EFFECTS OF UNCERTAINTY IN HYDROLOGIC MODEL CALIBRATION ON EXTREME EVENT SIMULATION

by

ANTHONY DAVID ROCHE

B.A.Sc., University of Waterloo, 2000

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(CIVIL ENGINEERING – HYDROTECHNICAL)

THE UNIVERSITY OF BRITISH COLUMBIA

April 2005

© Anthony David Roche, 2005

Abstract

Computer models representing the hydrologic cycle as a simplified system have become a preferred tool for estimating floods. However, scientific understanding of the uncertainty inherent in these models has not kept pace with their development and application. In many cases it is incorrectly assumed that all uncertainty in model structure, input data, and parameters is minimized or eliminated through calibration. The end result is a ubiquitous but unknowable degree of model predictive uncertainty that may or may not significantly affect the outcome of any given application. Extrapolation of a model beyond its calibration range (i.e., for extreme event simulation) invariably results in a substantial increase in this uncertainty.

This work aims to promote qualitative and quantitative understanding of model predictive uncertainty in extreme event simulation. It therefore begins with a review of the many sources contributing to model predictive uncertainty, an analysis of their origins and interdependencies, and a synthesis of various methods for analyzing uncertainty.

As a pre-requisite step towards the larger goal of reducing overall model predictive uncertainty, this work investigates the variability in estimates of extreme floods (e.g., peak flow, timing, and volume) introduced by subjective decisions made during calibration. Multiple automatic calibrations of a conceptual hydrologic model are conducted using different objective functions to evaluate calibration performance, resulting in a collection of non-inferior parameter sets. Each parameter set is then used to simulate an extreme event based on hydrologic data for the Coquitlam Lake watershed in British Columbia, which is developed for hydropower by BC Hydro. The combined output of these extreme event simulations characterizes the relative variability in the hydrographs.

Simulations are conducted using the University of British Columbia Watershed Model (UBCWM), which is widely used to describe and forecast watershed behaviour in mountainous areas of British Columbia. Calibrations of the UBCWM utilize the Shuffled Complex Evolution Algorithm (SCE-UA), an effective and efficient optimization-based automatic calibration routine. Because automatic calibrations fail to capture the different kinds of expert knowledge

inherent in a manual calibration, extreme event hydrographs obtained using calibrated parameter sets are compared on a relative rather than absolute basis.

Results show that automatic calibration may provide a straightforward method of identifying potential areas where subject models are over-parameterized with respect to the calibration data. More importantly, preliminary results show that the variability is relatively constrained amongst simulations based on a Probable Maximum Flood (PMF) scenario, with coefficient of variation for peak flow, event volume, and time to peak of 4%, 1%, and 1% respectively. This value is negligible in comparison with other uncertainties that dominate extreme events like the PMF. Thus, the PMF-based simulations are relatively insensitive to the different measures of calibration performance used. Similar trials using other models would permit an estimate of the extent to which one could expect to resolve divergent estimates through implementing different but equally valid calibrations.

Observations of this work are applicable for the management of hydropower production and flood control for these watersheds. These observations will provide insights into uncertainty in extreme event simulation and may contribute to the improved management of water, hydropower systems, and public safety in Canada and around the world.

Table of Contents

List of Tables

List of Figures

Acknowledgements

While this thesis is nominally the product of a sole author, it could never have been completed without the advice and support of many colleagues and friends.

Firstly, I would like to thank my supervisor, Dr. Barbara Lence of UBC Civil Engineering. Her input, guidance, and enthusiasm were indispensable in keeping me on course and focussed. She also reviewed this entire document repeatedly throughout its evolution, and for that has captured my admiration as well as my gratitude.

Secondly, I would like to thank my informal thesis committee at BC Hydro – Murray Kroeker, Graham Lang, and Dr. Des Hartford. They helped to set my course and are the "target audience" for whom this thesis is prepared. I would also like to thank Dr. Markus Weiler of UBC Forestry, who provided an experienced hydrologist's review and excellent insight as my official second reader.

Several people assisted with the applied portion of this thesis, and in doing so made my job very much easier. I benefited greatly from the assistance of UBC Professor Emeritus Dr. Michael Quick, principal author of the University of British Columbia Watershed Model, and Edmond Yu, who diligently maintains and upgrades the UBCWM source code. Dr. Zoran Micovic, who puts the "expert" in BC Hydro's "expert calibrations", provided advice on calibration of both the UBCWM and hydrologic models in general.

I salute Dr. Qingyun Duan of the United States National Oceanic and Atmospheric Administration and Drs. Soroosh Sorooshian and Hoshin Gupta of the University of Arizona, who developed, support, and continue to advance the remarkable SCE-UA optimization algorithm used extensively in my work. Dr. Steve Burges of the University of Washington, Dr. Nick Kouwen of the University of Waterloo, and many other industry and research leaders also provided commentary and guidance as I was starting out

Of course, I am eternally grateful to those whose assistance kept me financially solvent during my work. Primary support for this endeavour was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) through a Post-Graduate Scholarship. Additional financial assistance was provided through scholarships and awards from the University of British Columbia, UMA Group Ltd., Mrs. Earl R. Peterson, the Canadian Water Resources Association, and a Research Assistantship with my supervisor, Dr. Barbara Lence of Civil Engineering at UBC. BC Hydro also supported this project through their Professional Partnership Program, providing advice, access to their formidable library of hydrologic studies, and of course the data that forms the basis of the experiment herein.

Finally, I would like to thank my colleagues at UBC, BC Hydro, and Kerr Wood Leidal, my family and friends, and most especially my wife Diane, who have collectively and exhaustively put up with me talking about this thesis for nearly four years. I love you all and am pleased to say that I look forward to moving on to the next big thing.

1. Introduction

"Where there is no water, there is no life... we live by the grace of water."
- National Geographic Special Edition, November 1993

There is no substance more relevant or more necessary to the continuation of life on Earth than water. It is a constant source of sustenance, convenience, and conflict, but rare extremes of either presence or absence can quite easily become a matter of life and death. As such, there should be little surprise that efforts to manage, move, or mitigate the benefits and hazards of water have been chronicled for thousands of years. Perhaps more surprising is the extent of uncertainty that persists in our understanding and prediction of the "when", "where", and "how much" of water, and the lack of tools available to aid in its characterization. Failure to appreciate the implications of such uncertainty can result in spectacular and awesome consequences. This thesis discusses the uncertainty in using computer models to predict watershed response. It specifically examines the influence of uncertainty in calibration objectives on the prediction of floods. Several non-inferior parameter sets are used to predict runoff from an extreme event based on a PMF scenario for the Coquitlam River watershed above Coquitlam Lake.

Floods are the most obvious of adverse consequences for a civilization characterized by lowland, riparian, and coastal habitation. In fact, floods are at once the most common and most devastating of natural disasters. Over two years (1994-1995), floods constituted 50% of global natural disasters, and were responsible for 8,500 casualties (Rosenhagen and Halpert, 1998). Resultant social and economic impacts can and do extend far beyond the directly flooded areas (Gregory et al., 1996); for the period above, total damages were estimated at approximately US $50B (Rosenhagen and Halpert, 1998). Closer to home, 20th century Canadian floods have resulted in several billion dollars in damages and at least 198 fatalities (Natural Resources Canada, 2003).

Protracted periods of low flow (drought), declining lake levels, and falling water tables are another common concern for industry, government, the environment, and the general population.

An accurate assessment of low flows is critical to fish passage, aquatic habitat, irrigation, and water use planning (Pike and Scherer, 2003). Although the primary focus of this work lies in addressing uncertainty in hydrologic modelling of extreme floods, it would be remiss not to mention the possibility for cross-application of many concepts herein for low-flow prediction.

1.1 Guidance for the Reader

Defining the uncertainty surrounding flood management, or even flood magnitude estimation, is far beyond the scope of any single work. This thesis focuses on the technical uncertainties of hydrologic modelling, with an emphasis on the uncertainty incorporated through subjective model calibration. In particular, this work presents a synthesis of the current state of knowledge with regard to hydrologic model predictive uncertainty in the estimation of extreme floods. To illustrate the impact of the concepts discussed, an investigation of how model calibration affects the estimation of extreme flood events is undertaken.

This introductory chapter takes an atypical form with the intent of placing later chapters in context. The chapter outlines some of the various ways that uncertainty can affect the selection of methods for flood estimation and the interpretation of their results. As this chapter is intended to provide the reader with an understanding of the larger context of uncertainty in which hydrologic models are applied, readers seeking a more focused discussion of applied hydrologic modelling may wish to proceed directly to Chapter 2 or 3.

Chapter 2 provides a literature-based review of hydrologic modelling, beginning with a review of hydrologic processes and their potential contributions to uncertainty in modelling. The chapter explores a brief evolutionary history of hydrologic modelling, as well as the basics of model classification and various approaches for model calibration. This chapter is intended as a background for less experienced modellers or those seeking a basic review of uncertainty in hydrologic modelling. Readers comfortable with a comfortable understanding of the fundamentals of hydrologic modelling may wish to proceed directly to Chapter 3.

Chapter 3 provides a more detailed literature-based discussion of the ways in which uncertainty in hydrologic modelling can be classified and explored. This chapter is intended for both beginning and experienced hydrologic modellers seeking to understand the various ways in which uncertainty is introduced into the model output. Those familiar with the body of literature on uncertainty in hydrologic modelling may wish to proceed directly to Chapter 4.

Chapter 4 describes an experiment that applies the University of British Columbia Watershed Model (UBCWM) to quantify the variability introduced in extreme event simulation through subjective assumptions made during calibration. Multiple automatic calibrations of a conceptual hydrologic model are conducted using different measures of performance to provide a collection of non-inferior parameter sets. Each parameter set is then used to simulate an extreme event based on hydrologic data provided by BC Hydro. The combined output of these extreme event simulations characterizes the relative variability in the hydrographs associated with the use of different calibration objectives.

Chapter 5 discusses the results of the experiment outlined in Chapter 4. Chapter 6 provides conclusions, and Chapter 7 gives recommendations for future work. These chapters will likely be of interest to all readers. Chapter 8 provides a glossary of commonly-used abbreviations for ease of reference.

1.2 Approaches for Estimating floods

The floods causing greatest harm are those associated with high magnitude and poor predictability. Extreme examples of this condition include debris flows and floods in steep terrain, which entrain rock fragments, earth, and air into the flood flow (Horton, 1999). Similarly, uncontrolled releases from natural (e.g., moraine-dammed) or anthropogenic reservoirs through failure of a containment structure are commonly associated with significant damages and fatalities. Large, unpredictable floods can also result from blockages of the downstream channel (e.g., ice, log or debris jams), and backwaters from other channels concurrently in flood can restrict flow and inundate significant upstream areas. All of these cases are difficult (if not impossible) to predict effectively, and all are generally beyond the scope of a watershed-scale hydrologic model. It is appropriate to commence a discussion of how to select an approach for flood estimation with a reminder of the insufficiency of hydrologic models in characterizing the complete flood risk for certain situations. Nonetheless, with rare exceptions, even these most extreme cases are initiated by a more predictable, purely hydrologic, high-flow event.

Almost every flood scenario requires heavy runoff from upstream areas. Therefore, the greatest part of scientific effort in flood management has been directed toward quantifying the effects of and relationship between precipitation and runoff. Many approaches have emerged for estimating the magnitude of a given hypothetical flood flow (e.g., a design event). Generic approximations such as the rational method or regionally-derived drainage area - discharge curves are generally acceptable in the absence of detailed data or as checks on more detailed calculations for small projects. Perhaps the simplest method for estimating flood quantiles involves examining historical events and inferring future events based on those observed in the past (e.g., flood frequency analysis or paleoflood hydrology). More complex methods use computer applications to represent one or more phases of the hydrologic cycle as a simplified system. Wherever possible, more than one method of estimating the design event should be used (National Research Council Canada, 1989). Regardless of the tool(s) implemented, it is crucial to recognize that even the most complex tools are only approximations of the natural system (McCuen, 1973). Each approach is naturally subject to a unique set of uncertainties.

Although any approach to estimating runoff may be loosely classified as a "hydrologic model", the neo-classical definition of hydrologic modelling involves the use of a computer model to predict flows under given input conditions. For this reason, a distinction is necessary between two different concepts of a "model". A fundamental hydrodynamic model or concept based on physical principles (e.g., kinematic or diffusive wave equation) is verifiable and repeatable in appropriate controlled experiments. A computer model effectively builds on more fundamental models, applying them to a complex reality with multiple assumptions and potential coding errors (Smith et al., 1994).

There are two established fundamental philosophies for defining a flood magnitude. Prior to the 1970s, the majority of floods computed for engineering design and analysis were defined probabilistically, i.e., an appropriate frequency for the design event was selected or mandated, and the magnitude defined as a co-requisite value (Berga, 1998). The most well-known example of this approach is flood frequency analysis. Klemeš (2000b) argues strongly against the hydrologist's longstanding dependence on flood frequency analysis to provide frequency and magnitude for an extreme event. Much-discussed problems with flood frequency analysis include the impact of plotting position and the applicability of the Independent and Identically

Distributed Random Variable (IIDRV) concept. Perhaps most importantly, a series of N data cannot reasonably be expected to provide reliable information about probabilities less than approximately $1/N$.

Subsequent development of mathematical and computer models has contributed to the emergence of a deterministic approach for estimating potential flood conditions. A deterministic analysis generates an estimate by combining a given set of initial conditions and modelling the resultant hydrologic response. The typical approach involves combining observed extreme hydrological factors using a hydrologic model to obtain a worst-case scenario. However, this *a priori* specification of input conditions has engendered much debate concerning just how extreme an event should be considered.

Since the 1950s, dam and spillway structures with severe consequences of failure have generally adopted the "probable maximum flood" (PMF) as a design criteria (Graham, 2000). The definition of the PMF became something of a holy grail for deterministic modellers, especially in the realm of dam safety. In rainfall-driven watersheds, a PMF event would be driven by the Probable Maximum Precipitation (PMP), defined as the greatest depth of precipitation theoretically possible for a given location, areal extent, and season (Hansen et al., 1988). Input conditions for watersheds that experience significant snowmelt runoff are somewhat more convoluted. In many snowmelt-dominated areas, the maximum flood arises from some critical combination of snowmelt (as a function of accumulated snowpack and temperature sequence) and heavy precipitation (possibly occurring as rain-on-snow).

The application of the PMF as a design event has recently begun to be questioned. Some scientists believe that the Probable Maximum Flood as defined above is too vague to allow for its generic use as an approach for engineering design and analysis. There is no current set of standardized procedures for calculating the PMF, nor is there a method for quantifying how probable the flood really is. Additionally, there is a dearth of information concerning behaviour of watersheds under flooding of PMF proportions. Faulkner notes that historic floods of record for some rivers are initiated by different processes than those presumed to cause the PMF (Faulkner, 2003). Several prominent researchers have noted that the pursuit of quantification ("how to route") should be superseded by attempts to gain a better understanding of the processes involved in extreme event runoff ("what to route") (Burges, 2003; Córdova and Rodríguez-Iturbe, 1983).

There is, therefore, significant uncertainty involved in selecting how any design flood should be defined. In fields such as structural engineering, government or other regulatory bodies provide guidance when calculating appropriate design conditions (e.g., load combinations); however, regulatory guidance on design flood selection is relatively limited. Since the 1970s, the move away from probabilistic flood estimation has resulted in a general shift away from prescriptive regulatory governance of flood selection (Berga, 1998). In the field of flood estimation, definitive and inflexible standards are of questionable value. In some cases, the design conditions suggested by prescriptive standards or guidelines do not include the critical case (Dumont and Dubé, 2003). One possible alternative is presented in the guidelines published by the Canadian Dam Association in 1999 for use as best management practices (CDA, 1999). These guidelines link the selection of design floods for dams to the consequences of failure based on loss of life, environmental damage, and financial loss. Although the guidelines represent current best practice in the industry, they still lack quantitative guidance for selecting an appropriate flood condition when the consequences of failure are neither severe nor negligible. Recent thought challenges the common practice of using overly-conservative "worst-case" scenarios in such cases as inconsistent, unwarranted, and philosophically pessimistic (Lohani et al., 1997). However, in the absence of an alternative with a strong legal precedent, their use is likely to continue.

1.3 Presenting Floods for Decision-making

Most computer-based hydrologic models present a single, deterministic estimate for a flood situation with each application. However, such a simple representation of a flood is often not sufficient for effective decision-making, especially where significant safety, financial, or environmental factors are involved. The problems with simple representation are obvious: if the flood has been over-estimated, inefficient or unnecessary designs translate into higher costs; if the flood is underestimated, an unsafe design results. Between these two extremes lies a range of values that would be "acceptable" if design conditions were known with certainty.

In most cases we have at best a limited understanding of where on the safety-vs.-efficiency continuum any given deterministic estimate lies. The presentation of "representative", "average", or "best" results without a description of their associated uncertainties gives an illusion of precision and objectivity, especially to those not familiar with the approaches or tools used in the estimate (Keeney and Winterfeldt, 1989). The literature review in Chapters 2 and 3 is concerned with identifying the various mechanisms that render precision and objectivity unattainable in hydrologic modelling, at least at present.

Those presenting flood estimates for analysis must be conscious of the growing involvement of the public in the decision-making process. The portrayal of flood-mitigation projects as preventing all or virtually all floods (i.e., by designing for the PMF or other extremely rare events) has had a polarizing effect. Typically, people are left either unaware of the potential for failure or skeptical of the experts' analysis and design (Linsley et al., 1992; Slovic, 1992). The discussion of low-probability events in the absence of numerical data is particularly difficult for the public to interpret, as qualitative concepts like "a small chance of being exceeded" can have large ranges of interpretation (Keeney and Winterfeldt, 1989).

Difficulties of presenting the results of a hydrologic simulation for discussion or decision are magnified when the impact of that information is unclear. Uncertainty and disagreement can arise in translating flows into water levels, failure probabilities for hydraulic structures, or consequences for aquatic resources (e.g., Caissie and El-Jabi, 2003). This is complicated when different stakeholders have different viewpoints and objectives (Gregory et al., 1996). Mutually-agreeable objectives for identifying, measuring, and understanding impacts are a pre-requisite for any group-oriented analysis of technical issues (Fiering, 1976).

Where non-experts must interpret uncertain flood risks, biases and beliefs can play a significant role in arguments and decisions. A variety of personal characteristics (e.g., personality, opinions, values, economic or cultural context) and situational factors (e.g., voluntariness of exposure, familiarity, control) have been noted to influence the relation between perceived risk, perceived benefit, and risk acceptance; few of these factors are explicitly quantifiable (Gregory et al., 1996; Slovic, 1987; Slovic, 1992). Therefore, experts have noted that the public assigns technically unsubstantiated perceptions of risk to certain situations (Gregory et al., 1996).

Experts must understand the impacts of any biases that stakeholders may have, because this can potentially have as much impact on the quality of a decision as the technical uncertainties.

Further difficulties can arise if non-expert stakeholders are required to use intuition to interpret statistical information. Even experienced researchers may avoid simple problems (e.g., the gambler's fallacy) while falling prey to the same biases under intuitive judgment of more intricate and less transparent problems (Tversky and Kahneman, 1974).

1.4 Expressing Uncertainty

Since uncertainty and risk are related at a fundamental level, a decision as to what is acceptable should address both the degree of residual risk that is "acceptable" to stakeholders and the uncertainty surrounding the risk analysis. Figure 1-1 illustrates why this is necessary; it shows the relationship between the risk and uncertainty, where risk is defined as the product of frequency and consequence. Note that each curve represents an equally valid but distinct interpretation of risk. Each curve is uniquely defined by its level of (un)certainty, expressed in this case as confidence level. The figure implies that any arbitrary frequency will be associated with a range of possible consequences, each with a different co-requisite level of uncertainty; likewise, each consequence will have a range of possible frequencies. Therefore, it is impossible to explicitly define an acceptable level of risk without at least implicitly defining a corresponding level of acceptable uncertainty.

One of the few less-explored areas for policy development in hydrology and other risk-related fields is the determination of what constitutes "acceptable uncertainty". Precedents exist for the selection of a single threshold value; for example, transportation engineering regularly implements projects designed to meet safety standards for a specific percentile of drivers. A more direct example is given by the United Kingdom Health and Safety Executive (HSE, 2001) through their adoption of 10^{-4} as an appropriate boundary between tolerable and unacceptable risk for situations in which the risk is imposed or involuntary.

As shown in Figure 1-1, different levels of uncertainty require different levels of protection and may therefore drive different choices among alternatives (NRC, 2000a). Projects having marginal or indeterminate benefit-cost ratios can be the most sensitive to uncertainty, as the

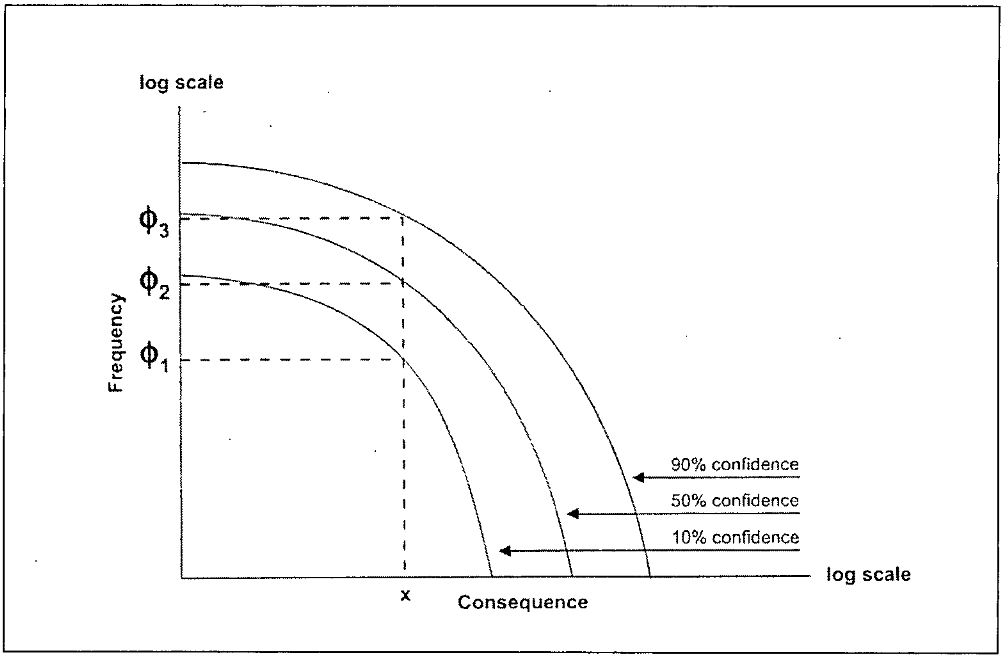


Figure 1-1: Risk – Uncertainty Interaction
(from p. 5-20, Lohani et al., 1997)

uncertainty may be sufficient to justify or deter their approval (NRC, 1995). For example, a small uncertainty in the conditions leading to failure of a levee can make a large difference in the overall system reliability and therefore can influence decisions pertaining to the project as a whole (NRC, 2000b). In many cases, it is reasonable to expect that the available information is insufficient to even begin to properly address cost-benefit studies. This is an undesirable state of affairs from both fiscal analysis and public safety perspectives.

For the various reasons above, an expression of the associated uncertainty should be required in all risk analyses (NRC, 1995). Too often results are limited to values calculated with the "best" estimates or by averaging over the final probability distribution. Even where uncertainties are acknowledged and accounted for, answers provided by scientists are often divorced from their uncertainty as they are passed "up the tree" to stakeholders or decision-makers (Grayson et al., 1992b).

The most thorough approach to expressing uncertainty involves presenting a full description of the uncertain results to decision-makers (e.g., a cumulative distribution showing the continuum of event magnitude and probability). Although this can sometimes be difficult to interpret, this full set of results effectively allows the policy decision (i.e., what constitutes acceptable risk) to be separated from the technical analysis. Alternatively, Lohani et al. (Lohani et al., 1997) argue for presenting mean values to describe magnitude while including low probability-high consequence information wherever the impacts warrant.

In cases where extensive quantitative information is not available, the use of successive approximations can sometimes provide acceptable bounds for the quantity of interest (Keeney and Winterfeldt, 1989). Examples include the limiting frequency of 10^{-6} implicit in the State of Washington's definition of a design flood for dams or the conceptual procedure for progressive refinement of "Ultimate Limit State" estimates (Faulkner, 2003; Hartford et al., 2001).

The above discussions have illustrated some of the ways that uncertainty can manifest in hydrologic analysis through subjective choices made in approach, event selection, acceptable risk, biases, and the inclusion of uncertainty in an analysis. However, the process of decision-making does not occur in a static environment, and the potential for change is a major source of uncertainty. Changes in policy, knowledge, technology, or potential consequences can all create a "moving target" effect when attempting to address floods and their impacts (Faulkner, 2003). In most cases, the factors are interdependent. For example, Jarrett (1990) used paleohydrologic evidence to show that large floods observed at higher elevations in Colorado are likely due to debris flows rather than intense high-elevation precipitation. This change in knowledge affects potential consequences through a certain but indeterminate increase in the expected frequency of flows of similar magnitude. The resulting increase in uncertainty might cause a policy shift in the region, leading to the implementation of additional structural or non-structural protective measures.

1.5 Model Predictive Uncertainty in Context

The broad-based nature of the hydrologic system leads Grayson et al. (1992b) to refer to the study of hydrology as "transscientific" - implying the pursuit of answers to questions asked of science which cannot be answered by science. The discussions of the preceding sections do not address the many ways that uncertainty can manifest itself quantitatively within the actual process of modelling - the "scientific" portion of any flood study. These more technical uncertainties must be combined with value judgments, biases, and other non-technical uncertainties. The ultimate goal is to define the role that uncertainty plays, and identify how it interacts with resulting risks, options, and decisions for any given situation.

Uncertainty is also a necessary consideration for hydrologic modelling itself (O'Connell and Todini, 1996). In addition to the caveats applicable to computer modelling in any field (e.g., garbage in – garbage out), hydrologic models create additional challenges. Most importantly, the results generated by computer simulations fuse together diverse types of uncertainty, such as those arising from natural processes and those related to mathematical representation.

Aleatory uncertainty, or natural variability, represents the variability of the physical world on the assumption that this variability cannot be mitigated. Conversely, the U.S. National Research Council (p. 41, NRC, 2000a) relates the concept of epistemic or knowledge-based uncertainty to "a lack of understanding of events and processes, or a lack of data from which to draw inferences". Knowledge-based uncertainty can be reduced under the correct conditions. Since most hydrologic models do not explicitly account for uncertainty in any form, an understanding of the magnitude and relationship of these two components is crucial. Understanding the uncertainty of the applied models and processes is an obvious prerequisite for assessing the uncertainty in any flood-related decision (NRC, 1995).

Within the category of epistemic uncertainty, potential sources for uncertainty include model structure, mathematical representation, model parameter values, and input data errors (Lei and Schilling, 1996). These different types of uncertainty in hydrologic modelling are often lumped together in an applied hydrological context to create an aggregate "model predictive uncertainty" (NRC, 1995). Any of the types of uncertainty listed above has the potential to influence results, and therefore all should be addressed in any uncertainty analysis (Vicens et al., 1975).

It has often been assumed that model calibration can resolve and reduce all components of model predictive uncertainty. However, calculating model predictive uncertainty with techniques such as sensitivity analysis is implicitly dependent on calibration; it is of only limited value unless the model and data errors are known to be insignificant (Lei and Schilling, 1996). Past analyses of "model predictive uncertainty" for hydrologic models have shown a wide range of variation in model accuracy. The accuracies of rainfall-runoff simulations reported in the literature are typically influenced by factors such as those presented by Michaud and Sorooshian (1994):

- the inclusion or exclusion of model calibration;
- the inclusion or exclusion of split-sample validation;
- the number, variety, and climatology of storms examined;
- different understandings of "good" and "poor" simulations;
- benchmark data reliability;
- the context of results (e.g., real-time forecasting, single historic results, or multiple peak flows from a set of historical storms);
- runoff dynamics (e.g., initial conditions, dominant processes); and
- model assumptions, parameter values, and spatial resolution.

Attempting to address many of the most common problems in the present era of high-powered digital computers has somewhat predictably led to the development of some immensely complex computational models of hydrology. However, theoretical rigour does not in and of itself limit uncertainty, and can imply a degree of accuracy that may not exist (Grayson et al., 1992b). Rather than increasing model complexity, progress in reducing model predictive uncertainty will likely depend on the establishment of a new paradigm that includes an acceptance of uncertainty in the results (Beven, 2002). A best practice approach is needed to allow professionals to move on to defining appropriate principles rather than arguing about how the impacts of uncertainty should be addressed (Faulkner, 2003).

2. Hydrologic Modelling

> *"Models are like maps: never final, never complete until they grow as large and complex as the reality they represent."*
>
> *- James Gleick, from "Genius: The Life and Science of Richard Feynman"*

This chapter is intended as background for less experienced modellers or for those seeking a basic review of the fundamentals of hydrologic modelling. Readers seeking a more advanced and focussed discussion of uncertainty may wish to proceed directly to Chapter 3.

A solid understanding of processes represented in hydrologic modelling is required for any discussion of model predictive uncertainty. Therefore, this chapter begins with a summary of various processes important to modelling. Although the concepts are basic, the novice modeller is encouraged to consider the discussion in terms of the potential uncertainty inherent in modelling the more complex aspects of the system. Those familiar with the complexity of these processes and the related simplifications and assumptions implicit in different hydrologic models will likely wish to proceed directly to Section 2.2. Section 2.2 provides the reader with insight into the approach and philosophy of hydrologic model development and application. In Section 2.3, a brief overview of the various kinds of models addresses their advantages and disadvantages. A detailed discussion of automatic and manual calibration follows in Section 2.4, highlighting the need for considering uncertainty in the calibration process. In particular, Section 2.4 introduces the Shuffled Complex Evolution method developed at the University of Arizona (SCE-UA), which is utilized in the quantitative experiment outlined in Chapter 4. This chapter concludes with a discussion of some of the factors limiting progress in hydrologic modelling, which leads into the discussion of uncertainty in Chapter 3.

2.1 Hydrologic Processes

2.1.1 Scale in Hydrology

In nature, scales of things are not arbitrary but tend to concentrate around discrete states as a function of their material substance and of the balance between the interacting forces (Klemeš, 1983). Scientific progress has typically been slower in disciplines attempting to work between dominant scale levels than in those working within a single scale (ibid.). Hydrology is a classic example; component processes can be active at many different spatial and temporal scales from minute to global. Klemeš (ibid.) generalizes the "characteristic" scale of hydrology as between 1 and 1000 km^2 in space and 100 seconds to 100 years in time. However, such generalizations serve only in philosophical discussions; the practicing hydrologic modeller must understand what is occurring at all scales to avoid making irresponsible simplifications. According to Song and James (1992), hydrologic processes can be characterized at five typical scales, including the following:

- laboratory scale – typically less than 10m, for describing detailed physics of water-surface interactions (e.g., infiltration) or subsurface processes;
- plot or hillslope scale – typically tens of metres, for describing runoff processes;
- catchment scale – typically hundreds to thousands of metres, for characterizing the interaction of various hillslopes feeding into a single channel;
- basin or watershed scale – typically tens to thousands of kilometres, for characterizing the generation, storage and translation routing of a channel network or river system; and
- continental or global scale – thousands of kilometres and greater, for characterizing the atmospheric processes that drive the hydrologic cycle.

Commonalities of vegetation and land use at each scale typically have associated commonalities of underlying hydrological mechanisms or behaviours (e.g., alpine vs. sub-alpine at the catchment scale; tropical vs. temperate at the basin scale) (Singh, 1995b). Viessman and Lewis

(1996) note that it is easiest to deal with hydrology at the watershed or river basin scale due to the relatively sharp boundaries of the runoff system. It is commonly accepted, however, that no adjustment of scale can place such well-defined boundaries on the other components of the water balance; the hydrologic cycle is a closed system only at the global scale (Beven, 2000; Klemeš, 1983; Viessman and Lewis, 1996).

2.1.2 Precipitation

Precipitation can take many forms, but the two most common (i.e., rain and snow) are of greatest import for hydrologic modelling. The main difference from a hydrological perspective is the delayed runoff response associated with snow. This difference is very important from the standpoint of hydrologic modelling, as rain-dominated basins promote a focus on the accurate capture of individual events while simulation of snowmelt-dominated basins is more concerned with seasonal precipitation totals (Micovic, 2003a). Other forms of precipitation (e.g., sleet, hail, graupel) are less common; their significance to hydrologic modelling is therefore limited.

There are three primary categories of precipitation events: convective, orographic, and cyclonic or frontal (Viessman and Lewis, 1996). Convective precipitation arises when moist air heated near the terrestrial interface rises and cools, and typically results in short-term high-intensity local precipitation in the area of the updraft (AMS, 2000; Horton, 1999). Orographic precipitation results from the lifting of moist air masses over natural barriers such as ridges or mountain ranges, and is controlled by barrier slope, height of barrier, and air mass stability (Quick, 1995; Viessman and Lewis, 1996). Frontal precipitation results from the interaction of two air masses of different density, almost invariably segregated by temperature (AMS, 2000). The term "frontal precipitation" is most significant in its sense of distinction from convective and orographic precipitation (Viessman and Lewis, 1996).

The different mechanisms of precipitation generation are distinct in their behaviour and should be modelled as such wherever possible, since factors such as precipitation and timing can often control the flood response of a basin (Konrad, 2001). For example, the various precipitation events that led to the Mississippi River flood of 1993 were not among the most extreme events of record at any spatial scale. Stationary or slow-moving storm systems can also lead to flood conditions (e.g., Spring Creek, Colorado in Ogden et al. (2000) and Kickapoo Creek, Texas in

Smith et al. (2000)). Successively increasing peaks of rainfall from a storm moving in the downstream direction represent the worst case scenario from a hydraulic standpoint. In this case, subsequent waves of runoff can propagate and overtake preceding waves (Ogden et al., 2000; Thapa and Khanal, 2001).

Since volume, timing, and distribution of precipitation are the most significant factors in determining flood magnitude, it is no surprise that techniques for the measurement of precipitation are well developed. The two main sources for rain data are gauge networks and radar measurement.

Recording precipitation gauges are the most common form of data used in hydrologic research, providing continuous point estimates of precipitation at a specific location (Duchon and Essenberg, 2001). The two most commonly used classes of recording gauges are tipping bucket and weighing gauges.

A tipping bucket gauge involves a small, bi-stable bucket having two chambers, each with a volume equivalent to a fraction of millimeter of rain. The presence of water in one side of the bucket will cause it to tip to that side, spilling the full chamber and aligning the empty chamber in position to collect precipitation. A datalogger records the number of tips within a specified period.

A weighing gauge records continuous or periodic mechanical measurements of the cumulative weight of precipitation by using a roll plot or electronic means, and therefore requires periodic calibration in the field (Duchon and Essenberg, 2001). For a more complete discussion of automatic gauges, the reader is referred to Nystuen et al. (1996).

Non-recording rain gauges are much more straightforward but do not allow the user to estimate rainfall rate or intensity. Regardless, such data have proven useful to hydrologic modellers in the past (e.g., Faurès et al., 1995). Non-recording rain gauges typically consist of an amplified collector which is measured manually against a fixed or removable scale. Their simplicity has led to the development of an extensive network of amateur meteorological stations across the United States (National Weather Service, 2003).

The expansion of Doppler radar across North America provides an alternative form of precipitation measurement. While recording rain gauges capture temporal variability for a given location, radar rainfall estimates can characterize spatial variability for a given series of temporal snapshots. Radar can provide three-dimensional observations over thousands of square kilometres. The physics involved in radar measurement limit its accuracy at greater distances, and usually an expert must review results prior to use to avoid incorporation of anomalies into the data set.

Solid precipitation (i.e., snow) is usually measured with a heated tipping bucket gauge or a weighing gauge treated with antifreeze to melt the snow on contact. While much less common, radar can also be used to estimate snowfall. Collier and Larke (1978) find that radar measurements can have as little as 13% error when compared to gauge-measured data. More generally, however, radar estimates of snowfall have not been very successful (Xiao et al., 1998). Studies involving radar and ground-based snow measurements should be interpreted with care, as radar estimation relationships can be dependent on diverse factors such as range, location, temperature, snowfall type, and season (ibid.; Hunter et al., 2001). In most practical cases, error is likely to be at least several times that reported by Collier and Larke (1978), and can easily exceed a factor of two (Krajewski, 2005).

Sonic snow depth sensors and snow pillows are also commonly used to record variations in snow depth and snow water equivalent over time. Manual snow depth sampling (i.e., snow courses) are routinely performed for the purposes of estimating the spring freshet. In a less traditional context, Matthews (1999) demonstrates how remote sensing can be applied to determine the snow-covered area for a basin and thus provide insight into snowpack generation and depletion.

There are two general approaches for estimating snowmelt: energy balance methods and index methods (Maidment, 1993). The physically-based energy balance method applies continuity principles to the various energy fluxes of a watershed (Viessman and Lewis, 1996). The main drawback of the energy balance method is that it requires significant amounts collected from within the basin (e.g., radiation, wind, vapour pressure).

The various index-based methods for estimating snowmelt use calibrated parameters and index variables to estimate snowmelt on a catchment-wide basis (Matthews, 1999). Index methods are generally less accurate but easier to apply, requiring only one or two data series for their index variable(s). Air temperature is the most commonly used variable due to its wide availability and strong correlation to snowmelt processes. The World Meteorological Organization's 1986 comparison of snowmelt models concludes that the choice of model usually depends on the intent of the application and the nature of the problem and hydrologic regime (WMO, 1986).

2.1.3 Runoff

A number of processes comprise the transition of precipitation to streamflow. A complete taxonomy of the processes active in any one catchment would show that over time, the composition of outflow by source changes depending on which processes are most active in a volumetric sense. In most areas, the *in situ* relative contributions cannot be measured directly; however, quantitative estimates are often possible. In cases where source chemistry can be identified for each mechanism, the utilization of chemical data and tracers allows researchers to explore quantitative estimates for the proportional origin of runoff waters (Hornberger and Boyer, 1995).

For decades, studies have attempted to characterize the properties of various runoff responses. For example, Dunne (1982) provides estimates of typical velocities for overland, subsurface, and channel flows. This section briefly describes the mechanics of and approaches for estimating the translation of precipitation into runoff.

Precipitation may be intercepted by vegetation or collect in minor depressions in the ground surface. These two intermediate processes are both sinks for precipitation, but neither contributes directly to runoff; most of the water detained by these processes either infiltrates into the ground or evaporates after the rain stops (Pike and Scherer, 2003). In both cases, even the most detailed hydrologic models approximate the processes on an areal basis, usually using empirical parameters. In particular, seasonal variations in interception are often not considered.

Infiltration refers to the process by which water moves from the surface into the soil matrix. Infiltration is constantly active in all wetted permeable areas except those from which active seepage is occurring. The rate of infiltration is a function of soil moisture and condition as well as soil type. Coarse soils, well-vegetated land, low soil moisture, and a macro-porous top layer (i.e., affected by burrowing insects and animals) all promote high infiltration rates (Fetter, 1994).

Initial soil moisture at a given location varies over time; as a dependent condition, the limiting infiltration rate does likewise. *In situ* soil moisture can be assessed using a variety of apparatus (e.g., tensiometers or TDR (Time-Domain Reflectometry) probes), and infiltration rates for soil samples under different moisture conditions can be measured in a laboratory. However, many factors can influence infiltration at a local level (e.g., leaf litter and local topography). This renders a detailed characterization of the infiltration process complicated even for ideal conditions (Viessman and Lewis, 1996).

Infiltrated water generally percolates vertically downward through the unsaturated zone until meeting the water table. However, it is common – especially where the catchment is dominated by steep hillslopes – to have horizontal flow above the nominal saturated zone, a process referred to as interflow. Preferential flow occurs through root voids and other macropores in the soil matrix under saturated or unsaturated conditions. However, interflow most commonly occurs on steep hillslopes when percolating water encounters a zone of lower vertical hydraulic conductivity (Fetter, 1994; Refsgaard and Storm, 1995). The resulting perched water table often triggers lateral subsurface flow, either through lower-permeability strata or soil macropores (Weiler et al., 2005). Thus, interflow can occur under saturated or near-saturated conditions even though the response occurs above the nominal water table (ibid.). This is an important process for delivering water to the valley bottom at the hillslope scale (Weiler and McDonnell, 2004).

Interflow is difficult to quantify due to its transient nature. An effective numerical description requires knowledge of subsurface strata topography, hydraulic conductivity, macroporosity, and connectivity, as well as antecedent soil moisture and groundwater conditions. The scientific understanding of interflow at the hillslope scale has evolved considerably over the past few decades through the application of new measurement techniques like isotope and chemical tracing (Weiler et al., 2005). Although the prevailing understanding of interflow has advanced, interflow is not well represented in most hydrologic models. Even relatively sophisticated hydrologic models like MIKE SHE (DHI Software's popular version of the Système Hydrologique Européen) calculate only vertical flow in the unsaturated zone (Refsgaard and Storm, 1995). In some cases (e.g., bulk hydraulic conductivity), the soil properties governing interflow (e.g., bulk hydraulic conductivity, including macropores) cannot be accurately assessed, since the act of sampling often changes the property in question (Beven, 2002).

Saturated-zone groundwater processes are comparatively well understood but in many cases just as difficult to quantify. Groundwater baseflow sustains surface water systems through dry periods by slow depletion of subsurface storage. Volumes and flow rates for groundwater are difficult to characterize because of their strong dependence on regional geology. Further complexity arises from the three-dimensional nature of subsurface flow. Although complex, groundwater can be modelled in three dimensions (e.g., MIKE SHE in Refsgaard and Storm, 1995). However, many models assume flow in the third dimension to be negligible, allowing for two-dimensional analysis within the saturated zone (Viessman and Lewis, 1996).

Groundwater measurements typically involve recording the depth to which an unrestricted column of water will rise at a number of individual locations. A series of such measurements can determine flow direction but cannot fully describe subsurface flow without additional information on hydraulic conductivity.

Typically, groundwater flow calculations are based on equations developed for a simple control volumes and benchmarked to the subject catchment using empirical parameters derived from field measurements. Regardless of the quality of the parameter estimation, equations derived on the basis of a homogeneous, isotropic control volume are often inapplicable at large scales due to heterogeneity and preferential flow pathways (Beven, 2002).

Modern hydrologic literature divides flow over the ground surface into two distinct mechanisms. The dominant mechanism for a given region depends on climate, topography, ground permeability, and rainfall intensity.

Horton or "infiltration excess" Overland Flow (HOF), named for hydrologic pioneer R. E. Horton, occurs when the precipitation rate exceeds the infiltration rate. Excess precipitation runs over the surface until it either infiltrates elsewhere or reaches an area of surface accumulation (e.g., depression, lake, stream). HOF can occur as a result of extremely intense precipitation or moderately-intense precipitation falling on low-permeability surface layers such as exposed bedrock, frozen ground, asphalt, or extremely dry soil. HOF is usually the dominant form of overland flow observed in arid environments due to the typical combination of a dry, low-permeability surface layer and infrequent but intense precipitation events.

Saturation Overland Flow (SOF) occurs when a precipitation event causes the water table to rise to the surface (e.g., at the base of a hillside), creating a seepage face (Dunne, 1982). The term saturation overland flow refers to the combination of direct precipitation on saturated soil and flow emerging from an ephemeral seepage face. SOF is more commonly observed than HOF in wet and temperate regions due to its dependence on topography rather than intense precipitation and low permeability. SOF can be observed along creeks and stream banks during most higher-intensity precipitation events. Flooding due to SOF is typically associated with prolonged periods of rain that substantially raise local water tables.

The volume of surface flow is typically calculated by subtracting estimated losses and infiltration from estimates of areal totals for rain and snowmelt (Michaud and Sorooshian, 1994). The actual mechanics of surface flow would be nearly impossible to simulate in detail, necessitating the use of substantial approximations. Where detailed simulation of overland flow is required, the most widely-applied approach is to use the St. Venant equations to route flow as a thin sheet of water moving over a homogenous landscape of constant or smoothly-varying roughness. More complex computations are used in leading-edge computer software such as DHI's MIKE SHE, which applies the Saint Venant equations in two horizontal dimensions (Refsgaard and Storm, 1995). These methods implicitly assume that surface runoff remains in the form of characteristically two-dimensional sheets rather than considering well documented but more complex real-world behaviour (e.g., rilling and backwatering around uneven micro-terrain features). These assumptions may ultimately provide the correct answer for timing and volume of water reaching the channel, but bear little resemblance to the *in situ* process on a hillslope (Burges, 2002).

All runoff eventually makes its way to the channel network of streams, rivers, and lakes. The acceptable representation of these surface water systems in most models arises from a fairly thorough understanding of open-channel hydraulics. However, many routing approaches still rely on empirical and subjective methods of approximation that are applicable only under specific conditions (e.g., Manning's formula). This suggests that our ability to model river hydraulics is arguably "good enough" (i.e., not unacceptable) rather than "good". Further, cases exist where the standard suite of assumptions are invalid. For example, channel losses can be important in arid regions where Hortonian runoff dominates, and ephemeral lakes can change size appreciably during runoff events, making modelling difficult (Michaud and Sorooshian, 1994; Ogden et al., 2000).

Streamflow measurement is critical to hydrology, as it is the most heavily used (and often the only available) indicator of the hydrological behaviour of a watershed. Fortunately, streamflow measurement can be relatively straightforward compared to measurement of other hydrological processes (e.g., areally-distributed precipitation and evapotranspiration). Structures such as weirs, flumes, and culverts can be used to measure flow directly through the well-documented relationship between their geometric properties and the location of the water surface at different flow rates. Accuracies of measurement vary with the type of structure.

A current meter, used to measure fluid velocity at a single point, can also be used to measure flow by employing a systematic pattern of velocity and depth measurements. The product of each pair of measurements is multiplied by its corresponding portion of the river cross-section. The results are then summed to attain an approximation of the streamflow. By repeating the above method at a variety of different flow levels, a relationship between ambient water surface elevation and discharge can be obtained. With this stage-discharge relationship, measurements of water stage (i.e., using a stilling basin or staff gauge) are easily converted into streamflow. This method is favoured by hydrometric agencies such as the Water Survey of Canada and the U.S. Geological Survey (Moore, 2004).

Geomorphologic changes can significantly affect the stage-discharge relationship; therefore, it must be updated on a regular basis to preserve its accuracy. This is the most common method for calculating streamflow; most gauging stations record only water level.

All of the above methods of measuring streamflow have conditions under which they are either infeasible or provide meaningless results. In basins where colder temperatures are associated with low streamflow, significant data can be lost or made suspect due to ice conditions (e.g., Environment Canada, 2001). General problems with current metering at low flows include streamflow velocities at or below the meter's stall speed, insufficient depth for required submergence, and insufficient stream width to sample a sufficient number of vertical sections (Pike and Scherer, 2003). Turbulent environments can also disrupt current metering, as velocities may not be consistent enough at any given point to provide a reliable measurement. Even relatively accurate measurement devices such as weirs and flumes are not exempt from problems; their installation can affect the local flow regime, and conditions must be allowed to stabilize before results can be considered broadly applicable (Michaud and Sorooshian, 1994).

Low flows, irregular cross-sections, and high turbulence may preclude the use of some or all of the above gauging techniques. However, in many such cases, discharge can be measured using a conservative tracer and the principle of mass balance. This approach, called salt dilution gauging, involves injecting into the flow a solution containing a chemical tracer of known concentration. The injection can be performed as either a continuous input or a slug injection. The dilution of the tracer is then measured at an appropriate distance downstream. Moore (2004) cites common table salt (NaCl) as the most popular tracer because it is inexpensive, readily available, easily measurable (using electrical conductivity), and non-toxic for the exposures currently associated with discharge measurements. Because salt dilution gauging relies on attaining a complete lateral mix of the tracer solution, it is ideally suited for the irregular and highly turbulent environments typical of mountain streams. For environmental and practical reasons, salt dilution gauging is less appropriate for flows greater than about 15 m^3/s (Kite, 1993).

However, salt dilution gauging also has limitations. Most importantly, the implicit requirement for steady-state flow generally limits salt dilution gauging to discrete (as opposed to continuous) measurements of discharge. There is also the potential for environmental impacts (e.g., Wood and Dykes, 2002), as high concentrations are typically observed near the point of injection. However, deleterious impacts are unlikely given the short-lived and localized nature of these higher concentrations; concentrations downstream of the mixing reach are usually far below commonly-accepted 48-hour toxicity thresholds (ibid.; Moore, 2005).

It is of great significance to hydrologic modellers that large floods often overwhelm fixed measuring structures, regardless of the measurement approach. This typically means that data collection is interrupted on the rising limb of a flood hydrograph, and is not resumed until repairs can be effected (e.g., Ogden et al., 2000). The result is that critical streamflow data – i.e., that of the flood peak and duration – is lost. In such cases, hydrologists must rely on expert analysis of qualitative data such as eye-witness accounts and high-water marks such as debris scatter and tree scarring. Approximate formulae such as the slope-area method, an approach commonly applied in such cases, incorporate substantial uncertainty into peak flow estimates.

Streamflow entering a reservoir can also be estimated where outflows are controlled or monitored and the volume of the reservoir is known. Inflows are estimated as the sum of total releases from the reservoir and any change in storage. Change in storage is typically estimated by measuring reservoir level and converting it to volume using a stage-storage curve; this often requires an implicit assumption that the lake or reservoir is groundwater-neutral (i.e., neither accumulating from nor discharging to the subsurface). Due to the potential impact of dynamic effects like wind set-up and waves on stage measurement, back-calculated estimates for reservoir inflows should be used with caution.

2.1.4 Evapotranspiration

Evaporation and transpiration together can account for as much as 80% of hydrologic activity in a typical basin (Klemeš, 1986a). Calculations of pure evaporation are commonly limited in application to determining losses from lakes or reservoirs. Approaches include applying empirical adjustments to pan evaporation measurements, or using more detailed data to employ water budget, energy budget, or mass transfer techniques. In some cases, evaporation has been pre-calculated from data collected at a climate station (Environment Canada, 2005). However, few hydrologic models attempt to explicitly calculate evaporation.

Transpiration is essentially the evaporation of water taken up by plants, shrubs, and trees. In addition to the physical factors governing evaporation (e.g., exposure, heat fluxes), transpiration is known to vary widely with plant species, density, and size (Fetter, 1994). Available moisture can also limit transpiration when soil moisture drops below the plant's wilting point (Viessman and Lewis, 1996). Changes in season will directly affect transpiration as plants respond to environmental stimuli (e.g., cf. deciduous and boreal forests). Measurement of transpiration alone is extremely difficult and must be undertaken in closely controlled laboratory conditions which eliminate external evaporation (e.g., using a potometer). The results of such experiments are naturally highly dependent on ambient and constituent conditions.

Due to the difficulty of obtaining distinct estimates for each process independently, the term evapotranspiration (ET) is used to represent the combined return of water to the atmosphere through evaporation and transpiration (Pike and Scherer, 2003). ET is the largest sink for precipitation in all but extremely humid, cool climates (Fetter, 1994). Refsgaard and Storm (1995) note that ET accounts for approximately 70% of annual precipitation in temperate zones. Klemeš (1986a) points out the discrepancy between the large fraction of hydrologic activity in a basin ascribed to ET and its relative dearth of treatment in hydrologic literature and practice.

ET plays a minor role during short-term storm events, since the air is typically at or near its saturation point. Following precipitation events, depletion of the soil moisture through evapotranspiration creates a soil moisture deficit. Infiltrated water must replenish this deficit before subsurface runoff will contribute to stormflow.

The dependence of transpiration on soil moisture is reflected in the parallel dependence of ET on soil moisture. To account for, Thornthwaite (1944) calls the upper limit of transpiration losses "potential ET", defined as "the water loss which would occur if at no time there is a deficiency of water in the soil for the use of vegetation". As soil moisture declines from field capacity to wilting point, there is some uncertainty as to the rate at which evapotranspiration is affected (Fetter, 1994). The reduced rate of ET resulting from reduced soil moisture is referred to as "actual evapotranspiration". Although soil water content can be measured with some precision (at least as a point value), there is persistent uncertainty associated with the conversion between potential and actual evapotranspiration.

Point approximations of evapotranspiration for a particular plant and location can be obtained using a lysimeter (Fetter, 1994), but such estimates are typically limited due to *in situ* variability of species, size, and density. Pan evaporation data have also been related to ET through empirical coefficients with varying degrees of accuracy (Allen et al., 1998).

Methods for calculating evapotranspiration in hydrologic modelling are generally either empirical or based on the mass balance and energy budget approaches noted above. The water budget approach is of little use in hydrologic modelling, since ET is commonly required to close a water balance when estimating runoff, and energy-based approaches require types of field data not commonly available in meteorologic data sets (Fetter, 1994). The end result is that many hydrologic models use purely empirical relationships that require a minimum amount of data and which rely on calibration to be locally applicable (e.g., Quick, 1995).

More advanced methods such as the Penman-Monteith equation (Monteith, 1965) directly calculate actual evapotranspiration using complex energy balance inputs such as net radiation and vegetative canopy resistance to heat and vapour transfer. However, the Penman Monteith equation is still only intended for use with uniform expanses of vegetation – a rare condition in the natural environment (Allen et al., 1998).

In general, even the most accurate methods for estimating ET have substantial margins of error. For a deeper discussion of evapotranspiration methods, the reader is referred to the work of Jensen et al. (Jensen et al., 1990).

2.2 The Evolution of Hydrologic Modelling

2.2.1 The Beginnings of Modelling

Musings on the nature of hydrology have been traced back as far as the philosophers of Ancient Greece; however, a scientific understanding of the hydrologic cycle did not begin to emerge until the fifteenth century. Early efforts to quantify hydrologic variables first began to appear in the 17^{th} century, while 18^{th} century developments in hydraulic theory and instrumentation led to extensive experimentation and empirical study throughout the 19^{th} century (Viessman and Lewis, 1996). Among these, Singh and Woolhiser (2002) identify the beginnings of mathematical modelling of hydrology in the rational method of Mulvany (1851) and the relationship between peak storm runoff and rainfall intensity developed by Imbeau (1892).

The 1930s saw systematic field experiments conducted in the US Midwest in an attempt to understand the physical processes involved in the rainfall-runoff transition (Woolhiser, 1996). Hydrologic modelling began to emerge as a process for developing concepts, theories, and models of individual components of the hydrologic cycle, such as overland flow, channel flow, infiltration, depression storage, evaporation, interception, subsurface flow, and base flow (Singh and Woolhiser, 2002). Early hydrology typically ignored informational uncertainty and adopted an either/or approach in its formulation rather than combining all available information (Vicens et al., 1975). Burges (2002) reviews the significance of individual contributions by Horton, Penman, Darcy, and others to models of the various component processes. However, their collective progress was limited by the intensity of the computations involved. Further difficulties emerged as researchers provided strong experimental evidence of the non-linear nature of the runoff process (Woolhiser, 1996).

The advent of the digital computer in the 1960s with its ability to manage calculations of previously prohibitive complexity led to an explosion of interest and research in hydrology (Woolhiser, 1996). The first "real" hydrologic models emerged (e.g., the Stanford Watershed Model, SWM), being conceptual in nature while retaining a degree of theoretical physical significance in the controlling parameters. Through these tools, modellers were first provided with the capability to comprehensively synthesize past events, predict future events, quantify extreme conditions, evaluate anthropogenic impacts on hydrology, and thereby improve the understanding of hydrology (p. 238, Freeze and Harlan, 1969).

Even while the earliest conceptual models were being refined, some researchers adopted a contrasting paradigm. They proposed that models link together existing but independent mathematical descriptions of the various hydrologic processes (e.g., Freeze and Harlan, 1969). The goal was the creation of a comprehensive physically-based digital hydrologic model whose output would completely describe the hydrologic system. The dominant role of spatial and sequential variations in the model input and output is reflected in Freeze and Harlan's vision of such models as "three-dimensional boundary-value problems with spatially and sequentially distributed inputs, solved by numerical methods" (p. 255, ibid).

Although researchers noted early on that data requirements for such models would be prohibitive even for small heavily-instrumented research catchments, the intervening decades have seen wide adoption of Freeze and Harlan's framework. Beven presents one possible reason for the acceptance of the framework through his observation that "difficult sciences [...] often aspire to demonstrate progress and maturity by more advanced mathematical descriptions" (p. 203, Beven, 2002).

Many different models are now based on variants of either the Stanford Watershed Model ("conceptual" models) or the FH69 blueprint ("physically-based" models). The various types of models are discussed in more detail in Section 2.3.3.

2.2.2 The Philosophy of Model Development and Application

The adoption of hydrologic models into mainstream science and engineering has resulted in the emergence of two distinct goals when developing (or applying) a hydrologic model. The first goal is research-oriented, attempting to increase our understanding of hydrology through exploration of different assumptions and theories. The second goal seeks effective prediction of real world behaviour for applied contexts such as water resources management (Grayson et al., 1992b). Experimentation advancing the first goal cannot be concluded *a priori* to extend the understanding of the second, and vice versa (Beven, 1989).

Numerous mathematical models of the last century have been developed in the second sense, i.e., to address only the hydrological variable of interest at the time (Franchini and Pacciani, 1991). In contrast, several authors contend that the prevailing "cost-effective", results-oriented application of complex numerical models is of limited significance to true progress in hydrology (e.g., Bergström et al., 2002; Braben, 1985; Klemeš, 1982, 2000a).

Beven (2000) contends that, in the past, technological rather than scientific progress has fuelled the demand for hydrologic models. Many acclaimed "advances" in hydrologic modelling more correctly reflect one or more of (Beck, 1987; Beven, 2000; Beven and Feyen, 2002):

- the technological capability for enhanced data collection and management;
- the widespread implementation of more complex models;
- the application of more, less, or different calibration; or
- the easy visualization of results.

Fundamental modelling technology developed decades ago is still in use in many parts of the world, in part because new techniques have proven unable to improve model accuracy or efficacy when the underlying model structure or data are fallible (Woolhiser, 1996; Singh and Woolhiser, 2002).

Since the development of computer models of hydrology, peripheral problems like model calibration have seemed to dominate the focus on process and perspective advocated by Horton in the 1930s (Klemeš, 1986b). In general, hydrologic models require a sound physical basis if they are to be scientifically credible. Therefore, it is viewed as a positive development that more recent work shows a growing return to process-oriented research and a focus on field work (Klemeš, 2000a).

Strong physical reasoning or empirical evidence should be used to determine which representations and simplifications are appropriate for a given natural system (O'Connell and Todini, 1996). Woolhiser (1996) notes that a good match between dominant *in situ* and modelled processes is a pre-requisite for success; neglecting such intuitive precepts can lead to structural inadequacies and calibration errors (Gan and Burges, 1990b). Faulkner et al. (1998) and Franchini and Pacciani (1991) present some specific examples of inconsistencies of methodology.

Very few models and data sets are substantial enough to allow modellers to explain why a given hydrologic model simulation is considerably different from its corresponding field measurements (Smith et al., 1994). Often, the lack of a parallel field program for many models interferes with advancing an understanding of the model structure and its influence on results (Grayson et al., 1992b; Michaud and Sorooshian, 1994; O'Connell and Todini, 1996; Song and James, 1992). In the absence of objectively verifiable truthing data, Woolhiser (1996) believes a modeller must constantly be asking questions such as "Does this make sense?", "What is the uncertainty of my prediction?", and "Does this level of uncertainty render the analysis meaningless?". Seibert and McDonnell (2002) argue that the modeller should also be checking the model for consistency and reasonableness against any "soft" (i.e., imprecise or qualitative) data that may be available. Loague and Freeze (1985) point out that the usefulness of results often depends on the modeller's understanding of the applied model and its relationship to the hydrologic nuances of the subject catchment.

Poor or uncertain model results should instinctively beget a return to the prototype rather than additional "fine-tuning" of an extant model. Klemeš (2000a) contends that, if the substance of the model is held sacrosanct, little insight can be gained into the prototype regardless of the effort expended. Unfortunately, poor model results are seldom reported (Beven, 2000; Grayson et al., 1992b). The model developer's emotional investment often makes it more appealing to search for a new application for an unsatisfactory tool than to search for a new and better way of dealing with the problem (Klemeš, 1983).

2.2.3 Model Complexity in Context

Until approximately a decade ago, limited computational capability was a substantive barrier to advanced hydrologic modelling. However, the ongoing and accelerating growth of processing power made it inevitable that computational ability would soon outpace and surpass our ability to model. With an increased ability to solve numerical problems, studies of the effects of parameter variation on model results became more viable (Gupta et al., 1999). Simultaneously, the value of investigation into simplifications and approximate solutions (e.g., Kuczera, 1997) diminished as computational advances negated the main advantage of such approaches. The complexity of most contemporary models is now constrained only by the degree of complexity appropriate to the model, subject, and research context.

The definition of an "appropriate" degree of complexity has been and continues to be a major focus for discussion and research. Some believe that the required level of accuracy should dictate the choice of model in any situation, since even simple, easy-to-use mathematical models can often explain a large part of streamflow variance (Beven, 1989; Dunne, 1982; Garen and Burges, 1981; Jakeman and Hornberger, 1993). Smith et al. (1994) point out that the theoretical rigor of complex models is not an *a priori* guarantee of accuracy. In many cases, simpler models may give answers of the same quality as their complex counterparts, often at a lower cost (Gan et al., 1997; Woolhiser, 1996). In general, there is a continuum of trade-offs between the limited process description of the simplest models and the uncertainty introduced by those of higher complexity (Hornberger et al., 1985).

The hydrologist should keep in mind that all models are, by definition, an abstraction of reality and are therefore to some extent incorrect (Woolhiser, 1996). If a model's representation of reality were perfect and exhaustive, the model would not be a model of the natural system but its duplicate (Klemeš, 2000a). Results for even the best models should therefore be viewed with a degree of skepticism. Grayson et al. (p. 855, 1994b) present a strong case that "models should not be applied as substitutes for knowledge". Rather, the authors contend that the proper use and interpretation of model results can require more knowledge and hydrologic insight than would otherwise be necessary.

2.3 Model Classifications

When exploring an emergent field of science, the initial relationships developed are generally simple and lead to only limited knowledge and understanding. Such relationships are labeled "empirical" to distinguish them from the "causal" relationships that characterize process dynamics (Klemeš, 1982). Empirical relationships are often used as convenient summaries of complex causal chains. In hydrology the empirical approach has many different labels, such as operational, prescriptive, analytical, and statistical, which all attempt to "understand" and "explain" the behaviour of any system in terms of a relatively few comprehensible elements (ibid.). Scientific disciplines usually evolve from the construction of empirical models to the development of causal models after reaching a fairly advanced stage of development; hydrology is no exception. Models in use today span a range of complexity from purely empirical to highly detailed.

Individual hydrologic models are most commonly classified according to their complexity. Three categories are typically considered, namely empirical and black box models, conceptual models, and physically-based reductionist models (Kuczera and Parent, 1998). Boundaries between the model types are not strict, and the roles of models in different categories can be complementary rather than competitive (O'Connell and Todini, 1996). Gan (1987) provides more information on the various classes of models.

Although many studies of the 1970's and early 80's compare results of models selected from within a single category, few compare or contrast the results generated by different types of models. There is even some debate about whether such comparisons are valid given the dissimilarities of process and application across model types; Smith et al. (p. 851, 1994) refer to one cross-comparison study as "a classic example of apples versus oranges". The lack of comparison studies makes it difficult to assemble a robust synthesis of experimental conclusions (Woolhiser, 1996). Further, the choice of model for a given situation is frequently dictated by available data or other factors not considered in a comparison study.

Singh (1995a) provides reviews of many models of varying complexity. A comprehensive and fairly up-to-date listing of all watershed models is enumerated by Singh and Woolhiser (2002).

2.3.1 Statistical, Empirical, and Black Box Models

From a scientific perspective, statistical, empirical, and black-box models represent the simplest class of hydrologic model. Singh and Woolhiser (2002) trace the roots of mathematical modelling in hydrology to the 19^{th} century works of Mulvany (1851) and Imbeau (1892). More recent examples of empirical models in hydrology include statistical Flood Frequency Analysis (FFA), regression models, and the transfer function models described by Jakeman and Hornberger (1993). Models in this category can also include the simple correlation of hydrologic quantiles (e.g., flood peak, characteristic unit hydrograph, or low flow) with geologic, geomorphic, and climatologic variables (Dunne, 1982).

Statistical models commonly consider a streamflow time series as a mathematical series having general descriptive properties such as central tendency, variance, and autocorrelation. Parameters are unlikely to have physical significance. When models are viewed as a purely mathematical construct, Beven (1989) notes that three to five parameters should be sufficient to reproduce most of the information in a hydrological record. Jakeman and Hornberger (1993) conclude that the "permissible model complexity" seems to be around six parameters.

Grigg et al. (1999) strongly caution against blind reliance on empirically-estimated floods (e.g., through FFA). Klemeš (1982) presents a stronger view, arguing that by utilizing the data to define an appropriate model, such models are self-limited and have no justification beyond their underlying data set. This does not mean that simple mathematical models are of little use to the hydrologist; on the contrary, they can be appropriate or even optimal in circumstances where the limitations of the methods are acknowledged.

The generic term "black-box" model refers to a model in which inputs are converted into outputs using formulas, calculations, or pre-programmed relationships that the end user does not need to see or understand to be able to use. Although all higher-order computer-based models of hydrology could therefore be considered "black-box" models, the more common interpretation refers to a model that – like statistical and empirical models – makes no attempt to employ the known physics of the hydrologic phenomenon (Kuczera and Parent, 1998). However, unlike statistical or simple empirical models, "black-box" models can possess significant technical complexity. Substantial knowledge is commonly required for set up, training, and analysis of process-oriented output characteristics like predictive uncertainty.

The concept of a "black-box model" in hydrology is perhaps most clearly understood when used to describe an artificial neural network (ANN). ANNs create a (potentially complex) set of relationships between input and output data that is fully dependent on (and thus variable with) the set of data used to "train" the model. In this way, an ANN model differs from all other model types, which combine a fixed set of formulas and algorithms with variable parameters. A working understanding of the complex "black-box" mechanics of an ANN hydrologic model would yield little to no hydrologic insight. Singh and Woolhiser (2002) observe that even the most complex ANNs do not model the internal processes of a catchment.

However, the ability of ANNs to recursively learn from data can save substantial time in model development, especially where traditional parameter estimation techniques are not convenient (Singh and Woolhiser, 2002). For more information on ANNs, the reader is directed to a two-part, detailed discussion of the role of ANNs in hydrology authored by the ASCE (2000a, 2000b).

2.3.2 Conceptual Models

Conceptual hydrologic models comprise an intermediate level of complexity. They generally represent the hydrologic cycle as several interconnected subsystems, each of which simulates a component process through empirically or heuristically-determined but physically-plausible functions (Duan et al., 1992). Conceptual models capture broad features of catchment response but are computationally and informationally straightforward, attempting to balance structural simplicity against the physics of the problem (p. 217, Franchini and Pacciani, 1991). The intended spatial and temporal scales of application often have a profound influence on model structure development (Singh, 1995b).

Beven (1989) reminds hydrologists that a conceptual model presents an approximation of the real world and therefore must introduce significant potential for error and uncertainty. In particular, Hornberger et al. (1985) note that some parts of a conceptual model may have a stronger basis in scientific or physical theory than others. Boundary conditions used to define the behaviours of the various component sub-models may also be neglected or changed, divorcing the model structure from its basic constitutive assumptions (Franchini and Pacciani, 1991).

There are different degrees of physical basis even within the category of conceptual models. Unsurprisingly, more complex models typically require greater effort in calibration and application than simpler models. In general, the level of calibration difficulty is directly related to the number of parameters and complexity of model structure. Franchini and Pacciani (1991) apply several conceptual models of varying complexity to a four-month simulation, reporting that all but one of the models generate acceptable simulations of the recorded discharges. Significantly, the more complex models require much more effort in calibration than the most abstract model, although the authors report that it is generally useless to attempt to identify linkages between the *in situ* and modelled runoff processes in the most abstract case (ibid.).

Certain structural features may limit the ability of a conceptual model to represent the hydrologic response of a catchment. For example, the typical "conceptual" representation of the subsurface involves partitioning it into two or more discrete zones with fixed storage capacities and infiltration thresholds. This typically results in a crude representation of processes such as infiltration, percolation, and evapotranspiration (Gan and Burges, 1990a). Klemeš (p. 102, 1982) criticizes such simplicity for "[taking] shortcuts to fill the void between the data and the goals with logically plausible assumptions that are sometimes correct but often wrong and, more often than not, individually untestable". Explicit validation of the model structure is usually not feasible due to the abstract nature of the model processes (Gan and Burges, 1990a).

The hydrologic characteristics of a catchment may also preclude good simulation by a conceptual model, as certain conditions are more easily realized in a conceptual framework than others. Gan and Burges (1990b) observe poor calibration results from the application of a conceptual model to a catchment having multiple hydrologically-distinct subcatchments with a common outflow. Most conceptual models have been developed for application to temperate or wet climates, making simulation of dry catchments (i.e., those in which less than 20% of rainfall becomes runoff) more difficult due to the greater significance of infiltration and evapotranspiration (Gan et al., 1997).

The abstraction of physical reality in a conceptual model limits the temporal resolution that can be realized. The sensitivity of a model to the chosen timestep is greatly increased as the timestep approaches and exceeds the basin residence time (Arnaud et al., 2002). Gan and Burges (1990b) find close agreement between observed and simulated mean flow rates at the daily scale, but the comparison deteriorates for lesser timesteps. Micovic (2003a) believes that satisfactory results for a watershed-scale conceptual model will prove unattainable for any time step less than one hour.

The abstract, "grey-box" nature of conceptual models is distinguished by the need to calibrate one or more parameters using observed data (Kuczera and Parent, 1998). Like empirical models, conceptual models are best suited to applications where conditions during the calibration period are hydrologically similar to those of the simulation period (p. 186S, Klemeš, 1986a. While identifying a unique and objectively verifiable set of "true" parameter values is usually impossible, Gan and Biftu (1996) argue that a conceptually-realistic parameter set is a pre-requisite for good performance in forecasting or prediction. However, earlier work by Gan (Gan, 1987; Gan and Burges, 1990a) shows that even a conceptually-realistic parameter set cannot guarantee good performance.

Because the parameters of conceptual models are defined by the data, the resulting structure is not general in its applicability (Beven, 2002). Like empirical models, there is no basis for assuming that a conceptual model can effectively extrapolate beyond its calibration experience (Gan and Burges, 1990a). Reducing dependence on calibration by decreasing the number of parameters in a conceptual model would have the undesirable effect of transforming the "grey-box" (i.e., conceptual) representation of the watershed into a purely black-box description (Seibert, 2000). Additional discussion concerning the extrapolation of calibrated models is provided in Section 3.4.

2.3.3 Physically-Based Models

In many fields including hydraulics, the most accurate form of modelling involves a "physical model" – an exact, scaled representation of the natural system designed and constructed under controlled conditions for investigative purposes. However, the large areal extent of most watersheds combines with small-scale hydrologic activity to preclude the use of physical models for all but the most basic hydrologic experiments. The most complex hydrologic models are instead based on highly detailed mathematical representations of the various *in situ* processes. Such models are referred to as physically-based. In contrast to physical models, physically-based mathematical models apply partial differential equations to represent the various processes within the boundary conditions of the basin (Freeze and Harlan, 1969). Therefore, while not as exact as a well-developed physical model, a physically-based hydrologic model is nominally capable of simulating in detail the internal processes and variables of a catchment.

There seems to be little argument that models of laboratory-scale systems are physically-based, in comparison with models operating at the catchment or basin scales (Woolhiser, 1996). Some argue that physically-based models are, in essence, an attempt to "scale up" the processes of the laboratory scale to larger contexts (e.g., Kuczera and Parent, 1998). The scaling and transposition of processes between dominant scales can lead to potential inconsistencies with their defining theoretical assumptions. The strengths and limitations of process and data scaling are discussed further in Section 3.1.

As a result of their complexity and detail, physically-based models have greater data requirements than their simplified counterparts. As a minimum, physically-based models generally require the following four kinds of input data (Freeze and Harlan, 1969):

- model definition information (e.g., grid size, time step, and topographic data);
- meteorological input data (e.g., the time-variant flux of water at the soil surface);
- flow parameter estimates (e.g., Manning's n, hydraulic conductivity); and
- mathematical basis and structure (i.e., the equations and functions that define the model).

Contemporary models can also incorporate diverse data from digital elevation models, remote sensing imagery, chemical tracer experiments, and groundwater level monitoring. The intensive data dependence of physically-based models implies that their performance is strongly dependent on the completeness, accuracy, and representativeness of the input data.

Physically-based models are arguably more "realistic" than models which must be calibrated to historical data in a curve-fitting exercise (Beven, 2001). Physically-based models offer the possibility of hydrologic relevance under conditions beyond those of the process-recorded history, and may identify efficient shortcuts which can then be used to improve empirical or conceptual models (Klemeš, 1982). In particular, research has shown that physically-based models offer improved performance over calibrated models for situations where physical and hydrologic descriptions of the watershed are available but a long-term gauging record is not (Michaud and Sorooshian, 1994).

Physically-based models are also able to generate distributed predictions and thereby evaluate changes in the constituent conditions of a watershed (Beven, 2002). However, internal calculations for points within the catchment are not necessarily representative of physical reality and are easily divorced from their uncertainty (Grayson et al., 1992b). Nonetheless, the distributed predictions of these models have proved popular in diverse fields where distributed hydrologic results act as inputs to other physical, chemical, biological, environmental, or ecological models (Singh, 1995b).

Two classes of criticism are commonly directed at physically-based models: firstly, strictly speaking, physically-based models almost never have an absolute physical basis (i.e., in the sense that it is highly unusual that all parameters can be determined *a priori* from research or field measurements); secondly, physically-based models are more likely to be mis-used than simpler models (Woolhiser, 1996). Grayson et al. (1994a) conclude that their inability to effectively represent the large- and small-scale spatial variability of rainfall is likely to eliminate any theoretical advantages that models of this type might otherwise possess. Others disagree, cautioning against "indicting" physically-based models on the basis of a single set of poor or non-representative results (Smith et al., 1994; Woolhiser, 1996).

In many cases, concern about the predictive capabilities of physically-based models appears to come in part from difficulties in application experienced by those who do not understand the model (Woolhiser, 1996). O'Connell and Todini (1996) propose that rather than abandoning physically-based modelling, one should re-align expectations of what can be achieved, and on what time scale.

In the most precise sense, a "physical basis" implies that a model does not require calibration. This is rarely the case in practice, resulting in an inability to completely abandon empirical, calibrated parameters (Madsen, 2003). For example, the prominent physically-based model MIKE SHE uses calibrated coefficients for calculating overland and channel flow, while flow through macropores in the unsaturated zone is modelled by an "empirical bypass function" (Refsgaard and Storm, 1995). For this reason, an uncalibrated physically-based model can be at a distinct disadvantage when compared directly with calibrated models (Loague and Freeze, 1985).

Calibration of physically-based models is difficult due to frequent mathematical over-parameterization, which in turn results from a lack of definitive *a priori* parameter value estimation. Woolhiser (1996) argues that the need for calibration often justifies a reduction in model dimensionality. In some cases, "excess" parameters can be replaced by simplifying assumptions. For example, the MIKE SHE model can be used with a lesser complement of parameters if the data are insufficient to support its full dimensionality, a feature specifically intended to reduce the risk of overparameterization (Refsgaard and Storm, 1995). The impact of reductions in model complexity or parameter dimensionality on the degree of "physical basis" has not been widely discussed in the literature.

Woolhiser (1996) identifies a dichotomy between the growing acceptance of physically-based models in the engineering community and the skepticism of many in the research community. Some have gone so far as to suggest that simpler models may be more appropriate in certain situations (ibid.). Regardless of their practical utility in a forecasting or predictive context, it is generally agreed that physically-based models have great potential for exploring the details of hydrologic interactions and investigating the fundamentals of runoff processes (Dunne, 1982; Gan and Burges, 1990a).

2.3.4 Other Classifications

Models can be classified by factors other than complexity of model structure. Singh (1995b) presents a fairly comprehensive overview of alternative classification systems. More prominent distinctions include the topographical representation of the watershed (lumped or distributed), the duration of the model (single event or continuous simulation), and the mathematical approach (stochastic or deterministic). Classification according to these factors is not necessarily strict.

Amongst the above, the classification most significant to the consideration of uncertainty is arguably the distinction between lumped and distributed models. If the spatial scale of model computation is large in comparison with the scale of *in situ* spatial variation, model parameters become physically unmeasurable and assume "effective values" (Woolhiser, 1996). Such models are referred to as "lumped". Most lumped models use a single set of representative values for each application (i.e., each watershed or catchment). Conceptual models are almost always lumped to some degree, whereas distributed models have been developed in response to the detailed representation required by physically-based models (Hornberger and Boyer, 1995). Typically, the distributed model breaks down a catchment into smaller response units or grid cells. A discussion of the relationship between variability, scale, and distributed modelling is included in Section 3.1.1.

Hydrologic models provide either continuous or single-event simulation. Both continuous and event-based models step through their application period using discrete timesteps. However, continuous models simulate hydrologic conditions for a prolonged period, whereas event-based models are limited to consideration of a single runoff event. Event-based models typically do not simulate ET, soil moisture depreciation or catchment recession.

Either type of model can be used for a given application, but one type may be better suited than the other in certain situations. For example, implementing the event-based model KINEROSR allows Faurès et al. (1995) to compile a statistical representation of peak estimates from multiple simulations of storm runoff without having to extract the information from a continuous time series. In a different study, Cooper et al. (1997) find that objective functions applied to the entire data set yield better approximations of the synthetic parameter set than objective functions that only examine peak or low flow quantiles. The results of Cooper et al. (ibid.) imply that a continuous model is more appropriate for their study.

Event-based models require a "snapshot" of watershed conditions at the beginning of each simulation. Measurements of these conditions are seldom available and therefore must be assigned *a priori* or calibrated as additional parameters. One could argue that event-based models exchange the uncertainty in modelling ET and soil moisture for uncertainty in initial conditions. For cases where initial conditions for an event-based simulation are obtained from a preliminary model, the user should be aware that the simulation is still subject to uncertainty in antecedent conditions; it is merely hidden as a computed variable (Grayson et al., 1992a). Michaud and Sorooshian (1994) present an example of a hydrologic model (KINEROS) applied with initial conditions excerpted from the output of another model.

Although continuous models also require initial conditions, an accurate estimate of the initial conditions is less important if data series is sufficiently long. In such cases, a "spin-up" (or "warm-up") period can be used. The idea is that, over time and regardless of initial conditions, modelled watershed conditions will gradually approach those *in situ*. Therefore, provided a sufficiently long data series is available, actual starting conditions are irrelevant (Houghton-Carr, 1999). Output from the spin-up period is not used to evaluate model performance.

The proper duration of the "spin-up" period should be determined from the time required for modelled watershed conditions to converge with *in situ* conditions. In practice, spin-up times range from several weeks to several years (Bingeman, 2001; Madsen, 2000; Thyer et al., 1999; Vrugt et al., 2003).

Finally, although this thesis is written largely in the context of deterministic hydrologic models, the reader should be aware of the range of stochasticity that can be accounted for in a hydrologic model. Most hydrologic models currently in use are fully deterministic; from the perspective of the model, all parameters and data are known with certainty. However, it is possible for a model to incorporate probabilistic representation of both parameters and data. In most cases, stochastic models use repeated Monte Carlo simulations to establish a probability-magnitude output which supplants the simple magnitude estimate provided by deterministic models.

Probability distributions for stochastic model parameters are usually either specified *a priori* based on expert knowledge or published data, or estimated from a field sampling program. The model is run repeatedly using random samples from the distribution of each parameter.

Stochastic models, such as the one espoused by Kuchment and Gelfan (2002), are typically utilized for extreme event prediction. Extreme events are most often characterized by significant uncertainty regarding antecedent conditions and flood-generating processes. Although stochastic models are designed to partially quantify model predictive uncertainty, the user should be aware that results will typically be limited by the least accurate constituent parameters, components, or probabilities.

2.4 Model Calibration

Every hydrologic modeller faces two fundamental challenges: the first is to select an appropriate model for the study site; the second is to determine parameter values such that the model closely simulates the *in situ* behaviour (Sorooshian and Gupta, 1995). Poor parameter specification can lead to poor simulation regardless of model sophistication (e.g., Michaud and Sorooshian, 1994). Therefore, careful set-up and initialization is necessary to maximize the reliability of the model (Gupta et al., 1999).

Model identification (e.g., system and boundary definition, data specification, and delineation of relationships between variables) cannot be treated as entirely distinct from model calibration (e.g., selection of evaluation criteria, identification of an informative data set, and parameter optimization). As the focus of this work involves neither designing nor altering a model, the model identification phase is largely set by the selection of an extant hydrologic model. However, even when applying a known model, the calibration phase should be pursued with due consideration of the context of the structure and properties of the individual model (Sorooshian and Gupta, 1985).

Three stages should be explored in model application: calibration, validation, and an assessment of reliability (Singh, 1995b). The bulk of this Chapter deals with model calibration. Validation is discussed further in Section 2.4.6, while reliability (in terms of the uncertainty in model results) is the subject of Chapter 3.

In general, applying a model requires estimating values for two types of parameters: "physical" parameters which can be measured (e.g., watershed area) and "process" parameters which represent watershed properties not directly measurable (e.g., time lag constants for runoff) (Sorooshian and Gupta, 1995). Calibration involves back-calculation of the "process" parameters through trial-and-error comparisons of predictions to runoff records (Dunne, 1982). Sorooshian and Gupta (1995) observe that *a priori* judgment and experience are frequently used to identify a range or bounds for each "process" parameter during the preliminary stages of calibration. In general, the use of expert judgment in such matters should complement rather than replace other evidence (Keeney and Winterfeldt, 1989).

The intention of calibration should be clear before commencing – i.e., is the goal simply to achieve the best possible fit for the calibration set, or does it involve the identification of a unique and realistic parameter set that closely represents the physical system (Sorooshian and Gupta, 1983). Such general considerations are discussed in Section 2.4.1, while Sections 2.4.2 and 2.4.3 address manual and automatic calibration, respectively.

2.4.1 The Nature of Calibration

Until recently, calibration has arguably been the most intensively investigated aspect of hydrologic modelling (Beven, 2000). Through many studies, threads of consistency have emerged identifying various pre-requisites for a successful calibration.

In the current era of reduced funding for hydrometric and meteorologic data networks, it is fitting that primary consideration be afforded to the data. Although even a single year of calibration data can assist in parameter specification, calibration is much more sensitive to data informativeness than to length of record (Gan et al., 1997; Gupta et al., 1998; Sorooshian et al., 1983). Refsgaard (1994) concludes that three to five years of streamflow data constitute an adequate basis for calibration of a continuous model. Yapo et al. (1996) find that informativeness of the data set is optimal using an 8-year span of streamflow data. While much less common, some guidance on the appropriate size of a calibration data set for event-based models is also available: Sorooshian and Gupta (1995) recommend that the number of data used in calibration exceed 20 times the number of parameters being estimated. Naturally, a model simultaneously calibrated against multiple data series (e.g., multiple streamflow measurement locations, or streamflow and groundwater together) will typically require a shorter period of data than the same model calibrated against catchment outflow alone.

While hydrologic model calibration generally implies comparison of observed and simulated catchment outflow, the use of additional information can increase confidence in the representativeness of the model. Calibration can also include comparisons to observed time series of depth to the water table, snow depth or snow water equivalent, soil moisture, stable natural isotopes (e.g., oxygen-18), alkalinity and pH, or nitrogen loadings (Bergström et al., 2002). Calibration against different data types may allow insights into different aspects of the model. For example, the application of conservative chemical tracer data provides an opportunity to improve descriptions of water pathways and residence times (ibid.).

At the outset, a modeller must decide what aspects of watershed behaviour will form the basis for calibration, and how those behaviours will be measured (Gupta et al., 1998). Further, the properties of the model (including model error) need to be recognized and considered when designing a calibration strategy (Yapo et al., 1996). Most importantly, all processes that are hydrologically relevant in the catchment must be activated during calibration. Unactivated process parameters cannot be considered calibrated and are likely to cause poor model performance if activated during later simulations or in other scenarios (Gan and Burges, 1990b; Lei and Schilling, 1996). Sorooshian and Gupta (1985) utilize data sets chosen specifically to ensure activation of all aspects of the model.

Data sequences containing a high degree of hydrologic variability are more likely to activate the various operational modes of the model and thus result in more reliable parameter estimates (Sorooshian and Gupta, 1995; Yapo et al., 1996). Wet years (or catchments) tend to be better suited for calibration against catchment outflow than dry years (or catchments). This is due to their more variable flow regime and the dominance of outflow – as opposed to ET – as a driving flux (Gan, 1987; Gan et al., 1997; Thyer et al., 1999). Quality of response is also less consistent for dry antecedent conditions (Gan and Burges, 1990b). Yapo et al. (1996) recommend using a data set consisting of the wettest period on record. This recommendation is supported by the experience of the US National Weather Service (Hogue et al., 2000). Gan (1987) attributes the acceptable performance of his model under extreme input to good representation of the processes important to extreme event runoff within the calibration data. He generally notes that including major flood events in the calibration data series can improve the adequacy of calibration for extreme conditions (ibid.).

One may be inclined to conclude that a model should be calibrated on conditions similar to those expected during its application. However, Klemeš (1986b) points out that this is not logical if the goal of calibration is to ascertain an appropriate model structure rather than merely provide the best fit for the observed data. If the goal is an appropriate model structure, Klemeš advises calibrating with hydrologically dissimilar data before validating on conditions similar to those of the target simulation.

Models relying solely on calibration data are often overparameterized. In such cases, the process of establishing appropriate interaction between the model and the available data can overshadow the search for a good representation of the system (Kuczera, 1997). In a large number of cases, varying parameter values obtained from multiple calibrations of the same model generate similar goodness-of-fit results. Seibert (2000) find that the aggregate parameter range defined by multiple "successful" calibrations encompasses approximately one-third of the feasible parameter space.

Non-identifiability (i.e., the inability to identify a unique and optimal parameter set) can manifest itself in two ways. Firstly, distinct output functions may yield similar results for quantitative evaluation criteria. Secondly, distinct parameter sets may generate identical output functions. The latter is generally of more concern to hydrologic modellers (Sorooshian and Gupta, 1985). Sorooshian and Gupta (1983) and Yapo et al. (1996) present several reasons why a model parameter may be poorly identifiable in the second sense:

- parameter interdependence: parameters interact strongly with other parameters, allowing for compensating variations in the feasible parameter space;
- parameter nonstationarity: parameter location and precision are correlated to varying watershed or data characteristics not accounted for during calibration;
- data noninformativeness: the data do not contain the hydrologic conditions required to properly identify the parameters;

- criterion inadequacy: the evaluation criteria cannot properly extract the information contained in the data;
- mathematical difficulty: the calibration is constrained by local features of the response surface; and
- insensitivity: variations in the values of the parameters do not significantly affect the model output or the calibration criterion.

Testing for global identifiability is impractical in most cases where it is not impossible (Sorooshian and Gupta, 1985). Therefore, in most situations, identifying stable and realistic parameter values is as important as finding values which produce a good fit to the observed data (Micovic, 1998). Later chapters of this work focus on the need to understand and quantify the potential impact of alternate solutions on model simulations.

2.4.2 Manual Calibration

The most common approach to model calibration is manual (sometimes called "expert") calibration (Hogue et al., 2000). Typically, an expert user follows a semi-intuitive trial-and-error process, adjusting parameter values to improve the "closeness" of fit between observed data and simulated output (Boyle et al., 2000). Although visual inspection of the observed and simulated hydrographs is the typical standard measure of performance, statistical comparisons and other factors are commonly used in conjunction with qualitative analysis of fit (Hogue et al., 2000). Emphasis is usually placed on matching peaks, flood volumes, recession slopes, and base flow (ibid.).

The evaluation process conventionally involves a qualitative synthesis of visual inspection and numerical criteria, and often proceeds through sequential calibration of parameter groups or objectives. The intent of sequential calibration is to simplify the process by considering fewer variables simultaneously. This is achieved by first isolating and optimizing those parameters or objectives having the greatest influence over the results while neglecting secondary objectives and less-sensitive parameters. However, the potential for parameter interaction implies that sequential calibration – unless highly iterative – can lock the parameter values in a trap set up by compensating errors (Bergström et al., 2002).

The process of manual calibration is highly labour-intensive. It calls for a considerable degree of training, experience, and effort, and requires a substantial commitment of time and expertise for even the most expert user (Boyle et al., 2000; Gan, 1987; Hogue et al., 2000). Achieving acceptable results requires that the modeller have a working understanding of the model, the data, and the watershed system (Boyle et al., 2000).

Many organizations have developed systematic procedures to be followed by users in an attempt to standardize the process. The US National Weather Service in particular has developed a manual procedure that provides excellent model calibrations but is knowledge-intensive and therefore not easily transferred between people or models (Boyle et al., 2000). While the calibration procedure followed by a user may be standardized, the individual's adjustments will be influenced on personal skills and experience. Houghton-Carr (1999) finds that various experts offer differing perspectives on what constitutes "good" or "bad" performance, and will typically evaluate performance differently.

Manual calibration relies on the conceptual relationship between model parameters and watershed characteristics to support the hydrologist's expectation that adjustment of a given parameter value in a particular direction will have a predictable effect on the model output (Gupta et al., 1999). Sorooshian and Gupta (1995) state that the main weakness of manual calibration is in not knowing when to quit, since even experts may not always be able to assess the potential for further improvement. The semi-intuitive nature of working through the complex process of manual calibration can lead to considerable frustration (Lan, 2001).

One of the greatest advantages of manual calibration is the ability to incorporate explicit allowance for potential data errors (Boyle et al., 2000). Where simulation quality is poor due to erroneous or non-representative data, the user can choose to ignore the erroneous data in favour of those believed to be more correct.

2.4.3 Automatic Calibration

Automatic calibration uses numerical comparisons between observed and simulated data to optimize parameter values without user intervention (Madsen, 2003). An automatic model calibration is typically objective in nature and relatively easy to implement, but focuses on numerical optimization and may thus result in a hydrologically unrealistic parameter set (Boyle et al., 2000; Hogue et al., 2000). Historically, this has led the hydrologic modelling community to be highly skeptical of automatic calibration in an applied context.

Nonetheless, a substantial body of opinion suggests that modern state-of-the-art automatic calibration methods can be considered a viable alternative to expert calibration for some purposes (Gan and Biftu, 1996; Gupta et al., 1999). One such situation is where multiple objectives add significant complexity to the calibration process (Madsen, 2003). Another example is the work of Seibert and McDonnell (2002), who use multi-objective calibration to demonstrate that automatic methods do not necessarily preclude the incorporation of qualitative or imprecise data. Automatic calibration can also be used as an adjunct to manual calibration; automatic optimization algorithms are commonly used for fine-tuning an expert hydrologist's manual calibration (Duan et al., 1994b).

The evolution of automatic calibration has been motivated by the need for simplicity, speed, objectivity, confidence, and less reliance on hard-to-find expert calibrators (Hogue et al., 2000; Sorooshian and Gupta, 1995). Most research regarding automatic calibration has been carried out for lumped conceptual models (Madsen, 2003). Recently-developed multi-objective automatic calibration tools may increase the application of automatic calibration to more complex distributed catchment models as they are adopted by distributed modellers seeking consistency with observed data throughout the catchment.

The implementation of an automatic calibration is defined by the selection of its various components: objective function, search algorithm, calibration data, and search termination criteria (Hogue et al., 2000). Typically, the automatic portion of any calibration routine has the following steps in common (Gupta et al., 1999):

- identify components;
- obtain an initial estimate of the values (or ranges) for the parameters;
- execute the model;
- measure performance; and
- apply a search algorithm to find parameter values that improve the value of the objective function.

The question of "when to quit" is addressed for automatic calibration through explicit specification of termination criteria. Appropriate termination criteria are necessary to achieve the desired balance between effectiveness and efficiency. Termination criteria typically include parameter convergence (meaning that unique values have been identified for each parameter), function convergence (meaning the objective function cannot be significantly improved), or a maximum number of iterations (to avoid endless loops) (Hogue et al., 2000).

Although parameter convergence is usually the best indicator of a mathematically-optimal parameter set, none of the above criteria can conclusively ascertain that the global optimum has been reached, since poorly-chosen criteria can halt the search procedure before it locates the best available parameter set (Singh and Woolhiser, 2002; Sorooshian et al., 1983).

The strength of an automatic procedure depends on how well its design reflects the various factors important to obtaining a successful calibration in each specific case. The identification of these factors has been the subject of significant research (Boyle et al., 2000). Much early research focused on improving search methods that converged to one of the multiple points or clusters comprising the local optima of the response surface (Kuczera, 1997). Although many advanced search techniques can now avoid being trapped by local optima, the problem is far from solved. While new procedures may be able to produce reliable estimates of global optima for complex problems (e.g., Duan et al., 1994b), the focus of these procedures is still directed at optimization rather than hydrologic model calibration.

Data uncertainty, model structure uncertainty, and dependence on the calibration procedure (e.g., objective function selection) imply that the best possible representation of the system will not necessarily coincide with the global optimum of the objective function (Vrugt et al., 2003). Gan

(1987) observes objective function values approaching their theoretical limit for some calibrations, despite "gross simulation errors" obvious in the corresponding hydrographs. Houghton-Carr (1999) emphasizes that optimization algorithms lack an expert's knowledge of model structure and cannot judge when a parameter set is unrealistic.

Automatic calibration methods are typically compared on two scales: accuracy of solution (i.e., effectiveness) and efficiency of convergence. Efficiency of convergence for an automatic calibration method refers to how quickly the final solution is identified, implicitly representing the amount of computational effort required (e.g., required number of function evaluations) (Thyer et al., 1999). Robustness of solutions (i.e., the variation in solutions across multiple trials) is another potential criterion for evaluation (Cooper et al., 1997).

The emergence of automatic calibration techniques has reduced the degree of user training required to obtain realistic-looking results from a hydrologic model. However, the simplified, "hands-off" model calibration process and the "objectivity" inherent in automatic calibration also increase the potential for error, as there is a net loss of expert guidance in intuitively assessing the feasibility of parameter values and sets. The knowledge contribution of the expert user in manual calibration avoids many of the pitfalls awaiting the incautious automatic calibrator. Consider the case of obvious data error: methods of automatic calibration implicitly assume that all data are equally correct. An automatic procedure may distort the calibration to compensate for data error that an expert would quickly and easily disregard.

The concept of an "objective" automatic calibration must also be viewed with caution, as results have been shown to depend on subjective decisions such as the criteria to be optimized and the period of calibration data (Bergström et al., 2002). Results of comparisons for various automatic calibration algorithms may also be affected by changing the dimensionality of the parameter space (Seibert, 2000). Gan et al. (p. 97, 1997) present a study of differences between automatic calibrations applied with a variety of models, optimization algorithms, data, and objective functions.

More specific limitations of automatic model calibration have been discussed by many authors. Some of the more widely acknowledged challenges include the following:

- the inability of any single performance measure to properly address all the important characteristics of the system (Gan, 1987; Madsen, 2000);
- interdependence of parameter values, resulting in uncertain and very slow improvements (Gan, 1987; Moore and Clarke, 1981);
- problems with model or parameter identifiability, such that appreciable changes in parameter value cause little or no change in objective function (Moore and Clarke, 1981; Sorooshian and Gupta, 1983);
- the impossibility of exhaustively exploring the feasible region due to discontinuities, non-differentiable regions, or other complexities of the response surface (Gan, 1987; Moore and Clarke, 1981);
- the potential for model divergence given poor initial parameter values (Gan, 1987); and
- treatment of each combination of parameter values as "theoretically possible" when some may be incompatible with science or physics (Klemeš, 2000a).

Obviously, automatic calibration is not a facile solution to model calibration, and cannot entirely replace manual involvement in the process. User information is required to customize the approach and evaluate the process (Madsen et al., 2002). Multiple studies have shown that global optimization performance improves as parameter ranges are reduced or constrained (e.g., Franchini et al., 1998; Gan and Burges, 1990a; Kuczera, 1997; Seibert and McDonnell, 2002). Thus, limiting the parameter space to be searched is one example of a key area for expert user input.

The above-noted potential inadequacy of a single measure of performance is not always insurmountable, since automatic calibration is not limited to the single-stage global optimization of general objectives. Often, the calibration problem is reduced to sub-problems, each of which may be solved manually or by using different optimization techniques (Gupta et al., 1999).

However, one must be careful to avoid the possibility of an optimization algorithm negating expert knowledge incorporated earlier in the process.

Model-specific knowledge-based expert systems are one example of a hybrid approach. A knowledge-based expert system uses optimization to automate various stages of a manual calibration (Madsen et al., 2002). These processes are designed to yield results at least comparable to the expert process while improving efficiency. Boyle et al. (2000) propose using expert knowledge to select feasible parameter ranges and values for dominant parameters, then implementing an automatic search to minimize parameter uncertainty with a focus on the potential for parameter interaction. Hogue et al. (2000) propose a Multistep Automatic Calibration Scheme (MACS) that sequentially applies automatic calibration to different subsets of model parameters. Subjective and statistical evaluations of MACS show a significant improvement in efficiency over the manual approach, with consistently comparable results. However, the authors caution that techniques such as MACS are not yet ready for operational use (ibid.).

It is important to distinguish between successful testing of automatic calibration techniques and a readiness for "real-world" application; the first does not necessarily imply the second. Much research and testing is pursued on "sterilized" data to allow better assessment of the effectiveness and efficiency of the calibration procedure. However, results from such tests are significant only in a research context. Before being approved for practical application, automatic techniques should be extensively tested with the type, quantity, and quality of data encountered in the field (Hogue et al., 2000).

2.4.4 Methods for Automatic Calibration

Methods for automatic calibration encompass a range of complexity. This section explores some of the many techniques available to the hydrologic modeller, beginning with some simple techniques and progressing to those of higher complexity.

The simplest of automatic calibration methods is Monte Carlo Simulation (MCS). "Monte Carlo" refers to the general process of random sampling from a pre-determined statistical population with the goal of generating an approximate solution to a quantitative problem. For a Monte Carlo procedure, the accuracy of the approximation is related to the number of trials. In the hydrologic modelling context, Monte Carlo Simulation typically involves repeated trials of the candidate model using parameter values randomly sampled from *a priori* parameter distributions. In most cases where MCS is used as a calibration tool, the probability distribution is assumed to be uniform for each parameter to reflect the ignorance of the modeller with regard to the "most likely" parameter values.

The MCS technique is straightforward, preserves nonlinear interactions within the calculations, and is not dependent on limiting assumptions about an appropriate distribution of parameter values (Binley et al., 1991). The process has no memory and each trial is independent. Therefore, a Monte Carlo approach is concerned with neither improving objectives nor updating the parameter distributions.

The chances that purely random sampling will quickly result in a near optimal solution are negligible. Even if found, there is little chance that the optimal solution could be recognized as anything more than the best member of a finite set; a random search without direction requires enumeration of a statistically large portion of the parameter space before near-optimality can be declared with confidence (Lan, 2001). Therefore, one can conclude that while MCS is a good tool for exploring the parameter space or identifying a good starting point for more detailed procedures, it is not itself a strong tool for automatic calibration. However, Monte Carlo theory plays a role in most advanced optimization algorithms.

Most commonly-used methods for automatic calibration involve a directed search in the form of optimization. The explicit goal of optimization methods is the systematic adjustment of parameter values in search of a set that minimizes (or alternatively, maximizes) the value of an objective function or set of functions (Cooper et al., 1997). Analytically-based solutions to optimization problems require response surface conditions such as continuity, convexity, and twice differentiability, which (at least in combination) are uncommon in the context of applied hydrologic modelling (ibid.). This precludes the use of analytical or derived solutions, leaving numerical solution as the only viable approach.

Numerically-based optimization can involve local-search or global search strategies. Local search strategies move in the direction of improving solutions (e.g., downward gradient) until no further improvement is possible. Examples of local search methods include the Pattern Search

Method, the Rosenbrock Method, Sequential Quadratic Programming (SQP), and the Nelder-Mead Algorithm, also called the Simplex Method (Franchini et al., 1998; Hogue et al., 2000).

The Pattern Search Method (Hooke and Jeeves, 1961) explores the response surface around the current point and then forces exploration in the direction of greatest improvement (Franchini et al., 1998). The Pattern Search Method deals only with the value of the objective function and therefore does not require continuity or differentiability. Termination occurs when either the response surface shows negligible improvement in any direction, or when the calculated step size of each improvement step reaches a specified limit (ibid.).

The Rosenbrock Method (Rosenbrock, 1960) approximates a gradient search without requiring the evaluation of derivatives. It systematically explores improvements in the objective function value along a set of orthogonal vectors in parameter space, rotating the ordinate system in the direction of the computed gradient once all vectors have been explored. Hornberger et al. (1985) apply the Rosenbrock method to the hydrologic model TOPMODEL, but find that parameter estimation is unreliable.

The SQP algorithm applies a constrained-optimization version of Newton's method and is extremely efficient at converging to a local optimum. Franchini et al. (1998) demonstrate the potential for using SQP as a "fine-tuning" step in calibrating conceptual rainfall-runoff models.

One of the most robust and popular of local-search techniques is the downhill simplex algorithm designed by Nelder and Mead (1965), often referred to as the Simplex Method. The Nelder-Mead downhill simplex algorithm should not be confused with the Simplex Method of Dantzig (1951) used in linear programming.

The Nelder-Mead algorithm begins with a simplex of $n+1$ points in n-dimensional parameter space and rank-orders them best to worst according to their objective function value. The worst point is then reflected through the centroid of all other points, replacing its parent if the reflection results in an improved objective function value. If the reflected point is better than the best member of the simplex, the simplex is expanded in the direction of the reflected point. A similar contraction step is pursued if the reflected point is worse than the worst member of the simplex.

Termination typically occurs when the standard deviation of objective function values across the simplex drops below a specified value.

Duan et al. (1992) find that the Nelder-Mead Simplex Method outperforms other local search methods because it adjusts to the response surface and can avoid being trapped by the local optima within a region of attraction. However, there are two noteworthy limitations with regard to the Simplex Method and its treatment of parameters. Firstly, its effectiveness is substantially diminished when dealing with more than five to seven concurrently optimized parameters (Burges, 2003). Secondly, because the original Nelder-Mead Algorithm is unconstrained, bounded parameters must be transformed before they can be optimized (Gan, 1987).

Local-search methods can be applied as either single-step processes or successive optimizations. There are two possible approaches for applying multi-step local search methods, namely repeated trials of a single algorithm (e.g., Duan et al., 1992) or consecutive application of different search strategies (e.g., Franchini et al., 1998). While such multi-step local searches are more robust than a single trial, any improvement over the single-trial case is most likely due to the reduced dominance of the various problems affecting each individual trial.

Local search methods are frequently challenged and often defeated by numerous systematic problems. Local optima can trap a search algorithm before it reaches the global solution, necessitating user intervention before the search can proceed (Sorooshian and Gupta, 1995; Thyer et al., 1999). Poor objective function sensitivity to changing parameter values and significant parameter interaction can also pose difficulties, especially as the search approaches the optimum. Other common problems include non-uniqueness of optimal solutions and constitutive assumptions that are incompatible with the hydrologic data (Cooper et al., 1997). These difficulties have led researchers to conclude that techniques designed to locate the optimum point of a localized region are mathematically insufficient for locating global optima in complex situations such as hydrologic modelling (e.g., Duan et al., 1992; McCuen, 1973).

Many optimization techniques have been designed to avoid the most common problems encountered by local methods (Cooper et al., 1997; Madsen, 2003). These global (sometimes called parallel) search strategies include, among others, Adaptive Random Search, Genetic Algorithms, Simulated Annealing, and the Shuffled Complex Evolution algorithm SCE-UA

(Hogue et al., 2000; Madsen et al., 2002). Global optimization methods (GOMs) are commonly viewed as having a local phase (typically replicated from well-studied pre-existing local search techniques) and a global phase that avoids entrapment at local optima. Therefore, the most significant differences between alternative GOMs are found in their exploration of the problem space (Cooper et al., 1997).

Adaptive Random Search (Masri et al., 1980) is arguably the simplest example of a GOM. Adaptive Random Search chooses a fixed number of random samples from each of several nested sub-regions of the parameter space. The nested sub-regions are typically separated in size by an order of magnitude but are all centred on the same focal. After each subset has been examined, the sampled parameter set yielding the best objective function value becomes the focal for the next iteration. The search continues until the best point is repeatedly identified at the smallest sub-region of the search. Duan et al. (1992) report that the effectiveness of ARS is greatly improved when coupled with a local Simplex search.

Simulated Annealing (Kirkpatrick et al., 1983) is based on observations of the energy-minimizing processes undergone by cooling metals. The process walks through the parameter space, exploring new parameter combinations by evaluating their objective function value. Any point that improves the objective function becomes the starting point for the next exploration. Being a derivative of the Metropolis process, Simulated Annealing also accepts inferior points with a specified probability. The algorithm must be set up to allow a large number of "uphill" steps, but not so large as to make the process completely random (Thyer et al., 1999).

Genetic Algorithms comprise a special case of GOMs that shares the fundamental components of global optimization but is distinguished by its search technique. Although research into Genetic Algorithms began in the 1970s, the work of Goldberg (1989) was the catalyst for their wider adoption in diverse fields of optimization.

Like other methods, Genetic Algorithms use quantitative measures of performance as objective functions. However, Genetic Algorithms search the parameter space by emulating genetics (i.e., through the probabilistic selection and recombination of extant parameter values) rather than using Monte Carlo or gradient-based search techniques to generate successive improvements (Seibert, 2000). Genetic selection allows individuals which are better suited to their environment (i.e., parameter sets which produce better model simulations) to have a higher probability of reproducing (Franchini et al., 1998).

Various "rules" are specified to govern the creation of new parameter sets. Common examples of different rules include duplication of a parent (reproduction), re-combination of elements from multiple parents into a new individual (crossover), and selection of random values from the feasible range or a sub-range bounded by the parent values (mutation) (Lan, 2001; Seibert, 2000). Rules are typically applied on a probabilistic basis. Implementing an appropriate balance of rules and probabilities is crucial for the success of the algorithm (Seibert, 2000).

As for many GOMs, parameter values can either be used in their original form (e.g., Seibert, 2000) or modified into a format more amenable to the processes used in genetic algorithms; Goldberg (1989) uses binary representations of parameter values. In either case, discrete rather than continuous parameters are required. Naturally, therefore, the GA's optimization reflects an integer rather than continuous-variable solution. Without careful consideration of the discretization process, any GOM that produces discrete solutions may yield near- or even sub-optimal solutions with artificially precise parameter values.

Results of studies applying GAs to hydrologic models are mixed (e.g., Franchini and Galeati, 1997; Lan, 2001; Seibert, 2000). Wang et al. (1995) recommend coupling a Genetic Algorithm with a secondary local-search algorithm to improve results. For a complete discussion of the development of genetic algorithms, the reader is referred to the seminal work by Goldberg (1989).

One particularly well-documented and successful global optimization technique for automatic calibration of hydrologic models is the Shuffled Complex Evolution algorithm (SCE-UA), developed by Duan et al. (1994a) at the University of Arizona. The SCE-UA method applies a version of the Nelder-Mead algorithm using concepts from natural biological evolution to address the inefficiencies of the Nelder-Mead process (Duan et al., 1992). The SCE-UA method is designed specifically for the calibration of hydrologic models and is based on a synthesis of ideas including a combination of deterministic and probabilistic approaches, competitive evolution, complex shuffling, and systematic evolution of a population spanning the parameter space (Duan et al., 1994b).

The well-known Nelder-Mead algorithm is the basis of the local search component for the SCE-UA method. However, it is the global framework that makes the SCE-UA method powerful and unique. Duan et al. (1992) describe the process in detail; their flowchart of the global search framework is included herein as Figure 2-1. The exchange of information between points – achieved by shuffling individuals between complexes – is what allows the search to avoid being trapped by local optima. The entire population will converge toward the neighborhood of the global optimum, provided the initial population size is sufficiently large (ibid.).

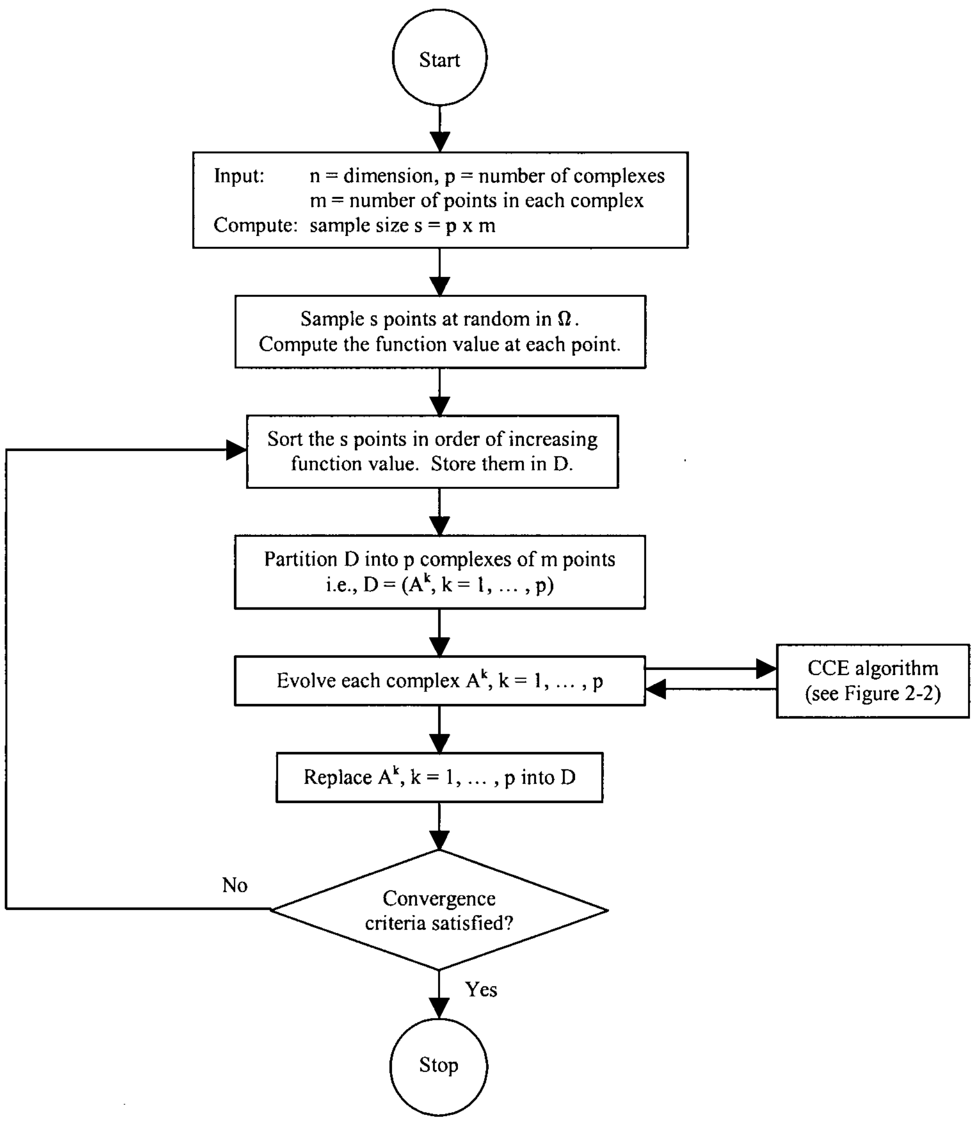


Figure 2-1: Flowchart of the SCE-UA Algorithm
(from p. 1027, Duan et al., 1992)

The separation of points into independent complexes allows each complex to explore the response surface in different directions. The evolution of each complex is driven by the "competitive complex evolution" (CCE) strategy, in which the majority of new offspring are generated by the Nelder-Mead procedure (Nelder and Mead, 1965). Duan et al. (1992) provide a detailed explanation of how each complex is evolved; Figure 2-2 is a flowchart of the CCE strategy excerpted from their work.

The various process parameters of the algorithm control various aspects of the optimization. These parameters can influence the efficiency and effectiveness of performance, and the user must ensure that all SCE-UA parameters are assigned appropriate values to maximize the probability that the algorithm will succeed. Duan et al. (1994b) describe the various parameters in detail and explore the sensitivity of the algorithm to parameter values.

The SCE-UA method may not always be able to reach global convergence, but usually produces at least near-optimal results (Gan and Biftu, 1996). In the decade since it was first developed, the SCE-UA method has been studied extensively with an emphasis on establishing its performance relative to other established approaches for optimization. The consensus in the literature is that the SCE-UA method provides solutions that are equal to or better than all other procedures for automatic calibration of hydrologic models (Burges, 2003; Gan and Biftu, 1996; Singh and Woolhiser, 2002; Sorooshian and Gupta, 1995). It has been compared favourably to Genetic Algorithms (Cooper et al., 1997; Kuczera, 1997), semi-automated methods (Gupta et al., 1999), Pattern Search and Sequential Quadratic Programming (Franchini et al., 1998), Simulated Annealing (Cooper et al., 1997; Thyer et al., 1999), and the Multi-Start Simplex Method (Duan et al., 1992; Gan and Biftu, 1996; Sorooshian et al., 1993). Although the algorithm typically requires tens of thousands of executions of the hydrologic model, most studies also conclude that the SCE-UA method is the most efficient automatic calibration tool available.

Despite the successes of the SCE-UA method in the realm of automatic calibration, Thyer et al. (p. 773, 1999) caution that "SCE-UA should not be viewed as a panacea". The SCE-UA method is still subject to general challenges like poor model formulation, unrepresentative or error-laden data, non-uniqueness of optimal solutions, and sensitivity to objective function.

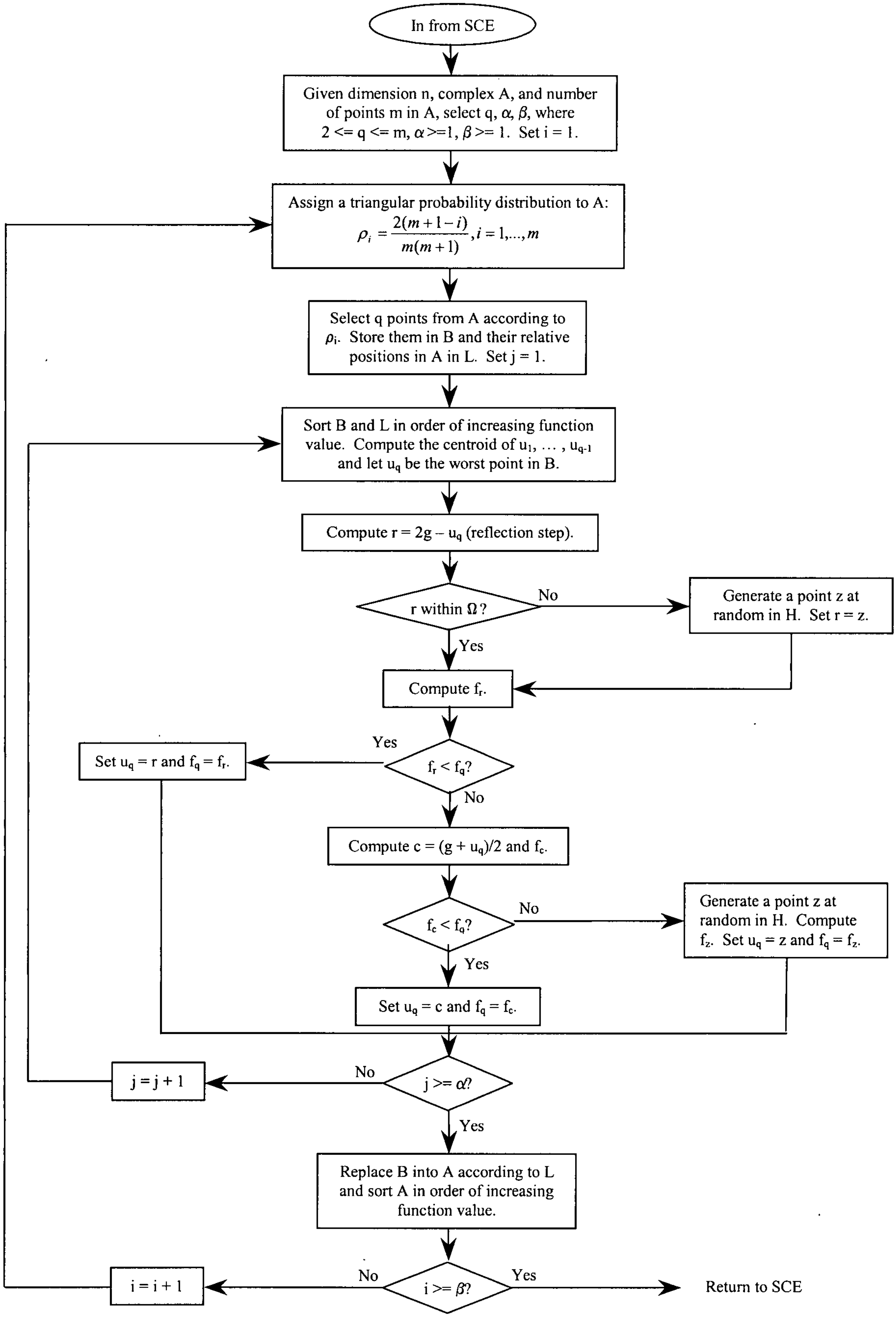


Figure 2-2: Flowchart of the SCE-UA CCE Strategy
(from p. 1028,Duan et al., 1992)

Gan and Biftu (1996) suggest that these issues imply a critical need to evaluate the global optimization features of the SCE-UA method when it is applied to a real-world catchment, since even the basic concept of a global optimum may not make sense in an applied context.

It is relevant to emphasize again the general distinction between methods which attempt to identify the global optima of an objective function and methods aimed at attaining the best possible representation of a natural system. Even where a unique global optimum can be found for a specific objective function, both the global optimum and the simulation quality still depend on the choice of objective function as well as the various model and data uncertainties. Vrugt et al. (2003) point out that, while the SCE-UA method is able to identify the global minimum for a problem, it has had little success in driving forward the quest for a unique "best" parameter set.

2.4.5 Evaluating Model Performance

Model performance is evaluated during calibration to provide direction to the search for a better representation of the natural system. Performance is then re-evaluated during validation to support the hypothesis that the model is correctly representing the prototype. Establishing "good" model performance usually involves attaining an "acceptable" goodness of fit between model output and historical record under the assumption that conditions of application will be similar to those experienced during calibration and validation (Klemeš, 1986b). Such evaluations of model "correctness" apply solely to the model output and cannot be extended to imply "hydrological soundness" of the model structure (ibid.).

Madsen (2000) proposes using four basic factors to evaluate a hydrologic model: water balance (average catchment runoff volume), hydrograph shape, timing / rate / volume of peak flows, and low flow agreement. Model calibration guidelines given by Burges (2002) focus on the following:

- maintaining annual, seasonal, weekly, and daily water balance;
- matching hydrograph shape, peak values, and peak timing;
- ensuring predicted ET $\leq$potential ET for the region;
- verifying simulated water storage fluctuations with precipitation patterns;
- confirming parameter values are consistent with catchment properties;

- matching surface flow to base flow ratio to soil and geological conditions;
- keeping comparisons consistent with the accuracy and errors of recorded data.

Objectives typically have different units of measurement and often cannot be aggregated into a single measure. Such objectives are called non-commensurate. Choosing a solution based on multiple objectives usually requires trade-offs between non-commensurate objectives (Revelle et al., 1997).

In hydrologic model calibration, no criteria or set of criteria for measuring model performance can be declared incontrovertibly superior to all other techniques (Klemeš, 1986b; Sorooshian et al., 1983). In a study assessing the performance of several models using multiple categories of criteria, Houghton-Carr (1999) finds that no model performs well in all categories, and no single category adequately describes model performance. The criteria chosen should therefore be related to the purpose to which the calibrated model will be applied (Klemeš, 1986b). Manual calibrations and some relatively recent automatic calibration algorithms (e.g., Bastidas et al., 2001) consider multiple criteria in an attempt to reduce the reliance on any single measure of performance.

Approaches to evaluation can be quantitative, qualitative, or both (Houghton-Carr, 1999). Often, a combination of statistical, graphical, and intuitive measures are used to evaluate a given model. For example, Gan (1987) utilizes statistical and graphical comparisons, while also considering physical plausibility of parameter values given the catchment characteristics and calibration data. Houghton-Carr (1999) finds that quantitative and qualitative approaches provide different information: while qualitative criteria are ambiguous, quantitative criteria demonstrate a relationship between flow regime and performance.

Reliance on any one approach to the exclusion of others is inadvisable. For example, subjective graphical analysis may not clearly reveal consistent biases. Perhaps more importantly, quantitative comparisons of paired values from two time can easily result in unjustified and misleading levels of error if model or data timing is wrong or uncertain. If timing is slightly off, large errors are created for rising and falling limbs. Statistical comparison of two such time series could result in a high level of error not supported by graphical analysis. As an alternative, Burges (2002) proposes calibration based on storm volumes and peak flows as a multi-objective approach. Bergström et al. (2002) make allowance for uncertainty in modelling nitrogen levels by considering the best match of observed and simulated nitrogen concentrations within ±3 days of the nominal date.

Although graphical comparisons are usually concerned with evaluations of simulated and observed hydrographs, other types of both quantitative and qualitative data can be expressed graphically. Different graphical techniques can illustrate the properties and patterns of a data set (e.g., box plots, scattergrams, transformations). Houghton-Carr (1999) makes good use of graphical comparison by plotting two different performance measures on the x- and y-axes, while Hogue et al. (2000) improve the representation of recessions while maintaining perspective for higher flows by applying a partial log transformation.

Graphical analysis can also help to clearly and concisely organize related data. Plotting the time series of rainfall, catchment outflow, and residual daily flow volume error on a single graph is an effective way of assessing performance and is particularly useful in highlighting systematic errors in the model (Burges, 2002; Gan, 1987).

While manual calibrations use both quantitative and qualitative criteria to evaluate model performance, automatic calibration algorithms rely exclusively on quantitative calculations. Quantitative measures can be as simple as direct comparison of flow volume, peak flow, or time to peak (Loague and Freeze, 1985). More complex quantitative measures such as maximum error, Root Mean Square Error (RMSE), Coefficient of Determination (R^2), efficiency, and coefficient of residual mass are typically based on the relationship between each quantile of the observed time series and its simulated counterpart. While usually applied to the entire time series, consideration of a subset of results can provide insight into the performance of a specific model component.

The standard approach to quantitative evaluation involves defining some measure of the length of a vector composed of the time series of residual errors (Gupta et al., 1998). The process of calibration is then viewed as an iterative attempt to find values of the model parameters that minimize the "length" of the vector. The process is complicated by the lack of an unambiguously correct way to define the length of the error vector (Gupta et al., 1998; Lan, 2001). Further, any statistical measure explicitly comparing the residuals of two time series arguably places more weight on timing than on matching flow patterns (Burges, 2002). Todini (1988) points out that statistical techniques based on analysis of residuals neglect the physical characteristics of the model, avoiding rather than taking advantage of prior expert knowledge intrinsic to the model structure.

Popular measures for quantitative evaluation include coefficients of linear correlation (e.g., the coefficient of determination, R^2), coefficients of efficiency (e.g., Nash-Sutcliffe efficiency, E!), and various equations related to regression and curve fitting theory (e.g., simple least squares, SLS) (Lan, 2001). A wide variety of statistics are available, although an individual user will generally have a set of preferred measures. Many users of the UBCWM prefer a modified version of Nash-Sutcliffe efficiency that corrects for overall volume error (e.g., Lan, 2001; Micovic, 1998). The US National Weather Service uses a wide variety of measures, including Daily Root Mean Squared Error, Total Mean Monthly Volume Squared Error, Mean and Maximum Absolute Error, Nash-Sutcliffe efficiency, Bias (mean daily error), Peak Difference, First Lag Autocorrelation, and Number of Sign Changes (i.e., the number of times the sequence of residuals changes sign) (Gupta et al., 1998).

Some of the more common functions used in evaluating hydrologic model performance are presented in Table 2.1, with descriptions, numerical formulations, and examples of studies in which each has been applied. These common functions generally each require a set of assumptions regarding the statistical distribution of the output data errors, while errors in input data are typically ignored (Gupta et al., 1998).

The coefficient of determination (R^2 or D!) is the square of the coefficient of linear correlation, a popular measure of the strength of relationship between two sets of data. For independent data with no correlation whatsoever, R and R^2 are zero; for highly-correlated data sets, R and R^2 approach unity. The coefficient of determination is defined as the quotient of explained variation and total variation. Lan (2001) and Seibert and McDonnell (2002) note the difference between the case of perfect correlation ($R^2 = 1$) and that of perfect simulation (i.e., zero error, where $q_{sim} = q_{obs}$). This difference suggests that R^2 alone is not a good measure of data similitude for an automatic calibration against a single data series. The low correlation (0.14) between the

Table 2-1: Common Measures for Quantitative Evaluation of Hydrologic Model Performance

Symbol / Acronym	Full Name	Description	Numerical Formulation	Example Applications
R^2 (D!)	Coefficient of Determination	Proportion of variance in q_{obs} (x) that can be explained by the modelled q_{sim} (y)	$R^2 = \frac{[n\Sigma xy - (\Sigma x)(\Sigma y)]^2}{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}$ (1)	(Lan, 2001) (Micovic, 1998) (Franchini and Pacciani, 1991)
HMLE	Heteroscedastic Maximum Likelihood Estimator	Sum of weighted squared errors ($w_t \cdot \epsilon_t$). Assumes errors are Gaussian and heteroscedastic, with zero mean and covariance matrix $\mathbf{V} = \sigma^2 \cdot \boldsymbol{I}$	$HMLE = \left\{\sum_{t=1}^{n} w_t \varepsilon_t^2\right\}\left\{n\left[\prod_{t=1}^{n} w_t\right]^{1/n}\right\}^{-1}$ (2)	(Gan, 1987) (Sorooshian et al., 1983)
LAD	Least Absolute Difference; Absolute Least Value	Sum of the absolute values of the errors (i.e., $q_{obs} - q_{sim}$) for each pair of q_{obs} and q_{sim}	$LAD = \sum_{t=1}^{n} \lvert q_{obs} - q_{sim} \rvert$ (3)	(Gan, 1987) (Hornberger et al., 1985) (Houghton-Carr, 1999)
E!	Nash – Sutcliffe Efficiency	A variant of the Coefficient of Determination that measures the relative magnitude of "noise" ($q_{sim} - q_{obs}$) to "information" ($q_{obs} - \overline{q_{obs}}$)	$E! = 1 - \frac{\sum_{t=1}^{n} [q_{sim} - q_{obs}]^2}{\sum_{t=1}^{n} [q_{obs} - \bar{q}_{obs}]^2}$ (4)	(Cooper et al., 1997) (Lan, 2001) (Micovic, 1998) (Gupta et al., 1999) (Houghton-Carr, 1999)
RMSE; DRMS	(Daily) Root Mean Square Error	The square root of the mean of squared errors (i.e., $q_{obs} - q_{sim}$). It is sometimes referred to as DRMS when calculated for a daily time series.	$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (q_{obs} - q_{sim})^2}$ (5)	(Cooper et al., 1997) (Gupta et al., 1999) (Hogue et al., 2000)
SLS or SSE	Simple Least Squares or Sum of Squared Errors	A simple summation of the squared errors (i.e., $q_{obs} - q_{sim}$) for each point in time series *t*	$SLS = \sum_{t=1}^{n} (q_{obs} - q_{sim})^2$ (6)	(Cooper et al., 1997) (Gan, 1987) (Hornberger et al., 1985) (Houghton-Carr, 1999) (Sorooshian et al., 1983)

coefficient of determination (D!) and residual volume error (dV) observed by Micovic (Micovic, 1998) supports this conclusion.

The Nash-Sutcliffe efficiency statistic (E!) measures the proportion of variability in the observed flow series that is explained by the hydrologic model (Loague and Freeze, 1985). Nash-Sutcliffe efficiency measures the relative magnitude of noise (q_{sim} - q_{obs}) to information (q_{obs} - q_{mean}). E! is better suited to evaluation of hydrologic simulations than the coefficient of determination because it evaluates the magnitude and shape of difference between observed and simulated hydrographs.

However, the Nash-Sutcliffe E! is biased for data sets with large total variance for which the residual variance (in the numerator) is dominated by the observed variance (in the denominator) (Lan, 2001). Large events are weighted more heavily in the sense that residual error for a large event will have much more effect on the statistic than the residual error for a small event, even if the two are the same as percentages of event magnitude. The poor reproduction of low flow periods in Seibert (2000) is attributed to the bias of the Nash-Sutcliffe efficiency statistic toward matching high-flow events at the expense of low-flow events. Loague and Freeze (1985) observe that, in general, Nash-Sutcliffe efficiencies are better when calculated for volume and peak flow than for peak flow timing. Recent work has shown that improvement is possible in some cases by calculating the Nash-Sutcliffe Efficiency for the natural logarithms of the discharge time series (i.e., E! = f[ln(Q)]) (Weiler, 2005).

In a Genetic Algorithm calibration of the UBCWM, Lan (2001) demonstrates that, across the 20th generation population, increasing E! is not necessarily associated with decreasing relative volume error *dV/V*. To emphasize conservation of mass and closing of the water balance, the UBCWM User's Manual (Quick, 1994) presents the modified Nash-Sutcliffe statistic EOPT!, defined as:

$$EOPT! = E! - | \sum q_{obs} - \sum q_{sim} | / \sum q_{obs} \quad (7)$$

The EOPT! statistic is used for evaluating the performance of the UBCWM.

The two quantitative criteria most widely used for model evaluation in the literature are the Simple Least Squares Error (SLS) and the Root Mean Square Error (RMSE) statistics. The SLS statistic is the sum of the squares of the individual residual errors (i.e., SLS = $\Sigma(q_{obs} - q_{sim})^2$). As for other statistics, data error can lead to problems when applying SLS since equal weight is given to both valid and erroneous quantiles in the time series. However, for SLS, any distortion is exacerbated through the squaring of large residual errors (Lan, 2001). Least absolute difference functions are better suited to erroneous data because the errors are not squared before summation.

The RMSE statistic is defined as the square root of the sum of the individual residual errors, effectively computing the standard deviation of the model prediction error (Gupta et al., 1999). When residual errors are calculated at a daily timestep, RMSE is sometimes labeled Daily Root Mean Square Error (DRMS). The RMSE statistic assumes the presence of Gaussian, independent error with homogeneous variance; the resulting implicit bias toward matching the highest events of record renders the RMSE statistic poorly applicable to data consisting mostly of low flows with a few large events (Gan and Biftu, 1996; Yapo et al., 1996). The tendency of the RMSE objective function to provide good peak estimates while allowing strong bias in other parts of the hydrograph is readily apparent at intermediate steps of the calibration process (Hogue et al., 2000).

By assuming homogeneous variance in error, SLS and RMSE imply acceptance of the argument proposed by Arnaud et al. (2002): that since the magnitude of error is not increasing with size, absolute error is constant and therefore relative error in the data should be expected to decrease for larger events.

If the accuracy of low-flow simulation is a priority, the Heteroscedastic Maximum Likelihood Estimator (HMLE), discussed in detail by Sorooshian et al. (1983), is a better choice than the RMSE or SLS statistics. Yapo et al. (1996) improve low-flow simulation by using the HMLE as an objective function instead of RMSE, although performance on higher flows deteriorates. The HMLE assumes output errors are Gaussian with zero mean and uncorrelated, heterogeneous variance. The heteroscedastic basis of the HMLE allows for the existence of magnitude-dependent errors in streamflow measurement (Sorooshian et al., 1983; Yapo et al., 1996). Both the RMSE and the HMLE require the independent and identically distributed designation for residual errors; in neither case is the assumption of error independence supported (Yapo et al., 1996). Independent calibration using the RMSE and HMLE criteria can result in drastically different values for many of the parameters (Sorooshian et al., 1993).

In the limiting case of homogeneous variance, the HMLE statistic reduces to the SLS. In a study comparing SLS and HMLE, Sorooshian et al. (1983) attribute the better calibration-period statistics obtained with the SLS calibration to its strong curve-fitting abilities; the HMLE statistics are found to be superior for the validation period. One implication of these findings is that automatic calibrations using the SLS statistic may sacrifice physical realism in favour of fitting the calibration data set. Measures with a poor physical basis should be avoided if possible, as they can easily lead to overfitting (Klemeš, 1982).

2.4.6 Hydrologic Model Validation

Many models can be calibrated to approximate a given data set without correctly representing the *in situ* processes. Therefore, a model must be tested on a distinct and independent data set to confirm that the uncertainty in applied model predictions is acceptable (Klemeš, 1986b; O'Connell and Todini, 1996). This process is most commonly referred to as "model validation". Elsewhere in the literature, the process is referred to as "confirmation" or "model performance evaluation" (e.g., Gupta et al., 1999).

Model validation is also sometimes referred to as "verification" (Gupta et al., 1999). Although the terms "validation" and "verification" are often used interchangeably, their respective meanings are distinct. A "validated" model must be internally consistent and have no known or detectable flaws. In contrast, a "verified" model is one whose truth has been demonstrated (Oreskes et al., 1994). It may be convenient to think of "verification" as the limit state reached as the quality of the validation process approaches perfection and the size and scope tend to infinity.

Hornberger and Boyer (1995) contend that establishing truth is not and should not be the goal of validation, since even a perfect match between observed and predicted data does not verify a model. Oreskes (p. 642, 1994) argues that verification is only possible for "closed systems in which all the components of the system are established independently and known to be correct". Hydrologic models cannot constitute closed systems due to factors such as incomplete knowledge, continuum theory, and scaling of non-additive properties (ibid.). Therefore, hydrologic models cannot be truly verified.

Typical criteria examined during hydrologic model validation include mathematical rigour, absence of bias, and closeness of fit. Obviously, success with these criteria does not necessarily imply scientific objectivity or hydrological insight (Klemeš, 2000a). For Lei and Schilling (p. 81, 1996), a successful validation means only that "the model is not rejected for this very task in this very situation". In this sense, a successful validation should be considered as supporting evidence for, rather than proof of, an acceptable representation of reality (Bergström et al., 2002). While the probability of a correct model representation improves with increasing diversity of validation data, there is little basis for extrapolating conclusions beyond the temporal, spatial, and magnitude limits of the validation data set (Beven, 1989; Klemeš, 1986b; Oreskes et al., 1994). Thus, the data used for validation should ideally be hydrologically similar to conditions expected in the applied simulations.

Although validation cannot conclusively verify a given model, poor model performance, as measured against observed data in the validation phase, is an obvious indicator of error. However, the results of validation are commonly insufficient to conclusively support or refute a model, especially where validation data are limited (Goodrich and Woolhiser, 1994; Oreskes et al., 1994). For models that purport to simulate internal catchment responses, comparison of outflow hydrographs is an insufficient test of validity (Beven, 1989). Bergström et al. (2002) recommend that measurements other than discharge be considered to further validate the model. Where validation is inconclusive, the corresponding lack of confidence should be reflected in results.

The quality of simulation may appear to deteriorate from calibration to validation given the curve-fitting nature of the calibration process and the assumed independence of the validation data set. Given a successful calibration, the change in performance from calibration to validation should be modest to negligible (e.g., Gan and Biftu, 1996; Gan et al., 1997; Madsen, 2003). Gan and Biftu (1996) find that model performance deteriorates more between calibration and validation when "stronger" automatic calibration approaches such as the SCE-UA method, are used. This may be an example of "overfitting" a data set with a powerful optimization tool.

There is often a temptation to return to the calibration phase or otherwise alter parameter values to improve the fit of the model to the validation data. However, such contamination of the validation process automatically precludes its success (Oreskes et al., 1994). Beven (1989) notes a further limitation for event-based models, since if boundary or initial conditions must be calibrated for a validation event, it cannot be considered a true validation. If a continuous model is being used in an independent validation test, allowance must be made for a "spin-up" period to avoid any influence of initial conditions on the validation period.

Loague and Freeze (1985) assert that, in the case of a successful calibration and in the absence of overfitting of parameters or validation beyond the calibration range, errors in calibration and validation data should be statistically similar. The most common approach to hydrologic model validation applies this principle to the so-called "split-sample" test. The split-sample test involves the subjective division of a single record into two parts with one used for calibration and the other reserved for validation. If the record length is not sufficient to be split 50/50, the record should be split twice, e.g., 70/30 and 30/70, and validation must pass on both sets for the model to pass this test (Klemeš, 1986b). Merz and Blöschl (2004) present a case study in which they use Klemeš' (1986b) concept of split-sample validation to good effect.

A split-sample test is not necessarily reliable in all cases (e.g., Gupta et al., 1999), especially given the typical lack of independence between the calibration and validation data series. If the split-sample test is suspect, a more in-depth analysis of model residuals is likely required. Such situations are not addressed herein. However, it is worth noting that qualitative analysis must also play a role in model validation, as parameter values must be feasible, realistic, and able to lend a degree of certainty to simulations and forecasts (Sorooshian et al., 1983).

3. Uncertainty in Hydrologic Modelling

"It is far better to foresee even without certainty than not to foresee at all."

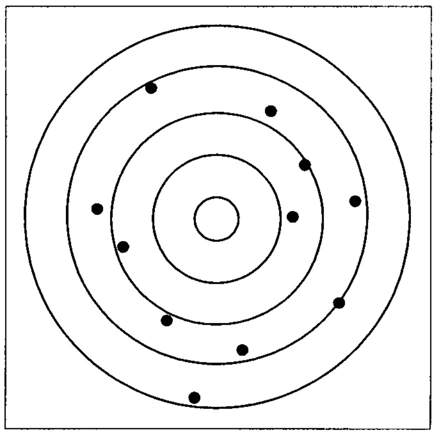
- Henri Poincare

This chapter is intended for both beginning and experienced hydrologic modellers seeking to understand the various ways in which uncertainty is introduced into the model output. Those familiar with the body of literature on uncertainty in hydrologic modelling may wish to proceed directly to Chapter 4.

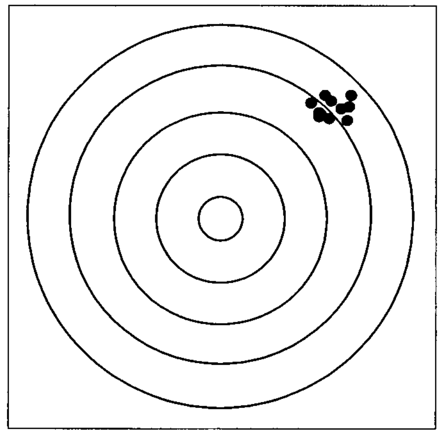
Despite significant progress, hydrologic models are still far from achieving desired levels of accuracy and certainty. Often, the best that can be achieved under ideal conditions is to predict behaviour with a certain degree of "confidence" – the complement of which is a degree of residual uncertainty. Any output from a hydrologic model should therefore be analyzed with regard to its significant in the context of the associated model predictive uncertainty (Binley et al., 1991). However, uncertainty itself is often indeterminate or unidentifiable; it can be quantitative, qualitative, or unknown in character (Lei and Schilling, 1996). Different aspects of model behaviour may have different degrees of uncertainty, and identifying the sources of uncertainties in model results can be extremely difficult (Garen and Burges, 1981). This chapter provides an overview of uncertainty in the context of hydrologic modelling.

A distinction must be made between the concepts of error (or accuracy) and uncertainty. Although sometimes used interchangeably, the two concepts are distinct in nature. "Error" is the difference between a computed or measured value and its true or theoretically correct counterpart. Therefore, the term "error" is properly used when the observed response, process, or outcome has been observed to be incorrect or invalid. "Uncertainty", on the other hand, is the condition that the computed or measured output *may* differ from the baseline response in magnitude, process, or probability. It is usually expressed in a relative or probabilistic sense, as opposed to error, which is an absolute quantity.

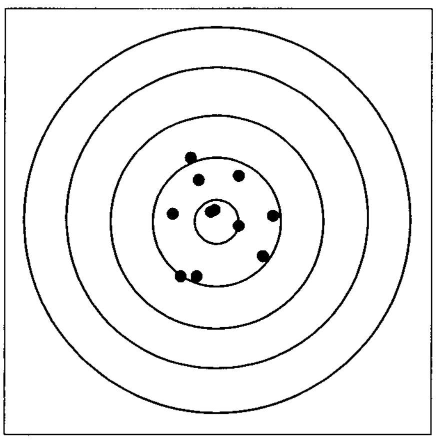
The concept of uncertainty is analogous in some ways to "precision" in the well-known example of accuracy in marksmanship. As illustrated in Figure 3-1, a result can be accurate but not precise; similarly, it can be precise but not accurate. By extension, model results can be accurate but uncertain, or certain but erroneous. Ideally, both certainty and accuracy will be present; more commonly, neither can be established conclusively.



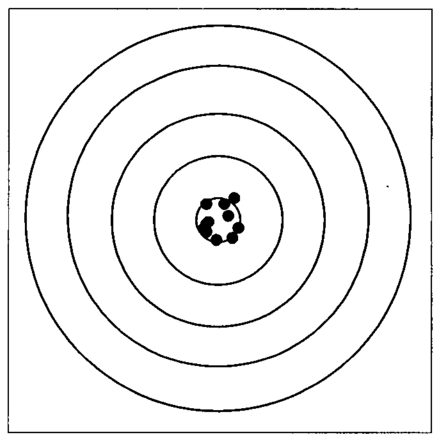
(a) inaccurate and imprecise



(b) precise but not accurate



(c) accurate but not precise



(d) accurate and precise

Figure 3-1: Accuracy and Precision as Analogues for Error and Uncertainty

Applied modellers sometimes tend to focus on the more easily-identifiable issue of model accuracy at the expense of giving due consideration to uncertainty. Indeed, if modelling experts struggle to understand the uncertainties implicit in their models, it would be naïve to assume that practicing hydrologists will give sufficiently comprehensive consideration to uncertainty in their results (Grayson et al., 1994b).

Large or difficult-to-resolve uncertainties in results can cause users to exert pressure on the model developer to "improve" the model (Klemeš, 1982). If "improvements" (e.g., more data, better understanding) are not easily achievable, some modellers resort to attempting to extract more information from the data through mathematical manipulation (ibid.). Model developers and promoters should ensure that there is an appropriate level of awareness and discussion of modelling uncertainties amongst the community of users (Grayson et al., 1994b).

A thorough understanding of the sources of uncertainty and a consistent taxonomy is necessary to facilitate awareness and discussion in the modelling community. These topics are explored in Section 3.1.

Substantial uncertainty will persist in the outputs of even a well-calibrated hydrologic model (Binley et al., 1991). A discussion of the interaction between the calibration process and model predictive uncertainty is contained in Section 3.2.

Although all aspects of uncertainty cannot be measured objectively, a number of approaches exist for exploring the impact of uncertainty on a model or decision. An overview of various methods for analyzing uncertainty is provided in Section 3.3.

Finally, the previously-discussed caveats of model extrapolation imply that uncertainty must increase conspicuously when simulating extreme events. Specific considerations of uncertainty for extreme event simulation are discussed in Section 3.4.

3.1 Classification of Uncertainty in Hydrologic Modelling

Over time, many classification schemes for uncertainty have been proposed. Lohani et al. (1997) provide perhaps the most fundamental, proposing that two basic types of uncertainty exist: what is not known at all, and errors in what is known. Although highly significant, such a perspective is philosophical in nature and does not address the more practical question of sources of uncertainty.

Beck (1987) proposes classifying uncertainty according to the processes through which it is introduced into the problem formulation, i.e., through prior assumptions and knowledge, model identification, and prediction. This approach is much better suited to understanding the sources of uncertainty in hydrologic modelling, but is still somewhat general in nature.

Vicens et al. (1975) consider uncertainty as belonging to one of two basic categories: natural and informational. The US National Research Council (p. 41, NRC, 2000b) has adopted the same perspective, defining natural variability and knowledge uncertainty as follows:

> *"Natural variability - sometimes called aleatory uncertainty - deals with inherent variability in the physical world; [...] In the water resources context, uncertainties related to natural variability include things such as stream flow, assumed to be a random process in time, or soil properties, assumed to be random in space. Natural variability is also sometimes referred to as external, objective, random, or stochastic uncertainty."*

> *"Knowledge uncertainty - sometimes called epistemic uncertainty - deals with a lack of understanding of events and processes, or with a lack of data from which to draw inferences; by assumption, such lack of knowledge is reducible with further information. The word epistemic is derived from the Greek "to know." Knowledge uncertainty is also sometimes referred to as functional, internal, or subjective uncertainty."*

The NRC (2000a) observes that these two uncertainties affect calculations of risk differently, and cautions that the two should be clearly distinguished in practice. Nonetheless, the distinction between the two is somewhat arbitrary and hypothetical in nature. Perception and context typically govern the distinction, since different assumptions may cause natural uncertainties to become knowledge uncertainties and vice versa (NRC, 2000b).

Vicens et al. (1975) divide informational (knowledge) uncertainty into two basic components, corresponding to uncertainty in model structure and in parameter values. Later studies add a third component corresponding to uncertainty arising from uncertainty or error in observed data (e.g., Garen and Burges, 1981; Loague and Freeze, 1985). Every hydrologic study is subject to varying degrees of all three components. These three components are collectively labelled model predictive uncertainty.

Melching et al. (p. 2275, 1990) summarize the resulting four basic classifications of uncertainty as follows:

- natural variability, which refers to "the random temporal and areal fluctuations inherent in natural processes";
- data uncertainty, which includes measurement inaccuracy and errors, the adequacy of the data to represent *in situ* conditions, and any data handling, transmission, or transcription errors;
- model parameter uncertainty, which reflects "variability in the determination of the proper parameter values to use in modelling a given causative event"; and
- model structure uncertainty, which characterizes "the ability of the model to accurately reflect the watershed's true physical runoff process".

Critical analysis of this taxonomy encourages exploration of relationships between specific sources of uncertainty in hydrologic modelling. The mind map shown in Figure 3-2 is the result of one such analysis. The figure shows the central goal of improving estimation of extreme events, surrounded by various issues that inhibit progress. Major issues connect to the goal as trunks, with sub-issues specified as branches. The major areas of focus are analogous to the types of uncertainty discussed above, with event uncertainty corresponding to natural variability, and model uncertainty including both model structural and parameter uncertainty. Data uncertainty is separated from the other knowledge uncertainties because of its dominant role in most situations. Four connections are shown between the various regions of the map (i.e., lack of extreme event data, model extrapolation, calibration, and physical modelling), representing concepts that span the distinction between uncertainties.

For example, while calibration is primarily used to reduce parameter uncertainty, the process implicitly attempts to account for any erroneous or uncertain data. Physically-based models attempt to reduce the uncertainty associated with lumping data by adopting a distributed framework and detailed process representation. Model extrapolation typically involves application of a validated structure of process dynamics under arbitrary conditions far

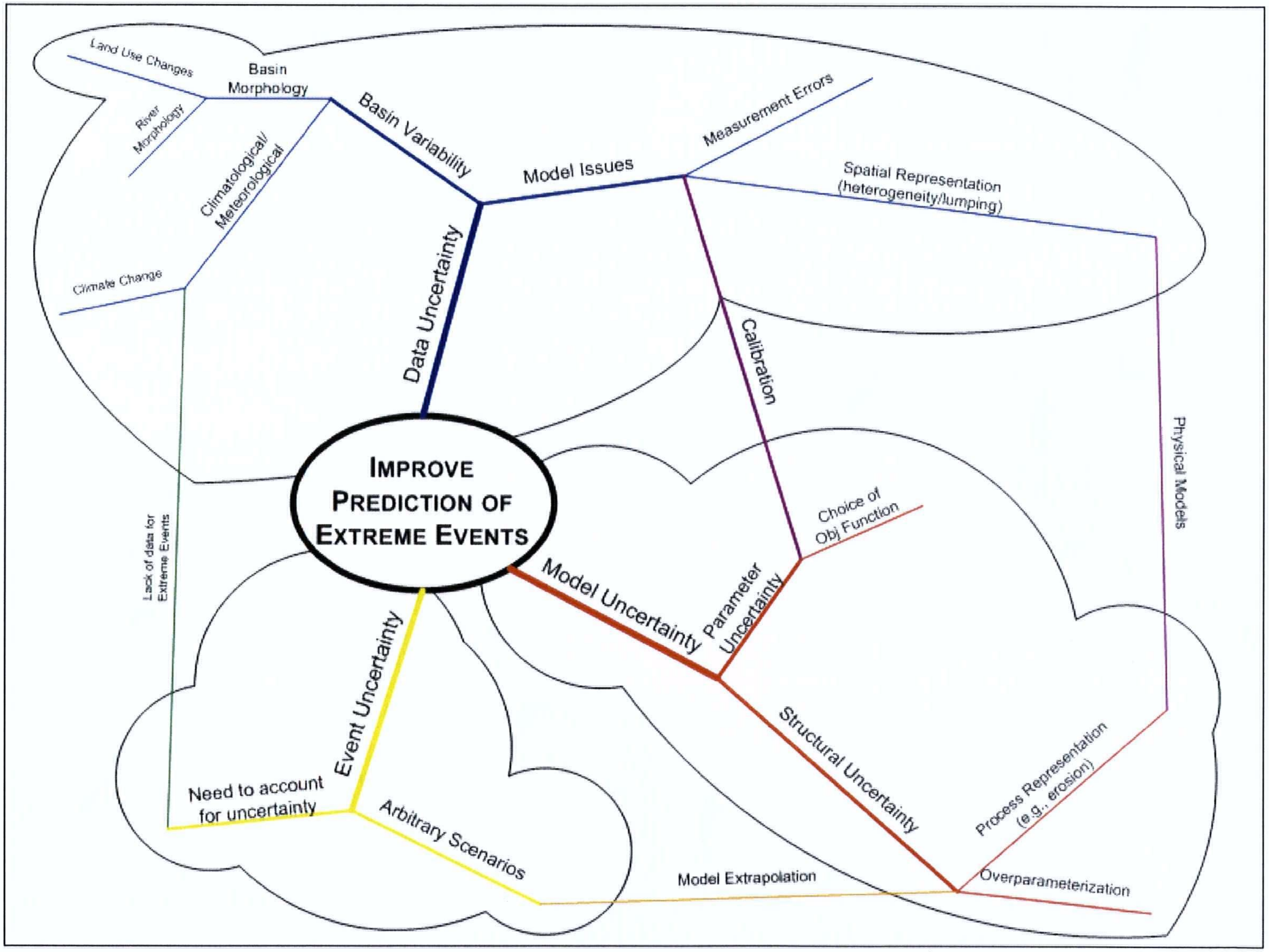


Figure 3-2: Mind Map of Uncertainties in Hydrologic Modelling

removed from those of validation. Finally, the lack of extreme event data can be classified as a function of both data uncertainty and natural variability.

Further discussion of how various sources of uncertainty fit into each of the four classifications of natural, data, parameter, and model uncertainty follows in Sections 3.1.1through 3.1.4.

3.1.1 Natural Variability

Natural variability occurs in both spatial and temporal dimensions. It is considered irreducible, since it is a natural characteristic of a given continuum of time and space and is therefore not subject to minimization or "improvement". Uncertainty is associated with natural variability because modellers lack the tools to fully understand, measure, and represent natural variability in hydrologic modelling. For example, Gan et al. (1997) conclude that dry catchments (i.e., those having streamflow/rainfall ratios less than 0.2) are generally more difficult to model than wet catchments due to their greater hydrologic complexity and variability. In this case, the watershed response of dry catchments is more variable. However, it cannot be said to be "less certain", unless one is discussing our ability to reflect or replicate the response through data collection and process modelling.

Natural variability across a catchment can result in data, parameter, or model uncertainties, or any combination thereof. Often, it is a combination of these resultant uncertainties that is responsible for anomalous behaviour. However, the natural variabilities of a given system may also have systematic elements. For example, Arnaud et al. (2002) note that peak flows are much more strongly influenced than average runoff volumes by the spatial and temporal characteristics of the watershed and the prevailing storm. Similarly, Woolhiser (1996) observes an inverse relationship between response variability and rainfall rate. Such observations can help reduce the knowledge uncertainties associated with natural variability.

Characterizing natural variability is a key focus for recent hydrologic research (Smith et al., 1994). Natural variability can be significant at many different scales (Woolhiser, 1996). It can arise from heterogeneous behaviour or properties (e.g., weather, topography), discontinuities (e.g., geological formations; land use), or processes (e.g., rainfall; infiltration) (Singh, 1995b; Song and James, 1992). Although the ability of science to measure, record, and analyze observations has improved, a growing awareness of small-scale hydrologic complexity has not resulted in an improved management approach (Beven, 2000). No authoritative guidance is available for assessments of heterogeneity, and few methods exist for measuring the spatial patterns of hydrologic behaviour (ibid.).

The most common approach to defining natural variability is to use a network or grid of point measurements. Although all point measurements represent an integration over some effective volume, this volume is often small compared with the macro-scale heterogeneity of the process being measured (Beven, 2000). Thus, in many cases the values obtained through field sampling are not representative of the larger hydrologic domain. Multi-point (e.g., grid or network) sampling programs, while generally preferable to single-point samples, may uncover greater heterogeneity and necessitate even more sampling (Grayson et al., 1992b). An unrealistic, statistically large number of samples could ultimately be required to obtain an adequate characterization of conditions (Beven, 2000).

While there are opportunities for using advanced measurement and analysis techniques such as GIS and remote sensing to characterize natural variability, such techniques still require a theory or method of spatial averaging, interpretation, and processing, and may not completely eliminate uncertainty (Beven, 1989).

Characterization of rainfall variability has been shown to have a particularly pronounced effect on the accuracy and quality of model results (e.g., Michaud and Sorooshian, 1994; O'Connell and Todini, 1996; Ogden et al., 2000; Steiner et al., 1999). In a study of rainfall variability across a high-density rain gauge network, Burges (2002) finds that no single gauge adequately describes either average depth (volume) or intensity, and no three gauges adequately describe spatial variability. Steiner et al. (1999) consider the potential for combining radar and rain gauge measurements, ultimately concluding that even in ideal circumstances, the different natures of the two measurements can still result in discrepancies. They find that, in general, the rain gauge data may not be spatially representative, while radar rain measurements may not be temporally representative (ibid.). In this way, natural variability could potentially defy complete and objective characterization even where an abnormally large amount of data is available.

Lumped hydrologic models attempt to account for natural variability by identifying a single set of effective parameter values that characterize the cumulative response of a heterogeneous catchment. This approach may be efficient where data are limited, but is unlikely to produce accurate results. Even distributed models lose detail and smaller-scale variability through integration and averaging over model element areas. Goodrich and Woolhiser (1994) show that the assumption of spatial uniformity, even at scales as small as 300m, is not supported for Walnut Gulch experimental watershed.

Wood et al. (1988) contend that the various processes of hydrology should, in theory, have a measurable catchment size at which variability is at a minimum. Singh (1995b) believes that an appropriate scale must be small enough to capture any significant hydrologic heterogeneity, but not so small as to be dominated by local physical features.

Numerous studies have been directed toward determining the optimal "scale" or area for averaging hydrologic heterogeneity. These attempts have met with varying degrees of success and reached a range of conclusions. Song and James (1992) find that a scale of approximately one square mile produces the best results, with the optimal scale size varying upwards and downwards for gentle and mountainous topography, respectively. Wood et al. (1988) find that variability of response stabilizes for catchment sizes greater than approximately one square kilometre. Grayson et al. (1992b) contend that the magnitude of a representative elementary area is not fixed, but rather will depend on the desired output.

In the final analysis, not all catchments have a characteristic scale, no single characteristic scale is applicable for all catchments, and output based on this characteristic scale will still be approximate. Residual uncertainty in the final result is often ignored for lack of a better alternative (Beven, 1989, 2001).

While spatial heterogeneity has received the bulk of attention in literature, a hydrologic modeller should also be aware of potential temporal effects. For example, bias in data may not be constant over time, and adjustments to the data must reflect this (Steiner et al., 1999). Garen and Burges (1981) identify the seasonal variance in streamflow measurement uncertainty as a prominent example.

3.1.2 Data Uncertainty

It is frequently impossible for hydrologists to resolve processes known to occur in the field with the input and output data available (Hornberger et al., 1985). This is not surprising, given that data uncertainty is usually the most significant component of uncertainty in hydrologic modelling (Kouwen, 2003). In general, data uncertainty can be said to exist when there is a gap between the data itself, either collectively or as individual quantiles, and what those data are assumed to represent. Naturally, this definition implicitly includes all situations where data are simply unavailable and must be inferred, extrapolated, transposed, or otherwise approximated from other locations or sources. Uncertainty in input data can lead to poor reliability or large prediction error for any otherwise-reasonable hydrologic model (Melching et al., 1990).

Obviously, the availability of data is of paramount importance, as uncertainty is highest in data-poor environments. In a study of the simple distributed model THALES, Grayson et al. (1992a, 1992b) use intensively-monitored research catchments having an unusually rich portfolio of data. However, the authors still report problems due to insufficient data. Often, data are simply not available in areas of concern to hydrologists and engineers, and subject catchments are chosen more for their data availability than for their appropriateness to the research purpose. For example, the work of Loague and Freeze (1985) is constrained to small upland catchments because these represent the only scale with sufficient data for their purposes.

Data collection is frequently constrained by logistical issues such as lack of funding or limited relevance to fields other than hydrologic modelling. In particular, Quick (1995) notes the value of both high and low elevation data for modelling orographic effects in mountainous watersheds. However, data collection stations are typically sited for proximity to communities and for easy maintenance access, and thus are usually confined to lower elevations. Eaton et al. (2002) clearly state their opinion that the hydrometric network in general is “grossly inadequate” in the face of increasing global, national, and regional pressures on water resources. To be truly effective, data collection networks should be designed to be of representative location, density, process, and scale, with due regard for the purposes to which the data will be applied.

In Canada and elsewhere, lack of funding often forces data collection agencies to reduce or rotate their monitoring networks. Gauge movement makes data analysis much more difficult, as the associated step-change may render pre- and post-relocation data irreconcilable as a single record (e.g., Hunter et al., 2002). Inhomogeneity can also be introduced into a hydrologic data series when one type of gauge is replaced by another (Sevruk, 1996). For example, Michaud and Sorooshian (1994) expect streamflow measurement accuracy to change by a factor of ten following replacement of broad-crested weirs with super-critical flumes. Other challenges arise where an experimental database spans administrative boundaries, since different governments and organizations commonly use different data collection methods with different standards (Sevruk, 1994).

The resulting (potentially fragmented) data series cannot always be resolved to the single continuous records usually required for hydrologic simulation. This often means that gaps in the data are filled with tenuous estimates or arbitrarily-transposed data (e.g., Michaud and Sorooshian, 1994; Seibert, 2000). Industry and professional associations are beginning to discuss the need for alternate means of maintaining and increasing data collection networks (e.g., Kulkarni and Blais-Stevens, 2003). Beven (2000) believes that one solution is for hydrologists to execute field programs as an integrated part of hydrologic modelling rather than simply adopting whatever data are available.

Steiner et al. (1999) cite data redundancy as the key to quality control. For example, the authors suggest using clusters of at least three rain gauges within a few hundred metres to establish data quality. When redundant data are not available, it is commonly assumed that any available data are of sufficient quality for modelling. This assumption tends to persist even where proof to the contrary can be considered a foregone conclusion (e.g., Michaud and Sorooshian, 1994). However, data error and uncertainty often prove difficult to detect and quantify (Steiner et al., 1999). Steiner et al. (ibid.) present a clear example of data error that is only identifiable because the study was conducted on a highly-instrumented experimental catchment. They report that, for 30 storms over the watershed, average radar rainfall (as measured by four WSR-88 stations) is less than the average rain gauge total for 80% of the storms. For 45% of the storms, the underestimation is at least 20%, and for 30% of the storms, the underestimation exceeds 30%.

Data uncertainties can arise from measurement errors, inconsistent or heterogeneous data, data handling and transcription errors, or non-representative sampling caused by temporal, spatial, or financial limitations (Binley et al., 1991; NRC, 2000b). Even in the absence of obvious mistakes, translation of the data into usable (often digital) format can create uncertainty. Faurès et al. (1995) find that subjectivity in digitizing plotted precipitation data (e.g., operator selection of breakpoints) can lead to peak flow coefficients of variability for simulated outflow in the 3-5% range. Harlin (1992) presents another example, describing how a single precipitation event occurring over two calendar days could be artificially split in a daily data series, making model results difficult to reconcile with observed streamflow.

Melching (1995) expands on the concept of measurement error, citing the possibility for equipment malfunction, non-representativeness of local conditions, and bias due to sampling location. Data uncertainty can also arise from a non-informative data set, where the data do not constitute a sufficient sample set for model calibration, validation, and evaluation (Sorooshian and Gupta, 1983).

The earliest data collection programs relied exclusively on manual data collection. Although manual data collection has not disappeared altogether (e.g., the network of amateur meteorological stations across the U.S. (National Weather Service, 2003)), most data is now collected electronically. The introduction of technology into data collection yielded both benefits and drawbacks: electronic data can be collected in more detail, often at less cost, and in remote locations. However, many believe that there has been a corresponding loss of qualitative information and data richness, and in some cases, a loss of accuracy (Weiler, 2005).

Electronic data sets collected before telemetry technology became available may be subject to additional uncertainties such as the potential for asynchronicity between precipitation and streamflow gauges (Loague and Freeze, 1985; Melching, 1995). In general, much early data from automatic stations cannot be assumed to exist with the same amount of certainty as modern data. In most cases the historical levels of uncertainty are indeterminate. Any assumptions made about the nature of data error are yet another possible source of uncertainty (e.g., Duchon and Essenberg, 2001).

Uncertainty can also be introduced through the processes of field sampling. For example, Beven (2000) shows that the hydraulic conductivity measured from a fist-sized soil sample in a laboratory, while correct for that sample, does not reflect the larger scale hydraulic conductivity due to the dominance of larger-scale flow pathways *in situ*. More generally, he cautions that measurement techniques can be "invasive with the potential to change the response of the system by the very process of observation" (p. 192, Beven, 2002). The effect of such a disturbance may or may not be permanent. Michaud and Sorooshian (1994) observe a temporary change in streamflow volume residuals following the introduction of new flow measurement structures, likely indicating an adjustment of the sediment regime.

The existence of data uncertainty (i.e., potential error) is often indicated by continual failure of a variety of models and calibrations to adequately represent the response of a catchment, or by abnormally poor model performance for a specific event or period of record. For example, Franchini and Pacciani (1991) report that all models in their comparison study have poor results for certain events. In their case study, the persistence of problems across a diverse set of models indicates that the problems likely originate in the data. Alternatively, Gupta et al. (1999) compare results for three different calibrations of SAC-SMA, observing poor performance for all calibrations in four of the eleven years simulated. In this case, data uncertainty is a likely contributor, but model uncertainty may also play a role. In another example, Hornberger et al. (1985) report that almost all of the squared error in the model residuals occurs in the largest event of record. Closer examination of this event reveals a large discrepancy between measured precipitation and runoff volumes, leading the authors to conclude that the error is a result of uncertainty in the data.

Perhaps the most important step in dealing with data uncertainty is to ensure that any preconceived expectations of model accuracy are realistic. Faurès et al. (1995) caution against expecting the accuracy of model outputs to exceed the resolution of input data. Smith et al. (1994) believe that computer model results will not be any more precise than the repeatability of a controlled physical experiment. The implications of their statement are wide-ranging given the impossibility of repeatable large-scale physical experiments in hydrology. Generally speaking, the best result attainable is one in which data uncertainty is insufficient to affect the basic results and conclusions of a study (e.g., Michaud and Sorooshian, 1994).

The acceptability threshold for data uncertainty is in part determined by the nature of the model being applied. For example, using the conceptual, quasi-distributed UBCWM, Lan (2001) finds that good precipitation data from nearby basins can generally be used in the study catchment if the elevation is preserved. However, the UBCWM is commonly applied in alpine regions with very scarce data, and thus must be fully calibrated for each study. One should not expect the same result for models that do not use calibration to attenuate any potential systematic aspects of data error.

As discussed in Section 2.3, simple hydrologic models require simple inputs, often only the dynamic fluxes of temperature, streamflow, and precipitation. More complex physically-based models usually require more complex and comprehensive static data (i.e., physical properties). Both static and dynamic data are subject to uncertainty. However, the uncertainty in the dynamic data dominates in most cases, since measurements of catchment properties are used only for estimating values for their corresponding model parameters. Parameter uncertainty, in its turn, can be reduced through calibration, as discussed in Section 3.1.3. Below are some specific aspects of data uncertainty related to the two hydrologic fluxes that dominate the literature on data uncertainty – streamflow and precipitation.

Streamflow data are generally considered to be more accurate than any other input data with the exception of temperature (Sorooshian and Gupta, 1995). Sorooshian and Gupta (ibid.) cite an expected overall accuracy on the order of ±10%, a figure substantiated by Faurès et al. (1995). Burges (2002) estimates that the best stream gauging stations in the U.S. are likely to be accurate to within ±5%. Moore (2004) provides a similar estimate of accuracy for current metering approaches for streamflow measurement. Even the replicate rainfall events on an impervious surface studied by Wu et al. (1982) measure differences of approximately 4% in peak discharge (Smith et al., 1994).

Because most hydrometric stations record water level and rely on a rating curve to convert measurements to streamflow, the changing accuracy of the rating curve over time is another source of uncertainty. Since large floods often result in an altered channel geometry, it is not surprising that errors in measurements of peak flow events – in cases where the flood did not interrupt data collection completely – are typically much greater than the more general values cited above. The error variance of streamflow measurements and residuals is typically heteroscedastic and tends to increase as the flow gets larger (Sorooshian et al., 1983).

Other methods for streamflow measurement can have their own unique sources of uncertainty. For example, Moore (2004) states that, under good conditions, salt dilution gauging can be accurate to within 5%. However, the accuracy of the method can be affected by the choice of mixing reach, dilution of the injection solution by rainfall, or the presence of vegetation, snow, or ice within the channel (Moore, 2005).

Methods for peak flow estimation such as the slope-area method are sometimes used to augment flow records, particularly in situations where the data are known to be unacceptable. For example, the gauging station at Kickapoo Creek, Texas was destroyed during the 1994 flood. However, peak flow could still be estimated because a video documentary allowed estimation of stage levels against elevation benchmarks on a bridge (Smith et al., 1996). As expected, these kinds of estimates are generally subject to even greater uncertainty than is present in computer modelling; the USGS estimates for the flood peak on Kickapoo Creek have an uncertainty of ±15% (Smith et al., 2000).

The various aspects of data uncertainty in streamflow records discussed above lead Burges (p. 284, 2002) to conclude that "any model that attempts to reproduce the measured hydrograph should include explicitly uncertainty bounds on the modelled streamflow hydrograph".

If data uncertainty can be said to dominate the literature on knowledge uncertainty in hydrologic modelling, then there is little doubt that uncertainty in precipitation plays the same dominant role amongst the components of data uncertainty. Burges (2002) notes that fundamental influence of precipitation, pointing out that systematic errors in the precipitation series preclude correct representation of the water balance, ultimately reducing the credibility of model output.

Precipitation gauges are a well-studied source of measurement error and uncertainty. Steiner et al. (1999) find that all rain gauges within their study catchment operate correctly in only four of the thirty storms examined, and no single rain gauge functions correctly for 100% of the two-year period. They report biological fouling and human interference as common causes of malfunctions for tipping-bucket rain gauges (ibid.). Duchon and Essenberg (2001) observe that in cases of high rain intensity, the time-to-tip for a tipping bucket gauge may become significant. This phenomenon generally is not addressed in the manufacturer's calibration, although after-market calibration is possible. For weighing-type gauges, gauge friction commonly causes the first response recorded during a rain event gauge to occur later than the first tip in the tipping bucket gauge. Similarly, a weighing gauge may register a slight increase in total precipitation hours after the end of the event that might not be captured by an event analysis. Thus, gauge friction is a common source of error for mechanical weighing gauges (ibid.).

An equally important degree of uncertainty in precipitation measurement is related to representativeness. Although the natural aspects of rainfall spatial variability are addressed in Section 3.1.1, one must also consider the ability of a measurement device to capture an accurate sample of local conditions at the gauge site. Even given moderately intense, uniform rainfall over a catchment with functioning, accurate rain gauges, there is no guarantee that results measured by the gauge will be truly representative of local conditions. Factors such as local topography, landscaping, and nearby buildings can have a strong influence. The initial wetting of the sides of the rain gauge bucket can also lead to underestimation of rainfall (Sevruk, 1996). Even the fall angle of precipitation with respect to the ground can have a non-negligible effect, since most hydrologic models assume that precipitation inputs are measured perpendicular to the ground surface. Experiments have shown, however, that the most significant cause of data error in precipitation measurement is wind (Larson and Peck, 1974). Despite its prevalence and potential significance, wind-induced losses are not typically accounted for in published precipitation data (Sevruk, 1996).

Larson and Peck (p. 857, 1974) explain that "as the air rises to pass over the gage, precipitation particles that would have passed through the gage orifice are instead deflected downwind" by turbulence and increased wind speed. The resulting increases in wind speed can exceed 40% (Sevruk, 1996). In this way, high wind conditions can induce significant bias (e.g., Burges, 2002). The magnitude of precipitation undercatch can vary with factors such as wind protection, height above ground, wind speed, precipitation size, precipitation form, and gauge properties (e.g., shape, diameter, and orifice rim thickness).

Different types of wind shields are sometimes attached to precipitation gauges in an attempt to minimize wind-induced undercatch. Although models such as the Atler or Nipher wind shields have been observed to substantially reduce undercatch-related errors for snow, none completely eliminate precipitation undercatch (Duchon and Essenberg, 2001).

Estimates of undercatch are themselves highly uncertain due to the difficulty in obtaining objectively "correct" baseline data. Steiner et al. (1999) advocate obtaining "baseline" data from gauges buried such that their aperture is contiguous with the ground surface. When appropriately protected against in-splash, such "pit" gauges represent the reference standard advocated by the World Meteorological Organization (Sevruk and Nešpor, 1998). However, pit gauges are not widely used even in experimental catchments, and are impractical for measuring snowfall.

Duchon and Essenberg (2001) utilize pit gauges to provide a baseline when estimating wind-induced undercatch for above-ground shielded and unshielded tipping-bucket and Belfort gauges. The authors conclude that it is impossible to establish a direct relationship between wind speed and undercatch, likely due to the absence of any drop-size information in their study. Various studies by Sevruk (e.g., Sevruk, 1996) and Sevruk and Nešpor (e.g., Sevruk and Nešpor, 1998; Nešpor and Sevruk, 1999) have identified a threshold value for rainfall intensity that varies with wind speed: below this threshold value, wind-induced error increases quickly; above it, the increase in wind-induced error is much slower.

Larson and Peck (1974) cite a variety of studies reporting liquid-phase precipitation undercatch from 5% to 20%, with more severe estimates of 40% to 80% for snow. More recently, Sevruk (1996) states that wind-induced loss is generally between 2% and 15% for rain, and up to 80% for snow, with even higher values observed in mountainous areas. Sevruk and Nešpor (1998) and Nešpor and Sevruk (1999) refer to average wind-induced errors of 2%-10% for rain and up to 50% for snow. Duchon and Essenberg (2001) observe the following undercatch values as a percentage of total rainfall (ibid.):

- Tipping bucket above-ground gauges show approximately 4% undercatch with respect to the buried tipping-bucket gauge;
- Weighing bucket above-ground gauges show approximately 5% undercatch with respect to the buried weighing gauge;
- In both cases, the reduction in undercatch realized by implementation of the Alter shield is less than 1%; and

- For extreme meteorological situations combining high wind and rainfall simultaneously (such as squall lines), the above figures increase to 15% undercatch for the tipping bucket gauge, 14% undercatch for the weighing gauge, and a 3% reduction in undercatch using the Atler shield.

Two basic approaches have been used to explore the impact of precipitation uncertainty on hydrologic model simulations. The first utilizes a dense rain gauge network, comparing model results based on a subset of gauges to output obtained using the full network. This approach assesses the reduction in uncertainty associated with increasing gauge density, and can lead to insights for situations in which a dense network is not available. The second approach compares on a relative basis results obtained with baseline precipitation data and results associated with a perturbed version of the baseline data. This second approach is analogous to sensitivity analysis and can provide quantitative insight into the impacts of uncertainty for any extant set of data. Following the second approach, Lumb and Linsley (1971) use mathematical augmentation of actual rainfall to understand the effects of small increases in precipitation on other hydrologic variables and processes. The effect of precipitation augmentation on annual streamflow volumes is found to vary (from zero to approximately 100% of the additional precipitation) with an increasing ratio of annual streamflow to annual precipitation.

To this point, discussion of uncertainty in precipitation measurement has been confined to gauge-based observations. However, uncertainties in radar measurement can be equally or even more significant. Burges (2002) writes that, in many situations, a dense network of reliable gauge data will tend to be more useful than even the best quality radar measurements. Smith et al. (1996) conclude that radar precipitation measurements for the 1995 flood on the Rapidan River are approximately one-third of actual values. The authors link the underestimation to three factors: growth of rainfall through and below the radar beam, inappropriate parameters in the reflectivity conversion calculations, and an inappropriately low hail threshold (ibid.). As the latter two factors indicate, radar estimates are highly dependent on the method used for conversion from reflectivity to rainfall. Measurement error for snow is much greater than for rain, in many cases exceeding a factor of two (Krajewski, 2005). In general, radar measurement does not perform well in mountainous regions (Weiler, 2005).

Radar data are often adjusted using rain gauge data. Arguably the most common procedure is to correct for bias in the radar data by removing the average difference of radar and rain gauge measurements under the assumption that the rain gauge data are reliable (Steiner et al., 1999). However, while this procedure can validate otherwise questionable data, one must recall that *both* sets of data are uncertain and, in all likelihood, neither measurement is accurate. Steiner et al. (ibid.) show that important differences remain even when bias adjustment of radar data is performed using reliable rain gauge data. Uncertainty is compounded where gauge-adjusted radar data are assumed applicable beyond the immediate vicinity of the rain gauge.

3.1.3 Model Uncertainty

The cartoon shown in Figure 3-3 is a good starting point for a discussion of model uncertainty. In this case, the modeller attempts to use a simple coin-toss model to forecast the weather, which is the result of a complex and dynamic set of natural processes. Obviously, the model has the potential to be in error a large percentage of the time, and is therefore subject to substantial model uncertainty.

Some basic properties of the coin-toss model contributing to model uncertainty include:

- the population of model outcomes consists exclusively of "rain" and "shine", each having a long-term probability of 50%, whereas natural conditions for any given day may include rain, shine, or both, mixed in varying proportions;
- successive trials (i.e., coin tosses for each new day) are independent, while natural conditions typically exhibit cyclical behaviour (e.g., frontal systems); and
- the long-term probability of the model forecasting "rain" or "shine" for any given day does not change over time or space, while natural conditions might have seasonal or regional components (e.g., wet / dry seasons or humid / arid climates).

In the context of hydrologic modelling, Gan (1987) lists the simplifying assumptions built into the physically-based Smith-Hebbert model, a hillslope-scale model intended only for research purposes. These assumptions are arguably as unrealistic from a hydrologic perspective as the expectation that a coin toss can be used to predict weather patterns. Thus, substantial model

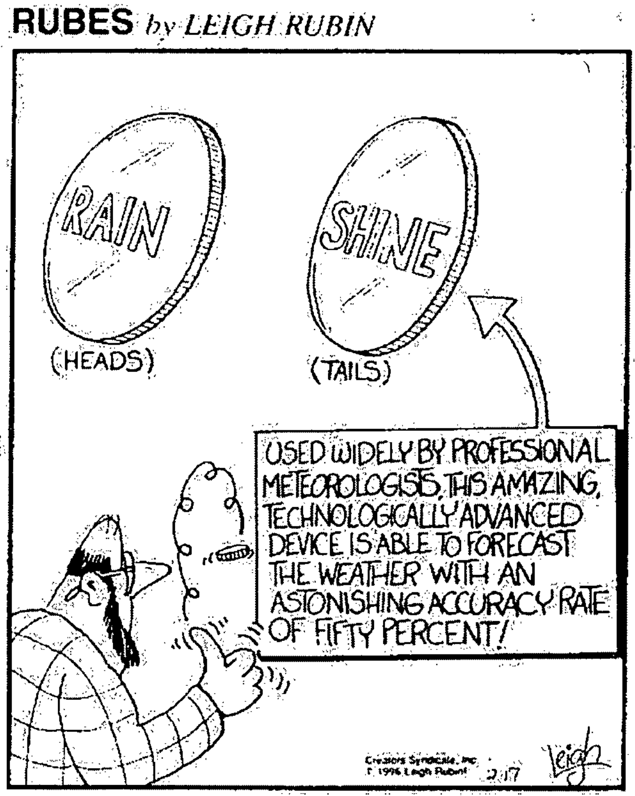


Figure 3-3: Model Uncertainty in Practice
Creators Syndicate Inc. © 1996 Leigh Rubin
Used by permission of Leigh Rubin and Creator's Syndicate Inc.

uncertainty would be present if this research-oriented model were subjected to practical application. Examples of simplifying assumptions include (Gan, 1987):

- the catchment is rectangular in area;
- a no-flow condition exists at the upper catchment boundary;
- drainage to the channel is perpendicular;
- all movement of water in the unsaturated zone is vertical;
- soils in each subsurface layer are homogeneous and isotropic;
- subsurface flow is parallel to the soil interface; and
- vegetation consists of uniform short grass, with rooting limited to the upper soil zone.

As is obvious from the examples above, model uncertainty exists where there is a gap between the assumptions, simplifications, processes, and mathematical representations that comprise a model and the natural processes that the model is attempting to replicate (Sorooshian and Gupta, 1983). It can manifest in different ways, including as a comprehensive inability to predict runoff accurately, even given correct parameters and input data, or as performance insensitivity to changes in relevant aspects of the model structure (e.g., Grayson et al., 1992a; Loague and Freeze, 1985).

In particular, a large degree of model uncertainty is common when either natural variability is neglected or processes known to be active *in situ* are not considered (Song and James, 1992). For example, Franchini and Pacciani (1991) note that the TANK and SSARR models lack a direct surface runoff component, but are nonetheless able to produce acceptable simulations of historical runoff. Naturally, less detailed or lumped models are subject to a higher degree of model uncertainty (Gan and Biftu, 1996; Gan and Burges, 1990a).

O'Connell and Todini (1996) believe that model uncertainty fundamentally results from an incomplete understanding of how various factors affect runoff at various spatial scales. Beven and Feyen (2002) offer a differing perspective, concluding that our understanding of hydrologic processes outstrips our ability to quantify them over a broad study area. Klemeš (1982) cautions that mathematical convenience has been known to take the place of scientific accuracy in defining a particular model structure, and that awareness of the arbitrary nature of the model fades away over time. These various viewpoints are not mutually exclusive. In general, some processes are well understood and are quantifiable, while others are understood but difficult to implement. Still others are effectively impossible to simulate due to limited knowledge or understanding.

Blöschl (2001) suggests that modellers should identify the dominant processes that control hydrologic response under different environmental conditions and at different scales, and focus their efforts on providing good simulations of these processes. However, modellers may be tempted to focus on those processes most amenable to modelling rather than those that are hydrologically significant for the catchment. For example, ET can represent up to 80% of hydrologic activity in a basin but receives disproportionately less treatment in hydrologic literature and practice (Klemeš, 1986a). Unsurprisingly, Gan and Burges (1990b) find that conceptual model predictive capability is poor in regions where ET accounts for a very significant proportion of the precipitation. Melching (p. 73, 1995) provides a relevant quote from Cornell (1972) that implies similar concerns, stating that "it is far better to have an approximate model of the whole problem than an exact model of only a portion of it".

Pre-defined methods, process models, or formulae used within a hydrologic model – such as those used for estimating ET – may have their own wide ranges of uncertainty (Binley et al., 1991). Many hydrologic models either calculate potential ET using previously-developed empirical formulae (e.g., Penman-Monteith) or require ET data as an input (e.g., Vrugt et al., 2003). The discussion of ET in Section 2.1.3 notes that the various estimation methods available are all subject to significant uncertainty. It is reasonable to consider this uncertainty to be irreducible, and thus the model uncertainty associated with this and other model components becomes subsumed into the model uncertainty of the larger hydrologic model.

Uncertainties in model structure can affect the properties of output information (e.g., volume, peak flow) in different ways and to different degrees (Melching, 1995). Model structure uncertainty may also play a large role in determining the behaviour of uncertainty in calibration (Yapo et al., 1996). Model uncertainty is often assumed to be insignificant unless and until a given calibration is proved unsuccessful, with attention instead focussed on uncertainty in the data. However, there is potential for model error to be "substantially larger" than measurement error, and many believe that model uncertainty is currently limiting model performance (e.g., Gupta et al., 1998; Yapo et al., 1996).

Analyzing model uncertainty is not as straightforward as analyzing uncertainty in input and output data. First and foremost, model errors will not necessarily have probabilistic properties that can be exploited to gain insight into the problem or potential solutions (Gupta et al., 1998). According to Singh and Woolhiser (p. 283, 2002), a lack of fundamental error analysis has limited our understanding of how different errors propagate through different model components and parameters. In evaluating model structure, general hydrologic laws are often self-evident but are difficult to verify due to uncertain data and the complex and variable nature of the system boundary conditions (Beven, 2002).

When considering multiple models for a task, Beven (1989) recognizes model uncertainty as the *potential* for error in any model, but seeks to avoid those models that are conclusively in error.

This philosophy, though potentially productive, must be applied with care. Although experimentation can identify errors in a given simulation, there is not necessarily any objective basis for attributing the error to model uncertainty. For example, Kuczera and Parent (1998) use Metropolis Sampling to identify apparent structural deficiencies in the CATPRO model. However, they do not eliminate the potential for data and parameter uncertainty, and therefore cannot declare with certainty that their observations are independent of other sources of error. At the other end of the continuum, attempts to show that model uncertainty is "negligible" are even less feasible, considering that such efforts are, in essence, verification, and are doomed to failure as discussed in Section 2.4.6.

The above discussion is not meant to imply that quantitative studies of model uncertainty are of limited importance. Harlin (1992) provides a strong example to the contrary. He finds that three different but equally acceptable realizations of the runoff-response function in the HBV conceptual model result in design flood simulations for four Swedish watersheds having an average uncertainty on the order of ±20%. Although the experiment does not represent the full extent of model uncertainty, it provides – at the very least – a good starting estimate.

The bulk of literature examining model uncertainty tends to focus on runoff and flow mechanics. This is not surprising, since the ability to accurately model fluid flows already exists in the form of the Navier-Stokes equations. However, Beven (2002) emphasizes a fundamental difference between flow equations in hydrology and other fluid dynamics disciplines. In hydrology, local geometry and boundary conditions replace fluid mechanics as the dominant constraint on small-scale flow pathways. A classic example is the tendency toward flow channelization even at small scales, which can distort physical representativeness at the model element scale (Woolhiser, 1996). Therefore, one could argue that model uncertainty in flow-related components of hydrologic models results from our inability to comprehensively apply the Navier-Stokes equations at the required level of detail (Beven, 2002).

The inability to relate actual boundary condition measurements and processes to even the smallest of element scales provides further evidence that the scale-dependence of model structures is a significant source of model uncertainty (Beven, 2001). The scaling of process dynamics, especially from laboratory to hillslope or catchment scale, is a contentious issue with surprisingly little scientific support. Many authors believe that the ability to perform calculations at the lumped catchment scale or even the model element or grid scale will continue to elude hydrologists for the foreseeable future (e.g., Beven, 2000; Blöschl, 2001).

In the absence of an acceptable, scaleable model framework, simplifications are necessary. Data uncertainty can be translated into additional model uncertainty where a lack of data requires further simplifying assumptions within the model structure. The conceptualization of overland flow as kinematic sheet flow is a common example, despite widespread documentation that kinematic sheet flow is rare in most hydrologic environments. In another example, Beven (2002) points out that Freeze and Harlan's blueprint for physically-based models is limited by its reliance on Darcian theory, which is not applicable for large scales and heterogeneous conditions.

Since the suite of model parameters is usually specified by the model structure, it should not be surprising that model uncertainty can be difficult to separate from parameter uncertainty. Consider the case where a model of forested areas includes parameters for litter-layer storage and interception storage but does not model the seasonal variation therein (e.g., Hornberger et al., 1985). In this case, the static nature of the litter-layer parameters is determined by the model structure. Therefore, even if the parameter values are correct for much of the year, processes based on the time-invariant litter-layer parameters introduce model uncertainty into the results. Conversely, if the model structure permitted seasonal variation of the litter-layer parameters but the parameter values were indeterminate, this would be a case of parameter uncertainty.

Model uncertainty can be easily mistaken for parameter uncertainty when there is no basis for choosing between parameter sets producing equally acceptable simulations (e.g., Grayson et al., 1992a). Sorooshian and Gupta (1983) present a systematic exploration of the SLS function response surface for the SMA-NWSRFS model. When objective function values are plotted against two of the percolation parameters, the authors identify a long, flat valley in the response surface. The authors conclude that the valley is a product of the structural representation of the percolation subprocess. Problems of non-identifiability can sometimes be solved through re-parameterization of the appropriate equation; in this case, a simple re-parameterization resulted in a marked improvement in parameter identifiability (Gupta and Sorooshian, 1983).

Model uncertainty may arise most commonly from process-related issues, but it can take other forms. Uncertainties can also arise from fundamental assumptions about how to interpret data and implement parameters within the model structure. Even the temporal increments and parameters chosen by the modeller can have a component of uncertainty. For example, modelling frequently disregards the physical basis for many time scales, mixing physical (e.g., day, season, year) and administrative (e.g., hour, week, month, decade) time intervals indiscriminately (Klemeš, 1983).

Arguably more important is the need to ensure that the modelling timestep is less than the runoff response time of the system. For example, the time taken for flood peaks on the Meuse and Rhine Rivers to reach the Netherlands are several hours and several days, respectively (van Hofwegen and Schultz, 1995). In this case, it is obvious that using a daily timestep model for flood forecasting on the Meuse River would be inappropriate. Less obvious is the answer to the question of whether or not an hourly timestep model is appropriate for the Rhine River basin. The answer would depend on the actual observed response time, since reasonable estimation of the flood hydrograph logically requires a modelling time step considerably smaller than the time of concentration. Harlin's study of extreme floods on six watersheds in Sweden using the HBV model is an example of the common practice of either assuming or failing to document that the chosen timestep is appropriate (Harlin, 1992). He uses a daily timestep without explaining how this timestep is selected, and whether it is, in fact, an appropriate choice.

3.1.4 Parameter Uncertainty

Given alternate sets of parameter values, a model may or may not achieve a good simulation of the historical data set. In this context, parameter uncertainty refers to the uncertainty associated with identifying a set of parameter values that the user believes best simulates *in situ* processes. Harlin and Kung (1992) illustrate the effects of parameter uncertainty by demonstrating substantial correlation between increased variance in model output and decreased parameter accuracy.

As noted in the preceding section, parameter uncertainty is closely related to model uncertainty in the sense that parameter definitions are a function of the chosen model structure. Accordingly, it is reasonable to expect that, if model uncertainty can be eliminated, then issues of parameter uncertainty and identification will become more tractable (Beven, 2001). However, every application of every model will still require user-defined parameters; thus, parameter uncertainty is and will undoubtedly continue to be a significant contributor to overall model predictive uncertainty.

The US National Research Council (NRC, 1995) notes that parameter uncertainty often takes a continuous format (e.g., through statistical functions like the probability density function (pdf) and cumulative distribution function (cdf)). Conversely, model uncertainty may involve "distinct and mutually exclusive choices" (ibid.). The NRC (1994) cautions that indiscriminate combinations of different types of uncertainty (e.g., averaging predictions from "equally probable" but distinct models) can yield results inconsistent with any of the alternative models.

Loague and Freeze (1985) contend that parameter uncertainty can have two distinct aspects, one related to natural variability and the other to knowledge uncertainty. The first, addressed in Section 3.1.1, deals with the potential for a poor representation of the areal distribution of catchment characteristics, and is strongly influenced by the innate variability of the catchment and the degree to which the model is distributed in nature. The second is a result of parameter interdependence, implying that the best set of parameter values may be unidentifiable due to non-uniqueness and correlation between parameters.

Kuczera and Parent (1998) caution modellers against naïve reliance on uniquely determined parameter values, the pursuit of which has historically commanded much attention in hydrologic literature. More recently, modellers are beginning to assess parameter uncertainty on a broader basis. Studies conclude almost universally that no single point in the parameter space can be uniquely defined as the global optimum, and in some cases, there may not even be a well-defined globally-optimum region (Vrugt et al., 2003). This aspect of parameter uncertainty has considerable implications for distributed modelling, since a calibrated parameter value must have small variance if it is to be compared with measured counterparts (ibid.). To fully address parameter uncertainty, such comparisons must be extended across multiple field measurements to ensure that the uncertainty distribution of the model parameter is representative of *in situ* measurement variability.

The crux of most ambiguity and identifiability problems is that the descriptions of process dynamics contained in the model structure are of a substantially higher order than typical observations of external system dynamics (Beck, 1987). The inability to accurately estimate parameters during calibration is a typical indicator that the structural complexity of the model cannot be resolved against the calibration data set (Hornberger et al., 1985). Grayson et al. (1992a) compare the case of an overparameterized model to a high-degree polynomial with too many degrees of freedom. Arguably, any model could simulate any flow series – regardless of scientific merit – by simply adding a sufficient number of parameters. The authors demonstrate this by using several combinations of parameter values to produce similar output series (ibid.). Hornberger et al. (1985) present similar results, using a broad range of parameters to generate "optimal" results.

For the reasons provided above, parameter uncertainty becomes very significant in the presence of substantial parameter complexity and nominal calibration data (Beven, 1989; Loague and Freeze, 1985; Jakeman and Hornberger, 1993). Jakeman and Hornberger (1993) go so far as to recommend that streamflow-calibrated models be limited to the half-dozen parameters that are mathematically required to represent the outflow series. Gan et al. (1997) caution that, despite growing awareness, over-parameterization could become more common as a result of advances in computing technology and optimization theory. Of course, parameter uncertainty can be reduced through the use of multiple data sets for calibration (Bergström et al., 2002; Burges, 2002; Jakeman and Hornberger, 1993). This approach is discussed further in Section 3.3.3.

Interdependence between a number of parameters implies that non-optimal or deviant parameter values can compensate for each other. In many cases, this makes it difficult to associate parameter uncertainty with individual parameters. Model performance is almost always associated with parameter combinations rather than individual parameter values; a given value for a particular parameter may result in either good or poor model performance depending on the values of other parameters in the model (Beven, 2000; Harlin and Kung, 1992). A set of calibrated parameters represents merely one combination that allows a particular model and calibration approach to produce results similar to the observed response. Parameters may only have meaning within the context of their particular parameter set and model, and are unlikely to work equally well with a different model or in a different parameter set (Beven, 2000, 2001;

Binley et al., 1991). Although the literature indicates that transfer of parameter values between models or hydrologic, climatic, or geographical regions may be possible in some situations, it must be undertaken carefully. In general, parameter values mean very little if removed from their original context.

Some research has explored the viability of applying "average" values for less sensitive parameters and focussing calibration and field measurement on those parameters to which the model is highly sensitive. Melching et al. (1990) conclude that the bulk of parameter uncertainty lies in calculating the quantity of runoff; if this is properly estimated, "average" routing parameters should suffice to calculate the peak and timing of the outflow.

Micovic (1998) investigates the use of average values for a subset of UBCWM parameters determined to have small variability across a range of catchments in British Columbia. He concludes that reasonably reliable simulations are possible using constant values. The use of "average" parameters may simplify simulations where only a rough estimate is required. However, the apparent limitation on parameter uncertainty must be interpreted with respect to the following:

- Micovic (1998) notes a 5% decrease in model efficiency when moving from case-specific to average parameters. This may not be insignificant; in some cases, even a 1% improvement can be quite difficult to achieve (e.g., Lan, 2001);

- Performance may not be as consistent on an event-by-event basis using average parameter values as it can be over the long term. Therefore, applying this approach for flood forecasting applications might not be advisable;

- "Averaging" of parameters can introduce systematic errors in model output for individual applications. For example, of the twelve basins reviewed by Micovic (1998), five consistently overestimate peak flows, while five others consistently underestimate peak flows; and

- The applicability of "average" parameter values must be considered carefully in the context of their potential dependence on other parameter values, catchment variability, and the prevailing hydrologic, climatic, or geographic region.

Merz and Blöschl (2004) present a more in-depth examination of parameter regionalization for the HBV model. They report that using a global parameter set (in this case, the mean calibrated parameter values for 308 catchments in Austria) results in generally poor performance. In reviewing the literature, they find that most case studies report low correlations between model parameters and catchment attributes. Although their study finds that regionalization methods based on spatial proximity perform significantly better than regionalization based on catchment attributes, their results imply that there is an upper limit of appropriateness for the spatial regionalization of model parameters.

Merz and Blöschl (2004) also propose exploring parameter uncertainty by cross-comparing the results from independent calibrations against two halves of a split-sample data series. In comparing their results to other studies in the literature, they suggest that uncertainty for a given set of parameters can have a significant degree of dependence on the catchment being studied and the characteristics of the available data.

Hornberger et al. (1985) investigate two methods to reduce parameter uncertainty. In the first case, as above, insensitive parameters are fixed at their median values. Although this leads to greater parameter stability, they find that objective function results are significantly inferior to those obtained with the full calibration. This procedure, while common, is obviously not ideal for this particular application (ibid.). In the second case, parameters are eliminated rather than arbitrarily fixed. This results in a much better defined minimum on the response surface, although the objective function is still inferior to the value obtained with the full calibration.

Takyi (1991) relates parameter interdependence to parameter uncertainty in a more general context, noting that disregard of parameter correlation can lead to gross underestimation of model predictive uncertainty during later uncertainty analyses. For example, Song and Brown (1990) find that the standard deviation for predicted dissolved oxygen deficit is 20-40% larger for correlated inputs than for uncorrelated inputs. Harlin and Kung (p. 211, 1992) point out that the difficulty of studying the interaction between three or more parameters is exacerbated by the inability to graph a response surface. They note that the shape of any two-parameter response surface will depend on the values of the other parameters.

Parameter uncertainty is perhaps most obvious when ostensibly static, single-value model parameters are observed to have a non-negligible dependence on the flow sequence and climatic data used for calibration. Gan and Burges (1990b) present one such example. Other examples are noted by Woolhiser (1996), who observes parameter sensitivities that are dependent on both basin scale and rainfall magnitude, and Arnaud et al. (2002), who find it difficult to resolve the "physical interpretation" of any parameter if it is found to be sensitive to the rainfall pattern. Gupta and Sorooshian (1983) point out that while the noted uncertainty could potentially be parameter uncertainty, the cause could equally be related to identifiability problems in the model structure, or to calibration data that do not adequately "activate" the relevant process. In general, dependence of parameter values on calibration data is often an indicator that changes to the model structure are necessary (Weiler, 2005).

Another common indicator of parameter uncertainty arises when the parameters of a "best fit" model assume values at or very near the extremities of their feasible range. Franchini et al. (1998) suggest that this phenomenon is due to undesirable compensation among parameters induced by restrictive specifications of the feasible parameter space. Preliminary investigations into the effect of widening the specified parameter ranges by Hogue et al. (2000) yield inconsistent and inconclusive results.

Of the three components of knowledge uncertainty, parameter uncertainty is arguably the most amenable to statistical analysis. For subjective assessments of parameters, a mean value can be used to represent the "best guess" of the parameter value and a variance can be assigned corresponding to the level of confidence in that guess (Vicens et al., 1975). Examples of statistical representation of parameter values abound. Binley et al. (1991) calibrate the IHDM against five storm events, using the resulting five sets of parameter values to estimate statistical properties for each parameter. Garen and Burges (1981) study model predictive uncertainty using the idea that Coefficients of Variation for the uncertain parameters can be estimated and used to describe the parameter variability. Such estimation is subjective, especially for parameters lacking a physical interpretation. Methods for estimating parameter uncertainty are much more defined than for model or even data uncertainty, and are discussed in Section 3.3.

3.2 Uncertainty and Calibration

It is unlikely that any hydrologic modeller could obtain a good simulation of observed data without recourse to calibration. The challenge of streamflow prediction in ungauged basins remains a major focus for research (e.g., IAHS, 2005). Even physically-based (i.e., field-measured) parameter values are subject to many problems and typically require some "adjustment" to improve performance (Binley et al., 1991). Therefore, the following discussion should not be viewed as being entirely limited to calibration-dependent conceptual and empirical models.

The most important aspect of uncertainty in calibration deals with the blurring of the different types of uncertainty discussed in Section 3.1 (e.g., data, parameter, and model uncertainty). Lei and Schilling (1996) highlight the widespread presumption that calibration can resolve data and model structure uncertainty as well as parameter uncertainty. While this is obviously not the case, the premise is not entirely groundless. Madsen (2000) points out that although calibration should ideally only affect parameter uncertainty, it may also compensate for errors in other areas of the simulation. The resulting complimentary errors can generate unbiased output that compares favourably with the observed data (Melching, 1995; Melching et al., 1990).

It would be more correct to state that model calibration partially compensate for uncertainties and errors in data and model structure by changing parameter values to force the model output to better fit the observed data. In this way, calibration translates some of the extant model and data uncertainty into parameter uncertainty. For example, a realistic model cannot resolve biased input data (e.g., rainfall undercatch) against unbiased calibration data (e.g., streamflow) without artificially distorting parameter values (Beck, 1987; Burges, 2003).

Conversely, if a model is able to reproduce an erroneous, non-representative or otherwise uncertain series of observed data, there is likely significant uncertainty in the model structure or parameter values. For example, Franchini and Pacciani (1991) find that a hydrologic model known to omit relevant processes can nonetheless be calibrated to obtain an acceptable simulation of observed streamflow.

Madsen et al. (2002) advise that, if data and model uncertainty are not otherwise addressed, the goal of calibration implicitly shifts from the minimization of parameter uncertainty to the balancing of compensating errors. Lei and Schilling (1996) suggest the use of "preliminary" uncertainty analysis to separate events and models with substantial uncertainties or errors in data and model structure from those that can properly benefit from calibration.

In essence, calibration adjusts parameters in an attempt to reproduce the observed transformation of a specific input series into a specific output series. Therefore, given the non-repeatability of hydrologic events, each subsequent application of a calibrated hydrologic model cannot help but involve extrapolation or interpolation from or between response modes observed during calibration.

The potential pitfalls of extrapolation are relatively well-known, and are addressed in an applied context in Section 3.4. However, even interpolation is not always acceptable. Klemeš (1986a) presents a classic example of the dangers of interpolation; he describes how if one man takes four hours to load a truck, two men take two hours, and four men take one hour, an observer might correctly assume that the loading process requires a total of four man-hours. However, it would be fundamentally inaccurate to interpolate that the same truck could be loaded over three hours by one and one third men.

The assumption that calibrations are transferable beyond the calibration period, while often valid, is sometimes tenuous. Gan et al. (1997) observe that a model calibrated against wet years tends to be biased toward over-estimation when validated against dry years, and toward under-estimation in the reverse case. In many cases, the level of uncertainty introduced during model application is related to the extent by which the nominal conditions differ from those for which the model was calibrated. For example, Gan (1987) finds that extreme flow forecasts generated by the SAC-SMA model are significantly better for surface-flow dominated watersheds than for those dominated by sub-surface processes. He contends that this is due to a better simulation of the processes involved in an extreme runoff event (i.e., overland flow) during calibration (ibid.).

The U.S. National Research Council (p.44, NRC, 2000b) notes that uncertainty can arise from "an inability to understand the objectives that society holds important or to understand how alternative projects or designs should be evaluated". While this description is highly generalized, it nonetheless captures the idea that there is uncertainty associated with the choice of objective for evaluating a situation. Measurement of hydrologic model performance during calibration is no exception. Boyle et al. (2000) caution that preemptive selection of a single criterion for calibration can potentially predispose the calibration process toward an inappropriate result. Insight about the nature of the various alternative objective-dependent response surfaces in the region of the final solution is necessary for quantifying uncertainty in model predictions (Kuczera, 1997).

A true set of globally-optimum parameter values should be independent of calibration data. However, in many cases the act of calibration implicitly relates the two (Gan and Biftu, 1996). The inability of calibration procedures to locate globally optimal parameter estimates with confidence translates into uncertainty regarding the accuracy of the model (Duan et al., 1992). Gupta and Sorooshian (1983) illustrate why achieving a good objective function value alone is not sufficient evidence to conclude that the calibration is successful. They prove that the use of additional data can alter the response surface such that only the global optimum point is shared by all potential distributions of local optima. Therefore, a "successful" model calibration that identifies only a local optimum could produce poor results when applied to other data sets (ibid.). This hypothesis is supported by studies of multi-objective calibration, which demonstrate that a "successful" calibration against one parameter is not necessarily associated with "successful" simulation of other aspects of the watershed (e.g., Bergström et al. (2002); Seibert (2000); Seibert and McDonnell (2002)).

Local optima may result from model or data uncertainty, the nature of the objective function itself, or any combination thereof. Their distribution may be scattered, clustered, or curvilinear. Through a process of exhaustive gridding, Duan et al. (1992) identify as many as 55 local optima when considering two parameters of the SIXPAR model, a simplified version of SAC-SMA. The number of local optima increases substantially as noise is introduced into the calibration time series, and exceeds 500 for several cases when considering a three-dimensional parameter subspace. The authors conclude that the response surface is quite flat in the vicinity of the global optimum. One particular set of parameter values is far from the "true" parameter values but has a function value virtually indistinguishable from the point nearest the true solution (ibid.).

Even a "global optimum" set of parameter values may not be correct; it is possible that a near-optimal point (as measured by the selected criterion) may be more representative of *in situ* processes than the mathematically-optimal answer. The choice between near-optimal alternative parameter values can be highly subjective and is the subject of some discussion in Section 3.3, most notably with regard to the concept of equifinality (Section 3.3.6).

The problems with calibration are well-documented, as evidenced by the dominant focus on model calibration within the literature of the past few decades. But, researchers are now beginning to question what more can be gained by setting aside studies of calibration in favour of exploring the root causes of uncertainty in modelling. This perspective aligns itself with the argument that progress in modelling – at least in terms of uncertainty reduction – must come from a better understanding of hydrology, and not from more intensive manipulation of the little knowledge we have (Klemeš, 1982; Melching et al., 1990).

3.3 Techniques for Exploring Uncertainty

Results from most hydrologic models are deterministic in nature, consisting of a single value or series of values presented without alternatives or probabilities. Beck (1987) contends that precise results from uncertain models offer a misleading sense of precision. This "arbitrary precision" may be acceptable in specific circumstances, but there are many cases in which an analysis of the accompanying uncertainty is critical. Scarce resources may restrict further investigation to factors or processes identified as important at the expense of other potentially important factors (NRC, 1995). A high degree of uncertainty may itself be sufficient to influence or bias sensitive decisions.

Beven (2001) notes that an assessment of predictive uncertainty improves the likelihood of success while providing mitigation in the event of error or miscalculation. Uncertainty analysis can also lead to insights concerning the value of additional information and thereby allow researchers to focus their efforts more effectively (NRC, 1995).

Simple solutions to the problem of how to reduce or quantify uncertainty have been proposed, including using a longer data series for calibration, reducing the dimensionality of a model by eliminating insensitive parameters, and utilizing different evaluation criteria. However, these approaches attempt to gain more information from the same knowledge and do not address fundamental issues. A more comprehensive approach is obviously required.

The U.S. National Research Council (NRC, 1995) outlines two approaches for quantifying uncertainty. The first, being confidence intervals, expresses uncertainty in terms of the probability with which repeated sampling is expected to yield outcomes within a given interval around the "true" solution (ibid.). Confidence intervals are the most commonly used method of describing sampling uncertainty, and are frequently employed even when they are not substantiated by the larger context of the problem. For example, confidence bands based on standard error are included in many flood frequency analyses, despite the fact that differences between predictions from alternative distributions can be of a much greater magnitude.

The NRC's second approach for considering uncertainty employs Bayesian statistics to generate probabilistic outcomes (NRC, 1995). The Bayesian approach treats model parameters as probabilistic variables whose pdfs represent the likelihood of each parameter assuming different values. Output probability distributions ("posterior" pdfs) are the product of two quantities: a likelihood function, which incorporates available calibration data, and a "prior" pdf which represents the modeller's *a priori* knowledge. In the end, the promise of this approach in quantifying uncertainty is not fulfilled because the likelihood functions and prior estimates are still subject to the issues raised in the first two sections of this chapter.

In general, statistically-based methods of uncertainty analysis suffer from the same fundamental lack of physical basis noted for empirical and black box models in Section 2.3.1. One must also be aware that statistical representation of uncertainty can be misleading in cases where some aspects of uncertainty are not included either by design or omission. For example, a 5% probability of exceedance determined from frequency analysis does not generally make allowance for the effects of climate change; the actual probability of exceedance could vary substantially from the "accepted" value.

Subsequent sections of this chapter summarize a number of the more widely-accepted techniques for managing uncertainty, including sensitivity analysis, reliability analysis, threshold techniques, equifinality, and uncertainty isolation. These approaches are by no means mutually exclusive. Madsen (2000) demonstrates the potential for different perspectives on uncertainty to lead to different insights. He evaluates the performance of the NAM model for single- and multi-objective calibrations using overall volume error, overall RMSE, average RMSE for peak flow events, and average RMSE of low flow events. Figure 3-4 shows the variability in parameter values obtained by considering a set of solutions for single-objective calibration (based on a balanced objective function) whose objective function values are all within 1% of the optimum value. In contrast, Figure 3-5 shows the variability in parameter values obtained by considering the Pareto set of optimal parameter values obtained through a multi-objective analysis of peak flow RMSE and low flow RMSE.

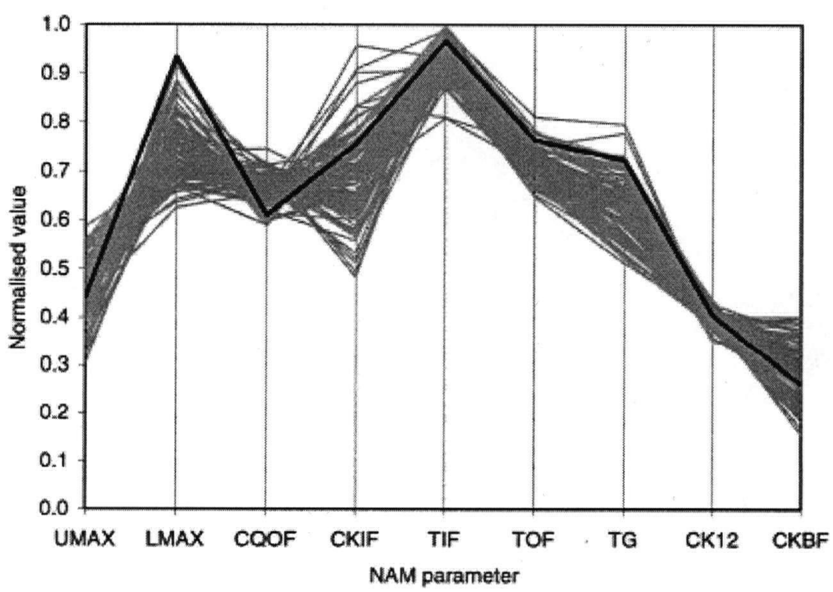


Figure 3-4: Normalized Parameter Sets for a Single-Objective Optimization where Objective Function Values Differ by < 1%

The bold line indicates the optimum parameter set. (from p. 286, Madsen, 2000)

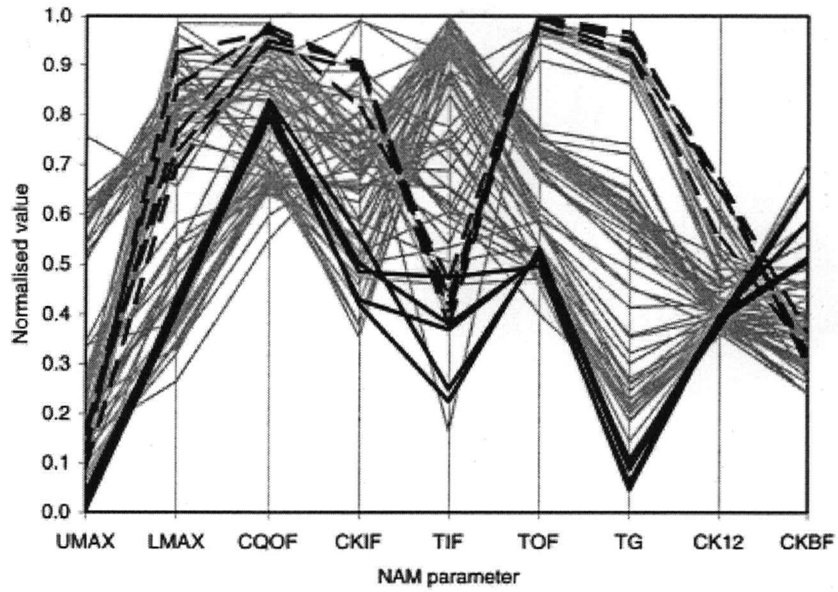


Figure 3-5: Normalized Pareto Optimal Parameter Sets for a Multi-Objective Optimization
Full and dashed bold lines indicate parameter sets associated with the five smallest RMSE values for peak flow and low flow, respectively. (from p. 284, Madsen, 2000)

Figures 3-4 and 3-5 are equally valid representations of different aspects of uncertainty for Madsen's application of the NAM model (Madsen, 2000). The importance of both perspectives to an analysis of uncertainty is obvious: the first approach provides insight into the parameter uncertainty (i.e., uniqueness) of the "optimal" solution for a single objective, while the second provides insight into parameter uncertainty across multiple objectives. A full uncertainty analysis would need to consider and account for both approaches.

3.3.1 Sensitivity Analysis

The most well-known and widely-used tool for exploring uncertainty in hydrology is sensitivity analysis. The USACE (1992) defines sensitivity analysis as "the systematic evaluation of the impacts on project formulation and justification resulting from changes in key assumptions underlying the analysis".

For the purpose of characterizing model predictive uncertainty, sensitivity analysis typically takes on a more specific meaning. Takyi (1991) defines sensitivity analysis as an investigation of how different aspects of the model contribute to model predictive uncertainty, expressed as the rate at which the model output changes with variations in the uncertain component.

Naturally, the criteria used to evaluate changes in the model output must be appropriate for the subject of the analysis.

Sensitivity analysis is most commonly used to identify the parameters that have the greatest impact on model performance. This is done by evaluating how model output changes with variations in the "best" set of parameter values (McCuen, 1973; Takyi, 1991). Typically, parameter values in the "best" parameter are varied one-by-one in small increments, with all other parameters fixed.

In the same fashion, sensitivity analysis can also be used to identify "insensitive" parameters. Gan and Burges (1990b) present an example, noting that a wide variety of values could be used for selected parameters in the Sacramento (SAC-SMA) model without affecting their simulation results. Lastly, sensitivity analysis can form the basis for a more intensive analysis; for example, sensitivity analysis can be used to identify parameters that affect model results in predetermined ways as a preliminary step to a more thorough investigation of those parameters.

The benefits of the qualitative and quantitative information gained from sensitivity analysis are fairly intuitive. Information related to the effects of parameter variation can be of considerable value in deriving relationships between parameter values and basin or storm properties (McCuen, 1973). Arguably more important is the role of sensitivity analysis in determining the confidence with which the user or decision-maker regards the model results. Sensitivity analysis effectively provides an estimate of the consequences of error in the analysis; confidence is increased if results are sensitive only to those parameter values or data series that are known with a high degree of certainty.

One major advantage of sensitivity analysis is its simplicity; it can be performed with relatively little effort as part of a larger study. If applied exhaustively across all dimensions of uncertainty, sensitivity analysis would no doubt prove to be a strong tool for characterizing uncertainty. However, in its most commonly-applied format, it examines the sensitivity of the model to only one factor at a time. This limitation implies that sensitivity analysis alone, while useful, cannot provide a complete description of model predictive uncertainty.

Users of sensitivity analysis techniques should be aware of several caveats. Firstly, sensitivity analysis assesses parameters under the often-invalid assumption that they are independent. The outcome of such an analysis may not reflect any correlations or interactions between parameters, and may therefore be misleading or invalid (Madsen, 2003). McCuen (p. 43, 1973) emphasizes that "the sensitivity of one factor depends, in the general case, on the magnitude of all factors of the system". The increase in model predictive uncertainty arising from independent perturbation of a single factor can far exceed that arising from the same perturbation accompanied by variations in other parameters. In some cases, the incautious variation of a single parameter value can result in a parameter set that is no longer appropriate for simulating the *in situ* conditions.

Performing sensitivity analysis for a calibrated model can add further complexity. In such cases, the analysis is limited to reflecting the uncertainty of the parameter near its "optimal" value and does not provide information about the overall importance of the parameter within the full parameter space (Takyi, 1991). Any attempt to reduce the dimensionality of the calibration problem based on the results of sensitivity analysis should proceed with caution and an eye to parameter correlations essential to model behaviour (Madsen et al., 2002). In seeking to develop a process-oriented calibration scheme for the HBV model, Harlin and Kung (1992) find that different calibration methods result in different "optimal" parameter sets. They conclude that the application of sensitivity analysis would not be meaningful since it is restricted to analysis of variation near a single "optimal" parameter set. The implicit argument against sensitivity analysis for calibrated models is that the results of the analysis can differ from one calibration to the next.

From another perspective, Beven (1989) questions the application of sensitivity analysis in deterministic simulation because probability is not paired with each outcome. The results of a sensitivity analysis essentially comprise non-commensurate results which must be assessed qualitatively. Without an accompanying assessment of the probability for variation in each parameter value, one must be aware of Melching's observation that insensitive but uncertain parameters may have a greater influence on the reliability of model output than highly sensitive parameters that are known with certainty (Melching, 1995). Where probability is paired with each aspect of a sensitivity analysis, the end result begins to take the form of a reliability analysis, discussed in the next section.

3.3.2 Reliability Analysis

It is often desirable to associate a probability or likelihood with the output of a hydrologic model as a means of assessing its reliability. Melching et al. (1990) believe that any approach for estimating reliability should be systematic, unbiased, flexible, consistent, simple, and comprehensive (i.e., accounting for all potential sources of uncertainty). While the realization of all of these goals simultaneously is obviously unattainable at present, several existing methods can provide a probabilistic distribution of uncertain results rather than a single deterministic estimate. For the purposes of this work, these methods are collectively labeled methods of reliability analysis.

Reliability analysis requires either a stochastic model or stochastic treatment of a deterministic model, as well as descriptions of variability for all basic quantities of interest (e.g., data and parameters). Reliability analyses for hydrologic models typically estimate individual quantiles such as peak flow in the form of a cumulative distribution function or cdf (i.e., as the likelihood $F(x)$ that the peak flow for a flood will be less than or equal to x).

Representing model results as a cdf allows the modeller a means of visually assessing the uncertainty in the simulation (Melching et al., 1990). In the limiting case of absolute certainty, the cdf and pdf converge to step and spike functions, respectively, such that a single unique answer is associated with 100% probability. Therefore, provided the model is representative, a steep cdf with minimal curvature at the extremes is an indicator of good reliability (i.e., fairly certain results) (ibid.). Conversely, a broad cdf of mild slope is an indicator that the results are highly uncertain or unreliable.

Reliability analysis usually addresses uncertainty at a more fundamental level than sensitivity analysis, and is therefore less concerned with isolating the impact of any single factor on the modelled results. Therefore, reliability analysis escapes most of the potential pitfalls discussed in the preceding section. However, other challenges take their place. For example, one must be aware of the potential for parameter values randomly selected on an individual basis to combine into unrealistic parameter sets. Simply removing any unrealistic scenarios from consideration after they have been identified is a poor solution, since this compromises the relationship between the set of "acceptable" alternatives and the prior probability distributions of their components. If, for example, parameters are found to be correlated, these relationships must be explored and, if necessary, incorporated into the analysis (e.g., Garen and Burges, 1981).

While resultant probabilities from a reliability analysis must, by definition, sum to unity, they almost never reflect the full extent of model predictive uncertainty. Reliability analyses rarely address any uncertainties associated with basic assumptions made during the formulation and execution of the model. Reliability analyses also rarely account for any indeterminate uncertainties that do not lend themselves well to mathematical representation. Modellers should respect the context of any analysis by documenting which aspects of uncertainty are included in their analysis, and which are not addressed. In the absence of a clear explanation of context, probabilistic representation of results could conceivably generate a disproportionate level of confidence in results (e.g., Song and James, 1992).

Monte Carlo Simulation (MCS) is generally the most accurate approach for reliability analysis; in particular, Takyi (1991) cites the advantages of MCS in developing the statistical properties of the aggregate model output, which in turn characterize the uncertainty in the model response. The MCS approach typically involves multi-thousand evaluations of a performance function such as $Z = Q_{max} - Q_{simulated}$, where $Q_{simulated}$ is the peak flow obtained from a deterministic model with a unique, randomly-selected combination of model, data, and parameters. The probability that the peak flow will be less than Q_{max} is approximated by the frequency with which the free variable Z is less than zero. By considering a range of values for Q_{max}, one can construct the cdf $F(Q_{sim} \leq Q)$ for all values of Q.

The main criticism levelled at MCS-based reliability analysis has historically involved its computational intensity. Subsequent advances in computing technology have led to more advanced investigations of uncertainty based on extensive random sampling from input distributions. However, it is widely recognized that MCS relies on the assumption of purely random sample selection. In most cases, random numbers are generated via one of many possible computer algorithms, and there is potential for the assumption of randomness to be violated. The user should be well aware of the famous quote attributed to John von Neumann, one of the fathers of MCS, which states that "anyone who considers arithmetical methods of producing random digits is, of course, in state of sin".

MCS has been widely accepted as the best method for reliability analysis over the past decade. However, alternatives exist in the form of less accurate but computationally simple methods of approximate solution. Naturally, their results are subject to additional uncertainty, which can be explored through comparison against results obtained via MCS (e.g., Tung, 1990). Melching (1995) outlines several approximation methodologies including Mean-Value First-Order Second-Moment (MFOSM) and Advanced MFOSM, the Point Estimation Methods of Rosenblueth and Harr, and Latin Hypercube Sampling. Brief discussion herein is afforded to the most common of these, being first-order approximation. Also, some discussion of more advanced methods involving Markov Chains and the Metropolis Algorithm is merited, as they represent great potential for exploring model predictive uncertainty and are likely to grow in popularity.

First-order methods for reliability analysis are computationally straightforward, requiring only the first and second statistical moments of the input variables. These methods use a linear approximation of the performance function, usually through truncation of a Taylor series expansion at the mean point of all uncertain variables. The linear approximation is used to estimate the expected value and variance of the performance function. The system reliability β can then be calculated as the inverse of the coefficient of variability for the performance function, and related to probability by $P(Q_{sim} \leq Q) = \Phi(\beta)$, where $\Phi()$ is the standard normal integral.

First-order methods involve several assumptions concerning the nature of the problem that can potentially result in misleading results. Most importantly, the linear approximation becomes less accurate with both increasing non-linearity in the system and increasing distance from the mean values, and can yield incorrect results (e.g., $g(E[x]) \neq E[g(x)]$) (Harlin and Kung, 1992). Garen and Burges (1981) compare the results of first-order approximations with results obtained through MCS. They confirm that first-order analyses underestimate means and standard deviations, as expected given the truncation of the Taylor Series expansion. The authors conclude that where first-order analysis is known to be a poor approximation of MCS, its results are likely subject to an unacceptable degree of uncertainty (ibid.).

Monte Carlo Markov Chain (MCMC) algorithms are growing in popularity as a means for assessing uncertainty in parameter values. MCMC algorithms, which began as models of physical systems seeking a state of minimal free energy, are a more advanced form of sampling that use both expert knowledge and simulation history to probabilistically guide Monte Carlo selection. Vrugt et al. (2003) explain that MCMC methods use stochastic techniques to successively explore solutions throughout the parameter space, updating their component probability distributions as they progress. Rather than seeking a probabilistic description of reliability based on *a priori* or unknown prior distributions, MCMC algorithms attempt to define the appropriate probability distribution for each parameter. Conventional expressions of reliability (e.g., 90% confidence interval for streamflow) are calculated from the posterior parameter distributions.

By far the most common instance of MCMC is the Metropolis-Hastings algorithm (Hastings, 1970). Vrugt et al. (2003) present a concise outline of the Metropolis-Hastings process. Kuczera and Parent (1998) use the Metropolis algorithm to investigate parameter uncertainty for a conceptual hydrologic model. The key to the adaptive ability of the algorithm is that it will always accept a solution of higher probability, but will also accept less probable solutions with a frequency obtained by randomly sampling from the interval [0,1] (ibid.).

It is important to preserve the ergodicity of the Markov Chain (i.e., its representation of low-probability areas) throughout MCMC processes. In this way, the intent of MCMC techniques contrasts with most other automatic calibration techniques, which typically abandon "sub-optimal" regions of the parameter space in favour of regions of higher probability. Vrugt et al. (2003) combine the Metropolis algorithm with the SCE-UA algorithm to form the hybrid Shuffled Complex Evolution Metropolis algorithm, abbreviated SCEM-UA. The SCEM-UA algorithm utilizes the demonstrated power of the SCE-UA algorithm while avoiding the tendency to collapse into a single "best" region of attraction. The authors find that parameter values identified as most probable in the SCEM-UA simulation are virtually identical to the unique solution generated by the SCE-UA method. Therefore, it is reasonable to assume that the SCEM-UA method can replace the two-step process of identifying a "global" optimum and then applying other techniques to investigate parameter uncertainty. The authors illustrate the utility of the SCEM-UA method by using its results to generate confidence limits for flood hydrographs (ibid.).

The results of an MCMC algorithm can be plotted as a series of hydrographs showing the range of solutions associated with a selected confidence level. For example, Vrugt et al. (2003) show that the band of hydrographs associated with the 95% confidence region does not include the observed records. They conclude that this bias is a strong indicator of uncertainty or error in the model structure. There is great potential for such techniques to contribute to our understanding of model predictive uncertainty. However, most relevant MCMC techniques are relatively recent, and the independence of their description of uncertainty has not been fully verified.

3.3.3 Multi-Objective Analysis

A large number of hydrologic model studies have focussed on the pursuit of a single "best" solution for the calibration problem. However, a single "best" solution is often unattainable. Gupta et al. (1998) explain that the multi-objective nature of model calibration makes any single-objective calibration necessarily subjective. Yapo et al. (1998) caution that even a carefully-chosen objective function can still fail to adequately measure important characteristics of, and differences between, observed and simulated data series. Much of the subjectivity in calibration results from the lack of an unambiguously correct way to minimize the length of an error vector containing non-commensurable components (Bastidas et al., 2001). And ultimately, even a perfect simulation of the observed data series of interest (e.g., an observed hydrograph) is not necessarily a robust indicator of an appropriate model for a system (Seibert and McDonnell, 2002).

If a modeller concludes that acceptable model performance for all required objectives cannot be obtained from a single unique solution, the recourse is multi-objective analysis. A multi-objective solution comprises the set of all non-inferior alternatives in the feasible space (Gupta et al., 1998; Yapo et al., 1998). Revelle et al. (p. 104, 1997) define a solution as non-inferior "if there exists no other feasible solution with better performance with respect to any one objective, without having worse performance in at least one other objective". Commonly-used synonyms for non-inferiority include dominance, efficiency, and Pareto optimality, after Italian economist Vilfredo Pareto (Revelle et al., 1997; Yapo et al., 1998).

The key assumption of multi-objective calibration is that the resulting non-inferior solutions collectively present a more reliable description of catchment behaviour than can be attained using any one single objective (Seibert, 2000). The concept of a set of non-inferior results can be compared with the set of calibrations that would result from a team of experts performing independent manual calibrations, where each expert uses a unique combination of knowledge, experience, insights, and priorities to generate equally good but different solutions. Because of the additional validation inherent in the process, models subjected to multi-objective calibration are typically considered more robust internally than their single-objective counterparts.

For a complete multi-objective analysis, residual uncertainty would arise only from concerns of subjectivity in selecting objectives and from statistical sampling issues. However, including alternative model structures and data sets in a multi-objective analysis would be difficult at best, and would likely result in the identification of discrete solutions rather than subjective trade-offs. For this reason, the bulk of the work in multi-objective analysis has focused on parameter uncertainty. Other dimensions of model predictive uncertainty are better addressed using the more philosophical techniques described in subsequent sections of this chapter. Yapo et al. (1998) suggest that further research into multiple-objective calibration is best directed toward understanding how to select the set of objective functions, the sensitivity of results to the number of objective functions, and the amount of data used in optimization.

The different "objectives" of multi-objective analysis need not conform to their familiar context as numerical measures of similarity between observed and simulated streamflow. A listing of various alternative "objectives" might include:

- different response characteristics (e.g., peak flows over threshold or annual water balance) (Boyle et al., 2000; Madsen, 2003);
- different model state variables or processes (e.g., groundwater level, chemical data, snowpack) (Bastidas et al., 2001; Bergström et al., 2002; Madsen, 2003);
- distributed or multi-site measurements (e.g., runoff measured at points other than the catchment outflow) (Madsen, 2003);

- multiple sets of calibration data (e.g., different events for an event-based model) (BC Hydro, 2004; Binley et al., 1991); and

- the use of both "hard" and "soft" (e.g., spotty, discontinuous, numerically approximate, or qualitative) data (Seibert and McDonnell, 2002; Merz and Blöschl, 2004).

Many objectives currently in use are themselves mathematical composites of multiple objectives; for example, the EOPT! statistic commonly used for evaluating the UBCWM sums the Nash-Sutcliffe efficiency E! and total volume error. Merz and Blöschl (2004) demonstrate the potential for improved performance using a compound objective. Their objective function combines runoff efficiency, volume error, and penalty functions for three kinds of "soft" data (expected parameter distributions, snow accumulation, and moisture accumulation). Using this compound objective function results in a decrease in E! for calibration compared to the conventional calibration against runoff. However, performance during validation improves, indicating a better representation of the catchment hydrology (ibid.).

Nevertheless, one could argue that a single quantitative combination of multiple objectives is still a single measure; it is not definitively "multi-objective" in the sense that it does not provide multiple alternative solutions.

As for multi-objective research in other fields, hydrologic model studies have shown that trade-offs in performance between members of the non-inferior set can be significant (e.g., Boyle et al., 2000; Loague and Freeze, 1985). Seibert (2000) finds that a multi-objective calibration will tend to simulate a single variable less accurately than a dedicated calibration. However, he contends that, for the right model, the decrease in performance should be small and will most likely be offset by improvements in the simulation of other variables. He goes on to demonstrate substantial improvement in groundwater level simulations (from $R^2 = 0.313$ to $R^2 = 0.834$) accompanied by a relatively minimal decrease in the quality of runoff simulation (from E! = 0.879 to E! = 0.834) when conventional calibration is expanded to include groundwater level data. Bergström et al. (2002) reach similar conclusions using snow depth, depth to the water table, runoff, and O^{18} groundwater content, as do Seibert and McDonnell (2002) using both "hard" and "soft" runoff and groundwater data. Seibert and McDonnell (ibid.) also note that their multi-objective calibration reduced parameter uncertainty by an average of 60% compared to the corresponding conventional single-objective calibration.

While near-unilateral improvement in objectives is possible, objectives and Pareto solutions can also be mutually exclusive. This situation means subjective preference must be used to choose between the various objectives and solutions (Lohani et al., 1997); the modeller must decide whether the required reduction in some aspect of performance is justified by the better representation of others (Seibert and McDonnell, 2002). In such cases, the non-inferior set is often plotted in two- or three-dimensional objective space to graphically represent the trade-offs between objectives (Revelle et al., 1997). Where trade-offs are significant, unacceptable trade-offs and impractical solutions are often eliminated as a preliminary step in paring down the non-inferior set (e.g., Boyle et al., 2000).

Analysis of the non-inferior set can reveal characteristic model or parameter properties and behaviours. This information can sometimes be used to guide structural improvements to the model itself (Boyle et al., 2000). For example, a parameter exhibiting small variability across the Pareto set is usually important to overall model behaviour, since parameters to which the model is highly sensitive will implicitly be estimated with greater accuracy (Madsen, 2003). Conversely, the Pareto set of parameter values may cover a large percentage of the feasible parameter space; in one case, Madsen (ibid.) observes Pareto parameter values varying by over 50% of their feasible range. A high degree of variability across the non-inferior set usually indicates that related aspects of the model require further study.

Gupta et al. (1998) plot all the hydrographs associated with the non-inferior set on a single graph to show "the space of possible hydrograph solutions" to the multi-objective problem. This type of plot is referred to herein as a Pareto hydrograph. Using a Pareto hydrograph, one can subjectively assess the fit between the set of non-inferior solutions and the observed data, as well as evaluating whether the fit can be improved through further calibration. This makes the Pareto hydrograph a good tool for identifying data error or model non-representativeness. Based on the plot shown in Figure 3-6, Yapo et al. (1998) conclude that the Pareto hydrograph for their multi-objective analysis does not bracket the observed data for all low-flow events.

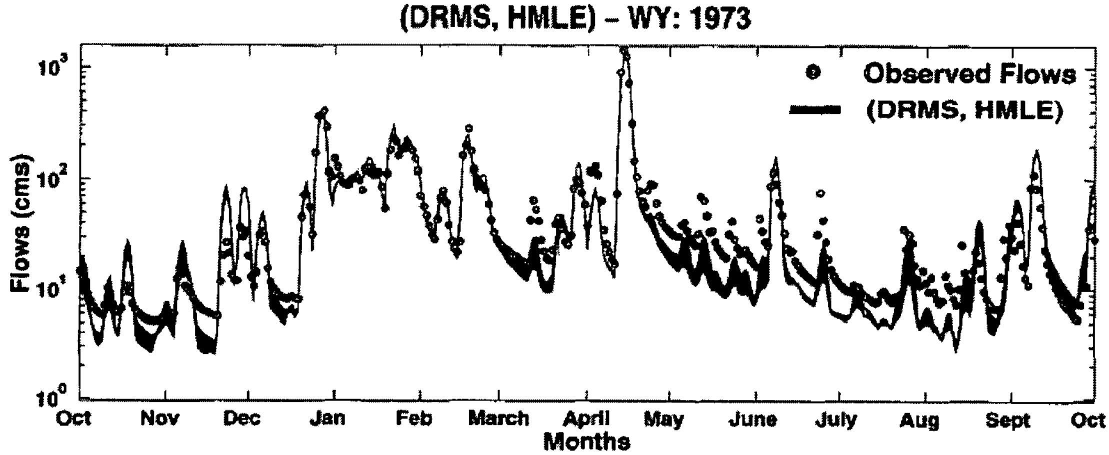


Figure 3-6: Hydrograph Ranges Associated with a Pareto Solution Set
(from p.94, (Yapo et al., 1998))

As for the MCMC calibrations of Vrugt et al. (2003) discussed above, the systematic nature of the deviations observed in Figure 3-6 is an indicator of possible model error. Respecting the non-commensurate nature of the results, the authors do not attempt to quantitatively combine the Pareto solutions. The final product is in fact the set of non-commensurate hydrographs, which collectively provide a qualitative description of uncertainty.

The most common approach to generating a non-inferior set is to separate the multi-objective calibration into a set of single-objective optimizations. This approach has the advantage of including in the Pareto set the "best" solution for each individual objective (Bastidas et al., 2001). The process can be computationally prohibitive, given that applied engineering problems may have hundreds of thousands of feasible extreme points and tens of possible objective functions (Revelle et al., 1997; Yapo et al., 1998). For this reason, Madsen (2003) advocates pursuing a good estimate of the Pareto frontier rather than an exhaustive identification of the full Pareto set.

Breaking the multi-objective problem down into a series of single-objective optimizations requires that each optimization be assigned a unique objective function. These objectives can be defined using one of two possible "generating techniques". The two methods are generally referred to as the constraint method and the weighting method.

The constraint method optimizes the objectives one by one, with all other objectives held at pre-selected values. The Pareto set is populated by successively optimizing each objective while iteratively varying the values of the constrained objectives. This method is most efficient when information on desired outcomes can be used to guide the selection of constraint values for each objective (Revelle et al., 1997). The constraint method yields only an approximation of the non-inferior set, since the *a priori* selection of constraint limits may preclude the identification of the extreme points of the feasible region in objective space.

The weighting method additively combines multiple objective functions using variable weighting factors to form a "grand objective function" (Revelle et al., 1997). To balance the grand objective function, transformation constants are applied to normalize the magnitudes of the component objective functions before the weighting factors are considered (e.g., Madsen, 2000). The non-inferior set is then obtained by repeatedly optimizing the grand objective function while systematically varying the weighting factor assigned to each objective. The weighting method is often more amenable to hydrologic modelling than the constraint method because of its suitability for automatic methods of optimization (Madsen, 2003).

Revelle et al. (1997) present two cautions pertaining to the weighting method. Firstly, if the variation of the weights is too coarse, the weighting method could fail to identify the full set of non-inferior solutions. Some solutions (e.g., "gap point solutions") may not be found with any combination of weights. Secondly, caution must be used in interpreting alternate optima. Not all alternate optima found while solving for an individual objective will be non-inferior; the non-inferior solutions must be identified by adding infinitesimal weights to the other objectives (ibid.).

Gupta et al. (1998) propose an alternate approach for optimizing multi-objective problems in the form of a dedicated algorithm called Multi-Objective COMplex evolution (MOCOM-UA). The authors contend that MOCOM-UA provides an effective and efficient estimate of the non-inferior solution space in a single run, without resorting to subjective weightings (ibid.). The MOCOM-UA method is based on an extension of the SCE-UA algorithm, combining controlled random search, competitive evolution, Pareto ranking, and a new strategy for multi-objective downhill simplex searches. The various strategies of "population evolution" produce rapid convergence without sacrificing global search ability, making the approach ideal for solving multi-objective problems such as hydrologic model calibration (Yapo et al., 1998). Comparisons to more standard methods show that MOCOM-UA provides a fairly consistent approximation to the Pareto set.

3.3.4 Generalized Sensitivity Analysis

Generalized Sensitivity Analysis (GSA) differs from the above techniques by attempting to identify factors governing model behaviour rather than quantitatively exploring the feasible region. GSA was originally developed by Spear and Hornberger (1980) as an attempt to identify key parameters in water quality models, but has since diversified into other fields, including runoff prediction. The GSA method uses MCS to randomly sample parameter values from a specified distribution, repeating the process for a number of simulations. The user then classifies each simulation in terms of whether or not it mimics known system "behaviour". Finally, tests of statistical similarity are used to determine whether parameter values associated with the "behaviour" and "non-behaviour" classes are statistically different. Naturally, additional steps are required if input parameters or data are correlated (Takyi, 1991).

Advantages of the GSA approach include simplicity, flexibility, and enforced declaration of arbitrary assumptions. However, it may be difficult to examine behavioural aspects of multivariate joint distributions. Also, the apparent ease of applying GSA may mask the need for underlying vigorous analysis of model hypotheses (Beck, 1987). Hornberger et al. (1985) demonstrate that GSA can introduce significant subjectivity through arbitrary selection of the "behavioural" and "non-behavioural" thresholds. Harlin and Kung (1992) also note the lack of a clear limit separating acceptable and unacceptable simulations. Seibert and McDonnell (2002) suggest that qualitative information and other forms of "soft" data could be of use in defining appropriate thresholds.

Widespread use of GSA in model-building is due to its ability to quickly and easily classify the relevance of certain parameters or processes for achieving the desired result (Beck, 1987). As such, it is most useful in cases where model structure uncertainty is dominant. For this reason, GSA has seen more application in water quality modelling than hydrologic modelling, since even the processes and parameters of environmental models are often unknown (Takyi, 1991).

A variety of techniques for sample generation, classification of behaviour, and statistical comparison have been used within the GSA framework. Because of its simplicity, it is easily implemented into studies of uncertainty in conjunction with other techniques. Hornberger et al. (1985) present a simple analysis which effectively demonstrates the GSA approach with multiple objectives. After several hundred simulations, they visually compare parameter value distributions for the lowest and highest 30% of objective function values. Significant differences between the distributions indicate that the model is sensitive to that parameter, as measured by the corresponding objective function. Certain parameters exhibit consistent trends across objective functions, clustering into the same sub-range of the parameter space for all "good" simulations. The "good" cluster regions for other parameters differ markedly from one objective function to the next (ibid.).

Beven and Binley (1992) take GSA a step further with their Generalized Likelihood Uncertainty Estimation (GLUE) procedure. The GLUE procedure uses GSA results to produce quantitative estimates under uncertain conditions. It proposes that all "non-behavioural" results be identified and discarded. For the remaining "behavioural" points, the objective used to evaluate behaviour is normalized such that its sum over all "behavioural" points is unity. Where predictions are required, simulations are run with each of the "behavioural" parameter sets. A weighted mean estimate and qualitative probability bounds for the variable of interest can then be calculated.

3.3.5 Equifinality

Limiting an investigation of model predictive uncertainty to a single predetermined model or group of parameters artificially constrains the scope of the work and, as such, can preclude a full exploration of the problem (Braben, 1985; Welles, 1984). Researchers such as Caissie and El-Jabi (2003) advocate exploring all available approaches (e.g., models) rather than exhaustively investigating any single method. The concept of "equifinality" proposed by Beven (2000) champions this view. The basic premise of equifinality is that no model that works acceptably well can be rejected; rather, the each acceptable model should be viewed as "equifinal" (i.e., non-inferior) to all other acceptable models.

Equifinality is less a quantitative approach than a conceptual philosophy. For example, traditional statistical theory prefers to minimize the risk of Type I error (rejection of a potentially true hypothesis) while accepting Type II error (acceptance of a potentially false hypothesis). In the same sense, the concept of equifinality minimizes the risk of rejecting a good model while accepting the risk of retaining a poor one.

Equifinality rejects the pursuit of a single "optimal" model as both impossible and undesirable (Beven, 2001). Beven (2000) explains that, for a given catchment, many models are often compatible with available data and observed processes; there is no basis for removing any such model from consideration. For example, in a study of eight different runoff-response functions for the HBV conceptual model, Harlin (1992) concludes that most of the functions could be calibrated to simulate floods as well as the original.

It is reasonable to expect that equifinal models will be scattered across the possible range of models and parameter sets. While all equifinal models will have some commonality of function, being in some sense similar to the real catchment, some predictive uncertainty will be irreducible (Beven, 2002).

The greatest practical difficulty in applying equifinality lies in the conceptual and computational intensity required to identify the suite of behavioural models (Beven, 2000). Specifically, establishing an exhaustive model space is conceptually difficult, and identifying behavioural models in the model space requires significant computational effort. While computationally intense, identifying an equifinal set of models and parameter combinations is conceptually straightforward, and is similar to the GSA procedure. Like the GLUE approach, equifinality essentially disregards any models found to be inappropriate or "non-behavioural".

Beven (p. 9, 2001) proposes the following blueprint for determining equifinality amongst physically-based models:

(i) Define the range of model structures to be considered.

(ii) Reject any model structures that cannot be justified as physically feasible for the catchment of interest.

(iii) Define an appropriate range for each parameter in each model.

(iv) Reject any parameter combinations that cannot be justified as physically feasible.

(v) Compare the predictions of each potential model with the available observed data and reject any models which produce unacceptable predictions, taking account of estimated error in the observations.

(vi) Make the desired predictions with the remaining successful models to estimate the range of possible outcomes.

The "blueprint" laid out above necessarily recognizes the importance of high-quality, representative data in evaluating model feasibility, since equifinality cannot address or compensate for poor spatial representation or data error (Beven, 2001, 2002). It is important to note that the blueprint contains neither assumptions nor equations. Instead, the equifinal set embraces all independently-developed models, regardless of their character or paradigm.

Equifinality analysis will not necessarily provide a quantitative outcome. Beven (2002) reminds us that all models could conceivably be rejected at Stage (ii) of the blueprint. Further testing or conditioning of the model set using additional information may also lead to the rejection of all models as non-behavioural, since good simulation of catchment outflow does not necessarily imply a good simulation of internal hydrological processes (Beven, 2000). Similar to multi-objective calibration, a trade-off will be required in many cases between successful reproduction of some observations and poor reproductions of others (ibid.). However, Beven (2001) cites the potential for insights and progress even in cases where all models are rejected as unsatisfactory. Complete rejection of all models requires that the study re-examine the suitability of each model structure and reconsider the relationships between measures of performance and desired outcomes.

Beven (2000) proposes combining the predictions of equifinal models into a cumulative distribution function. While substantial insight could be gained from this approach, it also has several weaknesses. Most significantly, it requires that arbitrary probabilities be assigned to each solution. As discussed for GSA in Section 3.3.4, there is also substantial subjectivity in defining the acceptability criteria. For example, accepting and rejecting models based solely on their simulation of catchment discharge can yield a broader range of "behavioural" models than if multiple diverse or distributed measurements were considered (Beven, 2002). The results of

Weiler et al. (2003) support this line of reasoning; the authors improve parameter identifiability for event water transfer functions by combining multiple methods for hydrograph separation in the TRANSEP model.

There is also potential for reviewers to assume that the final set of non-rejected equifinal models represents a comprehensive and exhaustive (or at least unbiased) range of outcomes; in reality, this is true only if the inventory of included uncertainties is exhaustive (ibid.). A better alternative for interpreting results is to view them as a non-commensurate set of multi-objective solutions.

3.3.6 Uncertainty Isolation

Model predictive uncertainty is usually far too complex to be tackled *en masse*. Uncertainty isolation is a term used herein to refer to approaches that attempt to isolate specific aspects or components of uncertainty and examine their influence on results under various conditions. Some studies accomplish this by ignoring or trivializing any uncertainties beyond their limited scope. In general, however, uncertainty isolation is achieved through simplification of some aspect of the hydrologic modelling process.

If a model cannot adequately emulate a simple system, it is questionable whether better performance can be expected for more complex systems (Gan, 1987). Therefore, one could argue that acceptable simulation of a simple system is a necessary but not sufficient milestone for model validation. The data, the model structure, or both can be simplified. Simplifying a model structure also usually simplifies its parameter representation.

Duan et al. (1992) use uncertainty isolation in their design testing of the SCE-UA method, applying it to calibrations of the SIXPAR model, a simplified six-parameter version of the SAC-SMA conceptual model. It is critical to recognize that, although the authors reduce parameter uncertainty, they neither attempt nor claim to achieve a net reduction of model predictive uncertainty.

Harlin (1992) applies the principle of uncertainty analysis differently, using data selection rather than model structure simplification to isolate a subset of model processes. His study focusses on rain floods to gain a clearer picture of the behaviour of his model's runoff-response function without the complicating influence of the model's snow routine.

Hydrologic models commonly exhibit pronounced sensitivity to data error, uncertainty, and spatial variability. If data uncertainty can be eliminated, any uncertainty in the output must result from parameter uncertainty and model structural limitations (Gan and Burges, 1990a). However, critical reviews have identified persistent data problems even for heavily-instrumented experimental catchments like Goodwin Creek (e.g., Steiner et al., 1999) and Hydrohill (e.g., Jakeman and Hornberger, 1993). It is not surprising that a large number of studies attempt to sidestep data uncertainty entirely.

One approach for avoiding data uncertainty is to focus on relative changes in model output resulting from controlled changes in the model input. This approach assumes that error and uncertainty are equally present in both the original and modified simulations, and therefore will not affect conclusions based on relative results. In one example of such a study, Bashford et al. (2002) investigate the effects of scaling and aggregating data by comparing output based on 30-metre gridded data to results obtained using the same input data discretized to a 1-km scale. Their focus is on examining the differences between the two sets rather than establishing absolute accuracy. Other examples include Lumb and Linsley's comparison of streamflow series based on measured and mathematically-augmented rainfall data (Lumb and Linsley, 1971), and the exploration of the effects of land-use changes on flow regime by Kuchment and Gelfan (2002).

Other studies elect to use virtual or synthetic data, which can be assumed error-free for experimental purposes. In such cases, model results are evaluated against a simulated "true" data series rather than against actual field data. In drawing conclusions from studies utilizing virtual data, Bashford et al. (p. 310, 2002) caution that a virtual hydrologic reality may not be complete, accurate or realistic in its process representation.

Klemeš (2000a) proposes that all models be subjected to testing with synthetic data produced by an exact model of a hypothetical system. In its simplest interpretation, this procedure has become known as synthetic calibration, and is the most easily identifiable example of uncertainty isolation. Synthetic calibration uses a specified set of parameter values to generate an arbitrary output series. These synthetic data then take the place of observed data in an automatic calibration. To be successful, the synthetic calibration must identify a definitively globally-optimal solution consisting of the parameter values used to generate the "observed" data. To ensure a realistic calibration process, synthetic calibration should be based on realistic input data and parameter values (Bashford et al., 2002; Seibert, 2000).

Synthetic calibration is most commonly used for evaluating the performance of automatic calibration algorithms. However, it can also provide insight into model structure uncertainty by highlighting any systematic failures with regard to various performance characteristics. If synthetic calibration fails to successfully reproduce the synthetic "truth" data, the pairing of calibration tool and model cannot be expected to perform well in the presence of real-world uncertainties. However, a synthetic calibration that reproduces the "true" streamflow time series without identifying the target parameter values is a half-success at best, and is likely an indicator of overparameterization in the model structure. For example, Seibert (2000) finds that some optimized parameters differ from their "true" values by 10% or more, despite near-perfect convergence of the objective function.

Beven (2002) reminds readers that processes in hydrologic models are not always an accurate reflection of their *in situ* counterparts. Uncertainty isolation approaches have been used in various attempts to explore the resulting model structure uncertainty. For example, the hypothetical case of a small, hydrologically simple catchment with uniform properties and processes could be modelled quite accurately using existing physically-based models (Gan and Burges, 1990a). Therefore, it should be possible to design a realistic virtual catchment for which a given physically-based model is absolutely accurate and completely certain (e.g., Gan, 1987).

Gan and Burges (1990a) explain how a mathematical model of a virtual catchment with known properties, inputs, and outputs can take the place of an ideal experimental catchment. By creating a virtual hydrologic reality which can be exhaustively explored through a mathematical model, the authors in effect create an ersatz duplicate of the controlled conditions of a field-scale, fully-contrived, repeatable laboratory experiment. Assuming the physically-based model is fully representative of the virtual catchment, fluxes calculated by the physical model detail what would actually occur in this catchment under the same input conditions. This effectively eliminates data uncertainty (Gan and Burges, 1990a).

Weiler and McDonnell (2004) pursue a similar goal using virtual experiments of hillslope-scale hydrologic processes. They refer to their virtual experiments as "numerical experiments with a model driven by collective field intelligence" (p. 6, ibid.). Rather than using a detailed physical model as a baseline truthing tool, the authors provide a qualitative discussion of their results in relation to other findings documented from field experiments. In general, results of the virtual experiment are found to correspond well with field observations.

Output from virtual experiments such as those of Gan and Burges (1990a, b) and Weiler and McDonnell (2004) can be used in the same manner as results from a laboratory experiment: to confirm the structure and performance of other models, and to evaluate the uncertainties introduced by various simplifying assumptions (e.g., Troch et al., 1993). Weiler and McDonnell (2004) suggest that the results of a virtual experiment be viewed in the context of equifinality (Section 3.3.5) as an "equifinality reducing instrument". In particular, a strong case can be made that extending the work of Gan and Burges (1990a) to include a rigourous multi-objective analysis (e.g., SCEM-UA) would account for both parameter and data uncertainty. If this were achieved, any residual deviations between the Pareto hydrograph and the "observed" data would theoretically be due to model uncertainty alone.

3.4 Uncertainty and Extreme Event Simulation

The crux of uncertainty in extreme event simulation is the relative dearth of detailed data and observations for historical events. The lack of records can be attributed to the infrequency, unpredictability, and power of extreme events, whose sheer magnitude often overwhelms observation equipment. In the absence of extensive quantitative study of extreme events, it is not surprising that there is no scientifically proven means of estimating them; hydrologic model simulations of extreme events are highly uncertain. Nonetheless, hydrologic models are widely used because they represent the best tool available, are based on scientifically plausible if not proven assumptions, and can make use of the best information available (Micovic, 2003a).

Modellers and decision-makers both need to be aware of the uncertainties involved in extreme event simulations. Decisions should make explicit allowance for uncertainty rather than accepting at face value the arbitrary precision of deterministic model results. Decision makers must fundamentally realize that precise estimates of extreme or design events are both unattainable and inappropriate. At the same time, modellers must act to prevent non-experts from adopting the view that extreme events like the PMF have a near-infinite return period and an infinitesimal probability of occurrence (NRC, 1985). To this end, Klemeš (1992) suggests replacing discrete estimates of extreme event magnitude and probability with the more defensible concept of a window of credibility in probability-magnitude space.

Deterministic estimates for extreme precipitation (e.g., the PMP) are particularly subject to problems of uncertainty. Smith et al. (1996) demonstrate that the storm transposition procedure historically recommended by the World Meteorological Organization for PMP analyses is physically implausible in certain situations. Similarly, the work of Bingeman (2001) supports the existence of an upper limit for precipitation and suggests that "traditional" PMP methods can result in overestimation of the PMF. Wind-induced precipitation undercatch also causes problems when parameter values calibrated against undercatch-laden data are used to generate predictions using deterministic precipitation inputs. All of the above considerations seem to support the argument of Arnaud et al. (2002) that extreme flows are generally overestimated.

Since the best parameter set arising from nominal calibration and validation is unlikely to be optimal for extreme or design flood conditions, the largest floods of record are often used in a final calibration stage to "fine tune" parameter values (Harlin, 1992; Harlin and Kung, 1992). Current practice at BC Hydro is to perform an event-based calibration over the largest discrete events of the historical record prior to simulating the PMF for any watershed (Micovic, 2003a). However, even the largest events of record are often far removed from the hydrologic conditions of the extreme event. One must therefore question to what extent an extreme event model can be said to be "validated".

Applying a hydrologic model to predict extreme floods implicitly presupposes that the mechanisms driving observed behaviours are simulated correctly (Seibert, 2000). However, there can be fundamental differences between processes controlling normal and extreme hydrologic responses. In fact, some researchers have argued outright that hydrologists do not have a clear idea of what to route in extreme situations (Córdova and Rodríguez-Iturbe, 1983).

Runoff during an extreme precipitation event (i.e., one of very low probability) is almost always dominated by overland flow; regardless of the significance of other processes under normal conditions, they are almost always relegated to minor roles (Kouwen, 2003). It is not surprising, therefore, that Gan and Burges (1990b) show that catchments dominated by surface runoff respond more "accurately" to extreme precipitation than their subsurface-dominated counterparts. Differences between uniform and non-uniform precipitation can become less important as precipitation events become spatially extensive and storage effects become small (ibid.; Arnaud et al., 2002). Woolhiser et al. (1996) urge caution in extrapolating models based on situations in which the active processes differ from those of the target application.

Alpine basins prone to debris flows and mass wasting are a good example of a situation for which the validity of model extrapolation is obviously questionable. Jakob and Jordan (2001) show that a purely hydrologic approach will underestimate the magnitude of design floods in mountainous regions due to the dominance of geomorphic processes for the highest observed peak discharges. Harlin and Kung (1992) present a more specific example, demonstrating how the influence of one particular parameter increases considerably when their model is extrapolated to extreme events. Garen and Burges (1981) also identify some parameters as insensitive except under extreme conditions. More generally, BC Hydro (Kroeker et al., 2003) has observed that different models, equally "well-calibrated" by their respective designers, can diverge significantly in extrapolation range to extreme events.

Klemeš (p. 17, 1986b) argues that the incautious extrapolation of models to problems beyond far beyond their established capabilities has "created a false impression that hydrology has answers to problems which may remain beyond its reach for decades to come or whose solution lies outside its framework". This situation may result in grossly incorrect estimates of hydrological conditions leading to poor decisions, which may in turn lead to a cynical waning of support for genuine hydrologic research. Ultimately, this cycle hinders the ability of hydrological science to provide better answers (ibid.).

4. Research Tools and Methods of Analysis

"Prediction is very difficult, especially if it's about the future."
- Nils Bohr

Any critical analysis of hydrologic modelling will undoubtedly identify model calibration as highly significant in its potential effects on uncertainty. Most hydrologic modellers share the curiosity of O'Connell and Todini (p. 7, 1996) as to whether "alternative but physically acceptable parameterizations, consistent with the available information, [can] still give rise to the same set of responses". In the context of extreme flood estimation, the question may be extended to include the degree of variability that these "alternative parameterizations" introduce into estimates of extreme floods such as the PMF. This thesis develops and demonstrates one possible approach for exploring these issues, focussing on the questions:

- What kind of variability in parameter values is associated with alternative "acceptable" calibrations?

- How much variability arises when the parameter values identified by these alternative calibrations are used to estimate extreme flood events?

As a pre-requisite step towards the larger goal of reducing overall model predictive uncertainty, this work investigates the impact of subjective decisions made during calibration on model predictive uncertainty for extreme events. Multiple automatic calibrations of a conceptual hydrologic model are conducted using different measures of performance, resulting in a collection of non-inferior parameter sets. This approach is demonstrated using hydrologic data for the Coquitlam Lake and Illecillewaet River watersheds in British Columbia.

Each non-inferior parameter set for the Coquitlam Lake watershed is then used to simulate an extreme event based on data provided by BC Hydro. The combined output of these extreme event simulations characterizes the relative variability in the hydrographs.

Simulations are conducted using the University of British Columbia Watershed Model (UBCWM), which is widely used to describe and forecast watershed behaviour in mountainous areas of British Columbia. Calibrations of the UBCWM utilize the Shuffled Complex Evolution Algorithm (SCE-UA), an effective and efficient optimization-based automatic calibration routine. Because automatic calibrations do not capture the level of knowledge inherent in a manual calibration, the extreme event hydrographs obtained using the alternative parameter sets are compared on a relative rather than absolute basis. This work focusses on exploring uncertainty, and as such does not attempt to produce a practicable calibration of the UBCWM for nominal or extreme-flood prediction.

Section 4.1 of this chapter presents a brief review of the UBCWM, SCE-UA, and the watersheds selected as case studies. A more detailed description of the experimental approach is given in Section 4.2.

4.1 Tools and Case Studies

4.1.1 The University of British Columbia Watershed Model

The UBCWM (Quick et al., 2003) has become BC Hydro's primary tool for forecasting watershed behaviour in mountainous areas. The UBCWM was chosen as the conceptual model for this study to ensure that results are useful both in an applied context as well as in the more general study of model predictive uncertainty. Also, technical support for the UBCWM is available locally through UBC and BC Hydro. The following brief overview of the model is intended to provide the reader with a working level of familiarity. For a more detailed description, the reader should refer to the UBCWM User's Manual (Quick, 1994).

The UBCWM is a conceptual model designed to balance simplicity against physical realism (Micovic, 1998). Topographic and land use data are required to define watershed properties. Once the watershed is set up, the model requires only precipitation and temperature data as inputs. Streamflow data are necessary for calibration; Micovic (2003a) recommends a minimum ten years of daily data to ensure reliable calibration and validation. The UBCWM is primarily intended to model mountain runoff resulting from snowmelt, glacial melt, and rainfall (Zhu, 1997). In example applications, the model exhibits better performance for a snowmelt-driven watershed (91% efficiency) than for a rainfall-driven watershed (80% efficiency) (Quick, 1995). As always, watershed complexity and data accuracy tend to govern model performance.

Although the UBCWM can be classified as a lumped model, it nonetheless has a distributed component. The UBCWM subdivides a watershed into elevation bands to capture orographic gradients of precipitation and temperature. Precipitation, snowmelt and ET are calculated independently for each elevation band. An analysis of the dependence of routing parameters on watershed area leads Micovic (1998) to conclude that the channel phase can be neglected in the routing calculations for small and medium-sized catchments. Therefore, a simple summation of runoff from all bands is used to calculate outflow. Since the UBCWM contains both distributed and lumped elements, it is useful to distinguish the model from either group by referring to it as "quasi-distributed".

Model timing is based on precipitation and temperature inputs, and can assume either a daily or hourly basis (Quick, 1995). UBCWM can operate as either an event-based or continuous model. For continuous simulations, a one-year spin-up period is commonly used with initial conditions set to zero and the first year of results not considered in the analysis. Initial conditions for event-based (hourly) calibrations are typically determined by excerpting data from longer-term continuous (daily) model simulations (Micovic, 2003b).

While primarily focused on streamflow estimation, the UBCWM can also provide estimates of snowpack, soil moisture, groundwater storage, geographic runoff contributions, and surface-subsurface contributions (Quick, 1995). The accuracy of the model in a forecasting role depends largely on accurate reproduction of snowpack, soil moisture and groundwater storage (Quick, 1995).

Every simulation using the UBCWM involves three major subcomponents of the model structure:

- The meteorological subroutine distributes input data throughout the catchment. This is the most important component due to its control of moisture input and snowmelt.

- The soil moisture subroutine calculates evaporation and runoff, apportioning the runoff between four response modes: fast (e.g., surface); medium (e.g., interflow); slow (e.g., upper groundwater); and very slow (e.g., deep groundwater). The response modes are conceptually analogous to their physical examples, but lack a physical basis in computation.

- The watershed routing subroutine calculates the time distribution of runoff from each of the four response modes apportioned above. Each response mode is subjected to storage using a cascade of linear reservoirs. Typical time constraints range from less than one day for fast runoff to 100-200 days for very slow runoff. An effective calibration must involve all four of the runoff response modes described above to ensure the respective timing parameters have all been optimized. Outflow from the four response modes is summed for each timestep.

These three components are applied consecutively and the output from one becomes the input to the next. Various other engines and utilities can be called on through the user interface to assist in data preparation and analysis.

Temperature data are used to determine snowmelt, evaporation, and precipitation form. Parameters required to calculate snowmelt, evaporation, and temperature lapse rates are all pre-calibrated. Snowmelt can be driven by either energy budget or degree-day approximations (Quick, 1995).

The UBCWM calculates ET in three stages. First, potential ET is calculated for each station based on observed temperatures. Potential ET is then distributed across the elevation zones using temperature lapse rates as a conversion mechanism. Finally, potential ET is converted to actual ET using soil moisture as a conditioning factor (Quick, 1995). Unlike many other models, the UBCWM does not attempt to explicitly calculate soil moisture. Instead, it relies on a running calculation of soil moisture deficit (SMD). The UBCWM calculates soil moisture deficit by adding the antecedent soil moisture deficit to the current ET demand at each current timestep, then subtracting infiltration (ibid.). The possibility for infinite soil moisture is avoided by conditioning actual ET toward zero as SMD increases.

Water distribution processes are controlled by user-calibrated parameters (Quick, 1995). Moisture priorities are managed to first satisfy any runoff from impervious areas, then replenish the soil moisture deficit, followed by apportioning to groundwater, and finally, interflow. The UBCWM assumes that "flash" behaviour (i.e., Horton overland flow) occurs when total rainfall for a timestep exceeds a user-specified value. The threshold must be set and activated manually for a storm event. Snowmelt does not contribute to the flash hydrograph, though it can affect whether or not flash behaviour occurs (ibid.).

The UBCWM User's Manual (Quick, 1994) proposes limiting a calibration to a subset of parameters (listed in Table 4-1) and setting all other parameters to constant, pre-determined values. Micovic (1998) contends that even some of the parameters in Table 4-1 may provide acceptable results when treated as pre-determined constants.

For this study, only those parameters listed in Table 4-1 were considered for calibration. All topographic parameters for the watershed case studies have been reviewed in detail by others and are assumed constant at the values specified in the expert calibrations. Where applicable, the maximum and minimum values for each parameter given in Table 4-1 correspond to limits recommended in the UBCWM User's Manual (Quick, 1994). These limits reflect ranges recommended by the model developers as typical for most catchments, but do not necessarily reflect mathematical or physical limits for each parameter. Wherever possible, these ranges are used to define the feasible parameter space for calibration.

The recommended approach for calibrating UBCWM is to use a three-stage sequential procedure, with each stage corresponding to one of the major subcomponents of the UBCWM identified above. Each stage of calibration is characterized by a different group of parameters (Micovic, 2003a; Quick, 1995). Firstly, the modeller should focus on resolving the water balance by adjusting parameters relating to the meteorologic subroutine (i.e., parameters P0GRADL, P0GRADM, P0GRADU, E0LMID, E0LHI, P0SREP, and P0RREP). Once the modeller has achieved an acceptable simulation of the water balance, the focus should then shift to matching the peaks and recessions of the observed hydrograph by adjusting the parameters of the soil moisture subroutine (i.e., parameters C0IMPA, Delta, P0AGEN, P0PERC, and P0DZSH). Finally, the modeller focuses on "fine tuning" model response by adjusting the

Table 4-1: Commonly Calibrated Parameters of the UBCWM

Name	Description	Typical Range
C0IMPA	Fraction of impermeable area in each elevation band	0.0 - 1.0
Delta	Increase in impermeable area between elevation bands	0.0 - 1.0*
P0GRADL	Precipitation gradient (lapse rate) below E0LMID (%)	0 - 20
P0GRADM	Precipitation gradient (lapse rate) below E0LHI (%)	0 - 20
P0GRADU	Precipitation gradient (lapse rate) above E0LHI (%)	0 - 20
E0LMID	Elevation separating P0GRADL and P0GRADM (m)	site specific
E0LHI	Elevation separating P0GRADM and P0GRADU (m)	site specific**
P0AGEN	Impermeable area modification factor	100
V0FLAS	Flash flood threshold (mm)	20 - 40
P0PERC	Groundwater percolation (mm/day)	0 - 50
P0DZSH	Deep zone fraction of groundwater	0.0 - 1.0
P0FRTK	Fast runoff time constant for rain (days)	0.0 - 2.0
P0FSTK	Fast runoff time constant for snowmelt (days)	0.0 - 2.0
P0GLTK	Fast runoff time constant for glacial runoff (days)	0.0 - 2.0
P0IRTK	Interflow time constant for rain (days)	1.0 - 10.0
P0ISTK	Interflow time constant for snowmelt (days)	1.0 - 10.0
P0UGTK	Time constant for upper groundwater runoff (days)	10 - 50
P0DZTK	Time constant for deep groundwater runoff (days)	100 - 300
P0SREP	Adjustment to measured precip when T < 0°C (snow)	-1.0 - 1.0
P0RREP	Adjustment to measured precip when T > 0°C (rain)	-1.0 - 1.0

* must be less than 1 – [C0IMPA / (number of elevation bands – 1)]
** must be greater than E0LMID

watershed routing parameters (i.e., parameters V0FLAS, P0FRTK, P0FSTK, P0GLTK, P0IRTK, P0ISTK, P0UGTK, and P0DZTK). A more detailed discussion of the recommended approach for calibrating the UBCWM is available in the UBCWM Manual (Quick, 1994).

The dominant role of precipitation distribution in the UBCWM necessitates some further explanation of the associated parameters. The two parameters controlling data representativeness (P0SREP for snow, P0RREP for rain) are used to compensate for any systematic trend or discrepancy between precipitation observed at the station and precipitation in the catchment (Quick, 1995). In general, these parameters are used to ensure that the water balance is essentially closed. Further control of precipitation is provided through parameters P0GRADL, P0GRADM, and P0GRADH, which, in combination with breakpoints E0LMID and E0LHI, allow up to three orographic lapse rates to be specified throughout the range of watersheds in the catchment. While the use of five parameters to control precipitation input theoretically allows for full representation of site-specific conditions, the absence of site-specific data to calibrate these values creates potential for compensating errors. This is recognized by the model authors, who suggest that the user begin with a single precipitation gradient and only consider a additional gradients if they prove necessary for obtaining an adequate representation of catchment response (Quick, 1994).

The primary input file for the UBCWM is text-based and contains values for all parameters as well as more general information such as watershed name, area, and location, as well as start and end dates for the simulation. This input file also contains references to input files for precipitation and temperature data. When calibrating and validating the UBCWM, a streamflow file is also specified. Precipitation, temperature, and streamflow files can be written in ASCII text format and converted to UBCWM input format using built-in utilities. A complete description of the input and output file structures is available in the UBCWM User's Manual (Quick, 1994).

4.1.2 The SCE-UA Method

The Shuffled Complex Evolution algorithm developed at the University of Arizona (SCE-UA, Duan et al., 1994a) was selected as the best approach for automatic calibration in this study. The merits and drawbacks of SCE-UA are reviewed in detail in Section 2.4.4. Most importantly, the algorithm has proved to be an effective and efficient means for identifying a global optimum solution. The algorithm is straightforward and logical, with an easy to understand structure. Therefore, both strengths and weaknesses of the algorithm are well-known and their effect on results can be anticipated. Considering the UBCWM in particular, the large number of parameters for which calibration is possible, combined with the possibility of compensating errors and the need for sequential treatment in manual calibrations, present an excellent situation to investigate the capabilities of the SCE-UA algorithm.

Source code for the SCE-UA algorithm was obtained from its principal author, Dr. Qingyun Duan of the Office of Hydrologic Development, National Weather Service, United States National Oceanic and Atmospheric Administration. The SCE-UA method requires a FORTRAN compiler and two text-based input files. The first input file provides parameters for SCE-UA such as the number of complexes, termination criteria, and upper and lower bounds for each model parameter being optimized. The second file contains information specific to the chosen hydrologic model, such as the number of parameters, the index for each parameter being calibrated, and the streamflow data against which the parameters are to be calibrated. A full description of input file structures and how to use SCE-UA is available as part of the SCE-UA distribution package (Duan et al., 1994a).

4.1.3 Interfacing UBCWM and SCE-UA

The original intent of this work was to incorporate the SCE-UA algorithm into the existing UBCWM semi-automatic calibration subroutine by dynamically linking the FORTRAN-based SCE-UA executable with the Visual Basic-based UBCWM. However, the complexity of the UBCWM source code precluded such a task. Instead, the SCE-UA source code was translated into Visual Basic and adapted such that the UBCWM core engine is called from within SCE-UA. Having been translated into Visual Basic, the SCE-UA source code remains available to be integrated into the UBCWM at a later date.

To facilitate automatic calibration, maximum and minimum values (i.e., the feasible range for each parameter) are taken from the recommended values set out by Quick (1994). For the purposes of this study, all parameters are assumed to have no physical basis and the prior pdf for each parameter takes the form of a uniform distribution over the feasible range.

Instructions detailing how to initialize and execute the SCE-UA / UBCWM interface are available from the author of this work.

4.1.4 The Coquitlam Lake and Illecillewaet River Watersheds

The most prolific user of the UBCWM, BC Hydro, has found that the model yields simulations of varying quality across their portfolio of watersheds. To ensure that the choice of watershed does not exert undue influence over the variance of the Pareto set of parameter values, it is desirable to consider at least two hydrologically different watersheds for calibration.

The Coquitlam River watershed is a flashy, rain-dominated watershed in southwestern British Columbia, located about 25 km east of downtown Vancouver, BC. About 15 km upstream of the confluence of the Coquitlam and Fraser Rivers, BC Hydro's Coquitlam Dam impounds the river to form Coquitlam Lake, which serves as a reservoir for both hydropower and municipal water supply. The Coquitlam River watershed above Coquitlam Dam (the Coquitlam Lake watershed) is selected for this study because hourly meteorological data and a completed PMF study are both available from BC Hydro. Perhaps because of the natural variability in its hydrologic response, the Coquitlam Lake watershed is not among the highest-quality simulations achieved by BC Hydro using the UBCWM, which was designed more for snowmelt-dominated catchments.

In contrast, the Illecillewaet River is a higher-elevation watershed in central British Columbia with a strong freshet that dominates the annual hydrograph. Consistently good results have been obtained applying the UBCWM to the Illecillewaet River watershed; in fact, the Illecillewaet River watershed is distributed as an example calibration with the UBCWM program (Quick et al., 2003). Therefore, the Illecillewaet River watershed is chosen to establish whether a watershed "better suited" to simulation with UBCWM would achieve more consistent performance across the various objective functions. Hourly and PMF data are not available for the Illecillewaet River watershed; therefore, the event-based application of the various calibrations is confined to the Coquitlam case study.

The Coquitlam Lake watershed is 187.8 km² in area and ranges in elevation from 153 m to 1773 m. The watershed does not include any glaciated areas. Characterized by steep alpine slopes, the watershed is mostly forested and has very little development. The mountains of the

Coquitlam basin present the first barrier to warm, moist flows of air approaching from the southwest. The resulting orographic effect is responsible for the most significant precipitation events (BC Hydro, 2004). Periodic fronts of warm moist air from the Pacific Ocean are particularly common during the winter months, and the heaviest precipitation events are observed during this period. Snowpacks can vary significantly throughout the season as snow built up during cooler periods is depleted by the warmer temperatures associated with Pacific disturbances (ibid.). The most significant streamflow events result from heavy precipitation with temperatures above zero at all elevations, and may include snowmelt-generated runoff. The dominant frontal nature of prevailing storm systems typically limits major streamflow events to less than four days (ibid.).

Meteorologic data for the Coquitlam Lake watershed are collected by two Data Collection Platforms (DCPs) maintained by BC Hydro. These are located at the Coquitlam Dam forebay (elevation 161 m) and above Coquitlam Lake (elevation 290 m). BC Hydro has constructed a more reliable data set by averaging values for the two DCPs into a single series, with missing data replaced through linear regression of DCP data from beyond the watershed boundaries. Only one value is calibrated for each of the UBCWM parameters P0SREP and P0RREP, given that there is only one data series for precipitation. Reservoir inflows are estimated by measuring the change in reservoir stage and computing the corresponding change in storage, then adjusting for measured releases and abstractions.

BC Hydro's UBCWM calibration for the Coquitlam Lake watershed uses a total of eleven elevation bands (BC Hydro, 2004). The synthetic (averaged) meteorological data set is assigned an elevation of 225 m for calculating precipitation and at elevation 550 m for calculating temperatures. The impermeable fraction of the lowest elevation band is set at 0.9 to represent the presence of Coquitlam Lake. The calibration algorithm (using C0IMPA and Delta) requires that the percentage of impermeable area increase by a constant, non-negative amount (Delta) for each successively higher elevation band. To allow flexibility in specifying impermeable area at the lowest and highest elevations, the user can specify low (starting) and high (ending) bands as limits for the calibration. The fraction of impermeable area will only be calibrated for those bands between the specified low and high benchmarks; impermeable area fractions for bands outside the specified range are fixed at pre-determined values. The algorithm assigns the lowest specified band an impermeable area percentage equal to C0IMPA, and adds Delta to a running sum for each band until the highest specified band is reached. In this study, the fraction of impermeable area for the Coquitlam Lake watershed is calibrated for elevation bands three through eleven, since the expert calibration does not follow the required pattern for elevation bands one, two, and three. This simplification is not believed to have a profound impact on the experiment.

The UBCWM algorithm is used to calculate temperature lapse rates and precipitation gradients. A full enumeration of the *a priori* parameter values can be found in the UBCWM input file produced by BC Hydro (2004).

The Illecillewaet River is a tributary to the Columbia River, originating at the Illecillewaet Glacier on the western flank of the Selkirk Mountains in Canada's Glacier National Park. The watershed is bounded by other glaciers, including Albert Glacier to the south and Dismal and Durrand Glaciers to the northwest (Micovic, 1998). The river flows southeast for approximately 60 km, dropping around 1500 m to meet the Columbia River at Revelstoke, British Columbia (Columbia Mountains Institute, 2003). BC Hydro's Arrow Lakes reservoir submerges the original confluence at full pool (ibid.). The rugged watershed occupies an area of approximately 1200 km², three quarters of which is covered by boreal forest (Micovic, 1998). Elevations within the watershed range from 520 m to 3260 m, providing a strong orographic effect on precipitation; mean annual precipitation increases from 950 mm at lower elevations to exceed 2500 mm at higher locations. In an average year, over two thirds of precipitation falls as snow, giving the typical annual hydrograph a strong freshet-driven peak in the late spring and early summer. The largest floods typically result from snowmelt and occur during this period (ibid.). However, the largest flood event of record occurred in July 1983, and was triggered by intense rainfall in the catchment. Thus, summer and early fall rainstorms are still an important flood-producing mechanism.

Meteorologic data for the Illecillewaet River watershed are available for three sites formerly maintained by Environment Canada. These are located at Revelstoke, BC (elevation 440 m), Rogers Pass (elevation 1330 m), and Mount Fidelity (elevation 1875 m). BC Hydro now maintains DCPs at Rogers Pass and Mount Fidelity. The UBCWM utilizes precipitation data for all three stations; three values of UBCWM parameters P0SREP and P0REP must therefore be calibrated, one of each parameter for each station. However, the expert calibration provided as an example file with the UBCWM utilizes only two precipitation gradients for the Illecillewaet, and therefore parameters P0GRADU and E0LHI are not used.

Over 40 years of streamflow records for the Illecillewaet River are available from Water Survey of Canada gauge 08ND013, Illecillewaet River at Greeley. The gauge is still active and has a drainage area of approximately 1170 km² (Environment Canada, 2001).

The example UBCWM calibration for the Illecillewaet River watershed uses a total of eight elevation bands. In the expert calibration, the impermeable fraction of lower elevation bands (one through five) is set at 0.1, possibly to represent the more permeable nature of the well-forested lower-elevation areas. The expert calibration shows an increasing fraction of impermeable area with elevation for elevation bands six through eight, presumably reflecting the disappearance of vegetation and soil cover in the upper watershed. For the calibrations performed in this study, the fraction of impermeable area for the Illecillewaet River watershed is held constant at 0.1 for elevation bands one through five and calibrated for elevation bands six through eight. The UBCWM algorithm is used to calculate temperature lapse rates and precipitation gradients. A full enumeration of the *a priori* parameter values can be found in the example input file distributed with the UBCWM installation package (Quick et al., 2003).

4.2 Experimental Design

Model predictive uncertainty must be minimized for hydrologic modelling to be an effective prediction tool. This section describes the experimental approach for characterizing a sub-component of model predictive uncertainty, specifically the variability in parameter values associated with alternative "acceptable" calibrations. The experiment also explores the effect of this "acceptable" parameter variability on extreme event simulation. A multi-objective calibration of the UBCWM provides alternative calibrations for the Coquitlam Lake and Illecillewaet River watersheds. The resulting Coquitlam Lake parameter sets are then used to estimate runoff from an extreme event based on a PMF scenario for Coquitlam Lake.

Section 4.2.1 describes the selection of calibration objectives, and Section 4.2.2 provides details regarding the calibration process. A brief discussion of validation is provided in Section 4.2.3. Sections 4.2.4 and 4.2.5 discuss the application of the Pareto set of calibrations to several observed and extreme events, respectively.

4.2.1 Selection of Objectives

Ideally, a multi-objective study such as this should address the full range of possible "objectives" presented in Section 3.3.3, i.e., different methods and measures for evaluating performance, as well as different types of calibration data collected at various points throughout the watershed.

However, the selection of the UBCWM and SCE-UA limits the options somewhat. The SCE-UA method is selected as the primary method for calibrating the model; other approaches (e.g., manual calibration or genetic algorithms) are not considered so as to maintain a manageable scope of work. Further, the UBCWM acts as a lumped model in its calculation of runoff; therefore, each application of the UBCWM must be calibrated against a single streamflow time series. Consideration of multi-site streamflow measurements is not possible for a single instance of the model.

Both the Coquitlam and Illecillewaet watersheds include active snowcourses, and these data play a significant role in the professional calibrations for these watersheds. However, calibration against the available snowpack data is neglected herein to simplify the procedure.

Measures of performance (i.e., objective functions) selected for multi-objective calibration include Simple Least Squares (SLS) and the Heteroscedastic Maximum Likelihood Estimator (HMLE), both of which were already established within the SCE-UA source code. Added for the purposes of this study are Nash-Sutcliffe Efficiency (E!) and the Coefficient of Determination (D! or R^2), both of which are widely used within the hydrologic literature. The Least Absolute Difference function is also considered. Mathematical definitions of these functions are provided in Table 2-1. Also included is the objective function EOPT!, a commonly used measure of performance for the UBCWM. However, review of the properties of the EOPT! statistic identified the potential for inconsistencies introduced by the use of absolute values. From Section 2.4.5, the formula for determining EOPT! is:

$$EOPT! = E! - | \Sigma q_{obs} - \Sigma q_{sim} | / \Sigma q_{obs} \quad (7)$$

In the above formula, E! refers to the Nash-Sutcliffe Efficiency statistic; q_{obs} and q_{sim} refer to the series of observed and simulated flows, respectively. As discussed in Section 2.4.5, the intent of the modification to E! is to compensate for volume error. However, mathematical analysis reveals that compensating errors can exist in the formula, since

$$| \Sigma q_{obs} - \Sigma q_{sim} | \neq | \Sigma q_{obs} | - | \Sigma q_{sim} | \quad (8)$$

Therefore, a new function, labelled EOPT', is defined and incorporated with the formula

$$EOPT' = E! - \Sigma | q_{obs} - q_{sim} | / \Sigma q_{obs} \quad (9)$$

This modified calculation essentially combines the Nash-Sutcliffe efficiency statistic (E!) with a normalized measure of least absolute difference. Recognizing that the least absolute difference term could easily dominate the value of EOPT', another modification of EOPT! was introduced to provide a function having a greater relative weighting on the Nash-Sutcliffe statistic. This last objective, defined as EOPT (no modifier), was included with the formula

$$EOPT = E! - \Sigma | q_{obs} - q_{sim} | / \{ \Sigma q_{obs} * n \} \quad (10)$$

where n is the number of data points in the summation. Thus, eight objective functions were ultimately used for the study. Objective functions were implemented as minimization problems at the recommendation of the developers of the SCE-UA method (Duan et al., 1992).

4.2.2 SCE-UA Calibrations

The test plan for SCE-UA calibrations begins with controlled conditions and progresses through more complex scenarios. This section describes the suite of tests.

Preliminary (or viability) testing, which focusses on demonstrating the basic functionality of the program, is undertaken to verify that the SCE-UA program and the UBCWM have been properly integrated. In the context of this study, preliminary testing could be carried out using synthetic data for any arbitrary watershed since the goal of this phase of testing is only to prove that the algorithm is able to converge to the correct solution given a specified parameter set and synthetic data. The UBCWM representation of the Campbell River watershed, a second example distributed along with the Illecillewaet River watershed in the UBCWM installation package, is chosen for preliminary testing because its calibration involves a smaller set of parameters than the Coquitlam or Illecillewaet models.

The preliminary testing phase is also used to identify the optimum number of SCE-UA complexes to use as a basis for automatic calibrations. Results from the preliminary testing are summarized in Section 5.1.

Following successful preliminary tests, synthetic calibrations for each of the eight candidate objective functions are conducted for the Coquitlam Lake watershed. The goal of this next phase is to determine which (if any) of the eight objective functions are suitable for calibration using actual data from the subject catchment. As described in Section 3.3.6, a successful synthetic calibration proves that a given automatic calibration approach is capable of finding a globally optimal parameter set, if one exists. The calibrations simulate a synthetic series of Coquitlam Lake inflow data for water years 1989-90 through 1993-94. An additional synthetic calibration is performed, applying objective function EOPT' over the entire period for which input data are available (i.e., water years 1985-86 through 1999-2000) to determine the effect of a longer data set on algorithm convergence. Results for the synthetic calibrations of parameters active in the Coquitlam Lake case study are presented in Section 5.2.

Simulation of the Illecillewaet River watershed also begins with a suite of synthetic calibrations. The calibration is performed for water years 1981-82 through 1988-89. Results for the eight synthetic calibrations are presented in Section 5.2. As for Coquitlam, a final synthetic calibration is performed utilizing all available data (water years 1970-71 through 1989-90) and objective function EOPT' to determine the degree to which lengthening the calibration data series affects the parameter estimates.

All objective functions producing successful synthetic calibrations are then used in calibrations against actual measurements of discharge for the Coquitlam and Illecillewaet Rivers. Results for these calibrations are presented in Section 5.3. Coquitlam calibrations are performed against BC Hydro's calculated inflows to Coquitlam Lake. Data from water years 1985-86 through 1998-99 are used for the calibrations. Since there are no observed step-changes in the data sets and no profound data non-stationarity documented in the BC Hydro calibration reports, it is believed that the use of longer period of calibration data is conservative in nature. The suite of calibrations for the Illecillewaet River are performed using available WSC streamflow data for Water Years 1981-82 through 1988-89.

Some parameter values in BC Hydro's calibrated model of the Coquitlam Lake watershed lie beyond the feasible region set out in the UBCWM User's Manual (Quick et al., 1994). Therefore, a second series of Coquitlam Lake watershed calibrations are performed with the feasible parameter space expanded for all parameters not constrained by mathematical limits (i.e., those that are not fractions, or subject to non-negativity assumptions, etc.). The revised ranges for the UBCWM parameters are shown in Table 4-2.

In any study of a single aspect of uncertainty in modelling, it is necessary to account for as much external subjectivity as possible to isolate the desired focus; in this case, the choice of objective function must be isolated. Therefore, a final set of simulations are necessary to confirm that results of a calibration for a given objective function are independent of the random seed used to initialize the SCE-UA algorithm.

To this end, multiple calibrations are conducted using a single objective function, with the only difference being the value of the SCE-UA random seed parameter ISEED. Within SCE-UA, the random seed itself is used only to initialize the random number generation routine of Press et al. (1992).

Five synthetic calibrations are run for each of the Coquitlam and Illecillewaet basins, and five more for each watershed use observed streamflow data. The Coquitlam calibrations use data from water years 1989-90 through 1993-94, while the Illecillewaet calibrations use water years 1981-82 through 1988-89. Each calibration uses a quasi-random value of ISEED either selected arbitrarily or from the intrinsic random-number generating function in Microsoft Excel. Results of these "seed sensitivity" trials are presented in Section 5.4.

The suite of SCE-UA calibrations carried out in this work is summarized in Table 4-3 below.

Table 4-2: Expanded Parameter Ranges for Coquitlam Lake Watershed Calibrations

Name	Description	Range
C0IMPA	Fraction of impermeable area in each elevation band	0 to 1
Delta	Increase in impermeable area between elevation bands	0 to 0.5
P0GRADL	Precipitation gradient for elevations below E0LMID (%)	0 to 30
P0GRADM	Precipitation gradient for elevations below E0LHI (%)	0 to 30
P0GRADU	Precipitation gradient for elevations above E0LHI (%)	0 to 30
E0LMID	Elevation separating P0GRADL and P0GRADM (m)	site specific
E0LHI	Elevation separating P0GRADM and P0GRADU (m)	site specific
P0AGEN	Impermeable area modification factor	1 to 200
V0FLAS	Flash flood threshold (mm)	0.1 to 100
P0PERC	Groundwater percolation (mm/day)	0 to 100
P0DZSH	Deep zone fraction of groundwater	0 to 1
P0FRTK	Fast runoff time constant for rain (days)	0 to 5
P0FSTK	Fast runoff time constant for snowmelt (days)	0 to 5
P0IRTK	Interflow time constant for rain (days)	0.1 to 10
P0ISTK	Interflow time constant for snowmelt (days)	0.1 to 10
P0UGTK	Time constant for upper groundwater runoff (days)	1 to 500
P0DZTK	Time constant for deep groundwater runoff (days)	1 to 500
P0SREP	Adjustment to measured precip when T < 0°C (snow)	-1 to 1

Table 4-3: Index of SCE-UA Calibrations

Phase	Subject Watershed	Period of Record	Total # of Calibrations	Objective Functions
Preliminary	Arbitrary (Campbell)	seven years (1984-1990)	24	all (10, 15, 20 complexes)
Synthetic	Coquitlam	five years (1990-1994)	8	all
	Coquitlam	full record (1986-2000)	1	EOPT'
	Illecillewaet	eight years (1982-1989)	8	all
	Illecillewaet	full record (1971-1990)	1	EOPT'
Actual Data	Coquitlam	1986-1999	8	all
	Coquitlam (expanded parameter space)	1986-1999	8	all
	Illecillewaet	1982-1989	5	SLS, EOPT', E!, LAD, EOPT
Seed Sensitivity	Coquitlam (synthetic)	five years (1990-1994)	10	EOPT', HMLE
	Coquitlam (actual)	five years (1990-1994)	5	EOPT'
	Illecillewaet (synthetic)	eight years (1982-1989)	6	EOPT'
	Illecillewaet (actual)	eight years (1982-1989)	5	EOPT'

4.2.3 Validation

Simulated hydrographs from all calibrations are compared to observed hydrographs for all available water years to ensure there are no aberrant results. Although this constitutes a form of split-sample validation (except in cases where the entire record was used for calibration), explicit validation against observed data is not the goal. Results were examined only for general compatibility.

Although Section 2.4.6 emphasizes the importance of validation for practical applications, theoretical studies of uncertainty in calibration are not – and should not be – constrained by validation to the same degree. There is little point in documenting the accuracy of absolute performance when the goal is to quantify variability across a set of "equally successful" calibrations (as measured by performance in the calibration stage). In such a study, validation leading to the rejection of some models as inferior could potentially influence the results by reducing the overall variability of the Pareto set.

In a staged investigation such as this, synthetic calibration can be used as a form of validation for the calibration routine itself, since any objective function failing to converge to the synthetic solution is unlikely to achieve a good simulation under normal levels of uncertainty and can thus be dropped from consideration. However, a good synthetic calibration is not necessarily indicative that calibration with that particular objective function will be successful. In recognition of this, and notwithstanding the caveat noted above, if model results and parameter sets clearly indicate that a calibration is unable to produce a good simulation of the watershed, the related parameter sets are typically not considered in further analysis.

4.2.4 Event Simulations

Only after completing all of the above processes can one produce a limited assessment of the variability arising from choice of calibration objective. As stated earlier, the goal of this study is to investigate the degree of variability associated with extrapolating these results to the scale of a PMF event. However, there is also insight to be gained from an analysis of lesser events; specifically, how the variability across a set of continuous calibrations changes when the daily timestep is reduced to one hour. Event-based simulations for this study are performed only for the Coquitlam Lake watershed. Initial conditions for each event are excerpted from a continuous daily simulation of the watershed.

BC Hydro's PMF study for Coquitlam Lake includes hourly data for several discrete storm events, including those dates as shown in Table 4-4 (BC Hydro, 2004). These events are among the largest events of record for the Coquitlam Lake watershed. In this study, each of the five events is simulated with all validated parameter sets. Results and comparisons against both observed data and BC Hydro's expert calibration are presented in Section 5.5.

Table 4-4: Start and End Dates for Major Inflow Events to Coquitlam Lake

Storm ID	Start Date	End Date	Observed Peak Flow (m^3/s)
Storm 1	7 November 1990	14 November 1990	601
Storm 2	21 November 1990	27 November 1990	737
Storm 3	6 November 1995	10 November 1995	650
Storm 4	15 March 1997	22 March 1997	716
Storm 5	31 December 2001	12 January 2002	528

Setting up the initial conditions for each event requires running the daily UBCWM calibration to generate an input (forecasting) file on each of the start dates listed above. If one proposes to use modelled (or otherwise approximate values) for initial conditions, a preliminary investigation of the corresponding sensitivity in the output can provide insights into how much effort should be invested in estimating the initial state of the system. To this end, a preliminary set of initial conditions is created using BC Hydro's long-term calibration.

Although initial conditions should arguably be independent of the parameter set being used for event simulation, this is not always the case; it is conceivable that different parameter sets could produce different forecasting files for the same date. Therefore, two simulations of Storm 1 are performed for the final calibration of each objective function, one using initial conditions calculated from its own calibration and one using initial conditions obtained by applying BC Hydro's parameter set. The results of this comparison are presented in Section 5.5. For simplicity, all other storms are simulated using initial conditions obtained using the BC Hydro parameter set. Any obviously aberrant simulations of Storms 2 through 5 are re-run with function-specific initial conditions to ensure that the observed behaviour is not a result of an inappropriate combination of initial conditions and parameter values.

In interpreting the results of event-based simulations, the reader should be aware that any reference to the "BC Hydro expert calibration" refers to the continuous parameter set produced and not the event-based, updated set of parameter values actually used to model the Coquitlam Lake PMF. There is little to be gained from comparing SCE-UA calibrations against simulations produced using BC Hydro's fine-tuned, event-oriented PMF parameter set, since the SCE-UA calibrations are not subjected to the extra calibration step.

4.2.5 Simulations based on a PMF Scenario

Because they are so extreme, events on the scale of the PMF require special consideration during hydrologic simulation. The modeller should be aware of the issues discussed in Section 3.4, but most especially the fact that few if any PMF-scale simulations can be adequately validated. Nonetheless, there is sometimes little alternative but to accept the uncertainties and proceed with applying a hydrologic model.

BC Hydro devotes significant resources to producing and periodically updating PMF estimates for all of their watersheds. Based on the guidelines produced by the Canadian Dam Association, Table 5-6 of BC Hydro's PMF study for the Coquitlam Lake Watershed (BC Hydro, 2004) lists the various potential PMF scenarios considered, including the following:

- a PMP event combined with a snow accumulation having an Annual Exceedance Probability (AEP) of 0.01 and prevailing storm temperatures;

- a precipitation event having an AEP of 0.01 combined with the Probable Maximum Snow Accumulation (PMSA) and prevailing storm temperatures;

- a precipitation event having an AEP of 0.01 combined with the PMSA and the critically severe temperature sequence; and

- antecedent precipitation (a pre-storm) falling on an average snowpack, followed by a drying period and subsequently by the PMP.

In this case, an estimate of the PMP for southwestern BC has been produced for BC Hydro by Water Management Consultants (2003). BC Hydro has also examined sensitivity to temporal distribution of the PMP to ensure that an appropriate (i.e., extreme, but not absolute worst-case) scenario has been identified. BC Hydro reports that the chosen PMF scenario for the Coquitlam Lake watershed is a late fall PMP event on top of a large (100-year) snowpack. A complete explanation of the derivation of the PMF is documented in the PMF study (BC Hydro, 2004). Inputs and initial conditions for the PMF scenario defined by BC Hydro are adopted as the extreme event case study for this work.

The PMP is a deterministic precipitation event, and in this case is produced as distributed (gridded) data (Water Management Consultants, 2003). BC Hydro adapted this information for the UBCWM by using GIS to overlay the gridded data onto the UBCWM elevation bands. Total band precipitation is calculated for each band, and precipitation parameters are adjusted until each band in the UBCWM simulation receives the amount of precipitation specified by the PMP study. The potential uncertainty in the estimate of the PMP produced by Water Management Consultants (ibid.) is noted to be on the order of 30%.

In a normal calibration, the precipitation parameters of UBCWM are altered to adjust location, state, and volume of precipitation. However, the deterministic nature of the PMP implies that its volume, distribution, and timing of precipitation should be identical for any simulation. Since the UBCWM cannot accept precipitation in distributed form for each elevation band, the model's precipitation parameters must be used to replicate the PMP distribution. Therefore, as part of their PMF studies, BC Hydro adjusts the precipitation parameters of the model until the UBCWM transformation of the representative meteorological input data mimic the distributed PMP (BC Hydro, 2004).

For the UBCWM, precipitation parameters have no role beyond calculating the volume of precipitation and distributing it to elevation bands. The UBCWM's disregard of the channel routing phase implies that runoff from all bands (i.e., rainfall plus snowmelt) is aggregated by simple addition before being divided into fast, medium, slow, and very slow components. Since the antecedent snowpack is specified *a priori* in the initial conditions and, like most major precipitation events in the watershed, the PMP is accompanied by temperatures above freezing at all elevations of the watershed, the precipitation parameters are used only for calculating the volume of precipitation reaching the watershed.

Given a pre-determined initial snowpack and a PMP consisting exclusively of rainfall, it can be argued that the precipitation parameters become disconnected from any other parameters in the calibration. Therefore, since BC Hydro's precipitation parameters are known to accurately reflect the PMP but will not influence any other portion of UBCWM, the values of parameters P0GRADL, P0GRADM, P0GRADU, E0LMID, E0LHI, P0SREP, and P0RREP are taken *en masse* from the final BC Hydro PMF calibration. These values replace their counterpart values (i.e., those determined through SCE-UA calibration) for the suite of PMF-based extreme event simulations. Limiting conditions for temperature lapse rates (i.e., parameters A0TLZZ and A0TLZP) are also taken from the BC Hydro calibration, as these are defined as a physical property of a PMP event.

Similarly, the snowpack having an AEP of 0.01 is nominally defined independent of any UBCWM simulation. Therefore, it is appropriate to adopt the initial snowpack conditions used in the BC Hydro PMF simulation.

The results of the various PMF-based extreme event simulations are presented in Section 5.6.1. While these simulations are based on BC Hydro's PMF scenario for Coquitlam Lake, they do not use BC Hydro's PMF parameter set, and therefore may differ significantly from the actual PMF for Coquitlam Lake.

Although this study requires an extreme storm as input, the choice of storm event is not limited to the PMP. The approach is equally valid using any extreme storm. In this case, the PMP produced by Water Management Consultants (2003) was selected only because it had already been distributed for the UBCWM by BC Hydro. To provide further insight into the behaviour of UBCWM under extreme input, each calibration is also run with the PMP meteorologic data and its own calibrated precipitation parameters. Although the resulting storm event may lack physical basis and is not directly comparable to the PMP, the variability of the hydrographs will be indicative of what the model might forecast were these calibrations used with the *in situ* measurements that would be recorded during the PMP or an event of similar scale. The results of these PMP-based simulations are presented in Section 5.6.2.

5. Results and Discussion

"The fact that there may be no unique answer does not mean that the approach is not science or scientific."

- Keith Beven

Chapter 4 introduces an experimental program to investigate the variability in runoff estimates introduced by alternative "acceptable" calibrations. This chapter discusses the results of that program.

The results of preliminary testing to confirm that the SCE-UA method is an appropriate tool for calibrating the UBCWM are presented in Section 5.1. Section 5.2 describes the results of synthetic calibrations utilizing multiple objective functions for the Coquitlam Lake and Illecillewaet River watersheds. Section 5.3 presents the applied multi-objective calibrations for the watersheds; Sections 5.3.1 and 5.3.2 focus on the Coquitlam Lake watershed (nominal and expanded parameter space, respectively), and Section 5.3.3 on the Illecillewaet River watershed. Section 5.4 explores changes in the final calibrated parameter set resulting from varying the initial random seed. Results for event-based and extreme event simulations for the Coquitlam Lake watershed are detailed in Sections 5.5 and 5.6, respectively.

5.1 Preliminary Tests

This phase of testing focusses on demonstrating the basic functionality of the UBCWM – SCE-UA interface. It requires a series of synthetic data for an arbitrary catchment. The Campbell River watershed (included as an example file with the UBCWM) is chosen for this purpose. Significant results for each calibration include the final parameter set, the final objective function value, and the best and worst function values as well as the parameter range (as a percentage of the original hypercube) for each shuffling loop.

As Table 5-1 shows, the algorithm is ultimately able to correctly and consistently identify the true solution set of parameter values. The successful calibration confirms that SCE-UA is an acceptable tool for use in calibrating UBCWM and is sufficiently robust to calibrate approximately twenty parameters simultaneously.

Table 5-1: Results of Synthetic Calibration for Campbell River Watershed
SLS Objective Function with 15 complexes

Name	Synthetic Value	Calibrated Value
C0IMPA	0.1	0.095
Delta	0.2	0.202
P0GRADL	1	0.999
P0GRADM	2	1.99
P0GRADU	N/A	N/A
E0LMID	927	923
E0LHI	N/A	N/A
P0AGEN	100	100
V0FLAS	32	32.2
P0PERC	21	21.0
P0DZSH	0.55	0.554
P0FRTK	0.39	0.389
P0FSTK	0.4	0.4
P0GLTK	N/A	N/A
P0IRTK	2	2.00
P0ISTK	3	2.98
P0UGTK	26	25.8
P0DZTK	74	73.7
P0SREP	-0.03	-0.03
P0RREP	0.18	0.18

A successful and efficient application of the SCE-UA method requires that the user select an appropriate number of complexes for the algorithm. Although 15 complexes are used in generating the results of Table 5-1, there is no reason to assume that 15 is the optimum value. Additional synthetic calibrations with 10, 15, and 20 complexes are conducted for the Campbell River data set. Figures 5-1(a) through 5-1(c) show example plots of the residual errors in parameter values for the best member of the evolving population at each shuffling step. The residual error for each parameter estimate is calculated as

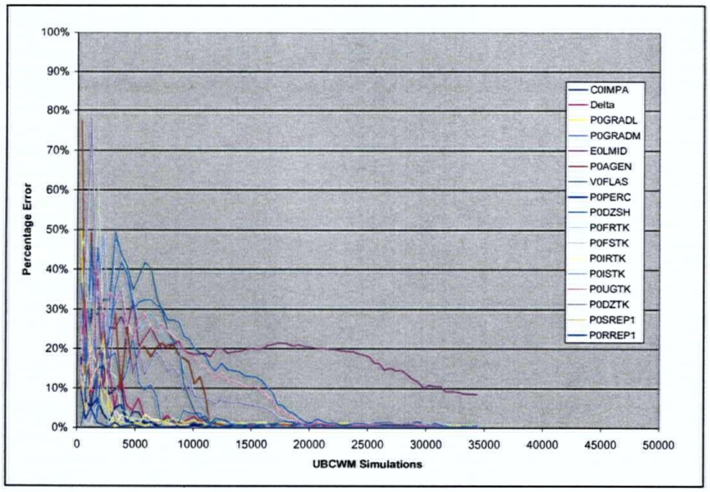
$$\%\,Err_x = | (x - x_{true}) / (x_{max} - x_{min}) | \tag{11}$$

where x is the calibrated parameter value, x_{true} is the synthetic target value, and x_{max} and x_{min} are the upper and lower bound of the feasible range, respectively. All calibrations terminate when the geometric mean of the normalized parameter ranges drops below 0.1%.

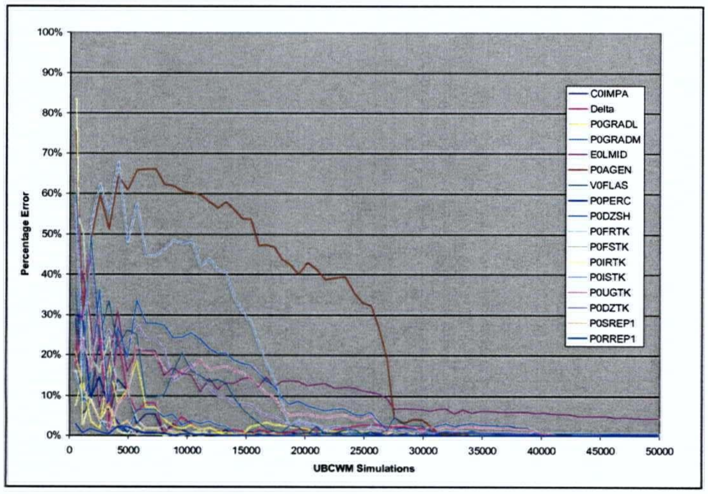
Utilizing ten complexes (Figure 5-1a), convergence is swift and largely satisfactory for all but one parameter. However, in consideration of the fact that the Campbell River model utilizes fewer parameters than the Coquitlam and Illecillewaet models, ten complexes is deemed insufficient for calibrating those catchments.

When 15 complexes are used (Figure 5-1b), the convergence is less swift, although all parameters converge to their target values at approximately 56,000 iterations of the UBCWM. This performance reflects the expected condition of full convergence, thereby proving that SCE-UA is capable of identifying a unique and globally-optimal parameter set for the UBCWM given the specified dimensionality, objective function, and data set.

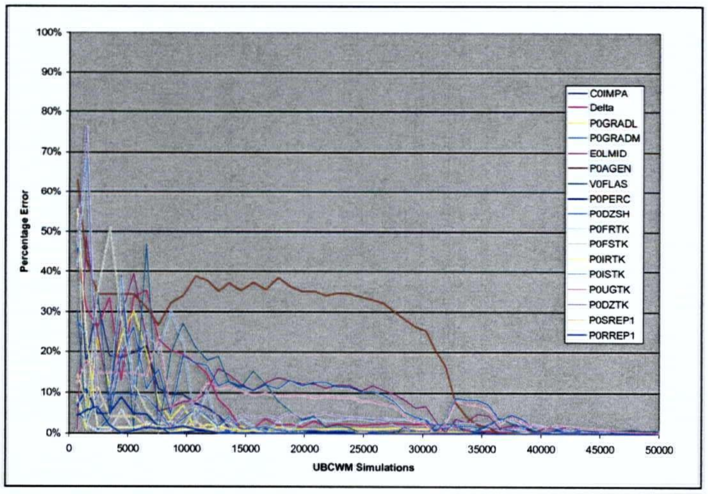
Adopting 20 complexes also leads to convergence of all parameters (Figure 5-1c), and thereby validates the SCE-UA method as an optimization tool for the UBCWM. However, the 43,000 UBCWM simulations required to attain substantial convergence of parameter values represent a substantial savings in efficiency over the 15-complex trial. The superiority of the 20-complex scenario is demonstrated in Figure 5-2, which compares the evolving objective functions for the 10, 15, and 20 complex trials. Although the 10-complex scenario appears arguably more efficient than the 20-complex scenario, the 20-complex scenario is indisputably the most effective of the three at achieving a low objective function value.



(a) with 10 complexes



(b) with 15 complexes



(c) with 20 complexes

Figure 5-1: Parameter Convergence for Synthetic Calibration in Preliminary Testing
Campbell River data set, EOPT' Objective Function

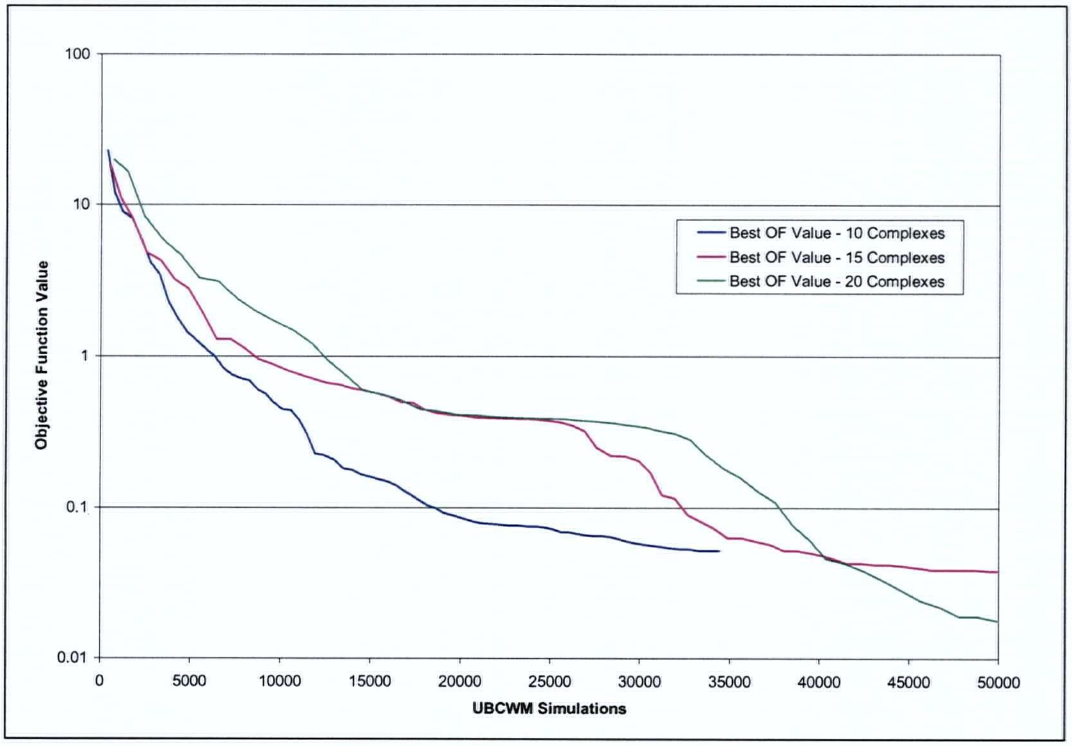


Figure 5-2: Objective Function Evolution for Calibrations with 10, 15, and 20 complexes
EOPT' Objective Function

For these reasons, and in consideration of the fact that the parameter set used to model the Campbell River has fewer calibration parameters than either of the Coquitlam or Illecillewaet models, all subsequent calibrations apply SCE-UA with a total of 20 complexes. This value is recommended by the authors of the SCE-UA method for the solution of very complex problems. Some additional discussion of this and other SCE-UA parameters is provided in Appendix A.

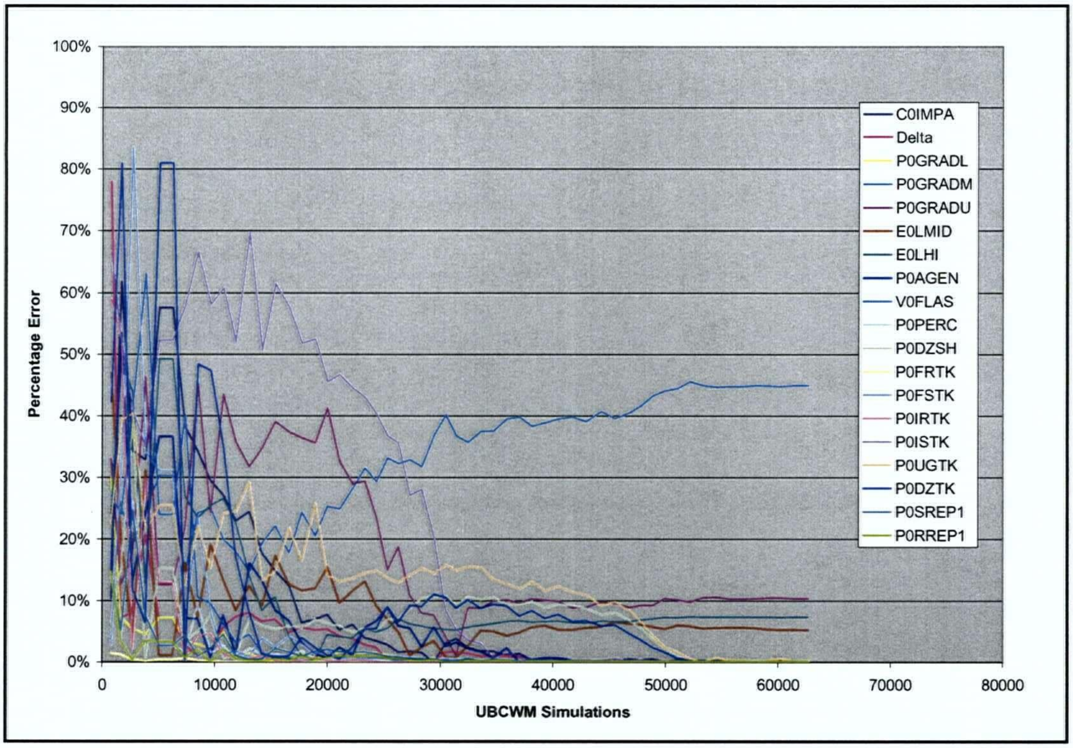
5.2 Synthetic Calibrations

Section 5.1 establishes the general capability of the SCE-UA method as a calibration tool for the UBCWM. In this section, each objective function is re-tested for each watershed to limit subsequent calibrations to only those objective functions capable of producing accurate solutions. All eight objective functions listed in Section 4.2.1 are evaluated independently in synthetic calibration of both the Coquitlam and Illecillewaet watersheds.

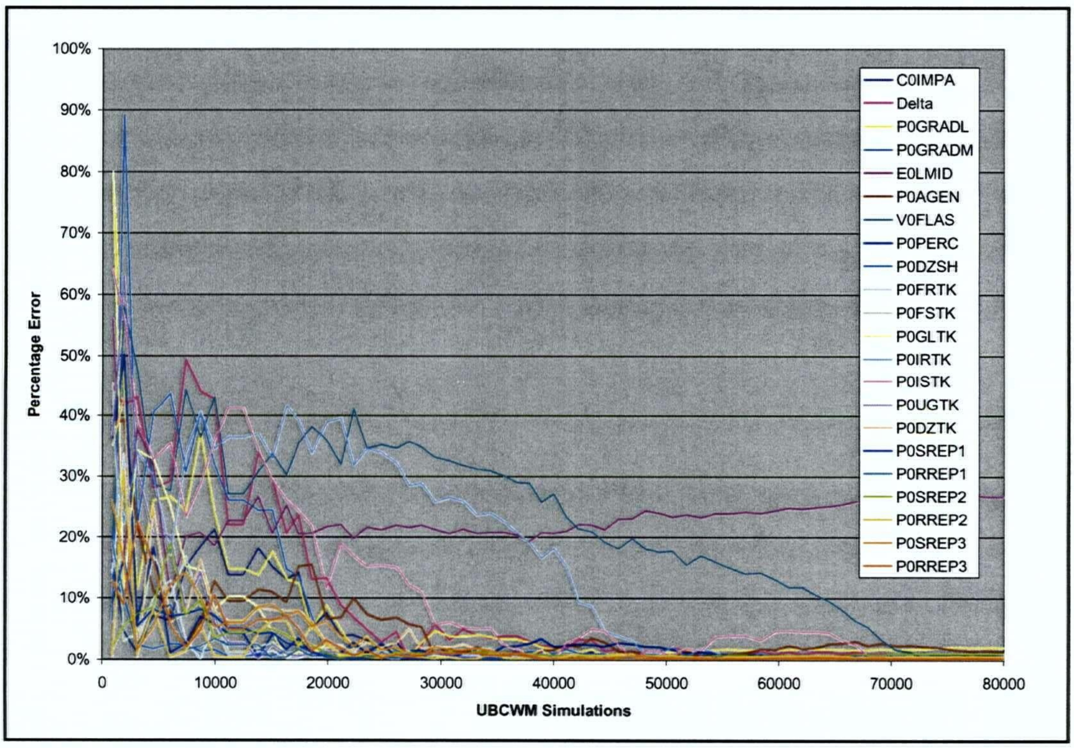
Figures 5-3(a) and (b) show examples of the residual errors in calibrating each parameter of the Coquitlam Lake and Illecillewaet River watersheds, respectively. These plots represent "successful" synthetic calibrations, and are generated using the EOPT' objective function. It is obvious for the Coquitlam and Illecillewaet watersheds that some parameters do not converge to their "true" values. However, the hydrographs shown in Figure 5-4(a) and (b) demonstrate that the synthetic and calibrated hydrographs are visually indistinguishable, and all statistical evaluations of simulation quality essentially reach their limiting values (e.g., $R^2 = 1$). These hydrographs are representative of all water years in the set of successful synthetic calibrations. Regardless of parameter accuracy, calibrations exhibiting such behaviour cannot be rejected without further investigation.

While the success of any synthetic calibration is not necessarily conclusive, the decision to reject unsuccessful calibrations is straightforward. Behaviour indicative of unsuccessful synthetic calibrations is immediately apparent from examination of Figure 5-5; as shown, more than half of the best estimates for individual parameter values in each case have errors exceeding 5%, with some exceeding 50%. Unsurprisingly, Figure 5-6 shows that the corresponding hydrographs are not consistent with the definition of success for a synthetic calibration (i.e., a very close match to the target hydrograph). Objective functions falling into this category include the EOPT! and D! (or R^2). The EOPT! calibrations terminate when the number of UBCWM simulations exceeded 100,000, while the D! calibrations terminate when SCE-UA could not improve the objective function value of the best member of the population over ten shuffling loops. Both EOPT! and D! have weaknesses discussed in previous chapters that account for their poor performance. Given their poor performance in synthetic calibration, these statistics are removed from further consideration.

Additionally, the HMLE calibration performs poorly for the Illecillewaet River watershed yet produces an acceptable synthetic calibration for the Coquitlam Lake watershed. The difference in performance is attributed to the difference in the parameter sets; the parameter space for the Illecillewaet River watershed has more dimensions, with the bulk of the difference focussed in the parameters controlling precipitation distribution. The synthetic calibration of the Illecillewaet River watershed using the HMLE objective function is classified as unsuccessful and the statistic is not considered further in that case study. However, the adequate performance of the HMLE objective function when applied to the Coquitlam Lake watershed requires that the HMLE objective function be carried forward for further study in the Coquitlam case study.

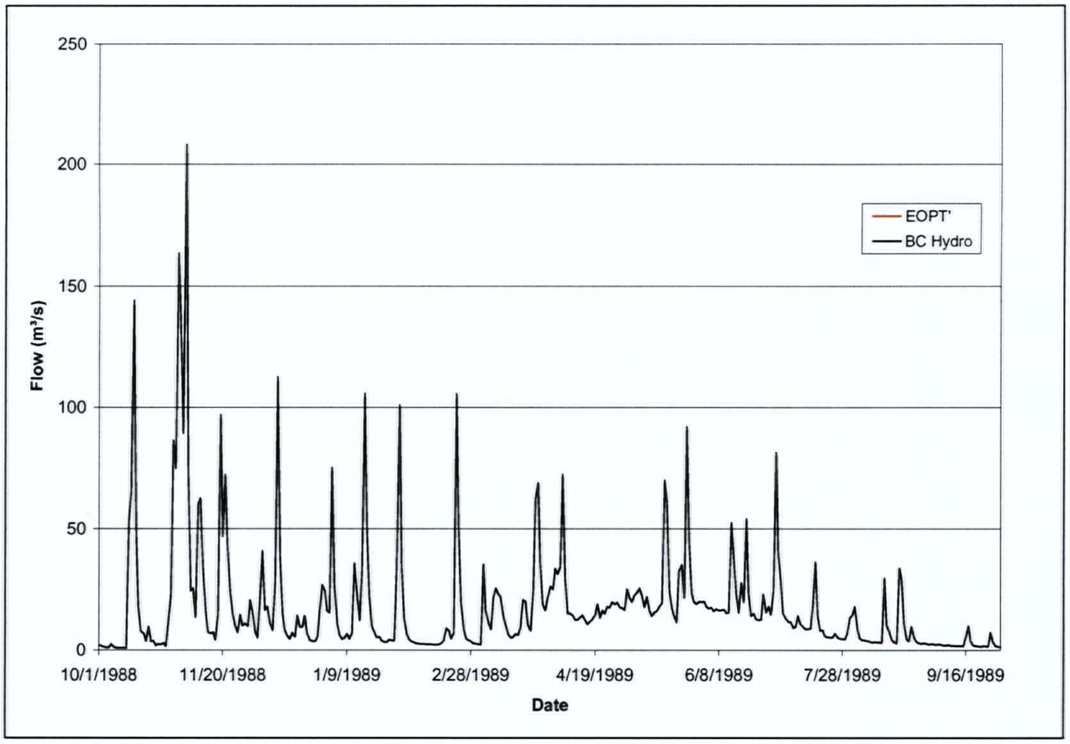


(a) Coquitlam Lake Watershed, EOPT' Objective Function

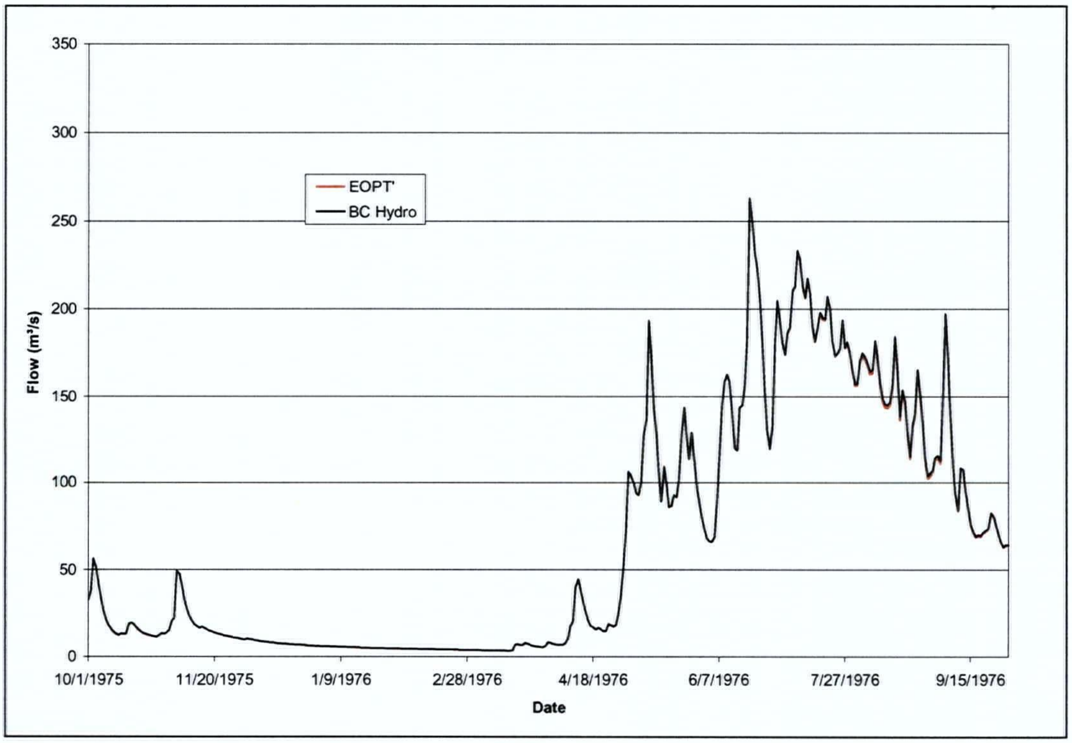


(b) Illecillewaet River Watershed, EOPT'Objective Function

Figure 5-3: Parameter Convergence for Successful Synthetic Calibrations

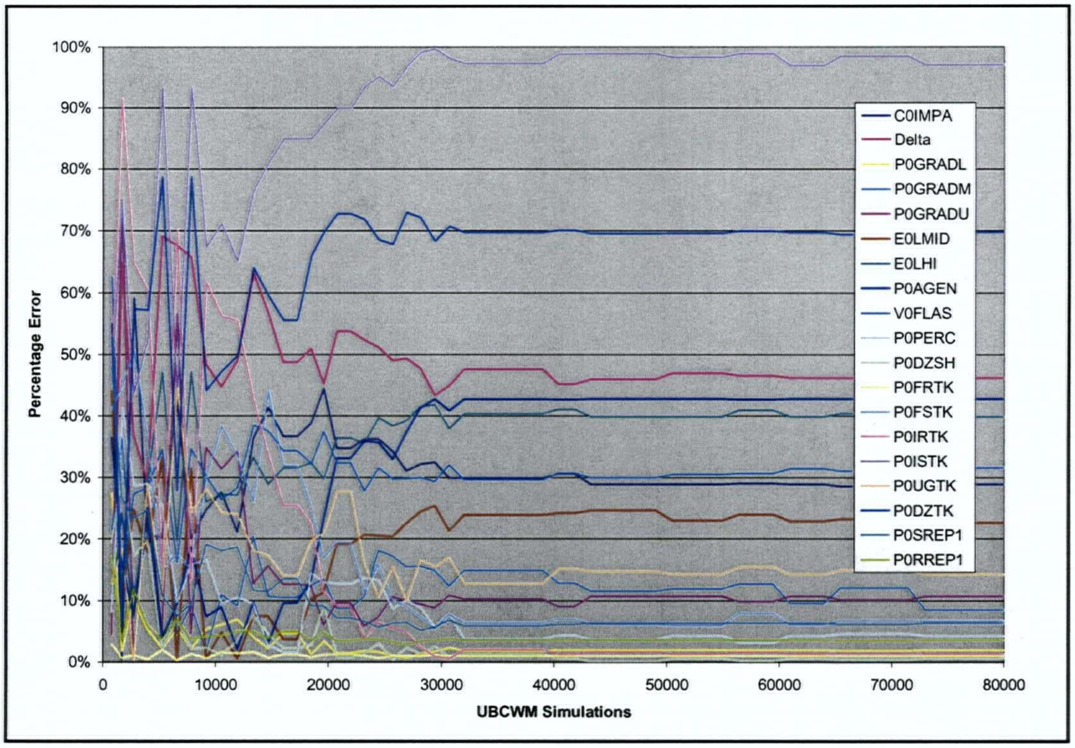


(a) Coquitlam Lake Watershed, EOPT' Objective Function

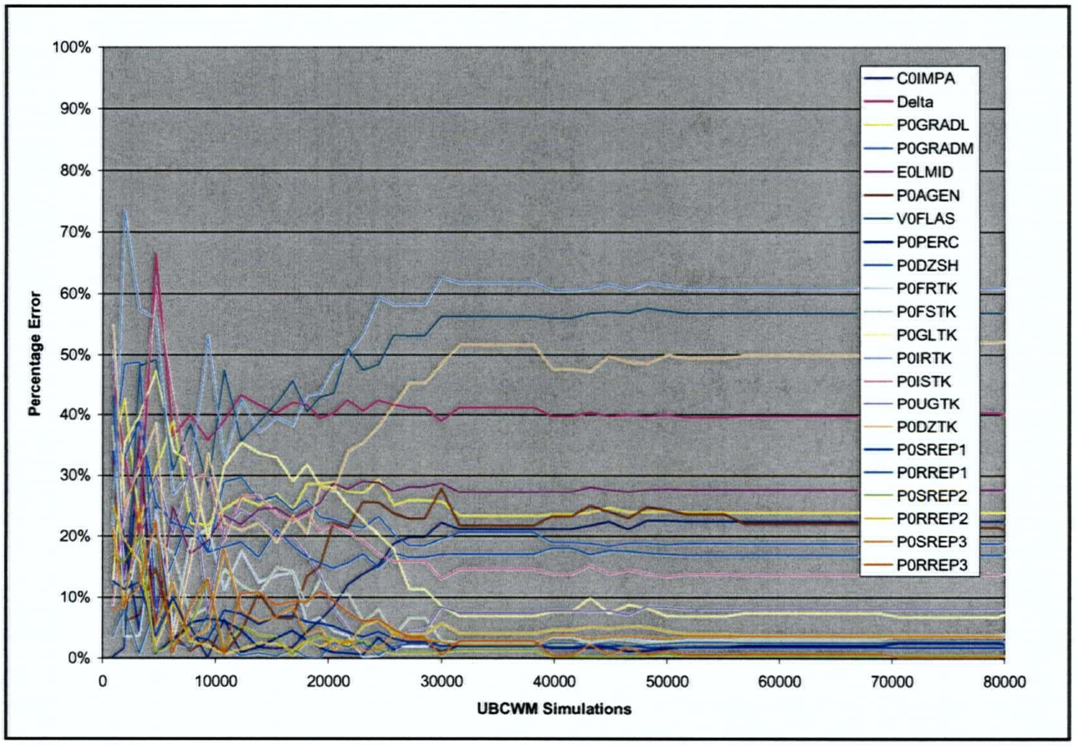


(b) Illecillewaet River Watershed, EOPT' Objective Function

Figure 5-4: Typical Annual Hydrographs for Successful Synthetic Calibrations

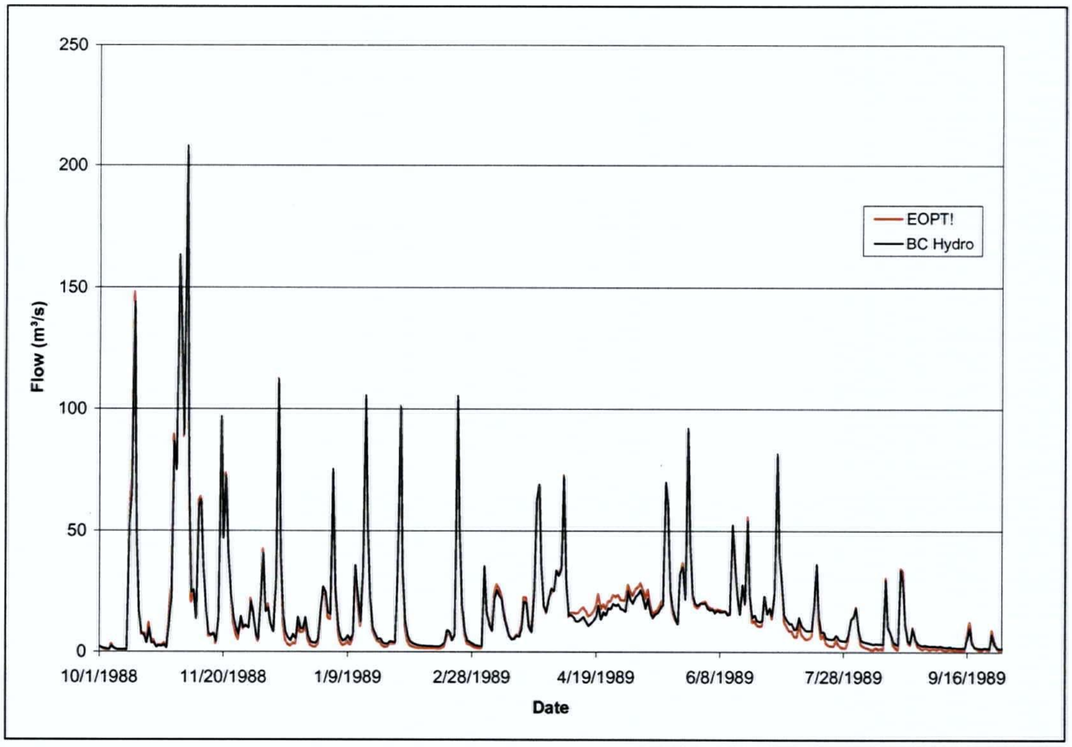


(a) Coquitlam Lake Watershed, EOPT! Objective Function

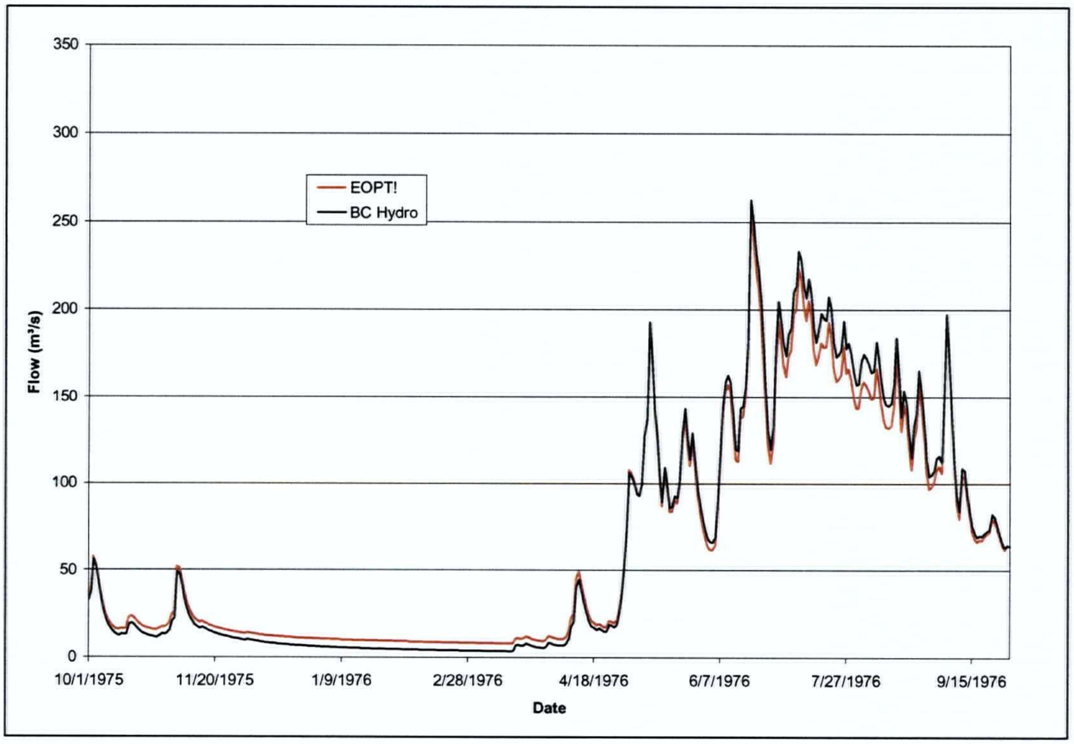


(b) Illecillewaet River Watershed, EOPT! Objective Function

Figure 5-5: Parameter Convergence for Unsuccessful Synthetic Calibrations



(a) Coquitlam Lake Watershed, EOPT! Objective Function



(b) Illecillewaet River Watershed, EOPT! Objective Function

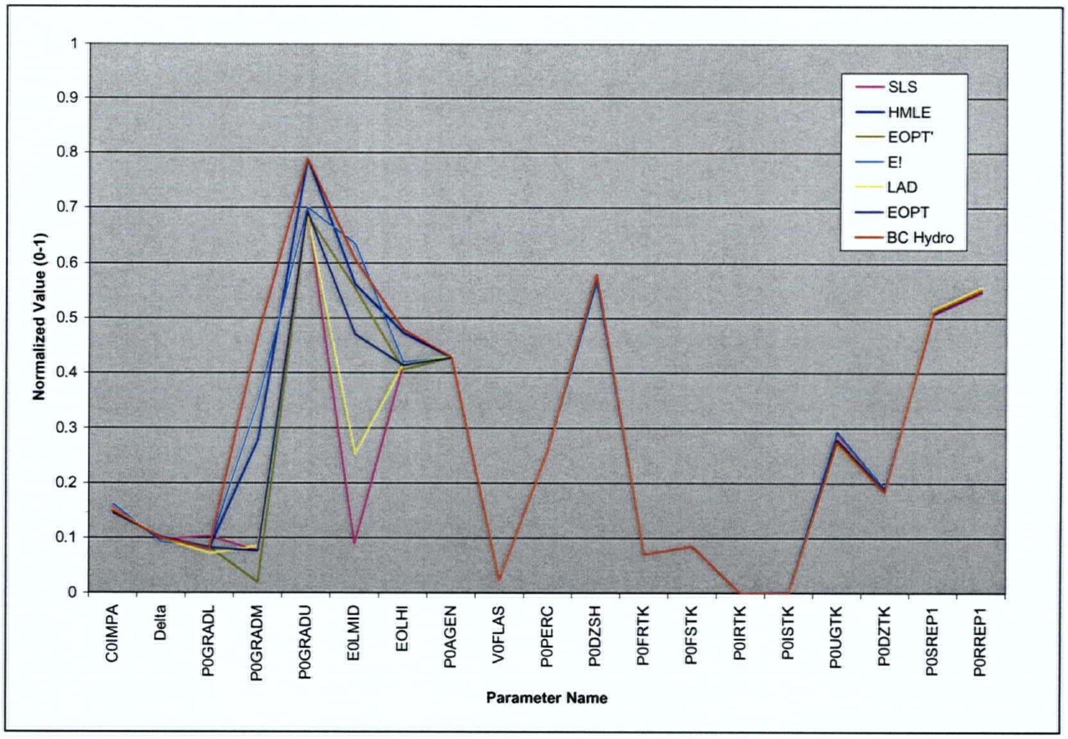
Figure 5-6: Typical Annual Hydrographs for Unsuccessful Synthetic Calibrations

Values of the final parameter sets for each calibration are presented graphically in Figures 5-7(a) and (b). For comparison, parameter values are normalized over their feasible range to the interval [0,1]. Comparing the figures, it is apparent that variability and error are more substantial for Coquitlam than for Illecillewaet. In both cases, variability and error are concentrated in the parameters controlling precipitation distribution (i.e., P0GRADL, P0GRADM, P0GRADU, E0LMID, and E0LHI). For comparison purposes, the "true" parameter sets (generated by the expert calibrations) are included for their respective watersheds. These are denoted as "BC Hydro" and "Expert" for the Coquitlam and Illecillewaet watersheds, respectively.

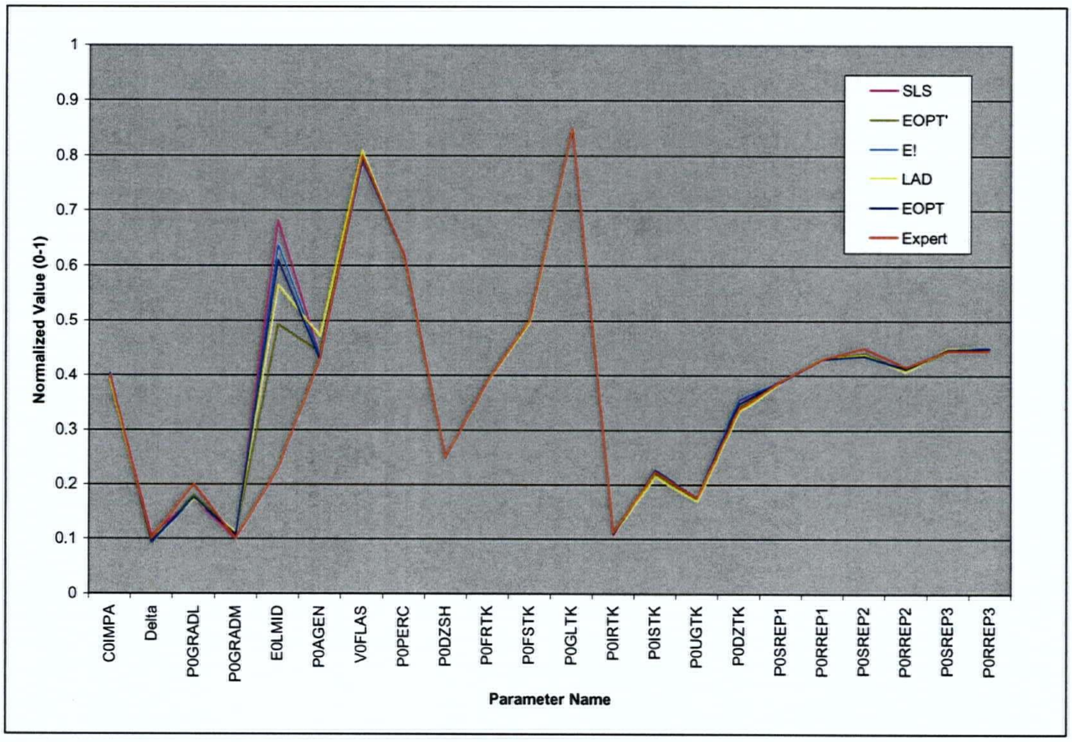
In light of this fact, it may seem surprising that the estimates for precipitation adjustment parameters P0SREP and P0RREP are both accurate and precise. The observed degree of accuracy and precision is believed to result from the strong role of P0SREP and P0RREP in closing the water balance of the watershed. This makes accurate estimation of P0SREP and P0RREP critical to achieving a successful synthetic calibration. This viewpoint is supported by consideration of the primary role these parameters assume in the sequential manual calibration typically used for UBCWM, as described in Section 4.1.1.

To determine whether additional data would yield a better synthetic calibration, a single calibration is performed for each watershed using all available data and the EOPT' objective function. In neither case do the additional data provide improvement. It can therefore be resolved that any lack of convergence is not driven by an insufficient duration of calibration data.

The SCE-UA method uses the random-number generation algorithm of Press et al. (1992). A random seed is required to initialize this algorithm. To explore the sensitivity of the SCE-UA method to this random seed, five synthetic calibrations are conducted for each watershed using the EOPT' objective function and unique random seeds. Despite the well-documented efficiency and effectiveness of the SCE-UA method, and notwithstanding its success in baseline convergence testing, the SCE-UA method demonstrates a consistent inability to accurately estimate the full complement of parameters for either watershed. No set of final parameter values exactly matches any other set. One must therefore conclude that, for this application, the calibration is sensitive to the value of the initial random seed. Figure 5-8 shows the normalized parameter sets obtained using different seed values.

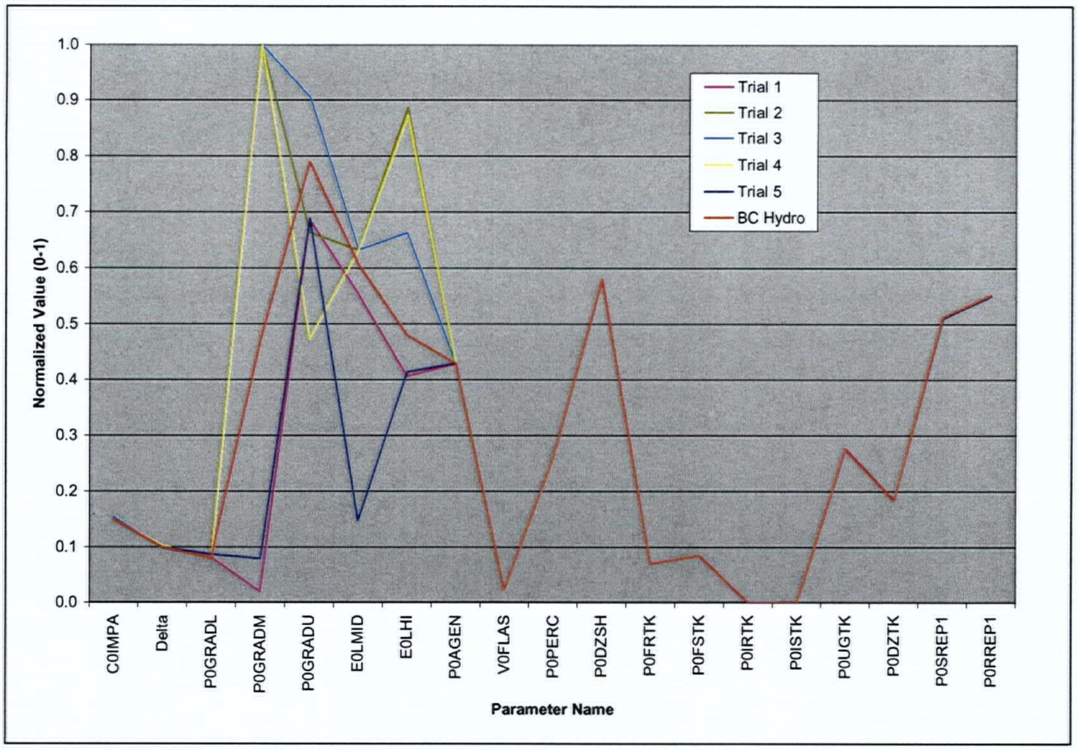


(a) Coquitlam Lake Watershed

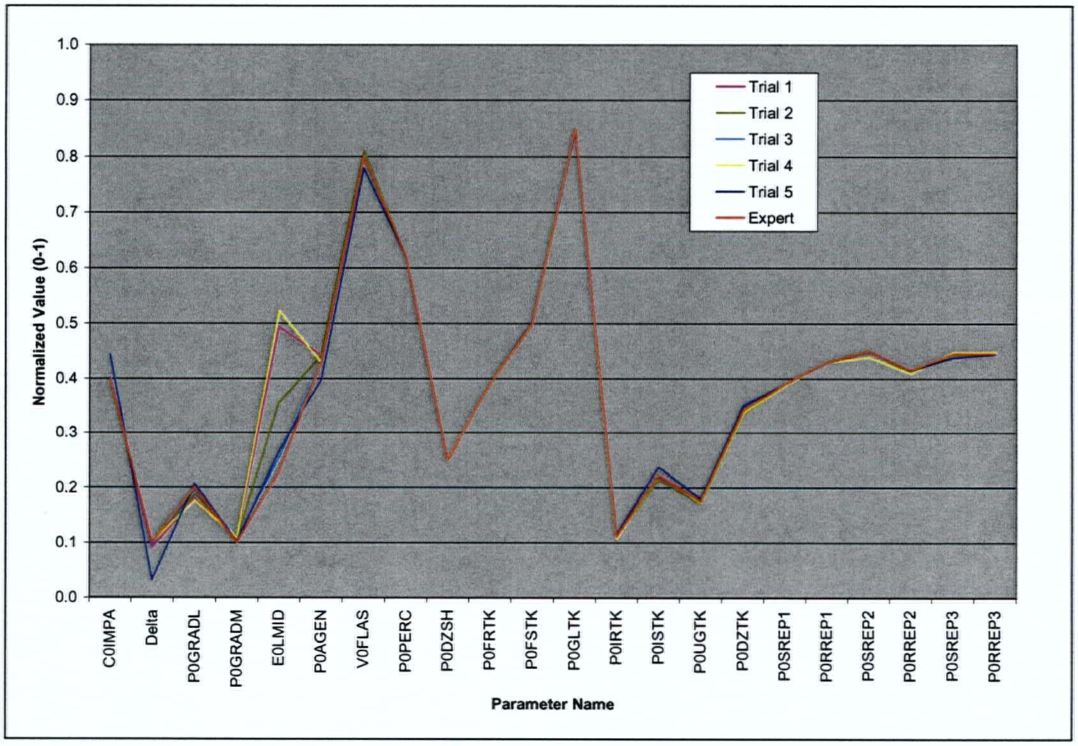


(b) Illecillewaet River Watershed

Figure 5-7: Normalized Parameter Values for Successful Synthetic Calibrations



(a) Coquitlam Lake Watershed, EOPT' Objective Function



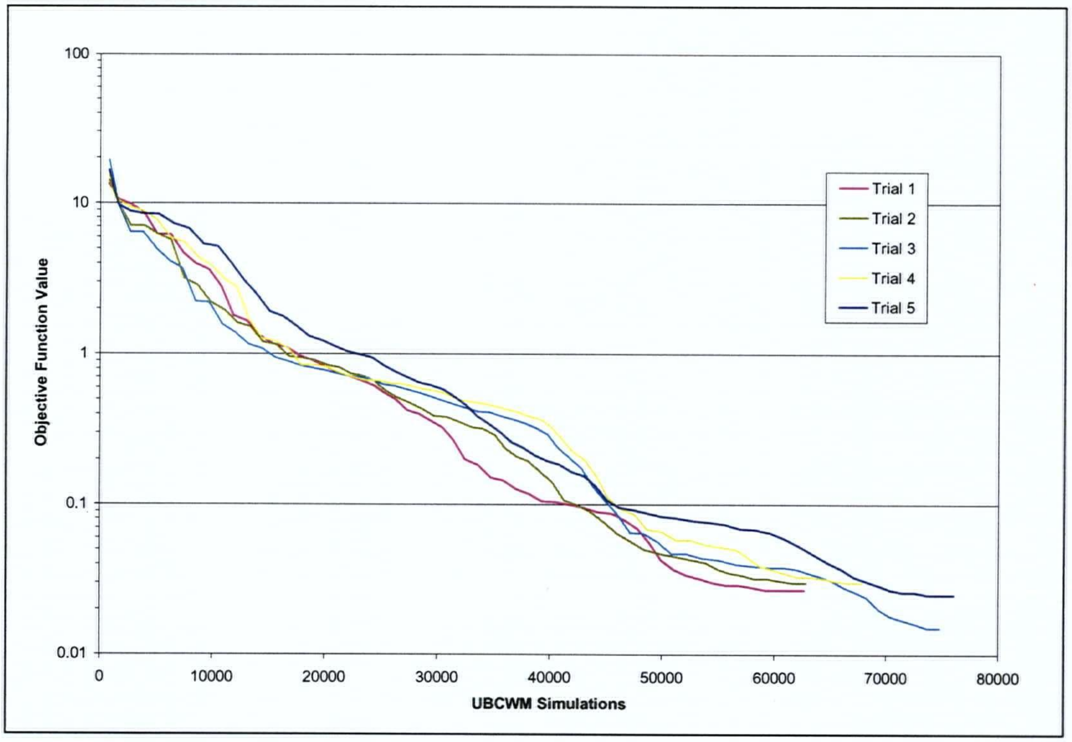
(b) Illecillewaet River Watershed, EOPT' Objective Function

Figure 5-8: Normalized Parameter Values for Synthetic Calibration Seed Sensitivity Trials

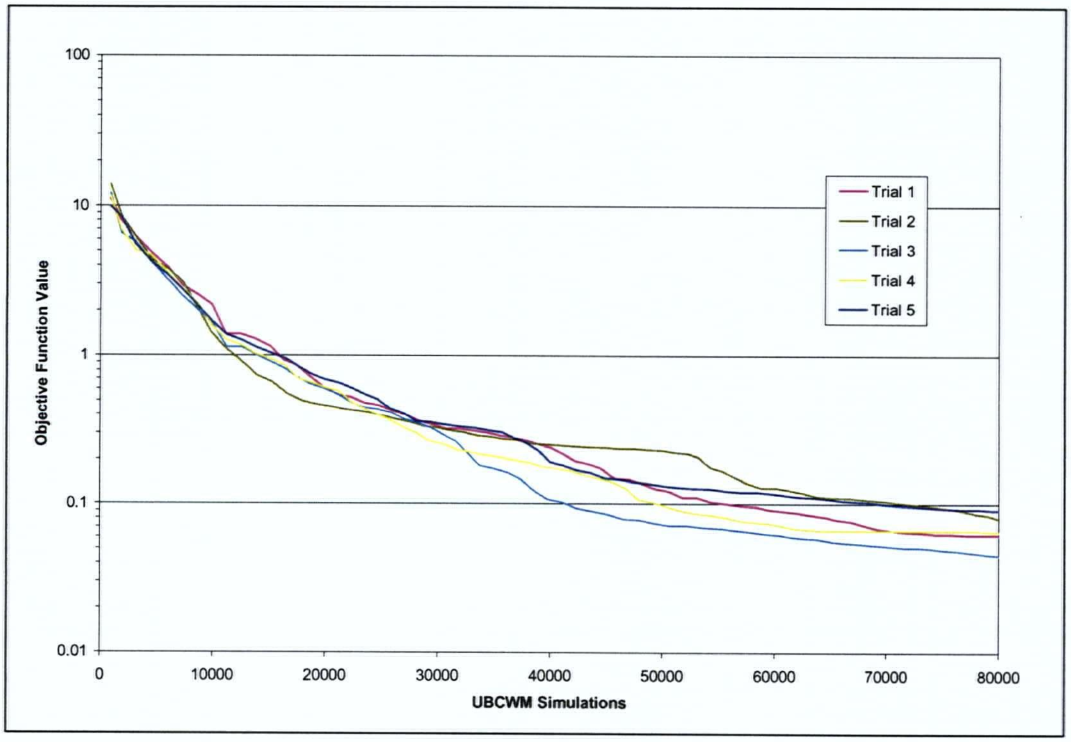
As Figure 5-8 shows, the variability of results for the Coquitlam Lake watershed is again greater than for the Illecillewaet River watershed. Equally clearly, variability is confined exclusively to the precipitation distribution parameters. Despite the obvious variability in parameter values, Figure 5-9 shows that the evolution of the objective function and its ultimate value are similar across the five trials conducted for each watershed. Any differences between the hydrograph traces are negligible, as shown in Figure 5-10. These findings suggest the possibility of over-parameterization in this aspect of the model. The lesser variability of the Illecillewaet River watershed (as compared to the Coquitlam Lake watershed) supports this hypothesis, since the Illecillewaet uses a simpler precipitation distribution with only two precipitation gradients.

The above discussion demonstrates that consistently good results (i.e., with respect to statistical and visual evaluation of output hydrographs) can be achieved with significant codependent variability in the precipitation distribution parameters. Therefore, the synthetic calibrations using objective functions SLS, HMLE (Coquitlam only), EOPT', E!, LAD, and EOPT can be accepted as "successful" notwithstanding the high residual errors in the estimates of precipitation distribution parameters. All of these successful synthetic calibrations terminate when the geometric mean of the normalized parameter ranges are reduced to less than 0.1%. The final parameter sets for the Coquitlam Lake and Illecillewaet River watersheds are given in Tables 5-2 and 5-3, respectively. As before, the "true" solutions for the synthetic calibrations are provided. These are labelled as "BC Hydro" and "Expert" calibrations for the Coquitlam and Illecillewaet watersheds, respectively.

Red-highlighted parameters in Tables 5-2 and 5-3 are rendered insignificant for that particular calibration by values of other parameters. The red-highlighted parameters could assume any value within their defined range without affecting the results. For example, the mid-elevation precipitation gradient P0GRADM for the E! calibration of Table 5-1 plays no role in the UBCWM simulation of the Coquitlam Lake watershed because E0LMID (the elevation breakpoint between the lower and middle precipitation gradients) is essentially equal to E0LHI (the elevation breakpoint between the middle and upper precipitation gradients). The range of application for the middle precipitation gradient P0GRADM is negligible. Red-highlighted parameters are not considered in the calculation of means and standard deviations across the set of calibrations.

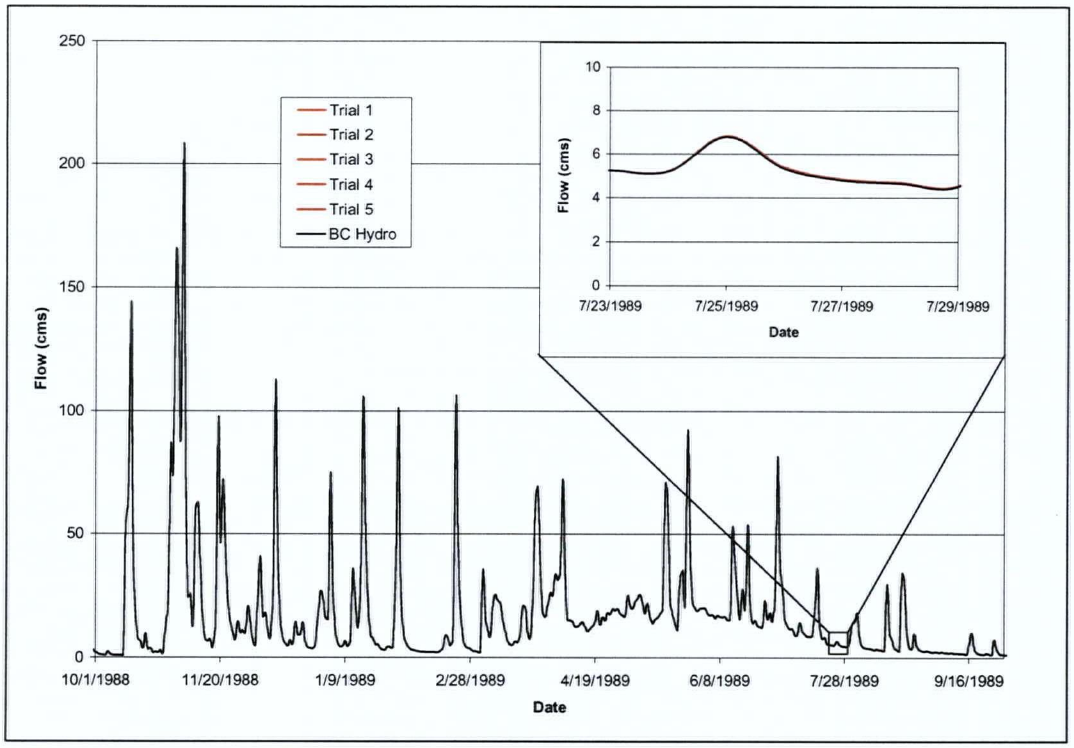


(a) Coquitlam Lake Watershed, Objective Function EOPT'

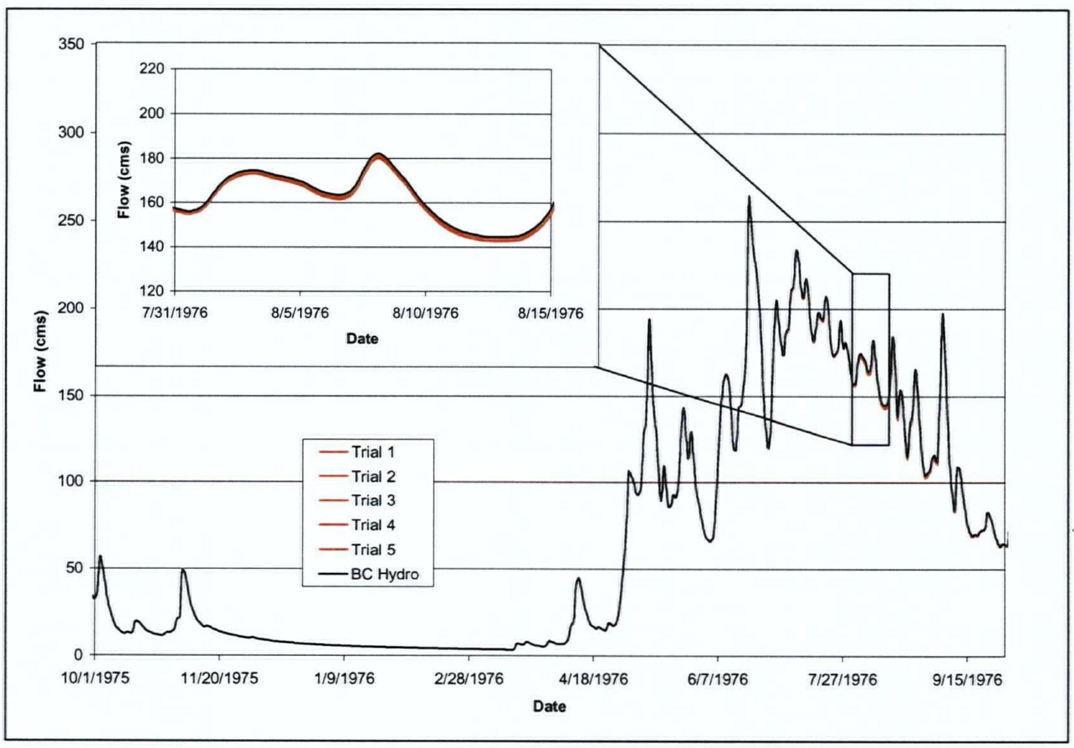


(b) Illecillewaet River Watershed, Objective Function EOPT'

Figure 5-9: Objective Function Evolution for Synthetic Calibration Seed Sensitivity Trials



(a) Coquitlam Lake Watershed



(b) Illecillewaet River Watershed

Figure 5-10: Example Annual Hydrographs for Synthetic Calibration Seed Sensitivity Trials

Table 5-2: Parameter Values for Successful Synthetic Calibrations - Coquitlam Lake Watershed

Objective	C0IMPA	Delta	P0GRADL	P0GRADM	P0GRADU	E0LMID	E0LHI	P0AGEN	V0FLAS	P0PERC	P0DZSH
SLS	0.16	0.049	2.1	1.5	21.0	542	1415	100.1	1.0	13.0	0.57
HMLE	0.15	0.050	1.6	5.6	23.7	1298	1473	99.9	1.0	13.0	0.58
EOPT'	0.15	0.050	1.6	0.4	20.6	1287	1407	99.9	1.0	13.0	0.58
E!	0.16	0.046	1.7	7.0	21.0	1415	1420	100.2	1.0	13.0	0.57
LAD	0.15	0.050	1.4	1.7	20.6	806	1416	100.4	1.0	13.0	0.58
EOPT	0.15	0.051	1.6	1.5	20.8	1153	1415	99.9	1.0	13.0	0.58
Std Dev	0.01	0.00	0.2	2.0	1.3	338	24	0.2	0.0	0.0	0.01
Mean	0.15	0.05	1.7	2.2	21.3	1083	1424	100.1	1.0	13.0	0.58
BC Hydro	0.15	0.050	1.6	9.4	23.7	1370	1480	100.0	1.0	13.0	0.58

Objective	P0FRTK	P0FSTK	P0IRTK	P0ISTK	P0UGTK	P0DZTK	P0SREP	P0RREP
SLS	0.14	0.17	1.00	1.01	21.68	97.70	0.014	0.094
HMLE	0.14	0.17	1.00	1.01	21.00	96.00	0.025	0.105
EOPT'	0.14	0.17	1.00	1.00	21.08	96.10	0.022	0.103
E!	0.14	0.17	1.00	1.02	21.80	97.40	0.018	0.100
LAD	0.14	0.17	1.00	1.00	20.87	96.00	0.032	0.112
EOPT	0.14	0.17	1.00	1.01	21.17	96.30	0.022	0.103
Std Dev	0.00	0.00	0.00	0.01	0.38	0.76	0.006	0.006
Mean	0.14	0.17	1.00	1.01	21.27	96.58	0.022	0.103
BC Hydro	0.14	0.17	1.20	1.00	21.00	96.00	0.025	0.105

Table 5-3: Parameter Values for Successful Synthetic Calibrations - Illecillewaet River Watershed

Objective	C0IMPA	Delta	P0GRADL	P0GRADM	E0LMID	P0AGEN	V0FLAS	P0PERC	P0DZSH	P0FRTK	P0FSTK
SLS	0.40	0.05	3.4	2.0	1828	101.1	35.8	31.09	0.25	0.78	1.00
EOPT'	0.40	0.05	3.6	2.1	1488	102.5	36.1	30.90	0.25	0.78	1.00
E!	0.41	0.05	3.4	2.1	1747	101.1	35.8	31.02	0.25	0.78	1.00
LAD	0.39	0.05	3.5	2.2	1616	107.6	36.2	30.86	0.25	0.78	0.99
EOPT	0.40	0.05	3.5	2.1	1699	99.9	35.9	30.99	0.25	0.78	1.00
Std Dev	0.01	0.00	0.1	0.1	130	3.0	0.2	0.09	0.00	0.00	0.00
Mean	0.40	0.05	3.5	2.1	1676	102.4	36.0	30.97	0.25	0.78	1.00
Manual	0.40	0.05	4.0	2.0	1009	100.0	36.0	31.00	0.25	0.78	1.00

Objective	P0GLTK	P0IRTK	P0ISTK	P0UGTK	P0DZTK	P0SREP1	P0RREP1	P0SREP2	P0RREP2	P0SREP3	P0RREP3
SLS	1.70	1.96	3.04	17.00	171.27	-0.220	-0.140	-0.129	-0.178	-0.102	-0.099
EOPT'	1.70	1.98	2.97	16.95	168.62	-0.223	-0.139	-0.119	-0.180	-0.110	-0.105
E!	1.70	1.98	3.04	17.04	171.68	-0.223	-0.141	-0.129	-0.175	-0.102	-0.099
LAD	1.70	1.98	2.93	16.76	166.89	-0.224	-0.139	-0.123	-0.185	-0.100	-0.103
EOPT	1.70	1.98	3.03	17.00	170.03	-0.223	-0.140	-0.130	-0.177	-0.106	-0.102
Std Dev	0.00	0.01	0.05	0.11	1.97	0.002	0.001	0.005	0.004	0.004	0.003
Mean	1.70	1.98	3.00	16.95	169.70	-0.223	-0.140	-0.126	-0.179	-0.104	-0.102
Manual	1.70	2.00	3.00	17.00	168.00	-0.220	-0.140	-0.100	-0.170	-0.110	-0.110

Parameters highlighted in yellow have assumed values at or within 1% of their boundary value. Boundary-value parameters are believed to be indicative of compensating errors in the calibration and should thus be viewed with caution (Franchini et al., 1998).

In summary, the objective functions SLS, HMLE (Coquitlam Lake watershed only), EOPT', E!, LAD, and EOPT produce successful synthetic calibrations and are considered for further testing. The objective functions HMLE! (Illecillewaet River watershed only), D!, and EOPT! failed to converge to an acceptable estimate of the synthetic parameter set and are therefore not considered further.

5.3 Calibrations against Observed Data

5.3.1 Coquitlam Lake Watershed

SCE-UA calibrations of the Coquitlam Lake watershed are performed with the objective functions SLS, HMLE, EOPT', E!, LAD, and EOPT. All but one of the objective functions terminate when SCE-UA is unable to improve the objective function value of the best member of the population over ten shuffling loops. The lone exception is the HMLE objective function, which terminates when the number of UBCWM simulations reaches the specified maximum.

For convenience, statistical analysis is limited to those statistics produced by the UBCWM, including Nash-Sutcliffe Efficiency (E!), Coefficient of Determination (D! or R^2), modified Nash-Sutcliffe Efficiency EOPT!, and the volume error dV/V, defined as:

$$dV/V = |\, 1 - (\Sigma q_{sim} / \Sigma q_{obs})\, | \qquad (12)$$

Summary statistics calculated for the period of record (with the first year discarded for spin-up) are presented in Table 5-4. Complete statistics, including both annual values and monthly averages, are included in Appendix B. Table 5-4 shows that all calibrations are statistically comparable to the BC Hydro expert calibration, with the exception of HMLE. After testing to ensure that the results are not an artifact of the initial random seed, the poor performance of HMLE is deemed sufficient to justify eliminating it from further analyses.

Table 5-4: Summary Statistics for Multi-Objective Calibrations of Coquitlam Lake Watershed
(Water Years 1986-87 through 1998-99)

Objective	E!	D! (R^2)	EOPT!	dV/V
SLS	0.77	0.77	0.76	0.02
HMLE	0.57	0.75	0.16	0.42
EOPT'	0.77	0.77	0.69	0.08
E!	0.77	0.77	0.76	0.02
LAD	0.76	0.77	0.64	0.12
EOPT	0.77	0.77	0.75	0.02
BC Hydro	0.75	0.76	0.71	0.04

A typical annual hydrograph for all objective functions is shown in Figure 5-11. As for the statistical results, it is difficult to conclusively prove that any one calibration is better than the others. It is interesting to note the differences between the hydrographs produced from SCE-UA and BC Hydro expert calibrations. In cases where both automatic and expert calibrations fail to reproduce a peak in the observed outflow series, the automatic calibrations attempt to shift the flat simulated hydrograph upwards or downwards to achieve better overall statistical performance; in contrast, the expert calibration typically ignores those data to better represent others. In general, the "better" statistical performance of automatic calibration – in comparison with manual calibration – may be artificial and should not necessarily be equated with a "better" simulation.

Results for the SCE-UA calibrations of the Coquitlam Lake watershed include the simulated hydrographs and final parameter set achieved with each objective function. A plot of normalized parameter values for the automatic calibrations is presented in Figure 5-12. There is a fair degree of similarity across the various objective functions for many of the parameters, with the exception of the parameters governing precipitation distribution. The observed variability for these parameters is likely an artifact of the over-parameterization of the precipitation distribution function.

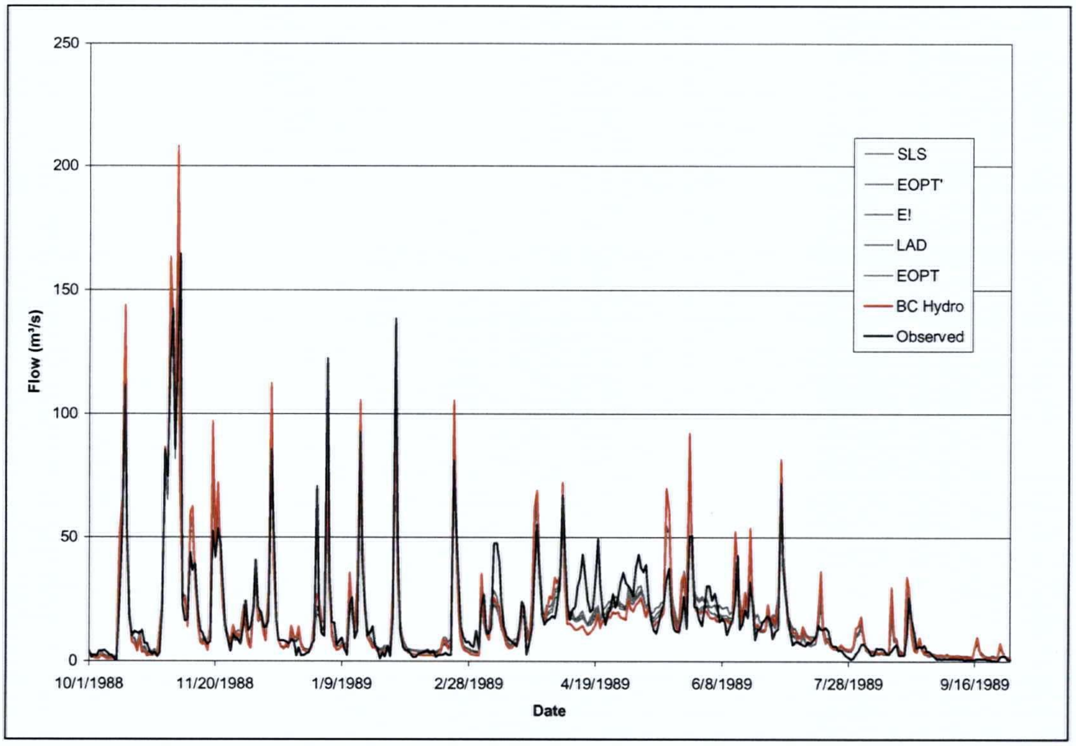


Figure 5-11: Typical Annual Hydrograph for Multi-Objective Calibration of Coquitlam Lake Watershed

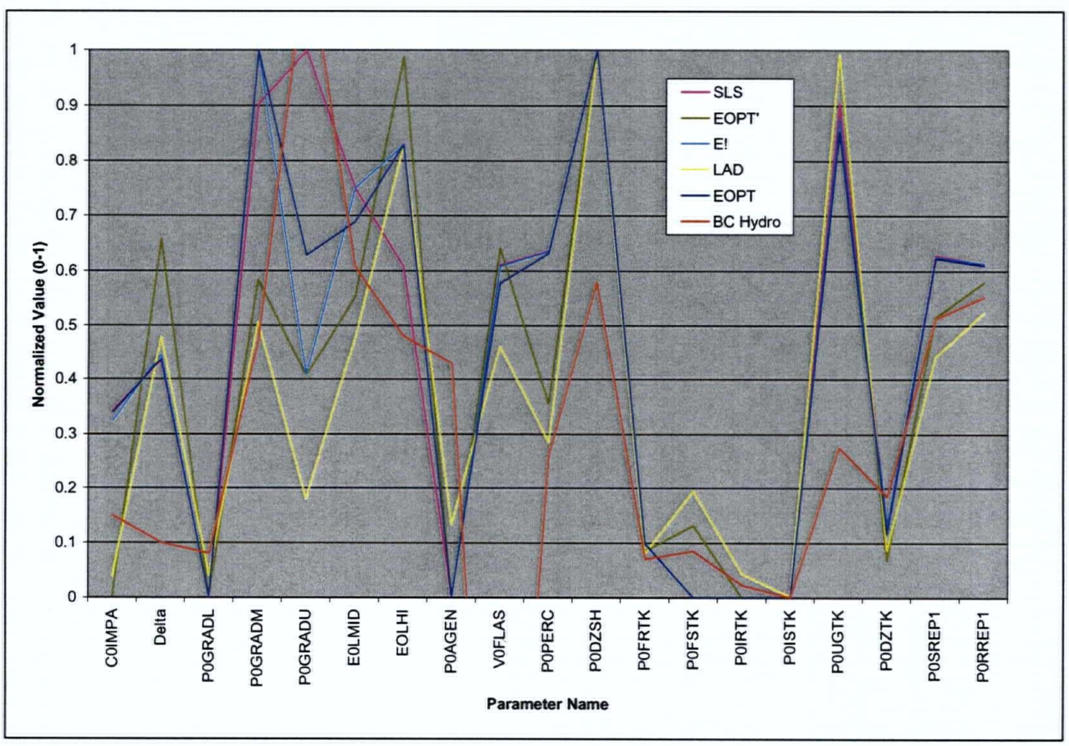


Figure 5-12: Normalized Parameter Values for Multi-Objective Calibration of Coquitlam Lake Watershed

Parameter values for BC Hydro's expert calibration are shown in Figure 5-12 for reference only. Since this experiment focuses on variability, it is not concerned with how well or how poorly the automatic calibrations reflect the BC Hydro values. Note that the calibrated BC Hydro values for parameters P0GRADU and V0FLAS are outside the ranges provided in Table 4-2 and thus also lie outside the normalized range [0,1].

Parameters C0IMPA, P0PERC, and P0FSTK all exhibit a tendency toward bi-modality, having values associated with SLS, EOPT, and E! in one cluster and EOPT' and LAD in a second. This observation is consistent with the bi-modality of the dV/V statistic presented in Table 5-4; the EOPT' and LAD calibrations result in significantly higher volume error over the period of record than the other objective functions.

Table 5-5 presents the actual parameter values obtained for the five multi-objective calibrations. Parameter P0UGTK (upper groundwater time constant) is effectively an inactive parameter for all calibrations because the value obtained for P0DZSH (1.00) results in 100% of groundwater being directed to the deep zone. Parameter P0GRADU is also not utilized in the final parameter set for all calibrations except the one using the SLS objective function. In each case, the breakpoint between middle and upper precipitation gradients is set above the mid-point of the highest elevation band in the model.

In addition to these two inactive parameters, a large number of parameters assume values at or near their boundary conditions. Several of the parameter values identified in the BC Hydro expert calibration lie outside the range of recommended values provided in the UBCWM User's Manual (Quick, 1994). Therefore, these values also lie outside the feasible parameter space used for the SCE-UA automatic calibrations. Specifically, the parameters in question include P0GRADU (calibrated value 23.7, recommended range 0-20) and V0FLAS (calibrated value 1.0, recommended value 20-40). Section 5.3.2 addresses subsequent calibrations that use an expanded parameter space for all parameters.

5.3.2 Calibrations with Expanded Parameter Space

The BC Hydro calibration report for the Coquitlam Lake watershed (BC Hydro, 2004) indicates that calibration is made difficult by the smaller size and flashy response of the catchment. The boundary values and inactive parameters of Table 5-5 suggest that the difficulty experienced by

Table 5-5: Parameter Values for Multi-Objective Calibration of Coquitlam Lake Watershed

Objective	C0IMPA	Delta	P0GRADL	P0GRADM	P0GRADU	E0LMID	E0LHI	P0AGEN	V0FLAS	P0PERC	P0DZSH
SLS	0.34	0.217	0.0	18.1	20.0	1599	1607	25.0	32.2	31.8	1.00
EOPT'	0.00	0.329	0.0	11.7	8.1	1284	1987	25.0	32.8	17.7	1.00
E!	0.32	0.223	0.0	20.0	8.1	1600	1828	25.0	32.2	31.7	1.00
LAD	0.04	0.239	0.8	10.1	3.6	1160	1829	48.3	29.2	14.2	1.00
EOPT	0.34	0.218	0.0	20.0	12.6	1502	1829	25.0	31.5	31.6	1.00
Std Dev	0.17	0.05	0.4	4.7	6.2	198	135	10.4	1.4	8.7	0.00
Mean	0.21	0.25	0.2	16.0	10.5	1429	1816	29.7	31.6	25.4	1.00
BC Hydro	0.15	0.050	1.6	9.4	23.7	1370	1480	100.0	1.0	13.0	0.58

Objective	P0FRTK	P0FSTK	P0IRTK	P0ISTK	P0UGTK	P0DZTK	P0SREP	P0RREP
SLS	0.20	0.00	1.00	1.00	46.17	82.96	0.256	0.224
EOPT'	0.17	0.26	1.00	1.00	49.52	67.03	0.030	0.159
E!	0.20	0.00	1.00	1.00	45.05	82.82	0.251	0.226
LAD	0.16	0.39	1.40	1.01	49.66	71.64	-0.116	0.048
EOPT	0.20	0.00	1.00	1.00	44.10	79.00	0.244	0.218
Std Dev	0.02	0.18	0.18	0.00	2.56	7.09	0.169	0.076
Mean	0.18	0.13	1.08	1.00	46.90	76.69	0.133	0.175
BC Hydro	0.14	0.17	1.20	1.00	21.00	96.00	0.025	0.105

BC Hydro could potentially arise because the "optimal" calibration lies beyond the limits of the feasible parameter space. This possibility is investigated in this study by calibrating the UBCWM using the expanded parameter space outlined in Table 4-2.

All calibrations terminate when SCE-UA is unable to improve the objective function value of the best member of the population over ten shuffling loops. Statistical results (E!, D!, EOPT!, and dV/V) achieved with the expanded parameter space are shown in Table 5-6. As in the preceding section, the statistics are calculated for the full period of record excluding the first year for spin-up. Also as before, the HMLE objective function fails to provide a sufficiently good simulation to justify its inclusion in further analyses. Statistical results for the remaining objective functions range from unchanged to marginally better than those achieved with the smaller parameter space (Table 5-4). The best objective function values for all expanded-space calibrations differ from their more constrained counterparts by less than 1%. Visual examination of the hydrographs offers no new insights. Table 5-7 presents the final parameter sets generated for each objective function.

Table 5-6: Summary Statistics for Multi-Objective Calibrations of Coquitlam Lake Watershed using Expanded Parameter Space
(Water Years 1986-87 through 1998-99)

Objective	E!	D!	EOPT!	dV/V
SLS	0.77	0.77	0.76	-0.01
HMLE	0.57	0.75	0.15	0.42
EOPT'	0.77	0.77	0.69	0.08
E!	0.77	0.77	0.76	0.01
LAD	0.76	0.77	0.64	0.12
EOPT	0.77	0.77	0.77	0.01
BC Hydro	0.75	0.76	0.71	0.04

Table 5-7: Parameter Values for Multi-Objective Calibrations of Coquitlam Lake Watershed using Expanded Parameter Space

Objective	C0IMPA	Delta	P0GRADL	P0GRADM	P0GRADU	E0LMID	E0LHI	P0AGEN	V0FLAS	P0PERC	P0DZSH
SLS	0.82	0.085	0.0	0.0	28.0	612	1586	4.9	36.0	31.7	0.94
EOPT'	0.40	0.198	0.0	10.4	8.3	1208	1730	18.4	29.6	29.0	0.64
E!	0.71	0.092	0.0	22.9	4.8	1526	1924	7.1	30.1	32.3	0.00
LAD	0.00	0.328	0.0	10.0	13.0	1143	1830	26.0	30.6	15.5	0.60
EOPT	0.98	0.007	0.0	28.5	1.8	1557	1828	4.5	36.2	31.7	1.00
Std Dev	0.39	0.12	0.0	11.3	10.3	381	128	9.6	3.3	7.1	0.40
Mean	0.58	0.14	0.0	14.4	11.2	1209	1780	12.2	32.5	28.1	0.64
BC Hydro	0.15	0.050	1.6	9.4	23.7	1370	1480	100.0	1.0	13.0	0.58

Objective	P0FRTK	P0FSTK	P0IRTK	P0ISTK	P0UGTK	P0DZTK	P0SREP	P0RREP
SLS	0.24	0.00	0.10	0.10	81.88	97.07	0.216	0.236
EOPT'	0.22	0.26	0.10	0.10	65.16	65.80	-0.017	0.152
E!	0.24	0.00	0.10	0.10	110.36	56.23	0.246	0.244
LAD	0.17	0.39	1.53	0.10	73.90	73.46	-0.049	0.090
EOPT	0.24	0.00	0.10	0.10	59.50	102.05	0.220	0.236
Std Dev	0.03	0.18	0.64	0.00	19.59	17.67	0.144	0.068
Mean	0.22	0.13	0.39	0.10	82.83	84.59	0.123	0.192
BC Hydro	0.14	0.17	1.20	1.00	21.00	96.00	0.025	0.105

The normalized parameter values shown in Figure 5-13 demonstrate a greater degree of variability than is observed in the corresponding figure for the constrained parameter space (Figure 5-12). A similar trend is reflected in comparing the standard deviations of Table 5-7 with those of Table 5-5. In general, there seems to be less consistency in the results achieved using the expanded parameter space. For example, in this case C0IMPA ranges in value from zero (LAD) to 0.98 (EOPT). As before, parameter values for BC Hydro's expert calibration are shown for reference only.

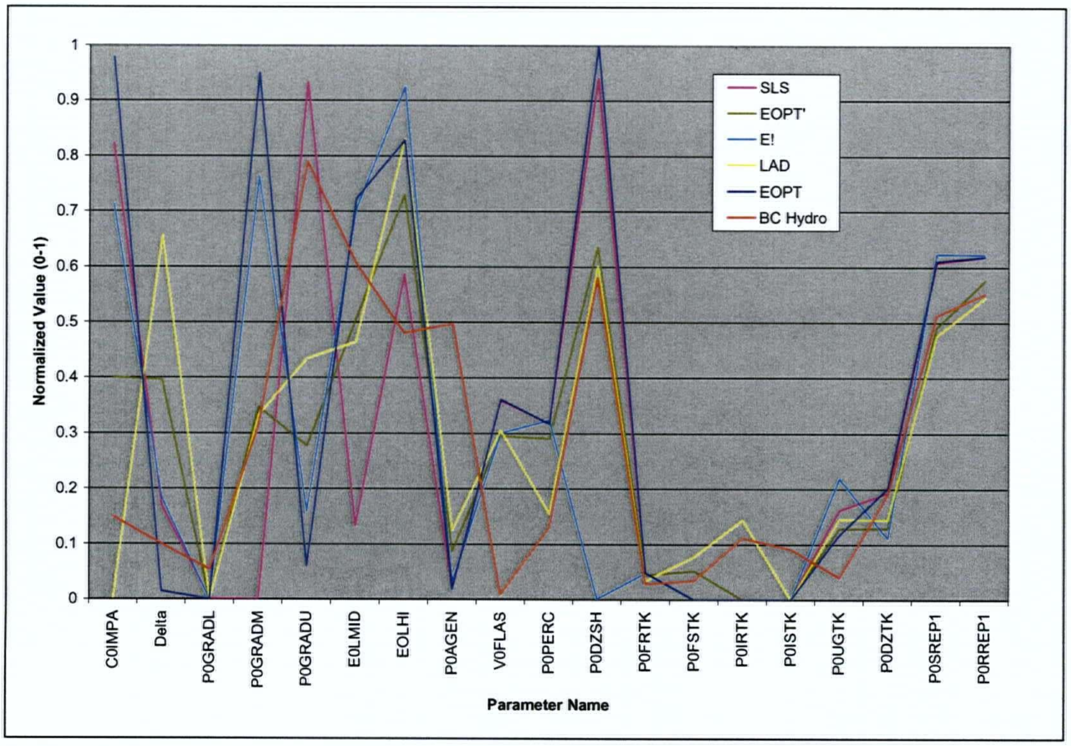


Figure 5-13: Normalized Parameter Values for Multi-Objective Calibrations of Coquitlam Lake Watershed using Expanded Parameter Space

Table 5-7 shows that expanding the parameter space reduces the number of irrelevant and boundary-value parameters but does not eliminate them. In particular, objective functions yielding boundary-value parameters in the preceding section simply tend to adjust to the new boundary condition when using an expanded parameter space. The objective functions E!, LAD, and EOPT once again render the upper precipitation gradient P0GRADU inactive by allowing the elevation breakpoint E0LHI to exceed the mid-point elevation of the highest elevation band.

E! and EOPT also render irrelevant the values of the lower (P0DZTK) and upper (P0UGTK) groundwater time constants, respectively, due to the corresponding values of 0.0 and 1.0 for P0DZSH which controls partitioning between the two zones.

A qualitative analysis of the parameter sets shown in Table 5-7 reveals that relative parameter realism is lacking in some cases. In particular, all objective functions except LAD yield "interflow" time constants for rainfall (P0IRTK) that are smaller (i.e., faster) than those corresponding to the ostensibly faster response of "surface" runoff (P0FRTK). All but one instance of the interflow time constants P0IRTK and P0ISTK assume values at the boundary condition of 0.1 days, a somewhat unrealistic value. Although UBCWM is a conceptual model and is therefore not strictly bound by a need for physical representativeness in parameter values, the parameters are intended to have a at least a relative degree of physical significance (e.g., P0FRTK < P0IRTK < P0UGTK < P0DZTK). While this relational significance is not explicitly required by the model structure, it should not be necessary to disregard physical or relational significance to attain a good calibration. This is evidenced by the success of the BC Hydro calibration for the Coquitlam Lake watershed. In contrast to some of the SCE-UA-generated parameter sets shown in Table 5-7, the BC Hydro calibration contains no parameter values that can be conclusively defined as physically unreasonable.

Of the two parameters in the BC Hydro expert calibration that are perturbed beyond their recommended ranges (P0GRADU and V0FLAS), Table 5-7 shows that in only one instance do either of these parameters achieve a value outside the initial parameter space. Parameter P0GRADU exceeds the recommended limit of 20% when calibrated using the SLS objective function. In this one case, values for the three precipitation gradients P0GRADL, P0GRADM, and P0GRADU are 0%, 0%, and 28% respectively. This represents a highly implausible physical condition, as it implies that the orographic precipitation gradient undergoes a step-change from negligible to extreme in the higher reaches of the watershed.

The similarity in quantitative and qualitative evaluations of the results for initial and expanded parameter spaces, combined with the questionable physical realism of some expanded-space parameter sets, implies that there is no basis for rejecting the results of Section 5.3.1 on the grounds of being too strongly constrained by the feasible region during calibration. Therefore, the event analyses are pursued utilizing the parameter sets contained in Table 5-5, and those of Table 5-7 are not considered further.

5.3.3 Illecillewaet River Watershed

Calibrations for the Illecillewaet River watershed use objective functions SLS, EOPT', E!, LAD, and EOPT. Each calibration terminates when SCE-UA is unable to improve the objective function value of the best member of the population within ten shuffling loops.

Summary statistics E!, D!, EOPT! and dV/V for the period of record excluding the first year of simulation (i.e., for water years 1971-72 through 1989-90) are presented in Table 5-8. It is apparent that the overall simulation quality is better than that achieved for the Coquitlam Lake watershed, although none of the Illecillewaet River calibrations are comparable in quality to the expert calibration in the UBCWM example file. In particular, significant volume errors in four of the five SCE-UA calibrations would likely prove problematic if the SCE-UA parameter sets were to be used in practice.

Table 5-8: Summary Statistics for Multi-Objective Calibrations of Illecillewaet River Watershed
(Water Years 1971-72 through 1989-90)

Objective	E!	D! (R^2)	EOPT!	dV/V
SLS	0.83	0.90	0.60	-0.24
EOPT'	0.85	0.88	0.68	-0.17
E!	0.87	0.87	0.80	-0.07
LAD	0.88	0.90	0.74	-0.14
EOPT	0.83	0.89	0.58	-0.25
Manual	0.90	0.93	0.82	-0.08

Significant results from the calibrations include the simulated hydrographs and final parameter set achieved with each objective function. The typical annual hydrograph shown in Figure 5-14 confirms that the simulation quality achieved with the SCE-UA method is inferior to that attained by the UBCWM example calibration. Again, data uncertainty likely plays a role, although there is no way to determine the extent of its impact.

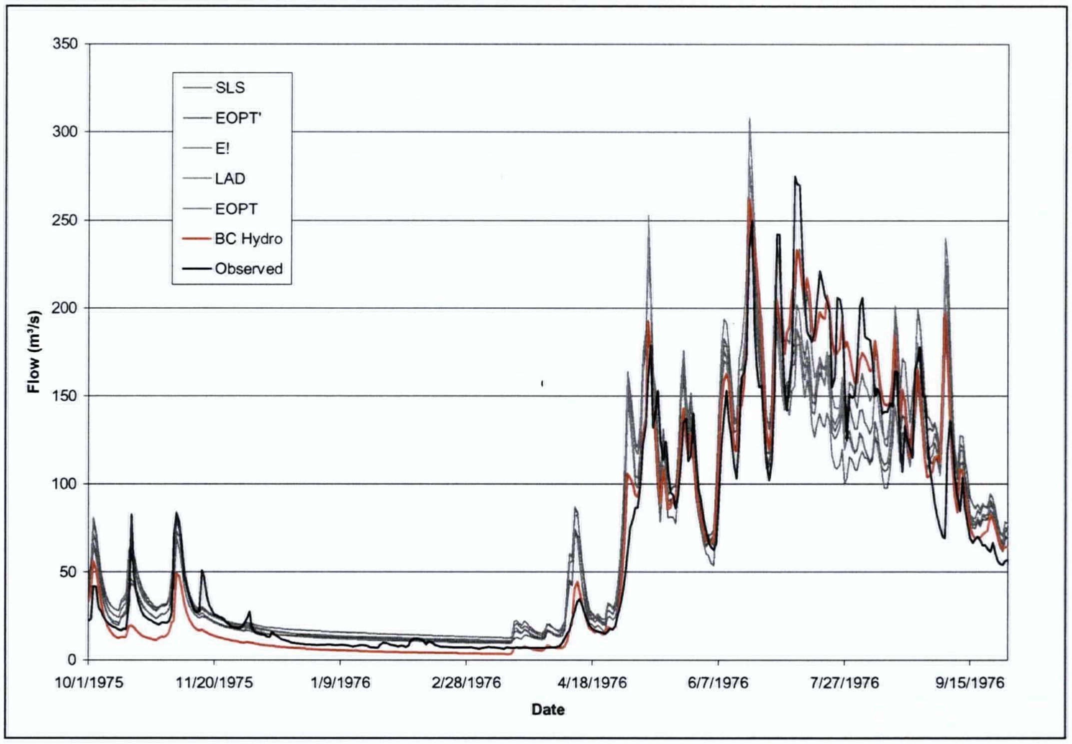


Figure 5-14: Typical Annual Hydrograph for Multi-Objective Calibration of Illecillewaet River Watershed

Figure 5-15 is a plot of normalized parameter values for the SCE-UA calibrations. Figure 5-15 shows that there is markedly more parameter variability than in either of the corresponding figures for the Coquitlam Lake watershed (Figures 5-12 and 5-14). In some cases, parameter values span the entire feasible range (e.g., P0FRTK). As for the Coquitlam Lake watershed calibrations, the parameter set from the expert calibration is included for reference.

Table 5-9 presents the final parameter sets for the SCE-UA calibrations. It is interesting to note that although Table 5-9 contains only two irrelevant parameters (P0GRADL and P0UGTK for objective function E!), the number of parameters assuming boundary-condition values is proportionately greater than those of Tables 5-5 or 5-7 for the Coquitlam Lake watershed. Since the example calibration does not contain any boundary-value parameters, their presence suggests that there are likely compensating errors among the parameter values.

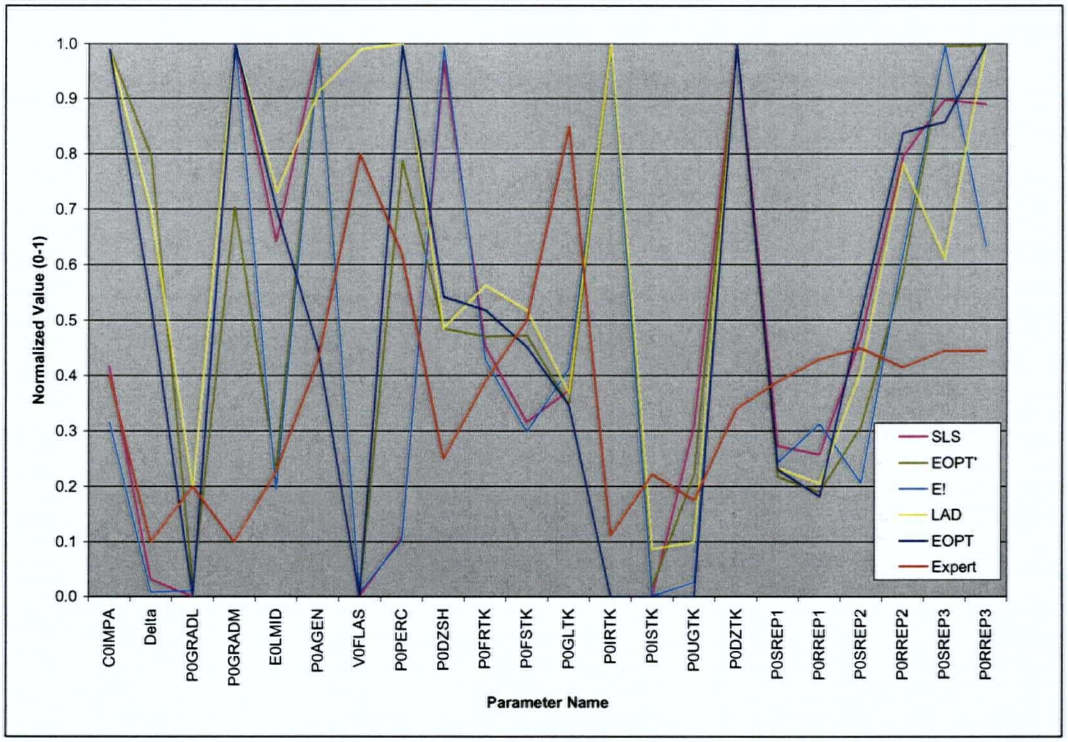


Figure 5-15: Normalized Parameter Values for Multi-Objective Calibrations of Illecillewaet River Watershed

In this aspect, the performance of SCE-UA is unexpectedly poor. Given the success of the synthetic calibrations in Section 5.2, the poor simulation quality cannot be directly attributed to the complexity of the Illecillewaet model parameter space. The precipitation distribution is in fact less complex than that of the Coquitlam model, leaving only the derivation of precipitation data from multiple meteorological stations as a complicating factor. In using data from three meteorological stations, the number of precipitation adjustment parameters (i.e., P0SREP and P0RREP) increases from two to six. Incorrect estimates of P0SREP and P0RREP will result in significant volume errors as compensating errors act to preserve a good overall response in the face of unrealistic adjustments of measured precipitation. This hypothesis is supported by the summary statistics in Table 5-8.

The values of P0SREP and P0RREP shown in Table 5-9 also support the above explanation; in comparison with the example calibration's relatively conservative range of -10% to -22%, the values of P0SREP and P0RREP determined using SCE-UA range from a highly questionable -64% (EOPT, P0RREP1) to a completely unrealistic +100% (EOPT, P0RREP3). Mean values

Table 5-9: Parameter Values for Multi-Objective Calibrations of Illecillewaet River Watershed

Objective	C0IMPA	Delta	P0GRADL	P0GRADM	E0LMID	P0AGEN	V0FLAS	P0PERC	P0DZSH	P0FRTK	P0FSTK
SLS	0.42	0.02	0.0	20.0	1755	199.9	20.0	5.48	0.97	0.91	0.63
EOPT'	0.99	0.40	0.4	14.1	1010	199.4	20.0	39.39	0.49	0.94	0.95
E!	0.32	0.00	0.2	20.0	950	196.8	20.1	5.20	1.00	0.86	0.60
LAD	0.99	0.35	3.9	20.0	1915	184.9	39.8	49.99	0.49	1.13	1.04
EOPT	0.99	0.27	0.0	20.0	1857	103.3	20.0	49.80	0.54	1.04	0.90
Std Dev	0.34	0.19	1.9	2.6	476	41.6	8.8	22.89	0.26	0.11	0.20
Mean	0.74	0.21	1.1	18.8	1497	176.8	24.0	29.97	0.70	0.97	0.82
Manual	0.40	0.05	4.0	2.0	1009	100.0	36.0	31.00	0.25	0.78	1.00

Objective	P0GLTK	P0IRTK	P0ISTK	P0UGTK	P0DZTK	P0SREP1	P0RREP1	P0SREP2	P0RREP2	P0SREP3	P0RREP3
SLS	0.75	10.00	1.00	22.37	299.97	-0.455	-0.486	-0.061	0.590	0.797	0.780
EOPT'	0.69	9.91	1.15	18.93	299.78	-0.566	-0.620	-0.386	0.161	0.992	0.995
E!	0.83	9.91	1.01	11.01	299.18	-0.513	-0.374	-0.588	0.234	0.993	0.264
LAD	0.74	10.00	1.78	13.95	299.94	-0.535	-0.589	-0.181	0.566	0.221	0.999
EOPT	0.69	1.00	1.00	10.00	299.94	-0.538	-0.636	0.023	0.675	0.715	1.000
Std Dev	0.05	4.00	0.33	5.4	0.33	0.042	0.110	0.249	0.231	0.316	0.318
Mean	0.74	8.16	1.19	16.3	299.76	-0.521	-0.541	-0.239	0.445	0.744	0.808
Manual	1.70	2.00	3.00	17.00	168.00	-0.220	-0.140	-0.100	-0.170	-0.110	-0.110

calculated across the objective functions range from -24% (P0SREP2) to +81% (P0RREP3); coefficients of variation range from -104% (P0SREP2) to 52% (P0RREP2).

5.4 Sensitivity to Initial Random Seed

Given the sensitivity to the initial random seed noted in the synthetic calibrations of Section 5.2, this section investigates the influence of the initial random seed on calibrations against observed data. Sensitivity of final parameter values to the seed value could potentially overshadow any insights into trade-offs between different calibration objectives. All calibrations in this section implement the EOPT' objective function.

5.4.1 Coquitlam Lake Watershed

Calibrations are carried out using the expanded feasible ranges for parameters P0GRADU and V0FLAS, the two parameters outside the recommended range in the BC Hydro expert calibration. All other parameters are constrained to the ranges outlined in Table 4-1. Allowing only P0GRADU and V0FLAS to vary over their expanded ranges will further investigate the potential for improvement over the results of Section 5.3.1. Five independent trials are used, each with a different arbitrarily-selected seed value. All calibrations terminate when the SCE-UA algorithm is unable to improve the objective function of the best member of the population within ten shuffling loops.

Summary statistics for the full period of record (excluding the first year for spin-up) are presented in Table 5-10. The statistics are stable, changing by only 2% in the most extreme case. The example annual hydrograph in Figure 5-16 supports the statistical data.

The calibrations of this section use different periods of data than those of Section 5.3.1. Comparison of Table 5-10 with Table 5-4 reveals that results are comparable, although the statistics in Table 5-4 are slightly better for E! and D!, and more significantly so for EOPT! and dV/V. This observation is consistent with expectations, since the summary statistics in Table 5-4 are based only on calibration data, while those of this section are calculated for a period including both calibration and validation data.

Table 5-10: Summary Statistics for Coquitlam Lake Watershed Seed Sensitivity
(Water Years 1986-87 through 1998-99)

Objective	E!	D! (R^2)	EOPT!	dV/V
Trial 1	0.76	0.77	0.67	0.10
Trial 2	0.76	0.76	0.65	0.10
Trial 3	0.76	0.77	0.67	0.09
Trial 4	0.76	0.77	0.67	0.09
Trial 5	0.76	0.77	0.67	0.09
BC Hydro	0.75	0.76	0.71	0.04

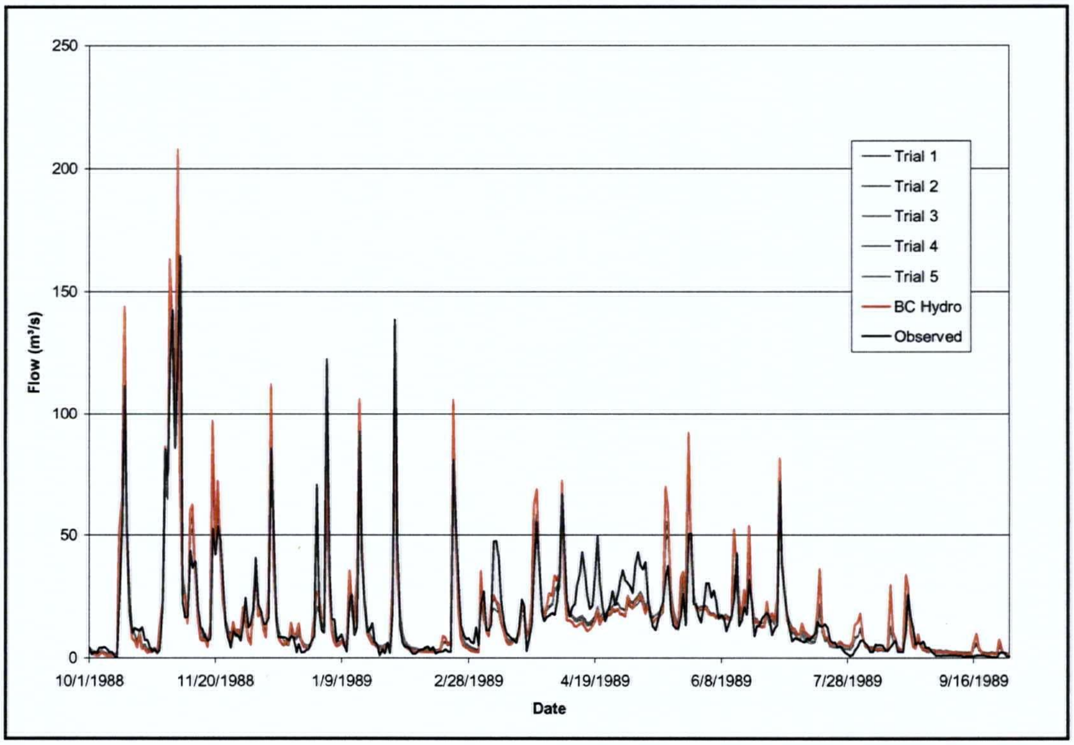


Figure 5-16: Typical Annual Hydrograph for Coquitlam Lake Watershed Seed Sensitivity

Objective function evolution for the set of calibrations is documented in Figure 5-17. The figure shows little basis for differentiating between the trials.

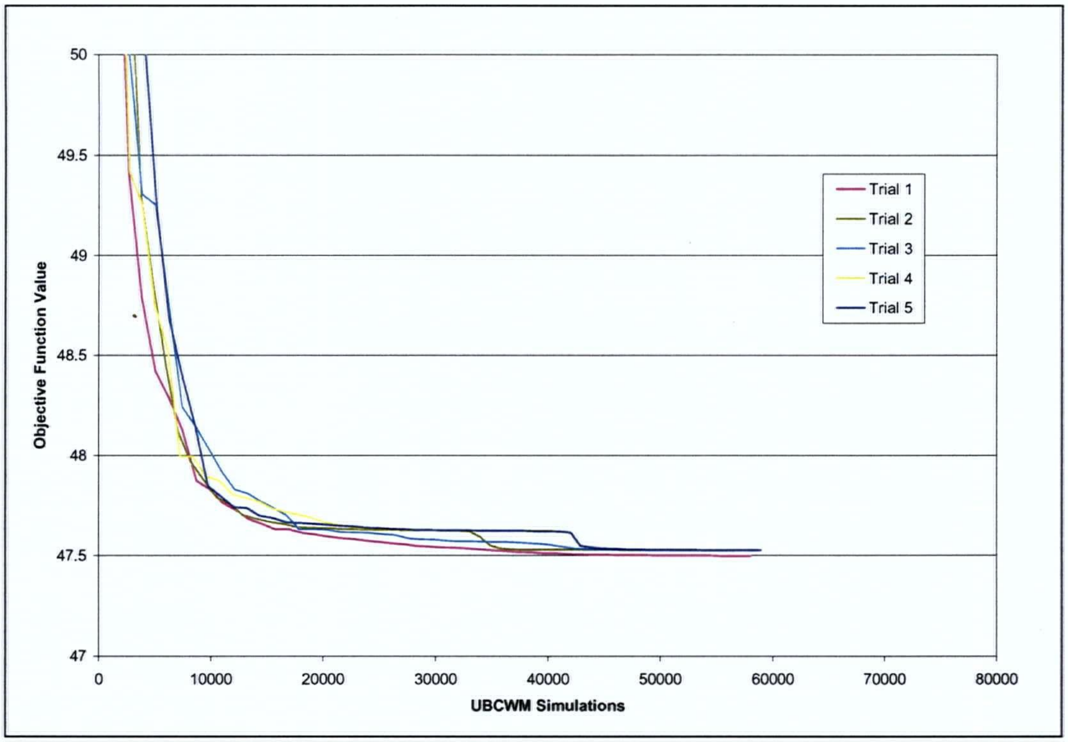


Figure 5-17: Objective Function Evolution for Coquitlam Lake Watershed Seed Sensitivity

The normalized parameter values shown in Figure 5-18 demonstrate variability comparable to that observed across the multiple objective calibrations of Figure 5-12. Notable variability is expected amongst the precipitation distribution parameters; however, there is also a bi-modal response from the upper zone groundwater time constant P0UGTK. No reason for this exception to the otherwise limited parameter variability is immediately apparent, beyond the fact that the P0UGTK parameter for EOPT', E!, and EOPT has assumed its boundary value and is therefore likely subject to compensating errors. However, all three objective functions that result in a boundary value for P0UGTK are based on the Nash-Sutcliffe Efficiency statistic.

Final parameter values for the five seed sensitivity trials are provided in Table 5-11. Of note, the middle precipitation gradient P0GRADM for the second trial is rendered irrelevant by the proximity of the values for E0LMID and E0LHI, since no elevation band midpoint falls within

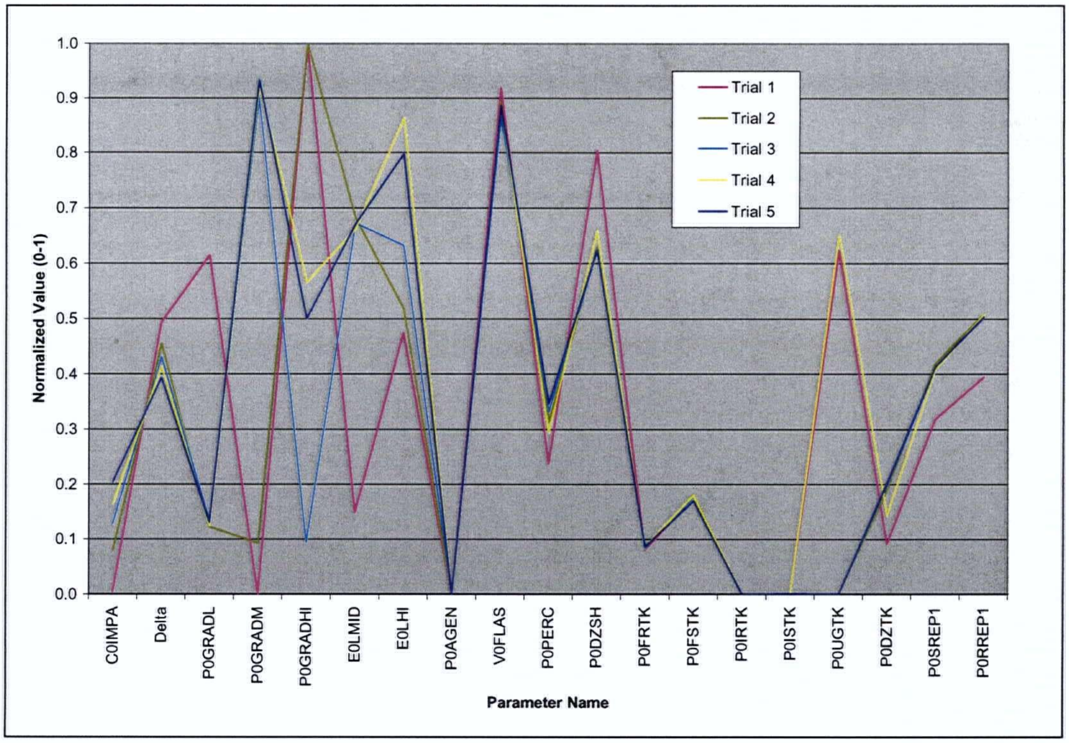


Figure 5-18: Normalized Parameter Values for Coquitlam Lake Watershed Seed Sensitivity

the range over which P0GRADM is applied. Trials four and five do not utilize the upper precipitation gradient P0GRADU, as it becomes effective at an elevation above that of the highest elevation band in the Coquitlam Lake watershed model. The boundary values of rain- and snow-based interflow time constants P0IRTK and P0ISTK is unsurprising given the near-boundary and boundary values calibrated by BC Hydro.

Only trials one and two result in parameter values beyond the range recommended in the UBCWM User's Manual (Quick et al., 1994), and in both cases it was the upper precipitation gradient P0GRADU. The lack of appreciable difference between the results of these calibrations and those of other trials implies that using an expanded parameter range for parameters P0GRADU and V0FLAS has no significant effect on SCE-UA calibration performance or the quality of the results obtained thereby.

5.4.2 Illecillewaet River Watershed

Six calibrations are performed to investigate seed sensitivity for the Illecillewaet River watershed. Each calibration uses the EOPT' objective function. All calibrations are performed

Table 5-11: Parameter Values for Coquitlam Lake Watershed Seed Sensitivity

Objective	C0IMPA	Delta	P0GRADL	P0GRADM	P0GRADU	E0LMID	E0LHI	P0AGEN	V0FLAS	P0PERC	P0DZSH
Trial 1	0.01	0.25	12.3	0.0	29.9	638	1475	25.0	36.8	11.87	0.80
Trial 2	0.08	0.23	2.5	1.9	30.0	1491	1519	25.0	35.8	15.69	0.63
Trial 3	0.13	0.22	2.7	18.0	2.8	1476	1633	25.0	34.6	16.55	0.63
Trial 4	0.16	0.21	2.5	18.4	17.0	1459	1864	25.0	35.5	14.67	0.66
Trial 5	0.20	0.20	2.6	18.7	15.0	1475	1799	25.0	35.5	17.33	0.63
Std Dev	0.08	0.02	4.4	9.6	19.2	374	170	0.0	0.8	2.12	0.08
Mean	0.12	0.22	4.5	11.4	16.4	1308	1658	25.0	35.6	15.22	0.67
BC Hydro	0.15	0.05	1.6	9.4	23.7	1370	1480	100.0	1.0	13.00	0.58

Objective	P0FRTK	P0FSTK	P0IRTK	P0ISTK	P0UGTK	P0DZTK	P0SREP	P0RREP
Trial 1	0.16	0.35	1.00	1.00	34.94	73.06	-0.364	-0.209
Trial 2	0.16	0.36	1.00	1.01	10.00	97.58	-0.160	0.019
Trial 3	0.17	0.35	1.00	1.00	10.01	99.48	-0.177	0.007
Trial 4	0.17	0.36	1.00	1.01	36.05	85.72	-0.182	0.016
Trial 5	0.17	0.34	1.00	1.01	10.00	100.71	-0.173	0.009
Std Dev	0.01	0.01	0.00	0.00	13.97	11.82	0.086	0.099
Mean	0.17	0.35	1.00	1.01	20.20	91.31	-0.211	-0.032
BC Hydro	0.14	0.17	1.20	1.00	21.00	96.00	0.025	0.105

on the same data set (i.e., water years 1981-82 through 1988-89). All six trials terminate when the SCE-UA algorithm is unable to improve on the objective function value of the best member of the population within ten shuffling loops.

Summary statistics for water years 1971-72 through 1989-90 are presented in Table 5-12. This period represents the full period of record for the watershed after discarding the 1970-1971 water year for model spin-up. The table indicates that statistical performance is fairly insensitive to initial random seed, resulting in a variation of at most 3% from the original calibration (Trial 1). However, differences are observable in the example hydrograph shown in Figure 5-19. A more detailed examination of the statistics for the simulation reveals that the seasonal and annual values do not exhibit the same degree of consistency across the trials. This is not necessarily surprising, since the calibration objective is to optimize performance against an aggregate measure of inter-annual performance.

Table 5-12: Summary Statistics for Illecillewaet River Watershed Seed Sensitivity
(Water Years 1971-72 through 1989-90)

Objective	E!	D! (R^2)	EOPT!	dV/V
Trial 1	0.85	0.88	0.68	-0.17
Trial 2	0.86	0.91	0.67	-0.19
Trial 3	0.86	0.91	0.68	-0.18
Trial 4	0.87	0.89	0.70	-0.17
Trial 5	0.85	0.88	0.67	-0.18
Trial 6	0.86	0.89	0.67	-0.19
Manual	0.90	0.93	0.82	-0.08

Figure 5-20 shows that even the evolution of the objective function across multiple trials does not display the degree of consistency that is observed for the Coquitlam Lake watershed. At termination, there is still residual disparity between the objective function values for different trials. It appears that there are three different locally-optimal values for the objective function, each of which captures two trials.

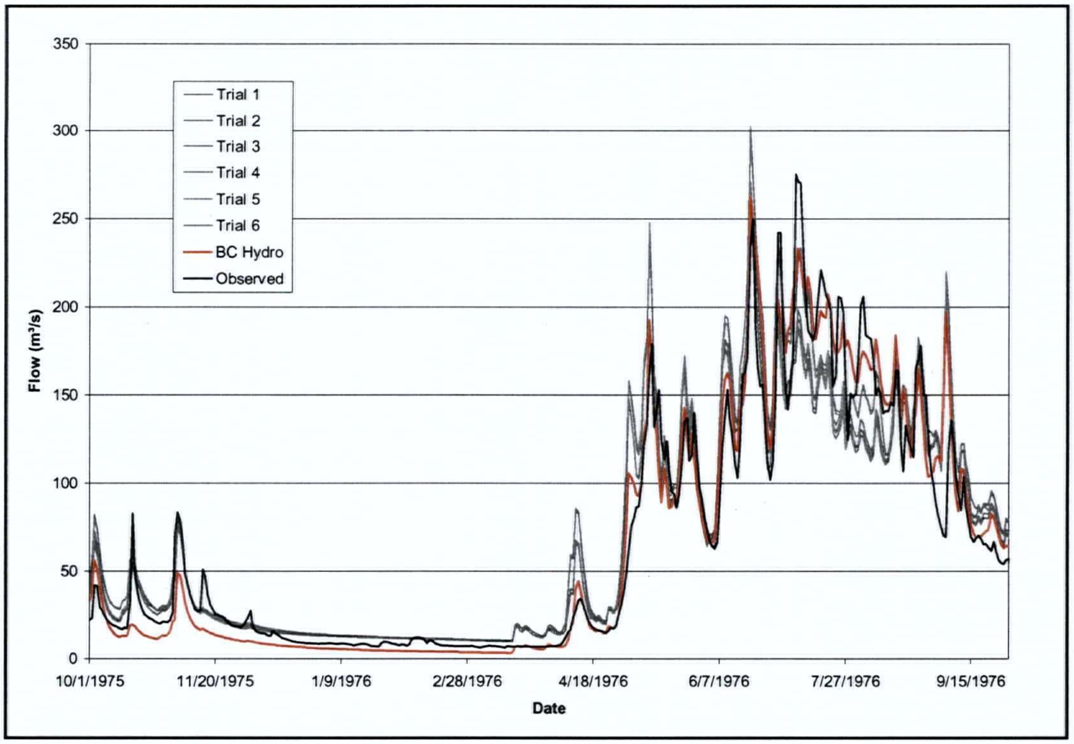


Figure 5-19: Typical Annual Hydrograph for Illecillewaet River Watershed Seed Sensitivity

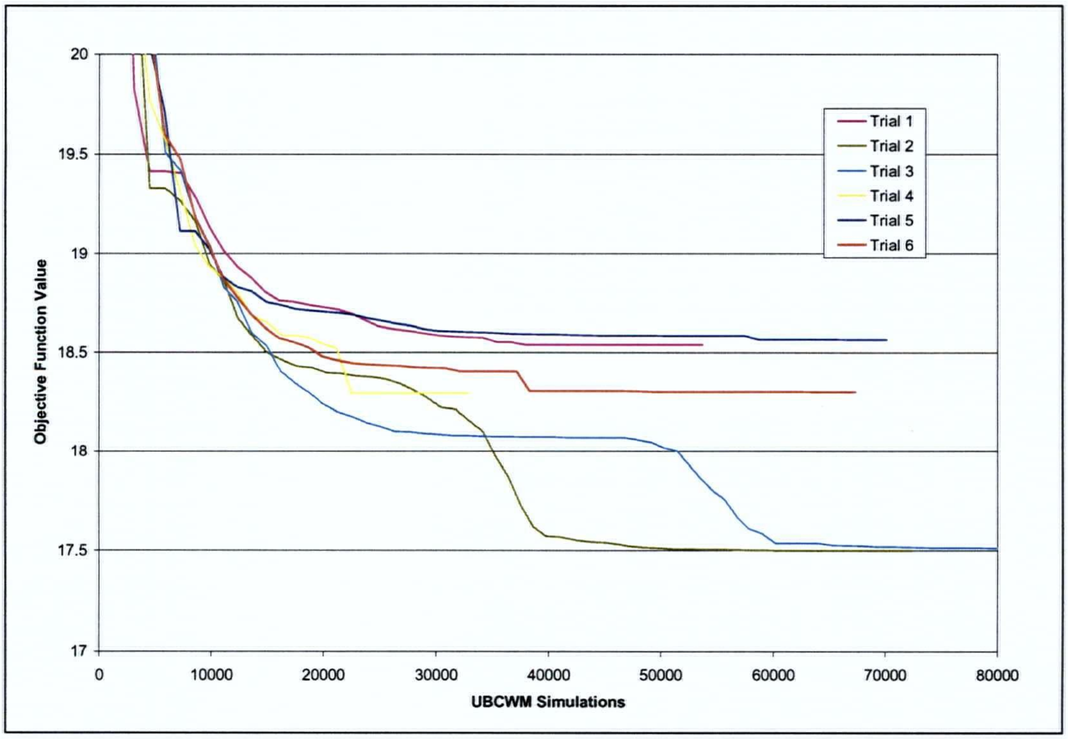


Figure 5-20: Objective Function Evolution for Illecillewaet River Watershed Seed Sensitivity

Figure 5-21 is a plot of the normalized parameter values obtained from each of the six calibrations. The variability therein is similar in magnitude to that observed for the corresponding multi-objective calibrations in Figure 5-15. Once again, there are a significant number of parameters having values at or near their boundary condition. Final parameter values for each trial can be found in Table 5-13. These results lead one to believe that the SCE-UA automatic calibration algorithm is incapable of producing consistent and optimal calibrations for the Illecillewaet River watershed, although the resulting statistical performance of the subsequent simulations is generally indicative of the quality of simulation possible with intensive expert calibration.

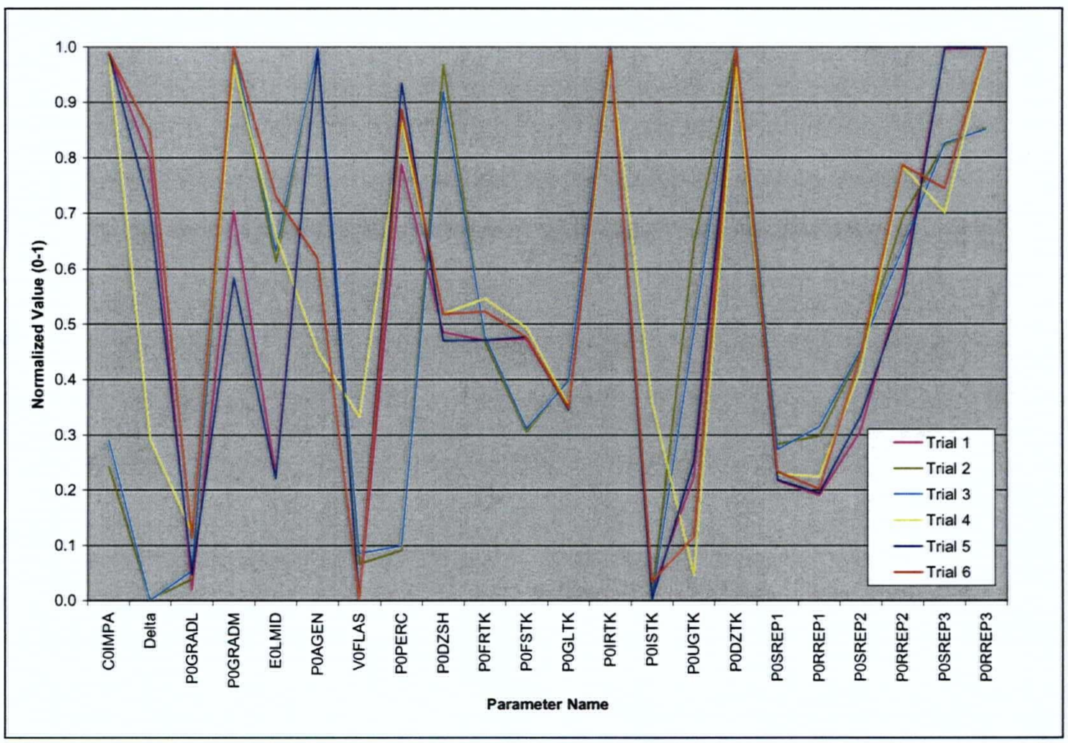


Figure 5-21: Normalized Parameter Values for Illecillewaet River Watershed Seed Sensitivity

5.5 Event Simulations

Having established multi-objective calibrations for the Coquitlam Lake watershed in Section 5.3.1, this section reviews the performance of the resulting parameter sets when applied to simulations of individual events of record at an hourly timestep. The start and end dates of the five events considered are given in Table 4-4.

Table 5-13: Parameter Values for Illecillewaet River Watershed Seed Sensitivity

Objective	C0IMPA	Delta	P0GRADL	P0GRADM	E0LMID	P0AGEN	V0FLAS	P0PERC	P0DZSH	P0FRTK	P0FSTK
Trial 1	0.99	0.40	0.4	14.1	1010	199.4	20.0	39.39	0.49	0.94	0.95
Trial 2	0.24	0.00	0.8	19.9	1702	200.0	21.3	4.64	0.97	0.94	0.61
Trial 3	0.29	0.00	1.1	20.0	1735	199.9	21.7	5.03	0.92	0.96	0.62
Trial 4	0.98	0.15	2.5	19.3	1790	103.9	26.7	43.16	0.52	1.09	0.99
Trial 5	0.99	0.35	0.9	11.7	996	199.3	20.0	46.74	0.47	0.94	0.95
Trial 6	0.99	0.42	2.3	20.0	1915	133.1	20.0	44.42	0.52	1.05	0.95
Std Dev	0.37	0.20	0.9	3.7	411	42.9	2.6	20.07	0.23	0.07	0.18
Mean	0.75	0.22	1.4	17.5	1525	172.6	21.6	30.56	0.65	0.99	0.85
Expert	0.40	0.05	4.0	2.0	1009	100.0	36.0	31.00	0.25	0.78	1.00

Objective	P0GLTK	P0IRTK	P0ISTK	P0UGTK	P0DZTK	P0SREP1	P0RREP1	P0SREP2	P0RREP2	P0SREP3	P0RREP3
Trial 1	0.69	9.91	1.15	18.93	299.78	-0.566	-0.620	-0.386	0.161	0.992	0.995
Trial 2	0.80	10.00	1.06	35.89	299.85	-0.433	-0.404	-0.099	0.386	0.645	0.707
Trial 3	0.80	10.00	1.01	29.48	299.00	-0.453	-0.370	-0.092	0.272	0.651	0.703
Trial 4	0.71	9.75	4.24	11.88	292.56	-0.542	-0.552	-0.160	0.568	0.402	0.994
Trial 5	0.69	9.99	1.01	20.09	299.97	-0.563	-0.611	-0.331	0.104	0.999	0.995
Trial 6	0.70	9.95	1.31	14.62	299.51	-0.533	-0.595	-0.113	0.573	0.489	1.000
Std Dev	0.05	0.10	1.29	9.15	2.90	0.057	0.110	0.129	0.200	0.250	0.150
Mean	0.73	9.93	1.63	21.82	298.45	-0.515	-0.525	-0.197	0.344	0.696	0.899
Expert	1.70	2.00	3.00	17.00	168.00	-0.220	-0.140	-0.100	-0.170	-0.110	-0.110

As noted in Section 2.3.4, every event-based model must be initialized with a suite of initial conditions. Initial conditions are calculated for the UBCWM according to the procedure outlined in Section 4.2.4. To check the sensitivity to initial conditions, Storm 1 is executed twice for each calibration, once with initial conditions determined from its own continuous simulation, and once with initial conditions derived from a continuous simulation using the BC Hydro expert-calibrated parameter set.

As Figure 5-22 illustrates, the hydrographs for the SCE-UA and BC Hydro-derived initial conditions are essentially identical, even at the beginning of the event. Although the figure shown illustrates only the results for EOPT', results for SLS, E!, LAD, and EOPT are similar. Since the event simulations appear almost completely insensitive to initial conditions, all further simulations use initial conditions taken from a continuous simulation using the BC Hydro parameter set. Although the sensitivity to initial conditions may be greater for other events (e.g., wetter or drier antecedent conditions, more or less snowpack), the lack of sensitivity in Figure 5-22 suggests that any increase in sensitivity to initial conditions will not be significant.

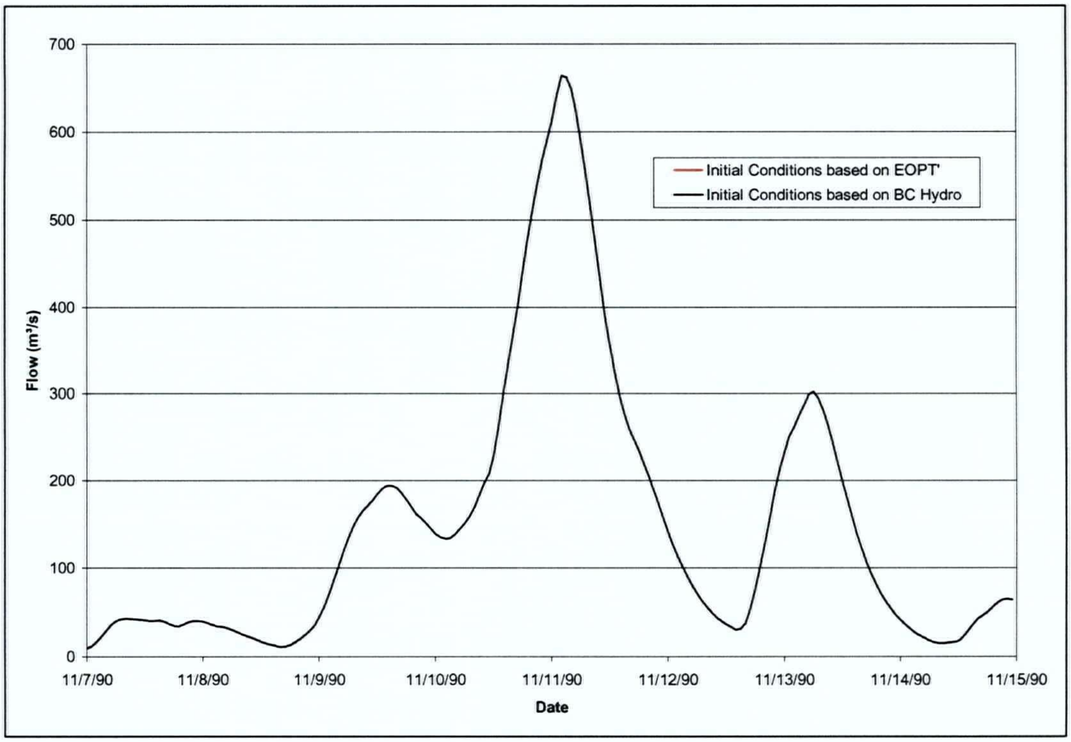


Figure 5-22: Typical Hydrograph Comparing of Initial Watershed Conditions for Storm 1
(Objective Function EOPT')

The five parameter sets given in Table 5-5 are used to simulate each of the five storm events. The Pareto hydrograph for Storm 1 (i.e., the set of hydrographs arising from the suite of multi-objective calibrations) is shown in Figure 5-23. Pareto hydrographs for the other storm events demonstrate a similar degree of variability, although the quality of the simulations vary. The under-estimation of peak flow with respect to the BC Hydro estimate is reflected in all events.

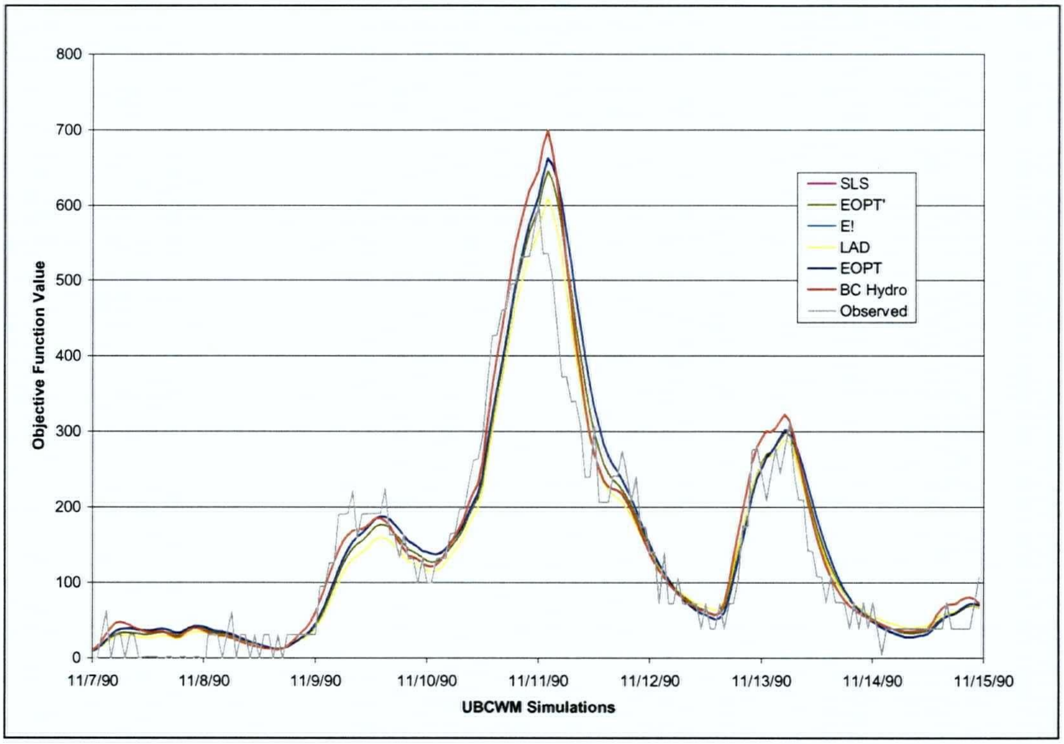


Figure 5-23: Coquitlam Lake Watershed Storm 1 Pareto Hydrograph

Summary results for Storm 1 are shown in Table 5-14; those for other storm events are included in Appendix B. Event volumes in Table 5-14 are expressed in cms-days, where 1 cms-day = 1 m^3/s * 1 day. Although Table 5-14 indicates that the SCE-UA calibrations outperform the BC Hydro calibration for Storm 1, results for the other events range from comparable to slightly favouring the BC Hydro simulations.

Table 5-15 contains summary statistics for the key attributes of all five storms. Mean errors for each quantile are calculated as the mean of the percentage errors for the various objective functions. The coefficient of variation (CoV), defined as the ratio of standard deviation to mean, is calculated using the same approach as for mean error.

Table 5-14: Summary Results for Storm 1 Simulations

Objective	Peak Flow (m^3/s)	Event Volume (cms · d)	Time to Peak (hours)
Observed	601	26891	96
BC Hydro	699	29930	98
SLS	662	29597	98
EOPT'	646	28592	98
E!	663	29608	98
LAD	610	27127	98
EOPT'	662	29578	98
Mean	649	28900	98
Std Dev	22.6	1082	0.0
CoV	3%	4%	0%
Mean Error	8%	7%	2%
BCH Error	16%	11%	2%

Since this thesis focuses on uncertainty rather than accuracy, CoV values are more relevant than mean error. Time to peak is estimated with the greatest precision; the only case where the peak flow does not occur at the same time for all objective functions was Storm 4. Event volumes are somewhat more variable, having CoVs of 4% to 5%. However, the CoV remains fairly constant for each storm scenario. As expected, peak flow estimates are most variable across the Pareto set, with CoVs ranging from 1% for Storm 5 to 5% for Storm 4.

Section 5.3.1 identifies LAD as the poorest performer amongst the objective functions. A recalculation of event statistics not including LAD reduces peak flow and event volume CoVs for each storm by at least 2%, with a lesser improvement for mean error.

Table 5-15: Average Results for All Multi-Objective Event Simulations

Storm Event	Peak Flow		Event Volume		Time to Peak	
	CoV	Mean Error	CoV	Mean Error	CoV	Mean Error
Storm 1	3%	8%	4%	7%	0%	2%
Storm 2	4%	-22%	4%	-4%	0%	3%
Storm 3	3%	-16%	5%	-1%	0%	2%
Storm 4	5%	-44%	4%	-3%	1%	7%
Storm 5	1%	-31%	4%	-14%	0%	-9%

5.6 Extreme Event Simulations

This section summarizes the multi-objective-based simulation of extreme events. Section 5.6.1 applies the multi-objective calibrations of section 5.3.1 to a PMF scenario for the Coquitlam Lake watershed. To re-create the pre-determined distributed PMP for the Coquitlam Lake Watershed, the calibrated precipitation parameters are replaced with pre-defined values. While these simulations are based on BC Hydro's PMF scenario, they do not use BC Hydro's PMF parameter set, and therefore may differ significantly from the actual PMF for Coquitlam Lake.

Section 5.6.2 provides the results of simulations in which the meteorologic input files for the PMP are considered as if they represent observed data and applied with the calibrated precipitation parameters. In both cases, performance is benchmarked against a simulation using BC Hydro's continuous daily calibration for Coquitlam Lake.

5.6.1 Simulations Based on a PMF Scenario

The Pareto hydrograph for the SCE-UA simulations is shown in Figure 5-24. The hydrographs of the SCE-UA objective functions cluster together, with a visible separation from the BC Hydro simulation occurring at peak flow.

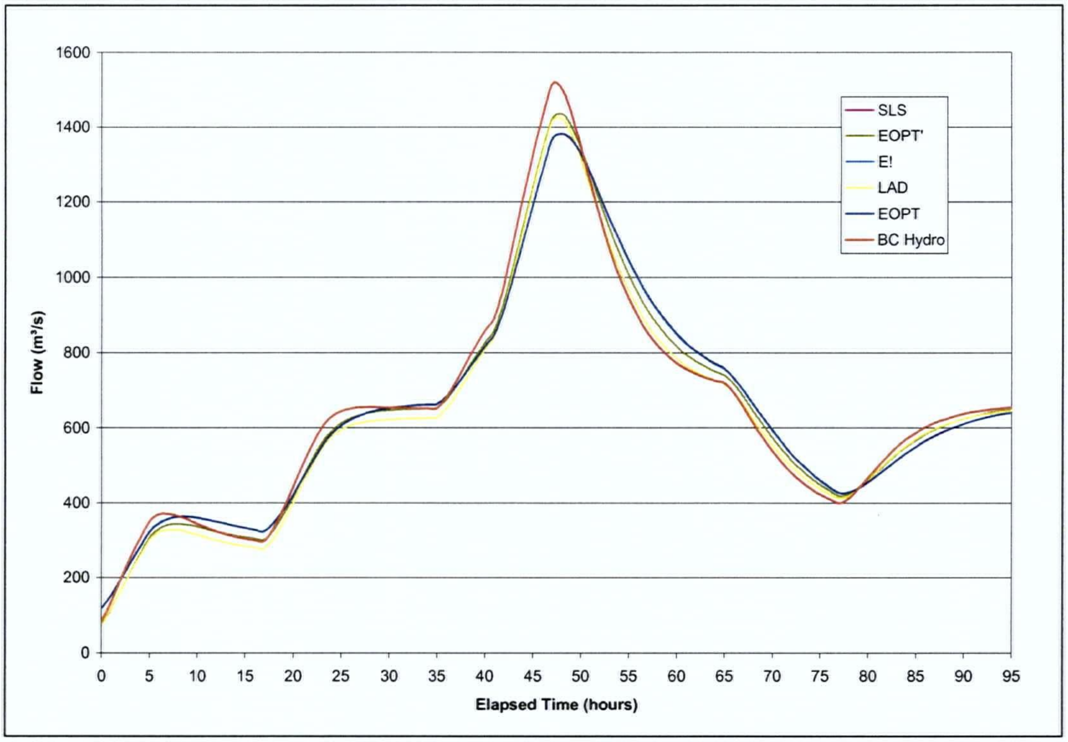


Figure 5-24: Pareto Hydrograph for PMF-based Extreme Event Simulations

Summary results including peak flow, event volume, and time to peak for each simulation are given in Table 5-16. Estimates of peak flow range from 1381 m³/s (SLS) to 1485 m³/s (EOPT'), with a mean of 1400 m³/s and a CoV of 2%. Event volumes fall between approximately 60,340 and 62,440 cms-hours, where 1 cms-hour = 1 m³/s * 1 h. The event volumes have a mean of 61,896 cms-hours and a CoV of 1%. All SCE-UA objective functions yield a time to peak of 48 hours. The small CoVs for peak flow and event volume suggest that the variability across the multiple automatic calibrations is small in comparison with the storm magnitudes.

Including the BC Hydro expert calibration as a member of the Pareto set slightly increases mean values for peak flow and event volume, and results in a non-zero CoV for time to peak. Most significantly, the CoV for event volume is unchanged and the CoV for peak flow increases from 2% to 4%. In general, the extreme event hydrographs remain fairly precise across the Pareto set. This implies that any uncertainty associated with different calibration objectives does not translate into large variability in the runoff estimate when considered in the context of the uncertainty surrounding the definition and calculation of events on the scale of the PMF.

Table 5-16: Summary Results for PMF-based Extreme Event Simulations

Objective		Peak Flow (m^3/s)	Event Volume (cms · h)	Time to Peak (hours)
BC Hydro		1517	62408	47
SLS		1381	62426	48
EOPT'		1435	61885	48
E!		1382	62388	48
LAD		1422	60342	48
EOPT		1382	62441	48
SCE-UA Calibrations	Mean	1400.4	61896	48
	Std Dev	26.1	899	0.0
	CoV	2%	1%	0%
SCE-UA & BC Hydro Calibrations	Mean	1420	61982	47.8
	Std Dev	53.0	831	0.4
	CoV	4%	1%	1%

5.6.2 Simulations Based on a PMP-Scale Event

The Pareto hydrograph for the PMP-based extreme event is shown in Figure 5-25. It is obvious from the hydrograph that the performance of LAD is significantly different from the other SCE-UA calibrations and thereby increases the variability of the Pareto set. Once again, the results from the BC Hydro expert calibration are distinct from those of the automatic calibrations pursued in this work.

Summary results are presented in Table 5-17, including peak flow in m^3/s, event volume in cms-hours, and time to peak in hours. Peak flow estimates range from 1295 m^3/s to 1356 m^3/s, with a mean of 1342 m^3/s and a CoV of 2%. Estimates of event volume fall between 55,159 cms-hours and 61,074 cms-hours. The mean estimate for event volume is 59,346 cms-hours, with a CoV of

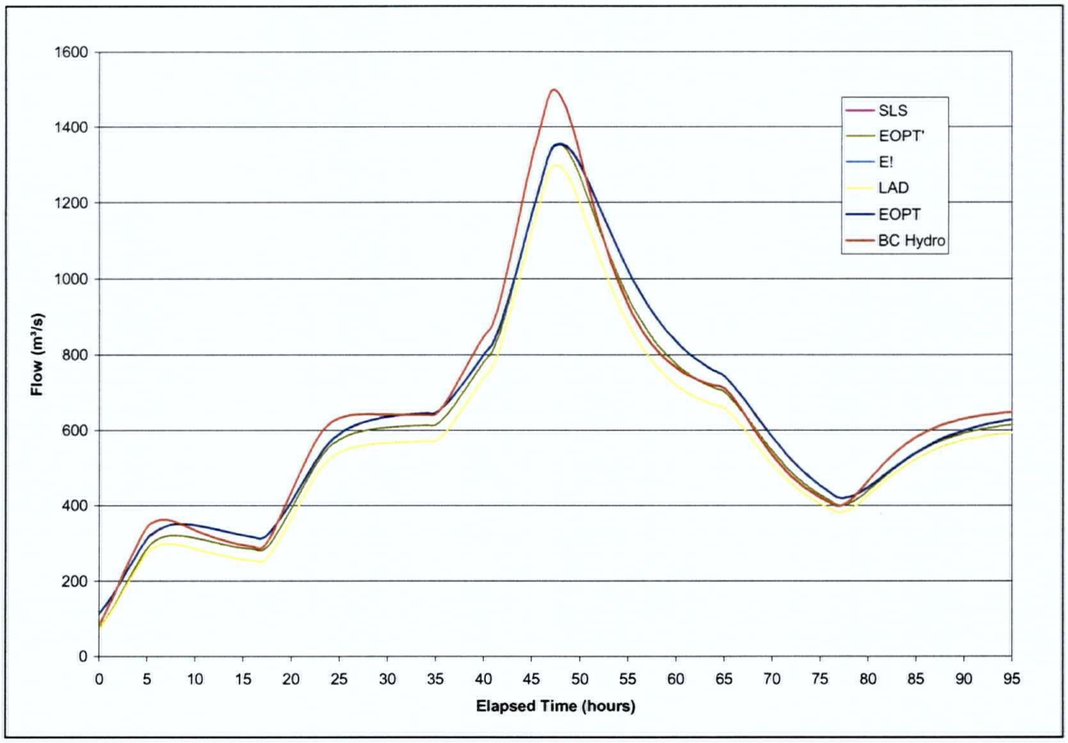


Figure 5-25: PMP-based Storm Pareto Hydrograph

4%. All SCE-UA simulations experience peak flow at hour 48 of the simulation. The obvious discrepancy between LAD and the other objective functions results in a significant decrease in CoVs for peak flow and event volume when the LAD calibration is removed from the analysis.

Considering the BC Hydro expert calibration as a member of the Pareto set slightly increases mean estimates for peak flow and event volume. The CoV for time to peak is once again influenced by the peaking of the BC Hydro simulation in hour 47. The CoV for peak flow undergoes the most significant change in magnitude, moving from 2% to 5% in response to the introduction of a peak flow 140 m^3/s (or approximately 10%) greater than the next largest member of the Pareto set. However, the 5% CoV in event peak flow and a 4% CoV in event volume are not necessarily significant when considering the other uncertainties affecting the extreme event estimate.

Table 5-17: Summary Results for PMP-based Storm Simulations

Objective		Peak Flow (m^3/s)	Event Volume (cms · h)	Time to Peak (hours)
BC Hydro		1496	61526	47
SLS		1353	61019	48
EOPT'		1353	58440	48
E!		1356	61074	48
LAD		1295	55159	48
EOPT		1353	61038	48
SCE-UA Calibrations	Mean	1342	59346	48
	Std Dev	26.3	2598	0.0
	CoV	2%	4%	0%
SCE-UA & BC Hydro Calibrations	Mean	1368	59709	47.8
	Std Dev	67.1	2488.6	0.4
	CoV	5%	4%	1%

6. Conclusions

> *"A good forecaster is not smarter than everyone else; he merely has his ignorance better organized."*
>
> *- Author Unknown*

Most water management and flood protection decisions require an estimate of expected flows, and many decision-makers rely on computer-based hydrologic models to supply the necessary information. However, the estimates available from these models are often highly uncertain. This work presents a literature-based review of uncertainty in hydrologic modelling and develops one possible approach for exploring the following questions:

- What kind of variability in parameter values is associated with alternative "acceptable" calibrations?
- How much variability arises when the parameter values identified by these alternative "acceptable" calibrations are used to estimate extreme flood events?

This chapter summarizes conclusions arising from the literature-based review and the investigation into calibration-related uncertainty for the Coquitlam Lake and Illecillewaet River watersheds.

The diverse nature and scale of hydrologic processes is the root cause of our inability to model hydrology with accuracy and precision. Where precise and accurate methods for measurement and modelling exist, they are typically confined to the point or laboratory scale. Where large-scale measurements and process models exist, detailed accuracy and precision are almost always compromised. Unlike the more heavily-studied field of hydraulics, there is currently no recourse to analog or scaled physical models, and catchment-scale field experiments are generally not repeatable.

The enormous increase of available computing power has caused some hydrologic models evolve more quickly as technical algorithms than as scientific tools. Different kinds of models have emerged, each having different attributes and inputs, different levels of simplification, and different characteristics of uncertainty. Each different class of model may be more or less applicable for a given purpose or scenario. The suitability of any model for a task is largely dependent on the resources and information available, the required degree of accuracy, and the purpose of the results.

Calibration and validation are widely accepted as necessary for optimizing the performance of a hydrologic model. Most practical calibrations are achieved through the manual efforts of expert users. However, automatic techniques such as the SCE-UA method have advanced to the point where they can identify a statistically – if not hydrologically – defensible approximation to the optimal solution for a given objective function, if such exists. In particular, the SCE-UA method has been widely shown to be capable of identifying the global optimum for a given response surface.

Even where a definitive global optimum is identified, it may not best represent the watershed from a hydrologic standpoint. It is not generally advisable to attempt to draw conclusions about the physical processes or properties of a watershed from a calibrated parameter set, especially where the calibration is automatic or the model highly simplified (e.g., statistical, empirical, or conceptual models).

Split-sample model validation is almost always used as a final check on the performance of a calibrated model. Successful validation increases the likelihood that a model is acceptable, but does not constitute proof or confirmation that the model is an accurate representation of *in situ* processes and results. The validation process constrains but does not eliminate uncertainty. While validation is necessary to prove that the model is generally capable of producing reasonable simulations, its success should not be interpreted as sufficient justification for extrapolating a model to simulations whose character or magnitude differ significantly from the calibration and validation data series.

Every comprehensive application or study of a hydrologic model should include uncertainty analysis during calibration. Many methods of uncertainty analysis have been widely recognized for decades as a vital step in understanding model performance (e.g., sensitivity analysis); others are just emerging in response to recent advances in computational technology (e.g., MCMC-based methods such as SCEM-UA). Most methods of uncertainty analysis are complimentary rather than competitive, as each typically explores a slightly different aspect of model predictive uncertainty.

Synthetic calibration explores model predictive uncertainty by investigating how well an automatic calibration can re-create a set of known parameter values and model outputs. Because it is known with certainty that the calibration data can be accurately and precisely reproduced by the subject model, model predictive uncertainty should be highly constrained. Any viable pairing of model and calibration algorithm should perform well under such simplified conditions. In this work, synthetic calibration is used to evaluate the ability of the SCE-UA method to identify an optimal parameter set for the UBCWM. The UBCWM is applied to synthetic data sets corresponding to three watersheds: the Campbell River, the Illecillewaet River, and the Coquitlam River above Coquitlam Dam (Coquitlam Lake).

Objective functions used include the Simple Least Squares (SLS) equation, the Heteroscedastic Maximum Likelihood Estimator (HMLE), the Nash-Sutcliffe Efficiency statistic (E!), the Coefficient of Determination (R^2 or D!), and the Least Absolute Difference (LAD) of the time series quantiles. The set of objectives also includes the EOPT! statistic, a variation of Nash-Sutcliffe Efficiency commonly used in evaluating the UBCWM, and two numerical variations thereof designated EOPT and EOPT'.

In general, tests showed that the SCE-UA method was able to achieve a calibration that essentially duplicates the original synthetic data series. Only the EOPT! and D! objective functions perform poorly in synthetic calibration; this is attributed to potential mathematical weaknesses in identifying a good match for the target calibration data. Optimal SCE-UA performance is achieved using a total of 20 complexes, and is not improved by increasing the length of the calibration data series.

Although the SCE-UA method duplicates the synthetic calibration data with all objective functions except EOPT! and D!, some parameters demonstrate pronounced variability across the Pareto set. "Optimum" values for these parameters are therefore non-unique, and the ultimate values of these parameters differ for each objective function. More surprisingly, a subset of parameters also demonstrates variability when different random seeds are used in the initialization phase.

The subset of parameters with non-unique "optimal" values includes the parameters governing precipitation distribution. One possible explanation is that this aspect of the model may be mathematically overparameterized with respect to the limited data available. The observed relationship between complex precipitation distribution in the UBCWM and increased variability in the Pareto set supports this hypothesis. Incorporating additional data (e.g., snowcourses) into the calibration process could potentially constrain this poorly-identifiable area of the model. Alternatively, MCMC-based algorithms may ultimately yield better performance for overparameterized components of a hydrologic model because they preserve ergodicity throughout the calibration.

Because the objective functions converge to their limiting values (i.e., zero error in the calibration data series), the synthetic calibrations definitively show that alternative but duplicate representations of the subject catchment's hydrologic response are possible using the UBCWM. The synthetic calibrations explored herein correspond to the eight "end points" of the Pareto set (i.e., the calibrations associated with 100% relative weight for each objective function). Additional calibrations are unlikely to generate substantial additional insight within the current framework given the observed discrepancy between convergence of the objective function and convergence of parameter values for the eight "end points".

Once the synthetic calibrations establish that the SCE-UA method is capable of calibrating the UBCWM, calibrations against actual data for the Coquitlam and Illecillewaet watersheds are produced using the SCE-UA method. This process addresses the first question of interest noted above – how much variability is associated with alternative representations of applied conditions?

For the Coquitlam Lake watershed, automatic calibrations of the UBCWM using observed data are of comparable statistical quality to BC Hydro's expert (manual) calibration. Visual inspection of annual hydrographs appears to confirm this result, as a "best" hydrograph is not necessarily obvious at first glance. Although many parameters assume boundary values in this calibration, relaxing their constraints provides no improvement in performance or simulation quality.

For the Illecillewaet River watershed, the results of automatic calibrations using observed data are statistically superior to those for the Coquitlam Lake watershed in some aspects, but are significantly less consistent. Visual inspection shows that the automatic calibrations are obviously of lower quality than the BC Hydro expert calibration. The additional variability is attributed to the increased complexity in scaling precipitation data.

The relative variability of different parameters across the Pareto sets for the Coquitlam and Illecillewaet watersheds is markedly similar to synthetic results. Although variability is naturally greater across the board for the observed-data calibrations, most parameters still demonstrate a fair degree of precision in Pareto values. The notable exceptions are again the parameters governing precipitation distribution. As before, this is a possible indicator of overparameterization in this aspect of the model.

Visual inspection of the calibrated hydrographs reveals situations where neither manual nor automatic calibration can effectively replicate the observed data. In such cases, the automatic calibration algorithm sacrifices intuitively-consistent performance in favour of numerical improvement. This is most obvious in cases where better statistical performance results from shifting a flat simulated hydrograph upward to match a spike in the observed data, or vice versa. While the overall statistics might improve, the model has obviously become less representative. These observations demonstrate the importance of establishing a high-quality set of calibration data. Where data are believed to be of high quality but no configuration of the model is able to provide an acceptable simulation, there is cause to consider removing the irresolvable data from consideration during an automatic calibration. However, this should only be considered once the modeller is convinced that the model is an appropriate representation of the subject catchment.

For most calibrations performed using observed data, the SCE-UA method terminates after being unable to further improve the value of the objective function. One can therefore conclude that achieving a significantly better solution for that particular objective function is unlikely using the SCE-UA method. Results from these calibrations confirm that automatic calibration can provide an approximate upper bound for the statistical quality of simulation that can be attained with a given data set and model. As observed, it should be possible to reproduce at least this level of success with expert calibration.

Ultimately, the statistical and graphical results of the various Coquitlam and Illecillewaet calibrations indicates that a significant amount of variability in parameter values for a subset of parameters can exist within a set of "acceptable" calibrations.

Consideration of uncertainty in calibration is critical when one is extrapolating a calibrated hydrologic model to simulate extreme events. This is particularly true in cases such as this where the initial context itself has a substantial degree of uncertainty. Using the event-based UBCWM representation of the Coquitlam Lake watershed, this work investigates the second question noted at the beginning of the chapter, namely how much variability arises from using the Pareto set of calibrations to simulate an extreme runoff event.

Event-based models require that hydrologic conditions be specified for the start of each simulation. In this work, differences between initial conditions generated by BC Hydro's accepted parameter set and the Pareto set of SCE-UA calibrations are found to be insignificant. This finding implies that an approximation of the initial conditions is sufficient when modelling major storm events for the subject watershed.

A series of large historical events are used to check performance of the Pareto set under event-based conditions. Although simulation quality varies (i.e., in comparison with observed data), all SCE-UA-calibrated parameter sets consistently underestimate the peak flows predicted by BC Hydro's expert calibration. Relative variability across the Pareto set is fairly constant. Time to peak is estimated with a high degree of precision, while event volumes have coefficients of variation in the four to five percent range.

Runoff volumes predicted by the Pareto set of PMF-based extreme event simulations for the Coquitlam Lake watershed are relatively less variable than the simulations of observed storm events (CoV for event volume = 1%). The CoV for the Pareto set of peak flow estimates is 2%, increasing to 4% when the BC Hydro expert calibration is included. Variability across the Pareto set is small in comparison with storm magnitudes. This implies that any uncertainty associated with different calibration objectives does not translate into large variability in estimates of extreme events on the scale of the PMF.

The limited variability in event volumes observed for the PMF-based extreme event simulations is partially explained by the deterministic nature of PMP input. Since a PMP is a deterministically-derived distributed rain event, input and precipitation parameters (e.g., lapse rates, gradient elevation breakpoints, and station adjustment factors) are adjusted to ensure the correct areal distribution of precipitation quantities and intensities. Thus, each trial receives exactly the same amount and distribution of precipitation in time and space.

To better evaluate the variability associated with the Pareto set of calibrations, each Pareto parameter set is applied to the PMP input with calibrated rather than calculated precipitation distribution parameters (i.e., using calibrated lapse rates, gradient elevation breakpoints, and adjustment factors). Using the calibrated Pareto parameter set increases the CoV for event volume from 1% to 4%, while the CoV for peak flow remains unchanged at 2%.

Overall, the results for the extreme event simulations indicate that alternative "acceptable" calibrations incorporating significant parameter variability do not necessarily introduce large variability into an extreme event forecast. Perhaps more importantly, one must consider the significance of a CoV of less than 5% in the context of other uncertainties surrounding the inputs and driving processes in a PMF-scale flood event, which can easily exceed 30%.

The results of this work support the hypothesis that data and model uncertainty are likely to overshadow parameter uncertainty in any computer-based simulation of extreme events such as the PMF, and that effort is best directed toward addressing those factors. Key conclusions include:

- Caution should be used to avoid mathematical overparameterization of precipitation distribution in the UBCWM.
- Objective functions EOPT!, D!, and HMLE are less reliable overall for automatic calibration than SLS, E!, or either of the two variants of EOPT! considered herein.

- Automatic calibration tools can be used to assess an approximate upper bound on the simulation quality attainable with a given model applied to a specific watershed.
- Considering alternative but statistically comparable calibrations introduces limited variability into extreme event simulations. Data and model uncertainty are likely to have a greater effect on extreme event simulations using the UBCWM, and are therefore better targets for further research.

7. Future Directions

To realize the full potential of a discovery will often require as much creativity as went into its creation.
- after Donald Braben

This work has identified several opportunities for further work to extend and significantly enhance the exploration of uncertainty in hydrologic modelling.

Firstly, one must examine on a philosophical and policy level what degree of certainty is ultimately required for each hydrologic simulation. The answer will likely change in different situations. To even explore each problem requires an understanding of the source of relevant hydrologic uncertainties and how they can affect decisions. To this end, one must ensure that the quest to understand and reduce model predictive uncertainty in hydrologic models is appropriately balanced by work in other areas that also feed the decision process (e.g., concepts such as flood protection and acceptable risk).

Risk-based approaches offer a strong means of synthesis for these diverse fields, often providing a common basis against which decision-makers can assess results. Stochastic models are beginning to gain recognition as one such tool. Ultimately, hydrographs will likely need to be presented in a multi-dimensional format, demonstrating risk by pairing probability with each flow estimate. Further work towards risk-based rainfall-runoff modelling will no doubt be necessary.

The various technologies of automatic calibration can also be used to assist in exploring and defining uncertainty. Models other than the UBCWM can easily be considered within the experimental approach outlined in Chapter 4. This would allow comparison of the relative variability associated with calibration for different models, alternative model structures, and various predictive contexts. Similarly, one could relatively quickly and easily use automatic calibration to evaluate the potential capabilities of a focus group of models for a given task. This would lead to flow forecasting based on more robust ensemble simulations from a variety of models rather than the single-model typical *status quo*. Modellers and decision-makers would be able to consider the benefits of increased reliability resulting from of a forecast or prediction based on a suite of models, and weigh them against the risks and cost savings associated with not considering alternative simulations. The emergence of divergent solutions from different models – provided the differences could not be reconciled through alternative calibrations – would be an indicator that a much deeper analysis is required.

Specific to the process herein, relational constraints (e.g. P0UGTK $\leq$P0DZTK) could be investigated as a means for improving automatic calibration of the UBCWM. More generally and more importantly, the work herein suggests that the main thrust of future research using the UBCWM should be directed at reducing data and model uncertainty rather than parameter uncertainty.

The insights gained from exploring the approach detailed in Chapter 4 highlight the potential for a more comprehensive approach to estimating model uncertainty for a hydrologic model. The larger framework given below has been formulated as an illustrative example of the potential for using this line of research to investigate a more complete representation of model predictive uncertainty:

(i) Apply a proven and intensive physical model to a small, hypothetical, hydrologically simple virtual catchment;

(ii) Assume the physical model is sufficiently accurate as to represent observations that would be obtained from a prototype or physical model of the virtual catchment;

(iii) Apply a conceptual model to the virtual catchment;

(iv) Calibrate the conceptual model against the physical model's "reality" using a multi-objective application of SCEM-UA;

(v) Validate the model results as a set based on hydrologically varied data;

(vi) Apply an extreme event of PMF-scale to both models under the assumption that the physical model's representation of active processes is correct; and finally,

(vii) Analyze results with emphasis on the impact and identification of uncertainty in the model structure.

8. Glossary of Acronyms and Watershed Models

ANN	Artificial Neural Network
ASCII	American Standard Code for Information Interchange
CATPRO	A lumped catchment variable source area hydrologic model developed by Kuczera et al. (1993)
CCE	Competitive Complex Evolution algorithm used in the SCE-UA method
cdf	Cumulative Distribution Function
CoV	Coefficient of Variation
D!	Coefficient of Determination, more commonly denoted R^2 (Eq. 1)
DCP	Data Collection Platform
E!	Nash-Sutcliffe efficiency statistic (Eq. 4)
EOPT	Modified version of EOPT! (Eq. 10)
EOPT!	A variant of the Nash-Sutcliffe efficiency statistic that accounts for total volume error (Eq. 7)
EOPT'	Modified version of EOPT! (Eq. 9)
ET	Evapotranspiration
FFA	Flood Frequency Analysis
GA	Genetic Algorithm
GLUE	Generalized Likelihood Uncertainty Estimation procedure of Beven and Binley (1992)
GOM	Global Optimization Method
GSA	Generalized Sensitivity Analysis
HBV	A semi-distributed conceptual hydrologic model developed by Bergström (1972) at the Swedish Meteorological and Hydrological Institute. HBV is the acronym for Hydrologiska Byråns Vattenbalansavdeling, the Institute's Hydrological Bureau Water Balance Section.
HMLE	Heteroscedastic Maximum Likelihood Estimator (Eq. 2)

HOF	Horton (also known as infiltration excess) Overland Flow
IHDM	The Institute of Hydrology Distributed Model, a hydrologic model developed by Beven, Calder and Morris (1987) at the U.K. Institute of Hydrology
IIDRV	Independent and Identically-Distributed Random Variable
KINEROS	KINematic runoff and EROSion model, a hydrologic model developed by Woolhiser, Smith and Goodrich (1990) at the U.S. Department of Agriculture.
KINEROSR	A research version of the KINEROS model adapted by Goodrich (1990)
LAD	Least Absolute Difference statistic (Eq. 3)
MACS	Multistep Automatic Calibration Scheme
MCMC	Monte Carlo Markov Chain
MCS	Monte Carlo Simulation
MFOSM	Mean-Value First-Order Second-Moment
MIKE SHE	A version of the SHE hydrologic model distributed by DHI Software as one of their MIKE series of programs, described by Refsgaard and Storm (1995)
MOCOM-UA	Multi-Objective COMplex evolution algorithm developed by Gupta et al. (1998)
NAM	The Nedbor-Afstromnings Model, a conceptual hydrologic model developed at the Technical University of Denmark (DHI, 1982)
pdf	Probability Density Function
PMF	Probable Maximum Flood
PMP	Probable Maximum Precipitation
R^2	Coefficient of Determination, sometimes denoted D! (Eq. 1)
RMSE	Root Mean Square Error (Eq. 5)
SAC-SMA	The Sacramento Catchment Model, a lumped, conceptual hydrologic model developed by Burnash, Ferral, and McGuire (1973) at the California-Nevada River Forecast Centre in Sacramento, CA.
SCEM-UA	The Shuffled Complex Evolution Metropolis algorithm, an MCMC-based variant of the SCE-UA method developed by Vrugt et al. (2003) at the University of Arizona

SCE-UA	The Shuffled Complex Evolution algorithm, an optimization method for hydrologic model calibration developed by Duan et al. (1994b) at the University of Arizona
SHE	Système Hydrologique Européen, a hydrologic modelling system initially developed by the Danish Hydraulic Institute, the U.K. Institute of Hydrology, and SOGREAH in France and described by Abbott et al. (1986a,b)
SIXPAR	A simplified six-parameter research version of the SAC-SMA model, introduced by Duan et al. (1992)
SLS	Simple Least Squares statistic (Eq. 6)
SMA-NWSRFS	Soil Moisture Accounting model of the National Weather Service River Forecast System described by Peck (1976)
SMD	Soil Moisture Deficit
SOF	Saturated Overland Flow
SQP	Sequential Quadratic Programming
SSARR	Streamflow Synthesis and Reservoir Regulation Model, a hydrologic model first developed by the U.S. Army Corps of Engineers and described by Rockwood (1982)
SWM	The Stanford Watershed Model IV developed by Crawford and Linsley (1966)
TANK	An early, simple hydrologic model described in detail by Sugawara (1995)
TDR	Time-Domain Reflectometry
THALES	A hydrologic model developed by Grayson, Moore, and McMahon (1992) and named after the Greek philosopher Thales of Miletos
TOPMODEL	A hydrologic model described by Beven et al. (p. 627, 1995) as "a set of conceptual tools that can be used to reproduce the hydrological behaviour of catchments in a distributed or semi-distributed way"
TRANSEP	Transfer Function Hydrograph Separation Model, developed by Weiler et al. (2003), which provides coupled but constrained representations of transport and hydraulic transfer functions.
UBCWM	The University of British Columbia Watershed Model developed by Quick et al. (2003)
VB	Microsoft Visual Basic

9. References

Abbott, M.B., J.C. Bathurst, J.A. Cunge, P.E. O'Connell, and J. Rasmussen, 1986a. "An introduction to the European Hydrological System - Système Hydrologique Européen, "SHE", 1: History and Philosophy of a Physically-Based, Distributed Modelling System." *Journal of Hydrology*, **87**: 45-59.

Abbott, M.B., Bathurst, J.C., Cunge, J.A., O'Connell, P.E. and Rasmussen, J., 1986b. "An Introduction to the European Hydrological System - Système Hydrologique Européen, "SHE", 2: Structure of a Physically-Based, Distributed Modelling System." *Journal of Hydrology*, **87**: 61-77.

Allen, R.G., L.S. Pereira, D. Raes and M. Smith, 1998. Crop evapotranspiration - Guidelines for computing crop water requirements. Food and Agriculture Organization (FAO) of the United Nations. FAO Irrigation and Drainage Paper No. 56. 290 pp.

AMS (American Meteorological Society), 2000. *Glossary of Meteorology*. Accessed on 2 December 2003. http://amsglossary.allenpress.com/glossary.

Arnaud, P., C. Bouvier, L. Cisneros and R. Dominguez, 2002. "Influence of rainfall spatial variability on flood prediction." *Journal of Hydrology*, **260**: 216-230.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000a. "Artificial neural networks in hydrology. I: Preliminary concepts." *Journal of Hydrologic Engineering*, **5**(2): 115-123.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000b. "Artificial neural networks in hydrology. II: Hydrologic applications." *Journal of Hydrologic Engineering*, **5**(2): 124-137.

Bashford, K.E., K.J. Beven and P.C. Young, 2002. "Observational data and scale-dependent parameterizations: explorations using a virtual hydrological reality." *Hydrological Processes*, **16**: 293-312.

Bastidas, L.A., P. Yapo, H. Gupta and S. Sorooshian, 2001. The Multi-Objective Complex Evolution Algorithm MOCOM-UA, Ver. 1.3. Software available through Department of Hydrology and Water Resources, University of Arizona.

BC Hydro, 2004. Coquitlam Lake Watershed Probable Maximum Flood (2003). Unpublished report prepared by BC Hydro Engineering. Report No. E196.

Beck, M.B., 1987. "Water Quality Modeling: A Review of the Analysis of Uncertainty." *Water Resources Research*, **23**(8): 1393-1442.

Berga, L., 1998. "New trends in hydrological safety." *In:* Dam Safety. L. Berga, Ed. A. A. Balkema. Rotterdam, Netherlands. 1099-1106.

Bergström, S., 1972. "Utveckling och tillämpning av en digital avrinningsmodell. (Development and application of a digital runoff model)." SMHI. Notiser och preliminära rapporter, serie HYDROLOGI, No. 22, Norrköping. In Swedish.

Bergström, S., G. Lindström and A. Pettersson, 2002. "Multi-variable parameter estimation to increase confidence in hydrological modelling." *Hydrological Processes*, **16**: 413-421.

Beven, K. J., 1989. "Changing Ideas in Hydrology - The Case of Physically-Based Models." *Journal of Hydrology*, **105**: 157-172.

Beven, K. J., 2000. "Uniqueness of place and process representations in hydrological modelling." *Hydrology and Earth System Sciences*, **4**(2): 203-213.

Beven, K. J., 2001. "How far can we go in distributed hydrological modelling?" *Hydrology and Earth System Sciences*, **5**(1): 1-12.

Beven, K. J., 2002. "Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system." *Hydrological Processes*, **16**: 189-206.

Beven, K. J. and A.M. Binley, 1992. "The future of distributed models: model calibration and uncertainty prediction." *Hydrological Processes*, **6**(3): 279-298.

Beven, K.J., A. Calver and E. Morris, 1987. The Institute of Hydrology Distributed Model. U.K. Institute of Hydrology Report No. 98. Institute of Hydrology, Wallingford, U.K.

Beven, K.J. and J. Feyen, 2002. "The Future of Distributed Modelling." *Hydrological Processes*, **16**(2): 169-172.

Beven, K.J., R. Lamb, P. Quinn, R. Romanowicz and J. Freer, 1995. "TOPMODEL." *In:* Computer Models of Watershed Hydrology. V.P. Singh, Ed. Water Resources Publications. Highlands Ranch, CO. 627-668.

Bingeman, A.K., 2001. Improving Dam Safety Analysis by Using Physically-Based Techniques to Derive Estimates of Atmospherically Maximum Precipitation. PhD Thesis, Civil Engineering, University of Waterloo, Waterloo, ON. 198 pp.

Binley, A.M., K.J. Beven, A. Calver and L.G. Watts, 1991. "Changing Responses in Hydrology: Assessing the Uncertainty in Physically Based Model Predictions." *Water Resources Research*, **27**(6): 1253-1261.

Blöschl, G., 2001. "Scaling in Hydrology." *Hydrological Processes*, **15**: 709-711.

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

Braben, D.W., 1985. "Innovation and academic research." *Nature*, **316**(6027): 401-402.

Burges, S.J., 2002. "Process Representation, Measurements, Data Quality, and Criteria for Parameter Estimation of Watershed Models." *In:* Calibration of Watershed Models. Q. Duan, H.V. Gupta, S. Sorooshian, A.N. Rousseau and R. Turcotte, Eds. American Geophysical Union. Washington, D.C. 283-299.

Burges, S.J., 2003. Personal communication re: D. Roche literature review to David Roche and Barbara Lence, 3 March 2003. Teleconference.

Burnash, R.J., R.L. Ferral and R.A. McGuire, 1973. "A generalized streamflow simulation system." *In:* Conceptual Modeling for Digital Computers. Technical Report, Joint Federal and State River Forecast Center, U.S. National Weather Service and California Department of Water Resources. Sacramento, CA. 204pp.

Caissie, D. and N. El-Jabi, 2003. "Instream Flow Assessment: From Holistic Approaches to Habitat Modelling." *Canadian Water Resources Journal*, **28**(2): 173-183.

CDA (Canadian Dam Association), 1999. Dam Safety Guidelines. Canadian Dam Association. 118 pp.

Collier, C.G. and P.R. Larke, 1978. "A case study of the measurement of snowfall by radar: an assessment of accuracy." *Quarterly Journal of the Royal Meteorological Society*, **104**: 615-621.

Columbia Mountains Institute, 2003. *About the Illecillewaet and Beaver watersheds.* Accessed on 20 February 2005. http://www.cmiae.org/mapping/about.htm.

Cooper, V.A., V.T.V. Nguyen and J.A. Nicell, 1997. "Evaluation of Global Optimization Methods for Conceptual Rainfall-Runoff Model Calibration." *Water Science and Technology*, **36**(5): 53-60.

Córdova, J.R. and I. Rodríguez-Iturbe, 1983. "Geomorphoclimatic estimation of extreme flow probabilities." *Journal of Hydrology*, **65**: 159-173.

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

DHI (Danish Hydraulic Institute), 1982. NAM model documentation. *82-892,* JCR/Skn. Updated versions available from DHI.

Dantzig, G.B., 1951. "II: The programming of interdependent activities: mathematical model." *In:* Activity Analysis of Production and Allocation. T.C. Koopmans, Ed. John Wiley & Sons Inc. New York. 19-32.

Duan, Q., S. Sorooshian and V. Gupta, 1992. "Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models." *Water Resources Research*, **28**(4): 1015-1031.

Duan, Q., S. Sorooshian and V. Gupta, 1994a. Shuffled Complex Evolution (SCE-UA) Method, Ver. 2.2. Software available through Department of Hydrology and Water Resources, University of Arizona.

Duan, Q., S. Sorooshian and V.K. Gupta, 1994b. "Optimal use of the SCE-UA global optimization method for calibration watershed models." *Journal of Hydrology*, **158**: 265-284.

Duchon, C.E. and G.R. Essenberg, 2001. "Comparative rainfall observations from pit and aboveground rain gauges with and without wind shields." *Water Resources Research*, **37**(12): 3253-3263.

Dumont, J. and R. Dubé, 2003. "Design of the East Clayton Stormwater Infiltration Systems." Presented at *Water Stewardship: How Are We Managing?* Canadian Water Resources Association 51st Annual Conference. June 11-13, 2003. Vancouver, BC. CWRA.

Dunne, T., 1982. "Models of Runoff Processes and Their Significance." *In:* Scientific Basis of Water Resource Management. National Research Council. National Academy Press. Washington, D.C. 17-30.

Eaton, B., M. Church and D. Ham, 2002. "Scaling and Regionalization of flood flows in British Columbia, Canada." *Hydrological Processes*, **16**(16): 3245-3263.

Environment Canada, 2001. *HYDAT National Water Data Archive.* Water Survey of Canada, Meteorological Service of Canada. Accessed February 20, 2005. http://www.wsc.ec.gc.ca/hydat/H2O/.

Environment Canada, 2005. *National Climate Data and Information Archive: Climate Data Online.* Environment Canada. Accessed February 20, 2005. http://www.climate.weatheroffice.ec.gc.ca/Welcome_e.html.

Faulkner, E., 2003. "Establishing Inflow Design Floods For High-Hazard Potential Dams: Current Practices and Future Prospects." Presented at *The Design Flood: Toward a Common Rationale?* July 29, 2003. Buffalo, NY. HCI Publications.

Faulkner, E.B., L.D. Morland, R. Juyal and J.W. Gotzmer, 1998. *Calculating the Probable Maximum Flood: How Well Do the Methods Work?* Accessed on November 4, 2001. Last Updated on September, 1998. www.meadhunt.com/news/flint.htm.

Faurès, J.-M., D.C. Goodrich, D.A. Woolhiser and S. Sorooshian, 1995. "Impact of small-scale spatial rainfall variability on runoff modeling." *Journal of Hydrology*, **173**: 309-326.

Fetter, C.W., 1994. Applied Hydrogeology. 3rd Edition. Prentice-Hall. Toronto, ON. 691 pp.

Fiering, M.B., 1976. "Reservoir planning and operation." *In:* Stochastic Approaches to Water Resources. H.W. Shen, Ed. Hsieh Wen Shen. Fort Collins, CO. **2:** 17.1-17.21.

Franchini, M. and G. Galeati, 1997. "Comparing several genetic algorithm schemes for the calibration of conceptual rainfall-runoff models." *Hydrological Sciences Journal*, **42**(3): 357-379.

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

Franchini, M. and M. Pacciani, 1991. "Comparative Analysis of Several Conceptual Rainfall-Runoff Models." *Journal of Hydrology*, **122**: 161-219.

Freeze, R.A. and R.L. Harlan, 1969. "Blueprint for a physically-based, digitally-simulated hydrologic response model." *Journal of Hydrology*, **9**: 237-258.

Gan, T.Y., 1987. Applications of Scientific Modeling of Hydrologic Responses From Hypothetical Small Catchments to Assess a Complex Conceptual Rainfall-Runoff Model. PhD Thesis, Civil Engineering, University of Washington, Seattle, WA. 259 pp.

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

Gan, T.Y. and S.J. Burges, 1990a. "An Assessment of a Conceptual Rainfall-Runoff Model's Ability to Represent the Dynamics of Small Hypothetical Catchments: 1. Models, Model Properties, and Experimental Design." *Water Resources Research*, **26**(7): 1595-1604.

Gan, T.Y. and S.J. Burges, 1990b. "An Assessment of a Conceptual Rainfall-Runoff Model's Ability to Represent the Dynamics of Small Hypothetical Catchments: 2. Hydrologic Responses for Normal and Extreme Rainfall." *Water Resources Research*, **26**(7): 1605-1619.

Gan, T.Y., E.M. Dlamini and G.F. Biftu, 1997. "Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling." *Journal of Hydrology*, **192**: 81-103.

Garen, D.C. and S.J. Burges, 1981. "Approximate Error Bounds for Simulated Hydrographs." *Journal of the Hydraulics Division, ASCE*, **107**(HY11): 1519-1534.

Goldberg, D.E., 1989. Genetic algorithms in search, optimization and machine learning. Addison-Wesley. Reading, MA. 412 pp.

Goodrich, D.C., 1990. Geometric simplification of a distributed rainfall-runoff model over a range of basin scales. PhD Thesis, Hydrology and Water Resources, University of Arizona, Tucson, AZ. 361 pp.

Goodrich, D.C. and D.A. Woolhiser, 1994. "Comment on "Physically based hydrologic modelling, 1, A terrain-based model for investigative purposes" by R.B. Grayson, I.D. Moore, and T.A. McMahon." *Water Resources Research*, **30**(3): 845-847.

Graham, W.J., 2000. "Should dams be modified for the probable maximum flood?" *Journal of the American Water Resources Association*, **36**(5): 953-963.

Grayson, R.B., I.D. Moore and T.A. McMahon, 1992a. "Physically Based Hydrologic Modeling: 1. A Terrain-Based Model for Investigative Purposes." *Water Resources Research*, **28**(10): 2639-2658.

Grayson, R.B., I.D. Moore and T.A. McMahon, 1992b. "Physically Based Hydrologic Modelling: 2. Is the Concept Realistic?" *Water Resources Research*, **26**(10): 2659-2666.

Grayson, R.B., I.D. Moore and T.A. McMahon, 1994a. "Reply." *Water Resources Research*, **30**(3): 849.

Grayson, R.B., I.D. Moore and T.A. McMahon, 1994b. "Reply." *Water Resources Research*, **30**(3): 855-856.

Gregory, R., P. Slovic and J. Flynn, 1996. "Risk perceptions, stigma, and health policy." *Health & Place*, **2**(4): 213-220.

Grigg, N.S., N.H. Doesken, D.M. Frick, M. Grimm, M. Hilmes, T.B. McKee and K.A. Oltjenbrums, 1999. "Fort Collins Flood 1997: Comprehensive View of An Extreme Event." *Journal of Water Resources Planning and Management*, **125**(5): 355-362.

Gupta, H.V., S. Sorooshian and P.O. Yapo, 1998. "Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information." *Water Resources Research*, **34**(4): 751-763.

Gupta, H.V., S. Sorooshian and P.O. Yapo, 1999. "Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration." *Journal of Hydrologic Engineering*, **4**(2): 135-143.

Gupta, V.K. and S. Sorooshian, 1983. "Uniqueness and Observability of Conceptual Rainfall-Runoff Model Parameters: The Percolation Process Examined." *Water Resources Research*, **19**(1): 269-276.

E. M. Hansen, E.M., D. D. Fenn, L. C. Schreiner, R. W. Stodt and J. F. Miller, 1988. Probable Maximum Precipitation Estimates - United States Between the Continental Divide and the 103rd Meridian. Report No. HMR-55A. National Weather Service, National Oceanic and Atmospheric Administration, US Department of Commerce. Silver Spring, MD. 242pp.

Harlin, J., 1992. "Modelling the Hydrological Response of Extreme Floods in Sweden." *Nordic Hydrology*, **23**: 227-233.

Harlin, J. and C.-S. Kung, 1992. "Parameter uncertainty and simulation of design floods in Sweden." *Journal of Hydrology*, **137**: 209-230.

Hartford, D.N.D., A.H. Estergaard and J.G. Lang, 2001. "Towards "Values-Based" Selection of Inflow Design Floods." Presented at *Canadian Dam Association 2001 Annual Conference*, September 30 - October 4, 2001. Fredericton, New Brunswick. Canadian Dam Association.

Hastings, W.K., 1970. "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika*, **57**: 97-109.

Hogue, T.S., S. Sorooshian, H. Gupta, A. Holz and D. Braatz, 2000. "A Multistep Automatic Calibration Scheme for River Forecasting Models." *Journal of Hydrometeorology*, **1**(6): 524-542.

Hooke, R. and T.A. Jeeves, 1961. " "Direct search" solution of numerical and statistical problems." *Journal of the ACM*, **8**: 212-229.

Hornberger, G.M., K.J. Beven, B.J. Cosby and D.E. Sappington, 1985. "Shenandoah Watershed Study: Calibration of a topography-based, variable contributing area hydrological model to a small forested catchment." *Water Resources Research*, **21**(12): 1841-1850.

Hornberger, G.M. and E.W. Boyer, 1995. "Recent advances in watershed modelling." U.S. National Report to IUGG, 1991-1994. *Reviews of Geophysics, July 1995 Supplement*: 949-957.

Horton, G.A., 1999. *Water Words Dictionary.* Accessed on 2 December 2003. Last Updated on 7 August 2000. http://water.nv.gov/Water%20planning/dict-1/ww-index.htm.

Houghton-Carr, H.A., 1999. "Assessment criteria for simple conceptual daily rainfall-runoff models." *Hydrological Sciences Journal*, **44**(2): 237-261.

HSE (Health and Safety Executive, U.K., 2001. Reducing Risks, Protecting People. Risk Assessment Policy Unit. Health and Safety Executive.

Hunter, F.G., D.B. Donald, B.N. Johnson, W.D. Hyde, J.M. Hanesiak, M.O.B. Kellerhals, R.F. Hopkinson and B.W. Oegema, 2002. "The Vanguard Torrential Storm (Meteorology and Hydrology)." *Canadian Water Resources Journal*, **27**(2): 213-227.

Hunter, S.M., E.W. Holroyd III, and C.L. Hartzell, 2001. "Improvements to the WSR-88D Snow Accumulation Algorithm." *Preprints, 30th International Conference on Radar Meteorology*. July 19-24, 2001. Munich, Germany. American Meteorological Society: 716-718.

IAHS (International Association of Hydrological Sciences), 2005. Freshwater: Sustainability within Uncertainty. VIIth IAHS Scientific Assembly. April 3-9, 2005. Foz do Iguaçu, Brazil.

Imbeau, M.E., 1892. "La Durance: Regime, crues et inundations." *Ann. Ponts Chaussees*, **3**(1): 5-18. In French.

Jakeman, A.J. and G.M. Hornberger, 1993. "How Much Complexity is Warranted in a Rainfall-Runoff Model?" *Water Resources Research*, **29**(8): 2637-2649.

Jakob, M. and P. Jordan, 2001. "Design flood estimates in mountain streams - the need for a geomorphic approach." *Canadian Journal of Civil Engineering*, **28**: 425-439.

Jarrett, R.D., 1990. "Paleohydrologic techniques used to define the spatial occurrence of floods." *Geomorphology*, **3**: 181-195.

Jensen, M.E., R.D. Burman and R.G. Allen, Eds. 1990. Evaporation and irrigation water requirements. ASCE Manuals and Reports on Engineering Practice. American Society of Civil Engineers, New York, NY. 332 pp.

Keeney, R.L. and D. von Winterfeldt, 1989. "On the Uses of Expert Judgment on Complex Technical Problems." *IEEE Transactions on Engineering Management*, **36**(2): 83-86.

Kite, G., 1993. "Computerized streamflow measurement using slug injection." *Hydrological Processes*, **7**(2): 227-233.

Kirkpatrick, S., C.D. Gelatt Jr. and M.P. Vecchi, 1983. "Optimization by simulated annealing." *Science*, **220**(4598): 671-680.

Klemeš, V., 1982. "Empirical and Causal Models in Hydrology." *In:* Scientific Basis of Water Resource Management. National Academy Press. Washington, D.C.: 95-104.

Klemeš, V., 1983. "Conceptualization and Scale in Hydrology." *Journal of Hydrology*, **65**: 1-23.

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

Klemeš, V., 1986b. "Operational testing of hydrological simulation models." *Hydrological Sciences Journal*, **31**(1): 13-24.

Klemeš, V., 1992. Probability of PMP for Coquitlam Lake: An Exploratory Analysis. Unpublished report prepared for BC Hydro. Report No. N763.

Klemeš, V., 2000a. "Of Carts and Horses in Hydrologic Modelling." *In:* Common Sense and Other Heresies: Selected Papers on Hydrology and Water Resources Engineering by Vit Klemeš. C.D. Sellars, Ed. Canadian Water Resources Association. Waterloo, ON.: 53-68.

Klemeš, V., 2000b. "Tall tales about tails of hydrological distributions, Parts I and II." *Journal of Hydrologic Engineering*, **4**(3): 227-239.

Konrad II, C.E., 2001. "The Most Extreme Precipitation Events over the Eastern United States from 1950 to 1996: Considerations of Scale." *Journal of Hydrometeorology*, **2**(3): 309-325.

Kouwen, N., 2003. Personal Communication re: Seeking Example Extreme Flood Events to Barbara Lence, 27 January 2003. Electronic Mail.

Krajewski, W.F., 2005. Personal Communication re: Snow and weather rata to Markus Weiler, 18 April 2005. Electronic Mail.

Kroeker, M., D. Hartford and G. Lang, 2003. Personal communication re: D. Roche MASc project to David Roche and Barbara Lence, 5 June 2003. Private meeting at BC Hydro (Edmonds), Burnaby, BC.

Kuchment, L.S. and A.N. Gelfan, 2002. "Estimation of Extreme Flood Characteristics Using Physically Based Models of Runoff Generation and Stochastic Meteorological Inputs." *Water International*, **27**(1): 77-86.

Kuczera, G., 1997. "Efficient subspace probabilistic parameter optimization for catchment models." *Water Resources Research*, **33**(1): 177-185.

Kuczera, G. and E. Parent, 1998. "Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm." *Journal of Hydrology*, **211**: 69-85.

Kuczera, G., G. P. Raper, N. S. Brah, and M. D. A. Jayasuriya, 1993. "Modelling yield changes following strip thinning in a mountain ash catchment: An exercise in catchment model validation." *Journal of Hydrology*, **150**(2-4): 433–457.

Kulkarni, T. and A. Blais-Stevens, 2003. *Community Vulnerability to Flooding and Climate Change*. Joint Workshop between C-CIARN Landscape Hazards and CWRA BC Branch, 5 December 2003. Vancouver, BC. C-CIARN.

Lan, Y.-H., 2001. <u>A Genetic-Algorithm Based Automatic Model Calibrator for the UBC Watershed Model</u>. MASc Thesis, Civil Engineering, University of British Columbia, Vancouver, BC. 117 pp.

Larson, L.W. and E.L. Peck, 1974. "Accuracy of Precipitation Measurements for Hydrologic Modeling." *Water Resources Research*, **10**(4): 857-863.

Lei, J.H. and W. Schilling, 1996. "Preliminary Uncertainty Analysis - A prerequisite for Assessing the Predictive Uncertainty of Hydrologic Models." *Water Science and Technology*, **33**(2): 79-90.

Linsley, R.K., J.B. Franzini, D.L. Freyberg and G. Tchobanoglous, 1992. <u>Water Resources Engineering</u>. 4th Edition. McGraw-Hill. San Francisco, CA. 841 pp.

Loague, K.M. and R.A. Freeze, 1985. "A Comparison of Rainfall-Runoff Modeling Techniques on Small Upland Catchments." *Water Resources Research*, **21**(2): 229-248.

Lohani, B.N., J.W. Evans, H. Ludwig, R.R. Everitt, R.A. Carpenter and S.L. Tu, 1997. "Risk and Uncertainty in Environmental Impact Assessment." *In:* <u>Environmental Impact</u>

Assessment for Developing Countries in Asia -- Vol. 1 -- Overview. Asian Development Bank. Manila, Philippines. **1:** 5.1-5.45.

Lumb, A.M. and R.K. Linsley, 1971. "Hydrologic Consequences of Rainfall Augmentation." *Journal of the Hydraulics Division, ASCE*, **97**(HY 7): 1065-1080.

Madsen, H., 2000. "Automatic calibration of a conceptual rainfall-runoff model using multiple objectives." *Journal of Hydrology*, **235**: 276-288.

Madsen, H., 2003. "Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives." *Advances in Water Resources*, **26**: 205-216.

Madsen, H., G. Wilson and H.C. Ammentorp, 2002. "Comparison of different automated strategies for calibration of rainfall-runoff models." *Journal of Hydrology*, **261**: 48-59.

Maidment, D.R. (Ed.), 1993. Handbook of Hydrology. McGraw-Hill. New York, NY. 1424 pp.

Masri, S.F., G.A. Bekey and F.B. Safford, 1980. "A global optimization algorithm using adaptive random search." *Applied Mathematics and Computation*, **7**: 353-375.

Matthews, H.M., 1999. Snowmelt Forecasting in a Mountainous Watershed. MSc Thesis, Civil Engineering, Colorado State University, Fort Collins, CO. 244 pp.

McCuen, R.H., 1973. "The role of sensitivity analysis in hydrologic modelling." *Journal of Hydrology*, **18**: 37-53.

Melching, C.S., 1995. "Reliability Estimation." *In:* Computer Models of Watershed Hydrology. V.P. Singh, Ed. Water Resources Publications. Highlands Ranch, Colorado. 69-118.

Melching, C.S., B.C. Yen and J. Harry G. Wenzel, 1990. "A reliability Estimation in Modeling Watershed Runoff With Uncertainties." *Water Resources Research*, **26**(10): 2275-2286.

Merz, R. and G. Blöschl, 2004. "Regionalisation of catchment model parameters." *Journal of Hydrology*, **287**: 95-123.

Michaud, J. and S. Sorooshian, 1994. "Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed." *Water Resources Research*, **30**(3): 593-605.

Micovic, Z., 1998. Regional Flow Estimation Using a Hydrologic Model. MASc Thesis, Department of Civil Engineering, University of British Columbia, Vancouver, BC. 106 pp.

Micovic, Z., 2003a. Personal communication re: UBCWM / hydrologic modelling to D. Roche, 24 January 2003. Private meeting at BC Hydro (Edmonds), Burnaby, BC.

Micovic, Z., 2003b. Personal communication to D. Roche, 26 February 2003. Phone conversation.

Monteith, J.L., 1965. "Evaporation and Environment." Presented at *The state and movement of water in living organisms*, 8-12 September, 1964. Swansea, UK. Cambridge University Press.

Moore, R.D., 2004. "Introduction to Salt Dilution Gauging for Streamflow Measurement: Part 1." *Streamline Watershed Management Bulletin*, **7**(4): 20-23.

Moore, R.D., 2005. "Introduction to Salt Dilution Gauging for Streamflow Measurement, Part III: Slug Injection Using Salt in Solution." *Streamline Watershed Management Bulletin*, **8**(2): 1-6.

Moore, R.J. and R.T. Clarke, 1981. "A distribution function approach to rainfall runoff modeling." *Water Resources Research*, **17**(5): 1367-1382.

Mulvany, T.J., 1851. "On the use of self-registering rain and flood gauges in making observations of the rainfall and flood discharges in a given catchment." *Transactions of the Institute of Civil Engineers of Ireland*, **4**(2): 18-31.

National Research Council Canada, 1989. Hydrology of Floods in Canada: A Guide to Planning and Design. Associate Committee on Hydrology, National Research Council Canada. 245 pp.

National Weather Service, 2003. *Cooperative Observer Program*. Accessed on 2 December 2003. Last Updated on 19 May 2003. http://www.nws.noaa.gov/om/coop/what-is-coop.html.

Natural Resources Canada, 2003. *Flood Disasters in Canada*. Accessed on 11 December 2003. Last Updated on 27 November 2003. http://sts.gsc.nrcan.gc.ca/clf/geoserv_floods.asp.

Nelder, J.A. and R. Mead, 1965. "A simplex method for function minimization." *Computer Journal*, **7**: 308-313.

Nešpor, V. and B. Sevruk, 1999. "Estimation of wind-induced error of rainfall gauge measurements using a numerical simulation." *Journal of Atmospheric and Oceanic Technology*, **16**(4): 450-464.

NRC (U.S. National Research Council), 1985. Safety of Dams, Flood and Earthquake Criteria. Committee on Safety Criteria for Dams, Water Science and Technology Board, Commission on Engineering and Technical Systems. 294 pp.

NRC (U.S. National Research Council), 1994. Science and Judgment in Risk Assessment. Committee on Risk Assessment of Hazardous Air Pollutants, Board on Environmental Studies and Toxicology and Commission on Life Sciences. National Academy Press. 286 pp.

NRC (U.S. National Research Council), 1995. Flood Risk Management and the American River Basin: An Evaluation. Committee on Flood Control Alternatives in the American River

Basin, Water Science and Technology Board, Commission on Geosciences, Environment and Resources. 235 pp.

NRC (U.S. National Research Council), 2000a. Risk Analysis and Uncertainty in Flood Damage Reduction Studies. Committee on Risk-Based Analysis for Flood Damage Reduction, Water Science and Technology Board, Commission on Geosciences, Environment, and Resources. 202 pp.

NRC (U.S. National Research Council), 2000b. "Risk Analysis Concepts and Terms." *In:* Risk Analysis and Uncertainty in Flood Damage Reduction Studies. National Academy Press. Washington, D.C. 40-50.

Nystuen, J.A., J.R. Proni, P.G. Black and J.C. Wilkerson, 1996. "A comparison of automatic rain gauges." *Journal of Atmospheric and Oceanic Technology*, **13**: 62-73.

O'Connell, P.E. and E. Todini, 1996. "Modelling of rainfall, flow and mass transport in hydrological systems: an overview." *Journal of Hydrology*, **175**: 3-16.

Ogden, F.L., H.O. Sharif, S.U.S. Senarath, J.A. Smith, M.L. Baeck and J.R. Richardson, 2000. "Hydrologic analysis of the Fort Collins, Colorado, flash flood of 1997." *Journal of Hydrology*, **228**(1-2): 92-100.

Oreskes, N., K. Shrader-Frechette and K. Belitz, 1994. "Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences." *Science*, **263**: 641-646.

Peck, E.L., 1976. Catchment modeling and initial parameter estimation for the National Weather System river forecast system. NOAA Tech. Memo NWS HYDRO-31. U.S. Department of Commerce. Silver Spring, MD.

Pike, R.G. and R. Scherer, 2003. "Overview of the potential effects of forest management on low flows in snowmelt-dominated hydrologic regimes." *BC Journal of Ecosystems and Management*, **3**(1): 1-17.

Press, W.H., S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, 1992. Numerical Recipes in Fortran 77: The Art of Scientific Computing. 2nd Edition. Cambridge University Press. New York, NY. 933 pp.

Quick, M.C., 1994. UBC Watershed Model Manual. Edition 3.0. University of British Columbia Mountain Hydrology Group. Vancouver, BC.

Quick, M.C., 1995. "The UBC Watershed Model." *In:* Computer Models of Watershed Hydrology. V.P. Singh, Ed. Water Resources Publications. Highlands Ranch, CO. 233-280.

Quick, M.C., A. Pipes, H. Assaf, E. Yu, D. Nixon, S. Loukas, R. Millar, P. Singh and Z. Micovic, 2003. UBC Watershed Model, Ver. 5 (Windows XP Edition). Software available through UBC Mountain Hydrology Research Group.

Refsgaard, J.C., 1994. "Model and data requirements for simulation of runoff and land surface processes in relation to global circulation models." *In:* Land Surface Processes in Hydrology: Trials and Tribulations of Modeling and Measuring. S. Sorooshian, H.V. Gupta and J.C. Rodda, Eds. Springer-Verlag. New York, NY.

Refsgaard, J.C. and B. Storm, 1995. "MIKE SHE." *In:* Computer Models of Watershed Hydrology. V.P. Singh, Ed. Water Resources Publications. Highlands Ranch, CO. 809-846.

Revelle, C.S., E.E. Whitlatch and J.R. Wright, 1997. Civil and Environmental Systems Engineering. Prentice Hall. Upper Saddle River, NJ. 507 pp.

Rockwood, D.M., 1982. "Theory and Practice of the SSARR Model as Related to Analyzing and Forecasting the Response of Hydrologic Systems." *In:* Applied Modeling in Catchment Hydrology, Proceedings of the International Symposium on Rainfall-Runoff Modeling, held May 1981. Mississippi State University. V.P. Singh, Ed.. 87-106.

Rosenbrock, H.H., 1960. "An automatic method for finding the greatest or least value of a function." *Computer Journal*, **3**: 175-184.

Rosenhagen, G. and M. Halpert, 1998. "Chapter 5: Flooding." *In:* The Global Climate System Review, December 1993 - May 1996. J.M. Nicholls, Ed. World Climate Data and Monitoring Programme, World Meteorological Organization **WMO No. 856:** 40-47.

Seibert, J., 2000. "Multi-criteria calibration of a conceptual runoff model using a genetic algorithm." *Hydrology and Earth System Sciences*, **4**(2): 215-224.

Seibert, J. and J.J. McDonnell, 2002. "On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration." *Water Resources Research*, **38**(11): 1241.

Sevruk, B., 1994. "Spatial and temporal inhomogeneity of global precipitation data." *In:* Global Precipitation and Climate Change. M. Desbois and F. Desalmand, Eds. Springer Verlag, Berlin, Germany. NATO ASI Series, **I 26**: 219-230.

Sevruk, B., 1996. "Adjustment of tipping-bucket precipitation gauge measurements." *Atmospheric Research*, **42**(1-4): 237-246.

Sevruk, B. and V. Nešpor, 1998. "Empirical and theoretical assessment of the wind induced error of rain measurement." *Water Science and Technology*, **37**(11): 171-178.

Singh, V.P., Ed. 1995a. Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, Colorado. 1130 pp.

Singh, V.P., 1995b. "Watershed Modelling." *In:* Computer Models of Watershed Hydrology. V.P. Singh, Ed. Water Resources Publications. Highlands Ranch, Colorado. 1-22.

Singh, V.P. and D.A. Woolhiser, 2002. "Mathematical Modeling of Watershed Hydrology." *Journal of Hydrologic Engineering*, **7**(4): 270-292.

Slovic, P., 1987. "Perception of Risk." *Science*, **236**(4799): 280-285.

Slovic, P., 1992. "Perception of Risk: Reflections on the Psychometric Paradigm." *In:* Social theories of risk. S. Krimsky and D. Golding, Eds. Praeger Publishers. Westport, Conn. 117-152.

Smith, J.A., M.L. Baeck, J.E. Morrison and P. Sturdevant-Rees, 2000. "Catastrophic Rainfall and Flooding in Texas." *Journal of Hydrometeorology*, **1**(1): 5-25.

Smith, J.A., M.L. Baeck, M. Steiner and A.J. Miller, 1996. "Catastrophic rainfall from an upslope thunderstorm in the central Appalachians: The Rapidan storm of June 27, 1995." *Water Resources Research*, **32**(10): 3099-3113.

Smith, R.E., D.R. Goodrich, D.A. Woolhiser and J.R. Simanton, 1994. "Comment on "Physically based hydrologic modeling, 2, Is the concept realistic?" by R. B. Grayson, I. D. Moore, and T. A. McMahon." *Water Resources Research*, **30**(3): 851-854.

Song, Q. and L.C. Brown, 1990. "DO model uncertainty with correlated inputs." *Journal of Environmental Engineering*, **116**(6): 1164-1180.

Song, Z. and L.D. James, 1992. "An Objective Test for Hydrologic Scale." *Water Resources Bulletin*, **28**(5): 833-844.

Sorooshian, S., Q. Duan and V.K. Gupta, 1993. "Calibration of Rainfall-Runoff Models: Application of Global Optimization to the Sacramento Soil Moisture Accounting Model." *Water Resources Research*, **29**(4): 1185-1194.

Sorooshian, S. and V.K. Gupta, 1983. "Automatic Calibration of Conceptual Rainfall-Runoff Models: The Question of Parameter Observability and Uniqueness." *Water Resources Research*, **19**(1): 260-268.

Sorooshian, S. and V.K. Gupta, 1985. "The Analysis of Structural Identifiability: Theory and Application to Conceptual Rainfall-Runoff Models." *Water Resources Research*, **21**(4): 487-495.

Sorooshian, S. and V.K. Gupta, 1995. "Model Calibration." *In:* Computer Models of Watershed Hydrology. V.P. Singh, Ed. Water Resources Publications. Highlands Ranch, Colorado. 23-68.

Sorooshian, S., V.K. Gupta and J.L. Fulton, 1983. "Evaluation of Maximum Likelihood Parameter Estimation Techniques for Conceptual Rainfall-Runoff Models: Influence of Calibration Data Variability and Length on Model Credibility." *Water Resources Research*, **19**(1): 251-259.

Spear, R.C. and G.M. Hornberger, 1980. "Eutrophication in Peel Inlet - 2: Identification of critical uncertainties via Generalized Sensitivity Analysis." *Water Research*, **14**: 43-49.

Steiner, M., J.A. Smith, S.J. Burges, C.V. Alonso and R.W. Darden, 1999. "Effect of bias adjustment and rain gauge data quality control on radar rainfall estimation." *Water Resources Research*, **35**(8): 2487-2503.

Sugawara, M., 1995. "TANK Model." *In:* Computer Models of Watershed Hydrology. V.P. Singh, Ed. Water Resources Publications. Highlands Ranch, CO. 165-214.

Takyi, A.K., 1991. Uncertainty Analysis for Environmental Management Models: Application of the Generalized Sensitivity Analysis. MSc Thesis, Civil Engineering, University of Manitoba, Winnipeg, MB. 225 pp.

Thapa, K.B. and N.R. Khanal, 2001. *Flood Hazard in Nepal & Need for Flood Forecasting in HKH Region*. Accessed on 27 January 2003. http://www.southasianfloods.org/document/floodhazard/.

Thornthwaite, C.W., 1944. "Report of the Committee on Transpiration and Evaporation." *Transactions of the American Geophysical Union*, **25**(Part V): 683-693.

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

Todini, E., 1988. "Rainfall-runoff modeling - past, present and future." *Journal of Hydrology*, **100**: 341-352.

Troch, P.A., M. Mancini, C. Paniconi and E.F. Wood, 1993. "Evaluation of a Distributed Catchment Scale Water Balance Model." *Water Resources Research*, **29**(6): 1805-1817.

Tung, Y.-K., 1990. "Evaluating the Probability of Violating Dissolved Oxygen Standard." *Ecological Modelling*, **51**: 193-204.

Tversky, A. and D. Kahneman, 1974. "Judgment under uncertainty: Heuristics and Biases." *Science*, **185**: 1124-1131.

USACE (U.S. Army Corps of Engineers), 1992. Guidelines for Risk and Uncertainty Analysis in Water Resources Planning. **Vol I: Principles with Technical Appendices** (IWR Report 92-R-1). Water Resources Support Center. Institute for Water Resources. 173 pp.

van Hofwegen, P. and B. Schultz, 1995. "Floodings in The Netherlands, what happened?" Delft Alumni Newsletter. August 1995 edition. Delft, The Netherlands.

Vicens, G.J., I. Rodriguez-Iturbe and J. John C. Schaake, 1975. "A Bayesian Framework for the Use of Regional Information in Hydrology." *Water Resources Research*, **11**(3): 405-414.

Viessman, W.J. and G.L. Lewis, 1996. Introduction to Hydrology. 4th Edition. HarperCollins College Publishers. New York, NY. 760 pp.

Vrugt, J.A., H.V. Gupta, W. Bouten and S. Sorooshian, 2003. "A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters." *Water Resources Research*, **39**(8): 1214.

Wang, Q.J., F.H.S. Chiew and T.A. McMahon, 1995. "Calibration of environmental models by genetic algorithms." Presented at *MODIM 95: International Congress on Modelling and Simulation*, November 1995. Newcastle, Australia. Modelling and Simulation Society of Australia.

Water Management Consultants, 2003. Analysis of Probable Maximum Precipitation for Southwest British Columbia. Unpublished report prepared for BC Hydro by CDN Water Management Consultants Inc. Report No. 7056/R2.

Weiler, M., 2005. Personal Communication re: thesis revisions to D. Roche and B. Lence, 15 April 2005. Thesis annotations and meeting with B. Lence.

Weiler, M. and J.J. McDonnell, 2004. "Virtual experiments: a new approach for improving process conceptualization in hillslope hydrology." *Journal of Hydrology*, **285**: 3-18.

Weiler, M., J.J. McDonnell, H.J. Tromp-van Meerveld, and T. Uchida, 2005. "Subsurface Stormflow." *In:* Encyclopedia of Hydrological Sciences. Wiley and Sons. *In press.*

Weiler, M., B.L. McGlynn, K.J. McGuire, and J.J. McDonnell, 2003. "How does rainfall become runoff? A combined tracer and runoff transfer function approach." *Water Resources Research*, **39**(11): 1315.

Welles, J.F., 1984. "The survival advantage of stupidity." *Speculations in Science and Technology*, **7**(1): 17-21.

WMO (World Meteorological Organization), 1986. Intercomparison of models of snowmelt runoff. WMO Operational Hydrology Report No. 23. 36 pp.

Wood, E.F., M. Sivapalan, K. Beven and L. Band, 1988. "Effects of Spatial Variability and scale with implications to hydrologic modeling." *Journal of Hydrology*, **102**: 29-47.

Wood, P.J. and A.P. Dykes, 2002. "The use of salt dilution gauging techniques: ecological considerations and insights." *Water Research*, **36**(12): 3054-3062.

Woolhiser, D.A., 1996. "Search for Physically Based Runoff Model - A Hydrologic El Dorado?" *Journal of Hydraulic Engineering*, **122**(3): 122-129.

Woolhiser, D.A., R.E. Smith and D.C. Goodrich, 1990. KINEROS, A Kinematic Runoff and Erosion Model: Documentation and User Manual. Agriculture Research Service, U.S. Department of Agriculture. Report No. ARS-77. 130 pp.

Wu, Y.-H., D.A. Woolhiser and V. Yevjevich, 1982. "Effects of spatial variability of hydraulic resistance of runoff hydrographs." *Journal of Hydrology*, **59**: 231-248.

Xiao, R., V. Chandrasekar, and H. Liu, 1998. "Development of a Neural Network Based Algorithm for Radar Snowfall Estimation." *IEEE Transactions on Geoscience and Remote Sensing*, **36**(3): 716-724.

Yapo, P.O., H.V. Gupta and S. Sorooshian, 1996. "Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data." *Journal of Hydrology*, **181**(1-4): 23-48.

Yapo, P.O., H.V. Gupta and S. Sorooshian, 1998. "Multi-objective global optimization for hydrologic models." *Journal of Hydrology*, **204**: 83-97.

Zhu, D., 1997. *Overview of the UBC Watershed Model.* Accessed on 6 October 2002. Last Updated on 30 January 1997. www.civil.ubc.ca/home/ubcmodel.

Appendix A

Selection of SCE-UA Parameter Values

Appendix A: Selection of SCE-UA Parameter Values

This Appendix briefly describes the selection of parameter values for the SCE-UA algorithm.

Most of the parameters of the SCE-UA algorithm have default values that are defined *a priori* by the developers or by the number of parameters to be calibrated. In general, these parameters are set to their default values. The parameter names and descriptions referred to below are taken from those found in the README.1ST file of the SCE-UA distribution package (Duan et al., 1994a).

The maximum number of trials (MAXN) (i.e., the number of executions of UBCWM) before the algorithm terminates is set at 50,000 for the first set of calibrations. However, some of this first set of calibrations terminate after reaching 50,000 iterations without converging to a solution. Therefore, MAXN is raised to 100,000 for subsequent tests. This adjustment ensures that other termination criteria will tend to dominate the later calibrations. In general, setting MAXN to 100,000 is concluded to be appropriate in the context of this study.

The number of shuffling loops over which the objective function must show improvement (KSTOP) is arbitrarily set to ten. Although the recommended value for KSTOP is five, a conservative value is selected to allow the greatest possible chance for the algorithm to locate an improved solution. The additional computation required is not overly significant when examined in context.

The required percentage improvement (PCENTO) is set to its default value of 0.01%.

There is no default value for the number of complexes to be determined (NGS). Duan (Duan et al., 1994a) recommends a value between two and twenty depending on the complexity of the optimization problem; obviously, a more complex problem is likely to require a higher number of complexes. No explanation is given for the origin of 20 complexes as an upper limit for NGS. Most likely, this value arises either from array-size limitations in the original version of SCE-UA or from undocumented experiments that concluded there was little to be gained by increasing NGS beyond 20.

It is assumed that the UBCWM has a maximum of 28 parameters to be calibrated, as outlined in Table 4-1 (allowing for up to five different precipitation stations). The calibration problem is sufficiently difficult to justify the use of a high value for NGS. However, no guidance is available in the literature as to whether fifteen or even twenty complexes can effectively calibrate a model with 28 parameters. Preliminary tests to determine an appropriate value for NGS were conducted using synthetic calibration. As the discussion in Section 5.1 indicates, a value of 20 was ultimately selected based on the results of the preliminary testing.

ISEED is a random integer used to initialize the internal random number generation algorithm of Press et al. (Press et al., 1992). As a random input to a random number generator, an arbitrarily-selected integer in the range 100 to 9999 should suffice. Each execution of SCE-UA utilizes a different arbitrarily-selected value of ISEED.

The IDEFLT value of zero used for all tests instructs the algorithm to use default values for all parameters specified in the second line of SCEIN_VB.DAT.

The default value for NPG, the number of points assigned to each complex, is set equal to $2n + 1$, where n is the number of parameters to be optimized. The number of points assigned to each sub-complex (NPS) is set to $n + 1$. NSPL , the number of evolution steps taken by each complex between shuffling stages, is set equal to NPG at $2n + 1$.

The parameter MINGS defaults to the value of NGS, implying that the number of complexes created by SCE-UA will not decrease as the calibration progresses.

A default value of zero for INIFLG means that the initial value specified for each parameter in SCEIN_VB.DAT will not be included in the analysis.

Finally, IPRINT, the Boolean variable controlling whether the entire population of points will be output to a text file each time the complexes are shuffled, is set to the default option of zero. This results in SCE-UA recording only the best set of parameter values obtained with each shuffling loop, since recording the entire population at each loop would result in an enormous amount of marginally-significant data.

Appendix B

Detailed Calibration Results

Appendix B.1

Complete Parameter Sets

Final Parameter Sets for Synthetic Calibrations of the Coquitlam Lake Watershed Data Set

OF	C0IMPA	Delta	P0GRADL	P0GRADM	P0GRADU	E0LMID	E0LHI	P0AGEN	V0FLAS	P0PERC	P0DZSH	P0FRTK	P0FSTK	P0IRTK	P0ISTK	P0UGTK	P0DZTK	P0SREP	P0RREP
SLS	0.16	0.049	2.1	1.5	21.0	542	1415	100.1	1.0	13.0	0.57	0.14	0.17	1.00	1.01	21.68	97.70	0.014	0.094
HMLE	0.15	0.050	1.6	5.6	23.7	1298	1473	99.9	1.0	13.0	0.58	0.14	0.17	1.00	1.01	21.00	96.00	0.025	0.105
EOPT'	0.15	0.050	1.6	0.4	20.6	1287	1407	99.9	1.0	13.0	0.58	0.14	0.17	1.00	1.00	21.08	96.10	0.022	0.103
E!	0.16	0.046	1.7	7.0	21.0	1415	1420	100.2	1.0	13.0	0.57	0.14	0.17	1.00	1.02	21.80	97.40	0.018	0.100
LAD	0.15	0.050	1.4	1.7	20.6	806	1416	100.4	1.0	13.0	0.58	0.14	0.17	1.00	1.00	20.87	96.00	0.032	0.112
D!	0.26	0.032	1.5	19.9	18.8	1408	1652	121.7	1.4	14.1	0.54	0.14	0.18	1.09	1.09	25.60	155.40	0.041	0.077
EOPT!	0.43	0.287	2.0	7.0	20.7	1875	1880	25.0	15.0	11.0	0.58	0.16	0.45	1.01	9.89	26.34	271.60	-0.098	0.034
EOPT	0.15	0.051	1.6	1.5	20.8	1153	1415	99.9	1.0	13.0	0.58	0.14	0.17	1.00	1.01	21.17	96.30	0.022	0.103
EOPT' full	0.15	0.050	1.6	20.0	21.6	1410	1613	100.2	1.0	13.0	0.58	0.14	0.17	1.00	1.00	21.05	96.10	0.025	0.105
Std Dev	0.10	0.08	0.2	8.8	1.5	384	163	26.9	4.7	0.8	0.01	0.01	0.09	0.03	2.96	2.12	59.21	0.042	0.024
Mean	0.20	0.07	1.7	7.2	21.0	1244	1521	94.1	2.6	12.9	0.57	0.14	0.20	1.01	2.00	22.29	122.51	0.011	0.093
True Value	0.15	0.050	1.6	9.4	23.7	1370	1480	100.0	1.0	13.0	0.58	0.14	0.17	1.20	1.00	21.00	96.00	0.025	0.105

Parameter has assumed a boundary value

Parameter does not affect performance due to values of other parameters

Final Parameter Sets for Coquitlam Seed Sensitivity Synthetic Calibrations

All trials use EOPT' Objective Function

OF	C0IMPA	Delta	P0GRADL	P0GRADM	P0GRADU	E0LMID	E0LHI	P0AGEN	V0FLAS	P0PERC	P0DZSH	P0FRTK	P0FSTK	P0IRTK	P0ISTK	P0UGTK	P0DZTK	P0SREP	P0RREP
Trial 1	0.15	0.05	1.6	0.4	20.6	1286	1407	99.9	1.0	13.00	0.58	0.14	0.17	1.00	1.00	21.08	96.15	0.022	0.103
Trial 2	0.15	0.05	1.6	20.0	19.9	1410	1887	100.2	1.0	13.00	0.58	0.14	0.17	1.00	1.00	21.00	95.99	0.025	0.105
Trial 3	0.15	0.05	1.6	20.0	27.2	1412	1662	100.0	1.0	13.00	0.58	0.14	0.17	1.00	1.00	21.01	96.02	0.025	0.105
Trial 4	0.15	0.05	1.6	20.0	14.2	1411	1873	100.2	1.0	13.00	0.58	0.14	0.17	1.00	1.00	21.02	95.99	0.025	0.105
Trial 5	0.15	0.05	1.7	1.6	20.7	636	1414	100.1	1.0	13.00	0.58	0.14	0.17	1.00	1.01	21.02	95.97	0.021	0.101
Std Dev	0.00	0.00	0.06	10.42	3.77	337	235	0.1	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.07	0.002	0.002
Mean	0.15	0.05	1.63	12.39	22.80	1231	1648	100.1	1.0	13.00	0.58	0.14	0.17	1.00	1.00	21.03	96.02	0.024	0.104
True Value	0.15	0.05	1.6	9.4	23.7	1370	1480	100.0	1.0	13.00	0.58	0.14	0.17	1.20	1.00	21.00	96.00	0.025	0.105

Parameter has assumed a boundary value

Parameter does not affect performance due to values of other parameters

Final Parameter Sets for Multi-Objective Automatic Calibrations of Coquitlam Lake Watershed

OF	C0IMPA	Delta	P0GRADL	P0GRADM	P0GRADU	E0LMID	E0LHI	P0AGEN	V0FLAS	P0PERC	P0DZSH	P0FRTK	P0FSTK	P0IRTK	P0ISTK	P0UGTK	P0DZTK	P0SREP	P0RREP
SLS	0.34	0.217	0.0	18.1	20.0	1599	1607	25.0	32.2	31.8	1.00	0.20	0.00	1.00	1.00	46.17	82.96	0.256	0.224
HMLE	0.96	0.001	5.0	8.3	0.0	763	1540	45.5	20.0	4.8	1.00	0.30	0.84	1.00	1.00	45.97	137.63	-0.418	-0.494
EOPT'	0.00	0.329	0.0	11.7	8.1	1284	1987	25.0	32.8	17.7	1.00	0.17	0.26	1.00	1.00	49.52	67.03	0.030	0.159
E!	0.32	0.223	0.0	20.0	8.1	1600	1828	25.0	32.2	31.7	1.00	0.20	0.00	1.00	1.00	45.05	82.82	0.251	0.226
LAD	0.04	0.239	0.8	10.1	3.6	1160	1829	48.3	29.2	14.2	1.00	0.16	0.39	1.40	1.01	49.66	71.64	-0.116	0.048
D!	0.70	0.002	0.0	1.5	20.0	1383	1390	111.6	20.0	19.3	1.00	0.19	0.00	1.00	1.00	23.70	143.06	-0.001	-0.220
EOPT!	0.50	0.137	0.0	8.3	13.8	1447	1655	25.0	26.4	39.8	0.75	0.20	0.00	1.01	2.94	37.12	123.72	0.274	0.235
EOPT	0.34	0.218	0.0	20.0	12.6	1502	1829	25.0	31.5	31.6	1.00	0.20	0.00	1.00	1.00	44.10	79.00	0.244	0.218
Std Dev	0.32	0.12	1.8	5.4	9.4	278	194	30.1	5.4	11.7	0.09	0.04	0.30	0.14	0.69	8.51	31.01	0.244	0.268
Mean	0.40	0.17	0.7	13.8	13.5	1342	1708	41.3	28.1	23.9	0.97	0.20	0.19	1.05	1.24	42.41	98.48	0.065	0.050
BC Hydro	0.15	0.050	1.6	9.4	23.7	1370	1480	100.0	1.0	13.0	0.58	0.14	0.17	1.20	1.00	21.00	96.00	0.025	0.105

Parameter has assumed a boundary value

Parameter does not affect performance due to values of other parameters

Final Parameter Sets for Multi-Objective Automatic Calibrations of Coquitlam Lake Watershed using an Expanded Parameter Space

OF	C0IMPA	Delta	P0GRADL	P0GRADM	P0GRADU	E0LMID	E0LHI	P0AGEN	V0FLAS	P0PERC	P0DZSH	P0FRTK	P0FSTK	P0IRTK	P0ISTK	P0UGTK	P0DZTK	P0SREP	P0RREP
SLS	0.82	0.085	0.0	0.0	28.0	612	1586	4.9	36.0	31.7	0.94	0.24	0.00	0.10	0.10	81.88	97.07	0.216	0.236
HMLE	0.92	0.009	6.8	0.0	16.6	1586	1932	45.1	11.9	4.5	0.00	0.31	0.85	0.90	0.10	176.66	46.43	-0.419	-0.517
EOPT'	0.40	0.198	0.0	10.4	8.3	1208	1730	18.4	29.6	29.0	0.64	0.22	0.26	0.10	0.10	65.16	65.80	-0.017	0.152
E!	0.71	0.092	0.0	22.9	4.8	1526	1924	7.1	30.1	32.3	0.00	0.24	0.00	0.10	0.10	110.36	56.23	0.246	0.244
LAD	0.00	0.328	0.0	10.0	13.0	1143	1830	26.0	30.6	15.5	0.60	0.17	0.39	1.53	0.10	73.90	73.46	-0.049	0.090
D!	0.99	0.004	0.0	0.0	22.6	401	1370	5.1	25.6	22.7	0.00	0.23	0.07	0.10	0.10	113.27	51.88	-0.033	-0.089
EOPT!	0.78	0.067	0.9	8.7	23.4	1368	1877	8.3	27.9	41.6	0.63	0.22	0.02	0.23	8.78	72.85	398.77	0.171	0.180
EOPT	0.98	0.007	0.0	28.5	1.8	1557	1828	4.5	36.2	31.7	1.00	0.24	0.00	0.10	0.10	59.50	102.05	0.220	0.236
Std Dev	0.34	0.11	2.4	10.8	10.2	446	194	14.4	7.6	11.6	0.42	0.04	0.30	0.53	3.07	38.98	141.34	0.224	0.260
Mean	0.70	0.10	1.0	10.1	19.6	1175	1760	14.9	28.5	26.1	0.48	0.23	0.20	0.39	1.19	99.16	147.43	0.042	0.067
BC Hydro	0.15	0.050	1.6	9.4	23.7	1370	1480	100.0	1.0	13.0	0.58	0.14	0.17	1.20	1.00	21.00	96.00	0.025	0.105

Parameter has assumed a boundary value
Parameter does not affect performance due to values of other parameters

Final Parameter Sets for Coquitlam Seed Sensitivity Calibrations

All trials use the EOPT' Objective Function

OF	C0IMPA	Delta	P0GRADL	P0GRADM	P0GRADU	E0LMID	E0LHI	P0AGEN	V0FLAS	P0PERC	P0DZSH	P0FRTK	P0FSTK	P0IRTK	P0ISTK	P0UGTK	P0DZTK	P0SREP	P0RREP
Trial 1	0.01	0.25	12.3	0.0	29.9	638	1475	25.0	36.8	11.87	0.80	0.16	0.35	1.00	1.00	34.94	73.06	-0.364	-0.209
Trial 2	0.08	0.23	2.5	1.9	30.0	1491	1519	25.0	35.8	15.69	0.63	0.16	0.36	1.00	1.01	10.00	97.58	-0.160	0.019
Trial 3	0.13	0.22	2.7	18.0	2.8	1476	1633	25.0	34.6	16.55	0.63	0.17	0.35	1.00	1.00	10.01	99.48	-0.177	0.007
Trial 4	0.16	0.21	2.5	18.4	17.0	1459	1864	25.0	35.5	14.67	0.66	0.17	0.36	1.00	1.01	36.05	85.72	-0.182	0.016
Trial 5	0.20	0.20	2.6	18.7	15.0	1475	1799	25.0	35.5	17.33	0.63	0.17	0.34	1.00	1.01	10.00	100.71	-0.173	0.009
Std Dev	0.08	0.02	4.4	9.6	19.2	374	170	0.0	0.8	2.12	0.08	0.01	0.01	0.00	0.00	13.97	11.82	0.086	0.099
Mean	0.12	0.22	4.5	11.4	16.4	1308	1658	25.0	35.6	15.22	0.67	0.17	0.35	1.00	1.01	20.20	91.31	-0.211	-0.032
BC Hydro	0.15	0.05	1.6	9.4	23.7	1370	1480	100.0	1.0	13.00	0.58	0.14	0.17	1.20	1.00	21.00	96.00	0.025	0.105

Parameter has assumed a boundary value
Parameter does not affect performance due to values of other parameters

Final Parameter Sets for Additional Coquitlam Seed Sensitivity Calibrations

All trials use the HMLE Objective Function

OF	C0IMPA	Delta	P0GRADL	P0GRADM	P0GRADU	E0LMID	E0LHI	P0AGEN	V0FLAS	P0PERC	P0DZSH	P0FRTK	P0FSTK	P0IRTK	P0ISTK	P0UGTK	P0DZTK	P0SREP	P0RREP
Trial 1	0.16	0.12	20.0	0.0	12.5	990	1448	25.0	25.7	1.63	1.00	0.13	1.45	1.86	1.22	47.19	117.45	-0.765	-0.715
Trial 2	0.24	0.11	9.9	10.6	30.0	403	1448	25.0	25.6	1.35	1.00	0.13	1.51	2.06	1.06	18.74	103.39	-0.723	-0.572
Trial 3	0.20	0.11	10.5	19.5	30.0	1448	1448	25.0	26.4	1.29	1.00	0.13	1.43	1.96	1.89	14.88	102.60	-0.725	-0.573
Trial 4	0.10	0.13	20.0	0.0	30.0	990	1448	25.0	26.5	1.30	1.00	0.12	1.36	1.79	1.08	21.20	109.77	-0.763	-0.718
Trial 5	0.18	0.10	15.5	19.9	3.4	508	1000	25.0	26.1	1.68	1.00	0.12	1.65	1.84	1.30	10.72	115.95	-0.779	-0.692
Std Dev	0.05	0.01	4.9	9.9	12.5	422	200	0.0	0.4	0.19	0.00	0.01	0.11	0.11	0.34	14.34	6.88	0.025	0.075
Mean	0.17	0.11	15.2	10.0	21.2	868	1358	25.0	26.0	1.45	1.00	0.13	1.48	1.90	1.31	22.55	109.83	-0.751	-0.654
Z_Coqtlm	0.15	0.05	1.6	9.4	23.7	1370	1480	100.0	1.0	13.00	0.58	0.14	0.17	1.20	1.00	21.00	96.00	0.025	0.105
HMLE tcs	0.96	0.00	5.0	8.3	0.0	763	1540	45.5	20.0	4.80	1.00	0.30	0.84	1.00	1.00	43.97	137.63	-0.418	-0.494

Parameter has assumed a boundary value
Parameter does not affect performance due to values of other parameters

Final Parameter Sets for Synthetic Calibrations of the Illecillewaet River Watershed Data Set

OF	C0IMPA	Delta	P0GRADL	P0GRADM	E0LMID	P0AGEN	V0FLAS	P0PERC	P0DZSH	P0FRTK	P0FSTK	P0GLTK	P0IRTK	P0ISTK	P0UGTK	P0DZTK	P0SREP1	P0RREP1	P0SREP2	P0RREP2	P0SREP3	P0RREP3
SLS	0.40	0.05	3.4	2.0	1828	101.1	35.8	31.09	0.25	0.78	1.00	1.70	1.96	3.04	17.00	171.27	-0.220	-0.140	-0.129	-0.178	-0.102	-0.099
HMLE	0.36	0.09	3.1	3.7	2386	104.8	32.8	31.03	0.25	0.78	1.00	1.72	2.30	4.26	16.86	167.31	-0.236	-0.145	-0.131	-0.171	-0.071	-0.082
EOPT'	0.40	0.05	3.6	2.1	1488	102.5	36.1	30.90	0.25	0.78	1.00	1.70	1.98	2.97	16.95	168.62	-0.223	-0.139	-0.119	-0.180	-0.110	-0.105
E!	0.41	0.05	3.4	2.1	1747	101.1	35.8	31.02	0.25	0.78	1.00	1.70	1.98	3.04	17.04	171.68	-0.223	-0.141	-0.129	-0.175	-0.102	-0.099
LAD	0.39	0.05	3.5	2.2	1616	107.6	36.2	30.86	0.25	0.78	0.99	1.70	1.98	2.93	16.76	166.89	-0.224	-0.139	-0.123	-0.185	-0.100	-0.103
D!	0.50	0.00	0.0	4.1	1726	78.1	28.6	29.47	0.00	0.76	1.01	1.57	1.50	1.84	14.67	100.31	-0.236	-0.137	0.063	-0.193	-0.082	-0.917
EOPT!	0.17	0.25	8.7	5.8	1505	65.4	24.6	32.20	0.42	0.85	1.05	1.56	7.49	4.23	13.67	276.21	-0.252	-0.092	-0.090	-0.097	-0.038	-0.104
EOPT	0.40	0.05	3.5	2.1	1699	99.9	35.9	30.99	0.25	0.78	1.00	1.70	1.98	3.03	17.00	170.03	-0.223	-0.140	-0.130	-0.177	-0.106	-0.102
EOPT' ifr	0.41	0.04	3.6	2.1	1304	102.6	36.4	30.87	0.25	0.78	0.99	1.70	1.99	2.89	16.88	167.40	-0.221	-0.139	-0.101	-0.178	-0.109	-0.106
Std Dev	0.09	0.07	2.2	1.3	303	14.2	4.2	0.69	0.11	0.02	0.02	0.06	1.85	0.73	1.24	37.94	0.011	0.016	0.062	0.028	0.024	0.272
Mean	0.38	0.07	3.7	2.9	1700	95.9	33.6	30.94	0.24	0.79	1.00	1.67	2.57	3.14	16.31	182.42	-0.229	-0.135	-0.099	-0.170	-0.091	-0.191
True Value	0.40	0.05	4.0	2.0	1009	100.0	36.0	31.00	0.25	0.78	1.00	1.70	2.00	3.00	17.00	168.00	-0.220	-0.140	-0.100	-0.170	-0.110	-0.110

Parameter has assumed a boundary value — Parameter does not affect performance due to values of other parameters

Final Parameter Sets for Illecillewaet Seed Sensitivity Synthetic Calibrations

All trials use the EOPT' Objective Function

OF	C0IMPA	Delta	P0GRADL	P0GRADM	E0LMID	P0AGEN	V0FLAS	P0PERC	P0DZSH	P0FRTK	P0FSTK	P0GLTK	P0IRTK	P0ISTK	P0UGTK	P0DZTK	P0SREP1	P0RREP1	P0SREP2	P0RREP2	P0SREP3	P0RREP3
Trial 1	0.40	0.05	3.6	2.1	1488	102.5	36.1	30.90	0.25	0.78	1.00	1.70	1.98	2.97	16.95	168.62	-0.223	-0.139	-0.119	-0.180	-0.110	-0.105
Trial 2	0.40	0.05	3.8	2.2	1239	102.5	36.2	30.97	0.25	0.78	1.00	1.70	1.99	2.94	16.84	166.52	-0.223	-0.139	-0.098	-0.179	-0.113	-0.110
Trial 3	0.40	0.05	3.9	2.1	1055	100.2	36.0	30.97	0.25	0.78	1.00	1.70	1.99	2.99	17.03	168.99	-0.223	-0.140	-0.099	-0.173	-0.109	-0.107
Trial 4	0.40	0.05	3.5	2.2	1540	100.0	36.0	31.01	0.25	0.78	1.00	1.70	1.97	3.01	16.91	166.97	-0.224	-0.140	-0.121	-0.181	-0.106	-0.104
Trial 5	0.44	0.02	4.1	2.0	1079	94.6	35.6	30.94	0.25	0.78	1.00	1.69	2.03	3.14	17.17	169.80	-0.220	-0.141	-0.102	-0.169	-0.123	-0.113
Std Dev	0.02	0.01	0.2	0.1	226	3.2	0.2	0.04	0.00	0.00	0.00	0.00	0.02	0.08	0.13	1.39	0.002	0.001	0.011	0.005	0.007	0.004
Mean	0.41	0.04	3.8	2.1	1280	100.0	36.0	30.96	0.25	0.78	1.00	1.70	1.99	3.01	16.98	168.18	-0.223	-0.140	-0.108	-0.176	-0.112	-0.108
True Value	0.40	0.05	4.0	2.0	1009	100.0	36.0	31.00	0.25	0.78	1.00	1.70	2.00	3.00	17.00	168.00	-0.220	-0.140	-0.100	-0.170	-0.110	-0.110

Parameter has assumed a boundary value — Parameter does not affect performance due to values of other parameters

Final Parameter Sets for Multi-Objective Automatic Calibrations of Illecillewaet River Watershed

OF	C0IMPA	Delta	P0GRADL	P0GRADM	E0LMID	P0AGEN	V0FLAS	P0PERC	P0DZSH	P0FRTK	P0FSTK	P0GLTK	P0IRTK	P0ISTK	P0UGTK	P0DZTK	P0SREP1	P0RREP1	P0SREP2	P0RREP2	P0SREP3	P0RREP3
SLS	0.42	0.02	0.0	20.0	1755	199.9	20.0	5.48	0.97	0.91	0.63	0.75	10.00	1.00	22.37	299.97	-0.455	-0.486	-0.061	0.590	0.797	0.780
EOPT'	0.99	0.40	0.4	14.1	1010	199.4	20.0	39.39	0.49	0.94	0.95	0.69	9.91	1.15	18.93	299.78	-0.566	-0.620	-0.386	0.161	0.992	0.995
E!	0.32	0.00	0.2	20.0	950	196.8	20.1	5.20	1.00	0.86	0.60	0.83	9.91	1.01	11.01	299.18	-0.513	-0.374	-0.588	0.234	0.993	0.264
LAD	0.99	0.35	3.9	20.0	1915	184.9	39.8	49.99	0.49	1.13	1.04	0.74	10.00	1.78	13.95	299.94	-0.535	-0.589	-0.181	0.566	0.221	0.999
EOPT	0.99	0.27	0.0	20.0	1857	103.3	20.0	49.80	0.54	1.04	0.90	0.69	1.00	1.00	10.00	299.94	-0.538	-0.636	0.023	0.675	0.715	1.000
Std Dev	0.34	0.19	1.9	2.6	476	41.6	8.8	22.89	0.26	0.11	0.20	0.05	4.00	0.33	5.4	0.33	0.042	0.110	0.249	0.231	0.316	0.318
Mean	0.74	0.21	1.1	18.8	1497	176.8	24.0	29.97	0.70	0.97	0.82	0.74	8.16	1.19	16.3	299.76	-0.521	-0.541	-0.239	0.445	0.744	0.808
Manual	0.40	0.05	4.0	2.0	1009	100.0	36.0	31.00	0.25	0.78	1.00	1.70	2.00	3.00	17.00	168.00	-0.220	-0.140	-0.100	-0.170	-0.110	-0.110

Parameter has assumed a boundary value — Parameter does not affect performance due to values of other parameters

Final Parameter Sets for Illecillewaet Seed Sensitivity Calibrations

All trials use the EOPT' Objective Function

OF	C0IMPA	Delta	P0GRADL	P0GRADM	E0LMID	P0AGEN	V0FLAS	P0PERC	P0DZSH	P0FRTK	P0FSTK	P0GLTK	P0IRTK	P0ISTK	P0UGTK	P0DZTK	P0SREP1	P0RREP1	P0SREP2	P0RREP2	P0SREP3	P0RREP3
Trial 1	0.99	0.40	0.4	14.1	1010	199.4	20.0	39.39	0.49	0.94	0.95	0.69	9.91	1.15	18.93	299.78	-0.566	-0.620	-0.386	0.161	0.992	0.995
Trial 2	0.24	0.00	0.8	19.9	1702	200.0	21.3	4.64	0.97	0.94	0.61	0.80	10.00	1.06	35.89	299.85	-0.433	-0.404	-0.099	0.386	0.645	0.707
Trial 3	0.29	0.00	1.1	20.0	1735	199.9	21.7	5.03	0.92	0.96	0.62	0.80	10.00	1.01	29.48	299.00	-0.453	-0.370	-0.092	0.272	0.651	0.703
Trial 4	0.98	0.15	2.5	19.3	1790	103.9	26.7	43.16	0.52	1.09	0.99	0.71	9.75	4.24	11.88	292.56	-0.542	-0.552	-0.160	0.568	0.402	0.994
Trial 5	0.99	0.35	0.9	11.7	996	199.3	20.0	46.74	0.47	0.94	0.95	0.69	9.99	1.01	20.09	299.97	-0.563	-0.611	-0.331	0.104	0.999	0.995
Trial 6	0.99	0.42	2.3	20.0	1915	133.1	20.0	44.42	0.52	1.05	0.95	0.70	9.95	1.31	14.62	299.51	-0.533	-0.595	-0.113	0.573	0.489	1.000
Std Dev	0.37	0.20	0.9	3.7	411	42.9	2.6	20.07	0.23	0.07	0.18	0.05	0.10	1.29	9.15	2.90	0.057	0.110	0.129	0.200	0.250	0.150
Mean	0.75	0.22	1.4	17.5	1525	172.6	21.6	30.56	0.65	0.99	0.85	0.73	9.93	1.63	21.82	298.45	-0.515	-0.525	-0.197	0.344	0.696	0.899
Manual	0.40	0.05	4.0	2.0	1009	100.0	36.0	31.00	0.25	0.78	1.00	1.70	2.00	3.00	17.00	168.00	-0.220	-0.140	-0.100	-0.170	-0.110	-0.110

(light shading) Parameter has assumed a boundary value

(dark shading) Parameter does not affect performance due to values of other parameters

Appendix B.2

Annual Statistics for All Calibrations

Annual Statistics for Multi-Objective Automatic Calibrations of Coquitlam Lake Watershed

Nash-Sutcliffe Efficiency E!

Period	Objective Function								
	SLS	HMLE	EOPT'	E!	LAD	D!	EOPT!	EOPT	BC Hydro
WY 86-87	0.80	0.65	0.81	0.80	0.82	0.68	0.80	0.80	0.78
WY 87-88	0.58	0.58	0.59	0.58	0.62	0.61	0.57	0.58	0.48
WY 88-89	0.80	0.66	0.81	0.80	0.80	0.68	0.80	0.80	0.73
WY 89-90	0.77	0.71	0.78	0.77	0.79	0.73	0.77	0.77	0.69
WY 90-91	0.89	0.65	0.88	0.89	0.87	0.70	0.89	0.89	0.89
WY 91-92	0.87	0.75	0.88	0.87	0.89	0.78	0.87	0.87	0.84
WY 92-93	0.63	0.62	0.66	0.63	0.68	0.62	0.63	0.63	0.58
WY 93-94	0.75	0.49	0.73	0.75	0.70	0.53	0.75	0.75	0.74
WY 94-95	0.67	0.44	0.67	0.67	0.64	0.50	0.67	0.67	0.66
WY 95-96	0.85	0.63	0.85	0.85	0.84	0.68	0.85	0.85	0.85
WY 96-97	0.69	0.38	0.67	0.69	0.64	0.45	0.69	0.69	0.71
WY 97-98	0.78	0.54	0.77	0.78	0.76	0.58	0.78	0.78	0.78
WY 98-99	0.63	0.36	0.60	0.63	0.59	0.40	0.64	0.63	0.59
Mean	0.75	0.57	0.75	0.75	0.74	0.61	0.75	0.75	0.72
Abs Mean	0.75	0.57	0.75	0.75	0.74	0.61	0.75	0.75	0.72
WY 86-99	0.77	0.57	0.77	0.77	0.76	0.61	0.77	0.77	0.75

Coefficient of Determination D! (R^2)

Period	Objective Function								
	SLS	HMLE	EOPT'	E!	LAD	D!	EOPT!	EOPT	BC Hydro
WY 86-87	0.81	0.79	0.81	0.81	0.82	0.81	0.81	0.81	0.81
WY 87-88	0.68	0.67	0.68	0.68	0.69	0.68	0.67	0.68	0.67
WY 88-89	0.81	0.83	0.81	0.81	0.81	0.81	0.81	0.81	0.77
WY 89-90	0.82	0.84	0.82	0.82	0.82	0.82	0.82	0.82	0.80
WY 90-91	0.90	0.88	0.90	0.90	0.90	0.90	0.90	0.90	0.89
WY 91-92	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.88
WY 92-93	0.69	0.70	0.69	0.69	0.70	0.69	0.69	0.69	0.68
WY 93-94	0.79	0.79	0.78	0.79	0.77	0.80	0.79	0.79	0.76
WY 94-95	0.68	0.63	0.68	0.68	0.67	0.68	0.68	0.68	0.66
WY 95-96	0.85	0.81	0.85	0.85	0.85	0.85	0.85	0.85	0.85
WY 96-97	0.76	0.68	0.77	0.76	0.76	0.77	0.76	0.76	0.77
WY 97-98	0.78	0.75	0.78	0.78	0.78	0.78	0.78	0.78	0.79
WY 98-99	0.72	0.72	0.71	0.72	0.71	0.72	0.72	0.72	0.71
Mean	0.78	0.77	0.78	0.78	0.78	0.78	0.78	0.78	0.77
WY 86-99	0.77	0.75	0.77	0.77	0.77	0.77	0.77	0.77	0.76

Modified Nash-Sutcliffe Efficiency EOPT!

Period	Objective Function								
	SLS	HMLE	EOPT'	E!	LAD	D!	EOPT!	EOPT	BC Hydro
WY 86-87	0.74	0.28	0.81	0.74	0.78	0.35	0.72	0.74	0.74
WY 87-88	0.46	0.23	0.54	0.46	0.60	0.31	0.44	0.47	0.39
WY 88-89	0.74	0.30	0.79	0.74	0.74	0.37	0.71	0.74	0.72
WY 89-90	0.74	0.33	0.73	0.74	0.70	0.41	0.72	0.75	0.69
WY 90-91	0.85	0.28	0.85	0.85	0.80	0.38	0.83	0.85	0.88
WY 91-92	0.77	0.39	0.85	0.77	0.88	0.49	0.75	0.78	0.76
WY 92-93	0.51	0.30	0.62	0.51	0.67	0.36	0.49	0.51	0.50
WY 93-94	0.65	0.01	0.57	0.65	0.51	0.11	0.67	0.65	0.62
WY 94-95	0.65	0.06	0.61	0.65	0.55	0.16	0.63	0.65	0.65
WY 95-96	0.83	0.23	0.81	0.83	0.76	0.33	0.82	0.83	0.85
WY 96-97	0.55	-0.11	0.48	0.55	0.41	0.01	0.57	0.55	0.54
WY 97-98	0.72	0.10	0.65	0.72	0.61	0.19	0.74	0.72	0.69
WY 98-99	0.34	-0.22	0.26	0.34	0.23	-0.14	0.36	0.34	0.28
Mean	0.66	0.17	0.66	0.66	0.63	0.26	0.65	0.66	0.64
Abs Mean	0.66	0.22	0.66	0.66	0.63	0.28	0.65	0.66	0.64
WY 86-99	0.76	0.16	0.69	0.76	0.64	0.25	0.77	0.75	0.71

Volume Error dV/V

Period	Objective Function								
	SLS	HMLE	EOPT'	E!	LAD	D!	EOPT!	EOPT	BC Hydro
WY 86-87	-0.06	0.38	0.00	-0.06	0.04	0.33	-0.08	-0.06	-0.04
WY 87-88	-0.11	0.36	-0.06	-0.11	-0.02	0.30	-0.13	-0.11	-0.10
WY 88-89	-0.06	0.36	0.02	-0.06	0.06	0.31	-0.08	-0.06	-0.01
WY 89-90	-0.03	0.38	0.05	-0.03	0.09	0.32	-0.05	-0.03	0.01
WY 90-91	-0.04	0.37	0.03	-0.04	0.07	0.32	-0.05	-0.04	-0.01
WY 91-92	-0.10	0.36	-0.03	-0.10	0.01	0.29	-0.12	-0.10	-0.09
WY 92-93	-0.12	0.33	-0.04	-0.12	0.01	0.27	-0.14	-0.12	-0.08
WY 93-94	0.10	0.48	0.16	0.10	0.19	0.42	0.08	0.10	0.12
WY 94-95	-0.02	0.38	0.05	-0.02	0.10	0.34	-0.04	-0.02	0.02
WY 95-96	-0.02	0.40	0.04	-0.02	0.08	0.34	-0.03	-0.02	0.00
WY 96-97	0.14	0.49	0.20	0.14	0.23	0.44	0.13	0.14	0.17
WY 97-98	0.06	0.44	0.12	0.06	0.15	0.39	0.04	0.06	0.09
WY 98-99	0.29	0.58	0.34	0.29	0.36	0.54	0.28	0.29	0.31
Mean	0.00	0.41	0.07	0.00	0.11	0.35	-0.01	0.00	0.03
Abs Mean	0.09	0.41	0.09	0.09	0.11	0.35	0.10	0.09	0.08
WY 86-99	0.02	0.42	0.08	0.02	0.12	0.36	0.00	0.02	0.04

Annual Statistics for Multi-Objective Automatic Calibrations of Coquitlam Lake Watershed Using an Expanded Parameter Space

Nash-Sutcliffe Efficiency E!

Period	Objective Function								
	SLS	HMLE	EOPT'	E!	LAD	D!	EOPT!	EOPT	BC Hydro
WY 86-87	0.79	0.65	0.81	0.80	0.82	0.74	0.79	0.80	0.78
WY 87-88	0.56	0.58	0.60	0.59	0.63	0.65	0.59	0.60	0.48
WY 88-89	0.80	0.66	0.81	0.80	0.81	0.76	0.79	0.81	0.73
WY 89-90	0.76	0.71	0.78	0.76	0.80	0.78	0.76	0.77	0.69
WY 90-91	0.89	0.64	0.88	0.89	0.87	0.78	0.89	0.89	0.89
WY 91-92	0.86	0.75	0.88	0.87	0.89	0.84	0.87	0.87	0.84
WY 92-93	0.62	0.62	0.66	0.64	0.68	0.66	0.64	0.64	0.58
WY 93-94	0.75	0.49	0.74	0.76	0.70	0.62	0.75	0.76	0.74
WY 94-95	0.67	0.44	0.66	0.68	0.65	0.58	0.68	0.68	0.66
WY 95-96	0.85	0.62	0.85	0.85	0.84	0.76	0.85	0.85	0.85
WY 96-97	0.71	0.38	0.67	0.70	0.64	0.52	0.69	0.70	0.71
WY 97-98	0.78	0.54	0.77	0.77	0.76	0.65	0.77	0.77	0.78
WY 98-99	0.64	0.36	0.60	0.63	0.59	0.49	0.64	0.63	0.59
Mean	0.74	0.57	0.75	0.75	0.74	0.68	0.75	0.75	0.72
Abs Mean	0.74	0.57	0.75	0.75	0.74	0.68	0.75	0.75	0.72
WY 86-99	0.77	0.57	0.77	0.77	0.76	0.68	0.77	0.77	0.75

Coefficient of Determination D! (R^2)

Period	Objective Function								
	SLS	HMLE	EOPT'	E!	LAD	D!	EOPT!	EOPT	BC Hydro
WY 86-87	0.81	0.79	0.81	0.80	0.82	0.80	0.80	0.81	0.81
WY 87-88	0.67	0.67	0.68	0.67	0.68	0.67	0.66	0.67	0.67
WY 88-89	0.82	0.83	0.82	0.81	0.81	0.82	0.81	0.82	0.77
WY 89-90	0.82	0.84	0.82	0.81	0.82	0.81	0.81	0.82	0.80
WY 90-91	0.89	0.87	0.90	0.90	0.90	0.90	0.90	0.90	0.89
WY 91-92	0.88	0.88	0.89	0.88	0.89	0.89	0.88	0.88	0.88
WY 92-93	0.69	0.70	0.69	0.68	0.70	0.68	0.68	0.69	0.68
WY 93-94	0.78	0.79	0.79	0.80	0.78	0.82	0.80	0.79	0.76
WY 94-95	0.68	0.63	0.68	0.69	0.67	0.69	0.69	0.69	0.66
WY 95-96	0.85	0.81	0.85	0.85	0.85	0.85	0.85	0.85	0.85
WY 96-97	0.77	0.69	0.77	0.77	0.76	0.77	0.77	0.77	0.77
WY 97-98	0.78	0.75	0.78	0.77	0.78	0.77	0.78	0.78	0.79
WY 98-99	0.72	0.72	0.71	0.72	0.71	0.72	0.72	0.72	0.71
Mean	0.78	0.77	0.78	0.78	0.78	0.78	0.78	0.78	0.77
WY 86-99	0.77	0.75	0.77	0.77	0.77	0.77	0.77	0.77	0.76

Modified Nash-Sutcliffe Efficiency EOPT!

Period	Objective Function								
	SLS	HMLE	EOPT'	E!	LAD	D!	EOPT!	EOPT	BC Hydro
WY 86-87	0.70	0.27	0.80	0.72	0.77	0.52	0.69	0.72	0.74
WY 87-88	0.41	0.22	0.54	0.48	0.62	0.44	0.46	0.48	0.39
WY 88-89	0.71	0.30	0.79	0.73	0.74	0.54	0.71	0.73	0.72
WY 89-90	0.71	0.33	0.74	0.73	0.71	0.53	0.71	0.73	0.69
WY 90-91	0.83	0.28	0.85	0.85	0.79	0.54	0.84	0.84	0.88
WY 91-92	0.72	0.39	0.85	0.76	0.87	0.64	0.74	0.75	0.76
WY 92-93	0.47	0.30	0.62	0.52	0.67	0.46	0.51	0.51	0.50
WY 93-94	0.69	0.01	0.57	0.67	0.50	0.27	0.68	0.67	0.62
WY 94-95	0.63	0.06	0.61	0.67	0.55	0.32	0.66	0.66	0.65
WY 95-96	0.81	0.22	0.82	0.82	0.76	0.50	0.83	0.82	0.85
WY 96-97	0.59	-0.10	0.47	0.56	0.40	0.14	0.55	0.56	0.54
WY 97-98	0.74	0.09	0.65	0.73	0.60	0.35	0.74	0.73	0.69
WY 98-99	0.37	-0.22	0.27	0.35	0.22	0.01	0.36	0.35	0.28
Mean	0.64	0.17	0.66	0.66	0.63	0.40	0.65	0.66	0.64
Abs Mean	0.64	0.21	0.66	0.66	0.63	0.40	0.65	0.66	0.64
WY 86-99	0.76	0.15	0.69	0.76	0.64	0.4	0.77	0.77	0.71

Volume Error dV/V

Period	Objective Function								
	SLS	HMLE	EOPT'	E!	LAD	D!	EOPT!	EOPT	BC Hydro
WY 86-87	-0.10	0.38	-0.01	-0.08	0.04	0.23	-0.11	-0.08	-0.04
WY 87-88	-0.15	0.36	-0.06	-0.11	-0.01	0.21	-0.12	-0.12	-0.10
WY 88-89	-0.09	0.36	0.02	-0.07	0.07	0.22	-0.09	-0.07	-0.01
WY 89-90	-0.05	0.38	0.05	-0.04	0.09	0.25	-0.05	-0.04	0.01
WY 90-91	-0.06	0.37	0.03	-0.04	0.07	0.24	-0.05	-0.04	-0.01
WY 91-92	-0.14	0.36	-0.04	-0.11	0.01	0.20	-0.13	-0.12	-0.09
WY 92-93	-0.14	0.32	-0.04	-0.12	0.01	0.19	-0.13	-0.12	-0.08
WY 93-94	0.06	0.48	0.16	0.09	0.20	0.35	0.07	0.08	0.12
WY 94-95	-0.04	0.38	0.06	-0.01	0.10	0.26	-0.02	-0.02	0.02
WY 95-96	-0.04	0.40	0.03	-0.03	0.08	0.26	-0.02	-0.03	0.00
WY 96-97	0.12	0.49	0.20	0.14	0.24	0.38	0.14	0.14	0.17
WY 97-98	0.04	0.44	0.11	0.04	0.16	0.30	0.03	0.04	0.09
WY 98-99	0.27	0.58	0.33	0.28	0.37	0.48	0.27	0.28	0.31
Mean	-0.02	0.41	0.07	0.00	0.11	0.27	-0.02	-0.01	0.03
Abs Mean	0.10	0.41	0.09	0.09	0.11	0.27	0.09	0.09	0.08
WY 86-99	-0.01	0.42	0.08	0.01	0.12	0.28	0.00	0.01	0.04

Annual Statistics for Coquitlam Seed Sensitivity Calibrations

All trials use the EOPT Objective Function

Nash-Sutcliffe Efficiency E!

Period	Objective Function					
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	BC Hydro
WY 86-87	0.81	0.81	0.81	0.81	0.81	0.78
WY 87-88	0.58	0.63	0.58	0.59	0.58	0.48
WY 88-89	0.80	0.81	0.80	0.81	0.81	0.73
WY 89-90	0.77	0.78	0.77	0.78	0.78	0.69
WY 90-91	0.88	0.87	0.88	0.88	0.88	0.89
WY 91-92	0.88	0.89	0.88	0.88	0.88	0.84
WY 92-93	0.66	0.68	0.66	0.66	0.66	0.58
WY 93-94	0.73	0.70	0.73	0.73	0.73	0.74
WY 94-95	0.66	0.65	0.66	0.66	0.66	0.66
WY 95-96	0.84	0.83	0.85	0.84	0.85	0.85
WY 96-97	0.67	0.63	0.66	0.66	0.66	0.71
WY 97-98	0.77	0.75	0.77	0.77	0.76	0.78
WY 98-99	0.58	0.58	0.59	0.59	0.59	0.59
Mean	0.74	0.74	0.74	0.74	0.74	0.72
Abs Mean	0.74	0.74	0.74	0.74	0.74	0.72
WY 86-99	0.76	0.76	0.76	0.76	0.76	0.75

Coefficient of Determination D! (R^2)

Period	Objective Function					
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	BC Hydro
WY 86-87	0.81	0.81	0.81	0.81	0.81	0.81
WY 87-88	0.68	0.69	0.68	0.68	0.68	0.67
WY 88-89	0.81	0.81	0.81	0.81	0.81	0.77
WY 89-90	0.82	0.81	0.82	0.82	0.82	0.80
WY 90-91	0.90	0.89	0.90	0.90	0.90	0.89
WY 91-92	0.89	0.89	0.89	0.89	0.89	0.88
WY 92-93	0.70	0.70	0.70	0.70	0.70	0.68
WY 93-94	0.79	0.78	0.78	0.78	0.78	0.76
WY 94-95	0.68	0.67	0.67	0.68	0.67	0.66
WY 95-96	0.85	0.84	0.85	0.85	0.85	0.85
WY 96-97	0.76	0.75	0.76	0.76	0.76	0.77
WY 97-98	0.78	0.77	0.78	0.78	0.77	0.79
WY 98-99	0.71	0.70	0.71	0.71	0.71	0.71
Mean	0.78	0.78	0.78	0.78	0.78	0.77
WY 86-99	0.77	0.76	0.77	0.77	0.77	0.76

Modified Nash-Sutcliffe Efficiency EOPT!

Period	Objective Function					
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	BC Hydro
WY 86-87	0.79	0.80	0.80	0.80	0.80	0.74
WY 87-88	0.54	0.59	0.53	0.54	0.53	0.39
WY 88-89	0.76	0.75	0.77	0.76	0.77	0.72
WY 89-90	0.70	0.71	0.71	0.71	0.71	0.69
WY 90-91	0.84	0.82	0.84	0.84	0.84	0.88
WY 91-92	0.87	0.87	0.86	0.86	0.86	0.76
WY 92-93	0.65	0.68	0.64	0.65	0.64	0.50
WY 93-94	0.55	0.52	0.55	0.55	0.55	0.62
WY 94-95	0.59	0.56	0.59	0.58	0.59	0.65
WY 95-96	0.80	0.79	0.80	0.80	0.80	0.85
WY 96-97	0.45	0.41	0.45	0.45	0.45	0.54
WY 97-98	0.64	0.63	0.64	0.64	0.64	0.69
WY 98-99	0.23	0.22	0.24	0.24	0.24	0.28
Mean	0.65	0.64	0.65	0.65	0.65	0.64
Abs Mean	0.65	0.64	0.65	0.65	0.65	0.64
WY 86-99	0.67	0.65	0.67	0.67	0.67	0.71

Volume Error dV/V

Period	Objective Function					
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	BC Hydro
WY 86-87	0.01	0.02	0.01	0.01	0.01	-0.04
WY 87-88	-0.05	-0.04	-0.05	-0.05	-0.05	-0.10
WY 88-89	0.05	0.05	0.04	0.04	0.04	-0.01
WY 89-90	0.07	0.07	0.06	0.07	0.06	0.01
WY 90-91	0.05	0.05	0.04	0.05	0.04	-0.01
WY 91-92	-0.02	-0.02	-0.02	-0.02	-0.02	-0.09
WY 92-93	-0.01	0.00	-0.02	-0.02	-0.02	-0.08
WY 93-94	0.18	0.19	0.18	0.18	0.18	0.12
WY 94-95	0.08	0.09	0.07	0.08	0.07	0.02
WY 95-96	0.05	0.05	0.04	0.04	0.04	0.00
WY 96-97	0.22	0.22	0.21	0.21	0.21	0.17
WY 97-98	0.13	0.13	0.12	0.13	0.12	0.09
WY 98-99	0.35	0.35	0.35	0.35	0.34	0.31
Mean	0.08	0.09	0.08	0.08	0.08	0.03
Abs Mean	0.10	0.10	0.09	0.10	0.09	0.08
WY 86-99	0.10	0.10	0.09	0.09	0.09	0.04

Annual Statistics for Additional Coquitlam Seed Sensitivity Calibrations

All trials use the HMLE Objective Function

Nash-Sutcliffe Efficiency E!

Period	Objective Function					
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	BC Hydro
WY 86-87	0.68	0.68	0.69	0.68	0.68	0.78
WY 87-88	0.60	0.61	0.61	0.60	0.60	0.48
WY 88-89	0.64	0.64	0.64	0.64	0.64	0.73
WY 89-90	0.70	0.70	0.70	0.70	0.70	0.69
WY 90-91	0.69	0.69	0.69	0.68	0.69	0.89
WY 91-92	0.78	0.79	0.79	0.78	0.78	0.84
WY 92-93	0.60	0.60	0.60	0.60	0.59	0.58
WY 93-94	0.50	0.49	0.49	0.49	0.49	0.74
WY 94-95	0.45	0.45	0.45	0.45	0.45	0.66
WY 95-96	0.68	0.68	0.69	0.67	0.68	0.85
WY 96-97	0.39	0.38	0.39	0.39	0.39	0.71
WY 97-98	0.56	0.56	0.56	0.56	0.56	0.78
WY 98-99	0.36	0.35	0.35	0.36	0.35	0.59
Mean	0.59	0.59	0.59	0.58	0.58	0.72
Abs Mean	0.59	0.59	0.59	0.58	0.58	0.72
WY 86-99	0.59	0.59	0.59	0.59	0.59	0.75

Coefficient of Determination D! (R^2)

Period	Objective Function					
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	BC Hydro
WY 86-87	0.81	0.81	0.81	0.81	0.81	0.81
WY 87-88	0.67	0.68	0.68	0.67	0.67	0.67
WY 88-89	0.80	0.80	0.80	0.79	0.79	0.77
WY 89-90	0.81	0.81	0.81	0.81	0.80	0.80
WY 90-91	0.89	0.89	0.89	0.88	0.89	0.89
WY 91-92	0.89	0.89	0.89	0.89	0.89	0.88
WY 92-93	0.69	0.70	0.70	0.69	0.68	0.68
WY 93-94	0.77	0.77	0.77	0.77	0.76	0.76
WY 94-95	0.64	0.64	0.64	0.63	0.64	0.66
WY 95-96	0.83	0.84	0.84	0.83	0.83	0.85
WY 96-97	0.71	0.71	0.71	0.71	0.71	0.77
WY 97-98	0.75	0.75	0.75	0.75	0.75	0.79
WY 98-99	0.69	0.69	0.69	0.70	0.69	0.71
Mean	0.77	0.77	0.77	0.76	0.76	0.77
WY 86-99	0.75	0.75	0.75	0.74	0.74	0.76

Modified Nash-Sutcliffe Efficiency EOPT!

Period	Objective Function					
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	BC Hydro
WY 86-87	0.32	0.32	0.33	0.32	0.32	0.74
WY 87-88	0.27	0.29	0.30	0.27	0.28	0.39
WY 88-89	0.26	0.23	0.23	0.26	0.24	0.72
WY 89-90	0.31	0.29	0.29	0.31	0.29	0.69
WY 90-91	0.31	0.30	0.30	0.31	0.30	0.88
WY 91-92	0.44	0.44	0.44	0.44	0.44	0.76
WY 92-93	0.25	0.23	0.23	0.26	0.23	0.50
WY 93-94	0.02	-0.01	0.00	0.02	0.01	0.62
WY 94-95	0.05	0.02	0.03	0.05	0.03	0.65
WY 95-96	0.30	0.31	0.32	0.30	0.31	0.85
WY 96-97	-0.10	-0.13	-0.13	-0.11	-0.12	0.54
WY 97-98	0.13	0.12	0.13	0.12	0.12	0.69
WY 98-99	-0.22	-0.24	-0.23	-0.22	-0.23	0.28
Mean	0.18	0.17	0.17	0.18	0.17	0.64
Abs Mean	0.23	0.23	0.23	0.23	0.22	0.64
WY 86-99	0.17	0.16	0.17	0.17	0.17	0.71

Volume Error dV/V

Period	Objective Function					
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	BC Hydro
WY 86-87	0.36	0.36	0.36	0.36	0.36	-0.04
WY 87-88	0.33	0.32	0.32	0.33	0.32	-0.10
WY 88-89	0.38	0.41	0.40	0.38	0.40	-0.01
WY 89-90	0.40	0.41	0.41	0.40	0.41	0.01
WY 90-91	0.37	0.39	0.39	0.37	0.39	-0.01
WY 91-92	0.34	0.35	0.35	0.34	0.34	-0.09
WY 92-93	0.35	0.37	0.37	0.35	0.36	-0.08
WY 93-94	0.48	0.49	0.49	0.48	0.49	0.12
WY 94-95	0.40	0.43	0.42	0.40	0.42	0.02
WY 95-96	0.37	0.37	0.37	0.37	0.37	0.00
WY 96-97	0.50	0.51	0.51	0.50	0.51	0.17
WY 97-98	0.43	0.44	0.44	0.43	0.44	0.09
WY 98-99	0.58	0.59	0.59	0.58	0.58	0.31
Mean	0.41	0.42	0.42	0.41	0.41	0.03
Abs Mean	0.41	0.42	0.42	0.41	0.41	0.08
WY 86-99	0.42	0.43	0.42	0.42	0.42	0.04

Annual Statistics for Multi-Objective Automatic Calibrations of Illecillewaet River Watershed

Nash-Sutcliffe Efficiency EI

Period	Objective Function					
	SLS	EOPT'	EI	LAD	EOPT	Manual
WY 71-72	0.93	0.93	0.93	0.95	0.92	0.94
WY 72-73	0.80	0.81	0.83	0.84	0.78	0.91
WY 73-74	0.90	0.88	0.86	0.91	0.89	0.97
WY 74-75	0.80	0.84	0.86	0.88	0.80	0.91
WY 75-76	0.86	0.84	0.84	0.86	0.84	0.93
WY 76-77	0.72	0.75	0.80	0.80	0.68	0.86
WY 77-78	0.80	0.85	0.84	0.85	0.82	0.92
WY 78-79	0.78	0.79	0.85	0.83	0.76	0.85
WY 79-80	0.84	0.86	0.89	0.90	0.85	0.93
WY 80-81	0.85	0.87	0.89	0.89	0.86	0.88
WY 81-82	0.80	0.86	0.89	0.88	0.83	0.87
WY 82-83	0.84	0.85	0.89	0.88	0.82	0.90
WY 83-84	0.88	0.89	0.87	0.90	0.87	0.93
WY 84-85	0.86	0.89	0.91	0.92	0.86	0.88
WY 85-86	0.92	0.94	0.94	0.94	0.93	0.90
WY 86-87	0.75	0.80	0.85	0.84	0.77	0.88
WY 87-88	0.69	0.71	0.77	0.76	0.67	0.84
WY 88-89	0.65	0.70	0.80	0.74	0.64	0.77
WY 89-90	0.81	0.87	0.89	0.91	0.83	0.84
Mean	0.81	0.84	0.86	0.87	0.81	0.89
Abs Mean	0.81	0.84	0.86	0.87	0.81	0.89
WY 71-90	0.83	0.85	0.87	0.88	0.83	0.9

Coefficient of Determination DI (R^2)

Period	Objective Function					
	SLS	EOPT'	EI	LAD	EOPT	Manual
WY 71-72	0.94	0.93	0.94	0.95	0.92	0.95
WY 72-73	0.89	0.87	0.85	0.88	0.89	0.95
WY 73-74	0.91	0.89	0.87	0.91	0.89	0.97
WY 74-75	0.90	0.87	0.87	0.90	0.88	0.95
WY 75-76	0.89	0.85	0.84	0.86	0.86	0.94
WY 76-77	0.89	0.87	0.84	0.89	0.88	0.95
WY 77-78	0.87	0.88	0.84	0.86	0.88	0.95
WY 78-79	0.91	0.86	0.87	0.89	0.89	0.95
WY 79-80	0.93	0.90	0.90	0.93	0.93	0.96
WY 80-81	0.91	0.89	0.89	0.90	0.90	0.91
WY 81-82	0.89	0.88	0.89	0.90	0.88	0.93
WY 82-83	0.92	0.89	0.90	0.90	0.91	0.93
WY 83-84	0.89	0.89	0.88	0.90	0.88	0.94
WY 84-85	0.92	0.92	0.92	0.94	0.92	0.93
WY 85-86	0.95	0.94	0.94	0.94	0.95	0.92
WY 86-87	0.91	0.88	0.89	0.90	0.90	0.94
WY 87-88	0.89	0.83	0.82	0.85	0.85	0.94
WY 88-89	0.90	0.87	0.87	0.89	0.89	0.95
WY 89-90	0.91	0.90	0.89	0.93	0.91	0.93
Mean	0.91	0.88	0.88	0.90	0.90	0.94
WY 71-90	0.9	0.88	0.87	0.9	0.89	0.93

Modified Nash-Sutcliffe Efficiency EOPTI

Period	Objective Function					
	SLS	EOPT'	EI	LAD	EOPT	Manual
WY 71-72	0.80	0.89	0.90	0.91	0.81	0.88
WY 72-73	0.51	0.56	0.71	0.65	0.45	0.80
WY 73-74	0.83	0.88	0.77	0.89	0.81	0.92
WY 74-75	0.47	0.61	0.72	0.70	0.48	0.76
WY 75-76	0.71	0.79	0.81	0.84	0.71	0.92
WY 76-77	0.35	0.43	0.62	0.53	0.26	0.68
WY 77-78	0.59	0.70	0.80	0.75	0.62	0.92
WY 78-79	0.49	0.55	0.71	0.62	0.46	0.69
WY 79-80	0.57	0.64	0.77	0.74	0.58	0.84
WY 80-81	0.67	0.73	0.86	0.78	0.67	0.86
WY 81-82	0.56	0.71	0.82	0.75	0.59	0.78
WY 82-83	0.58	0.63	0.78	0.70	0.53	0.82
WY 83-84	0.80	0.84	0.82	0.88	0.75	0.86
WY 84-85	0.62	0.72	0.83	0.79	0.62	0.78
WY 85-86	0.74	0.84	0.93	0.88	0.76	0.88
WY 86-87	0.41	0.53	0.68	0.61	0.43	0.72
WY 87-88	0.30	0.39	0.57	0.47	0.28	0.63
WY 88-89	0.25	0.33	0.57	0.41	0.20	0.53
WY 89-90	0.59	0.69	0.83	0.76	0.58	0.74
Mean	0.57	0.66	0.76	0.72	0.56	0.79
Abs Mean	0.57	0.66	0.76	0.72	0.56	0.79
WY 71-90	0.6	0.68	0.8	0.74	0.58	0.82

Volume Error dV/V

Period	Objective Function					
	SLS	EOPT'	EI	LAD	EOPT	Manual
WY 71-72	-0.13	-0.04	0.03	-0.04	-0.11	-0.05
WY 72-73	-0.28	-0.25	-0.12	-0.19	-0.33	-0.11
WY 73-74	-0.07	0.00	0.08	0.02	-0.08	0.04
WY 74-75	-0.33	-0.23	-0.14	-0.18	-0.32	-0.15
WY 75-76	-0.15	-0.05	0.03	-0.02	-0.13	0.01
WY 76-77	-0.36	-0.32	-0.19	-0.28	-0.42	-0.18
WY 77-78	-0.21	-0.15	-0.04	-0.10	-0.21	-0.01
WY 78-79	-0.29	-0.24	-0.14	-0.21	-0.30	-0.16
WY 79-80	-0.27	-0.22	-0.12	-0.16	-0.27	-0.08
WY 80-81	-0.18	-0.14	-0.03	-0.10	-0.19	-0.03
WY 81-82	-0.25	-0.15	-0.06	-0.14	-0.24	-0.08
WY 82-83	-0.26	-0.22	-0.10	-0.18	-0.29	-0.09
WY 83-84	-0.08	-0.05	0.05	-0.02	-0.12	0.07
WY 84-85	-0.24	-0.17	-0.08	-0.14	-0.24	-0.10
WY 85-86	-0.18	-0.09	0.00	-0.06	-0.17	-0.02
WY 86-87	-0.34	-0.27	-0.17	-0.23	-0.34	-0.16
WY 87-88	-0.38	-0.32	-0.20	-0.28	-0.39	-0.20
WY 88-89	-0.41	-0.37	-0.23	-0.33	-0.44	-0.24
WY 89-90	-0.22	-0.18	-0.06	-0.15	-0.25	-0.10
Mean	-0.24	-0.18	-0.08	-0.15	-0.25	-0.09
Abs Mean	0.24	0.18	0.10	0.15	0.25	0.10
WY 71-90	-0.24	-0.17	-0.07	-0.14	-0.25	-0.08

Annual Statistics for Illecillewaet Seed Sensitivity Calibrations

All trials use the EOPT Objective Function

Nash-Sutcliffe Efficiency E!

Period	Objective Function						
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Manual
WY 71-72	0.93	0.95	0.95	0.94	0.93	0.94	0.94
WY 72-73	0.81	0.83	0.83	0.82	0.81	0.81	0.91
WY 73-74	0.88	0.92	0.91	0.90	0.89	0.90	0.97
WY 74-75	0.84	0.83	0.84	0.86	0.84	0.85	0.91
WY 75-76	0.84	0.88	0.88	0.86	0.85	0.86	0.93
WY 76-77	0.75	0.79	0.79	0.78	0.74	0.76	0.86
WY 77-78	0.85	0.82	0.83	0.85	0.85	0.84	0.92
WY 78-79	0.79	0.82	0.83	0.81	0.79	0.80	0.85
WY 79-80	0.86	0.88	0.88	0.89	0.85	0.88	0.93
WY 80-81	0.87	0.86	0.87	0.88	0.87	0.87	0.88
WY 81-82	0.86	0.84	0.84	0.86	0.86	0.85	0.87
WY 82-83	0.85	0.87	0.87	0.87	0.84	0.86	0.90
WY 83-84	0.89	0.90	0.90	0.89	0.89	0.89	0.93
WY 84-85	0.89	0.89	0.89	0.91	0.89	0.90	0.88
WY 85-86	0.94	0.93	0.93	0.94	0.93	0.94	0.90
WY 86-87	0.80	0.80	0.80	0.82	0.80	0.81	0.88
WY 87-88	0.71	0.73	0.74	0.73	0.71	0.72	0.84
WY 88-89	0.70	0.71	0.73	0.71	0.70	0.70	0.77
WY 89-90	0.87	0.84	0.85	0.88	0.87	0.87	0.84
Mean	0.83	0.84	0.85	0.85	0.83	0.84	0.89
Abs Mean	0.83	0.84	0.85	0.85	0.83	0.84	0.89
WY 71-90	0.85	0.86	0.86	0.87	0.85	0.86	0.9

Coefficient of Determination D! (R^2)

Period	Objective Function						
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Manual
WY 71-72	0.93	0.95	0.95	0.94	0.93	0.94	0.95
WY 72-73	0.87	0.90	0.89	0.88	0.87	0.88	0.95
WY 73-74	0.89	0.92	0.91	0.91	0.89	0.91	0.97
WY 74-75	0.87	0.91	0.91	0.90	0.87	0.89	0.95
WY 75-76	0.85	0.90	0.89	0.86	0.85	0.87	0.94
WY 76-77	0.87	0.91	0.90	0.89	0.87	0.89	0.95
WY 77-78	0.88	0.86	0.86	0.87	0.88	0.88	0.95
WY 78-79	0.86	0.92	0.92	0.89	0.87	0.89	0.95
WY 79-80	0.90	0.94	0.94	0.93	0.90	0.93	0.96
WY 80-81	0.89	0.91	0.91	0.90	0.89	0.90	0.91
WY 81-82	0.88	0.90	0.90	0.89	0.88	0.89	0.93
WY 82-83	0.89	0.92	0.92	0.91	0.89	0.91	0.93
WY 83-84	0.89	0.90	0.90	0.89	0.90	0.90	0.94
WY 84-85	0.92	0.94	0.93	0.93	0.92	0.93	0.93
WY 85-86	0.94	0.96	0.96	0.95	0.94	0.95	0.92
WY 86-87	0.88	0.92	0.92	0.90	0.88	0.90	0.94
WY 87-88	0.83	0.90	0.89	0.85	0.83	0.85	0.94
WY 88-89	0.87	0.91	0.91	0.89	0.87	0.89	0.95
WY 89-90	0.90	0.92	0.92	0.92	0.91	0.92	0.93
Mean	0.88	0.91	0.91	0.90	0.88	0.90	0.94
WY 71-90	0.88	0.91	0.91	0.89	0.88	0.89	0.93

Modified Nash-Sutcliffe Efficiency EOPT!

Period	Objective Function						
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Manual
WY 71-72	0.89	0.84	0.85	0.88	0.88	0.86	0.88
WY 72-73	0.56	0.59	0.60	0.59	0.55	0.55	0.80
WY 73-74	0.88	0.88	0.88	0.89	0.88	0.87	0.92
WY 74-75	0.61	0.55	0.56	0.63	0.60	0.59	0.76
WY 75-76	0.79	0.77	0.77	0.80	0.79	0.78	0.92
WY 76-77	0.43	0.50	0.50	0.46	0.42	0.43	0.68
WY 77-78	0.70	0.69	0.70	0.72	0.70	0.70	0.92
WY 78-79	0.55	0.57	0.58	0.58	0.54	0.55	0.69
WY 79-80	0.64	0.66	0.67	0.70	0.63	0.67	0.84
WY 80-81	0.73	0.72	0.74	0.75	0.73	0.73	0.86
WY 81-82	0.71	0.63	0.64	0.69	0.70	0.66	0.78
WY 82-83	0.63	0.66	0.66	0.65	0.62	0.62	0.82
WY 83-84	0.84	0.85	0.86	0.85	0.84	0.83	0.86
WY 84-85	0.72	0.69	0.70	0.74	0.71	0.70	0.78
WY 85-86	0.84	0.80	0.81	0.85	0.83	0.82	0.88
WY 86-87	0.53	0.51	0.52	0.56	0.52	0.52	0.72
WY 87-88	0.39	0.40	0.41	0.42	0.38	0.38	0.63
WY 88-89	0.33	0.36	0.38	0.35	0.32	0.32	0.53
WY 89-90	0.69	0.65	0.67	0.71	0.68	0.68	0.74
Mean	0.66	0.65	0.66	0.67	0.65	0.65	0.79
Abs Mean	0.66	0.65	0.66	0.67	0.65	0.65	0.79
WY 71-90	0.68	0.67	0.68	0.7	0.67	0.67	0.82

Volume Error dV/V

Period	Objective Function						
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Manual
WY 71-72	-0.04	-0.11	-0.10	-0.06	-0.05	-0.08	-0.05
WY 72-73	-0.25	-0.24	-0.23	-0.23	-0.26	-0.26	-0.11
WY 73-74	0.00	-0.04	-0.03	-0.01	-0.01	-0.03	0.04
WY 74-75	-0.23	-0.28	-0.27	-0.23	-0.23	-0.26	-0.15
WY 75-76	-0.05	-0.11	-0.11	-0.06	-0.06	-0.08	0.01
WY 76-77	-0.32	-0.29	-0.29	-0.31	-0.33	-0.33	-0.18
WY 77-78	-0.15	-0.13	-0.12	-0.13	-0.15	-0.15	-0.01
WY 78-79	-0.24	-0.25	-0.24	-0.23	-0.25	-0.25	-0.16
WY 79-80	-0.22	-0.22	-0.21	-0.19	-0.22	-0.21	-0.08
WY 80-81	-0.14	-0.14	-0.13	-0.13	-0.14	-0.14	-0.03
WY 81-82	-0.15	-0.21	-0.20	-0.17	-0.16	-0.19	-0.08
WY 82-83	-0.22	-0.21	-0.21	-0.22	-0.23	-0.23	-0.09
WY 83-84	-0.05	-0.04	-0.04	-0.04	-0.05	-0.06	0.07
WY 84-85	-0.17	-0.20	-0.19	-0.17	-0.18	-0.19	-0.10
WY 85-86	-0.09	-0.14	-0.13	-0.09	-0.10	-0.12	-0.02
WY 86-87	-0.27	-0.29	-0.28	-0.27	-0.28	-0.29	-0.16
WY 87-88	-0.32	-0.33	-0.33	-0.32	-0.33	-0.34	-0.20
WY 88-89	-0.37	-0.35	-0.34	-0.37	-0.37	-0.38	-0.24
WY 89-90	-0.18	-0.19	-0.18	-0.18	-0.18	-0.20	-0.10
Mean	-0.18	-0.20	-0.19	-0.18	-0.19	-0.20	-0.09
Abs Mean	0.18	0.20	0.19	0.18	0.19	0.20	0.10
WY 71-90	-0.17	-0.19	-0.18	-0.17	-0.18	-0.19	-0.08

Appendix B.3

Mean Monthly Statistics for All Calibrations

Mean Monthly Statistics for Multi-Objective Automatic Calibrations of Coquitlam Lake Watershed

Nash-Sutcliffe Efficiency E!

	Period	Objective Function								
		SLS	HMLE	EOPT'	E!	LAD	D!	EOPT!	EOPT	BC Hydro
Arithmetic Mean	October	-5.38	-0.02	-4.76	-5.40	-4.21	-0.39	-5.64	-5.34	-5.68
	November	0.71	0.59	0.71	0.71	0.72	0.60	0.70	0.71	0.66
	December	0.64	0.40	0.66	0.64	0.65	0.48	0.63	0.64	0.58
	January	0.68	0.42	0.67	0.68	0.65	0.48	0.68	0.68	0.67
	February	0.57	0.28	0.55	0.57	0.54	0.39	0.56	0.57	0.47
	March	0.40	0.39	0.45	0.40	0.48	0.44	0.40	0.40	0.38
	April	0.22	-0.02	0.28	0.22	0.29	-0.05	0.22	0.22	0.07
	May	0.32	-0.01	0.37	0.32	0.37	0.00	0.31	0.32	0.17
	June	-0.40	-1.06	-0.27	-0.39	-0.23	-0.69	-0.39	-0.38	-0.39
	July	-0.04	-0.58	-0.05	-0.04	-0.05	-0.30	-0.05	-0.04	-0.20
	August	-1.88	-1.12	-1.25	-1.92	-1.52	-0.44	-2.46	-1.81	-3.45
	September	-4.62	-1.78	-2.92	-4.71	-3.35	-2.70	-6.50	-4.31	-3.93
Absolute Mean	October	6.89	1.08	6.27	6.91	5.72	1.55	7.14	6.85	7.14
	November	0.71	0.59	0.71	0.71	0.72	0.60	0.70	0.71	0.66
	December	0.77	0.60	0.74	0.77	0.72	0.61	0.78	0.77	0.80
	January	0.68	0.44	0.67	0.68	0.65	0.49	0.68	0.68	0.67
	February	0.57	0.51	0.56	0.57	0.56	0.45	0.57	0.57	0.56
	March	0.56	0.51	0.59	0.56	0.56	0.49	0.56	0.56	0.65
	April	0.54	0.68	0.50	0.54	0.49	0.71	0.53	0.54	0.60
	May	0.42	0.38	0.49	0.42	0.50	0.39	0.41	0.43	0.50
	June	0.98	1.51	0.97	0.97	0.99	1.19	0.95	0.95	0.88
	July	1.17	1.14	1.26	1.17	1.22	1.23	1.16	1.16	1.12
	August	2.31	1.47	1.72	2.35	1.94	0.94	2.83	2.23	3.71
	September	5.29	2.29	3.58	5.37	3.95	3.51	7.11	4.96	4.40

Coefficient of Determination D! (R^2)

	Period	Objective Function								
		SLS	HMLE	EOPT'	E!	LAD	D!	EOPT!	EOPT	BC Hydro
Arithmetic Mean	October	0.81	0.78	0.81	0.81	0.82	0.82	0.81	0.81	0.81
	November	0.76	0.77	0.76	0.76	0.77	0.76	0.76	0.76	0.75
	December	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.73
	January	0.83	0.82	0.83	0.83	0.83	0.82	0.82	0.83	0.81
	February	0.70	0.70	0.71	0.70	0.72	0.69	0.70	0.70	0.68
	March	0.72	0.74	0.74	0.72	0.74	0.72	0.72	0.73	0.72
	April	0.56	0.61	0.59	0.56	0.61	0.57	0.56	0.56	0.56
	May	0.59	0.59	0.61	0.59	0.61	0.59	0.58	0.59	0.58
	June	0.59	0.59	0.61	0.59	0.60	0.59	0.59	0.59	0.58
	July	0.77	0.75	0.77	0.77	0.77	0.77	0.77	0.77	0.74
	August	0.57	0.58	0.55	0.57	0.54	0.56	0.56	0.56	0.50
	September	0.59	0.64	0.57	0.59	0.57	0.62	0.59	0.60	0.56

Modified Nash-Sutcliffe Efficiency EOPT!

	Period	Objective Function								
		SLS	HMLE	EOPT'	E!	LAD	D!	EOPT!	EOPT	BC Hydro
Arithmetic Mean	October	-5.79	-0.47	-5.13	-5.82	-4.60	-0.90	-6.09	-5.75	-6.09
	November	0.58	0.26	0.60	0.58	0.61	0.28	0.57	0.58	0.50
	December	0.50	-0.02	0.52	0.50	0.51	0.13	0.48	0.50	0.44
	January	0.40	-0.06	0.39	0.40	0.35	0.06	0.40	0.40	0.37
	February	0.36	-0.18	0.34	0.36	0.31	0.02	0.36	0.36	0.24
	March	0.11	0.05	0.20	0.11	0.24	0.15	0.11	0.11	0.15
	April	0.04	-0.35	0.10	0.04	0.11	-0.37	0.04	0.04	-0.11
	May	0.14	-0.36	0.21	0.14	0.21	-0.36	0.14	0.15	0.01
	June	-0.71	-1.50	-0.53	-0.70	-0.49	-1.08	-0.70	-0.68	-0.66
	July	-0.36	-1.10	-0.30	-0.35	-0.31	-0.63	-0.39	-0.35	-0.51
	August	-2.66	-1.68	-1.78	-2.71	-2.04	-0.98	-3.38	-2.59	-4.13
	September	-5.21	-2.36	-3.35	-5.31	-3.78	-3.19	-7.20	-4.89	-4.46
Absolute Mean	October	7.16	0.83	6.52	7.19	5.95	1.46	7.44	7.12	7.41
	November	0.58	0.31	0.60	0.58	0.61	0.31	0.57	0.58	0.52
	December	0.64	0.45	0.63	0.64	0.63	0.45	0.65	0.64	0.67
	January	0.51	0.38	0.49	0.51	0.48	0.35	0.51	0.51	0.50
	February	0.49	0.51	0.50	0.49	0.48	0.40	0.48	0.49	0.54
	March	0.54	0.42	0.47	0.54	0.49	0.40	0.53	0.53	0.60
	April	0.58	0.72	0.53	0.58	0.53	0.72	0.58	0.58	0.70
	May	0.40	0.48	0.44	0.40	0.44	0.40	0.39	0.40	0.48
	June	1.17	1.76	1.09	1.16	1.10	1.34	1.14	1.14	1.04
	July	1.12	1.35	1.18	1.11	1.13	1.17	1.12	1.11	1.19
	August	2.84	1.75	2.05	2.88	2.28	1.26	3.50	2.74	4.26
	September	5.47	2.49	3.75	5.56	4.13	3.67	7.42	5.15	4.71

Volume Error dV/V

	Period	Objective Function								
		SLS	HMLE	EOPT'	E!	LAD	D!	EOPT!	EOPT	BC Hydro
Arithmetic Mean	October	-0.29	0.39	-0.20	-0.29	-0.19	0.19	-0.34	-0.28	-0.30
	November	-0.07	0.33	-0.03	-0.07	0.02	0.31	-0.09	-0.07	-0.11
	December	-0.01	0.42	0.04	-0.01	0.10	0.35	-0.03	-0.01	-0.02
	January	0.08	0.47	0.13	0.08	0.18	0.41	0.07	0.08	0.09
	February	0.02	0.46	0.09	0.02	0.15	0.37	0.01	0.02	0.10
	March	-0.15	0.34	-0.07	-0.15	0.00	0.26	-0.15	-0.15	-0.03
	April	-0.05	0.33	0.03	-0.05	0.07	0.32	-0.05	-0.04	0.08
	May	0.02	0.36	0.11	0.02	0.13	0.36	0.01	0.02	0.12
	June	0.06	0.44	0.18	0.06	0.19	0.39	0.03	0.06	0.13
	July	0.00	0.52	0.14	0.00	0.13	0.32	-0.07	-0.02	0.00
	August	-0.66	0.55	-0.41	-0.67	-0.40	-0.35	-0.80	-0.66	-0.57
	September	-0.45	0.58	-0.22	-0.46	-0.22	-0.17	-0.58	-0.44	-0.35
Absolute Mean	October	0.42	0.44	0.37	0.42	0.39	0.51	0.45	0.41	0.41
	November	0.13	0.33	0.12	0.13	0.11	0.31	0.14	0.13	0.16
	December	0.14	0.42	0.14	0.14	0.15	0.35	0.15	0.14	0.14
	January	0.28	0.48	0.28	0.28	0.29	0.42	0.28	0.28	0.30
	February	0.21	0.46	0.22	0.21	0.23	0.37	0.21	0.21	0.23
	March	0.29	0.34	0.25	0.29	0.24	0.29	0.29	0.29	0.23
	April	0.18	0.33	0.19	0.18	0.18	0.32	0.18	0.18	0.18
	May	0.18	0.36	0.17	0.18	0.16	0.36	0.18	0.18	0.16
	June	0.31	0.44	0.27	0.30	0.26	0.39	0.31	0.30	0.28
	July	0.31	0.52	0.25	0.31	0.26	0.32	0.34	0.31	0.31
	August	0.78	0.56	0.53	0.79	0.52	0.54	0.91	0.78	0.68
	September	0.59	0.58	0.43	0.60	0.43	0.49	0.70	0.58	0.53

Mean Monthly Statistics for Multi-Objective Automatic Calibrations of Coquitlam Lake Watershed Using an Expanded Parameter Space

Nash-Sutcliffe Efficiency E!

	Period	SLS	HMLE	EOPT'	E!	LAD	DI	EOPT!	EOPT	BC Hydro
Arithmetic Mean	October	-6.78	-0.07	-4.33	-4.94	-3.72	-0.33	-5.03	-5.46	-5.68
	November	0.70	0.59	0.71	0.70	0.72	0.68	0.70	0.70	0.66
	December	0.61	0.40	0.64	0.61	0.66	0.58	0.62	0.60	0.58
	January	0.69	0.41	0.67	0.69	0.65	0.59	0.69	0.69	0.67
	February	0.55	0.28	0.56	0.57	0.55	0.50	0.57	0.56	0.47
	March	0.39	0.39	0.45	0.40	0.48	0.49	0.42	0.40	0.38
	April	0.23	-0.01	0.27	0.22	0.29	0.09	0.24	0.24	0.07
	May	0.28	0.01	0.38	0.37	0.37	0.13	0.37	0.33	0.17
	June	-0.31	-1.06	-0.24	-0.25	-0.26	-0.49	-0.23	-0.26	-0.39
	July	-0.11	-0.56	-0.03	-0.05	-0.06	-0.25	-0.08	-0.04	-0.20
	August	-5.85	-1.08	-1.24	-3.88	-1.17	-0.85	-5.84	-4.73	-3.45
	September	-12.60	-1.69	-2.95	-11.58	-2.92	-3.44	-20.48	-11.49	-3.93
Absolute Mean	October	8.25	1.11	5.84	6.44	5.22	1.60	6.53	6.95	7.14
	November	0.70	0.59	0.71	0.70	0.72	0.68	0.70	0.70	0.66
	December	0.80	0.60	0.77	0.80	0.72	0.70	0.79	0.81	0.80
	January	0.69	0.44	0.67	0.69	0.65	0.59	0.69	0.69	0.67
	February	0.55	0.51	0.56	0.57	0.57	0.53	0.57	0.56	0.56
	March	0.58	0.52	0.59	0.55	0.56	0.53	0.56	0.55	0.65
	April	0.53	0.67	0.50	0.54	0.47	0.66	0.51	0.53	0.60
	May	0.39	0.39	0.51	0.41	0.51	0.40	0.41	0.40	0.50
	June	0.82	1.53	0.97	0.85	1.03	1.03	0.78	0.86	0.88
	July	1.14	1.15	1.27	1.18	1.22	1.23	1.15	1.18	1.12
	August	6.09	1.42	1.74	4.21	1.64	1.35	6.10	5.01	3.71
	September	13.01	2.21	3.68	12.10	3.56	4.21	20.95	11.95	4.40

Coefficient of Determination D! (R^2)

	Period	SLS	HMLE	EOPT'	E!	LAD	DI	EOPT!	EOPT	BC Hydro
Arithmetic Mean	October	0.81	0.78	0.82	0.81	0.82	0.81	0.81	0.81	0.81
	November	0.76	0.77	0.76	0.76	0.77	0.76	0.76	0.76	0.75
	December	0.74	0.73	0.74	0.74	0.75	0.74	0.74	0.74	0.73
	January	0.82	0.81	0.82	0.82	0.83	0.82	0.83	0.82	0.81
	February	0.70	0.70	0.70	0.70	0.72	0.70	0.70	0.70	0.68
	March	0.73	0.74	0.74	0.72	0.74	0.72	0.72	0.72	0.72
	April	0.56	0.62	0.58	0.54	0.60	0.56	0.55	0.55	0.56
	May	0.59	0.60	0.60	0.57	0.61	0.58	0.57	0.59	0.58
	June	0.58	0.60	0.61	0.59	0.61	0.60	0.59	0.59	0.58
	July	0.76	0.77	0.77	0.78	0.77	0.79	0.77	0.78	0.74
	August	0.54	0.55	0.56	0.56	0.54	0.55	0.55	0.55	0.50
	September	0.54	0.62	0.57	0.57	0.57	0.56	0.55	0.55	0.56

Modified Nash-Sutcliffe Efficiency EOPT!

	Period	SLS	HMLE	EOPT'	E!	LAD	DI	EOPT!	EOPT	BC Hydro
Arithmetic Mean	October	8.19	-0.52	-4.69	-5.40	-4.10	-0.81	-5.58	-5.95	-6.09
	November	0.56	0.26	0.60	0.57	0.61	0.45	0.56	0.58	0.50
	December	0.47	-0.02	0.50	0.46	0.51	0.35	0.47	0.46	0.44
	January	0.41	-0.07	0.39	0.42	0.36	0.25	0.42	0.42	0.37
	February	0.35	-0.18	0.36	0.37	0.32	0.19	0.38	0.37	0.24
	March	0.10	0.05	0.20	0.12	0.25	0.24	0.14	0.12	0.15
	April	0.05	-0.35	0.09	0.03	0.11	-0.18	0.06	0.06	-0.11
	May	0.09	-0.34	0.22	0.20	0.20	-0.17	0.20	0.15	0.01
	June	-0.60	-1.49	-0.49	-0.54	-0.52	-0.83	-0.52	-0.55	-0.66
	July	-0.42	-1.09	-0.27	-0.40	-0.32	-0.54	-0.47	-0.36	-0.51
	August	-6.97	-1.63	-1.81	-5.08	-1.65	-1.48	-7.26	-5.91	-4.13
	September	3.87	-2.26	-3.39	-12.49	-3.33	-3.93	6.72	3.17	-4.46
Absolute Mean	October	8.61	0.86	6.07	6.74	5.44	1.57	6.89	7.28	7.41
	November	0.56	0.31	0.60	0.57	0.61	0.45	0.56	0.58	0.52
	December	0.67	0.45	0.65	0.67	0.64	0.58	0.66	0.68	0.67
	January	0.51	0.37	0.49	0.51	0.48	0.40	0.52	0.51	0.50
	February	0.51	0.52	0.49	0.48	0.48	0.41	0.48	0.49	0.54
	March	0.55	0.42	0.45	0.52	0.48	0.42	0.51	0.53	0.60
	April	0.57	0.71	0.54	0.60	0.53	0.64	0.57	0.58	0.70
	May	0.37	0.48	0.46	0.35	0.46	0.38	0.35	0.38	0.48
	June	0.99	1.76	1.07	0.97	1.14	1.17	0.93	0.98	1.04
	July	1.13	1.35	1.17	1.15	1.14	1.19	1.13	1.14	1.19
	August	7.08	1.69	2.09	5.21	1.96	1.72	7.38	6.03	4.26
	September	13.57	2.39	3.84	12.70	3.73	4.39	21.79	12.51	4.71

Volume Error dV/V

	Period	SLS	HMLE	EOPT'	E!	LAD	DI	EOPT!	EOPT	BC Hydro
Arithmetic Mean	October	-0.45	0.39	-0.19	-0.33	-0.15	0.13	-0.45	-0.37	-0.30
	November	-0.09	0.33	-0.03	-0.08	0.02	0.22	-0.09	-0.08	-0.11
	December	-0.03	0.42	0.03	-0.05	0.09	0.23	-0.04	-0.05	-0.02
	January	0.07	0.47	0.12	0.06	0.18	0.31	0.07	0.06	0.09
	February	0.03	0.46	0.08	0.01	0.14	0.28	0.02	0.02	0.10
	March	-0.15	0.34	-0.08	-0.14	0.00	0.16	-0.13	-0.13	-0.03
	April	-0.06	0.33	0.02	0.00	0.07	0.24	0.00	-0.02	0.08
	May	0.00	0.35	0.12	0.06	0.14	0.30	0.06	0.03	0.12
	June	0.04	0.43	0.18	0.04	0.21	0.33	0.01	0.04	0.13
	July	-0.08	0.51	0.13	-0.11	0.16	0.26	-0.18	-0.08	0.00
	August	-1.02	0.53	-0.45	-1.09	-0.36	-0.46	-1.31	-1.07	-0.57
	September	-0.73	0.57	-0.24	-0.80	-0.19	-0.22	-1.06	-0.76	-0.35
Absolute Mean	October	0.53	0.45	0.36	0.45	0.38	0.48	0.54	0.49	0.41
	November	0.14	0.33	0.11	0.13	0.11	0.22	0.14	0.13	0.16
	December	0.14	0.42	0.14	0.15	0.14	0.23	0.15	0.15	0.14
	January	0.28	0.48	0.28	0.27	0.29	0.34	0.27	0.27	0.30
	February	0.20	0.46	0.21	0.20	0.23	0.30	0.19	0.20	0.23
	March	0.29	0.34	0.26	0.28	0.23	0.24	0.27	0.28	0.23
	April	0.18	0.33	0.18	0.19	0.19	0.27	0.18	0.18	0.18
	May	0.19	0.35	0.16	0.16	0.17	0.30	0.17	0.18	0.16
	June	0.30	0.44	0.26	0.29	0.26	0.34	0.29	0.29	0.28
	July	0.32	0.52	0.24	0.35	0.27	0.29	0.39	0.32	0.31
	August	1.12	0.55	0.57	1.20	0.48	0.63	1.41	1.18	0.68
	September	0.81	0.57	0.44	0.91	0.41	0.50	1.14	0.86	0.53

Mean Monthly Statistics for Coquitlam Seed Sensitivity Calibrations

All trials use the EOPT' Objective Function

Nash-Sutcliffe Efficiency E!

	Period	Objective Function					
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	BC Hydro
Arithmetic Mean	October	-3.91	-2.97	-4.03	-4.00	-4.00	-5.68
	November	0.72	0.72	0.71	0.72	0.71	0.66
	December	0.66	0.66	0.66	0.66	0.66	0.58
	January	0.66	0.64	0.66	0.66	0.67	0.67
	February	0.53	0.55	0.54	0.54	0.54	0.47
	March	0.46	0.48	0.47	0.46	0.47	0.38
	April	0.24	0.18	0.24	0.24	0.24	0.07
	May	0.34	0.33	0.34	0.35	0.34	0.17
	June	-0.42	-0.23	-0.39	-0.37	-0.39	-0.39
	July	-0.11	-0.06	-0.10	-0.09	-0.10	-0.20
	August	-0.52	-0.76	-0.55	-0.54	-0.54	-3.45
	September	-1.09	-3.09	-1.39	-1.25	-1.38	-3.93
Absolute Mean	October	5.41	4.46	5.54	5.50	5.51	7.14
	November	0.72	0.72	0.71	0.72	0.71	0.66
	December	0.73	0.72	0.73	0.73	0.73	0.80
	January	0.66	0.64	0.66	0.66	0.67	0.67
	February	0.57	0.56	0.56	0.57	0.56	0.56
	March	0.61	0.57	0.60	0.60	0.60	0.65
	April	0.49	0.62	0.50	0.50	0.50	0.60
	May	0.49	0.50	0.50	0.50	0.50	0.50
	June	1.14	0.99	1.11	1.09	1.12	0.88
	July	1.25	1.15	1.21	1.22	1.22	1.12
	August	1.16	1.26	1.19	1.17	1.19	3.71
	September	1.83	3.80	2.11	1.97	2.11	4.40

Coefficient of Determination D! (R^2)

	Period	Objective Function					
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	BC Hydro
Arithmetic Mean	October	0.82	0.81	0.81	0.82	0.81	0.81
	November	0.77	0.77	0.77	0.77	0.77	0.75
	December	0.75	0.74	0.74	0.74	0.75	0.73
	January	0.83	0.82	0.83	0.83	0.83	0.81
	February	0.72	0.71	0.72	0.71	0.72	0.68
	March	0.74	0.73	0.74	0.74	0.74	0.72
	April	0.58	0.57	0.58	0.58	0.58	0.56
	May	0.60	0.58	0.59	0.59	0.59	0.58
	June	0.62	0.61	0.62	0.62	0.62	0.58
	July	0.77	0.77	0.76	0.77	0.76	0.74
	August	0.58	0.56	0.57	0.58	0.58	0.50
	September	0.63	0.61	0.61	0.62	0.62	0.56

Modified Nash-Sutcliffe Efficiency EOPT!

	Period	Objective Function					
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	BC Hydro
Arithmetic Mean	October	-4.24	-3.31	-4.38	-4.33	-4.35	-6.09
	November	0.59	0.61	0.58	0.59	0.58	0.50
	December	0.52	0.52	0.52	0.52	0.52	0.44
	January	0.37	0.35	0.37	0.37	0.37	0.37
	February	0.30	0.33	0.31	0.31	0.31	0.24
	March	0.22	0.25	0.23	0.22	0.23	0.15
	April	0.06	-0.01	0.06	0.06	0.06	-0.11
	May	0.16	0.16	0.18	0.18	0.18	0.01
	June	-0.71	-0.49	-0.67	-0.66	-0.68	-0.66
	July	-0.38	-0.33	-0.39	-0.36	-0.39	-0.51
	August	-0.91	-1.35	-0.93	-0.94	-0.93	-4.13
	September	-1.42	-3.60	-1.76	-1.61	-1.76	-4.46
Absolute Mean	October	5.62	4.66	5.76	5.72	5.73	7.41
	November	0.59	0.61	0.58	0.59	0.58	0.52
	December	0.62	0.61	0.63	0.63	0.63	0.67
	January	0.49	0.48	0.49	0.49	0.49	0.50
	February	0.52	0.47	0.51	0.51	0.51	0.54
	March	0.54	0.51	0.53	0.52	0.53	0.60
	April	0.56	0.64	0.56	0.57	0.57	0.70
	May	0.43	0.45	0.45	0.45	0.45	0.48
	June	1.26	1.10	1.23	1.21	1.23	1.04
	July	1.18	1.10	1.13	1.14	1.13	1.19
	August	1.40	1.65	1.46	1.42	1.46	4.26
	September	1.96	4.03	2.26	2.13	2.27	4.71

Volume Error dV/V

	Period	Objective Function					
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	BC Hydro
Arithmetic Mean	October	-0.16	-0.15	-0.18	-0.16	-0.19	-0.30
	November	-0.04	-0.04	-0.06	-0.04	-0.06	-0.11
	December	0.05	0.03	0.04	0.04	0.04	-0.02
	January	0.15	0.15	0.15	0.14	0.15	0.09
	February	0.12	0.11	0.12	0.11	0.12	0.10
	March	-0.03	0.00	-0.02	-0.03	-0.03	-0.03
	April	0.06	0.12	0.06	0.06	0.06	0.08
	May	0.14	0.16	0.14	0.14	0.14	0.12
	June	0.23	0.21	0.22	0.22	0.22	0.13
	July	0.20	0.11	0.20	0.19	0.20	0.00
	August	-0.24	-0.47	-0.24	-0.26	-0.25	-0.57
	September	-0.11	-0.34	-0.14	-0.13	-0.15	-0.35
Absolute Mean	October	0.33	0.34	0.35	0.34	0.34	0.41
	November	0.13	0.12	0.13	0.13	0.13	0.16
	December	0.15	0.14	0.14	0.14	0.14	0.14
	January	0.30	0.29	0.30	0.29	0.30	0.30
	February	0.23	0.22	0.23	0.23	0.23	0.23
	March	0.24	0.23	0.24	0.24	0.24	0.23
	April	0.19	0.19	0.18	0.19	0.18	0.18
	May	0.18	0.17	0.17	0.17	0.17	0.16
	June	0.29	0.26	0.29	0.28	0.29	0.28
	July	0.27	0.27	0.29	0.28	0.29	0.31
	August	0.39	0.59	0.39	0.40	0.39	0.68
	September	0.34	0.51	0.37	0.36	0.37	0.53

Mean Monthly Statistics for Additional Coquitlam Seed Sensitivity Calibrations

All trials use the HMLE Objective Function

Nash-Sutcliffe Efficiency E!

	Period	Objective Function					
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	BC Hydro
Arithmetic Mean	October	-0.80	-0.50	-0.52	-0.80	-0.81	-5.68
	November	0.61	0.63	0.63	0.61	0.62	0.66
	December	0.42	0.44	0.44	0.41	0.42	0.58
	January	0.42	0.42	0.42	0.41	0.41	0.67
	February	0.27	0.27	0.27	0.26	0.26	0.47
	March	0.33	0.33	0.32	0.33	0.32	0.38
	April	-0.12	-0.23	-0.23	-0.11	-0.18	0.07
	May	-0.09	-0.32	-0.32	-0.08	-0.16	0.17
	June	-1.36	-1.54	-1.54	-1.28	-1.48	-0.39
	July	-0.76	-0.76	-0.76	-0.68	-0.80	-0.20
	August	-1.23	-1.18	-1.19	-1.18	-1.21	-3.45
	September	-1.72	-1.90	-1.91	-1.78	-1.66	-3.93
Absolute Mean	October	1.96	1.69	1.72	1.96	1.98	7.14
	November	0.61	0.63	0.63	0.61	0.62	0.66
	December	0.60	0.59	0.60	0.60	0.60	0.80
	January	0.45	0.45	0.46	0.44	0.45	0.67
	February	0.54	0.53	0.53	0.54	0.54	0.56
	March	0.52	0.51	0.51	0.52	0.51	0.65
	April	0.77	0.88	0.88	0.77	0.83	0.60
	May	0.37	0.43	0.44	0.38	0.38	0.50
	June	1.72	1.72	1.73	1.66	1.76	0.88
	July	1.21	1.20	1.20	1.20	1.23	1.12
	August	1.54	1.56	1.57	1.53	1.54	3.71
	September	2.23	2.42	2.43	2.28	2.18	4.40

Coefficient of Determination D! (R^2)

	Period	Objective Function					
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	BC Hydro
Arithmetic Mean	October	0.80	0.81	0.81	0.80	0.80	0.81
	November	0.77	0.78	0.78	0.77	0.77	0.75
	December	0.74	0.74	0.74	0.75	0.74	0.73
	January	0.84	0.84	0.84	0.84	0.84	0.81
	February	0.74	0.74	0.74	0.74	0.74	0.68
	March	0.72	0.73	0.73	0.72	0.72	0.72
	April	0.59	0.58	0.57	0.59	0.58	0.56
	May	0.55	0.55	0.55	0.56	0.55	0.58
	June	0.58	0.58	0.58	0.58	0.57	0.58
	July	0.74	0.76	0.76	0.75	0.74	0.74
	August	0.56	0.59	0.59	0.57	0.55	0.50
	September	0.66	0.70	0.70	0.68	0.66	0.56

Modified Nash-Sutcliffe Efficiency EOPT!

	Period	Objective Function					
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	BC Hydro
Arithmetic Mean	October	-1.24	-0.92	-0.94	-1.24	-1.25	-6.09
	November	0.31	0.34	0.34	0.31	0.32	0.50
	December	0.01	0.04	0.05	0.00	0.01	0.44
	January	-0.07	-0.06	-0.06	-0.08	-0.08	0.37
	February	-0.18	-0.18	-0.18	-0.19	-0.20	0.24
	March	-0.02	-0.04	-0.05	-0.03	-0.05	0.15
	April	-0.46	-0.60	-0.61	-0.47	-0.54	-0.11
	May	-0.46	-0.77	-0.77	-0.46	-0.56	0.01
	June	-1.87	-2.10	-2.10	-1.78	-2.01	-0.66
	July	-1.39	-1.37	-1.37	-1.26	-1.45	-0.51
	August	-1.82	-1.76	-1.76	-1.75	-1.79	-4.13
	September	-2.28	-2.46	-2.47	-2.34	-2.20	-4.46
Absolute Mean	October	1.77	1.50	1.53	1.77	1.81	7.41
	November	0.35	0.38	0.38	0.35	0.36	0.52
	December	0.48	0.47	0.47	0.48	0.49	0.67
	January	0.42	0.41	0.41	0.42	0.43	0.50
	February	0.58	0.56	0.56	0.58	0.58	0.54
	March	0.46	0.45	0.45	0.45	0.46	0.60
	April	0.83	0.93	0.94	0.83	0.88	0.70
	May	0.53	0.77	0.77	0.53	0.58	0.48
	June	2.03	2.12	2.12	1.95	2.11	1.04
	July	1.46	1.41	1.41	1.38	1.49	1.19
	August	1.87	1.83	1.84	1.81	1.86	4.26
	September	2.47	2.63	2.64	2.53	2.39	4.71

Volume Error dV/V

	Period	Objective Function					
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	BC Hydro
Arithmetic Mean	October	0.33	0.33	0.32	0.33	0.32	-0.30
	November	0.30	0.29	0.28	0.30	0.29	-0.11
	December	0.41	0.40	0.39	0.41	0.41	-0.02
	January	0.47	0.47	0.47	0.48	0.47	0.09
	February	0.45	0.45	0.45	0.45	0.45	0.10
	March	0.36	0.37	0.37	0.36	0.37	-0.03
	April	0.35	0.38	0.38	0.35	0.36	0.08
	May	0.38	0.45	0.45	0.38	0.40	0.12
	June	0.50	0.56	0.56	0.49	0.54	0.13
	July	0.63	0.61	0.61	0.58	0.65	0.00
	August	0.58	0.55	0.55	0.56	0.57	-0.57
	September	0.56	0.56	0.56	0.56	0.54	-0.35
Absolute Mean	October	0.44	0.42	0.42	0.44	0.44	0.41
	November	0.30	0.29	0.28	0.30	0.29	0.16
	December	0.41	0.40	0.39	0.41	0.41	0.14
	January	0.49	0.48	0.48	0.49	0.49	0.30
	February	0.45	0.45	0.45	0.46	0.46	0.23
	March	0.36	0.37	0.37	0.36	0.37	0.23
	April	0.35	0.38	0.38	0.35	0.36	0.18
	May	0.38	0.45	0.45	0.38	0.40	0.16
	June	0.50	0.56	0.56	0.49	0.54	0.28
	July	0.63	0.61	0.61	0.58	0.65	0.31
	August	0.59	0.57	0.58	0.57	0.58	0.68
	September	0.56	0.56	0.56	0.56	0.54	0.53

Mean Monthly Statistics for Multi-Objective Automatic Calibrations of Illecillewaet River Watreshed

Nash-Sutcliffe Efficiency E!

	Period	Objective Function					
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Manual
Arithmetic Mean	October	-14.95	-21.03	-8.60	-12.17	-21.83	-3.92
	November	-35.50	-35.99	-19.53	-16.16	-52.29	-6.65
	December	-13.80	-12.66	-8.32	-7.93	-29.16	-7.19
	January	-41.53	-28.64	-28.16	-21.63	-73.90	-21.75
	February	-60.92	-36.45	-40.82	-30.81	-110.63	-39.06
	March	-133.05	-89.33	-64.21	-91.45	-210.04	-21.58
	April	-0.91	-1.33	-0.46	-1.20	-1.87	0.10
	May	0.01	0.19	0.33	0.11	-0.06	0.50
	June	0.06	0.40	0.28	0.40	0.34	0.31
	July	-0.56	-0.10	-0.38	0.01	-0.34	-0.69
	August	-2.23	-2.16	-0.74	-0.64	-2.86	-0.62
	September	-3.27	-3.47	-1.69	-1.29	-3.52	-0.27
Absolute Mean	October	15.00	21.09	8.75	12.30	21.88	4.01
	November	35.85	36.32	19.98	16.54	52.56	6.87
	December	13.90	12.73	8.51	8.12	29.16	7.19
	January	41.53	28.64	28.16	21.63	73.90	21.75
	February	60.92	36.45	40.82	30.84	110.63	39.06
	March	133.06	89.41	64.29	91.45	210.04	21.64
	April	1.22	1.60	0.94	1.48	2.09	0.75
	May	0.63	0.61	0.57	0.66	0.72	0.66
	June	0.53	0.46	0.39	0.48	0.48	0.44
	July	1.01	0.71	1.03	0.68	0.84	1.06
	August	2.49	2.34	1.15	1.17	3.01	1.21
	September	3.35	3.55	1.85	1.55	3.61	0.86

Coefficient of Determination D! (R^2)

	Period	Objective Function					
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Manual
Arithmetic Mean	October	0.59	0.55	0.57	0.53	0.56	0.47
	November	0.75	0.74	0.74	0.74	0.73	0.69
	December	0.67	0.67	0.67	0.67	0.67	0.67
	January	0.33	0.34	0.36	0.32	0.32	0.32
	February	0.30	0.28	0.30	0.26	0.28	0.28
	March	0.40	0.42	0.40	0.42	0.41	0.43
	April	0.77	0.78	0.77	0.82	0.79	0.85
	May	0.79	0.77	0.77	0.78	0.78	0.82
	June	0.60	0.57	0.55	0.59	0.57	0.75
	July	0.52	0.60	0.57	0.60	0.58	0.57
	August	0.61	0.64	0.65	0.67	0.60	0.71
	September	0.76	0.74	0.76	0.72	0.75	0.72

Modified Nash-Sutcliffe Efficiency EOPT!

	Period	Objective Function					
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Manual
Arithmetic Mean	October	-15.68	-21.85	-9.12	-12.76	-22.67	-4.21
	November	-35.92	-36.42	-19.80	-16.49	-52.79	-6.91
	December	-14.17	-13.03	-8.59	-8.19	-29.71	-7.56
	January	-42.10	-29.13	-28.64	-22.06	-74.71	-22.13
	February	-61.51	-36.92	-41.31	-31.24	-111.46	-39.49
	March	-133.62	-89.77	-64.62	-91.90	-210.78	-21.95
	April	-1.34	-1.78	-0.78	-1.66	-2.44	-0.13
	May	-0.19	0.01	0.21	-0.10	-0.30	0.35
	June	-0.07	0.34	0.20	0.35	0.27	0.17
	July	-0.72	-0.22	-0.52	-0.09	-0.49	-0.88
	August	-2.52	-2.47	-0.96	-0.84	-3.19	-0.82
	September	-3.84	-4.06	-2.13	-1.67	-4.11	-0.49
Absolute Mean	October	15.70	21.90	9.22	12.84	22.69	4.27
	November	36.22	36.69	20.17	16.80	53.01	7.08
	December	14.25	13.09	8.76	8.35	29.71	7.56
	January	42.10	29.13	28.64	22.06	74.71	22.13
	February	61.51	36.92	41.31	31.28	111.46	39.49
	March	133.62	89.82	64.67	91.90	210.78	21.98
	April	1.56	1.96	1.09	1.85	2.53	0.80
	May	0.73	0.65	0.58	0.70	0.79	0.57
	June	0.55	0.43	0.37	0.45	0.44	0.46
	July	1.09	0.74	1.06	0.74	0.90	1.18
	August	2.72	2.60	1.27	1.27	3.30	1.31
	September	3.91	4.14	2.21	1.80	4.19	0.93

Volume Error dV/V

	Period	Objective Function					
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Manual
Arithmetic Mean	October	-0.73	-0.82	-0.52	-0.60	-0.83	-0.16
	November	-0.41	-0.42	-0.22	-0.29	-0.50	0.26
	December	-0.37	-0.37	-0.27	-0.26	-0.55	0.37
	January	-0.57	-0.50	-0.48	-0.43	-0.81	0.37
	February	-0.60	-0.48	-0.50	-0.43	-0.83	0.43
	March	-0.57	-0.43	-0.38	-0.44	-0.75	0.37
	April	-0.42	-0.44	-0.28	-0.46	-0.57	0.13
	May	-0.20	-0.17	-0.07	-0.21	-0.24	-0.03
	June	-0.13	0.02	0.07	0.00	-0.05	-0.13
	July	-0.10	-0.01	0.09	0.00	-0.08	-0.18
	August	-0.29	-0.29	-0.17	-0.17	-0.33	-0.19
	September	-0.57	-0.59	-0.44	-0.38	-0.59	-0.19
Absolute Mean	October	0.73	0.82	0.52	0.60	0.83	0.29
	November	0.41	0.43	0.27	0.33	0.50	0.26
	December	0.37	0.37	0.27	0.26	0.55	0.37
	January	0.57	0.50	0.48	0.43	0.81	0.37
	February	0.60	0.48	0.50	0.43	0.83	0.43
	March	0.57	0.44	0.40	0.44	0.75	0.37
	April	0.43	0.45	0.31	0.46	0.57	0.23
	May	0.20	0.18	0.11	0.21	0.24	0.15
	June	0.13	0.06	0.09	0.06	0.08	0.13
	July	0.15	0.12	0.13	0.10	0.15	0.18
	August	0.30	0.31	0.21	0.20	0.33	0.20
	September	0.57	0.59	0.44	0.38	0.59	0.22

Mean Monthly Statistics for Illecillewaet Seed Sensitivity Calibrations

All trials use the EOPT' Objective Function

Nash-Sutcliffe Efficiency E!

	Period	Objective Function						
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Manual
Arithmetic Mean	October	-21.03	-9.29	-9.50	-14.32	-21.76	-15.85	-3.92
	November	-35.99	-22.04	-22.49	-24.13	-37.62	-29.21	-6.65
	December	-12.66	-7.30	-7.52	-12.95	-13.19	-14.72	-7.19
	January	-28.64	-23.36	-23.56	-34.20	-28.43	-37.06	-21.75
	February	-36.45	-34.65	-34.24	-49.85	-34.74	-52.24	-39.06
	March	-89.33	-102.25	-99.00	-122.67	-86.96	-125.23	-21.58
	April	-1.33	-0.52	-0.57	-1.43	-1.27	-1.41	0.10
	May	0.19	0.06	0.11	0.07	0.19	0.05	0.50
	June	0.40	0.10	0.18	0.39	0.41	0.38	0.31
	July	-0.10	-0.31	-0.31	-0.06	-0.10	-0.11	-0.69
	August	-2.16	-1.08	-1.16	-1.13	-2.35	-1.64	-0.62
	September	-3.47	-1.93	-1.98	-1.81	-3.62	-2.23	-0.27
Absolute Mean	October	21.09	9.43	9.63	14.41	21.82	15.93	4.01
	November	36.32	22.43	22.89	24.52	37.92	29.59	6.87
	December	12.73	7.59	7.80	13.04	13.25	14.79	7.19
	January	28.64	23.37	23.57	34.20	28.43	37.06	21.75
	February	36.45	34.65	34.24	49.85	34.74	52.24	39.06
	March	89.41	102.29	99.05	122.67	87.05	125.25	21.64
	April	1.60	0.97	1.00	1.69	1.55	1.66	0.75
	May	0.61	0.60	0.58	0.66	0.61	0.66	0.66
	June	0.46	0.52	0.50	0.48	0.48	0.50	0.44
	July	0.71	0.82	0.81	0.73	0.70	0.73	1.06
	August	2.34	1.59	1.64	1.55	2.51	1.96	1.21
	September	3.55	2.11	2.15	1.95	3.71	2.33	0.86

Coefficient of Determination D! (R^2)

	Period	Objective Function						
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Manual
Arithmetic Mean	October	0.55	0.59	0.58	0.55	0.54	0.55	0.47
	November	0.74	0.75	0.75	0.73	0.74	0.73	0.69
	December	0.67	0.67	0.67	0.67	0.67	0.67	0.67
	January	0.34	0.33	0.33	0.32	0.34	0.32	0.32
	February	0.28	0.29	0.29	0.26	0.28	0.27	0.28
	March	0.42	0.40	0.40	0.41	0.42	0.40	0.43
	April	0.78	0.78	0.78	0.81	0.78	0.79	0.85
	May	0.77	0.80	0.80	0.78	0.77	0.78	0.82
	June	0.57	0.62	0.61	0.58	0.58	0.58	0.75
	July	0.60	0.54	0.54	0.59	0.60	0.60	0.57
	August	0.64	0.66	0.66	0.64	0.65	0.64	0.71
	September	0.74	0.75	0.75	0.73	0.74	0.74	0.72

Modified Nash-Sutcliffe Efficiency EOPT!

	Period	Objective Function						
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Manual
Arithmetic Mean	October	-21.85	-9.87	-10.08	-14.98	-22.59	-16.56	-4.21
	November	-36.42	-22.36	-22.81	-24.50	-38.06	-29.60	-6.91
	December	-13.03	-7.55	-7.78	-13.30	-13.57	-15.10	-7.56
	January	-29.13	-23.78	-23.99	-34.75	-28.92	-37.63	-22.13
	February	-36.92	-35.09	-34.68	-50.41	-35.21	-52.81	-39.49
	March	-89.77	-102.72	-99.46	-123.21	-87.39	-125.77	-21.95
	April	-1.78	-0.90	-0.95	-1.93	-1.71	-1.90	-0.13
	May	0.01	-0.13	-0.07	-0.15	0.01	-0.17	0.35
	June	0.34	-0.03	0.07	0.33	0.36	0.32	0.17
	July	-0.22	-0.45	-0.44	-0.18	-0.23	-0.24	-0.88
	August	-2.47	-1.30	-1.39	-1.36	-2.67	-1.91	-0.82
	September	-4.06	-2.37	-2.43	-2.25	-4.23	-2.72	-0.49
Absolute Mean	October	21.90	9.95	10.16	15.05	22.63	16.62	4.27
	November	36.69	22.72	23.18	24.85	38.30	29.94	7.08
	December	13.09	7.81	8.02	13.37	13.63	15.16	7.56
	January	29.13	23.78	23.99	34.75	28.92	37.63	22.13
	February	36.92	35.09	34.68	50.41	35.21	52.81	39.49
	March	89.82	102.76	99.51	123.21	87.44	125.79	21.98
	April	1.96	1.14	1.19	2.08	1.90	2.05	0.80
	May	0.65	0.69	0.65	0.72	0.65	0.72	0.57
	June	0.43	0.55	0.50	0.45	0.45	0.46	0.46
	July	0.74	0.88	0.87	0.76	0.74	0.78	1.18
	August	2.60	1.71	1.77	1.69	2.80	2.15	1.31
	September	4.14	2.46	2.51	2.34	4.31	2.80	0.93

Volume Error dV/V

	Period	Objective Function						
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Manual
Arithmetic Mean	October	-0.82	-0.58	-0.58	-0.66	-0.83	-0.70	-0.16
	November	-0.42	-0.31	-0.30	-0.34	-0.44	-0.38	0.26
	December	-0.37	-0.25	-0.25	-0.35	-0.39	-0.38	0.37
	January	-0.50	-0.42	-0.42	-0.55	-0.50	-0.57	0.37
	February	-0.48	-0.44	-0.44	-0.56	-0.47	-0.57	0.43
	March	-0.43	-0.47	-0.45	-0.54	-0.41	-0.54	0.37
	April	-0.44	-0.36	-0.36	-0.50	-0.43	-0.49	0.13
	May	-0.17	-0.18	-0.17	-0.22	-0.17	-0.22	-0.03
	June	0.02	-0.13	-0.11	-0.02	0.02	-0.03	-0.13
	July	-0.01	-0.08	-0.07	-0.02	-0.03	-0.05	-0.18
	August	-0.29	-0.21	-0.22	-0.21	-0.31	-0.26	-0.19
	September	-0.59	-0.44	-0.44	-0.44	-0.61	-0.48	-0.19
Absolute Mean	October	0.82	0.58	0.58	0.66	0.83	0.70	0.29
	November	0.43	0.32	0.32	0.36	0.44	0.39	0.26
	December	0.37	0.25	0.26	0.35	0.39	0.38	0.37
	January	0.50	0.42	0.42	0.55	0.50	0.57	0.37
	February	0.48	0.44	0.44	0.56	0.47	0.57	0.43
	March	0.44	0.47	0.46	0.54	0.43	0.54	0.37
	April	0.45	0.38	0.38	0.50	0.44	0.49	0.23
	May	0.18	0.19	0.18	0.22	0.18	0.22	0.15
	June	0.06	0.13	0.11	0.06	0.06	0.07	0.13
	July	0.12	0.13	0.13	0.12	0.13	0.13	0.18
	August	0.31	0.22	0.23	0.24	0.32	0.27	0.20
	September	0.59	0.44	0.45	0.44	0.61	0.48	0.22

Appendix B.4

Summary Results from Event-based Simulations

Storm 1 Summary Results

Objective	Peak Flow (m^3/s)	Event Volume (cms-d)	Time to Peak (hours)
Observed	601	26891	96
BC Hydro	699	29930	98
SLS	662	29597	98
EOPT'	646	28592	98
E!	663	29608	98
LAD	610	27127	98
EOPT'	662	29578	98
Mean	649	28900	98
Std Dev	22.6	1082	0.0
CoV	3%	4%	0%
Mean Error (%)	8%	7%	2%
BCH Error (%)	16%	11%	2%

Storm 2 Summary Results

Objective	Peak Flow (m^3/s)	Event Volume (cms-d)	Time to Peak (hours)
Observed	737	17559	65
BC Hydro	615	17369	66
SLS	592	17345	67
EOPT'	573	16623	67
E!	592	17345	67
LAD	537	15786	67
EOPT'	592	17326	67
Mean	577	16885	67
Std Dev	24.0	688	0.0
CoV	4%	4%	0%
Mean Error (%)	-22%	-4%	3%
BCH Error (%)	-17%	-1%	2%

Storm 3 Summary Results

Objective	Peak Flow (m^3/s)	Event Volume (cms-d)	Time to Peak (hours)
Observed	650	11421	57
BC Hydro	601	11718	58
SLS	552	11703	58
EOPT'	543	11186	58
E!	553	11707	58
LAD	514	10501	58
EOPT'	552	11686	58
Mean	543	11357	58
Std Dev	16.6	527	0.0
CoV	3%	5%	0%
Mean Error (%)	-16%	-1%	2%
BCH Error (%)	-8%	3%	2%

Storm 4 Summary Results

Objective	Peak Flow (m^3/s)	Event Volume (cms-d)	Time to Peak (hours)
Observed	716	19162	97
BC Hydro	426	18931	101
SLS	412	19056	105
EOPT'	390	18202	103
E!	412	19064	105
LAD	364	17191	102
EOPT'	411	18993	105
Mean	398	18501	104
Std Dev	21.1	817	1.4
CoV	5%	4%	1%
Mean Error (%)	-44%	-3%	7%
BCH Error (%)	-41%	-1%	4%

Storm 5 Summary Results

Objective	Peak Flow (m^3/s)	Event Volume (cms-d)	Time to Peak (hours)
Observed	528	20070	183
BC Hydro	436	17195	165
SLS	366	17637	166
EOPT'	369	16944	166
E!	367	17628	166
LAD	356	16170	166
EOPT'	366	17603	166
Mean	365	17196	166
Std Dev	5.0	645	0.0
CoV	1%	4%	0%
Mean Error (%)	-31%	-14%	-9%
BCH Error (%)	-18%	-14%	-10%