear regression models with application to a groundwater flow model. Water Resources Research 23: 7, 1237-1250.

Willmott C J. 1981. On the validation of models. Phys. Geogr. 2: 184-194.

CHAPTER 6

IMPACT OF SPATIAL SCALING AND RESOLUTION ON SENSITIVITY AND UNCERTAINTY OF A WATERSHED MODEL

Igor Iskra and Ronald Droste
Department of Civil Engineering, University of Ottawa
161, Louis Pasteur, Ottawa, K1N 6N5, Canada

The novel contribution of the presented paper is in studying scaling dependency of HSPF parameters and possibility of transferring their values to the neighboring watersheds for prediction peak flows and low flows. In addition, the impact of DEM resolution on HSPF topographical parameters (area, cross-section geometry, length and slope of overland plane, stream length, and elevation differences) was studied.

6.1 Abstract

Spatial scaling of 14 watersheds in Eastern Ontario was assessed using the product of moments. The logarithms among the *k*th moments of average annual runoff, river length, peak flows, and watershed area were found to be linear, which indicates applicability of simple scaling laws for the studied watersheds. Hydrologic Simulation Program Fortran (HSPF) parameters were studied at different spatial scales and it was found that HSPF parameters are not scale dependent and can be transferred from neighboring watersheds. Simple scaling and HSPF were used to simulate daily flows for ungaged watersheds. It was found that HSPF can be an efficient tool in predicting annual and peak flows in ungaged watersheds; however, the model cannot be used to quantify low-flow conditions and generally assess stress states in the watershed. The impact of digital elevation model (DEM) resolution on HSPF topographical parameters (area, cross-section geometry, length and slope of overland plane, stream length, and elevation differences) was studied. It was found that small watersheds are more sensitive to DEM resolution. Area is a sensitive parameter for simulation of minimum and maximum flows. Assuming that the error in maximum flows should not exceed 6% and the error in minimum flows should be within 12%, it is critical to have DEM at 35 m resolution for small watersheds (<200 km^2) and 100 m resolution for larger areas.

Keywords: Watershed modeling; HSPF; Spatial scaling; DEM; Spatial resolution

6.2 Introduction

6.2.1 Scaling in watershed modeling

Natural processes behave differently at different scales. Recent research has produced orderly ways, scaling laws, to connect various phenomena in biology, ecology, hydrology, and other sciences *(Klarreich, 2005; Savage et al., 2004; West et al., 1997)*. Although the scaling concept in hydrology is relatively new (*Smith, 1992; Yue and Gan, 2004)*, it provides a tool for modeling hydrological processes in ungaged watersheds.

Gupta and Dawdy (1995), Gupta et al. (1994), and Smith (1992) explored spatial characteristics of flood peaks. *Burlando and Rosso (1996), Kumar and Foufoula-Georgiou (1993), and Gupta*

and Waymire (1990) applied scaling theory to analyze the spatial distribution of rainfall events. *Yue and Gan (2004)* studied scaling properties of annual streamflow. One of the conclusions from these researchers is that if a hydrological process shows simple scaling or self-similarity, it is possible to transform the process via a scaled probability distribution to wide spatial scales. Often geographical and climatic factors are important to take into account when hydrological processes are to be predicted in neighbouring ungaged watersheds.

Scaling issues become more complicated when watershed models are involved. Although hydrological processes at larger scales are the resultant of smaller scale processes, models representing these processes vary considerably. Thus a major objective in watershed modeling is to understand the impact of changes in scale on hydrological processes and model parameters. The importance of the scale has long been recognized by water resources engineers; however, many questions related to modeling scale and model parameters remain unresolved. Do the mathematical equations developed and validated for laboratory scale apply at the watershed scale? Where are the upper and lower limits of applicability for a watershed model? What approaches does the model use for spatial scaling?

Testing similarity of hydrological parameters on different scales originated from soil physics *(Miller, 1980)* and now it is enthusiastically being explored by some watershed modelers. Two fundamental hypotheses underlie their research. The first is that larger watersheds are more linear (governed by simpler laws) in their response than smaller watersheds *(Armorocho, 1963)*. Another view of this is that input data are often absent or of a much lower quality at a larger scale, which results in a tendency to use simpler, empirical models at a larger scale *(Heuvelink, 1998)*. The second hypothesis is that different processes dominate at different scales, and as a result, different questions arise at different scales. Also, it is important to remember that different processes are ignored in the simplification steps of model development.

Refsgaard and Butts (1999) have suggested a classification of scaling approaches in hydrological modeling, which is schematically presented in Figure 6.1. According to their classification, upscaling means the equations and model parameters are modified or substituted when moving from a smaller to a larger watershed. Aggregation is a process when smaller scale equations are aggregated into a larger scale. The least reliable and thus rarely used approach is to assume that small scale equations are valid for large scales. HSPF, the model used in the current research, exploits embedded large-scale equations. Thus HSPF equations and parameters are explicitly developed for watershed conditions and not necessarily for point or hill-slope scale.

Scaling is a significant problem in hydrologic modeling and it will likely remain a problem for a long time. The main reasons for this are: the governing equations in the code are difficult to solve effectively for different scales. Temporal lumping and spatial lumping cause uncertainty which is difficult to estimate, resolution of input data often needs to be different at various scales, and scaling issues are likely to influence the search for the global optimum in model calibration. This paper aims to improve understanding of spatial scaling and resolution issues in watershed modeling, particularly in HSPF.

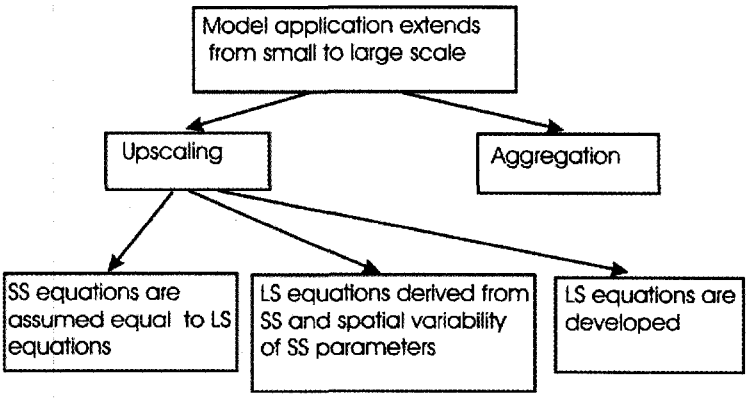


Figure 6.1 Methods of extending hydrological processes from small-scale (SS) to large-scale (LS) model application

6.2.2 Spatial resolution

The fundamental element of spatial definition of a model as well as one of its controlling boundary conditions is topography. Topography plays an important role in flux and energy distribution in the watershed. In many geographical information system (GIS)-based models, including Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) coupled with HSPF, landscape topography is represented by the DEM. DEM provides easy and reliable measurement of slope, aspect, flow length, drainage divide, and channel network.

The impact of spatial resolution on model simulation has not been widely studied. Often a simulation is done with the most detailed data available and the question of how the model performs if these data are absent remains unanswered.

In the current research the impact of DEM on model response is explored through spatial aggregation of the detailed data. DEM aggregation leads to the coarsening of terrain attributes (elevation, slope, curvature, etc.). *Kuo et al. (1999)* highlighted that variation of curvature with the grid size is an important constraint in hydrological simulation. As the grid size increases,

details of lateral flow pathways are lost causing subsequent effects on transport and accumulation of water across the watershed. Also, *Wolock and McCabe (2000)* noted that since watersheds vary in topography, it is expected that the aggregation effect is site specific and related to terrain roughness and drainage density.

Vivoni et al. (2005) studied the impact of DEM on the hydrograph simulations and spatial distribution of surface-subsurface watershed response in MIKE-SHE. They concluded that over a broad range of DEM resolution (30 to 90 m) many processes exhibit weak resolution dependence; total runoff and evapotranspiration (ET) are approximately constant. Their results further suggest that the critical DEM size is about 100 m and it is related to the mean hill slope length of the studied watershed. *Vazquez et al. (2002)* found that the effective grid resolution to achieve reasonably accurate prediction in MIKE-SHE is 600 m.

For watershed modeling, two aspects are important for DEM selection: the quality and resolution of data. Quality refers to the accuracy of elevation data, and resolution refers to the horizontal and vertical grid spacing. In the current research a square-grid DEM with horizontal resolution of 25 $\times$ 25 m and vertical resolution of 0.5 m was used.

From a spatial resolution perspective, the purpose of this research is two-fold. Firstly, it is to identify how DEM resolution impacts HSPF topographical parameters and which of them are the most sensitive to the changes in DEM. The second objective is to understand the sensitivities of the maximum and minimum flows at various size watersheds to small and large changes of HSPF topographical parameters.

6.2.3 Size of the watershed

Hydrological processes are often studied and modeled at the point scale. As a result, many hydrological models are originated and tested at the hill slope scale. Examples are TOPMODEL *(Beven, 1995)* and MIKE-SHE *(Bathurst, 1986)*. Transferring a field measurement or governing parameter to a watershed scale is problematic with these models. Various aggregating and upscaling approaches exist, the main principle of which is to perform another calibration when the scale is changing *(Vazquez et al., 2002)*. Some other physically-based models such as HSPF and SWAT are developed and tested explicitly for the watershed scale. Thus, the equations and parameters are generally representative at watershed scale and not necessarily at point scale. For example, the simple conceptual Swedish Hydrological Model (HBV) is reported to be stable over

a wide range of scales from several tens of thousands of kilometres to the continental scale of the Baltic Sea *(Bergstrom and Graham, 1998)*.

Intuitively the physically-based models suggest that the finer grid (landuse segment in HSPF, hydrological response unit in SWAT) size and time steps result in better representation of a watershed. However, owing to computation limitations, data availability, and model structure, these models can only be applied when effective grid (or segment) parameters are used.

Understanding model behaviour at different scales opens the possibility of transferring model parameters to neighbouring ungaged watersheds. The hypothesis of scale-dependency of physically-based parameters is explored in this study. Using neighbouring watersheds with similar climate, soil, and landuse conditions for the approximation of model parameters in ungaged watersheds is examined. Scaling relations for annual stream flow and watershed size suggested by *Yue and Gan (2004)* were tested.

6.3 Materials and methods

6.3.1 Study area and data

Fourteen watersheds in Eastern Ontario with similar climatic and geomorphologic conditions were analyzed. The drainage area of the watersheds varies between 75 and 3800 km^2. The study period covers 11 years from 1990 to 2000. Extensive meteorological data were used to set up the model. Hourly data included precipitation, potential evaporation, air temperature, wind speed, solar radiation, and dewpoint temperature. Landuse data were taken from Landsat imagery with a resolution of 30 m. Five landuse categories were identified: agriculture, forest, pasture, bareland, and urban territories. GIS topographical maps, DEM with 25 m resolution, and hydrography data were used. Flow data were obtained from Environment Canada.

6.3.2 Watershed model and optimization techniques

HSPF, a widely-used watershed model, was used in the current research. HSPF is a part of the software system BASINS, which is developed and supported by the USEPA. HSPF is a comprehensive continuous simulation watershed model that simulates flows, point and non-point pollution, and performs routing in stream reaches. Detailed description of the model is available at *Bicknell et al. (2001)*.

Two nonlinear methods were used for model calibration: Gauss-Marquardt-Levenberg (GML) method and Shuffled Complex Evolution (SCE-UA) method developed by *Duan et al. (1992)*.

GML is implemented in the Parameter Estimation Tool (PEST) that was used for calibration and uncertainty analysis *(Doherty, 2004)*. SCE-UA was employed to ensure that the global minimum of the objective function was found.

6.3.3 Model calibration and objective function

Hydrological calibration consists of adjusting hydrological and snow parameters and obtaining the best fit of the simulated and observed daily flows, monthly volumes, and snow depths in the watershed. Based on the literature review and sensitivity analysis, 10 hydrological and 2 snow parameters were used for calibration. The physically possible ranges for selected parameters were taken from *USEPA (2000)*. Model performance comparison was based on the following statistical measures: sum of squared errors (SSE), coefficient of determination (r^2), Nash and Sutcliffe (NS) model fit, index of agreement (IA) proposed by *Willmott (1981)*, and daily root mean square error (RMSE).

Model calibration is an iterative procedure of parameter evaluation and adjustment. The essence of calibration is to minimize differences between observed and simulated values. The sum of squared errors that is minimized during the calibration is called an objective function (OF). It is assumed that the optimal parameter set corresponds to the minimum value of the OF. In the current research OF (ϕ) was expressed as a sum of log transformed daily flows.

$$\Phi = \sum_{i=1}^{N} \left(\log Q^i - \log q^i \right)^2 \tag{6.1}$$

where Q – observed and q – simulated daily flows.

Log transformation results in more equal contributions of high and low flows into the OF. In a non-log OF, high values dominate and the fit of low flows is poor. In order to avoid numerical instability in small watersheds where flow at summer can reach zero, the OF was expressed in a slightly different way:

$$\Phi = \sum_{i=1}^{n} [\log(Q^i + 1) - \log(q^i + 1)]^2 \tag{6.2}$$

6.3.4 Sensitivity and uncertainty analysis

In a broad sense sensitivity refers to the rate of change in one factor with respect to change in another; basically, sensitivity is a ratio of changes. In the current research the finite difference or perturbation sensitivities were obtained by perturbing each parameter at ±5, ±25 and ±75% from

the calibrated value. The combination of small and large perturbations allows identification of non-linear behavior of a parameter. Normalized sensitivity values (S_{uv}) for output Y_v to input X_u are computed according to *Brown and Barwell (1987)*:

$$S_{uv} = \left(\Delta Y_v / Y_v\right) / \left(\Delta X_u / X_u\right) \tag{6.3}$$

where X_u is calibrated value of input parameter, ΔX_u is magnitude of input perturbation, Y_v is calibrated value for output variable, and ΔY_v is sensitivity of output variable.

The normalized sensitivity represents the percentage change in the output variable resulting from a unit percentage change in input parameter. This approach helps to detect a non-linear dependence when parameters of different orders of magnitude are to be compared. Also, normalized sensitivities allow quick assessment of which parameters are important for simulations and ranking them accordingly.

The first-order variance estimation method (FOVE), which approximates the function by a flat plane tangent to the curved surface of a function at the mean value, was used for uncertainty analysis. In a case with two variables, the variance of model outcome z was approximated by the tangent planes with slope $\partial z/\partial x$ in the x dimension and $\partial z/\partial y$ in the y dimension.

$$Var(z) \approx \left(dz/dx\right)^2 Var(x) + \left(dz/dy\right)^2 Var(y) + 2\ \left(\partial z/\partial x\right)\ \left(\partial z/\partial y\right)\ Cov(x,y) \tag{6.4}$$

where x and y are model parameters and $Cov(x,y)$ is the covariance of x and y. If x and y are uncorrelated (as assumed in the current research), the third term disappears and Eq. 6.4 becomes the Gaussian error propagation formula *(Kirchner, 2001)*. The FOVE is a straightforward method which does not require knowledge of the probability distribution function of the input parameters. The only requirement is function differentiability.

6.3.5 Analysis of scaling

The watersheds can exhibit simple or multiple scaling. The fundamentals of scaling in watersheds are discussed by *Smith (1992), Yue and Gan (2004), and Gupta et al. (1994)*. The simple scaling between two watersheds with areas A_i and A_j and flow series Q_i and Q_j can be represented *(Smith, 1992)* by

$$Q_i^k \overset{d}{=} Q_j^k \left(A_i / A_j\right)^{k\theta} \tag{6.5}$$

The d above the equals sign denotes that the probability distributions of the random variables on both sides are the same. θ is a scaling constant, which varies for different climatic regions. k is an order of product of moments. Taking the natural logarithm of Eq. 6.5,

$$\ln Q_i^k = \ln Q_j^k + k\theta \ln\left(A_i / A_j\right) \tag{6.6}$$

Assuming that the reference watershed has area of 1, the equation becomes

$$\ln Q_i^k = a_k + b_k \ln A_i \tag{6.7}$$

$$\text{where } a_k = \ln Q_j^k \text{ and } b_k = k\theta \tag{6.8}$$

In a simple scaling model, as seen from Eq. 6.8, there is a linear relationship between the logarithm of drainage area and logarithm of the k-order moment of flood flows. The slope is a linear function of the moment order, k.

The watershed has multi-scaling properties when Eq. 6.8 does not hold, i.e., the slope varies with the return period. In terms of the physical processes in the watershed, simple scaling reflects additive mechanisms, while multi-scaling is tied to multiplicative generating mechanisms *(Gupta and Waymire, 1990)*. Results obtained by *Gupta and Dawdy (1995)* suggest that simple scaling is typical for watersheds where the floods are generated mostly by snowmelt, whereas multi-scaling is present in watersheds with rainfall generated floods.

Besides the product of moment method described above, other methods can be used for identifying the behavior of the processes in a watershed. *Kumar and Foufoula-Georgiou (1993)* developed the probability-weighted moment (PWM) method which exploits probability of flows instead of flows themselves. PWM is more robust and less affected by outliers; however, it requires more data.

Regression analysis was conducted for 14 watersheds. The method of ordinary least squares was applied to estimate parameters a_k and b_k, and θ in Eqs. 6.7 and 6.8. The two tailed t-test with a significance level of 0.05 was used to assess the significance of the linear regression. The residuals of the regression were checked to determine if they were normally distributed.

6.3.6 DEM resampling methods

DEM with a cell size of 25 m and vertical resolution of 0.5 m was used. One of the research objectives was to test the impact of DEM resolution on HSPF parameters and flow simulations.

A DEM is a continuous raster data set. The resampling method for producing different DEM resolutions can have significant impact on watershed characteristics. Three methods are the most common in GIS: nearest, bilinear, and cubic resampling. Each resampling scheme has distinct properties and affects further watershed delineation.

In the nearest resampling scheme each output cell value is the unchanged value from the closest input cell. It is the fastest and the least computationally intensive interpolation method. However, a blocky appearance can result from the duplication (smaller output cell size) or dropping (large cell size) of input cell values *(Chan et al., 2005)*. Bilinear interpolation computes the output cell value based on the weighted average of four closest input cell values. This method produces a smoother appearance than the nearest neighbour approach and it is well suited for continuous data. Cubic resampling determines the new cell value based on fitting a smooth curve through the nearest 16 input cell centers. The output value is a distance-weighted average where the weight varies as a nonlinear function of distance *(Chan et al., 2005)*. It is appropriate for continuous data, although it may result in the output raster containing values outside the range of the input raster. It is the most computationally intensive resampling method.

After comparative analysis of the resampling techniques, the bilinear method was chosen for application in the current study due its smooth and realistic representation of the DEM.

6.4 Results

6.4.1 Spatial scaling

Watershed models are often calibrated and validated where observed data are available; however, they are lacking in ungaged watersheds and calibration is impossible. The understanding of spatial scaling properties of the studied characteristic and model parameters is needed if such models are to be useful in ungaged areas.

Spatial scaling in watershed modeling is a broad topic. Three problems are addressed in this study. First of all, the existence of the scaling laws in all 14 tested watersheds was investigated. These laws can provide indicative predictions in ungaged watersheds where model calibration is not possible. Secondly, the HSPF parameters were studied as to whether they exhibit scale dependency. An attempt to explore and possibly derive the probability distribution of different model parameters versus watershed sizes was made. Thirdly, the simulation of daily maximum and minimum flows using a simple scaling function and HSPF parameter transfer from the neighbouring gaged watersheds was compared and discussed.

6.4.1.1 Test of scaling laws

The simple scaling similarities in the tested watersheds can be expected based on their climatic resemblance and snowmelt dominated floods *(Gupta and Dawdy, 1995)*. Moreover, all of the studied watersheds are located close to each other and have similar topography, soil, and landuse practices. Average annual runoff and watershed area were plotted and the linear regression parameters a_k and b_k, the coefficient of determination (r^2), and the t - statistics for the first 3 orders of the product moment were estimated (Table 6.1). It can be seen that r^2 is high and $t_s >> t_c$ at a significance level of 0.05. The regression lines are shown graphically in Figure 6.2. Both graphical and statistical tests provide grounds to conclude that a linear relationship between watershed area and average annual runoff exists. Thus, most of the variability in the annual runoff can be described by the scale of the watershed alone. *Yue and Gan (2004)* investigated the scaling properties of Canadian annual average runoff and watershed area and their regression coefficients for the Great Lakes region are very similar to those obtained in this study. The intercept a_k and the moment order k were tested and it appears that Eq. 6.8 holds indicating that the simple scaling law holds for the studied watersheds.

Table 6.1 Regression coefficients and statistics for annual runoff for 14 watersheds in Eastern Ontario

Moment order, k	a	b	r^2	t statistic	t critical
1	–7.62	0.98	0.99	287.49	2.16
2	–15.25	1.96	0.99	29.37	2.16
3	–22.87	2.93	0.99	17.83	2.16

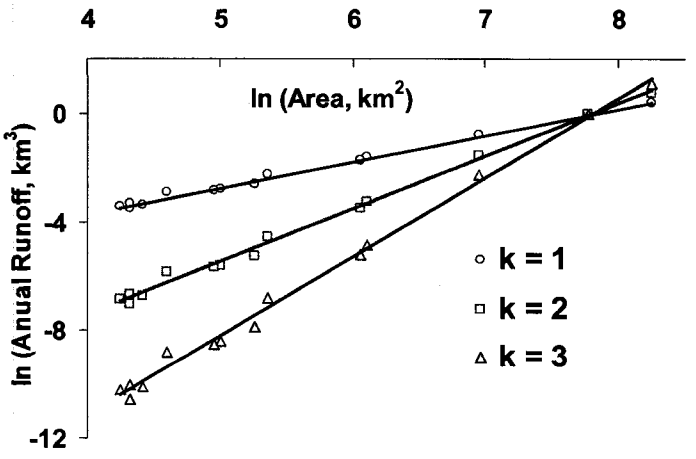


Figure 6.2 Log relationship between the expectations for the first three orders of product of moments and their watershed area

The river length and peak daily flows versus the drainage area were explored and plotted for the 14 watersheds (Figures 6.3-6.4). The same approach as for the annual runoff was used and the simple scaling law holds. It is worth noting that when larger watersheds demonstrate linear scaling, the smaller watersheds exhibit sizable variability, which is the reason that uncertainty of flow simulations in small watersheds using the scaling principles can be expected to be large.

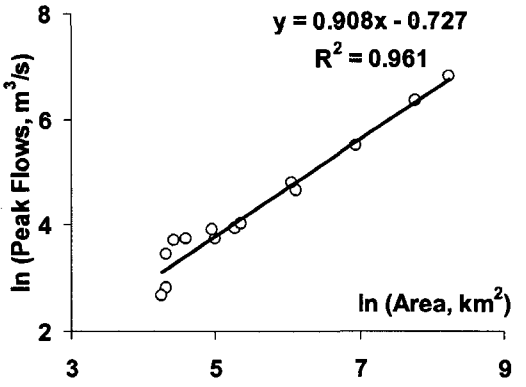


Figure 6.3 Log relationship between watershed area and peak flow

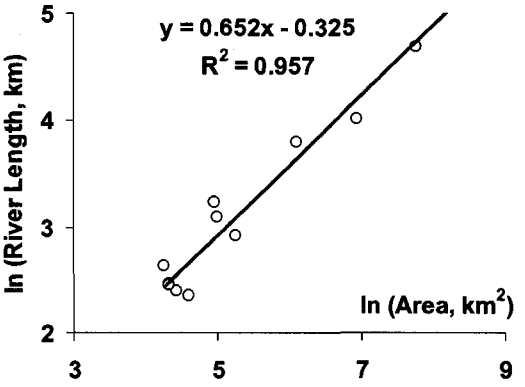


Figure 6.4 Log relationship between watershed area and river length

Simple scaling principles may be used to predict daily flows in neighbouring ungaged watersheds. Flows at different times of the year are influenced by different factors. This suggested examining scaling relations independently in spring, summer-autumn, and winter seasons. Spring is categorized as March-May, summer-autumn is June-October, and the winter season is November-February. It can be seen (Figure 6.5) that the best regression with r^2 =0.93 is in spring, when the high flows are typical. The weakest correlation between flows in

neighbouring watersheds is in the summer-autumn period when highly variable low flows prevail. The regression was plotted for 9 years of data (1990-1998).

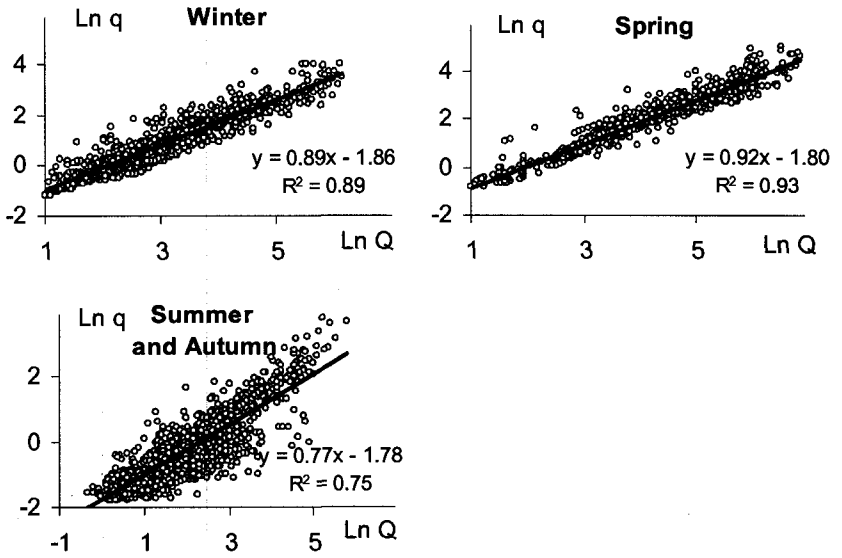


Figure 6.5 Log relationship of daily flows at SNR Plantagenet (Q) and Castor River at Russell (q) at various seasons. Their watershed areas are 3780 and 428 km^2 respectively

Scatter plots of spring flows for different watersheds are presented in Figure 6.6. The least squares principle was used to build the smoothed line. It can be seen that the linear regression can explain the variability of flows at the neighbouring watersheds. The slopes of the regression lines are very similar for all tested watersheds. The smallest watersheds exhibit the weakest scaling relationship.

Based on the analysis, it can be stated that, in absence of more sophisticated techniques, the simple scaling laws may be applied in the studied watersheds. Having two watersheds located close by with the same climatic and geomorphologic conditions it is possible to predict daily flows, especially in spring time, in an ungaged watershed based on flows in the gaged watershed. However, uncertainty of these predictions can be very high.

6.4.1.2 Scale dependency of HSPF parameters

Two approaches were used to test the scale dependency of HSPF hydrological parameters. The first approach includes calibration of various sized watersheds with similar hydrology and meteorology. Seven gaged watersheds with areas from 75 to 3780 km^2 and 3 artificial watersheds with areas of 25, 50, and 10,000 km^2 were tested (Table 6.2). The artificial watersheds were

calibrated against flows obtained from the simple scaling principles described above. The SCE-UA method was used for calibration.

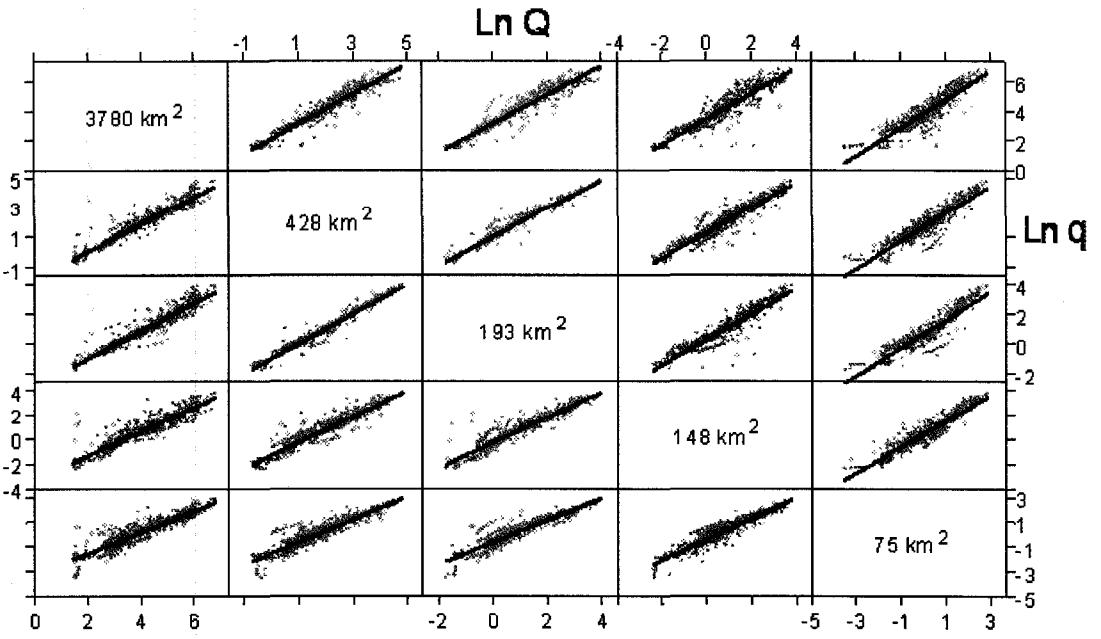


Figure 6.6 Scatter plot matrix of spring flows for various-size watersheds

Table 6.2 Calibrated parameter values for different size watersheds

HSPF parameters	Area, km^2									
	25	50	75	148	193	212	428	449	3,780	10,000
SNOWCF	1.015	1.401	1.938	1.304	1.208	2.078	1.251	1.292	1.281	1.262
CCFACT	1.038	1.321	1.397	0.772	0.806	1.683	0.963	0.646	0.953	0.904
LZSN	8.466	8.250	5.142	3.001	3.275	2.872	3.548	2.520	2.290	2.008
INFILT	0.200	0.200	0.200	0.041	0.046	0.190	0.045	0.032	0.035	0.054
KVARY	3.497	2.303	1.381	1.678	0.923	1.018	1.082	1.746	1.404	1.090
DEEPFR	0.500	0.500	0.500	0.500	0.499	0.459	0.374	0.478	0.500	0.195
BASETP	0.074	0.037	0.055	0.038	0.027	0.069	0.066	0.029	0.200	0.044
AGWETP	0.200	0.200	0.200	0.056	1.515	0.200	0.042	0.060	0.034	0.040
CEPSC	0.010	0.400	0.400	0.400	0.400	0.369	0.348	0.282	0.017	0.336
UZSN	1.134	1.502	1.995	1.946	1.999	1.818	1.978	1.865	1.987	1.999
IRC	0.620	0.649	0.548	0.612	0.712	0.711	0.716	0.726	0.753	0.753
LZETP	0.900	0.861	0.446	0.333	0.410	0.189	0.344	0.258	0.339	0.157

The second approach was to take the same watershed and reduce and increase its area by 5, 25, and 75%. Calibration was done using SCE-UA against the flows obtained according to the simple scaling (Table 6.3).

From both approaches it can be concluded that there is no evident and clear scale dependency in HSPF hydrological parameters. As well, uncertainty and sensitivity of the parameters do not show much scale dependency.

The lack of scale dependency can be explained by the nature of the HSPF parameters. They are not point scale values. Mostly they have been developed as spatially mean values. For example, spatially averaged infiltration and interflow capacity over the land segment is calculated according to the hyperbola-like soil moisture ratio in the lower zone. Then, infiltration capacity is assumed to vary linearly based on a probability density function, which varies in time to represent changes in soil moisture in the unsaturated zone *(Bicknell et al., 2001)*.

The proposed scheme was suggested in order to take into consideration some parts of the subwatershed that have very low infiltration capacity and some that have very high infiltration. The total volume of infiltrated water from the land segment in a specified time step is basically a function of current soil-moisture ratio in the lower zone and the water supply rate. More details about HSPF treatment of spatial variability are given in *Doherty (2004)* and *Bicknell et al. (2001)*.

6.4.1.3 Flow simulations for ungaged waters hed

Flows at an ungaged watershed can be estimated from the neighbouring gaged watershed assuming both have similar meteorological and geomorphologic conditions. The gaged watershed of the Castor River at Russell (area is 428 km^2) was used to predict daily flows of the South Nation River (SNR) at Plantagenet (area is 3780 km^2). Two approaches were used. The first was simple scaling. Eqs. 6.7-6.8 were used to develop regression equations for each season and then to simulate flows for the SNR at Plantagenet. The second approach was to calibrate the gaged watershed using HSPF and to transfer the optimal set of hydrological parameters into the ungaged watershed. Now, HSPF can de used for simulation daily flows in ungaged watershed assuming that detailed DEM data exist. These two approaches are shown schematically on Figures 6.7 and 6.8.

Table 6.3 Calibrated parameter values and their 95% confidence limits (Castor River at Russell, 428 km^2)

	−75%	−25%	−5%	0%	5%	25%	75%
SNOWCF	1.209 ± 0.065	1.227 ± 0.015	1.229 ± 0.015	1.228 ± 0.015	1.222 ± 0.037	1.228 ± 0.015	1.227 ± 0.015
CCFACT	0.881 ± 0.057	0.916 ± 0.045	0.918 ± 0.046	0.917 ± 0.045	0.923 ± 0.051	0.920 ± 0.046	0.918 ± 0.046
LZSN	2.839 ± 0.467	2.758 ± 0.640	2.728 ± 0.648	2.766 ± 0.643	2.878 ± 0.670	2.775 ± 0.638	2.768 ± 0.645
INFILT	0.051 ± 0.005	0.051 ± 0.008	0.048 ± 0.008	0.051 ± 0.008	0.051 ± 0.008	0.052 ± 0.008	0.053 ± 0.008
KVARY	0.503 ± 0.143	0.902 ± 0.162	0.958 ± 0.167	0.949 ± 0.165	0.930 ± 0.173	0.975 ± 0.167	1.007 ± 0.168
DEEPFR	0.352 ± 0.043	0.354 ± 0.038	0.325 ± 0.038	0.358 ± 0.038	0.366 ± 0.040	0.356 ± 0.038	0.372 ± 0.038
BASETP	0.050 ± 0.013	0.067 ± 0.022	0.068 ± 0.021	0.067 ± 0.021	0.066 ± 0.021	0.071 ± 0.021	0.068 ± 0.021
AGWETP	0.038 ± 0.018	0.030 ± 0.013	0.037 ± 0.013	0.033 ± 0.013	0.036 ± 0.013	0.035 ± 0.012	0.036 ± 0.012
CEPSC	0.400 ± 0.071	0.400 ± 0.063	0.400 ± 0.064	0.400 ± 0.063	0.400 ± 0.064	0.400 ± 0.063	0.400 ± 0.064
UZSN	2.000 ± 0.210	2.000 ± 0.203	2.000 ± 0.205	2.000 ± 0.202	2.000 ± 0.205	2.000 ± 0.202	2.000 ± 0.201
IRC	0.747 ± 0.019	0.748 ± 0.018	0.750 ± 0.018	0.749 ± 0.018	0.754 ± 0.018	0.749 ± 0.018	0.748 ± 0.018
LZETP	0.334 ± 0.039	0.339 ± 0.038	0.329 ± 0.038	0.330 ± 0.038	0.322 ± 0.039	0.328 ± 0.038	0.321 ± 0.038

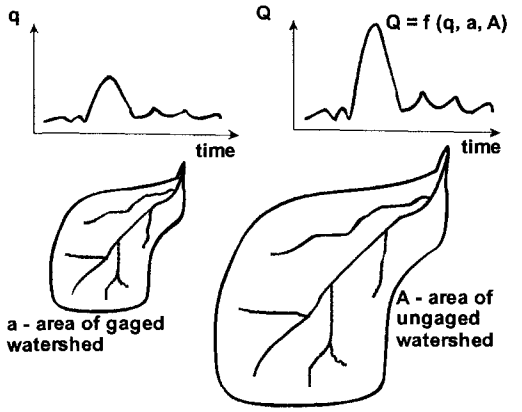


Figure 6.7 Simple scaling relationship between two watersheds

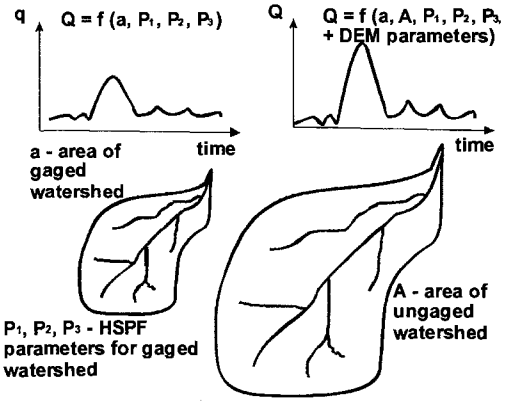


Figure 6.8 HSPF parameter transfer from a gaged to an ungaged watershed

For comparison, the observed flow data for SNR at Plantagenet were used to calibrate the model and obtain the true values of the hydrological parameters. Based on these parameters flows were simulated and statistics were calculated (Table 6.4). It can be seen that simple scaling results in a rough estimation, because the flow is a function of area only. The parameter transfer approach is more sophisticated since HSPF utilizes topography information, landuse, hydrography, meteorological data, etc. The performance statistics for parameter transfer and HSPF calibration with the measured data showed results that were as good as the calibration. It can be concluded that HSPF with parameter transfer from a calibrated neighboring watershed can be used successfully for daily flow simulations in ungaged watersheds with a high quality DEM. It important to note that the conducted analysis was performed on relatively flat subwatersheds in Eastern Ontario and the results may not be necessarily applicable to the watersheds with higher slopes.

6.4.2 DEM resolution

Resolution of DEM impacts the HSPF simulation in three different ways. First of all, area is computed during the watershed delineation based on DEM. As a coarser resolution of DEM is used, a larger error is introduced into the area measurement. Area is a key factor for computing the amount of precipitation on the watershed and as a result it impacts the entire model simulations. Secondly, DEM impacts calculation of channel characteristics such as stream depth, width, and length, as well as channel slopes. Stream geometry is obviously crucial for handling floods and describing the area-discharge relationships for each cross-section. Thirdly, DEM affects the overland plane features of the watershed. Slope and length of the overland plane determine the estimate of travel time of water to the stream or another drainage path. The individual contribution of each factor to the uncertainty and sensitivity of flow simulations was estimated in this study.

Table 6.4 Statistics for daily flows obtained by simple scaling, parameter transfer, and HSPF calibration for SNR at Plantagenet

		SSE	RMSE	IA	NS	r^2
All flows	Simple scaling	1.31×10^{11}	6321	–2.320	0.619	0.367
	Parameter transfer	8.90×10^{9}	1645	0.775	0.928	0.788
	HSPF real data calibration	8.58×10^{9}	1615	0.783	0.938	0.786
Maximum flows	Simple scaling	5.87×10^{9}	13339	0.497	–5.739	0.266
	Parameter transfer	2.26×10^{9}	8271	0.615	–1.591	0.479
	HSPF real data calibration	1.44×10^{9}	6608	0.738	–0.654	0.563

6.4.2.1 Area of watershed

Various resolutions of DEM impact the size of the watershed during the delineation process. Automatic delineation was performed using BASINS and ArcView capabilities with the option "cells on the edge flow towards the inner cell". The minimum drainage area (1,000 ha threshold area) required to form the beginning of a stream was used for the delineation. The mean errors of various DEMs were computed relative to 25 m DEM, the highest resolution available for the studied watersheds. As shown above, misrepresenting the watershed area will, of course, result in flow simulation errors.

Based on delineation of the 14 watersheds it was observed that smaller watersheds are more sensitive to the DEM resolution (Table 6.5). For example, 5% error was observed for small watersheds (<200 km^2) at 35 m DEM, for medium watersheds ($<1{,}000$ km^2) at 100 m DEM, and for large watersheds ($<4{,}000$ km^2) at 150 m DEM. It is worth noting that increasing DEM resolution can introduce positive as well as negative error. The value of the error progressively

rises with increasing DEM scale. Thus, for small watersheds mean error could be 10% at 150 m DEM and as high as 50% at 300 m DEM . DEM is a critical factor for correct area determination, especially for small watersheds.

Then the impact of watershed area underestimation or overestimation on the maximum and minimum daily flows was analyzed. The maximum flows were defined as the top 1% daily flows and the minimum flows were defined as the bottom 1% daily flows for all periods of research (1990-2000). A mid-size watershed (Castor River at Russell, drainage area is 428 km^2) was chosen to study the effect of area on daily flows (Figure 6.9). It is seen that underestimating area of the watershed by 10% may result in 8-12% error in maximum flows and 10-21% error in minimum flows. Overestimating the watershed area by 20% may result in overpredicting maximum flow by 14-25% and minimum flows by 22-52%. The minimum flow appeared to be more sensitive to changes in watershed area.

Table 6.5 Maximum percent mean error of watershed area estimated using different resolutions of DEM (based on 14 watersheds)

Area, km^2	Resolution of DEM, m								
	30×30	35×35	40×40	50×50	100×100	150×150	200×200	250×250	300×300
<200	2.0	4.6	6.7	5.6	6.2	9.9	27.5	35.7	51.4
200-1,000	1.7	1.5	5.0	2.8	5.5	5.1	6.8	9.5	10.7
1,000-4,000	3.1	3.5	3.1	2.9	4.4	5.2	5.5	7.7	8.3

It can be concluded that DEM impacts the area of watershed and as a result effects the flow simulations. Small watersheds are more sensitive to DEM resolution. Mean errors of minimum flows are typically higher than mean errors in maximum flows. Assuming that the error in maximum flows should not exceed 6% and error in minimum flows should be within 12%, it is critical to have 35 m DEM resolution for small watersheds (<200 km^2), 100 m DEM resolution for medium watersheds (200-1,000 km^2), and 150 m DEM resolution for relatively large watersheds (1,000-4,000 km^2).

6.4.2.2 Cross-section geometry

The optimal approach to obtain stream geometry characteristics such as stream depth, width, length, cross section area, and side slopes is from a detailed cross-section survey. However, in

most cases this is not possible and the DEM is used to extract these channel data. Obviously, resolution of DEM directly impacts precision of the values obtained.

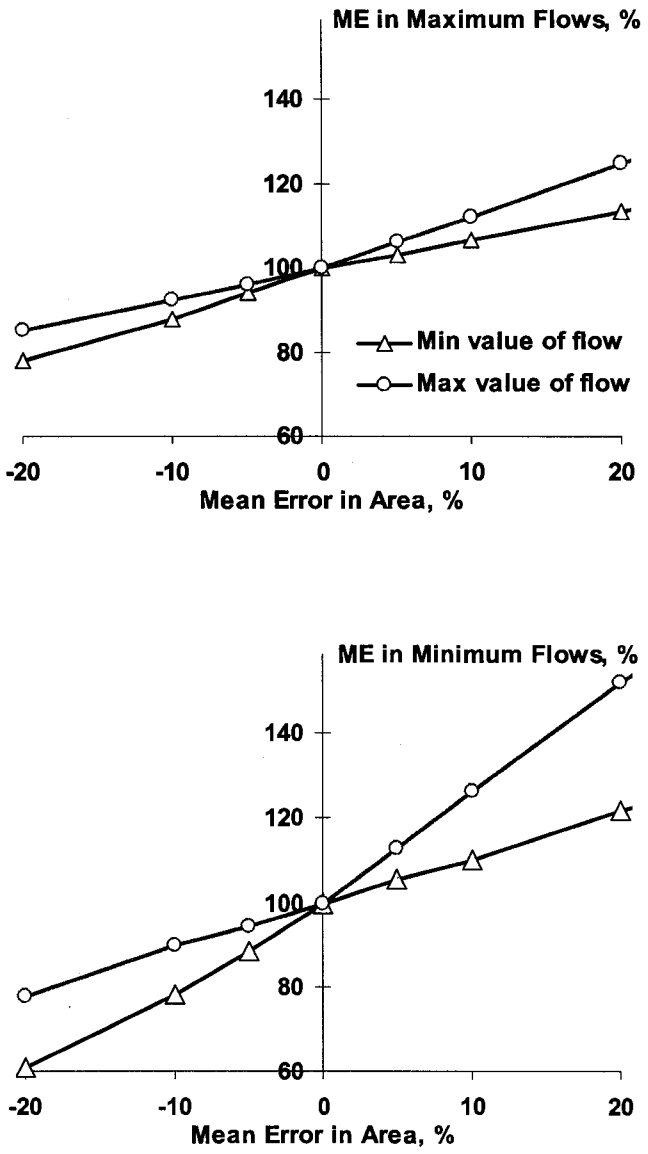


Figure 6.9 Impact of changes in watershed area on maximum and minimum daily flows in HSPF

During watershed delineation BASINS extracts available information from the DEM and creates channel geometry file with the extension *.ptf. This file contains length of the reach, stream mean depth and width, side slopes of the channel, channel depth, maximum depth, and other information. Various techniques are used for extracting this information. For example, mean

width (W) and depth (D) in BASINS are found from the regression equations $W=1.22 \cdot Q^{0.557}$ and $D = 0.34 \cdot Q^{0.341}$ *(Allen et al., 1994)*, where Q is a two year peak stream discharge estimated by a neural network based on drainage area and elevation as described in *Muttiah et al. (1997)*.

Some cross-section geometry information obtained at various DEM resolutions is presented in Table 6.6. It can be seen that error increases with increasing the cell size of DEM. The smaller watershed is more sensitive to changes in DEM.

Table 6.6 BASINS calculated mean depth and width of two cross-sections at different DEM resolutions

		DEM, m						
		25×25	30×30	40×40	50×50	100×100	200×200	500×500
SNR at Heckston 75 km^2	depth, m	0.72	0.74	0.74	0.74	0.74	0.81	0.83
	width, m	16.80	17.38	17.53	17.66	17.65	20.16	20.81
SNR at Plantagenet 3780 km^2	depth, m	3.50	3.50	3.50	3.51	3.53	3.54	3.85
	width, m	180.50	180.60	180.46	180.77	182.47	183.61	207.81

Based on the channel geometry file, HSPF generates a hydraulic function table (F-table) to represent the geometric and hydraulic properties of the stream reaches *(Bicknell et al., 2001)*. The assumptions for deriving the F-tables are listed at *USEPA (1999)*. The F-tables describe relationships between water depth, surface area, water volume, and discharge. The impact of F-tables on maximum flow simulations was investigated (Table 6.7). The percent mean error was calculated relative to 25 m DEM. It was observed that the range of flow variation is higher for small watershed and it tends to increase with the larger cell size. Even for 30 m DEM, error in flow simulations can be 11.5 and 4.1% for small and large watersheds, respectively. The mean values of maximum flows are quite similar for various resolutions with an average tendency to decrease with the coarser DEM. The minimum flows are impacted by the F-tables to a larger degree. Due to its small values the mean error reaches 150% at various DEM resolutions.

F-tables play a critical role in sediment transport simulation. F-tables are used to represent hydraulic radius, and subsequent shear values as a function of depth. Inaccurate F-tables can generate erroneous shear and scour conditions during flood events and impair the simulations.

6.4.2.3 Overland plane parameters

DEM impacts a number of overland plane parameters in HSPF which influence flow routing. These parameters describe the overland slope, minimum and maximum elevations of the watershed, channel slope and length, and elevation difference between start and end of the reach. An attempt was made to investigate how various DEM resolutions impact these parameters and the sensitivity of the simulated flows to these parameters.

Table 6.7 Percent mean error of daily maximum flows due to F-tables computed at different DEMs

	DEM, m					
	30×30	40×40	50×50	70×70	100×100	300×300
	SNR at Heckston, area 75 km^2					
Mean	−0.7	−1.2	−1.1	−1.1	−3.1	−2.5
Limits	−10.2 to 11.5	−15.6 to 11.6	−15.3 to 11.6	−13.7 to 11.9	−17.6 to 11.3	−15.9 to 11.8
	SNR at Plantagenet, area 3780 km^2					
Mean	−0.3	−0.1	−0.3	−0.8	−2.0	−4.2
Limits	−4.1 to 3.9	−2.0 to 5.8	−2.6 to 5.0	−4.0 to 5.8	−5.6 to 5.8	−10.6 to 9.2

Overland slope (SLSUR) is a slope of surface overland flow plane expressed in elevation change per distance, e.g., ft/ft or m/m. SLSUR affects total surface and subsurface runoff volumes and timing. Its value is estimated for each land use area either using GIS capabilities or detailed topographical maps.

Among the most common approaches for SLSUR estimation are the graphical techniques that impose a grid pattern on the watershed and calculate slope values for each grid point for each land use *(USEPA, 2000)*. *Johnson (2005)* has suggested finding the difference in maximum and minimum elevations and to divide it by the longest flow path for each subwatershed. Another useful method for determining the land surface slope involves tracing multiple flow paths from equally spaced starting points around the perimeter of the subwatershed and recording elevation change and flow distance to the point where flow enters the reach. A good measure of the land surface slope for a typical subwatershed area is derived as an average from about 10 to 20 flow paths *(SFWMD, 2000)*.

In the current research, internal BASINS capabilities were used to delineate the watershed and estimate the overland slope. Different resolutions of DEM led to different estimations of SLSUR. It was observed that increasing the DEM cell size leads to a decreasing SLSUR value (Table 6.8).

That is quite logical since a coarser DEM smoothes elevation differences. It is seen that the increasing DEM from 25 to 35 m can lead to error in SLSUR of 15-20%; at a DEM resolution of 300 m the error can reach 80%.

The sensitivity of the daily flows to the SLSUR values was then investigated. The calibrated value of 0.04 was increased and decreased by 5, 50, and 75% and the simulated flows were recorded (Table 6.9). It was found out that increasing SLSUR generally leads to increasing peak flows and decreasing minimum flows. This is logically expected. The larger variability of flow is observed for maximum flows, for example, decreasing SLSUR by 75% may reduce the peak flow by 21%.

Table 6.8 SLSUR values at different DEM resolutions

	Area, km^2	DEM, m					
		25×25	35×35	70×70	100×100	200×200	300×300
Scotch river at St. Isidore	31	0.021	0.018	0.014	0.011	0.008	0.006
Bear Brook at Bourget	440	0.020	0.016	0.011	0.009	0.006	0.005
SNR at Plantagenet	3810	0.018	0.015	0.010	0.008	0.006	0.005

Table 6.9 Percent mean error of daily flows due to SLSUR variability. SNR at Plantagenet

		Deviation of LSUR from the calibrated value, %					
		−75%	−50%	−5%	+5%	+50%	+75%
Maximum flows:	Mean	−2.5	−0.5	−0.1	0.1	0.3	0.8
	Limits	−21.1 to 2.8	−4.2 to 0.5	−0.7 to 0.1	−0.1 to 0.7	−0.3 to 3.1	−0.9 to 7.4
Minimum flows:	Mean	2.9	0.5	0.1	−0.1	−0.4	−0.9
	Limits	−4.8 to 6.7	−0.8 to 1.2	0.0 to 0.6	−0.5 to 0.2	−1.0 to 0.6	−2.2 to 1.5

Length of assumed overland flow plane (LSUR) approximates the average length of travel of water to a stream reach or any drainage path within the subwatershed. LSUR is a truly physically based parameter, which can be estimated from high resolution topographical maps, DEM, or a field survey. LSUR directly affects the timing of the overland flow. LSUR is often assumed to vary with slope such that flat slopes have larger LSUR values and vice-versa. Typical values for pervious areas range from 200 to 500 ft for slopes ranging from 15 to 1%, respectively *(USEPA, 2000)*. For impervious areas the usual range for LSUR is 50 – 150 ft.

Watts (2006) has suggested estimating the length of overland flow plane from the drainage density of the subwatersheds (Eq. 6.9). The drainage density (km/km^2) is defined as the sum of all drainage path lengths divided by the area.

$$LSUR = 1/(2 \times \text{Drainage Density}) \quad (6.9)$$

LSUR is also often estimated from topographic data by dividing the watershed area by twice the length of all streams, gullies, ditches, etc., that move the water to the stream *(USEPA, 2000)*. LSUR values derived from topographic data of insufficient resolution are not able to display all small streams and as a result, such values are often overestimated *(USEPA, 2000)*.

In the current research BASINS was used to derive different LSUR values from DEMs of different resolutions (Table 6.10). It appears that BASINS uses an increment of 50 ft for LSUR. It can be seen that the small watershed becomes sensitive to LSUR at 35 m DEM and larger watersheds become sensitive at 100 m DEM.

Table 6.10 LSUR values at different DEM resolutions

	Area, km^2	DEM, m					
		25×25	35×35	70×70	100×100	200×200	300×30
Scotch river at St. Isidor	31	300	350	350	350	400	400
Bear Brook at Bourget	440	350	350	350	400	400	500
SNR at Plantagenet	3810	350	350	350	400	400	500

Then, the impact of LSUR on daily flows was investigated (Table 6.11). The mean error was computed relative to calibrated value of LSUR, which is 350 ft for SNR at Plantagenet. It is seen that the higher LSUR values generate higher maximum and lower minimum flows. The maximum flows are more sensitive to changes in LSUR than minimum flows. For example, reducing LSUR by 5% results in average decrease of maximum flow by 0.4% and in increase of minimum flows by 0.2%. Limits in Table 6.11 are referred to upper and lower bounds of variation of the percent mean error. The limits are wider for maximum flows.

Table 6.11 Percent mean error of daily flows due to different LSUR values

		Deviation of LSUR from the calibrated value, %					
		−75%	−25%	−5%	+5%	+25%	+75%
Maximum flows:	Mean	−8.3	−1.9	−0.4	0.3	1.6	3.8
	Limits	−58.9 to −4.6	−14.7 to 1.2	−2.9 to 0.2	−0.2 to 2.7	−1.1 to 12.0	−2.9 to 27.9
Minimum flows:	Mean	4.6	1.3	0.2	−0.2	−1.2	−3.3
	Limits	−10.4 to 10.5	−2.2 to 2.8	−0.2 to 0.6	−0.7 to 0.5	−2.8 to 1.5	−7.1 to 3.7

Elevation difference (DELTH) is the drop in water elevation between the upstream and downstream sections of the reach. Impact of DEM resolution of 14 various watersheds on DELTH is shown in Table 6.12. It can be seen that small watersheds are more sensitive to DEM cell size. A 20% mean error of DELTH is observed for small, medium, and large watersheds at DEM 30, 35, and 150 m, respectively.

Table 6.12 Maximum absolute values of percent mean error of DELTH at different DEM resolutions

Watershed area, km^2		Resolution of DEM, m						
		30×30	35×35	40×40	50×50	100×100	200×200	300×300
<200	(4 watersheds)	19.2	22.9	23.1	26.1	26.9	65.4	69.2
200-1,000	(6 watersheds)	12.8	21.3	55.3	39.5	36.8	61.7	60.5
1,000-4,000	(4 watersheds)	10.3	11.1	4.9	13.9	12.1	33.3	45.9

It was found that DELTH does not impact flow simulations in HSPF. It is used only for dissolved oxygen calculations by the Tsivoglou-Wallace method *(Tsivoglou and Wallace, 1972)* and for sand transport by the Toffaleti and Colby method *(Bicknell et al., 2001)*.

Length of the stream reach (LEN) is used for calculations of F-tables, particularly for width and hydraulic radius. Once the F-tables are computed, parameter LEN does not impact flow simulations in HSPF. LEN is also used in the sediment subroutine for calculating amount of bed sediments and in dissolved oxygen simulation for computing stream reaeration.

The 14 different watersheds were used to estimate the impact of DEM on the calculated reach length (Table 6.13). The mean error was computed relative to 25 m DEM . It was observed that the coarser DEM resulted in higher error in the LEN determination. BASINS tends to underestimate LEN, which is especially evident at larger DEM cell size.

Table 6.13 Maximum absolute values of percent mean error of LEN at different DEM resolutions

Area, km^2		Resolution of DEM, m						
		30×30	35×35	40×40	50×50	100×100	200×200	300×300
<200	(4 watersheds)	8.4	7.7	7.2	9.2	16.4	14.6	27.1
200-1,000	(6 watersheds)	4.0	4.6	5.8	6.5	14.9	19.3	18.5
1,000-4,000	(4 watersheds)	0.8	2.1	2.1	2.2	6.9	11.4	19.6

6.4.2.4 Normalized sensitivities of topographic parameters

In order to study sensitivities of various topographic parameters to flow simulations the normalized sensitivities according to Eq. 6.3 were computed (Table 6.14). It is seen that the flows are the most sensitive to area, and the dependence is almost linear, especially in the case for maximum flows. Larger watershed area generates higher maximum peaks. The almost linear relationship is likely determined by the HSPF method of computing amount of available water. The available watershed water is almost a linear function of precipitation and the watershed area.

Table 6.14 Normalized sensitivities of topographic parameters. SNR at Plantagenet

A. Maximum flows

Parameter		Parameter deviation from the calibrated value, %					
		−75%	−25%	−5%	+5%	+25%	+75%
AREA:	Mean	1.000	1.002	1.006	0.999	1.003	1.008
	Limits	0.990 to 1.009	0.972 to 1.036	0.982 to 1.038	0.975 to 1.047	0.981 to 1.048	0.990 to 1.056
LSUR:	Mean	−0.111	−0.091	−0.048	−0.069	−0.056	−0.051
	Limits	−0.786 to 0.061	−0.647 to 0.054	−0.354 to 0.029	−0.542 to 0.045	−0.420 to 0.041	−0.373 to 0.039
SLSUR:	Mean	0.033	0.019	0.017	0.015	0.014	0.011
	Limits	−0.037 to 0.282	−0.019 to 0.168	−0.017 to 0.148	−0.015 to 0.14	−0.014 to 0.124	−0.011 to 0.098
DEPTH:	Mean	−0.001	0.005	0.006	0.004	−0.014	−0.008
	Limits	−0.032 to 0.011	0.006 to 0.062	0.003 to 0.078	0.004 to 0.069	−0.022 to 0.036	−0.010 to 0.028
WIDTH:	Mean	0.005	−0.008	−0.013	−0.011	−0.004	−0.008
	Limits	−0.064 to 0.133	−0.004 to 0.163	−0.012 to 0.239	−0.010 to 0.127	−0.004 to 0.037	−0.005 to 0.064
B. Minimum flows							
AREA:	Mean	1.040	1.148	1.105	1.253	1.217	1.248
	Limits	0.940 to 1.067	0.952 to 1.310	0.667 to 1.60	0.656 to 1.600	0.918 to 1.395	0.918 to 1.422
LSUR:	Mean	0.061	0.057	0.035	0.039	0.046	0.044
	Limits	−0.139 to 0.140	−0.107 to 0.126	−0.029 to 0.068	−0.097 to 0.140	−0.054 to 0.098	−0.049 to 0.095
SLSUR:	Mean	−0.039	−0.020	−0.015	−0.020	−0.015	−0.012
	Limits	−0.090 to 0.064	−0.048 to 0.046	−0.120 to 0.0	−0.101 to 0.042	−0.040 to 0.025	−0.030 to 0.019
DEPTH:	Mean	−0.020	−0.026	−0.019	−0.019	−0.819	0.402
	Limits	−0.196 to 0.657	−0.158 to 0.468	−0.154 to 0.427	−0.139 to 0.378	−2.603 to −0.443	0.060 to 0.850
WIDTH:	Mean	−0.160	−0.160	−0.158	−0.163	−0.159	0.160
	Limits	−0.255 to −0.008	−0.257 to −0.012	−0.263 to 0.0	−0.279 to 0.0	−0.257 to −0.012	0.008 to 0.257

The next sensitive parameter for the maximum flows is LSUR, which has a nonlinear behavior. It is most sensitive at its lower values and less sensitive at its high values, which means LSUR is an important parameter in hilly watersheds with short overland plains. SLSUR exhibits almost the same sensitivity pattern; however, its sensitivity is at least 3–4 times less. The least sensitive parameters for maximum flows are mean width and depth. The HSPF algorithm has little reliance on these parameters, possibly due to their imperfect approximation from the regression equations as described above. The mean depth and width are more sensitive for small watersheds (<200 km^2) than for relatively large watersheds (1,000 - 4,000 km^2).

A slightly different picture emerges for the minimum flows. The second most sensitive parameter is mean stream width. These are followed by LSUR, DEPTH, and SLSUR. If accurate simulation of the minimum flow is a concern, attention should be paid to precise estimation of WIDTH.

The impact of DEM on topographical parameters can be summarized as follows. The descending order of sensitivities to the DEM changes is SLSUR >LSUR >AREA >WIDTH >DEPTH. The sensitivity of topographical parameters to the maximum flows reduces as AREA >LSUR > SLSUR >WIDTH >DEPTH. The sensitivity of parameters to the minimum flows is AREA > WIDTH >LSUR >DEPTH >SLSUR.

6.4.3 Impact of watershed area on HSPF parameters and their uncertainty

Since watershed area is the most sensitive parameter for determining maximum and minimum flows, its impact on hydrological parameters, their sensitivities, and uncertainties was investigated. The areas of all landuse segments were decreased and increased by 5, 25, and 75%. All other parameters were unchanged. This situation was simulated when area of the watershed is over- or underestimated. Then, the model was calibrated using SCE-UA and GML methods and the calibrated parameter values, their uncertainties, and sensitivities were recorded (Table 6.15). The Castor River at Russell was used as a test watershed.

It can be seen that changing watershed area impacts the parameter values in different ways. The values of snow parameters (SNOWCF and CCFACT) and their uncertainties are reduced as the watershed area becomes more highly overestimated. This is logical since with a reduced snow gage correction factor (SNOWCF) the amount of fallen snow decreases. The rate of heat transfer (CCFACT) is also directly proportional to the loss of water due to condensation and convection. The same pattern is exhibited for AGWETP and IRC. Small values of AGWETP indicate that there is a small amount of water in active groundwater storage for potential evapotranspiration.

All available water either goes to overland flow or to the deep aquifer. IRC is the ratio of the one day's interflow discharge to the previous day's interflow discharge. As the watershed area increases, IRC values become smaller indicating that interflow does not hold the excess water; it quickly passes it to the overland flow. Interestingly, IRC uncertainty increases while its value decreases.

Table 6.15 Calibrated hydrological parameters and their 95% confidence intervals when watershed area is underestimated or overestimated

Parameter	AREA deviation from the real value, %					
	−75%	−25%	−5%	+5%	+25%	+75%
SNOWCF	3.000 ± 0.134	1.353 ± 0.073	1.243 ± 0.068	1.224 ± 0.034	1.059 ± 0.060	1.009 ± 0.010
CCFACT	2.000 ± 0.092	0.969 ± 0.063	0.918 ± 0.057	0.913 ± 0.016	0.886 ± 0.061	0.783 ± 0.036
LZSN	2.000 ± 0.807	2.502 ± 0.656	2.693 ± 0.654	2.835 ± 0.634	3.545 ± 0.776	5.432 ± 0.765
INFILT	0.033 ± 0.018	0.035 ± 0.006	0.049 ± 0.007	0.052 ± 0.008	0.056 ± 0.008	0.081 ± 0.011
KVARY	0.443 ± 0.095	0.649 ± 0.129	0.852 ± 0.163	1.010 ± 0.166	1.242 ± 0.009	2.709 ± 0.357
DEEPFR	0.001 ± 0.072	0.086 ± 0.061	0.321 ± 0.046	0.402 ± 0.039	0.496 ± 0.240	0.500 ± 0.035
BASETP	0.001 ± 0.012	0.088 ± 0.028	0.072 ± 0.020	0.063 ± 0.021	0.053 ± 0.034	0.200 ± 0.102
AGWETP	0.164 ± 0.034	0.046 ± 0.018	0.039 ± 0.014	0.030 ± 0.013	0.028 ± 0.019	0.007 ± 0.007
CEPSC	0.010 ± 0.016	0.335 ± 0.063	0.400 ± 0.065	0.400 ± 0.064	0.400 ± 0.064	0.400 ± 0.065
UZSN	1.565 ± 0.216	2.000 ± 0.198	2.000 ± 0.198	2.000 ± 0.191	2.000 ± 0.220	1.471 ± 0.207
IRC	0.838 ± 0.015	0.768 ± 0.017	0.754 ± 0.018	0.749 ± 0.018	0.751 ± 0.018	0.638 ± 0.030
LZETP	0.100 ± 0.039	0.311 ± 0.038	0.318 ± 0.039	0.332 ± 0.037	0.349 ± 0.040	0.504 ± 0.036

The opposite behaviour is exhibited by LZSN, INFILT, KVARY, DEEFR, CEPSC, and LZETP. Overestimation of the watershed area leads to increased values of these parameters. Higher values of LZSN increase lower zone soil moisture storage, making the "holding capacity" of the soil larger. By increasing LZSN the model tries to accommodate the excess precipitation (due to increased watershed area). It is quite logical that infiltration (INFILT) and water lost to the deep aquifer (DEEPFR) increase when the area is overestimated. Amount of rainfall retained by vegetation (CEPSC) reaches its maximum value of 0.4 and afterwards it does not change with increase of area. Uncertainties of LZSN and INFILT are minimal around the calibrated values and increase when area is underestimated or overestimated. Uncertainty does not change much for CEPCS, USZN, and LZETP.

6.5 Conclusions

Spatial scaling properties of annual runoff, river length, peak flows, and watershed area in 14 closely located watersheds of Eastern Ontario were investigated. Based on the product of moments, the logarithmic relationships between kth moment order and changes in scale were found to be linear. It suggests that simple scaling laws hold for the studied watersheds. It was

observed that, variability of annual runoff, river length, and peak flows can be explained by the watershed area alone.

The experiments conducted indicate that there is no clear scale dependency in HSPF hydrological parameters. Parameter uncertainty and sensitivity do not show much scale dependency either. This is supported by the fact that HSPF parameters were developed as watershed mean or spatially averaged values rather than point scale values.

Simulations of daily flows using simple scaling did not appear to be effective. The parameter transfer approach is more sophisticated since HSPF utilizes topography information, hydrography, meteorological data, etc. The performance statistics of parameter transfer and HSPF calibration with the real data showed close results. It can be concluded that HSPF with parameter transfer from a calibrated neighbouring watershed can be used for daily flow simulations in neighbouring ungaged watersheds assuming that detailed DEM data exist.

The impact of DEM resolution on HSPF topographical parameters (area, cross-section geometry, length and slope of overland plane, stream length, and elevation differences) was studied. It was found that small watersheds are more sensitive to DEM resolution. Area is the most sensitive parameter for simulation of minimum and maximum flows. Errors in minimum flows due to area calculations are typically higher than errors in maximum flows. Assuming that the error in maximum flows should not exceed 6% and error in minimum flows should be within 12%, it is critical to have DEM resolution of 35 m for small watersheds (<200 km^2), DEM resolution of 100 m for medium watersheds (200-1,000 km^2), and DEM resolution of 150 m for relatively large watersheds (1,000- 4,000 km^2).

DEM resolution impacts the cross section geometry (stream depth, width, length, cross section area, and side slope) of the reaches and as a result simulated flows. Small watersheds are more sensitive to changes in cross section geometry. Increasing DEM resolution from 25 to 30 m, the mean error in maximum flows can reach 11% for small (<200 km^2) and 4% for large (<4,000 km^2) watersheds. The minimum flows are impacted by the cross section geometry even to a larger degree. Due to small values of low flows, the mean error can reach 150% with various DEM low resolutions.

The impact of DEM on topographical parameters can be summarized as follows. The descending order of sensitivities to the DEM changes is SLSUR >LSUR >AREA >WIDTH >DEPTH. The sensitivity of topographical parameters on the maximum flows decreases in the following order:

AREA >LSUR >SLSUR >WIDTH >DEPTH. The sensitivity of parameters to the minimum flows is AREA >WIDTH >LSUR >DEPTH >SLSUR.

Impact of watershed area on HSPF parameters and their uncertainty was studied. Because area is proportional to flows it was observed that overestimating the area leads to increasing infiltration parameters, lower zone moisture storage and ET, and interception storage parameter. Oppositely, the snow and interflow recession parameters decrease when the area increases. Uncertainties of SNOWCF, CCFACT, DEEPFR, AGWETP decrease and uncertainties of KVARY, BASETP, CEPCS, IRC tend to increase when area of the watershed increases. Uncertainties of LZSN and INFILT are minimal around the calibrated values and increase when area is under- or overestimated.

6.6 References

Allen, P.M., Arnold, J.G., Byars, B.W., 1994. Downstream channel geometry for use in planning-level models. Water Resour. Bull. 30 (4), 663-671.

Amorocho, J., 1963. Measures of the linearity of hydrologic system. J. Geophys. Res. 68 (8), 2237-2249.

Bathurst, J.C., 1986. Physically-based distributed modeling of an upland catchment using the Systeme Hydrologique Europeen. J. of Hydrology 87, 79-102.

Bergstrom, S., Graham, L.P., 1998. On the scale problem in hydrological modeling. J. of Hydrology 211 (1-4), 253-265.

Beven, K.J., 1995. Chapter 18: TOPMODEL. Computer models of watershed hydrology, V.P. Singh, ed., Water Resources Publications, Littleton, Colo.

Bicknell, B.R., Imhoff, J.C., Kittle, J.L., Jobes, T.H., Donigian, A.S., 2001. Hydrological simulation program HSPF. Version 12. User's Manual. USEPA, Athens, GA.

Brown, L.C., Barnwell, T.O., 1987. The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS: documentation and user manual. Env. Res. Laboratory. USEPA 600/3-87/007, Athens, GA.

Burlando, P., Rosso, R., 1996. Scaling and multiscaling depth-duration-frequency curves of storm precipitation. J. of Hydrology 187 (1-2), 45-64.

Chan, W.C., Yu, K., Krull, U.J., Hornsey, R.I., Wilson, B.C., Weersink, R.A., 2005. Photonic applications in biosensing and imaging. SPIE Proceedings, 90-95.

Doherty, J., 2004. PEST – model-independent parameter estimation, User manual: 5th edition. Watermark Computing, Corinda, Australia.

Duan, Q.Y., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimisation for conceptual rainfall-runoff models. Water Resour. Res. 28, 1015-1031.

Gupta, V.K., Dawdy, D., 1995. Physical interpretation of regional variations in the scaling exponents in flood quintiles. Hydrological Processes 9 (3-4), 347-361.

Gupta, V.K., Waymire, E., 1990. Multiscaling properties of spatial rainfall and river flow distributions. J. Geophys. Res. 95, 1999-2009.

Gupta, V.K., Mesa, O.J., Dawdy, D., 1994. Multiscaling theory of floods: regional quantile analysis. Water Resour. Res. 30 (12), 3405-3421.

Heuvelink, G.B.M., 1998. Uncertainty analysis in environmental modelling under a change of spatial scale. Nutrient Cycling in Agroecosystems 50, 255-264.

Johnson, N.W., 2005. ArcGIS and HSPF model development. Center for Research in Water Resources, Austin, Texas.

Kirchner, J., 2001. Data analysis toolkit: uncertainty analysis and error propagation. Course notes, University of California, Berkeley.

Klarreich, E., 2005. Life on the scales. Science News 167 (7), 106-107.

Kumar, P., Foufoula-Georgiou, E., 1993. A multicomponent decomposition of spatial rainfall fields: 2. self-similarity in fluctuations. Water Resour. Res. 29 (8), 2533-2544.

Kuo, W.L., Steenhuis, T.S., McCulloch, C.E., Mohler, C.L., Weinstein, D.A., DeGloria, S.D., Swaney, D.P., 1999. effect of grid size on runoff and soil moisture for a variable-source-area hydrology model. Water Resour. Res. 35, 3419-3428.

Miller, E.E., 1980. Similitude and scaling of soil water phenomena. In applications of soil physics, ed. D. Hillel, Academic Press, New York, 300-318.

Muttiah, R.S., Srinivasan, R., Allen, P.M., 1997. Prediction of two-year peak stream discharges using neural networks. J. Am. Water Resour. Assoc. 33, 625-630.

Refsgaard, J.C., Butts, M.B., 1999. Determination of grid scale parameters in catchment modelling by upscaling of local scale parameters. International workshop on modelling of transport processes in soils at various scales in time and space, Leuven.

Savage, V.M., Gillooly, J.F., Woodruff, W.H., West, G.B., Allen, A.P., Enquist, B.J., Brown, J.H., 2004. The predominance of quarter-power scaling in biology. Functional Ecology 18, 257-282.

Smith, J. A., 1992. Representation of basin scale in flood peak distributions. Water Resour. Res. 28 (11), 2993-2999.

SFWMD, 2000. Estero Bay and watershed assessment. Volume E: Watershed monitoring program. http://www.sfwmd.gov/org/exo/ftmyers/report-text/vole/vole.html.

Tsivoglou, E.C., Wallace, J.R., 1972. Characterization of stream reaeration capacity. EPA R3-72-012.

USEPA, 1999. BASINS Technical Note 1. Creating hydraulic function tables (FTABLES) for reservoirs in BASINS. EPA-823-R-99-066. http://www.epa.gov/waterscience/BASINS/tecnote1.pdf.

USEPA, 2000. BASINS Technical Note 6. Estimating hydrology and hydraulic parameters for HSPF. EPA-823-R00-012. http://www.epa.gov/waterscience/basins/tecnote6.pdf.

Vazquez, R., Feyen, J., Feyen, L., Refsgaard, J., 2002. Effect of grid size on effective parameters and model performance of the MIKE-SHE code. Hydrological Processes 16, 355-372.

Vivoni, E.R., Ivanov, V.Y., Bras, R.L., Entekhabi, D., 2005. On the effects of triangulated terrain resolution on distributed hydrologic model response. Hydrological Processes 19, 2101-2122.

Watts, J.L., 2006. GeoSpatial HSPF model of the Sandies and Elm watershed, Texas. Report 05-14, Center for research in water resources, University of Texas, Austin http://www.crwr.utexas.edu/online.shtml.

West, G.B., Brown, J.H., Enquist, B.J., 1997. A general model for the origin of allometric scaling models in biology. Science 276, 122-126.

Willmott, C.J., 1981. On the validation of models. Phys. Geog. 2, 184-194.

Wolock, D.M., McCabe, G.J., 2000. Differences in topographic characteristics computed from 100 and 1000 meter resolution digital elevation model data. Hydrological Processes 14, 987-1002.

Yue, S., Gan, T.Y., 2004. Simple scaling properties of Canadian annual average streamflow. Advan. in Water Resour. 27 (5), 481-495.

CHAPTER 7

DEFINITION OF THE OBJECTIVE FUNCTION FOR MODEL CALIBRATION

The novel contribution of the presented paper is in studying various single and compound OFs in multi-objective optimization of HSPF. The optimal single and compound OFs were suggested. Recommendations on the OF weighting were made.

Model calibration is a critical component in model application which consists of four consecutive steps (Figure 7.1). Firstly, the parameters for calibration should be selected based on the results of sensitivity analysis, objectives of the study, and experience of the modeller. The number of parameters depends on the model, simulation objectives, and computer resources. Typically the number should not exceed 15 due to increased chance of parameter intercorrelation *(Gupta, 2003)*. Secondly, a GOM or LOM should be chosen for optimization of the model parameters. Numerous optimization methods, their advantages and disadvantages are discussed in Section 2.3. Thirdly, the OF must be specified. Fourthly, the goodness-of-fit criteria for model result comparison and evaluation are to be selected.

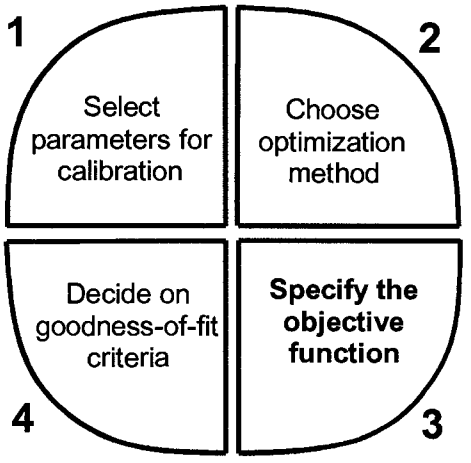


Figure 7.1 Steps in model calibration

This chapter deals with the third step – defining the OF for model calibration. Proper definition of the OF is a key element of any calibration. Even the best optimization technique with a poorly defined OF will not produce the expected fit of the modeled and observed data.

The OF is a function to be optimized during calibration and it constitutes the problem to be solved. The OF measures the misfit between observed and simulated data. The OF is most commonly a weighted sum of squared residuals in watershed models *(Madsen, 2000)*. Thus, the

objective of model calibration is to minimize the OF as low as possible using various LOMs and GOMs. Two approaches are commonly used: single objective and multi-objective optimization *(Schoups et.al., 2005; Dunn and Colohan, 1999; Doherty and Johnston, 2003)*. Both of them were tested in the current research.

Different definitions of the OF used in the current research are presented in Figure 7.2. A number of weighted and non-weighted OFs were tested and the simulation results were compared. A multi-objective optimization was obtained by combining the single OFs as well as using the Pareto solution.

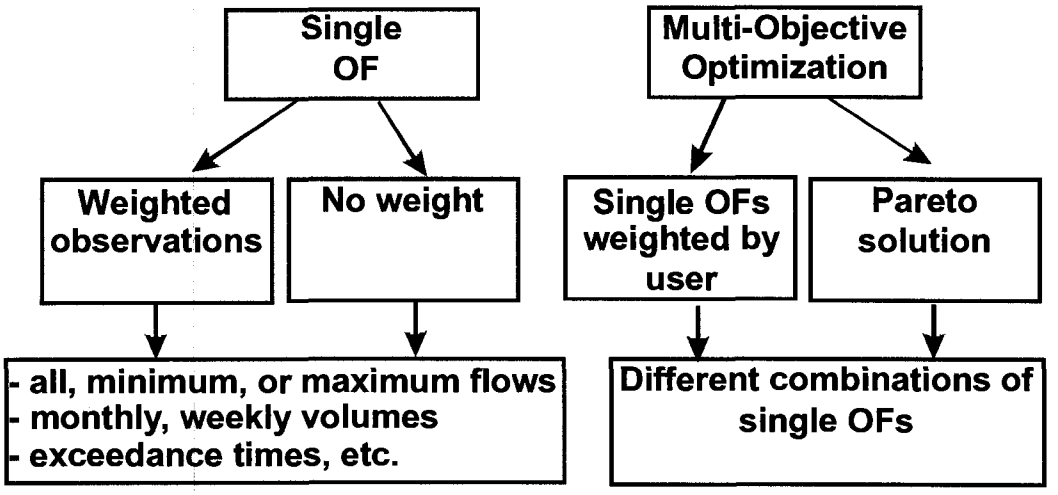


Figure 7.2 Definitions of the OF for watershed model calibration

7.1 Single OF

The most simplistic and therefore most frequently used approach in automatic model calibration is formulation of a single OF. In the majority of cases it is a simple sum of squared residuals between observed and simulated values. However, in hydrological modeling small and large values of flows can differ significantly and as a result the simple OF may produce biased simulation penalizing the minimum flows. The problem can be partially solved by taking log values of flow, assigning different weights to individual flows, or a combination of both.

Besides the squared residuals of point observations (daily, hourly flows) of other single OFs can be used: exceedance times (number of times the specific threshold is exceeded) or volume differences (weekly, 10-days, and monthly). Methods of calculation of these OFs are shown in Figure 7.3.

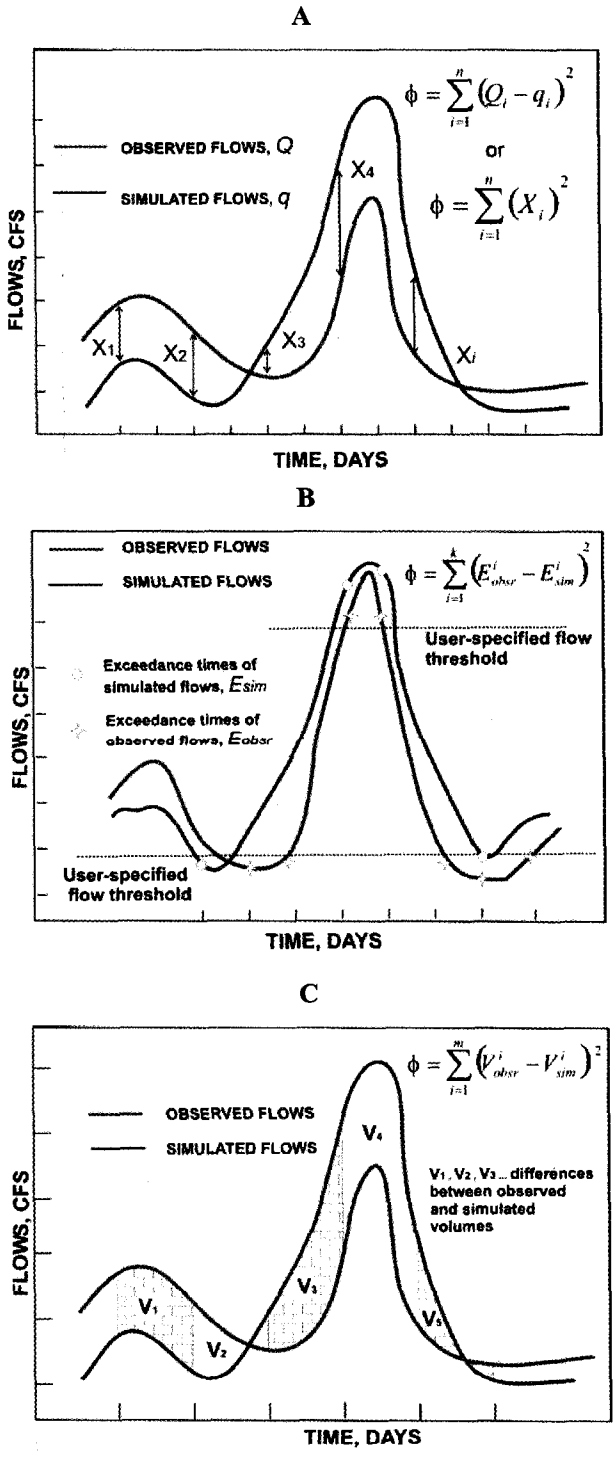


Figure 7.3 **OF expressed as flow differences (A), exceedance times (B), and volume differences (C)**

Data for the South Nation River at Plantagenet (1990-1998) were used for HSPF calibration throughout this chapter. Q and q are the values of daily observed and simulated flows, respectively; V_{obsr} and V_{sim} are the values of monthly observed and simulated volumes, respectively; E_{obsr} and E_{sim} are the observed and simulated exceedance times, respectively, for seven flow thresholds: 5; 10; 100; 1,000; 2,000; 10,000; and 20,000 ft^3/s. In this research more than 20 weighted and non-weighted single OFs were tested. Four examples of different definitions of the OF and resulting scatter plots and duration curves are presented in Figure 7.4. The performance statistics and calibrated parameter values for the same four cases are presented in Tables 7.1 and 7.2. More examples are shown in Appendix E.

Table 7.1 Performance statistics for four cases of single OF

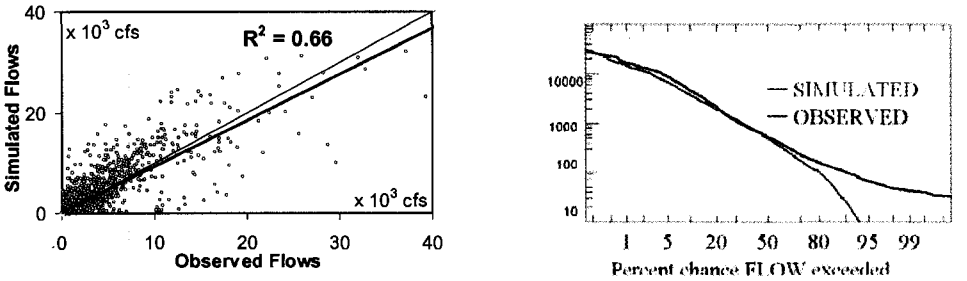
	Case 1	Case 2	Case 3	Case 4
Mean Error	−35.5	−99.7	−31.4	228.0
Root Mean Square Error	1496	1852	2057	3837
Nash and Sutcliffe Coefficient	0.813	0.713	0.646	0.232
Coefficient of Efficiency	0.631	0.626	0.580	0.446
Index of Agreement	0.808	0.803	0.780	0.734

It is clearly seen that various formulations of a single OF lead to different calibrations. The calibration which minimizes monthly volumes typically produces good overall water balance; however, minimum and especially maximum flows are often poorly predicted. The calibration which minimizes exceedance times does not change volumes of the peaks; however, it efficiently controls heights of the peaks. As well, it ensures that flows do not fall below certain limits.

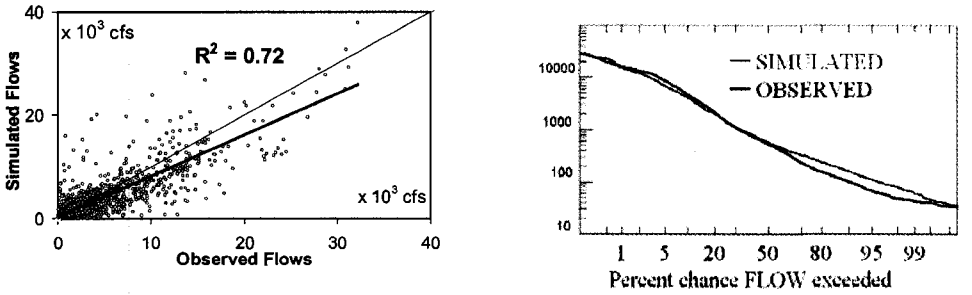
The SCE-UA method was used for model calibration. As such, the method produces not one, but several parameter sets with the same low value of the OF.

Based on statistics, the best calibration is obtained for Case 1. However, comparing the scatter plots and especially flow duration curves it is seen that low flows are poorly predicted and this definition of the OF probably is not the best choice. Case 2, when OF is expressed as the log of observed and simulated flows weighted as a log of observed flows, produces better graphics and may be chosen as a preferable way of defining the OF, although the statistics are not the best.

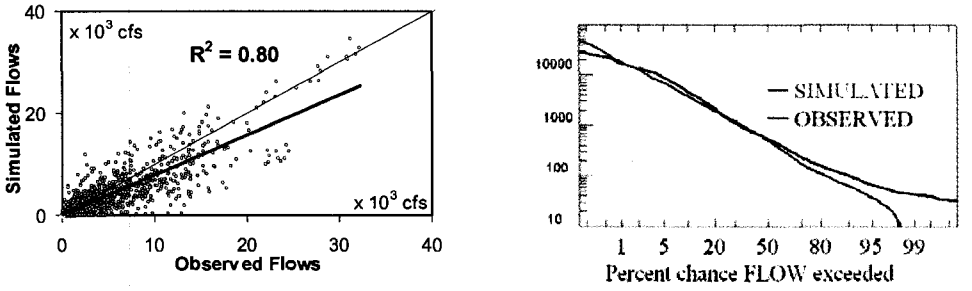
Case 1: $\phi = \sum_{i=1}^{n} (Q_i - q_i)^2$



Case 2: $\phi = \sum_{i=1}^{n} (\log Q_i \times (\log Q_i - \log q_i))^2$



Case 3: $\phi = \sum_{i=1}^{m} (V_{obsr}^i - V_{sim}^i)^2$



Case 4: $\phi = \sum_{i=1}^{k} (E_{obsr}^i - E_{sim}^i)^2$

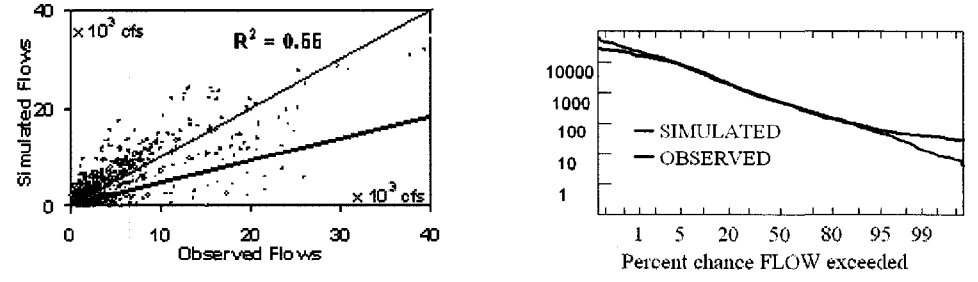


Figure 7.4 Definitions of the single OF and resulting scatter plots and flow duration curves

Table 7.2 Values of HSPF parameters for four cases of single OF

	Case 1	Case 2	Case 3	Case 4
SNOWCF	1.317	1.282	1.097	1.483
CCFACT	1.000	1.000	1.000	1.180
LZSN	3.598	2.000	2.852	3.960
INFILT	0.025	0.070	0.079	0.049
AGWRC	0.850	0.982	0.850	0.926
DEEPFR	0.076	0.500	0.500	0.206
BASETP	0.200	0.136	0.028	0.183
AGWETP	0.200	0.071	0.200	0.012
CEPSC	0.250	0.250	0.040	0.050
IRC	0.729	0.759	0.787	0.641
LZETP	0.623	0.106	0.461	0.469

In general single OFs tend to capture only one optimization criterion and may not be adequate to represent distinctions and similarities between the simulated and observed data. Another problem with single OFs is that modern distributed and semi-distributed models involve more than a single quality of measurement which necessitates several measurements of the OF *(Khu, 1998)*. This underlines the need to consider multiple criteria functions for model calibration.

7.2 Compound OF

Calibration based on a single performance measure is often insufficient for proper simulation of all important characteristics of the system. In a multi-objective context model calibration can be performed on the following basis.

- multi-variable measurements, i.e., groundwater level, river flow, snow depth, and other types of measurements,
- multi-site measurements, i.e., several groundwater level and flow measurement sites distributed within the watershed, and
- multi-response modes, i.e., the OFs that measure various responses of hydrological processes such as water balance, peak flows, low flows, etc.

The last type of multi-objective optimization was performed in this research. It was approached through a compound OF and the Pareto optimum. The compound OF is expressed as a combination of single OFs (*Schoups et al., 2005; Doherty and Johnston, 2003; Gupta et al., 2003).* Depending on the study objectives the single OFs can be equally weighted or non-equally weighted. In this research only equally weighted OFs were employed.

Although there has been notable progress in the past years regarding the development of multi-objective optimization techniques, limited experience exists on practical applications in hydrological models, and most of them are restricted to two-dimensional objective spaces dues to the complexity of three-dimensional spaces.

The intention of obtaining the best fit in maximum, minimum flows, volumes, and individual flows is virtually impossible to encapsulate in one OF. It is particularly important to have accuracy over the breadth of flows when water quality and associated ecological stress states are a concern. Besides the various combinations of single OFs discussed in Section 7.1, the compound OF was defined as a sum of three individual OFs. One OF was responsible for optimizing the high flows, the second OF optimized minimum flows, and the third optimized mid-flows.

The best fit of the observed and simulated flows was obtained when the maximum flows were expressed as a sum of squared residuals for the top 1% of flows. Minimum flows were expressed as a sum of squared residuals for the 20% lowest flows. The middle flows lay between the top and the bottom. The compound OF with encapsulated exceedance times and monthly volumes does not generally produce as good results.

More than 40 weighted compound OFs were tested. Some examples are presented in Figure 7.5. Q_{max} and q_{max} are the respective values of observed and simulated top 1% of daily flows. Q_{min} and q_{min} are the respective values of observed and simulated bottom 20% of daily flows, and Q_{mid} and q_{mid} are the respective values of observed and simulated flows between the maximum and minimum flows. More examples of tested compound OFs are presented in Appendix F. The selection of the best OF was based on statistical and graphical comparison of daily flows, monthly volumes, and minimum and maximum flows.

Weights in the compound OF were assigned to achieve approximately equal contributions from all three single OFs. For example, for Case 7 the assigned weight for OF1 was 0.1; the assigned weight for OF2 was $0.008/\sqrt{V_{obsr}}$. The exceedance times in days were calculated for seven flow thresholds: 5; 10; 100; 1,000; 2,000; 10,000; 20,000 ft^3/s. The weighting factor for OF3 was assigned a value of 15,000.

Performance statistics and calibrated parameter values for the three cases of compound OFs are presented in Tables 7.3 and 7.4. It is seen that in Case 6 scatter plot and duration curves show very good simulation results; however, the statistics are not the best.

Table 7.3 Performance statistics for various compound OFs

	Case 5	Case 6	Case 7
Mean	−286.4	−56.3	−107.0
Root Mean Square Error	1682	1803	1795
Nash and Sutcliffe Coefficient	0.763	0.595	0.730
Coefficient of Efficiency	0.622	0.607	0.628
Index of Agreement	0.806	0.801	0.808

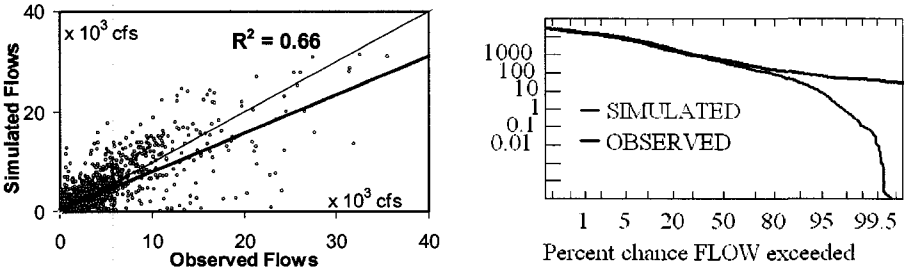
Table 7.4 Values of HSPF parameters for various compound OFs

	Case 5	Case 6	Case 7
SNOWCF	1.293	1.259	1.300
CCFACT	1.000	1.000	1.000
LZSN	2.000	2.000	2.000
INFILT	0.101	0.029	0.059
AGWRC	0.850	0.972	0.928
DEEPFR	0.500	0.500	0.431
BASETP	0.125	0.116	0.102
AGWETP	0.197	0.200	0.200
CEPSC	0.250	0.250	0.250
IRC	0.788	0.748	0.791
LZETP	0.708	0.283	0.323

An example of the shape of compound OF (Case 7) in the parameter space of INFILT and LZSN is shown in Figure 7.6. The plot is based on 12,000 runs. Locally weighted scatterplot smoothing (LOESS) with 3rd-order polynomial regression was used to build the error surface. It can be seen that the surface or search space is very complex with many local minima where optimization techniques can be easily trapped. The problem can be partially solved by reducing the parameter boundaries.

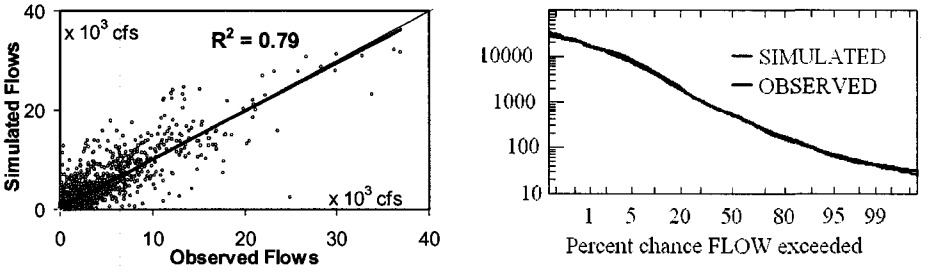
Case 5

$$\phi = \sum_{i=1}^{n}\left(w_1\left(Q_{mid}^i - q_{mid}^i\right)\right)^2 + \sum_{i=1}^{k}\left(w_2\left(Q_{max}^i - q_{max}^i\right)\right)^2 + \sum_{i=1}^{j}\left(w_3\left(Q_{min}^i - q_{min}^i\right)\right)^2$$



Case 6

$$\phi = \sum_{i=1}^{n}\left(w_1\left(\log Q_{mid}^i - \log q_{mid}^i\right)\right)^2 + \sum_{i=1}^{k}\left(w_2\left(\log Q_{max}^i - \log q_{max}^i\right)\right)^2 + \sum_{i=1}^{j}\left(w_3\left(\log Q_{min}^i - \log q_{min}^i\right)\right)^2$$



Case 7

$$\phi = \sum_{i=1}^{n}\left(w_1\left(\log Q^i - \log q^i\right)\right)^2 + \sum_{i=1}^{k}\left(w_2\left(V_{obsr}^i - V_{sim}^i\right)\right)^2 + \sum_{i=1}^{j}\left(w_3\left(E_{obsr}^i - E_{sim}^i\right)\right)^2$$

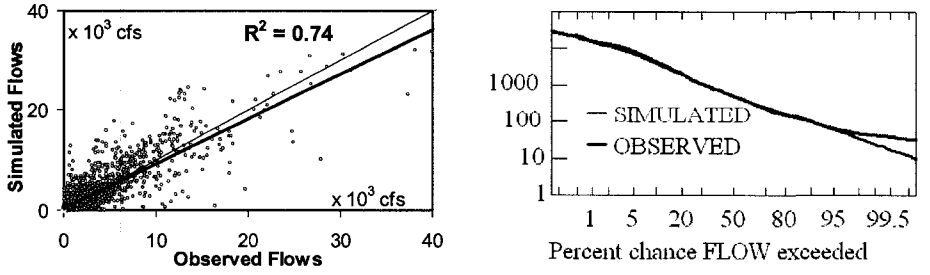


Figure 7.5 Different definitions of compound OF and resulting scatter plots and flow durations curves

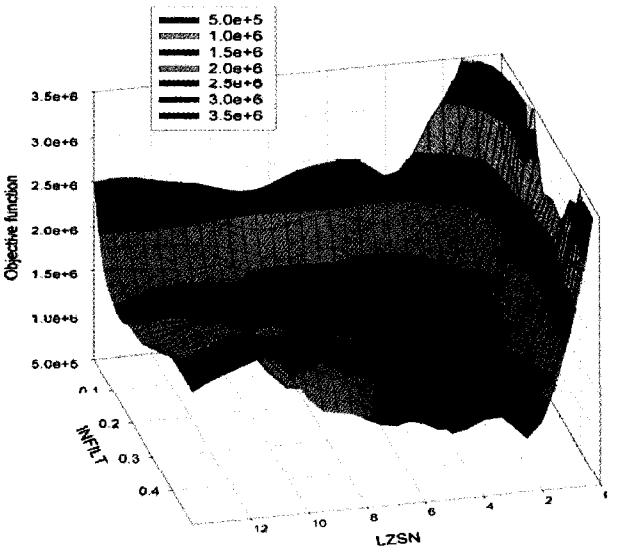


Figure 7.6 Surface of the compound OF in parameter space of LZSN and INFILT

7.3 Pareto optimality

Another multi-objective approach is to use the concept of Pareto optimality, where a set of Pareto optimal solutions are generated (*Zitzler and Thiele, 1999; Rozoset al., 2004; Schoups et al., 2005*). Several single OFs are selected and minimized individually over the feasible parameter space. Then the best solution based on various subjective techniques is chosen.

The solution, in general, is not a single unique set of parameters but a so-called Pareto set of solutions (non-dominated solutions), which is obtained according to various trade-offs between the different OFs. The trade-offs in a multi-objective optimization are governed by research objectives. The modeller should decide which objective must be sacrificed to more closely approach another objective.

In a simple case with two objective functions (OF1 and OF2) the thick line in Figure 7.7 identifies the Pareto front. All points on the Pareto front have the characteristic that no other points have both a smaller value of OF1 and a smaller value of OF2. Moving along the Pareto front from top to bottom, results in successively smaller values of OF2 at the expense of larger values of OF1.

When solving a multi-objective calibration problem, the problem can be transformed into a single-objective optimization problem by defining a scalar that aggregates the various OFs (e.g., the weighted average of individual OFs). Also, the solution can be found as the geometrically closest distance from the utopia point.

A proper scaling of single OFs can be defined as a transformation function of different objectives to a common scale. *Madsen (2000)* used a Euclidean distance function in which all the objective functions are transformed to have approximately the same distance to the origin near the optimum. *Van Griensven and Bauwens (2001)* adapted a probability distribution function for transformation of the OFs into a probability scale.

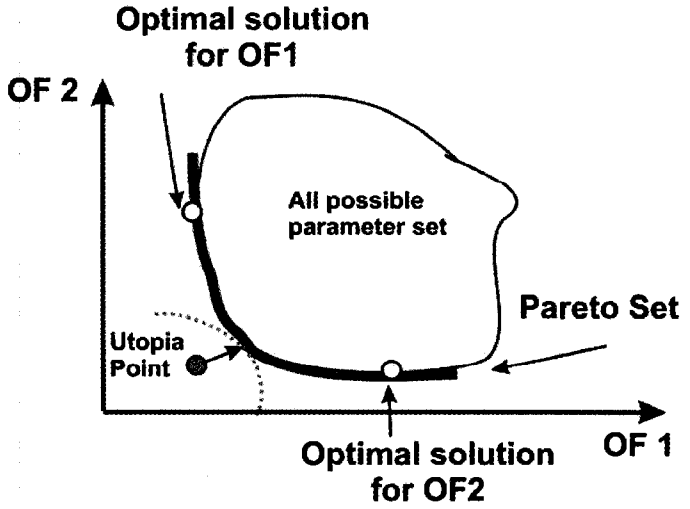


Figure 7.7 Pareto solutions for two in dependent OFs

Actual results of populating the Pareto solution are shown in Figure 7.8. Fifteen thousand optimization runs were performed using SCE-UA to generate parameter sets around the minimum values of OF1, OF2, and OF3. OF1 was formulated as a sum of squared residuals of 1% top flows, OF2 was formulated as a sum of squared residuals of 20% bottom flows, and OF3 accounted for middle flows. The thick line indicates the Pareto front. The best solution for the optimization problem is located somewhere on the Pareto front. The star on the graph indicates the same parameter set shown in different dimensions. It can be seen, for example, that the best solution for OF1 is far from the best solution for OF2.

Besides the technique of independent optimization of several OFs, there is another approach in the Pareto analysis. It consists of performing a number of optimization runs with different OF weights, and as a result of multiple simulations the Pareto front can be drawn. Often normalization or standardization helps to obtain equal representation of various single OFs in the compound OF.

The weighting of the individual OFs is a complex problem for all multi-objective problems. It is subjective and depends on research objectives, value and units of OFs, optimization algorithm,

and so forth. The modeller can specify the weights of individual OFs to reflect the relative priorities given to certain objectives. The problem is often even more complicated because individual OFs have their own weighted observations. Also, weights often reflect measurement errors, i.e., smaller weights are given to measurements with larger errors.

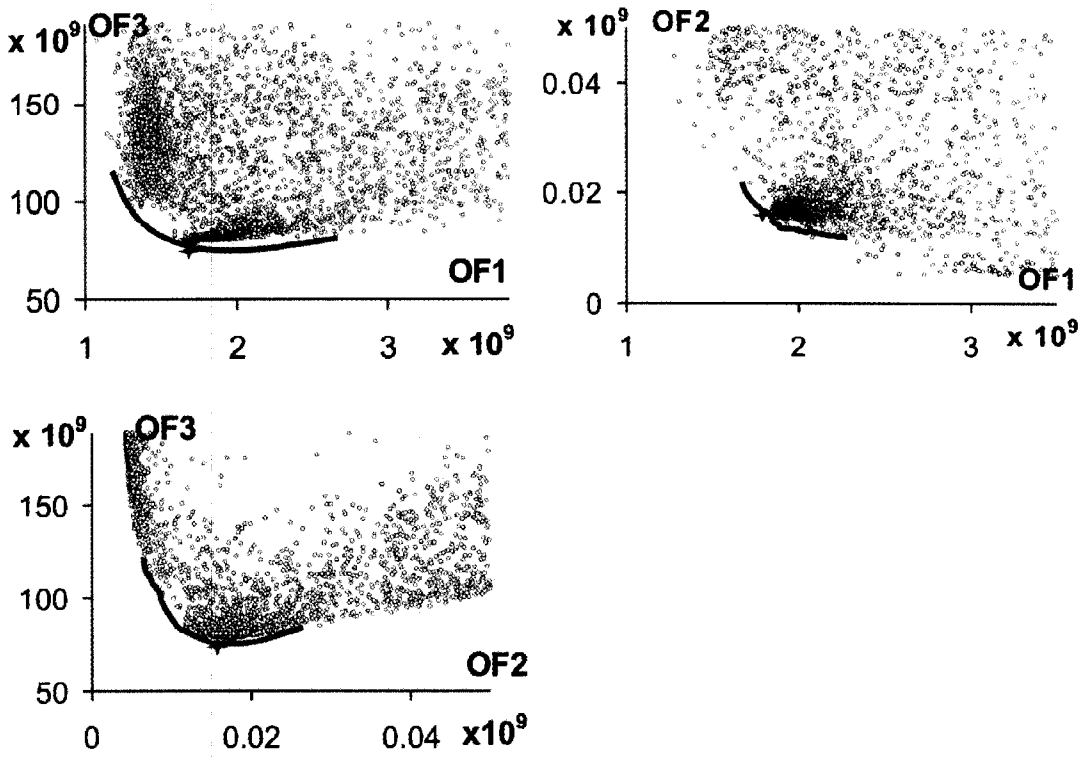


Figure 7.8 Results of populating the Pareto solution for three OFs

The result of populating the Pareto solution in 3 dimensions for the same three OFs is shown in Figure 7.9. Having assigned the same weights to three OFs and using SCE-UA the Pareto produces a large set of equally probable (non-dominated) optimum parameters, which also include the results obtained in the standard optimization.

The advantage of using Pareto lies mostly in a research plane. The Pareto front shows which parameter sets dominate if the weight of an individual OF is changed. The Pareto analysis provides understanding of the relations among the best solutions, which trade-offs in research objectives will lead to domination of one or another parameter set. From the research perspective, the Pareto optimization is a unique opportunity to understand the whole picture of multi-objective

optimization. Also, the Pareto front provides the modeler with a choice of optimum solutions for each combination of objective functions.

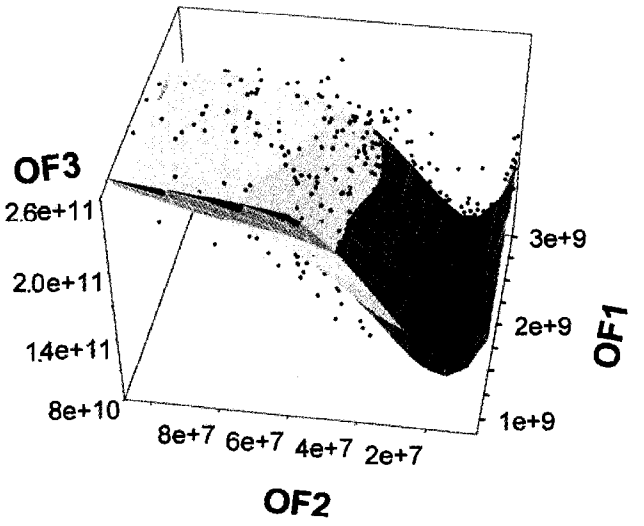


Figure 7.9 Pareto solutions plotted in 3D

Obtaining the complete Pareto front is a very computationally expensive procedure. In a routine optimization the Pareto approach is probably not the optimum calibration technique due to the huge number of runs required to obtain the complete Pareto front. In practical, every day calibrations it is more logical to do straightforward multi-objective optimization assigning specific weights to the OFs.

7.4 Conclusions

OF is the function to be optimized during model calibration and it constitutes the problem to be solved. The OF measures the misfit between the observed and simulated data, and it is typically expressed as a weighted sum of squared residuals. Proper definition of the OF is a key element of model calibration. The best optimization techniques and detailed data cannot compensate a poorly defined OF.

A single OF is most commonly used in model calibration. Various formulations of the single OF produce different optimal parameter sets. Selection of an acceptable OF should not be solely based on performance statistics which can be misleading. Graphical comparisons (percent exceedance flow curves, scatter plots, graphs of residuals, and plots of cumulative differences) of

observed and simulated values should be made. The OF expressed as a log of observed and simulated flows weighted as a log of observed flows was found to be the most appropriate single OF.

Because single OFs tend to capture only one optimization criterion multi-objective optimization (compound OF and Pareto optimum) was investigated to take into account the balance among several conflicting objectives to find the best solution.

A number of compound OFs were tested. The best fit between observed and simulated flows was obtained when the compound OF was expressed as a sum of three equally weighted OFs. The first OF improves maximum flows by minimizing a sum of squared residuals of 1% top flows. The second OF improves minimum flows by minimizing a sum of squared residuals of 20% bottom flows. The third improves the middle flows which lie between the top and bottom flows.

Visualization of Pareto frontiers can assist in selecting the optimal parameter set. In Pareto optimization several single OFs are selected and minimized individually over the feasible parameter space. Then, the best solution based on various subjective techniques is chosen. The solution, in general, is not a single unique set of parameters but consists of a Pareto set of solutions, according to various trade-offs between the different OFs.

When fitting HSPF with numerous observed responses, irregular Pareto fronts were observed, possibly due to structural and data errors.

7.5 References

Gupta, H.V., Sorooshian, S., and Yapo, P.O. (1998). Toward improved calibration of hydrological models: multiple and noncommensurable measures of information, Water Resour. Res., 34, 4, 751-763.

Gupta, H.V., Sorooshian, S., Hogue, T., and Boyle, D.P. (2003). Advances in automatic calibration of watershed models. In Advances in calibration of watershed models, Edited by Duan, Q., Sorooshian, S., Gupta, H.V., Rousseau, A.N., and Turcotte, R. Monograph Series on Water Resources, American Geophysical Union, Washington DC.

Gupta, H.V., Bastidas, L.A., Vrugt, J.A., and Sorooshian, S. (2003). Multiple criteria global optimization for watershed model calibration. In Calibration of watershed models, Edited by Duan, Q., Sorooshian, S., Gupta, H.V., Rousseau, A.N., and Turcotte, R., American Geophysical Union, Washington DC.

Khu, S.T. (1998). Automatic calibration of NAM model with Multi-objectives consideration, D2K Technical report 1298-1, National University of Singapore/Danish Hydraulic Institute.

Madsen, H. (2000). Automatic calibration of a conceptual rainfall-runoff model using multiple objectives, Journal of Hydrology, 235, 276-288.

Madsen, H. and Jacobsen, T. (2001). Automatic calibration of the MIKE SHE integrated hydrological modelling system, 4th DHI Software Conference, Helsingør, Denmark.

Schoups, G., Hopmans, J.W., Young, C.A., Vrugt, J.A., and Wallender, W.W. (2005). Multi-criteria optimization of a regional spatially-distributed subsurface flow model, Journal of Hydrology, 311, 20-48.

Dunn, S.M. and Colohan, R.J.E. (1999). Developing the snow component of a distributed hydrological model: a step-wise approach based on multi-objective analysis, Journal of Hydrology, 223, 1-2, 1-16.

Doherty, J. and Johnston, J.M. (2003). Methodologies for calibration and predictive analysis of a watershed model, J. American Water Res. Association, 39, 2, 251-265.

Zitzler, E.K. and Thiele, L. (1999). Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach, IEEE Transactions on Evolutionary Computation, 3, 4, 257-271.

Refsgaard, J.C. and Henriksen, H.J. (2004). Modelling guidelines - terminology and guiding principles, Advances in Water Resources 27, 71-82.

Rozos, E.A., Efstratiadis, I., Nalbantis, I., and Koutsoyiannis, D. (2004). Calibration of a semi-distributed model for conjunctive simulation of surface and groundwater flows, Hydrological Sciences Journal, 49, 5, 819-842.

Van Griensven, A. and Bauwens, W. (2001). Integral modelling of catchments, Water Sci. Technol., 43, 7, 321-328.

CHAPTER 8

CONCLUSIONS AND RECOMMENDATIONS

Based on the objectives and obtained results the following conclusions can be made:

- The effectiveness of three nonlinear optimization methods through detailed study of their variables was enhanced. The techniques that ensure that all regions of attraction within the parameter space are tested at various stages of optimization process were compared. Specific methods of dealing with parameter intercorrelation, insensitivity, and matrix singularity were suggested. They include elimination of or fixing insensitive and correlated parameters for each OF, log-transformation of parameters, start of optimization process from random locations, proper selection of optimization variables (e.g., Marquardt lambda), and optimum definition of the OF.

- Efficiency of HSPF for watershed modeling was improved through detailed study of data requirements (amount and resolution), robustness, and model limitations. The overparameterization problem and its impact on model performance and uncertainly was addressed. Detailed sensitivity analysis of HSPF parameters to daily flows, monthly volumes, minimum, and maximum flows was performed. Parameter sensitivity changes when a different number of parameters are optimized simultaneously. Sensitivity transformations during the optimization process were evaluated.

- The propagation of uncertainty from input data and parameter uncertainty to uncertainty of model predictions was investigated. For example, parameter uncertainty at a 95% confidence interval propagates into up to ± 8% variation in maximum flow predictions and up to ±85% variation in minimum flow predictions. FOVE, MC, and RS methods were applied and compared for uncertainty estimation. Uncertainty was expressed as confidence intervals and PDFs.

- The impact of parameter uncertainty on simulated peak flows was quantified. Total predictive interval was computed as a sum of predictive confidence intervals (uncertainty in parameter estimates) and random errors in the dependent variable (daily flows). Predictive intervals were estimated for maximum spring and minimum autumn flows using single and compound OFs. It was found that total predictive confidence intervals for maximum and minimum flows are not high. The largest portion of the confidence interval is predictive noise. Typically, the compound OF is more desirable for prediction due to its smaller predictive confidence intervals.

- Spatial scaling properties of annual runoff, river length, peak flows, and watershed area in 14 closely located watersheds of Eastern Ontario were investigated. Based on product of moments, the logarithmic relationship between k^{th} moment order and changes in scale was found to be linear suggesting that simple scaling laws hold for these watersheds. It was observed that variability of annual runoff, river length, and peak flows can be explained by the watershed area alone. There was no clear scale dependency in HSPF hydrological parameters. As well, parameter uncertainty and sensitivity do not show much scale dependency. Although simple scaling holds for annual runoff, simulation of daily flows using simple scaling appeared to be not effective and should be avoided. Transferring parameters from a calibrated neighboring watershed to an ungaged watershed and using DEM data showed results for simulation of daily flows as good as the calibrated model since HSPF utilizes topography information, hydrography, meteorological data, etc.

- Impact of input data resolution on sensitivity of model predictions was studied. It was shown how different resolutions of DEM propagate into uncertainty of topographical parameters (area, cross-section geometry, length and slope of overland slope, stream length, elevation differences), and further to simulated maximum and minimum flows. For example, assuming that the error in maximum flows should not exceed 6% and error in minimum flows should be within 12%, it is critical to have DEM resolution of 35 m for small watersheds (<200 km^2), DEM resolution of 100 m for medium watersheds (200 - 1,000 km^2), and DEM resolution of 150 m for relatively large watersheds (1,000 - 4,000 km^2). Also, the impact of DEM on hydrological parameters was investigated.

- Definitions of the OF were studied. Given the various response characteristics of the observed data, it is important to include all relevant information for a particular problem in the OF. In general, relevant information encompasses maximum, minimum, and middle flows. However, in every individual case (based on research objectives) the information incorporated in the OF can be different. Also, it is important to assign appropriate weight factors to different OFs. The weights should be assigned proportional to their perceived significance in the overall optimization. Minimum values of the OF do not necessarily correspond to the optimal solution. A number of parameter sets close to the global minimum have the same or even better performance statistics and graphical fit. Optimum formulations of single and weighted compound OFs in multi-objective optimization were suggested. Visualization of Pareto frontiers can greatly assist selection of the optimal parameter set from a Pareto set of solutions taking into account various trade-offs between different single OFs.

Recommendations for the optimal model calibration can be summarized as following:

- The method of optimization is important for model calibration. In most cases global optimization methods perform better than local ones. SCE-UA proved to be an efficient and suitable method for most model optimizations. The number of simultaneously optimized parameters should not be more than 12, maximum 15. A larger number of parameters cause higher parameter uncertainty, intensified correlation among the parameters, and significantly lengthens the optimization process. A complete SCE-UA optimization run with 12 parameters on a Pentium IV with 2.0 Ghz usually takes about 30-40 hours. An optimization run with 18 parameters lasts about 7-8 days; 4 parameters are optimized in 5-6 hours.
- The best formulation of the OF for most applications is a compound OF that is expressed as the sum of three equally weighted OFs. The first OF is the log-transformed squared residuals of the top 1 % of flows; the second OF is the log-transformed squared residuals of the bottom 20 % of flows; the third is the log-transformed residuals associated with the middle flows. For specific applications, the weights of individual single OFs can be assigned proportional to their perceived significance in the overall optimization.
- Availability of detailed meteorological data is important. For daily flow prediction at least three hourly time series are required: precipitation, potential evaporation, and air temperature. Slight improvement in calibration is observed if wind speed, solar radiation, cloud cover, and dewpoint temperature data are available. The most important meteorological data are precipitation. For an adequate flow prediction, missing precipitation data should not exceed 20-30 %, and they should be substituted with data from a nearby meteorological station.
- The global minimum of the OF is an indication of good calibration; however, often it is not the sole indicator. The final decision about the best parameter set should be based on graphical and statistical performance comparisons of observed and simulated time series. The most efficient graphical tools are scatter plots, graphs of residuals, and plots of cumulative differences. The most useful statistical characteristics are NS model fit, CE, and IA.

APPENDIX A Widely used watershed models

Acronym and version	Model name	Author, year	Comments	Availability
AGNPS 3.42	Agricultural Nonpoint Source Pollution Model	Bingner et al., 2001	examine water quality as it is affected by soil erosion from agriculture and urban areas	free download at http://www.sedlab.olemiss.edu/agnps.html
ANSWERS 2000	Areal Nonpoint Source Watershed Environment Response Simulation	Dillaha et al., 2001	distributed parameter, physically-based, continuous simulation, watershed scale, upland planning model developed for evaluating the effectiveness of agricultural and urban BMP	free download at http://dillaha.bse.vt.edu/answers/index.htm
CREAMS	Chemicals, Runoff, and Erosion from Agricultural Management System	Foster et al., 1980	a field scale model for predicting runoff, erosion, and chemical transport from agricultural management systems	free download at ftp://ftp.nrcs.usda.gov/centers/itc/applications/wqmodels/
GLEAMS 3.0	Groundwater Loading Effects of Agricultural Management Systems	Knisel and Williams, 1995	simulate hydrologic, sedimentation, and nutrient and pesticide transport in a large, complex rural watershed	free download at http://sacs.cpes.peachnet.edu/sewrl/
HEC-HMS 2.2.2	Hydrologic Modeling System	USACE, 2000	simulate precipitation-runoff processes of dendritic watershed systems. Physically based, distributed.	free download at http://www.bossintl.com/html/hec-hms-overview.html
HSPF 12.0	Hydrologic Simulation Package-Fortran IV	Bicknell at al., 2001	continuous, dynamic simulation of hydrology, hydraulic and water quality	free download athttp://www.epa.gov/ceampubl/swater/hspf/index.htm
HYDROTEL	INRS-Eau watershed model	Fortin, et al., 1995	a spatially distributed hydrological model with physical bases specifically developed to facilitate the use of remote sensing and geographical information system data	Need to contact developers http://www.inrs-eau.uquebec.ca/activites/modeles/hydrotel/en/accueil.htm
IHDM	Institute of Hydrology Distributed Model	Beven et al., 1987	physically based, distributed, continuous rainfall-runoff modeling system	available at 95 USD from http://www.wrpllc.com/books/cmwhc.html
Mike SHE 2003	Mike Systeme Hydrologique Europeen	Bathurst et al., 1995	a physically based, spatially distributed model system for water flow and sediment transport. The model simulates all major processes in the land phase of the hydrologic cycle	available at 10,000 USD from: http://www.dhisoftware.com/mikeshe/
QUAL2EU 3.22	Enhanced Stream Water Quality Model	Brown and Barnwell, 1987	a steady state model for conventional pollutants in branching streams and well mixed lakes	free download at http://www.epa.gov/ceampubl/swater/qual2eu/index.htm
SAC-SMA	Sacramento Soil Accounting Model of National Weather Service River Forecast System	Burnash, 1995	spatially-lumped, continuous conceptual runoff model that represents an attempt to parameterize the soil moisture characteristics that affect streamflow production.	Available at 95 USD from http://www.wrpllc.com/books/cmwhc.html
SWAT 2000	Soil and Water Assessment Tool	Di Luzio et al., 2002	river basin scale model developed to quantify the impact of land management practices in large, complex watersheds. Distributed, conceptual,	free download at http://www.brc.tamus.edu/swat/swat2000.html

			continuous simulation model	
SWMM 4.3	EPA Stormwater Management Model	Huber, 1995	process-oriented, semi-distributed continuous storm water model	free download at http://www.epa.gov/ceampubl/swater/swmm/index.htm
SWRRB/WQ	Simulator for Water Resources in Rural Basins (with Water Quality)	Williams, 1995	process-oriented, semi-distributed, runoff, and sediment yield simulation model	free download at http://www.cee.odu.edu/model/swrrbwq.php
TOPMODEL	Physically Based Runoff Production Model	Beven, 1995	rainfall-runoff model that bases its distributed predictions on an analysis of catchment topography. Continues hydrologic simulation model	free download at http://www.es.lancs.ac.uk/hfdg/topmodel.html
UBC	University of British Columbia Model	Quick, 1995	estimates streamflow runoff using meteorological inputs of precipitation and temperature. The model is designed for mountain runoff prediction, in which runoff is generated by snowmelt, glacier melt, and rainfall	public domain, available from http://www.civil.ubc.ca/home/ubcmodel/main.htm
WASP 6.0	Water Quality Analysis Simulation Program	Wool et al., 2001	a dynamic compartment-modeling program for aquatic systems, including both the water column and the underlying benthos. The time-varying processes of advection, dispersion, point and diffuse mass loading, and boundary exchange are represented.	free download at http://www.epa.gov/ceampubl/swater/wasp/index.htm
WATFLOOD	Waterloo Flood System	Kouwen, 2000	forecast flood flows. Process-oriented, semi-distributed continuous flow simulation model.	free download at http://www.civil.uwaterloo.ca/watflood/
WMS 7	Watershed Modeling System	Scientific Software Group	comprehensive graphical modeling environment for all phases of watershed hydrology and hydraulics	full version 4600 USD http://www.scisoftware.com/products/wms_overview/wms_overview.html
Xinanjiang Model	Chinese Watershed Model	Zhao, 1992	conceptual, semi-distributed hydrological model applicable for arid and semi-arid regions	available at 95 USD from http://www.wrpllc.com/books/cmwhc.html

APPENDIX B Characteristics of selected watershed models

Model	Source code	Infiltration	Runoff	Land surface	Subsurface treatment	GIS linked	Advantages	Disadvantages
Lumped models								
CREAMS	Fortran	Green-Ampt Eq.	Hortonian	Single crop, soil	Multiple layers	No	Public domain, simplicity	Output large ASCII files,
GLEAMS	Fortran	n/a	SCS Curve Number method	Single crop, soil	Multiple layers	No *	Public domain, simplicity	Output large ASCII files,
Semi-distributed models								
HSPF	Fortran ***	Philip Eq.	Hortonian	Homogeneous subwatersheds	2 vadose zones	Yes	All hydrological process modeled, public domain	Extensive data demand
HEC-HMS	Java	Can be selected	Can be selected	subwatersheds	Multiple layers	Yes	State-of art in public domain. Flexible structure. Compatible with US-ACE packages	Snow accumulation and melt under development. No water quality analysis
SWAT	Fortran	Green-Ampt Eq.	SCS Curve Number method	Homogeneous subwatersheds	Multiple layers	Yes	Comprehensive model structure, public domain, expensive documentation	No event simulation, extensive data demand, limited available information
SWMM	Fortran	Green-Ampt Eq.	Hortonian	Urban subwatersheds	2 zones: vadose and saturated	No **	All hydrological process, GUI (third party vendors): PCSWMM, Visual SWMM, MIKE-SWMM	No event simulations, extensive data demand, small basins, time step not flexible. No snomelt, no groundwater. Commercial graphical interface is expensive (5,000 USD)
TOP-MODEL	Fortran ***	Beven & Kirkby method	Beven & Kirkby method	subwatersheds		No	Public domain, broad coverage in literature	No final version, Win version simplified, no support
Distributed models								
HYDROTEL	C++			grid	Multiple layers	Yes	All hydrological processes. Integrated program for deriving distributed inputs. Applied in Canada.	Technical support is likely needed
MIKE/SHE	Fortran, C++	Can be selected	Can be selected	grid	Multiple layers	Yes	All hydrological process, state of art product, highly flexible, various add-on modules.	High price, technical support is likely needed, annual subscription
WATFLOOD	Visual Basic	Philip Eq.		grid	Multiple layers	Yes	All hydrological process, good support, reasonable data demand	A step behind commercial products. Simplified channel routing
ANSWERS	Fortran	Holtan Eq.	Hortonian	grid	n/a	Yes	Public domain	Model documentation and user support is very limited, only suitable for expert modelers

* - academic integration was done, * * - commercial products exist, *** - code publicly available

APPENDIX C Description and data requirement of main HSPF modules

PERLND (pervious land)

Module	Description	Required data	Comments
ATEMP	corrects air temperature for elevation differences	- elevation difference between temperature gage and land segment; - initial air temperature above land segment	module is needed when the elevation of the land segment is significantly different than the elevation at the temperature gage
SNOW	simulates accumulation and melting of snow. Module consists of 13 subroutines (METEOR, EFFPRC, COMPAC, SNOWEV, EXCHR, DEGDAY, COOLER, WARMUP, MELTER, LIQUID, ICING, GMELT, NOPACK)	- precipitation; - air temperature; - solar radiation; - dewpoint; - wind velocity; - cloud cover	two methods are available: energy balance approach and temperature index, or "degree-day" approach
PWATER	calculates water budget, primarily total runoff from a pervious area. Module consists of 8 subroutines (ICEPT, SURFAC, DISPOS, INTFLW, UZONE, LZONE, GWATER, EVAPT)	- evapotranspiration; - precipitation, infiltration; - air temperature; - rainfall; - snow cover; - water yield; - ice content of the snowpack	method of calculation is based on the original research for the LANDS subprogram of the Stanford Watershed Model IV *(Crawford and Linsley, 1966)*
SEDMNT	production and removal of sediment from a pervious area	- sediment detachment by rainfall; - fraction of the land covered by snow and other cover; - soil properties; - rainfall; - detached sediment storage; - scour of matrix soil	the equations used to produce and remove sediment are based on the predecessor models ARM and NPS (Donigian and Crawford, 1977)
PSTEMP	soil temperatures for the surface, upper, and lower/ groundwater layers of a land segment	- surface layer temperature; - slope; - air temperature	simulation is done by layers: surface layer, and subsurface layers (upper, lower, and groundwater)
PWTGAS	estimates water temperature and concentrations of dissolved oxygen and carbon dioxide in surface, interflow, and groundwater outflows from a land segment	- concentration of DO; - water temperature	
PQUAL	simulates water quality pollutants in the outflows from a pervious land segment Module consists of 4 subroutines (QUALSD, QUALOF, QUALIF, QUALGF)	- pollutant concentration - storage of pollutant on the surface; - total atmospheric deposition flux; - precipitation depth; - infiltration; - surface flow; - wash off of the pollutant from the surface; - accumulation and removal rate	for calculation the simple relationships with water and/or sediment yield is used. Up to 10 any pollutants can be simulated by this module section. The user supplies the name, units and parameter values appropriate to each of the constituents that are needed in the simulation

Agro-chemical	introduction of agricultural chemicals into streams, lakes, and groundwater from agricultural land Module consists of 5 subroutines (MSTLAY, PEST, NITR, PHOS, TRACER)	agrochemical parameters	

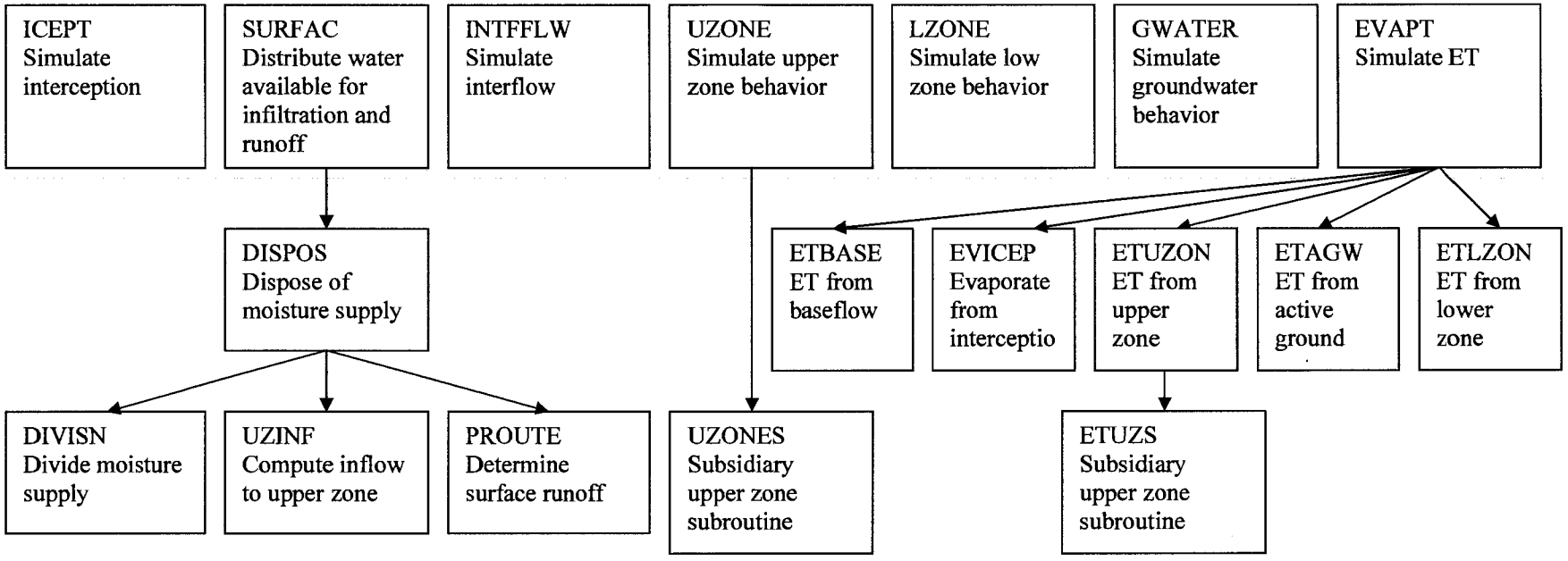
IMPLAND (impervious land)

Module	Description	Required data	Comments
ATEMP	same as PERLN	same as PERLN	
SNOW	same as PERLN	same as PERLN	
IWATER	simulates the retention, routing, and evaporation of water from an impervious land segment	- evapotranspiration; - precipitation; - air temperature; - rainfall; - snow cover;	the module has one subroutine – IROUTE which describes how Much Moisture Runs Off - IROUTE
SOLIDS	simulates the accumulation and removal of solids by runoff and other means from the impervious land segment. Module consists of 3 subroutines (SOLIDS1, SOLIDS2, ACCUM)	-capacity for removing solids; -coefficient for transport of solids; -surface water storage (inches); -surface outflow of water; -exponent for transport of solids	the equations are based on the NPS Model *(Donigian and Crawford, 1977).*
IWTGAS	estimates the water temperature and concentrations of dissolved oxygen and carbon dioxide in the outflow from the impervious land segment	-impervious surface runoff temperature; -slope; -air temperature; - concentration of DO, CD	
IQUAL	simulates water quality pollutants in the outflows from an impervious land segment. Module consists of 2 subroutines (WASHSD, WASHOF)	- pollutant concentration; - storage of pollutant on the surface; - dry or total atmospheric deposition flux - precipitation depth; - atmospheric deposition; - surface flow; - wash off of the pollutant from the surface; - accumulation and removal rate	basic algorithms used are from the NPS Model *(Donigian and Crawford, 1977)*

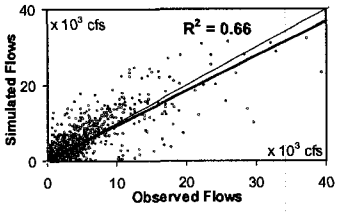
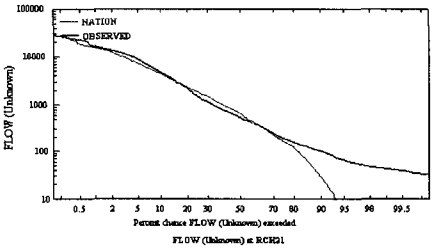
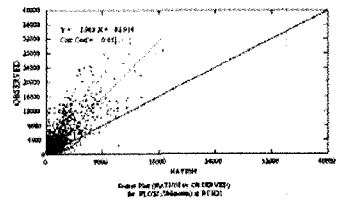
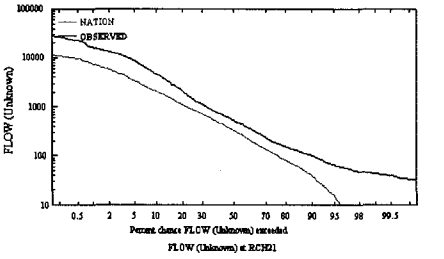
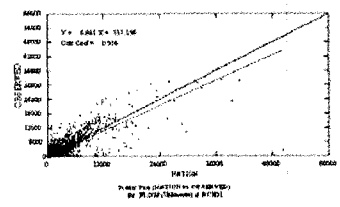
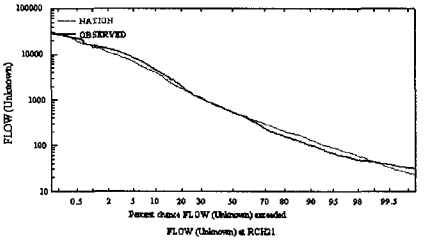
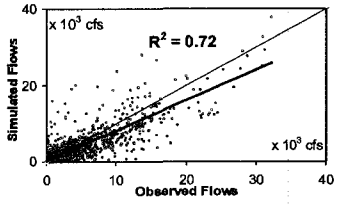
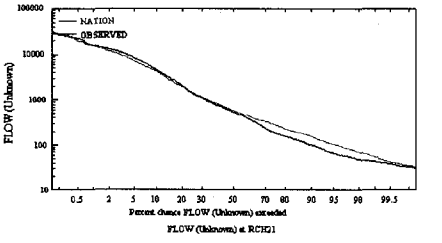
RCHRES (water body or a stream reach)

Module	Description	Required data	Comments
SINK	calculates the quantity of material settling out of a RCHRES and determines the resultant change in concentration of the material within the RCHRES	-fraction of material which settles out; - concentration of material before deposition; - sinking rate; - average depth of water	
HYDR	simulates the hydraulic processes occurring in a reach. Module consists of 4 subroutines (ROUTE, NOROUT, AUX, SHEAT)	- all hydraulic data	allows to route floods, study reservoir behavior, and analyze constituents dissolved in the water
ADCALC	calculates values for variables which are necessary to simulate longitudinal advection of dissolved or entrained constituents	- volume and outflow values; calculated in the hydraulics section (HYDR)	
CONS	simulates constituents which do not decay with time or leave water body by any mechanism other than advection. Module consists of one subroutine (ADVECT)	may applied for: - total dissolved solids; - chlorides; -pesticides and herbicides which decay very slowly	
HTRCH	simulates the processes which determine the water temperature in a reach. Module consists of one subroutine (RATEMP)	- solar radiation in langley; - cloud cover; - air temperature; - dewpoint temperature; - wind speed	
SEDTRN	simulates the transport, deposition, and scour of inorganic sediment in free flowing reaches. Module consists of 5 subroutines (COHESV, BDEXCH, SANDLD, TOFFAL, COLBY)	- particle diameter; - particle settling velocity; - particle density; - critical shear stress for deposition - critical shear stress for scour - erodibility coefficient	
GQUAL	simulates constituents which are present only in a dissolved state, or sediment-associated. Module consists of 4 subroutines (DDECAY, ADVQAL, ADECAY, ADSDES)	- decay data of dissolved material; - advection data of material on sediment; - decay of adsorbed material; - adsorption and adsorption data	
RQUAL	Simulates constituents involved in biochemical transformation	- primary DO and BOD balance; - inorganic nitrogen and phosphorus balances; - plankton population (zoo-, phyto-, benthic); - pH, carbon dioxide, total inorganic carbon, alkalinity	

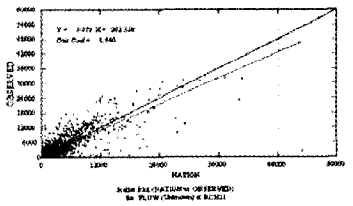
APPENDIX D Structure of subroutines in the water budget section of HSPF



APPENDIX E Performance statistics, flow duration curves, scatter plots, and parameter values for various single OFs

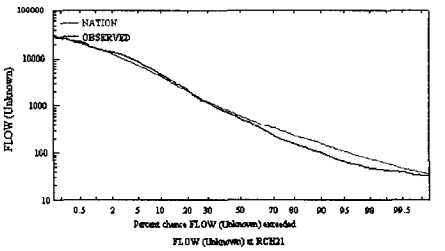
Formulation of OF and its scatter plot	Statistics	Flow duration curves	Parameters
CASE 1  $\phi = \sum_{i=1}^{n} \left(q_{obsr}^{i} - q_{sim}^{i}\right)^2$	Bias =-35.509 St.Er. =1496.461 Rel. Bias =-0.020 Rel.St.Er. =0.432 NS =0.812 CA =0.631 IA =0.808	 	SNOWCF =1.317 CCFACT =1.000 LZSN =3.598 INFILT =0.249 AGWRC =0.850 DEEPFR =0.076 BASETP =0.200 AGWETP =0.200 CEPSC =0.250 UZSN =2.000 NSUR =0.500 INTFW =9.900 IRC =0.729 LZETP =0.623
CASE 2  $\phi = \sum_{i=1}^{n} \left[\frac{1}{q_{sim}^{i}} \times \left(q_{obsr}^{i} - q_{sim}^{i}\right) \right]^2$	Bias =-916.531 St.Er. =2494.897 Rel. Bias =-0.517 Rel.St.Er. =0.721 NS =0.479 CA =0.494 IA =0.707	 	SNOWCF =1.106 CCFACT =1.000 LZSN =4.251 INFILT =0.441 AGWRC =0.898 DEEPFR =0.500 BASETP =0.200 AGWETP =0.010 CEPSC =0.050 UZSN =1.339 NSUR =0.500 INTFW =2.952 IRC =0.824 LZET =0.885
CASE 3  $\phi = \sum_{i=1}^{n} (\log q_{obsr}^{i} - \log q_{sim}^{i})^2$	Bias =-197.804 St.Er. =1898.176 Rel.Bias =-0.111 Rel.St.Er. =0.548 NS =0.698 CA =0.622 IA =0.800	 	SNOWCF =1.196 CCFACT =1.000 LZSN =2.200 INFILT =0.054 AGWRC =0.979 DEEPFR =0.500 BASETP =0.131 AGWETP =0.052 CEPSC =0.250 UZSN =2.000 NSUR =0.500 INTFW =9.900 IRC =0.751 LZETP =0.187
CASE 4  $\phi = \sum_{i=1}^{n} [\log q_{obs} \times (\log q_{obsr}^{i} - \log q_{sim}^{i})]^2$	Bias =-99.728 St.Er. =1851.729 Rel.Bias =-0.056 Rel.St.Er. =0.535 NS =0.713 CA =0.626 IA =0.803	 	SNOWCF =1.282 CCFACT =1.000 LZSN =2.00 INFILT =0.071 AGWRC =0.982 DEEPFR =0.500 BASETP =0.136 AGWETP =0.071 CEPSC =0.250 UZSN =2.000 NSUR =0.500 INTFW =9.900 IRC =0.759 LZETP =0.106

CASE 5



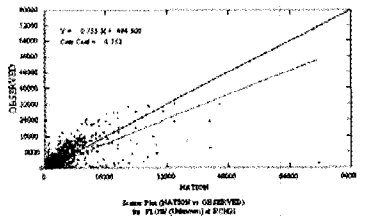
$$\phi = \sum_{i=1}^{n} [q_{obs} \times (\log q^{i}_{obsr} - \log q^{i}_{sim})]^{2}$$

Bias =-77.074
St.Er. =1921.959
Rel.Bias =-0.043
Rel.St.Er. =0.555
NS =0.691
CA =0.619
IA =0.800



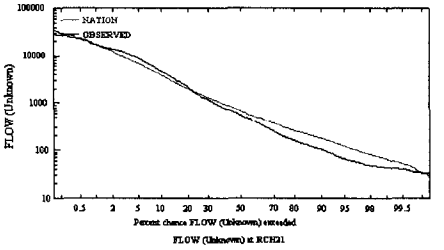
SNOWCF =1.291
CCFACT =1.059
LZSN =2.000
INFILT =0.0640
AGWRC =0.981
DEEPFR =0.500
BASETP =0.104
AGWETP =0.061
CEPSC =0.250
UZSN =2.000
NSUR =0.500
INTFW =9.900
IRC =0.760
LZETP =0.119

CASE 6



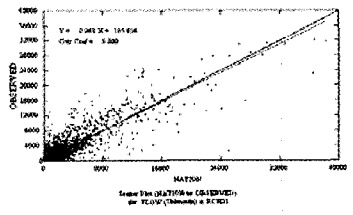
$$\phi = \sum_{i=1}^{n} [\sqrt{q_{obs}} \times (\log q^{i}_{obsr} - \log q^{i}_{sim})]^{2}$$

Bias =-80.317
St.Er. =2433.146
Rel.Bias =-0.045
Rel.St.Er. =0.703
NS =0.504
CA = 0.555
IA =0.762



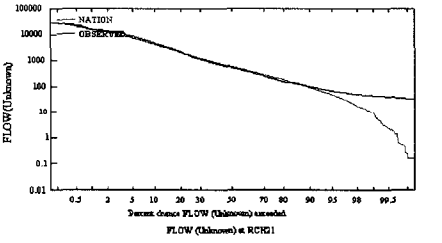
SNOWCF =1.173
CCFACT =1.111
LZSN =3.621
INFILT =0.071
AGWRC =0.938
DEEPFR =0.313
BASETP =0.111
AGWETP =0.010
CEPSC =0.050
UZSN =1.417
NSUR =0.500
INTFW =3.493
IRC =0.654
LZETP =0.261

CASE 7



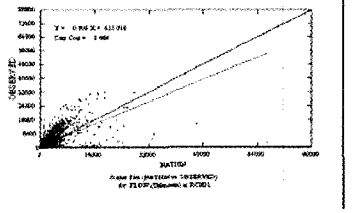
$$\phi = \sum_{i=1}^{n} (\sqrt{q^{i}_{obsr}} - \sqrt{q^{i}_{sim}})^{2}$$

Bias =-124.879
St.Er. =1602.232
Rel Bias =-0.070
Rel.St.Er. =0.463
NS =0.785
CA =0.641
IA =0.812



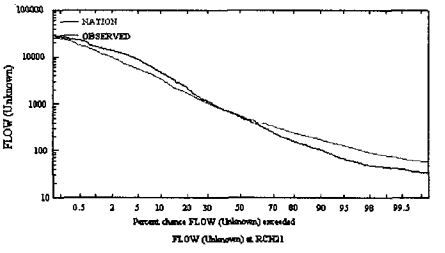
SNOWCF =1.264
CCFACT =1.000
LZSN =3.750
INFILT =0.110
AGWRC =0.876
DEEPFR =0.188
BASETP =0.200
AGWETP =0.010
CEPSC =0.0400
UZSN =1.628
NSUR =0.50
INTFW =9.900
IRC =0.724
LZETP =0.504

CASE 8



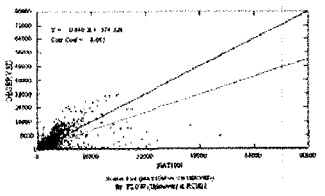
$$\phi = \sum_{i=1}^{n} \frac{1}{\log q_{obs}} (\sqrt{q^{i}_{obsr}} - \sqrt{q^{i}_{sim}})^{2}$$

Bias =-360.585
St.Er. =2670.011
Rel Bias =-0.203
Rel.St.Er. =0.772
NS =0.403
CE =0.508
IA =0.727



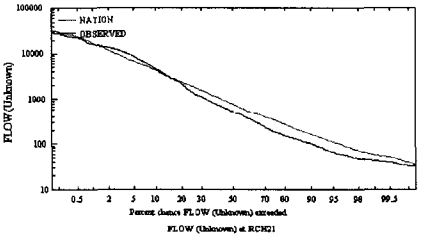
SNOWCF =1.00
CCFACT =1.207
LZSN =2.600
INFILT =0.0539
AGWRC =0.977
DEEPFR =0.500
BASETP =0.048
AGWETP =0.023
CEPSC =0.050
UZSN =1.589
NSUR =0.233
INTFW =3.880
IRC =0.758
LZETP =0.236

CASE 9



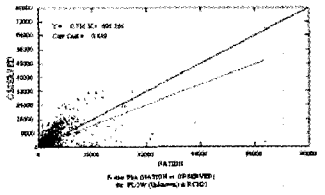
$$\phi = \sum_{i=1}^{n} [\frac{1}{q_{obs}} (\log q^i_{obsr} - \log q^i_{sim})]^2$$

Bias =75.355
St.Er. =2869.338
Rel.Bias =0.042
Rel.St.Er. =0.829
NS =0.311
CE =0.494
IA =0.730



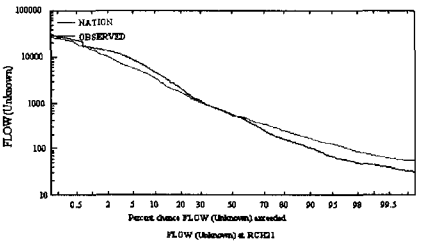
SNOWCF =1.230
CCFACT =1.530
LZSN =4.000
INFILT =0.043
AGWRC =0.945
DEEPFR =0.107
BASETP =0.200
AGWETP =0.015
CEPSC =0.050
UZSN =1.654
NSUR =0.500
INTFW =3.674
IRC =0.751
LZETP =0.348

CASE 10



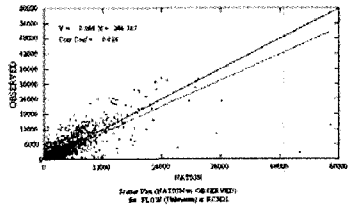
$$\phi = \sum_{i=1}^{n} (\frac{1}{\sqrt{q^i_{obsr}}} - \frac{1}{\sqrt{q^i_{sim}}})^2$$

Bias =-360.585
St.Er. =2670.010
Rel.Bias =-0.203
Rel.St.Er. =0.772
NS =0.403
CA =0.508
IA =0.727



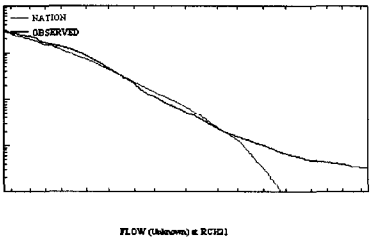
SNOWCF =1.000
CCFACT =1.207
LZSN =2.600
INFILT =0.0539
AGWRC =0.977
DEEPFR =0.500
BASETP =0.0484
AGWETP =0.023
CEPSC =0.050
UZSN =1.589
NSUR =0.233
INTFW =3.880
IRC =0.758
LZETP =0.236

CASE 11



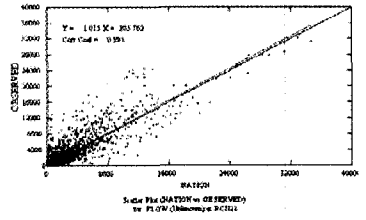
$$\phi = \sum_{i=1}^{n} (V^i_{obsr} - V^i_{sim})^2$$

Bias =-31.418
St.Er. =2057.108
Rel.Bias =-0.017
Rel.St.Er. =0.594
NS =0.646
CA =0.580
IA =0.780



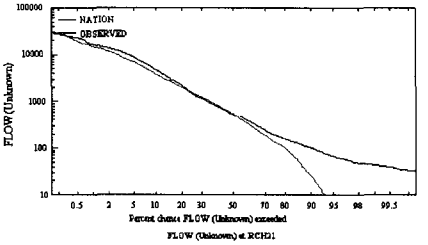
SNOWCF =1.097
CCFACT =1.000
LZSN =2.852
INFILT =0.0788
AGWRC 0.850
DEEPFR 0E-02
BASETP =0.0277
AGWETP =0.200
CEPSC =0.040
LZSN =2.000
NSUR =0.500
INTFW =3.517
IRC =0.787
LZETP =0.461

CASE 12



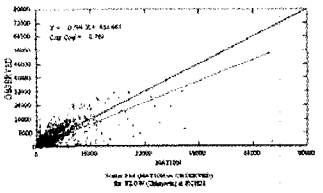
$$\phi = \sum_{i=1}^{n} [\frac{1}{\sqrt{V^i_{obsr}}} \times (V^i_{obsr} - V^i_{sim})]^2$$

Bias =-228.976
St.Er. =1572.121
Rel Bias =-0.129
Rel.St.Er. =0.454
NS =0.793
CA =0.642
IA =0.811



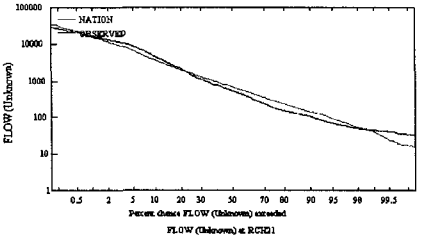
SNOWCF =1.502
CCFACT =1.071
LZS =4.620
INFILT =0.290
AGWRC =0.862
DEEPFR =0.250
BASETP =0.200
AGWETP =0.178
CEPSC =0.040
UZSN =1.754
NSUR =0.500
INTFW =5.914
IRC =0.751
LZETP =0.717

CASE 13



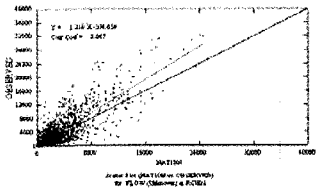
$$\phi = \sum_{i=1}^{n} (\log\ V_{obsr}^{i} - \log\ V_{sim}^{i})^2$$

Bias =-89.881
St.Er. =2329.812
Rel Bias =-0.050
Rel St.Er. =0.673
NS =0.545
CE =0.562
IA =0.765



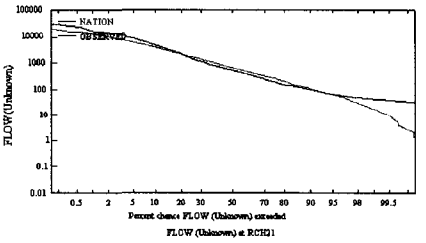
SNOWCF =1.295
CCFACT =1.185
LZSN =4.103
INFILT =0.121
AGWRC =0.919
DEEPFR =0.153
BASETP =0.200
AGWETP =0.010
CEPSC =0.050
UZSN =1.766
NSUR =0.099
INTFW =2.498
IRC =0.674
LZETP =0.510

CASE 14



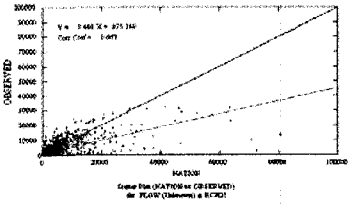
$$\phi = \sum_{i=1}^{n} \log V_{obsr} \times (\log V_{obsr}^{i} - \log V_{sim}^{i})^2$$

Bias =-230.368
St.Err. =1820.161
Rel Bias =-0.130
Rel St.Er. =0.526
NS =0.722
CE =0.580
IA =0.766



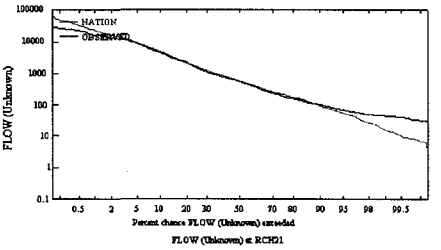
SNOWCF =1.283
CCFACT =1.485
LZSN =5.751
INFILT =0.226
AGWRC =0.893
DEEPFR =0.194
BASETP =0.200
AGWETP =0.012
CEPSC =0.040
UZSN =1.355
NSUR =0.500
INTFW =9.900
IRC =0.850
LZETP =0.669

CASE 15



$$\phi = \sum_{i=1}^{k} (E_{obsr}^{i} - E_{sim}^{i})^2$$

Bias =227.989
St.Er. =3837.319
Rel.Bias =0.1287
Rel.St.Er. =1.109
NS =0.231
CA =0.445
IA =0.734

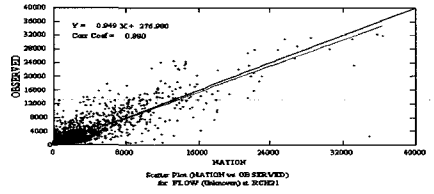
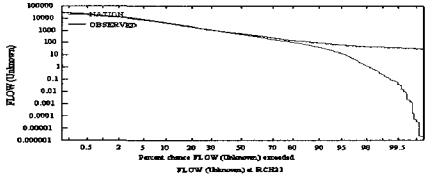


SNOWCF =1.483
CCFACT =1.180
LZSN =3.960
INFILT =0.0495
AGWRC =0.926
DEEPFR =0.206
BASETP =0.183
AGWETP =0.012
CEPSC =0.050
UZSN =0.908
NSUR =0.185
INTFW =2.071
IRC =0.641
LZETP =0.469

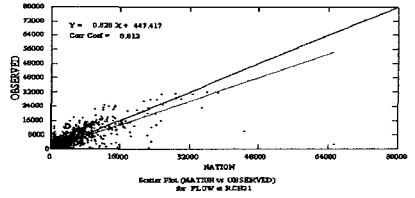
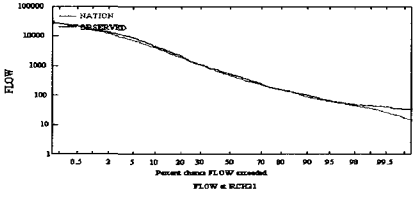
APPENDIX F Performance statistics, flow duration curves, scatter plots, and parameter values for various compound OFs

Maximum flows + Minimum flows + Monthly volumes

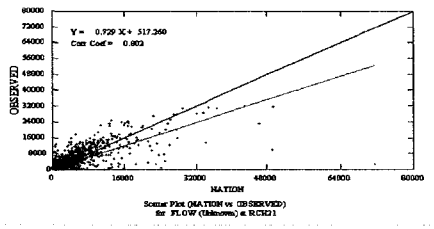
Case 1 $$\phi = \sum_{i=1}^{n} [2 \times (q \max^{i}_{obsr} - q \max^{i}_{sim})]^2 + \sum_{i=1}^{k} [12 \times (q \min^{i}_{obsr} - q \min^{i}_{sim})]^2 + \sum_{i=1}^{j} [\frac{0.25}{\sqrt{V_{obsr}}} \times (V^{i}_{obsr} - V^{i}_{sim})]^2$$

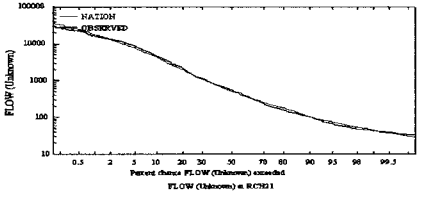
	Daily flows	Monthly volumes	March-April	August-September	Top 1% flows	Parameters
	Bias = -196.657 St.Er. = 1662.741 Rel.Bias = -0.111 Rel.St.Er. =0.480 NS = 0.768 CE = 0.632 IA = 0.811	Bias = -5.17E+08 St.Er. = 2.08E+09 Rel.Bias = -0.111 Rel.St.Er. = 0.330 NS = 0.890 CE = 0.689 IA = 0.846	Bias = -4.61E+08 St.Er. = 3.92E+09 Rel.Bias = -3.0E-02 Rel.St.Er. = 0.463 NS = 0.785 CE = 0.548 IA = 0.770	Bias = 1.60E+08 St.Er. = 7.35E+08 Rel.Bias = -0.226 Rel.St.Er. = 1.311 NS = -0.720 CE = -0.130 IA = 0.461	Bias = -3724.382 St.Er. =6508.515 Rel.Bias= -0.161 Rel.St.Er. = 1.405 NS = -0.976 CE = -0.462 IA = 0.507	SNOWCF = 1.258 CCFACT = 1.000 LZSN = 2.000 INFILT = 0.079 AGWRC = 0.850 DEEPFR = 0.306 BASETP = 0.177 AGWETP = 0.012 CEPSC = 0.250 UZSN 2.00 NSUR = 0.500 INTFW = 9.900 IRC = 0.802 LZETP = 0.753
						

Case 2 $$\phi = \sum_{i=1}^{n} [2 \times (q \max^{i}_{obsr} - q \max^{i}_{sim})]^2 + \sum_{i=1}^{k} ([5000 \times \log q \min \times (\log q \min^{i}_{obsr} - \log q \min^{i}_{sim})]^2 + \sum_{i=1}^{j} [\frac{0.25}{\sqrt{V_{obsr}}} \times (V^{i}_{obsr} - V^{i}_{sim})]^2$$

	Bias = -172.6159 St. Er. = 2105.949 Rel.Bias = -0.097 Rel.St.Er. = 0.609 NS = 0.629 CE = 0.599 IA = 0.793	Bias = -4.53E+08 St.Er. = 2.16E+09 Rel.Bias = -0.097 Rel.St.Er. = 0.343 NS = 0.881 CE = 0.675 IA = 0.837	Bias = -7.13E+08 St.Er. = 4.28E+09 Rel.Bias = -0.046 Rel.St.Er. = 0.505 NS = 0.744 CE = 0.488 IA = 0.734	Bias = -1.09E+08 St.Er. = 7.45E+08 Rel.Bias = -0.155 Rel.St.Er. = 1.328 NS = -0.766 CE = -0.111 IA = 0.452	Bias = -3796.884 St.Er. = 6795.714 Rel.Bias = -0.164 Rel.St.Er. = 1.467 NS = -1.154 CE = -0.500 IA = 0.506	SNOWCF = 1.262 CCFACT = 1.102 LZSN = 2.000 INFILT = 0.027 AGWRC = 0.975 DEEPFR = 0.500 BASETP = 0.128 AGWETP = 0.010 CEPSC = 0.250 UZSN = 2.000 NSUR = 0.500 INTFW = 9.900 IRC = 0.808 LZETP = 0.446
						

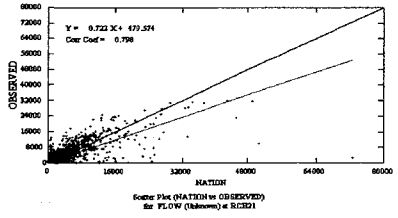
Case 3 $\phi = \sum_{i=1}^{n} [8.5 \times (\log q \max {}^{i}_{obsr} - \log q \max {}^{i}_{sim})]^2 + \sum_{i=1}^{k} [1 \times (\log q \min {}^{i}_{obsr} - \log q \min {}^{i}_{sim})]^2 + \sum_{i=1}^{j} [4 \times (\log V^{i}_{obsr} - \log V^{i}_{sim})]^2$

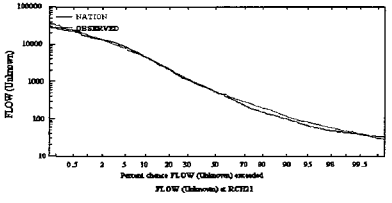




Bias = -51.117	Bias = -1.34E+08	Bias = -7.95E+07	Bias = 6.98E+07	Bias = -1450.530	SNOWCF = 1.276
St.Er. = 2308.125	St.Er. = 2.13E+09	St.Er. = 4.31E+09	St.Er. = 7.96E+08	St.Er. = 7763.240	CCFACT = 1.020
Rel Bias = -0.028	Rel Bias = -0.028	Rel Bias = -5.1E-03	Rel.Bias = 0.099	Rel Bias = -0.063	LZSN = 2.001
Rel St.Er. = 0.667	Rel St.Er. = 0.337	Rel.St.Er. = 0.508	Rel.St.Er. = 1.419	Rel.St.Er. = 1.676	INFILT = 0.023
NS = 0.554	NS = 0.885	NS = 0.741	NS = -1.016	NS = -1.811	AGWRC = 0.976
CE = 0.598	CE = 0.698	CE = 0.515	CE = -0.205	CE = -0.627	DEEPFR = 0.500
IA = 0.797	IA = 0.848	IA = 0.746	IA = 0.418	IA = 0.506	BASETP = 0.100
					AGWETP = 0.010
					CEPSC = 0.250
					UZSN = 2.000
					NSUR = 0.500
					INTFW = 9.900
					IRC = 0.740
					LZETP = 0.286

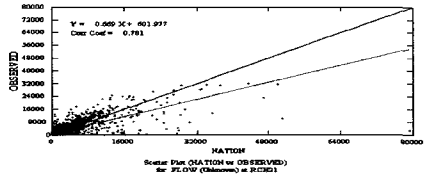
Case 4 $\phi = \sum_{i=1}^{n} [8.5 \times (\log q \max {}^{i}_{obsr} - \log q \max {}^{i}_{sim})]^2 + \sum_{i=1}^{k} [0.5 \times (\log q \min {}^{i}_{obsr} - \log q \min {}^{i}_{sim})]^2 + \sum_{i=1}^{j} [4 \times (\log V^{i}_{obsr} - \log V^{i}_{sim})]^2$

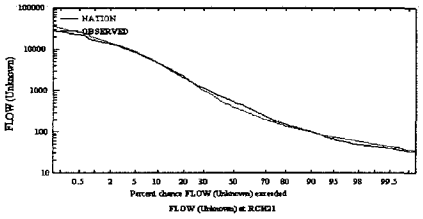




Bias = 17.277	Bias = 4.55E+07	Bias = 1.46E+08	Bias = 1.85E+08	Bias = -1657.152	SNOWCF = 1.289
St.Er. = 2339.906	St.Er. = 2.13E+09	St.Er. = 4.33+09	St.Er. = 8.65E+08	St.Er. = 7690.836	CCFACT = 1.010
Rel Bias = 0.009	Rel.Bias = 0.009	Rel. Bias = 0.009	Rel.Bias = 0.263	Rel.Bias = -0.071	LZSN = 2.004
Rel.St.Er. = 0.676	Rel.St.Er. = 0.339	Rel.St.Er. = 0.510	Rel.St.Er. = 1.544	Rel.St.Er. = 1.661	INFILT = 0.023
NS = 0.542	NS = 0.885	NS = 0.738	NS = -1.385	NS = -1.759	AGWRC = 0.982
IA = 0.593	IA = 0.702	IA = 0.516	IA = -0.276	IA = -0.617	DEEPFR = 0.500
IA = 0.794	IA = 0.850	IA = 0.750	IA = 0.423	IA = 0.512	BASETP = 0.044
					AGWETP = 0.028
					CEPSC = 0.250
					UZSN = 2.000
					NSUR = 0.500
					INTFW = 9.900
					IRC = 0.758
					LZETP = 0.195

Case 5 $\phi = \sum_{i=1}^{n} [8.5 \times (\log\ q\ \max\ {}^{i}_{obsr} - \log\ q\ \max\ {}^{i}_{sim})]^2 + \sum_{i=1}^{k} [1.0 \times (\log\ q\ \min\ {}^{i}_{obsr} - \log\ q\ \min\ {}^{i}_{sim})]^2 + \sum_{i=1}^{j} [2 \times (\log\ V^{i}_{obsr} - \log\ V^{i}_{sim})]^2$

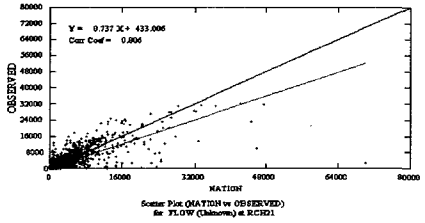


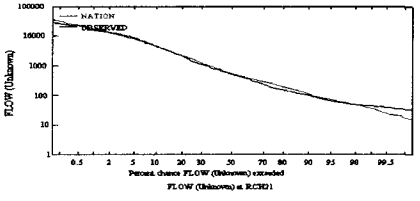


Daily flows	Monthly volumes	March-April	August-September	Top 1% flows	Parameters
Bias = -22.706	Bias = -5.95E+07	Bias = 2.80E+08	Bias = -1.12E+08	Bias = -1230.448	SNOWCF = 1.266
St.Er. = 2541.782	St.Er. = 2.43E+09	St.Er. = 4.68E+09	St.Er. = 7.28E+08	St.Er. = 7828.268	CCFACT = 1.001
Rel.Bias = -0.012	Rel.Bias = -0.012	Rel.Bias = 0.018	Rel.Bias = -0.159	Rel.Bias = -0.053	LZSN = 2.000
Rel.St.Er. = 0.735	Rel.St.Er. = 0.386	Rel.St.Er. = 0.553	Rel.St.Er. = 1.299	Rel.St.Er. = 1.690	INFILT = 0.023
NS = 0.459	NS = 0.850	NS = 0.694	NS = -0.688	NS = -1.858	AGWRC = 0.989
CE = 0.569	CE = 0.652	CE = 0.465	CE = -0.036	CE = -0.636	DEEPFR = 0.500
IA = 0.789	IA = 0.830	IA = 0.722	IA = 0.474	IA = 0.500	BASETP = 0.200
					AGWETP = 0.017
					CEPSC = 0.250
					UZSN = 1.006
					USUR = 0.500
					INTFW = 9.900
					IRC = 0.751
					LZETP = 0.606

Maximum flows + Middle flows + Exceedance of minimum flows

Case 6 $\phi = \sum_{i=1}^{n} [0.8 \times (\log\ q\ mid\ {}^{i}_{obsr} - \log\ q\ mid\ {}^{i}_{sim})]^2 + \sum_{i=1}^{k} [10 \times (\log\ q\ \max\ {}^{i}_{obsr} - \log\ q\ \max\ {}^{i}_{sim})]^2 + \sum_{i=1}^{j} [350 \times (E\ \min\ {}^{i}_{obsr} - E\ \min\ {}^{i}_{sim})]^2$



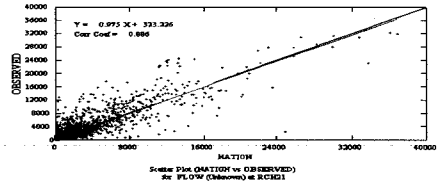


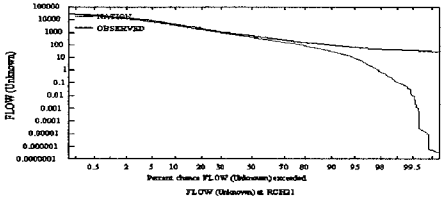
Daily flows	Monthly volumes	March-April	August-September	Top 1% flows	Parameters
Bias = 44.157	Bias = 1.16E+08	Bias = -2.50E+07	Bias = 2.37E+08	Bias = -1801.572	SNOWCF = 1.277
St.Er. = 2276.757	St.Er. = 2.18E+09	St.Er. = 4.41E+09	St.Er. = 9.25E+08	St.Er. = 7525	CCFACT = 1.002
Rel.Bias = 0.024	Rel.Bias = 0.024	Rel.Bias = -0.001	Rel.Bias = 0.336	Rel.Bias = -0.078	LZSN = 2.000
Rel.St.Er. = 0.658	Rel.St.Er. = 0.346	Rel.St.Er. = 0.520	Rel.St.Er. = 1.650	Rel.St.Er. = 1.625	INFILT = 0.030
NS = 0.566	NS = 0.880	NS = 0.729	NS = -1.723	NS = -1.641	AGWRC = 0.984
CE = 0.595	CE = 0.693	CE = 0.500	CE = -0.318	CE = -0.595	DEEPFR = 0.497
IA = 0.795	IA = 0.844	IA = 0.739	IA = 0.443	IA = 0.511	BASETP = 0.139
					AGWETP = 0.058
					CEPSC = 0.250
					UZSN = 1.999
					NSUR = 0.500
					INTFW = 9.900
					IRC = 0.756
					LZETP = 0.100

Middle flows + Maximum flows + Minimum flows

Case 7

$$\phi = \sum_{i=1}^{n} [1.1 \times (q\ mid^{\,i}_{obsr} - q\ mid^{\,i}_{sim})]^2 + \sum_{i=1}^{k} [2 \times (q \max^{\,i}_{obsr} - q \max^{\,i}_{sim})]^2 + \sum_{i=1}^{j} [18 \times (q \min^{\,i}_{obsr} - q \min^{\,i}_{sim})]^2$$

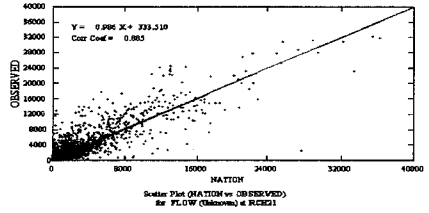


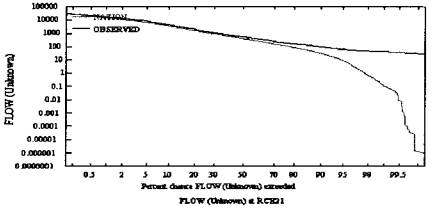


Daily flows	Monthly volumes	March-April	August-September	Top 1% flows	Parameters
Bias = -286.355	Bias = -7.52E+08	Bias = -1.02E+09	Bias = -2.24E+08	Bias = -4156.640	SNOWCF = 1.299
St.Er. = 1627.614	St.Er. = 2.19E+09	St.Er. = 4.09E+09	St.Er. = 6.99E+08	St.Er. = 6924.228	CCFACT = 1.000
Rel.Bias = -0.161	Rel.Bias = -0.161	Rel.Bias = -6.67E-02	Rel.Bias = -0.318	Rel.Bias = -0.180	LZSN = 2.000
Rel.St.Er. = 0.470	Rel.St.Er. = 0.348	Rel.St.Er. = 0.482	Rel.St.Er. = 1.248	Rel.St.Er. = 1.495	INFILT = 0.120
NS = 0.778	NS = 0.878	NS = 0.767	NS = -0.558	NS = -1.236	AGWRC = 0.850
CE = 0.636	CE = 0.679	CE = 0.541	CE = -0.102	CE = -0.488	DEEPFR = 0.367
IA = 0.813	IA = 0.841	IA = 0.764	IA = 0.471	IA = 0.527	BASETP = 0.200
					AGWETP = 0.010
					CEPSC = 0.250
					UZSN = 1.883
					NSUR = 0.500
					INTFW = 9.900
					IRC = 0.776
					LZETP = 0.869

Case 8

$$\phi = \sum_{i=1}^{n} [0.8 \times (q\ mid^{\,i}_{obsr} - q\ mid^{\,i}_{sim})]^2 + \sum_{i=1}^{k} [2 \times (q \max^{\,i}_{obsr} - q \max^{\,i}_{sim})]^2 + \sum_{i=1}^{j} [20 \times (q \min^{\,i}_{obsr} - q \min^{\,i}_{sim})]^2$$

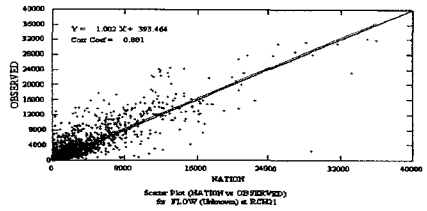


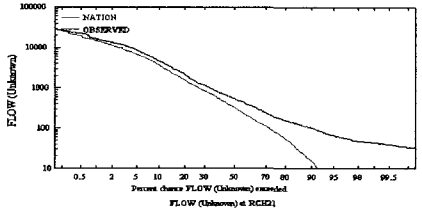


Bias = -313.155	Bias = -8.23E+08	Bias = -1.22E+09	Bias = -2.32E+08	Bias = -4289.270	SNOWCF = 1.281
St.Er. = 1641.759	St.Er. = 2.22E+09	St.Er. = 4.16E+09	St.Er. = 6.97E+08	St.Er. = 6879.478	CCFACT = 1.012
Rel.Bias = -0.176	Rel.Bias = -0.176	Rel.Bias = -0.080	Rel.Bias = -0.329	Rel.Bias = -0.186	LZSN = 2.000
Rel.St.Er. = 0.474	Rel.St.Er. = 0.352	Rel.St.Er. = 0.491	Rel.St.Er. = 1.244	Rel.St.Er. = 1.485	INFILT = 0.105
NS = 0.774	NS = 0.875	NS = 0.758	NS = -0.549	NS = -1.207	AGWRC = 0.850
CE = 0.633	CE = 0.673	CE = 0.534	CE = -0.087	CE = -0.524	DEEPFR = 0.398
IA = 0.811	IA = 0.838	IA = 0.759	IA = 0.472	IA = 0.505	BASETP = 0.200
					AGWETP = 0.013
					CEPSC = 0.250
					UZSN = 1.963
					NSUR = 0.500
					INTFW = 9.900
					IRC = 0.781
					LZETP = 0.801

Case 9

$$\phi = \sum_{i=1}^{n} [1.1 \times (q\ mid\ ^{i}_{obsr} - q\ mid\ ^{i}_{sim})]^2 + \sum_{i=1}^{k} [2 \times (q \max\ ^{i}_{obsr} - q \max\ ^{i}_{sim})]^2 + \sum_{i=1}^{j} [30 \times (q \min\ ^{i}_{obsr} - q \min\ ^{i}_{sim})]^2$$

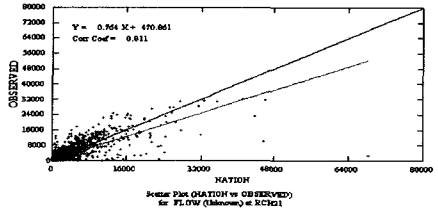


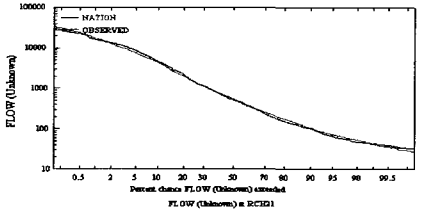


Bias = -396.593	Bias = -1.04E+09	Bias = -1.64E+09	Bias = -3.57E+08	Bias = -4842.596	SNOWCF = 1.293
St.Er. = 1681.951	St.Er. = 2.34E+09	St. Er. = 4.39E+09	St.Er. = 7.25+08	St.Er. = 7238.908	CCFACT = 1.000
Rel.Bias = -0.223	Rel.Bias = -0.223	Rel.Bias = -0.107	Rel.Bias = -0.507	Rel.Bias = -0.210	LZSN = 2.000
Rel.St.Er. = 0.486	Rel.St.Er. = 0.372	Rel.St.Er. = 0.518	Rel.St.Er. = 1.292	Rel.St.Er. = 1.563	INFILT = 0.100
NS = 0.763	NS = 0.861	NS = 0.731	NS = -0.671	NS = -1.444	AGWRC = 0.850
CE = 0.622	CE = 0.649	CE = 0.506	CE = -0.117	CE = -0.598	DEEPFR = 0.500
IA = 0.806	IA = 0.827	IA = 0.744	IA = 0.477	IA = 0.496	BASETP = 0.125
					AGWETP = 0.197
					CEPSC = 0.250
					UZSN = 1.953
					NSUR = 0.500
					INTFW = 9.900
					IRC = 0.788
					LZETP = 0.708

Case 10

$$\phi = \sum_{i=1}^{n} [1.0 \times (\log\ q\ \ mid\ ^{i}_{obsr} - \log\ qmid\ ^{i}_{sim})]^2 + \sum_{i=1}^{k} [12 \times (\log\ q\ \ \max\ ^{i}_{obsr} - \log\ q \max\ ^{i}_{sim})]^2 + \sum_{i=1}^{j} [1.5 \times (\log\ q \min\ ^{i}_{obsr} - \log\ q \min\ ^{i}_{sim})]^2$$

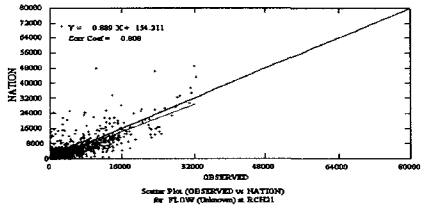


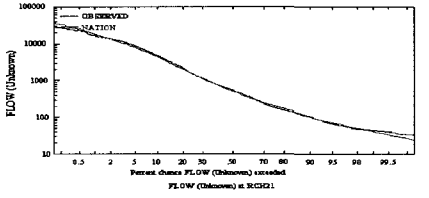


Bias = -68.262	Bias = -1.79E+08	Bias = -3.63E+08	Bias = 5.83E+07	Bias = -2152.683	SNOWCF = 1.259
SE = 2200.745	SE = 2.14E+09	SE = 4.32E+09	SE = 7.937E+08	SE = 7360.970	CCFACT = 1.021
Rel.Bias = -0.038	Rel.Bias = -0.038	Rel.Bias = -0.023	Rel.Bias = 0.082	Rel.Bias = -0.093	LZSN = 2.000
Rel.St.Er. = 0.636	Rel St.Er. = 0.340	Rel.St.Er. = 0.510	Rel.SE = 1.415	Rel.St.Er. = 1.589	INFILT = 0.029
NS = 0.594	NS = 0.884	NS = 0.739	NS = -1.003	NS = -1.527	AGWRC = 0.972
CE = 0.607	CE = 0.693	CE = 0.510	CE = -0.209	CE = -0.579	DEEPFR = 0.500
IA = 0.800	IA = 0.845	IA = 0.741	IA = 0.413	IA = 0.507	BASETP = 0.116
					AGWETP = 0.010
					CEPSC = 0.250
					UZSN = 1.834
					NSUR = 0.500
					INTFW = 9.900
					IRC = 0.748
					LZETP = 0.283

Case 11

$$\phi = \sum_{i=1}^{n} [0.8 \times (\log\ q\ \ mid\ {}^{i}_{obsr} - \log\ q\ \ mid\ {}^{i}_{sim})]^2 + \sum_{i=1}^{k} [14 \times (\log\ q\ \max\ {}^{i}_{obsr} - \log\ q\ \max\ {}^{i}_{sim})]^2 + \sum_{i=1}^{j} [1.5 \times (\log\ q\ \min\ {}^{i}_{obsr} - \log\ q\ \min\ {}^{i}_{sim})]^2$$

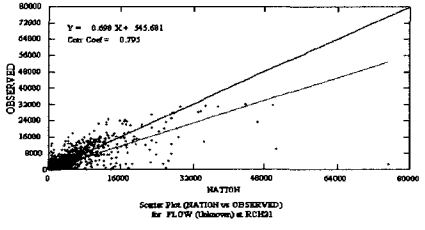


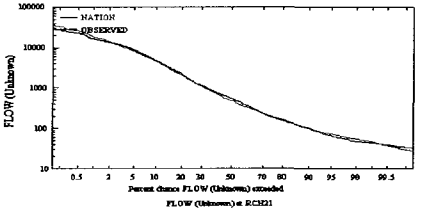


Bias = -43.137	Bias = -1.13E+08	Bias = 2.50E+07	Bias = 3.85E+07	Bias = -1470.204	SNOWCF = 1.303
St.Er. = 2275.493	St.Er. = 2.16E+09	St.Er. = 4.37E+09	St.Er. = 7.87E+08	St.Er. = 7782.584	CCFACT = 1.023
Rel.Bias = -0.024	Rel.Bias = -0.024	Rel.Bias = 1.63E-03	Rel.Bias = 0.054	Rel.Bias = -0.063	LZSN = 2.000
Rel.St.Er. = 0.658	Rel.St.Er. = 0.343	Rel.St.Er. = 0.515	Rel.St.Er. = 1.404	Rel.St.Er. = 1.680	INFILT = 0.027
NS = 0.566	NS = 0.882	NS = 0.734	NS = -0.973	NS = -1.825	AGWRC = 0.972
CE = 0.601	CE = 0.693	CE = 0.508	CE = -0.200	CE = -0.627	DEEPFR = 0.500
IA = 0.799	IA = 0.846	IA = 0.745	IA = 0.418	IA = 0.512	BASETP = 0.159
					AGWETP = 0.012
					CEPSC = 0.251
					UZSN = 1.915
					NSUR = 0.500
					INTFW = 9.900
					IRC = 0.747
					LZETP = 0.297

Case 12

$$\phi = \sum_{i=1}^{n} [0.7 \times (\log\ q\ \ mid\ {}^{i}_{obsr} - \log\ q\ \ mid\ {}^{i}_{sim})]^2 + \sum_{i=1}^{k} [12 \times (\log\ q\ \max\ {}^{i}_{obsr} - \log\ q\ \max\ {}^{i}_{sim})]^2 + \sum_{i=1}^{j} [1.0 \times (\log\ q\ \min\ {}^{i}_{obsr} - \log\ q\ \min\ {}^{i}_{sim})]^2$$

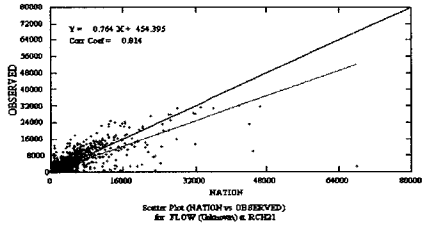


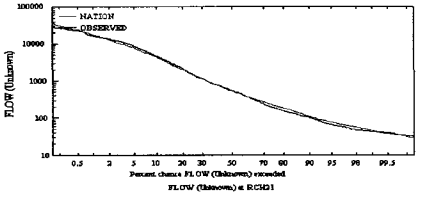


Bias = -15.828	Bias = -4.14E+07	Bias = 2.7549E+08	Bias = -7.05E+07	Bias = -1089.566	SNOWCF = 1.285
St.Er. = 2411.342	St.Er. = 2.26E+09	St.Er. = 4.48E+09	St.Er. = 7.64E+08	St.Er. = 7836.041	CCFACT = 1.002
Rel.Bias = -0.008	Rel.Bias = -0.008	Rel.Bias = 0.017	Rel.Bias = -0.010	Rel.Bias = -0.04	LZSN = 2.000
Rel.St.Er. = 0.697	Rel.St.Er. = 0.359	Rel.St.Er. = 0.529	Rel.St.Er. = 1.363	Rel.St.Er. = 1.692	INFILT = 0.024
NS = 0.513	NS = 0.871	NS = 0.719	NS = -0.859	NS = -1.864	AGWRC = 0.980
CE = 0.588	CE = 0.678	CE = 0.490	CE = -0.149	CE = -0.632	DEEPFR = 0.500
IA = 0.795	IA = 0.841	IA = 0.737	IA = 0.436	IA = 0.5061	BASETP = 0.200
					AGWETP = 0.010
					CEPSC = 0.250
					UZSN = 1.471
					NSUR = 0.500
					INTFW = 9.900
					IRC = 0.752
					LZETP = 0.369

Case 13

$$\phi = \sum_{i=1}^{n} [1.2 \times (\log q\ mid^{i}_{obsr} - \log q\ mid^{i}_{sim})]^2 + \sum_{i=1}^{k} [14 \times (\log q \max^{i}_{obsr} - \log q \max^{i}_{sim})]^2 + \sum_{i=1}^{j} [1.5 \times (\log q \min^{i}_{obsr} - \log q \min^{i}_{sim})]^2$$

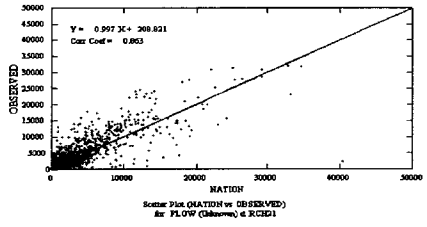


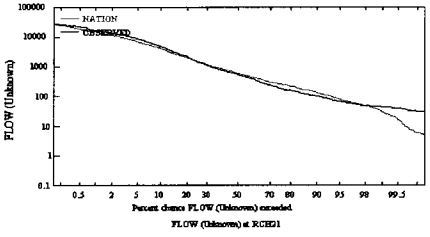


Bias = -46.715	Bias = -1.22E+08	Bias = -2.03E+08	Bias = 7.94E+07	Bias = -2038.154	SNOWCF = 1.273
St.Er. = 2190.105	St.Er. = 2.13E+09	St.Er. = 4.29E+09	St.Er. = 8.03E+08	St.Er. = 7402.917	CCFACT = 1.012
Rel.Bias = -0.026	Rel.Bias = -0.026	Rel.Bias = -0.013	Rel.Bias = 0.112	Rel.Bias = -0.088	LZSN = 2.001
Rel.St.Er. = 0.633	Rel.St.Er. = 0.338	Rel.St.Er. = 0.506	Rel.St.Er. = 1.433	Rel.St.Er. = 1.598	INFILT = 0.029
NS = 0.598	NS = 0.885	NS = 0.743	NS = -1.053	NS = -1.556	AGWRC = 0.971
CE = 0.607	CE = 0.696	CE = 0.516	CE = -0.223	CE = -0.585	DEEPFR = 0.486
IA = 0.800	IA = 0.847	IA = 0.746	IA = 0.409	IA = 0.510	BASETP = 0.110
					AGWETP = 0.011
					CEPSC = 0.250
					UZSN = 1.855
					NSUR = 0.500
					INTFW = 9.900
					IRC = 0.752
					LZETP = 0.275

Case 14

$$\phi = \sum_{i=1}^{n} [1.1 \times (q\ mid^{i}_{obsr} - q\ mid^{i}_{sim})]^2 + \sum_{i=1}^{k} [2 \times (q \max^{i}_{obsr} - q \max^{i}_{sim})]^2 + \sum_{i=1}^{j} [4000 \times \log q \min \times (\log q \min^{i}_{obsr} - \log q \min^{i}_{sim})]^2$$

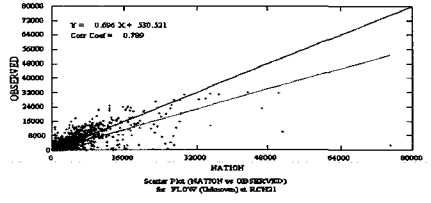


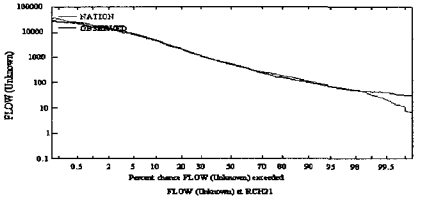


Bias = -166.354	Bias = -4.37E+08	Bias = -1.81E+09	Bias = 1.22E+08	Bias = -5632.351	SNOWCF = 1.258
St.Er. = 1739.221	St.Er. = 2.26E+09	St.Er. = 4.76E+09	St.Er. = 8.25E+08	St.Er. = 7653.375	CCFACT = 1.012
Rel. Bias = -0.093	Rel.Bias = -0.093	Rel.Bias = -0.118	Rel.Bias = 0.173	Rel.Bias = -0.244	LZSN = 2.000
Rel.St.Er. = 0.503	Rel.St.Er. = 0.358	Rel.St.Er. = 0.561	Rel.St.Er. = 1.472	Rel.St.Er. = 1.653	INFILT = 0.085
NS = 0.746	NS = 0.871	NS = 0.684	NS = -1.167	NS = -1.732	AGWRC = 0.970
CE = 0.631	CE = 0.681	CE = 0.438	CE = -0.263	CE = -0.760	DEEPFR = 0.500
IA = 0.803	IA = 0.833	IA = 0.698	IA = 0.394	IA = 0.449	BASETP = 0.200
					AGWETP = 0.070
					CEPSC = 0.250
					UZSN = 2.000
					NSUR = 0.500
					INTFW = 9.900
					IRC = 0.788
					LZETP = 0.122

Maximum flows + Exceedance of minimum flows +Monthly volumes

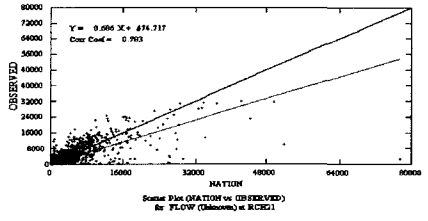
Case 15 $\phi = \sum_{i=1}^{n} [8.5 \times (\log q \max {}^{i}_{obsr} - \log q \max {}^{i}_{sim})]^2 + \sum_{i=1}^{k} [250 \times (E \min {}^{i}_{obsr} - E \min {}^{i}_{sim})]^2 + \sum_{i=1}^{j} [3.5 \times (\log V^{i}_{obsr} - \log V^{i}_{sim})]^2$

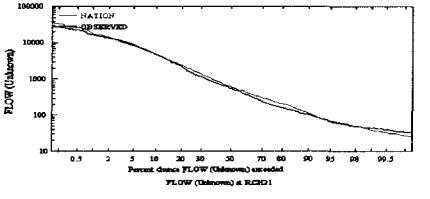




Daily flows	Monthly volumes	March-April	August-September	Top 1% flows	Parameters
Bias = 11.565	Bias = 3.05E+07	Bias = 1.23E+08	Bias = 1.91E+08	Bias = -1184.835	SNOWCF = 1.289
St.Er. = 2435.258	St.Er. = 2.20E+09	St.Er. = 4.49E+09	St.Er. = 8.72E+08	St.Er. = 7963.355	CCFACT = 1.011
Rel.Bias = 0.006	Rel.Bias = 0.006	Rel.Bias = 0.008	Rel.Bias = 0.271	Rel.Bias = -0.051	LZSN = 2.000
Rel.St.Er. = 0.704	Rel.St.Er. = 0.349	Rel.St.Er. = 0.530	Rel.St.Er. = 1.556	Rel.St.Er. = 1.719	INFILT = 0.022
NS = 0.503	NS = 0.877	NS = 0.718	NS = -1.423	NS = -1.958	AGWRC = 0.990
CE = 0.585	CE = 0.697	CE = 0.496	CE = -0.253	CE = -0.662	DEEPFR = 0.500
IA = 0.792	IA = 0.848	IA = 0.738	IA = 0.447	IA = 0.501	BASETP = 0.084
					AGWETP = 0.079
					CEPSC = 0.250
					UZSN = 2.000
					NSUR = 0.500
					INTFW = 9.900
					IRC = 0.746
					LZETP = 0.161

Case 16 $\phi = \sum_{i=1}^{n} [8.5 \times (\log q \max {}^{i}_{obsr} - \log q \max {}^{i}_{sim})]^2 + \sum_{i=1}^{k} [400 \times (E \min {}^{i}_{obsr} - E \min {}^{i}_{sim})]^2 + \sum_{i=1}^{j} [3.5 \times (\log V^{i}_{obsr} - \log V^{i}_{sim})]^2$

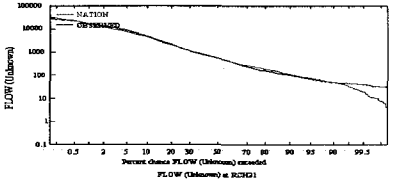
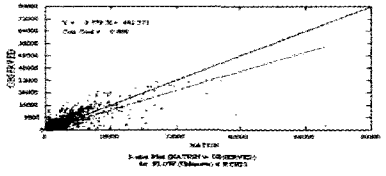




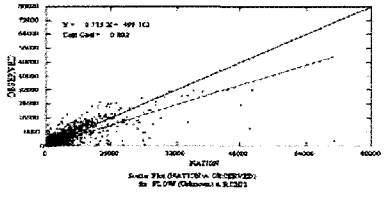
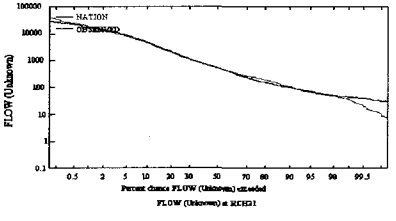
Bias = 11.565	Bias = 3.05E+07	Bias = 1.23E+08	Bias = 1.91E+08	Bias = -1184.840	SNOWCF = 1.273
St.Er. = 2435.258	St.Er. = 2.20E+09	St.Er. = 4.49+09	St.Er. = 8.72E+08	St.Er. = 7963.354	CCFACT = 1.004
Rel.Bias = 0.006	Rel.Bias = 0.006	Rel.Bias = 0.008	Rel.Bias = 0.271	Rel.Bias = -0.051	LZSN = 2.000
Rel.St.Er. = 0.704	Rel.St.Er. = 0.349	Rel.St.Er. = 0.530	Rel.St.Er. = 1.556	Rel.St.Er. = 1.719	INFILT = 0.020
NS = 0.503	NS = 0.877	NS = 0.718	NS = -1.423	NS = -1.958	AGWRC = 0.977
CE = 0.585	CE = 0.697	CE = 0.496	CE = -0.253	CE = -0.662	DEEPFR = 0.500
IA = 0.792	IA = 0.848	IA = 0.738	IA = 0.447	IA = 0.501	BASETP = 0.045
					AGWETP = 0.016
					CEPSC = 0.250
					UZSN = 2.000
					NSUR = 0.500
					INTFW = 9.900
					IRC = 0.766
					LZETP = 0.100

Daily flows + Monthly volumes + Exceedance times

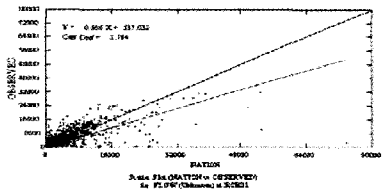
Case 17 $$\phi = \sum_{i=1}^{n}[\log q_{obs}^{i} \times (\log q_{obsr}^{i} - \log q_{sim}^{i})]^2 + \sum_{i=1}^{k}[\frac{2\times10^{-5}}{\sqrt{V_{obs}}} \times (V_{obsr}^{i} - V_{sim}^{i})]^2 + \sum_{i=1}^{j}[150\times(E_{obsr}^{i} - E_{sim}^{i})]^2$$

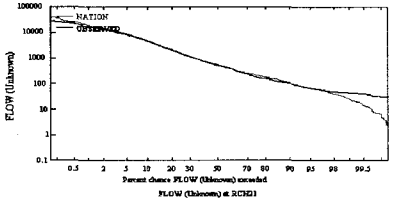
	Daily flows	Monthly volumes	March-April	August-September	Top 1% flows	Parameters
 	Bias = -65.488 St.Er. = 2185.312 Rel. Bias = -0.036 Rel.St.Er. = 0.632 NS = 0.600 CE = 0.604 IA = 0.797					SNOWCF = 1.277 CCFACT = 1.004 LZSN = 2.004 INFILT = 0.026 AGWRC = 0.983 DEEPFR = 0.500 BASETP = 0.200 AGWETP = 0.048 CEPSC = 0.250 UZSN = 2.004 NSUR = 0.500 INTFW = 9.900 IRC = 0.787 LZETP = 0.239

Case 18 $$\phi = \sum_{i=1}^{n}[\log q^{i} \times (\log q_{obsr}^{i} - \log q_{sim}^{i})]^2 + \sum_{i=1}^{k}[2\times\log V_{obs} \times (\log V_{obsr}^{i} - \log V_{sim}^{i})]^2 + \sum_{i=1}^{j}[150\times(E_{obsr}^{i} - E_{sim}^{i})]^2$$

 	Bias = 9.023 St. Er. = 2341.240 Rel. Bias = 0.005 Rel.St.Er. = 0.677 NS = 0.541 CE = 0.594 IA = 0.796					SNOWCF = 1.305 CCFACT = 1.002 LZSN = 2.001 INFILT = 0.0264 AGWRC = 0.984 DEEPFR = 0.500 BASETP = 0.110 AGWETP = 0.060 CEPSC = 0.250 UZSN = 2.000 NSUR = 0.500 INTFW = 9.90 IRC = 0.742 LZETP = 0.168

Case 19 $$\phi = \sum_{i=1}^{n}[\log q^i \times (\log q_{obsr}^i - \log q_{sim}^i)]^2 + \sum_{i=1}^{k}[2.5 \times \log V_{obs} \times (\log V_{obsr}^i - \log V_{sim}^i)]^2 + \sum_{i=1}^{j}[200 \times (E_{obsr}^i - E_{sim}^i)]^2$$

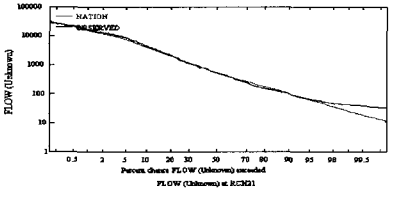


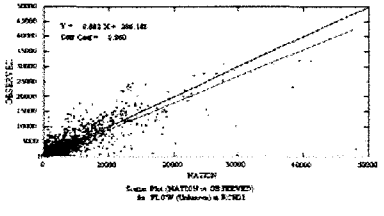


Bias = 28.1591
St.Er. = 2448.942
Rel.Bias = 0.015
Rel.St.Er. = 0.708
NS = 0.498
CE = 0.584
IA = 0.793

SNOWCF = 1.314
CCFACT = 1.001
LZSN = 2.001
INFILT = 0.023
AGWRC = 0.990
DEEPFR = 0.501
BASETP = 0.086
AGWETP = 0.094
CEPSC = 0.250
UZSN = 2.000
NSUR = 0.500
INTFW = 9.900
IRC = 0.738
LZETP = 0.149

Case 20 $$\phi = \sum_{i=1}^{n}[\log q^i \times (\log q_{obsr}^i - \log q_{sim}^i)]^2 + \sum_{i=1}^{k}[\frac{3 \times 10^{-5}}{\sqrt{V_{obs}}} \times (V_{obsr}^i - V_{sim}^i)]^2 + \sum_{i=1}^{j}[180 \times (E_{obsr}^i - E_{sim}^i)]^2$$



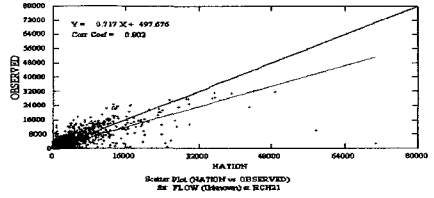


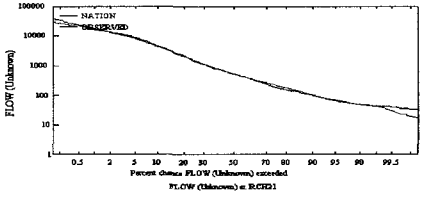
Bias = -87.343
St.Er. = 1810.750
Rel.Bias = -0.049
Rel.St.Er. = 0.523
NS = 0.725
CE = 0.627
IA = 0.808

SNOWCF = 1.316
CCFACT = 1.000
LZSN = 2.004
INFILT = 0.059
AGWRC = 0.928
DEEPFR = 0.436
BASETP = 0.135
AGWETP = 0.01
CEPSC = 0.250
UZSN = 1.963
NSUR = 0.501
INTFW = 9.900
IRC = 0.791
LZETP = 0.312

Case 21

$$\phi = \sum_{i=1}^{n} [\log q^i \times (\log q^i_{obsr} - \log q^i_{sim})]^2 + \sum_{i=1}^{k} [\frac{3 \times 10^{-5}}{\sqrt{V_{obs}}} \times (V^i_{obsr} - V^i_{sim})]^2 + \sum_{i=1}^{j} [300 \times (E^i_{obsr} - E^i_{sim})]^2$$

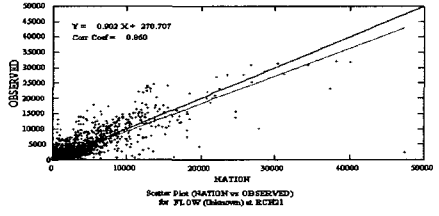


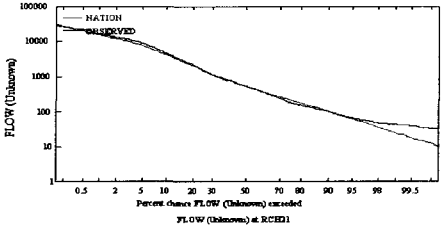


Bias = 5.971	Bias = 1.58E+07	Bias = 7.38E+08	Bias = 3.83E+07	Bias = -1135.798	SNOWCF = 1.362
St.Er. = 2337.337	St.Er. = 2.22E+09	St.Er. = 4.53E+09	St.Er. = 7.99E+08	St.Er. = 7785.492	CCFACT = 1.004
Rel.Bias = 0.003	Rel.Bias = 0.003	Rel.Bias = 4.81E-02	Rel.Bias = 0.054	Rel.Bias = 0.049	LZSN = 2.000
Rel.St.Er. = 0.676	Rel.St.Er. = 0.351	Rel.St.Er. = 0.534	Rel.St.Er. = 1.425	Rel.St.Er. = 1.681	INFILT = 0.022
NS = 0.543	NS = 0.876	NS = 0.713	NS = -1.032	NS = -1.827	GWRC = 0.978
CE = 0.585	CE = 0.687	CE = 0.483	CE = -0.210	CE = -0.691	DEEPFR = 0.500
IA = 0.791	IA = 0.845	IA = 0.744	IA = 0.422	IA = 0.504	BASETP = 0.200
					AGWETP = 0.016
					CEPSC = 0.250
					UZSN = 2.000
					NSUR = 0.500
					INTFW = 9.900
					IRC = 0.796
					LZETP = 0.310

Case 22

$$\phi = \sum_{i=1}^{n} [\log q^i \times (\log q^i_{obsr} - \log q^i_{sim})]^2 + \sum_{i=1}^{k} [\frac{2.9 \times 10^{-5}}{\sqrt{V_{obs}}} \times (V^i_{obsr} - V^i_{sim})]^2 + \sum_{i=1}^{j} [180 \times (E^i_{obsr} - E^i_{sim})]^2$$





Bias = -106.990	Bias = -2.81E+08	Bias = -3.83E+08	Bias = 7.038E+07	Bias = -3605.688	SNOWCF = 1.299
St.Er. = 1795.065	St.Er. = 2.06E+09	St.Er = 4.05E+09	St. Er. = 7.88E+08	St.Er. = 7065.343	CCFACT = 1.000
Rel.Bias = -0.060	Rel.Bias = -0.060	Rel Bias = -0.025	Rel Bias = 0.099	Rel.Bias = -0.156	LZSN = 2.000
Rel.St.Er. = 0.519	Rel.St. Er. = 0.327	Rel St.Er. = 0.478	Rel.St.Er. = 1.406	Rel.St.Er. = 1.526	INFILT = 0.059
NS = 0.730	NS = 0.892	NS = 0.771	NS = -0.979	NS = -1.328	AGWRC = 0.928
IA = 0.628	IA = 0.699	IA = 0.540	IA = -0.190	IA = -0.546	DEEPFR = 0.431
IA = 0.807	IA = 0.849	IA = 0.767	IA = 0.411	IA = 0.514	BASETP = 0.102
					AGWETP = 0.012
					CEPSC = 0.250
					UZSN = 2.000
					NSUR = 0.500
					INTFW = 9.900
					IRC = 0.791
					LZETP = 0.323

References

Bathurst, J.C., Wicks, J.M., and O'Connell, P. E. (1995). Chapter 16: The SHE/SHESED basin scale water flow and sediment transport modeling system, computer models of watershed hydrology, Singh, V.P. ed., Water Resources Publications, Littleton, Colo.

Beven, K.J. (1995). Chapter 18: TOPMODEL, Computer models of watershed hydrology, V. P. Singh, ed., Water Resources Publications, Littleton, Colo.

Beven, K.J., Calver, A., and Morris, E.M. (1987). Institute of hydrology distributed model, Internal Report, Institute of Hydrology, Wallingford.

Bicknell, B.R., Imhoff, J.C., Kittle, J.L., Jobes, T.H., and Donigian, A.S. (2001). Hydrological simulation program HSPF. Version 12. User's Manual, U.S. Environmental Protection Agency, Athens, Georgia.

Bingner, R.L., Theurer, F.D., Cronshey, R.G., and Darden, R.W. (2001). AGNPS 2001 Web Site, Internet at http://www.sedlab.olemiss.edu/AGNPS.html.

Brown, L.C. and Barnwell, T.O. (1987). The enhanced stream water quality model QUAL2E and QUAL2E-UNCAS: Documentation and User Manual, EPA-600/3-87/007, US EPA, Athens, GA, 30605.

Burnash, R.J.C. (1995). The NWS river forecast system – catchment modeling, in Computer Models of Watershed Hydrology, Singh, V.P., ed., Water Resources Publications, Littleton, Colo.

Crawford, N.H. and Linsley, R.K. (1966). Digital simulation in hydrology: Stanford Watershed Model IV, Technical Report No. 39, Dept. of Civil Engineering, Stanford University, Stanford, CA.

Di Luzio, M., Srinivasan, R., Arnold, J.G., and Neitsch, S.L. (2002). Soil and water assessment tool. ArcView GIS Interface Manual: Version 2000, GSWRL Report 02-03, BRC Report 02-07, Texas Water Resources Institute, TR-193, College Station, TX.

Dillaha, T.A., Wolfe, M.L., Shirmohammadi, A., and Byne, F.W. (2001). ANSWERS-2000, in agricultural non-point source water quality models: their use and a application, ed. Parsons, J.E., Thomas, D.L., and Huffman, R.L., Southern Coop. Ser. Bul. 398, 28-45.

Donigian, A.S. and Crawford, N.H. (1977). Simulation of nutrient loadings in surface runoff with the NPS model. Athens, GA: U.S. Environmental Protection Agency; EPA-600/3-77-065.

Fortin, J.P., Moussa, R., Bocquillon, C., and Villeneuve, J.P. (1995). HYDROTEL, un modèle hydrologique distribué pouvant bénéficier des données fournies par la télédétection et les systèmes d'information géographique, Revue des sciences de l'eau, 8, 1, 97-124.

Foster, G.R., Lane, L.J., Nowlin, J.D., Laflen, J.M., and Young, R.A. (1980). A model to estimate sediment yield from field-sized areas: development of model, in CREAMS: a field scale model for chemicals, runoff, and erosion from agricultural management systems, ed. Knisel, W.G., US Dept. of Agric., Sci. and Educ. Admin., Conser., Rep. 26, 36-64.

Huber, W. C. (1995). Chapter 22: EPA storm water management model SWMM, Computer models of watershed hydrology, Singh, V.P. ed., Water Resources Publications, Littleton, Colo.

Knisel, W. G. and Williams, J. R. (1995). Chapter 28: Hydrology components of CREAMS and GLEAMS models, computer models of watershed hydrology, Singh, V.P. ed., Water Resources Publications, Littleton, Colo.

Kouwen, N. (2000). WATFLOOD/SPL: Hydrological model and flood forecasting system, Dept. of Civil Engineering, Univ. of Waterloo, Waterloo, ON.

Quick, M. C. (1995). Chapter 8: The UBC watershed model, computer models of watershed hydrology, Singh, V.P. ed., Water Resources Publications, Littleton, CO.

USACE. (2000). Hydrologic Engineering Center, Hydrologic modeling system HEC-HMS, Technical Reference Manual.

Williams, J. R. (1995). Chapter 24: SWRRB – A watershed scale model for soil and water resources management, Computer models of watershed hydrology, Singh, V.P. ed., Water Resources Publications, Littleton, Colo.

Wool, T.A., Ambrose, R.B., Martin, J.L., and Comer, E.A. (2001). Water Quality Analysis Simulation Program (WASP) Version 6.0. DRAFT: User's Manual, US Environmental Protection Agency, Atlanta, GA.

Zhao, R.J. (1992). The Xinanjiang model applied in China, J. Hydrology, 135, 371-381.