10.1 MULTIOBJECTIVE, MANIFOLDLY CONSTRAINED MONTE CARLO OPTIMIZATION AND UNCERTAINTY ESTIMATION FOR AN OPERATIONAL HYDROLOGIC FORECAST MODEL

Sean W. Fleming*, Frank A. Weber, and Scott Weston
BC Hydro, GRM, 6911 Southpoint Drive, Burnaby, British Columbia V3N 4X8 Canada
* Corresponding author: 604.528.2242, sean.fleming@bchydro.com

1. ABSTRACT

River forecasts have two broad uncertainty classes: errors associated with meteorological forecasts, and those associated with the hydrologic model. We developed a technology (dubbed Absynthe) to address the latter error class in a practical and defensible way. The technique merges the proven, Monte Carlo-based Generalized Likelihood Uncertainty Estimation (GLUE) concept for model parameter identification with: (i) multiple performance goals defined by operational and physical considerations, including matching daily, seasonal, and annual flows as well as snowpack, as expressed via individual behavioural criteria and a net likelihood function; (ii) several moving (rank-based) constraints to assure non-pathological parameter sets, containing values that are physically plausible not only for each parameter individually but also collectively; and (iii) a hard constraint on snow-free elevation bands to force the surface meteorological component of the watershed model toward correct solutions. The result is an ensemble of parameter sets reflecting model uncertainty as captured in a loosely Bayesian framework. BC Hydro will combine these with ensemble NWP weather forecasts to generate uncertainty estimates for operational hydroelectric reservoir inflow forecasts.

2. INTRODUCTION

Operational forecasts of streamflows (or derived quantities, such as reservoir inflow volumes or flood height) have two main classes of uncertainty. These are errors associated with meteorological forecasts, and those associated with the hydrologic model. The former include limitations to NWP capabilities in terms of both the accuracy and forecast horizon of meteorological forecasts, arising (broadly speaking) from deterministic chaos and data availability constraints. Additional issues like heterogeneous boundary-layer effects in mountainous terrain, downscaling of NWP forecasts to local ground locations, and the limited spatial and/or elevation coverage of an existing surface station meteorological network, can all play important roles as well. Errors associated with hydrologic modelling include model structural errors (such as a failure to represent all required terrestrial hydrologic processes for a given catchment at a level appropriate for a particular application), and parameter uncertainty as it arises in the process of model calibration.

Calibration is the procedure of adjusting the values of certain fixed parameters within a mathematical-computational watershed model such that it is optimized for application to a particular river. Further general background on calibration is provided below in Section 3. In principle, calibration might seem to be a mundane procedure having only second- or third-order importance to hydrologic forecasting. In practise, however, matters are usually quite different.

A good calibration is a key source of hydrologic forecast skill, and conversely, a poor calibration can easily render a watershed model largely useless. Indeed, calibration quality can overwhelm the completeness or accuracy of process physics representation within the model as a source of forecast skill, or lack thereof. The question of identifying an optimum parameter set also turns out to be technically challenging, due to model nonlinearity, tradeoffs between model parameters, tradeoffs between multiple objectives, tradeoffs between quantitative objectivity and soft knowledge, and data constraints. One consequence is the principle of equifinality (again discussed in further detail below in Section 3), which reflects the nonuniqueness of parameter values in terms of a given error metric. Further, hydrologic model calibration is simply one example of the broader and very sophisticated problem of mathematical optimization, and has thus attracted applications of some of the most intellectually esoteric concepts to be found in the environmental sciences, including widespread adoption of bio-mimicry approaches such as evolutionary algorithms and particle swarm optimization. At the same time, although there are clearly wrong (or at least poor) approaches to hydrologic model optimization, there does not appear to be a single right answer either. Rather, the collective experience of the hydrologic modelling community has been that the most appropriate tool for a particular job is determined by a range of application-specific constraints and goals, both theoretical and applied in nature. Additionally, uncertainties in other parts of the hydrologic forecasting chain may be propagated into parameter values. For instance, many hydrologic models contain a parameter which adjusts meteorological station data in the basin for biases relative to the (in general, imprecisely known or occasionally, completely unknown) true total precipitation in the basin. That parameter value will obviously be sensitive to the representativeness of the surface meteorological station network; further, as that representativeness is often unclear, additional uncertainty is generated in the parameter value.

The net result of all the above considerations is that there is usually considerable uncertainty regarding the optimum parameter set, and such uncertainty is often in turn a major source of hydrologic forecast error. By extension, obtaining a good quantitative understanding of that parameter uncertainty enables quantitative estimation of confidence bounds around hydrologic forecasts.

This article summarizes work completed at BC Hydro to (1) review existing watershed model optimization and uncertainty estimation technologies for applicability to BC Hydro's operational hydrological forecast system, (2) develop a new method, consisting of a series of modifications of well-accepted existing technique, and (3) apply that technology to the re-calibration of operational inflow forecast models for 21 hydroelectric reservoirs.

3. BACKGROUND

3.1 *BC Hydro operational forecast system*

BC Hydro is a power utility and British Columbia crown corporation that generates 43,000 to 54,000 GWh annually, primarily through 31 hydroelectric facilities, providing electricity to an area containing over 94% of BC's population. The responsibilities of BC Hydro's Hydrology and Technical Services group (H&TS) include hydroclimate data collection and management, operational weather and reservoir inflow forecasting for BC Hydro engineers and planners, and improvement of the systems used to accomplish these duties. We also are called upon to lead efforts to address the impacts of climatic variability and change upon reservoir inflows.

There are four main types of hydrologic regimes in the BC Hydro watersheds for which H&TS issues operational forecasts (Figure 1). (1) Rainfall-dominated (pluvial): these regimes are found mainly in low-elevation coastal regions. Inflow values closely follow those of rainfall. The maximum monthly flows usually occur in November and December, while the minimum monthly flows usually occur in July and August. (2) Snowmelt-dominated (nival): snowmelt regimes commonly occur in the mountain regions and the interior of the province. Winter precipitation dominantly falls as snow and remains in storage as snowpack until the spring melt freshet. The highest flows occur in May, June and July, while the lowest flows occur in the winter months, or very late summer. (3) Combined rainfall and snowmelt (hybrid): many watersheds near the coast of British Columbia demonstrate characteristics of both rainfall and snowmelt dominated streamflow regimes. High inflows will occur in winter months (November, December, and January) due to rainfall events, and again during the spring snowmelt (May, June, and July). Lower inflows occur in the months in-between. (4) Combined snowmelt and glacial melt (glacionival): High elevation watersheds can exhibit both snowmelt and glacial melt characteristics. These watersheds typically exhibit their highest inflows in the late summer (July, August), and lowest flows occur in the winter months when precipitation primarily falls as snow and remains in storage.

H&TS issues both short-term and long-term forecasts of reservoir inflows. Short-term inflow forecasts are made, once each working day or more often if deemed necessary, of daily mean inflow rates in cubic metres per second (cms), over a rolling 5-day forecast horizon, for each of 21 BCH projects. The forecasts are issued by H&TS by noon. There are four main components to the daily forecast process. The first is data quality control (QC). H&TS data analysts ensure that the incoming meteorological and inflow data are gap-free and of sufficient quality for use in the short-term forecast process. These data stem from BC Hydro's Data Collection Platforms (DCPs), which are automated weather and hydrometric stations with telemetry; Environment Canada's observer network; Water Survey of Canada's hydrometric network; and BC Hydro's FLOCAL inflow calculation program, which generates estimates of local (generally, unregulated) reservoir inflows on the basis of reservoir water level elevation, electric generation, and other information sources. The second component is the weather forecast. In-house meteorologists consider solutions from various NWP models run by Environment Canada and others, with bias correction performed on contract by the University of British Columbia, and then form site-specific temperature and precipitation forecasts for our basins. The third component is the inflow forecast product itself. On a basin-by-basin basis, H&TS hydrologists run a process-oriented watershed model, called the UBC Watershed Model (see below), integrating past (observed) and future (forecast) weather to generate reservoir inflow forecasts for each day over the next five days. The model is run within the River Forecast System (RFS), a software package custom-built for H&TS by Accenture, containing the UBC Watershed Model recoded in Visual Basic, a graphical user interface (GUI), and report generation features, coupled to an Oracle database (Weiss 2001). The fourth component is forecast product communication. Weather and inflow forecast summaries are compiled in tabular and graphical form, posted on the H&TS website, and disseminated via text files and emails. Note that, although many of the constituent tasks have been automated, each step in the foregoing process requires a healthy manual injection of professional judgment and experience. This applies also the forecast updating process for the UBC Watershed Model. Forecast updating is critical to increase the accuracy of operational runoff forecasts generated with physically-based hydrologic models. At BC Hydro, input and simulated state variables many be manually adjusted based on supplementary hydroclimate data available and past model performance. Calibrated parameters are left unchanged.

Long-term water supply (also called long-term seasonal or, simply, seasonal) forecasts are made at the beginning of every month, from November onward to August, for each of 25 BCH basins. Although details vary between basins, with additional steps taken (for example) to satisfy Columbia River Treaty requirements, the overall emphasis is on making forecasts of total inflow volumes in millions of cubic metres (Mcm) (but usually expressed as % of normal conditions) over the upcoming February-September period. Two techniques are used. One is statistical (the

VoDCa system, based on principal components regression; Garen, 1992). The second is the process-oriented UBC Watershed Model, run in an Ensemble Streamflow Prediction (ESP) framework, again within the RFS environment (see above). Both yield probabilistic volume predictions, consisting of a best estimate along with confidence bounds. Because the usable forecast horizon for detailed NWP models in this region is only a few days to a week at most, long-term weather forecasts are not available for the long-term water supply forecast. Instead, the reservoir inflow predictions mainly reflect initial conditions in each basin as of the forecast date, including antecedent precipitation and flows, along with snowpack for those forecast dates when it is available. For seasonal forecasts based on the UBC Watershed Model, accuracy can again heavily depend on skillful operational forecast updating, especially snowpack updating. As used at BC Hydro, snowpack "updating" is a semi-quantitative snow water equivalent (SWE) data assimilation method, which corrects model-simulated SWE when measured SWE become available. BC Hydro's statistical forecast models additionally incorporate some climate state (specifically, El Niño-Southern Oscillation) information in some cases. This study focuses on calibration of the UBC Watershed Model used for both short-term and seasonal forecasting; the VoDCa system is briefly described above only for context.

The UBC Watershed Model is a process-oriented, modestly physically based, "conceptual" streamflow model (e.g., Quick, 1995). It is largely semi-distributed in space, using elevation bands to distribute precipitation, track snowpack accumulation and depletion, and describe variability in some physical parameters like impermeable area, but other terrestrial hydrologic characteristics are treated as spatially lumped. Its current operational implementation in H&TS is fully deterministic for daily forecasting, but probabilistic (using an ESP approach, as noted above) for seasonal forecasting. The UBC Watershed Model is considered appropriate to operational forecasting in British Columbia as, relative to many other models, it offers an appropriate mixture of properties, including accommodation of mountainous & forested terrain, complex meteorological gradients, and key regional runoff sources (rainfall, snowmelt, and glacial melt); it imposes comparatively modest day-to-day operational data requirements; and it is considered to have a lengthy and solid track record in practical applications.

3.2 *Hydrologic models and calibration*

Parameter uncertainty refers to issues associated with the hydrological model calibration process.

A watershed model mathematically maps, in one way or another, inputs (meteorological data) to outputs (terrestrial hydrologic data, and in particular, streamflow rates or volumes). Process-oriented models, which form the focus of the discussion here, accomplish this task by capturing the physics of runoff generation. Note, however, that many of the concepts discussed in this article apply equally well to strictly empirical techniques like statistical or soft-computing models. Because runoff generation involves many physical and biological processes, and the goals and therefore needs of any given modelling exercise may be different (e.g., fundamental research vs. operational forecasting), many different hydrologic modelling approaches have been developed. These range from spatially lumped to fully spatially distributed representations of a watershed, and from approximate back-of-the-envelope mathematical relationships to full numerical solutions of governing partial differential equations.

In virtually all cases, however, some degree of parameterization is involved. That is, some salient characteristic of the terrestrial hydrologic environment is captured using a parameter, and the value(s) of that parameter must be determined for a particular watershed. Physical considerations, direct field measurements, and personal experiential knowledge of the performance of a given watershed model in a given hydroclimatic environment all inform this procedure. Ultimately, however, it is usually necessary to iteratively adjust various parameter values such that the observed streamflow hydrograph is matched as well as possible by the hydrologic model output for the catchment, over some training period with known weather inputs. That process is known as calibration.

There are, essentially, two basic philosophies of hydrologic model calibration. The first is manual calibration. The procedure involves manually adjusting model parameters until an acceptable fit between the observed and modelled streamflow hydrographs is obtained. The primary advantage of this technique is that it readily accommodates "soft knowledge," i.e., the experiential knowledge and intuition of the hydrologic scientist or engineer. One disadvantage is that the yin of accommodating soft knowledge is accompanied by the yang of strong sensitivity to modeller skill and opinion, that is, a lack of objectivity. Two modellers may obtain very different parameter sets for precisely the same model and data. A second, related disadvantage is that, in general, it is very unlikely that the parameter space will be thoroughly sampled in the course of a manual calibration. There is simply not enough time to explore every combination of each "free" parameter (the parameters to be adjusted in the course of calibration). There presumably is therefore a good chance that the final parameter set may not be truly optimal, or perhaps even a reasonable approximation to the optimal parameter set. Manual calibration has also been criticized as being a slow, painstaking process. The validity of this criticism is likely context-dependent. Optimization algorithms, as discussed below, might be reasonably expected to speed the calibration process for fully distributed watershed models due to the hundreds or even thousands of parameter values that may have to be adjusted, depending on how the specific model operates and the exact calibration approach adopted. On the other hand, practical experience with the application of optimization routines to environmental models will, in general, quickly educate a user that the process is in general not straightforward or quick, and for simpler hydrologic models, it may be unclear whether manual calibration is slower or faster than so-called auto-calibration. Perhaps the single most problematic aspect of manual calibration, however, is that it is inherently incapable of generating quantitative uncertainty estimates, in terms of either parameter uncertainty or the hydrologic prediction uncertainty that arises from it. Granted, some feeling for parameter identifiability can be obtained through post-hoc sensitivity analyses. However, these in general involve only minor adjustments of parameter values, in the immediate vicinity of the ostensible best set, on a parameter-by-parameter basis without addressing parameter interactions and trade-offs.

The second philosophy, optimization, involves the application of computational algorithms to determine the best parameter set, where "best" is quantitatively described by some prescribed fit or error metric. Common measures include root mean squared error (RMSE) and Nash-Sutcliffe efficiency (NE) of the modelled streamflow hydrograph, relative to the observational dataset. A broad variety of optimization algorithms are available, and the development of refined or new approaches and routines is an active field of research in mathematics. These methods are reviewed below in Section 4.1. The advantages and disadvantages of this automated calibration philosophy are essentially the opposite of those listed above for manual calibration. Automated optimization offers potential to thoroughly sample the parameter space, identify the best parameter set, and quantitatively describe parameter uncertainty as a means for both assessing parameter identifiability and numerically estimating hydrologic prediction uncertainty. The disadvantages are that it is unclear whether many optimization techniques fully meet that potential, soft knowledge is considerably more challenging to incorporate into the process, a plethora of techniques exists, and in practise, higher information technology (IT) requirements are often involved. The question of objectivity is a somewhat grey area. Subjective, or only partially objective, choices must be made regarding the optimization technique and objective function(s) to be used. However, once placed into motion, an optimization algorithm is essentially objective, and there seems little doubt that the overall process is more objective than manual calibration.

Although optimization algorithms have been applied for decades to watershed model calibration in a research context, use of these techniques within the practitioner community remains very limited. There are undoubtedly a number of reasons for this relative lack of adoption, but likely the dominant problem reflects the lack of soft knowledge in the automated process, with a common criticism being that there are no guarantees of physically and/or intuitively plausible results.

3.3 *Project context, goals, and approach*

The UBC Watershed Model as employed for inflow forecasts within the RFS was last calibrated about eight years ago. Manual calibration techniques were employed for all basins. A decision was made in 2006 to initiate a process for recalibration of the UBC Watershed Model for all basins within the RFS. The general goals of the recalibration exercise are as follows:

1. *Update the models using more recent hydrometeorological and land cover data*. An additional half-decade to decade of meteorological data are now available. Given that hydrometeorological records for many basins in British Columbia are short, this additional data may offer significant opportunity to expand the database used for calibration (and for ESP forecasting following incorporation of the recalibrated UBCWM into the operational RFS). Further, nonstationarities in hydrologically salient land surface properties, such as glacial and forest cover, are such that updating the model setup with more recent estimates of these quantities will provide a more accurate representation of the watershed, potentially yielding improved forecasts.
2. *Improve the efficiency of the calibration process.* Standardize the calibration process to minimize trial-and-error procedures. Procedures for input station selection are an important example. Calibration in general requires many hours or days on a manual basis (and even using the automated procedures in Absynthe discussed below), as calibration is not a quick and easy process. Repeating the calibration multiple times for each watershed using each of a variety of meteorological input station combinations would therefore slow down the recalibration project considerably. In contrast, a standardized protocol would require far less time (both personnel and CPU) per basin.
3. *Improve quality of the models by working to remove known biases and errors*. For example, consistent and substantial biases are known to exist in modelled flow volumes for certain months in certain basins. Further, a few errors are known to be present in the UBC Watershed Model input parameter (.WAT) files for certain basins. Such known biases and errors will be addressed in the recalibration.
4. *Obtain a measure of the prediction uncertainty arising from model parameter uncertainty*. To reiterate, many watershed model parameters, though having clear physical meaning, cannot be estimated directly by field or laboratory observation. Rather, these parameters must primarily be estimated by calibrating the model to existing watershed-scale datasets (i.e., meteorological time series to drive the model, and streamflow time series as model output). The calibration process involves finding a set of

parameter values that minimizes the difference between observed and model-simulated streamflows. Even in hindcast mode, however, various error sources will lead to imperfect hydrologic predictions and to imperfect parameter value estimates. Further, simulation error as a function of parameter value is typically a complex and nonlinear function for watershed models, and it is often challenging to find the global (rather than local) minimum. Finally, watershed models are nonunique in terms of error as a function of parameter values. This phenomenon does not necessarily imply physical nonuniqueness of watershed response, but rather refers to the fact (termed equifinality in the hydrology literature) that equally good predictions, as measured using some error metric such as the Nash-Sutcliffe efficiency for instance, can usually be produced using different parameter sets for a given model and dataset. Thus, an entire suite of parameter sets may be identified as optimal or near-optimal, yet each may actually yield a different hydrograph. Thus, it is very useful to obtain (i) quantitative estimates of parameter identifiability to support model diagnosis, and (ii) quantitative estimates of prediction uncertainty for operational forecast use.

At an early stage in the process, it was decided that a reasonable "best" approach would be initially identified. Due to time and resource constraints, experimentation with a variety of techniques (e.g., a range of algorithms for automated watershed model calibration) would not be feasible in practice. A project roadmap was therefore developed on the basis of the following considerations:

- *Evaluation of technical information and hands-on experiential insights gained from a variety of sources*. These information sources included discussions with BC Hydro personnel, such as brief interviews with forecast product users and experienced modellers; extensive literature reviews; discussions with outside personnel having significant experience with calibration of semi-distributed watershed models in British Columbia's challenging hydrologic landscapes, coupling the UBC Watershed Model to automated calibration schemes, optimization technique development, and/or automated calibration of watershed models for the purpose of operational flood and/or hydroelectric inflow forecasting in both BC and other Nordic hydrologic environments, including but not limited to discussions at various conferences; and prior personal experience with the implementation and use of optimization techniques for quantitative earth systems modeling.
- *The necessity for satisfying all four general project goals* as outlined above in this section.
- *The necessity for balancing sophistication against practical considerations*, such as methodological robustness, project and timeline risk minimization, and relative ease of implementation under time and resource constraints.

4. METHODS

4.1 *Optimization review*

As noted in Section 2, the first major project step was to perform a review and assessment of watershed model parameter optimization and uncertainty estimation techniques. A summary of results is provided below.

4.1.1 *General concepts*

Optimization has been a distinct subdiscipline of mathematics for generations, reflected in such classical puzzles as the backpack problem and the traveling salesman problem. The latter belongs to a class of mathematical problems known as NP-complete or NP-hard problems, which are notoriously difficult to solve because the computational cost involved increases exponentially with the dimension of the problem (e.g., Press et al., 1992). Indeed, the traveling salesman problem continues to be employed as a research test question to this day, for instance in quantum computing investigations.

Optimization problems arise throughout the physical, biological, and social sciences. Although a variety of different problems fall under the general rubric of optimization, including the combinatorial (discrete) optimization considered by the traveling salesman problem, and goal-seeking subject to constraints as addressed through linear programming, the basic idea in many cases (and which is considered here) is as follows.

One begins with a mathematical and computational model of some system of interest. That model contains parameters which cannot be directly measured in a meaningful way. The "best" values for the parameters are therefore found through an iterative process, whereby many values are tried in the model, and the values that give model predictions best-matched to observational data are identified as correct, or at least sufficiently good for a given purpose. As noted above, in the case of watershed modeling, the data to be matched are (in general) streamflow observations, and the overall process is often known as automated calibration; inverse modeling and parameter estimation are other synonyms.

Automated watershed calibration can be a tricky application of optimization techniques, because the optimization problem is in this case usually ill-posed. A well-posed inverse problem involves a mathematical model having the following properties: (i) a solution exists; (ii) the solution is unique; and (iii) the solution depends continuously on the data. The main problem lies with (ii) vis-à-vis the equifinality of watershed models, which refers to the fact that, in general, a given degree of error is not uniquely associated with a particular parameter set. The problem is particularly severe for high-dimensional models, which contain a large number of free parameters to be estimated during the optimization process. These facts can have profound practical implications for watershed model calibration, as summarized below.

4.1.2 *Early methods*

An early attempt at systematic optimization of watershed models was the random search, along with more sophisticated variants. The overall idea in this case is simply to randomly search the feasible parameter space – that is, the range of values thought physically plausible for each free parameter in the optimization – until a best set is found. Problems with this approach are that it is essentially a blind search, and therefore computationally inefficient; and that one can never be absolutely certain that the exact best parameter values have been found, because there is little or no imposed movement in the search toward an error minimum. The latter problem, however, might be mitigated by two considerations at a practical level: a sufficiently extensive search yielding a reasonable final calibration may prove quite adequate for a given modeling task, and subsequent work has questioned the meaningfulness and usefulness of a single, absolute-best parameter set (see discussion of GLUE methodology below). Raw random search algorithms have not been suggested for some time as a preferred approach to watershed calibration. However, due in part to their ability to sample essentially the entire feasible parameter space, similar concepts play a central role in several modern techniques, as discussed below.

Another early line of work employed various types of gradient-descent algorithm. The best-known of these within the environmental sciences, though also used much more broadly, is the Levenberg-Marquardt algorithm and its variants. The basic idea behind gradient-descent methods is as follows. An objective function, also known as a cost or error function, is defined. This function gives a measure of the mis-match between observed and model-predicted streamflows, such as RMSE. One wishes to find model parameters which minimize this cost function. As a function of all the parameters, the objective function forms an error landscape, and the best-fit set of parameter values corresponds to the lowest point in that surface. Individual techniques vary, but gradient-descent approaches are based in one way or another on the partial derivatives of the objective function with respect to the parameters; one thus uses the derivative to find the direction “downhill” from some starting position to the lowest point in the error landscape. Under certain assumptions, statistical confidence bounds on the best-fit parameter values can be directly inferred from the results (e.g., Fleming and Haggerty, 2001), and these can in turn be used to generate Monte Carlo model predictions reflecting the effects of parameter uncertainty.

A problem with gradient-descent techniques is that in complex nonlinear problems, such as most types of watershed modeling, the error terrain has a complicated form, with many individual dips and hollows apart from the global minimum. Gradient-descent techniques tend to go downhill until the first low point is found, and then stop. They therefore get trapped in such local error minima, rather than finding the global optimum. Such methods are therefore known as local search techniques.

A potential work-around is multi-start local optimization, whereby the entire procedure is independently repeated several times with different (in general, randomly chosen) initial starting points, and the best-performing of the resulting parameter estimates are selected as the optimum parameter set. As is the case for random searches, difficulties with multi-start gradient-descent algorithms include computational inefficiency and a lack of assurance that the precise global optimum has been attained.

There are three additional potential issues with gradient-descent techniques; these appear to have been at least partially overcome in more recent implementations, but are still worth noting. First, a feasible parameter space is not specified a priori. Consequently, physically unrealistic parameter values may be returned by the algorithm. However, methods have been developed to help circumvent this problem, such as regularization. Second, gradient-based optimization techniques are also occasionally criticized because analytical expressions for the derivatives of the objective function with respect to the model parameters can be difficult or impossible to obtain for computational watershed models (e.g., Duan, 2003). However, it is unclear whether this critique is still relevant to modern applications. In practice, algorithms implementing finite-difference approximations to the Jacobian matrix have progressed sufficiently to be used with success for difficult environmental modeling problems; some fairly recent examples include Fleming and Haggerty (2001) and Haggerty et al. (2001). Third, local-search methods are normally applied only to a single objective function. However, a composite objective function containing several optimization goals can be constructed, with varying degrees of success (see also Section 4.1.4 below).

Although often dismissed in the modern automated watershed calibration literature, local-type gradient-descent algorithms are perhaps more widely used in practise than ever. The Parameter ESTimation (PEST) software application uses a cutting-edge implementation of the Levenberg-Marquardt routine. It is available in both non-GUI freeware and GUI-driven proprietary forms, and in both conventional and parallelized versions. Capabilities for regularization, sensitivity analysis, and parameter confidence bound estimation are incorporated. A major advantage of this application is that it is intended to be model- and platform-independent, so that in principle it can be coupled to any pre-existing environmental modeling application without the need for software development or revision. While references to the software in the peer-reviewed literature (e.g., Gallagher and Doherty, 2007) are limited relative to some of the techniques discussed below, it is perhaps the most widely employed tool for automated calibration of environmental models in practice. The technique is used regularly in private-sector consulting and by governmental regulatory bodies, with applications ranging from hydrogeological to rainfall-runoff to water quality modeling. The freeware version is available for download through the US EPA's website (http://www.epa.gov/ceampubl/tools/pest/), which might be viewed as a de facto endorsement of its use. It is also to be noted that gradient-descent and related approaches, including but not limited to the Levenberg-Marquardt algorithm, are routinely used today to train neural networks (e.g., Li, 2005; Fleming, 2007), and the various solutions obtained using different random initial network biases and weights (in what amounts to a multi-start local-type optimization procedure) can be used to form the basis of ensemble ANN models (e.g., Hsieh, 2001).

However, as noted above, the Levenberg-Marquardt and related or analogous algorithms may not deal efficiently with the issue of local vs. global error minima, and they do not appear especially well-suited to directly addressing certain key issues that have emerged in watershed modeling, such as multiple objective functions or equifinality. Further, conversations with other modellers and IT personnel seemed to suggest that considerable project overhead might be involved in specifically coupling the UBC Watershed Model to PEST. Thus, although local optimization techniques like Levenberg-Marquardt remain in general a valuable workhorse for a wide variety of optimization problems, and the PEST package appears to be a generally well-respected and widely-used approach, local search methods are not considered further here.

4.1.3 *Global search methods*

The watershed calibration problem of finding the global optimum with some confidence, and in reasonable CPU time, was largely solved by the 1990s. Two key methods were genetic algorithms (GA) and Markov Chain Monte Carlo (MCMC). These and other successes led to the development and very widespread adoption of the shuffled complex evolution algorithm (SCE; sometimes denoted SCE-UA to denote its origin at the University of Arizona).

Genetic algorithms involve a view of optimization as a (literally) evolutionary process. GAs are reported (Wagener et al., 2004) to have been first applied to watershed model calibration by Wang (1991). Chromosomes are defined as the set of free parameters in the optimization problem, and individual genes on those chromosomes represent individual free parameters. A population of chromosome values is randomly generated. Then, in successive generations, the genetic content of the chromosomes is altered through such analogs to biological evolution as breeding, crossover, and mutation. The overall process involves random changes superimposed upon an evolution to the fittest genetic structure, where fitness is defined as the global minimum in the error landscape. Thus, the method includes both a thorough random sampling of the feasible parameter space, helping ensure that a global rather than local optimum is achieved, and a driving force toward the optimum, achieving greater efficiency and better assurance that a precise error minimum has indeed been attained.

Markov Chain Monte Carlo methods similarly consist of what might be loosely viewed as a directed random search. Two relatively early and still-important implementations are the Metropolis algorithm and simulated annealing (SA) (e.g., Press et al., 1992). The Metropolis algorithm appears to have been first applied to watershed model calibration by Kuczera and Parent (1998), although the method dates to the 1950s, and simulated annealing had previously been applied to other optimization problems in environmental and geophysical model fitting. The overall idea is as follows. The feasible parameter space is randomly sampled and the objective function is evaluated; this may be viewed as the Monte Carlo component. The objective function value is then compared to that in the last iteration of the optimization routine; such first-order memory corresponds to the Markov chain component of the algorithm. If the prediction error is lower in the new iteration, the new parameter set is always retained. If the prediction error is instead higher in the current iteration, there is still a small but non-zero probability that the new parameter set is retained. That is, improvements are always accepted; degraded performance is occasionally accepted. Thus, as in GA, the probabilistic aspect of the method avoids trapping in local minima, but there is an overall impetus within the algorithm to descend the error gradient, affording greater efficiency than (for instance) a purely random search.

The shuffled complex evolution algorithm was developed by Duan et al. (1992). The method evidently consists, in essence, of a clever mélange of existing techniques which provides relatively fast convergence to a global optimum. Guidance on practical implementation of the technique was provided by Duan et al. (1993). Tolson and Shoemaker (2007) described SCE as "the dominant optimization algorithm in the watershed model automatic calibration literature over the past 10 years, given that more than 300 different publications reference the original set of SCE publications." This characterization seems accurate, and testifies to the status of SCE as the culmination of decades of research and practice in the global optimization of watershed models within a traditional (single-objective, single-solution) context.

These three global optimization techniques share many characteristics, but there are some distinctions. MCMC yields posterior probability distributions for the free model parameters (i.e., in a Bayesian framework, giving the distribution conditioned upon the available data through iterative modification of a generally simple prior distribution). These distributions serve as important measures of parameter identifiability, and also as the basis for estimating the component of forecast uncertainty due to parameter uncertainty. The basic versions of GAs and SCE do not inherently have this capability. However, useful ad hoc estimates might be obtained by recalibrating a model within a GA or SCE framework to several subsets of the data, or with different initial parameter values in the calibration, yielding a suite or distribution of parameter sets (D. Roche, Kerr Wood Leidal, pers. com., 2007; see also immediately below). Further, there now exist various different versions of these techniques more directly offering estimates of parameter uncertainty (see next subsection). There may be differences in efficiency as well. Although far more efficient than random searches, for instance, neither genetic algorithms nor Markov Chain Monte Carlo appear to be known for speedy convergence, whereas shuffled complex evolution was devised with efficiency in mind.

Notably, genetic algorithms and shuffled complex evolution have been applied to automated calibration of the UBC Watershed Model in British Columbia catchments (Lan, 2001; Roche, 2005). Both applications can be regarded as generally successful. Discussions with the authors of these studies indicated that neither provided quantitative information on parameter uncertainty that can be directly employed to generate forecast bounds, although in principle there may be ways around this limitation (see above). Interestingly, starting the SCE algorithm from different initial, random parameter values was consistently found to yield substantially different final calibrated values for certain UBC Watershed Model parameters (Roche, 2005). This behaviour is superficially similar to that encountered in a multi-start local-type search method (see previous subsection), but given the SCE algorithm's global search abilities, it appears to arise from a different source – specifically, the equifinality discussed in further detail in the next subsection. Behaviour was somewhat different between the two watersheds (Illecillewaet and Coquitlam) considered by Roche (2005), but in general UBC Watershed Model parameters associated with the vertical distribution of rainfall (e.g., E0LMID, P0GRADHI, P0GRADM) tended to show the greatest indeterminism, whereas final estimates for terrestrial hydrologic parameters, such as time constants for various reservoir/pathway systems, appeared relatively stable irrespective of the random parameter values used to initialize the SCE optimization.

4.1.4 *Modern directions*

At about the same time that the global optimization problem was more-or-less solved, it began to grow clear that this solution, although important, might not be quite as useful or complete as had been hoped. In response to a number of considerations and concerns, watershed model optimization research has taken a variety of directions since the early 1990s or so. Many of these issues and proposed solutions are closely inter-related.

The importance of generating estimates of parameter identifiability and the forecast error due to parameter uncertainty was discussed above. A number of earlier techniques incorporated such capabilities, but many did not. Considerable work has been devoted to furthering the development of these capabilities. Some well-known and well-regarded examples include the shuffled complex evolution Metropolis (SCEM) algorithm, which essentially combines the MCMC and SCE techniques (Vrugt et al., 2003), and the Monte Carlo toolbox and "dotty plot" techniques of Wagener et al. (2003) and Wagener and Kollat (2007).

A number of key insights were provided in the seminal work of Beven and Binley (1992), relating to the concept of equifinality and how to best deal with this challenge. This term denotes the fact that several different sets, or a continuous distribution of sets, of watershed model parameters can all yield prediction qualities which are, all things considered, essentially indistinguishable. That is, watershed model parameterizations are in general nonunique in terms of prediction error. Note that this does not necessarily require physical nonuniqueness, as two model parameterizations may yield identical objective function values but hydrographs which are substantially different. Nevertheless, physical nonuniqueness (identical hydrographs from different parameter sets for the same meteorological forcing and initial conditions) would also lead to equifinality. Equifinality may arise from a variety of sources, including similar/competing effects (i.e., trade-offs) between different model parameters, or lack of sensitivity of the objective function to a given parameter. In terms of optimization, the impact of equifinality is that the global minimum in the error landscape is not a well-defined and localized crater, but instead a low broad plain (or potentially, a field of separated craters of approximately equal depth). Any set of parameters which yields a prediction error lying upon that plain or within those craters is an acceptable model parameterization. Equifinality therefore amounts to parameter indeterminism, and the widespread occurrence of equifinality in watershed models in turn implies that the concept of a highly precise global optimum may not be hugely relevant in practice.

In response to these observations, Beven and Binley (1992) suggested a new paradigm for hydrologic model optimization, termed generalized likelihood uncertainty estimation (GLUE). Although computationally relatively simple, the method involved a fundamental change in perspective. An exact global optimum is not considered possible, or at least hydrologically meaningful, and no attempt is made to identify one per se. The technique employs a random search through the feasible parameter space. All parameter sets which yield model predictions deemed to be acceptable in terms of some error metric are denoted behavioural. By weighting each behavioural parameter set by a normalized variant of the corresponding error metric, a posterior distribution reflecting parameter uncertainty or indeterminism may be defined on the basis of the optimization. Subsequent work (Kuczera and Parent, 1998) has identified this technique as a form of importance sampling, and has criticized the identification of formal (e.g., 95%) confidence bounds on its basis unless an unreasonably large number of realizations is employed. Nevertheless, the method is widely credited with providing a direct and reasonable measure of the overall ranges in parameter values and their impact on hydrologic predictions, and is perhaps the only technique that directly addresses equifinality. In line with the overall concept of a distribution of acceptable parameters, the initial work of Beven and Binley (1992) did not emphasize the production of a single best hydrograph prediction. However, the resulting suite of parameter sets constitutes an ensemble, and an ensemble mean may be readily produced (normally, with members weighted by their individual performances), as summarized by Wagener et al. (2003). Similarly, the original work of Beven and Binley (1992) was primarily directed toward the parameterization of distributed hydrologic models, as equifinality issues tend to grow with the number of free parameters, of which there can be many in a distributed model. However, subsequent work has shown the concepts and techniques to be general in nature, and applicable to semi-distributed or lumped watershed models. Additionally, the approach has been applied to other environmental modeling contexts (e.g., Binley et al., 1997; Jia and Culver, 2008), and continues to be widely and profitably used in a variety of hydrometeorological problems (e.g., Unkrich et al., 2007). Within the technical literature, references to the GLUE method appear to run into the hundreds, likely exceeding even those to SCE. A recent review of the equifinality concept and GLUE is provided by Beven (2006).

An additional issue to which considerable attention has been devoted over the last several years is the choice of objective function. In the presence of an imperfect model and imperfect data, perfect hydrograph matches are not possible. A consequence of this fact is that a calibrated model will tend to best match the component of the hydrograph to which it is (in effect) fitted, and the component to which it is fitted is determined by the form of the objective function. For instance, minimization of root mean square error (RMSE) or maximization of Nash-Sutcliffe efficiency are two common objective functions. In practice, such objective functions tend to emphasize a successful match to peak flows. Whether this is desirable depends on the purpose to which the model is put. An alternative is minimization of error in the logarithms of the streamflow data, which tends to emphasize low flows at the expense of matching higher flows, and which may be useful for aquatic habitat or water resource studies (e.g., Hogue et al., 2003; Samaniego et al., 2007; Fleming, 2007). Noting that streamflow time series are second-order nonstationary, with variance that tends to increase with flow magnitude, the Heteroscedastic Maximum Likelihood Estimator (HMLE) was introduced as a new error metric (for reviews, see Gupta et al., 2003 and Wagener et al., 2004). In practice, using an objective function wherein HMLE is minimized leads to poorer prediction of high flows (Wagener et al., 2004), and was found to fail entirely for the Illecillewaet basin when used in conjunction with shuffled complex evolution and the UBC Watershed Model (Roche, 2005). Another possibility, not yet thoroughly explored in the rainfall-runoff modeling literature, is to minimize frequency-domain rather than time-domain error; work to date (e.g., Duffy and Al-Hassan, 1988; Fleming et al., 2002) suggests that this approach might yield better matches to hydrograph timing.

Multi-objective (MO) optimization, also known as multi-criteria optimization, is another fast-emerging field in hydrologic model calibration (see, for example, Hay and Umemoto (2006) and references therein). In effect, this method includes several different approaches; the common thread is that all involve simultaneous optimization of several objective functions. One technique focuses on using multiple model fit metrics, such as Nash-Sutcliffe efficiency of both streamflow and the logarithm of streamflow (e.g., Samaniego et al., 2007), intended to simultaneously capture prediction quality for both high and low flows. Another approach to achieving the same goal is to split the streamflow time series into different components (e.g., driven and non-driven components) in a data pre-processing step, and then perform a MO optimization in which RMSE in both is simultaneously minimized (e.g., Gupta et al., 2003b). Yet another implementation considers multiple target time series, such as both streamflow and groundwater level (e.g., Seibert, 2000; Beldring, 2002), or streamflows from several nested subcatchments (e.g., Seibert et al., 2000), and simultaneously minimizing prediction error in all of these. The overall idea in all cases is that a better-rounded model will result. In the particular case of multiple target time series, improvement in physical meaningfulness of the model may be highly significant, although this advantage comes at the expense of greater (and perhaps in many cases, unrealistic) data requirements. Another property of MO techniques in general is that the solution is phrased in terms of Pareto optimality, whereby an improvement in one objective function can only be attained by poorer performance in another. By the same token, the result for any given objective function within the MO framework will generally be inferior relative to that which would be obtained using a conventional single-objective optimization on that objective of interest alone. That is, MO optimization solutions are usually compromises, which (again) may or may not be desirable, depending on the use to which the model is put.

MO techniques are commonly developed from pre-existing methods. A relatively simple and broadly used approach is a composite objective function, whereby the objective function in a traditional single-objective optimization consists of some combination of error metrics. An example is the use of a combined metric capturing both correlation skill and volume error (Lindström, 1997). Masking of improvements in one objective component by decreased skill in another may be particularly severe in this approach and evidently may limit ability to find a genuine Pareto optimum, but it appears this may be overcome using compromise programming, which in practice evidently consists of a straightforward weighting scheme (e.g., Samaniego et al., 2007). Another approach is to combine various individual error metrics using fuzzy logic (Seibert, 1997). An advantage of composite objective functions is that they can be implemented within any optimization scheme; a disadvantage is that the effects of different components of the objective function might be challenging to disentangle. Another general approach is to modify existing techniques to explicitly perform MO optimization and clearly define a Pareto optimum. Methods developed in this way and applied to watershed model calibration include a multi-objective version of SCE, dubbed MOCOM-UA (for reviews, see Beldring, 2002; Gupta et al., 2003b), and a multi-objective version of SCEM, dubbed MOSCEM (Zhang et al., 2007).

A method closely related to, but conceptually distinct from, MO optimization is the staged or multi-step calibration (e.g., Hogue et al., 2003; Hay and Umemoto, 2006). The overall idea is to emulate, within an automated framework, the techniques commonly employed in manual hydrologic model calibration. Typically, recommended manual calibration procedures involve first estimating a certain subset of parameters to match some portion or aspect of the hydrograph, then another subset of parameters believed to be primary controls on some other portion or aspect of the hydrograph, and so forth (e.g., Quick, 1995; Smith et al., 2003). In the multi-step automated calibration, the same approach is taken, except that values of the parameters to be estimated at a given step are obtained using a formal optimization technique, such as SCE (Hay and Umemoto, 2006). This method is an intriguing hybrid between manual and automatic calibration approaches, but does not appear to have been extensively and broadly tested yet. One potential problem is that estimates of parameter identifiability and uncertainty may be less meaningful, as (at best) they would be obtained on a batch basis during each step. The results would solely represent the sensitivity of a particular parameter subset corresponding to a particular hydrograph component, and generally at a calibration stage when the parameter set as a whole is sub-optimal. By the same token, comprehensive quantitative assessments of the effects of parameter inter-relationships and parameter set uncertainty as a whole, as so clearly illuminated using a GLUE approach for example, would seem difficult or impossible to meaningfully extract in the multi-step framework.

Another direction in very recent research is the development of more efficient search strategies. The SCE approach was lauded as efficient in its day, and in most cases convergence will indeed be acceptably fast for any reasonable semi-distributed or lumped watershed model. However, interest in distributed models has grown significantly, and run times for such models can be long. Minimizing the number of runs required in the iterative optimization process thus becomes more important. A significant step forward is the Dynamically Dimensioned Search (DDS) algorithm of Tolson and Shoemaker (2007). The premise is to maximize the optimization performance obtained within a user-specified number of runs, by reducing the number of parameters inverted as the optimization progresses (i.e., to obtain the biggest bang for the buck in terms of CPU time). Results to date have been promising, and work continues on the DDS algorithm (Tolson et al., 2007). Similarly, more computationally efficient versions of SCEM have also been recently proposed (Vrugt et al., 2007).

Potentially relevant new optimization techniques continue to emerge, both within the hydrologic modeling literature and in other fields. Particular effort continues on finding more efficient, feasible, or physically reasonable ways to parameterize distributed hydrologic models and models in ungauged basins, yielding diverse methodological outcomes (a recent sample includes Shi et al., 2007; Chu et al., 2007; Pokhrel et al., 2007; Wagener et al., 2007; McMillan et al., 2007; Kling et al., 2007; Liu et al., 2007; Samaniego et al., 2007). Multi-model ensembles (e.g., Georgakakos et al., 2004) may subsume the parameter uncertainty present in individual ensemble members. Reverse-jump (transdimensional) Markov Chain Monte Carlo techniques, an MCMC approach in which the algorithm decides the dimensionality of the optimization problem, have been applied to models of groundwater flow and contaminant transport (Mendes and Draper, 2007). Multi-objective differential evolution techniques, consisting of a differential evolution method (Storn and Price, 1997) related to traditional genetic algorithms, but incorporating multiple objective functions, have been applied to chemical engineering problems for instance (Babu et al., 2005) and are being investigated for use in hydrometeorology (A. Cannon, Environment Canada, pers. com., 2007). Particle swarm optimization (Goswami and O'Connor, 2007; Gill et al., 2007) and delayed rejection adaptive Metropolis (Smith and Marshall, 2007) algorithms have also seen application to watershed model calibration. More broadly, fully interactive modelling software that may be run in forward or inverse modes with user-selectable free parameters in as few or as many stages of calibration the user wishes, loosely akin to the multi-step approach outlined above but far more flexible, has been commercially available and broadly applied in the field of geophysical modeling for some time (e.g., Fleming and Trehu, 1999), though the technique may share the multi-step method's limitations with respect to generating meaningful estimates of parameter uncertainty/identifiability and prediction error bounds. Further, the general problem of optimization under challenging conditions (including but not limited to those encountered in hydrology) remains an active field of research within the mathematics community.

4.2 *Optimization method choice/development*

As noted in Section 2, the second major step was to select and/or build an optimization and uncertainty estimation system specific to our needs.

4.2.1 *Selection criteria*

On the basis of the considerations and processes listed in Section 3.3 and in light of the issues and findings discussed in Section 4.1, the following nine criteria were identified for screening, selecting, and/or developing candidate optimization techniques and software:

- *Global*. The method should be robust to local minima in the error landscape.
- *Multiple/flexible optimization goals*. From the perspectives of both H&TS, and BC Hydro operations planning engineers responsible for various projects, it is important that the recalibrated model simultaneously satisfies multiple inflow forecasting goals. These include both daily forecast reliability (particularly for small, flashy, coastal basins with limited reservoir storage) and monthly and seasonal forecast volumes (particularly for larger basins in Columbia/Kootenay and Peace regions with large reservoir storage and major individual contributions to overall British Columbia power generation). Two model parameterizations, each obtained using a single corresponding objective within their respective independent optimizations, could be developed for each basin, and used separately for daily and seasonal forecasting. However, it is far preferable for a number of reasons – recalibration project efficiency, ease of incorporation of the final result into the existing RFS, and general conceptual elegance, for example – to develop a single model parameterization for each basin, capable of adequately performing multiple operational forecasting tasks. Note that for similar reasons, and the added complication of establishing initial conditions at seasonal model change-over, a set of seasonal calibrations for each basin (i.e., yielding different model parameter sets for different seasons) was ruled out. However, a capability to adjust the optimization goals somewhat from one basin to the next, reflecting local hydrometeorological characteristics and/or project-specific user priorities, is also desirable.
- *Direct and meaningful uncertainty estimation*. This selection/development criterion follows from the main goals of the recalibration exercise, as discussed in Section 3.3. The method should provide quantitative estimates of parameter identifiability, and ultimately hydrologic forecasts with error bounds reflecting the effects of parameter uncertainty, in a manner which explicitly addresses such issues as equifinality and parameter interdependence, preferably within a Bayesian framework. Additionally, hydrograph prediction uncertainty has to be captured specifically via an ensemble of individual hydrograph traces, rather than solely as a statistical summary metric like the standard error of prediction, because individual hydrograph forecast traces are required as input by some software applications employed by forecast product users.
- *Physical reasonableness of estimated parameters*. Methods should be in place for ensuring that the hydrological model parameters estimated using the optimization algorithm, and the corresponding hydrographs and hydrograph components (e.g., relative contributions of rainfall, snowmelt, glacial melt, and groundwater seepage) are physically plausible and defensible.
- *Utilitarian but user-friendly*. Large investments in developing a finely honed, commercial-grade optimization software package are not appropriate or feasible at this time. However, given that several personnel will likely be using the optimization code, a reasonable level of user-friendliness, including a sensible user interface for input, optimization control, and output visualization and interpretation, is desirable.
- *Transparency and accessibility*. Although not a strict requirement, it would be preferable if the optimization code employed is open to scrutiny by BC Hydro, and written in a manner that facilitates understanding, and revision if desired.
- *Amenable to coding in Visual Basic for Applications (VBA)*. This requirement is a product of other criteria or constraints, including: available IT support; facilitation of the transparency and accessibility described above; realization of project efficiencies through the use of some existing H&TS code resources; and technical details of the development of computationally efficient interfaces between various software components. The VBA code must integrate user control, optimization, UBC Watershed Model parameter (.WAT) file creation, calls to the UBC Watershed Model (in this application, an external executable file, compiled previously from Fortran) for each function evaluation, ASCII conversion and importation of binary-file model results, and visualization and summary capabilities, within a single application.
- *Capitalize on existing UBC Watershed Model & BC Hydro knowledge, capabilities, and results*. The current (manual) calibration of the UBC Watershed Model used in the RFS is the result of considerable effort by experienced and knowledgeable professionals. Further, practical time and resource constraints must also be borne in mind. Thus, a complete, ground-zero model re-development did not seem prudent or appropriate at this time, and much of the structure of the current UBC Watershed Model parameter files is to be maintained. As just one example, the number and placement of elevation bands for each basin in the current calibration is the result of basin-specific hypsometric analyses and extensive experience with the UBC Watershed Model; modification of the number of elevation bands currently employed is reported to be unlikely to result in significant hydrograph prediction or model run-time gains; and number and placement of bands would be challenging to incorporate as free parameters into an automated calibration procedure. Consequently, it was decided that the number and placement of elevation bands would not be altered from the current calibration, unless (for a given basin) there would be a clear

operational forecasting benefit to be gained by doing so. Similar considerations apply to several other questions of model parameter structure. Rather, the emphasis in the optimization was to refine a number of free parameters clearly amenable to such techniques, conditional upon updated hydrometeorological and land surface (e.g., glacial and forest cover) data, and to obtain corresponding parameter and prediction uncertainty estimates.

- *Proven track record*. Although some innovation and development could be appropriate, and indeed ultimately proved necessary, an optimization and parameter/prediction uncertainty estimation technique with a well-demonstrated prior history of successful and wide application to watershed model calibration was strongly preferable in order to minimize project and timeline risks.

It is expected that similar or analogous criteria would be imposed by many operational hydrologic forecasting groups.

4.2.2 *Selected approach: modified GLUE*

The method chosen (dubbed Automated Behavioural SYNThetic Hydrograph Estimation, or Abysnthe) consists of generalized likelihood uncertainty estimation, with a number of modest but important modifications introduced here to better satisfy our requirements. Some aspects of the method bear some similarities to the signature index approach proposed by Yilmaz et al. (2007), though the indices or criteria considered here are defined more on the basis of operational forecasting requirements than the underlying physics of the model per se. Overall logic of the Absynthe technique is illustrated schematically in Figure 2.

Equifinality and the GLUE method were described at a conceptual level in Section 4.1.4 above. In practise, the standard GLUE method – which forms the platform upon which Absynthe is built – proceeds as follows.

1. *Define prior distributions.* Define which parameters, for a given basin, are to be left free in the optimization, and choose physically reasonable ranges for each. Selection of the feasible space for each free parameter will be made using, for guidance, both general hydrologic knowledge as well as catchment-specific modelling experience.
2. *Input distribution sampling.* Perform random sampling of the feasible parameter space for each free parameter in the optimization. This Monte Carlo sampling is completed using a uniform distribution as the Bayesian prior for each parameter. The uniform distribution is bounded by the parameter's feasible space. Among other things, this framework ensures that each parameter can take on only values which are deemed to be physically reasonable.
3. *Model run.* Run the model using the randomly sampled parameter set. For the specific case of the UBC Watershed Model, this involved rewriting the .WAT file using the selected parameter set, running the UBC Watershed Model, convert the resulting output binary file to ASCII, and returning the simulated hydrograph, using (in part) VBA script cannibalized from an existing H&TS application.
4. *Behavioural test.* Deem the parameter set behavioural in the usual GLUE sense if the quality of the hydrograph is sufficient. A common approach is to specify a minimum acceptable Nash-Sutcliffe efficiency of daily flow predictions.
5. *Loop.* Repeat steps 2 through 4. The condition for convergence is normally the accumulation of some specified number of behavioural parameter sets, often numbered in the thousands. The resulting parameter sets may then be used to (i) define the posterior distributions of the individual parameters in a (loosely) Bayesian framework, providing estimates of parameter identifiability, and (ii) generate ensemble flow predictions on the basis of the ensemble of parameter sets, yielding parameter uncertainty-based confidence bounds on the forecasts.

A working prototype Absynthe system (Fleming and Weber, 2008) incorporated the following changes and additions to the standard GLUE procedure:

- *Multiobjective test for behavioural sets.* Based on collective operational forecasting experience with the UBC Watershed Model, interviews with forecast product users, general physical hydrologic considerations, and the practical requirement for a single model for both daily and seasonal forecasting, a multi-criteria framework was adopted. The three criteria were matches to daily, monthly, and annual flows. The daily objective entailed a minimum acceptable RMSE improvement over the persistence forecast, and is particularly important for mountainous, temperate maritime watersheds on the southwest BC coast, as these (in general) have flashy hydrologic responses and small reservoirs, and can be subjected to large rainfall and rain-on-snow events. The monthly objective in fact included 12 sub-objectives, specifically, reduction of monthly volume biases for each month. Imposing a maximum acceptable % bias for all months is particularly operationally important for interior BC projects with large reservoirs and significant total system planning impacts, and is also important in terms of technical forecasting issues insofar as it avoids the necessity for bias correction. The annual objective involved a minimum acceptable improvement in yearly total volume RMSE over the mean, and is important for seasonal inflow volume forecasting. Unlike a conventional MO optimization, however, a Pareto compromise is not an explicit goal in the MO GLUE procedure, and the foregoing were instead hard targets for parameter set acceptance or rejection. A generalized likelihood function was generated as the mean of the likelihood function functions for the three error components. These in turn were simple mathematical expressions to capture the foregoing error components in such a manner as to have increasing values as the fit quality increased. Note that a likelihood function was devised for the volume bias for each month individually, for a total of 12 likelihood functions, and then these were averaged to obtain a single likelihood function for the month-scale optimization target. Thus, daily-, monthly-, and annual-scale calibration performances were each given equal weight in the calculation of the global likelihood function. However, the importance of each calibration target can be controlled by the user through suitable adjustment of the behavioural criteria. For instance, the stronger importance of daily flow forecast quality for small, flashy coastal basins is such that the bar may be set high for the daily performance, but lower for monthly or annual performances. There is some precedent for the incorporation of multiple calibration goals into the GLUE framework (e.g., Jia and Culver, 2008).

- *Screening for pathological sets*. In most or all Monte Carlo-driven methods, a feasible parameter space is predefined, in part or in full on the basis of physical reasonableness; final estimates are guaranteed to fall within that range. Even in other methods (e.g., gradient descent), a penalty can assigned to force the cost function toward a physically reasonable parameter value. But in many cases, and in particular for medium-to-high dimensional inverse problems, that is not sufficient to ensure physical plausibility of the estimated parameter set, or therefore acceptability of the final model. In mathematics and physics, a pathological value or function is loosely defined to be one which is strictly mathematically correct but is in some way profoundly atypical, irrelevant, misbehaved, or senseless. We adapt this concept to watershed model calibration. In our context, we define a pathological parameter set as one which (i) satisfies, or may satisfy, the mathematical optimization goal (i.e., pass behavioural tests or minimize an objective function), (ii) satisfies the mathematical feasible range constraints set out in the specification of the Bayesian prior distributions for each parameter, and (iii) makes no physical hydrologic sense when considered as a whole. That is, we define a parameter set as pathological if the values satisfy the optimization problem and are physically realistic individually but not collectively. For example, say our watershed model has time constants for an upper groundwater system and a lower groundwater system. In the UBC Watershed Model, these parameters are specifically denoted P0UGTK and P0DZTK with units of days, but similar ideas apply to other models. One might reasonably set the prior distributions to P0UGTK ~ *U*[5 35] and P0DZTK ~ U[25 85], and obtain (through some calibration method) the parameter set (P0UGTK = 33.4, P0DZTK = 29.7). Such a set would meet conditions (i), (ii), and (iii) outlined above and thus be identified as pathological, as it is a legitimate answer to the mathematical optimization problem but implies shallow groundwater responding slower than deep groundwater, which for most watersheds (karst environments being a possible exception) is physically unreasonable. The advantages of process-oriented models are their abilities to explain hydrologic behaviour in terms of underlying physical processes, and predict hydrologic behaviour under circumstances not sampled by the hydrometeorological record available for the study basin. Hence, both the utility and credibility of a process-oriented model require that the parameter set be fully physically reasonable; otherwise, one might as well use a statistical or soft-computing model, which sacrifices process description for easier implementation and (often) superior prediction performance. One simple solution to the pathologic-set problem is to specify feasible ranges such that they do not overlap, but this limits the parameter space that can be explored and is useful only if the true value for each parameter can be reasonably well-estimated a priori. This is often not the case. Here, we implement a more general solution: test each randomly sampled parameter set for pathological relationships, and only perform modelling for those that pass the test. The emphasis in our work is on correct ordering of time constants for various terrestrial hydrologic pathways (in the specific context of the UBC Watershed Model, the constraint is P0FRTK < P0FSTK < P0GLTK < (P0IRTK and P0ISTK) < P0UGTK < P0DZTK). Due to the emphasis on correct ordering of values across parameters rather than the specific value taken on by a particular parameter, the pathology test is a rank-based or sliding constraint on the optimization problem. The importance of such correct ordering is rarely if ever discussed in the watershed model optimization literature, but seems key to establishing the hydrologic credibility of the optimization process and its results. To our knowledge, the explicit definition, identification, and screening of pathological parameter sets is a novel feature of Absynthe.
- *Set selection and ensemble size*. The aim in Absynthe, as in any GLUE-based technique, is to generate an ensemble of parameter sets. However, we make two departures from standard GLUE procedures in this respect. Both are motivated by BC Hydro-specific operational considerations (see also Section 3.1), though these considerations might apply elsewhere as well. (i) We generate only 100 behavioural (and non-pathological; see above) parameter sets for a given basin. This figure appears considerably smaller than that typically used in GLUE applications, but operationally, it would be difficult to perform (say) 5000 simulations for each of 21

basins every morning by noon, particularly if modeller adjustments and re-runs are required. This volume problem is compounded by the fact that for each GLUE ensemble member, an ensemble of weather realizations (historical climate traces in the existing ESP seasonal forecasting framework, or NWP ensembles in a forthcoming planned probabilistic daily system) would be run. Even for the short individual run times of the UBC Watershed Model, the total CPU time would be prohibitive. Additionally, the software applications currently employed by some in-house forecast product users for reservoir operations or planning purposes cannot accommodate large ensembles. The implication of the smaller Absynthe ensembles is that they do not provide statistically rigorous estimates of flow distributions. For current operational purposes, and in the context of other practical forecast issues like NWP uncertainty or the relatively short duration of available historical climate records available for BC Hydro ESP runs, this problem is likely of secondary importance. (ii) From among the Absynthe ensemble we choose one set to serve as the official calibration for the basin, with the ensemble serving only for uncertainty estimation purposes. This approach appears somewhat contrary to the spirit of equifinality and the GLUE approach, but is unavoidable for current operational purposes in H&TS. Viewing each parameter set as constituting a distinct model, and drawing on work in meteorology, climatology, and hydrology (e.g., Christensen et al., 2007; Block et al., 2007; Georgakakos et al., 2004; Block et al., 2007; Goddard et al., 2007), such a forecast could be generated as one of several types of ensemble mean: for example, an arithmetic average of each of the ensemble traces, or a weighted mean of those traces with the weights drawn from the Bayesian global likelihood function value associated with each parameter set. However, using the single, nominal “best” parameter set out of the ensemble for deterministically generating a single trace (with ensembles run separately if desired) was preferable for our purposes, in part due to run-time concerns. Looking forward, however, parallelization of computing resources and increasing user comfort levels with probabilistic modelling methods may ultimately lead to more fully ensemble-based implementations in H&TS operational practise.

This initial Absynthe implementation proved successful, and its performance is described in Section 5.1 below. However, during the course of application to a range of basins with different characteristics and lying in different hydroclimatic regions across the BC Hydro system, it became clear that some additional modifications would prove useful. These further changes to the GLUE procedure and/or Absynthe prototype were as follows.

- *Secondary data targets*. The three calibration targets listed in the multi-criteria discussion above are all based on the streamflow hydrograph. In the second major version of Absynthe, the potential calibration targets were extended to include two other data sources. The first is April 1 snow water equivalent (SWE). April 1 snow measurements approximately capture the total snowpack available for water production during the spring-summer freshet. Additionally matching to SWE data can therefore help guide the parameter sets toward better solutions, and in particular, may in principle help mitigate equifinality issues insofar as a correct partitioning of general runoff generation mechanisms is better supported. The second is total annual precipitation. The UBC Watershed Model, like most or perhaps all process-oriented watershed models, contains a de facto local-scale meteorological model which, among other things, uses various scaling factors to help adjust point meteorological measurements from surface weather stations to better represent catchment-wide meteorological conditions, including total basin precipitation. Thus, ClimateBC-estimated total annual precipitation values integrated over the watershed were obtained in a GIS environment and also used as a secondary data target. ClimateBC (Wang et al., 2006) is a higher-resolution, British Columbia-specific data product developed at the University of British Columbia, based on the well-known PRISM dataset (Daly et al., 2002) from Oregon State University. The global likelihood functions are adjusted to additionally incorporate the adequacy of SWE or total basin precipitation fits. Also note that while in theory the inclusion of additional, non-hydrograph calibration targets like April 1 SWE or total annual basin precipitation should dramatically improve the physical plausibility of the calibration, in practise results were found to be somewhat mixed, reflecting to a large degree the local accuracy, representativeness, and completeness of the available SWE or ClimateBC data.
- *Increased user control*. Two general sets of changes were made to give Absynthe users fuller control over the optimization and uncertainty estimation process. The first was flexible calibration metrics. The overall hydrograph calibration goals remain phrased in terms of daily, monthly, and annual hydrograph matches, but it was found that flexibility in the metrics employed for each was useful. Thus, a user-selectable menu of different fit measures was incorporated into the second-generation Absynthe software. For instance, several measures of daily hydrograph fits are now available; Nash-Sutcliffe efficiency is the most commonly used of them. The second was increased flexibility in free parameter selection. In the prototype, a substantial but finite number of UBC Watershed Model parameters were deemed potentially free. In the second-generation version of Absynthe, virtually any parameter in the model can be

opened up for estimation and uncertainty assessment. This capability has to be used with discretion, of course, as issues with both parameter identifiability and indeterminism, and the CPU time required to adequately sample the feasible parameter space, grow quickly with increasing numbers of free parameters.

- *Hard constraint on snow-free bands*. The highest elevation band in the UBC Watershed Model that must be snow-free at least one day per year can be specified. Any simulation which fails to meet that constraint is rejected. It was found that this constraint could be useful in some cases for correctly partitioning summer freshet flows between snow and glacier melt, and to avoid progressive build-up of permanent snow over successive model seasons at elevations where it is known not to occur. In the context of glacier-fed catchments, this elevation band corresponds to the equilibrium line altitude (ELA). Whereas screening for pathological parameter sets (see above) is a flexible or moving constraint, the highest snow-free band stipulation is a fixed constraint.

The Absynthe technique as described above satisfies all nine selection criteria (Section 4.2.1). It is perhaps particularly noteworthy that using GLUE as a starting platform afforded great flexibility; for instance, both the sliding (pathology) and fixed (snow-free band) constraints, which are key to ensuring that the optimization method returns parameter sets that represent a physically plausible hydrologic reality, are very simple to implement in the relatively straightforward and robust GLUE framework. These benefits are believed to more than compensate for the disadvantages, which include substantial computational inefficiency, no guarantee of locating the precise minimum in the error surface, and the perhaps non-rigorous nature of the quasi-Bayesian likelihood functions (but see Beven, 2006). The practical question of how to incorporate the results into fast and robust generation of hydrograph ensembles in an operational forecast setting remains only partially resolved, but this important challenge is in no way unique to the method selected.

4.3 *Other methodological considerations*

Although the primary emphasis of this paper is on model parameter optimization and uncertainty estimation, a number of other key technical issues had to additionally be considered in the model recalibration process. These are briefly summarized below.

4.3.1 *Hydroclimatic data quality control*

Unregulated, local reservoir inflows are not directly measured by BC Hydro, but are instead calculated from measured changes in reservoir storage and discharge using the principle of conservation of mass.

Due to the nature of reservoir inflow calculations, data with a high temporal resolution, such as daily data, are noisy (i.e., contain random errors). Noisy inflows can arise due to wind set-up on the reservoir (i.e., seiches) and the coarse resolution of elevation measurements. Noise becomes more obvious at times of low inflows, such as the winter low flow period, when the signal-to-noise ratio is low. Other error sources are measurement errors in reservoir elevations, turbine flow, spillway or valve readings, errors in stage-storage curves and errors in rating curves for various outlet facilities.

BC Hydro performs quality control on raw daily inflows using custom-built VBA Excel tools. The process involves comparison of inflow data with data from nearby unregulated Water Survey of Canada streamflow gauges. The objective is to remove noise while conserving overall inflow volume and correcting for obvious errors. The process is semi-automated and divided into two major steps. In the first step, a 5-day moving window is run over the data and linear regression relationships with reference streamflow data determined. The algorithm logic is such that raw inflow data are deemed to be good and left unchanged, if the best coefficient of determination for inflows and reference streamflows is greater than or equal to 0.5. The reasoning is that the reservoir inflow estimates are good if they respond in a similar manner as observed streamflow data from comparable reference stream gauges in the region. If the coefficient of determination is poor, i.e., between 0 and 0.5, then inflow data are estimated based on the best-correlating streamflow data. If the correlation coefficient is below 0, i.e., when the relationship between raw inflows and streamflows is extremely poor, then an estimate is calculated from the 5-day moving average of raw inflows. In a second step, long periods of poor data can be estimated using an additive relationship with reference streamflow data based on unit yield. With user intervention being possible at every step and additional data, such as temperatures, precipitation, reservoir elevations and outflows being available, the final quality of the data depends to a large degree on the hydrologic experience of the operator (Weber 2001).

Water Survey of Canada's streamflow data have undergone quality control by the agency itself and are assumed to be of good quality.

Raw climate data, specifically daily precipitation and daily minimum and maximum temperatures from BC Hydro-operated automated weather stations and Environment Canada's station network, can experience problems due to a variety of reasons. These include snow-capping, freezing, full precipitation standpipes, radiation effects on temperature sensors, or equipment malfunctions.

Quality control of raw data, as well as record extension, was done using a custom-built VBA Excel tool. First, precipitation and temperature data are screened for gaps, absolute outliers and regional outliers. Then, missing or poor data are estimated based on simple monthly (for temperature) and annual (for precipitation) linear relationships with predetermined reference stations. Additional features to estimate long-term gaps in precipitation records include the matching of a temporal pattern of precipitation at a selected reference station. Temperature data can also be estimated using the temperature lapse rates for the day in question as calculated from regional stations, or by adding the temperature change recorded at a reference station since the previous day to the previous day's temperature at the station estimated.

4.3.2 *Land surface data*

Many of the land surface data used in the previous set of calibrations of the UBC Watershed Model at BC Hydro were derived from paper maps, particularly for the larger watersheds in the province. For this calibration, all data were derived from digital GIS sources, using better and more up-to-date sources than were previously employed. The number and distribution of elevation bands was maintained from the previous calibration whenever possible. Pre-existing digital watershed boundaries were also used whenever available, but had to be delineated using digital elevation models (DEMs) for most of the larger watersheds. Band slope, orientation, mid-elevation and area were also derived from digital elevation models. 2005 glacier data for British Columbia were provided by the Western Canadian Cryospheric Network (B Menounos, University of Northern British Columbia, pers. comm., 2005). Forest and crown cover data were derived from the ca. 2006 Vegetation Resource Inventory and supplemented by ca. 1992 Baseline Thematic Mapping (BTM) land use data, both available from the Government of British Columbia's Land and Resource Data Warehouse (www.lrdw.ca).

4.3.2 *Meteorological station selection*

The UBC Watershed Model as implemented in the RFS requires time series of daily total precipitation, daily maximum temperature, and daily minimum temperature. Meteorological station data were incorporated into the model through the use of composite stations, which consist of a combination of data from various real weather stations within and/or near the watershed into a single "index" or "virtual" time series and location.

The choice to solely use composite stations (one per basin, and generated using a standard procedure, outlined below in this section) was motivated by both calibration and operations considerations. The following advantages were thought to outweigh the drawbacks, which are an inability to model inversions, and no guarantee of optimal meteorological data selection and representation:

- Relatively good objectivity, consistency, and standardization across multiple Absynthe users, in comparison to the trial-and-error approach;
- Efficiency, again relative to a trial-and-error approach to station selection;
- Defensibility, insofar as a reasonable and standardized approach was taken;
- Reduced optimization dimensionality, and thus reduced calibration effort/run time and mitigated equifinality, because additional parameters requiring calibration arise with additional meteorological stations;
- Operational robustness, in the sense that when using several station inputs but integrating them into a single composite value, the day-to-day practical consequences of data collection or transmission failures, or poor data quality, for a given individual station may be somewhat mitigated insofar as the effects of erratic data can be partially reduced by lumping;
- Operational simplicity, because by having only a single meteorological station (i.e., a single time series for each meteorological data type), the clarity and speed with which changes to the historical record (a common requirement in operational forecasting, generally arising from monitoring network incompleteness) can be made and tracked should be improved;
- Conceptual consistency with existing approaches, such as the index stations used in the existing RFS calibration for some basins, and methods used by the US National Weather Service models;
- Mitigated potential for physically unrealistic elevation-SWE distributions that are known to occur in the existing calibration when the horizontal heterogeneity in spatial meteorological fields is transposed into spurious vertical gradients in an elevation band-based semi-distributed watershed model;
- Relative amenability to stochastic weather generation, insofar as a multi-site weather generator would not be required, at least within a given watershed.

Selection of stations for inclusion in the composite proceeded on the basis of the following criteria. (i) Pre-screening with data QC analysts for station reliability and accuracy, as different sites and equipment have different performances; (ii) record length; (iii) proximity to the basin; (iv) a significant elevation range span across the suite of selected stations; and (v) linear cross-correlations between total annual flow volume and total annual rainfall volume. In general, no more than two or three stations were selected for inclusion in the composite, due to both meteorological network limitations, and the observation that including a larger number of stations into the linear combination employed to generate a synthetic record (see immediately below) tended to unrealistically damp out day-to-day variations. A composite precipitation daily precipitation record was then generated as:

$$P_{composite}^{daily}(t) = \frac{a_1 P_1^{daily}(t) + a_2 P_2^{daily}(t) + \ldots + a_N P_N^{daily}(t)}{\sum_{n=1}^{N} a_n}$$

where the coefficients, $a_{1,N}$, were obtained from a multiple linear regression of total annual flow volume upon the total annual precipitation recorded at the candidate stations:

$$Q^{annual}(t) = a_1 P_1^{annual}(t) + a_2 P_2^{annual}(t) + \ldots + a_N P_N^{annual}(t) + b$$

A daily maximum temperature time series was generated for the composite station by averaging the daily maximum temperature records from the constituent stations. A similar procedure was employed for daily minimum temperature. Because temperature variations are in general far more spatially coherent than those in precipitation, it is believed this simpler composite generation technique may reasonably be used for temperature. The station was nominally accorded an elevation equal to the mean of the elevations of the constituent stations, but in some cases this elevation was treated as a free parameter in the optimization, as doing so may allow the calibration routine to adjust for temperature biases in the data resulting from incomplete coverage of the surface meteorological observation network.

5. RESULTS

The final project component, as outlined in Section 2, was to implement the Absynthe technique for BC Hydro forecasting basins. This task was performed in two phases. The first was an initial benchtesting phase, described below in Section 5.1. This was following by some code modification to develop the second-generation Absynthe application (as described above in Section 4.2.2), and then the second phase of implementation, described in Section 5.2 below, consisting of production-line, team-based application of Absynthe across 21 basins. This second phase is not yet finished, but the number of basins completed to date is sufficient to provide a good feeling for the use and effectiveness of the method.

5.1 *Preliminary stand-alone Absynthe results*

The Alouette catchment was used as an initial proving ground for Absynthe. Alouette is a small coastal watershed, dominated by steep terrain and temperate rain forest, driven mainly by rainfall with moderate snowmelt and minor glacial melt contributions. To assess specifically the performance of the Absynthe algorithm, without the confounding influences of other recalibration project steps as outlined in Section 4.3, the initial implementation used the same land cover, meteorological, and streamflow data as did the existing RFS calibration. Instead, we simply allowed 21 parameters in the existing operational Alouette model to be free in an Absynthe parameter estimation. The first-generation version of the Absynthe software was employed, without calibration to snowpack or a hard constraint on the lowest snow-free elevation band; this is not unreasonable for Alouette.

An Absynthe ensemble for a typical year is illustrated in Figure 3. About 72% of the observations lie within the envelope of Absynthe hydrograph traces, which appears to be a middle-of-the-road value for GLUE-based methods (see, for example, the "containing ratio" values reported by Xiong et al., 2009). About 17% of the observations lie above the upper bound and 11% below the lower bound, which is reasonably symmetrical.

Performance of the nominal best parameter set is summarized in Table 1. Two "best" sets are illustrated: the individual parameter set having the highest value of the global likelihood function, and the individual parameter set with the best daily Nash-Sutcliffe efficiency, which is particularly important operationally for small, flashy coastal projects like Alouette. The Absynthe recalibration set yields daily and annual performance statistically similar to that of the existing calibration, while dramatically reducing the maximum monthly volume bias (i.e., of the 12 months of the year, the worst found in the calibration; for Alouette, this occurred during summer baseflows, when a modest absolute error can yield a substantial % error).

The foregoing results suggest that Absynthe produces results which are useable for deterministic and probabilistic forecasting operations. However, an additional key consideration in applying this method was to obtain a better understanding of the properties of the UBC Watershed Model. Some useful insights are provided by diagnostic features which were incorporated into Absynthe.

Some examples of "dotty plots" (projections of points on a likelihood surface onto a single parameter axis; Wagener et al., 2003; Beven, 2006; Wagener and Kollat, 2007) for the preliminary Alouette optimization are provided in Figure 4. Caution must be exercised when interpreting these figures, as each point on a given plot corresponds to an entire parameter set, not just the parameter referred to on that plot. Nevertheless, these plots are a highly useful diagnostic feature. A shotgun pattern indicates poor identifiability – a wide range of values could work for that parameter. Conversely, the appearance of a structured and peaked distribution on the dotty plot for a given parameter suggests good identifiability – irrespective of the values of the other parameters, high global likelihood values (i.e., good constrained multiobjective fits) arise for particular values of that parameter. In this case, UBC Watershed Model parameter P0RREP (a rainfall representation factor, which linearly adjusts rainfall amounts observed at a given meteorological station) shows fairly good identifiability. Parameter P0ISTK (interflow time constant for snowmelt, in days) shows murky structure: there is no clearly correct value, and equifinality is therefore in play, but some parts of the feasible parameter space can indeed be ruled out. The dotty plot for parameter P0GRADL (precipitation gradient factor, in % per 100 m elevation increase, for elevations below a threshold value E0LMID) shows no structure in this example: equifinality is in this case in full swing. Overall, of the

21 free parameters in this calibration, only 3 showed clear structure; 4 showed murky structure; and 14 showed no apparent structure. The results hence suggest widespread parameter indeterminism and therefore parameter uncertainty, duly transmitted into the prediction uncertainty apparent in Figure 3.

It is perhaps worth pointing out that if the minimum and maximum limits to the feasible parameter space are pre-defined to lie very close to each other, i.e., the specified feasible range for a given parameter is very narrow, an appearance of lack of structure is unavoidable but may not speak to gross equifinality. This is not an issue with the results discussed above, however, as the limits of the uniform prior distributions were set far apart for each parameter in order to fully explore the parameter space.

The source of this parameter indeterminism was tentatively explored by calculating a linear cross-correlation matrix between the values of the free parameters across behavioural sets. Results are illustrated in Figure 5. There are some clear trade-offs between certain parameters. One example is the strong negative correlation between P0RREP (the rainfall adjustment discussed above) and P0GRADM (like P0GRADL discussed above, but for a higher elevation range): as one increases, the other decreases. On the other hand, some other parameters which exhibited poor identifiability in the first-generation Absynthe application to Alouette do not show correlations with other parameters. An example is P0PERC (groundwater percolation in mm/day). In such a case, parameter indeterminism presumably arises instead from a straightforward insensitivity of the model performance to the value taken on by that parameter. Bear in mind, of course, that specific findings will vary from basin to basin, and also that only linear associations are considered in this matrix.

Some parameter trade-offs may also be imposed or catalyzed by the imposition of the non-pathology constraint. Recall that this constraint mainly relates to the physically plausible ordering of values across parameters. Thus, if one of the parameters upon which a constraint is placed takes on a certain value, then that must limit the range available to another parameter related by the same constraint (within a given parameter set). Some potential examples are illustrated in Figure 6. These "bubble plots" provide an alternative to conventional dotty plots, permitting illustration of (and thus facilitating comparison between) multiple parameters on a single graph. The parameters included on Figure 6 are all time constants for various terrestrial hydrologic pathways, with units of days, and are all related to each other explicitly through the imposed non-pathology constraint (see Section 4.2.2). As one proceeds from left to right across this figure, the parameters correspond to physical processes that should normally be progressively slower (the details need not be delved into here), and this is reflected in the pathology screening. We can see from the figure that this leads (as expected) to a grouping of values: the identified time constants for P0FRTK, for example, are generally lower than those for P0GLTK, though having identical feasible ranges. This is accompanied by a large positive correlation coefficient (Figure 5): within the ensemble, as one goes up, the other will often also go up, because a parameter set containing P0FRTK > P0GLTK is deemed pathological and is rejected for inclusion in the ensemble.

Some other potential impacts of these moving constraints were explored by re-running the Absynthe optimization for Alouette exactly as before, but omitting screening for pathological parameter sets. In general, one might conjecture a continuum of potential effects, having the following two end members. (i) At one extreme, if physically pathological parameter sets are inherently incapable of yielding behavioural hydrographs, turning off the pathology screening should slow down the Abysnthe run time, because more CPU time would be wasted on exploring barren parts of the parameter space. (ii) At the other extreme, if pathological sets are equally capable of yielding behavioural hydrographs, then many of the parameter sets identified in a conventional, non-screened run would show physically incorrect mutual relationships. Results are given in Table 2. In this particular case, the findings lean strongly toward scenario (ii). Indeed, if these constraints are not explicitly imposed as part of the optimization procedure, the vast majority (over 90%) of the parameter sets that would be accepted as behavioural in a conventional GLUE sense are physically unrealistic.

5.2 *Final results*

Calibrations have also been completed on seven more basins and sub-basins with at least five others in various stages of completion. Unlike the benchtesting Alouette results discussed above, these were in general performed using a match to Apr 1 SWE as part of the behavioural criteria and lowest permissible snow-free band as a hard constraint, in addition to updated meteorological, hydrometric, and land surface data, as well as composite stations, and are intended for operational forecasting use. Further, Alouette has been finalized using this second-generation version of Absynthe, the updated data sources, and the composite stations. Aside from Alouette, the completed and soon-to-be completed basins mainly include the larger, snow and glacial melt dominated watersheds on the Columbia, Kootenay and Peace Rivers of the BC Interior. The remaining watersheds to be completed in the future will mainly include the smaller rainfall-dominated basins of the Lower Mainland and Vancouver Island, as well as the Bridge River system.

Of the watersheds completed thus far, Alouette is a pluvial, rainfall-dominated basin with a relatively small snowmelt contribution. The Columbia River upstream of Mica Dam is split into two sub-basins: Donald upstream of the Columbia River at Donald Water Survey of Canada gauge and Mica Local, downstream of the gauge. These two sub-basins, along with Duncan and Revelstoke are glacionival basins with a significant amount of glacial contribution to flow in the late summer months. Special attention was given to these watersheds to ensure that the glacial contribution was reasonably simulated on an annual basis. The other three completed watersheds are largely nival (snowmelt-dominated), with very little glacial contribution.

Overall, the new calibrations have performed substantially better than the previous calibration, completed in 2001. Table 3 provides a statistical comparison of the results of both calibrations over their total periods (including both calibration and verification periods). Direct comparison between the two calibrations is challenging mainly due to non-overlapping time periods for the two calibrations, the varying length of the calibration period for the 2001 calibrations and the absence of verification runs for many of the 2001 calibrations. Those 2001 calibrations that did not include a verification period, and were calibrated over the entire record length evidently available at that time, are differentiated with an asterisk in Table 3. Although not presented here, the results are much better for the new calibration periods and even further exceed those from the old calibration.

In most cases the daily Nash-Sutcliffe efficiency and yearly percent volume error are comparable between the two sets of calibrations. Where the new set of calibrations has in particular provided significant improvement is in terms of the average monthly biases, and for some basins also in terms of annual volume R^2. The new calibrations also appear to provide a more reasonable amount of annual glacial contribution when compared to approximate calculations based on very limited glacier mass balance data.

Table 4 shows the percentage of observations that fall within the Absynthe ensemble inclusion envelope. For all calibrations, that value is about 70%, fairly typical for applications of GLUE-based optimization and uncertainty estimation technologies (see Section 5.1 above). Though variable from basin to basin, the distributions of confidence bounds around the correct value appear at least roughly symmetrical.

6. CONCLUSIONS

A watershed model parameter optimization and uncertainty estimation technique was developed and applied to several river basins in British Columbia, Canada, which are variously rainfall-, snowmelt-, and/or glaciermelt-influenced. The Absynthe technique is based on the proven generalized likelihood uncertainty estimation (GLUE) approach, but includes a number of important modifications: (i) multiple objectives; (ii) moving constraints to eliminate the possibility of “pathological” parameter sets, a concept introduced in our work to denote sets of parameter values which are physically reasonable individually but not collectively; and (iii) a hard constraint on the lowest elevation band that must be snow-free at least one day per year, which helps guide the de facto internal meteorological model toward physically correct solutions, particularly with respect to the split between glacier- vs. snowpack-sourced summer meltwaters.

Detailed benchtesting application of the first-generation Absynthe application to one catchment suggested good performance. Production-line applications in various different hydroclimatic environments led to further changes and improvements to the technique, and these second-generation Absynthe calibrations are producing parameter sets that will be used operationally. In addition, it is planned that the parameter set ensembles will be used in conjunction with NWP weather ensembles (for daily forecasting) and ESP climate traces and/or stochastic weather generators (for seasonal forecasting) to generate fully probabilistic forecasts.

The key advantages of the Absynthe method are that it (i) effectively provided some useful insights into the properties and performance of the watershed model, (ii) it yields good hydrologic predictions, where “good” is defined on a multi-objective statistical basis, and (iii) it accomplishes this statistical performance while ensuring physical hydrologic plausibility. The assurance of physical reasonableness is an important asset that distinguishes this method from many other optimization and uncertainty estimation techniques.

The main disadvantages stem from the purely random search method employed by this GLUE-based method. Specifically, the technique tends to be very slow to converge on good parameter sets. By the same token, for given specified behavioural criteria, feasible parameter ranges, and run times or ensemble sizes, it is difficult to ascertain whether the best-performing parts of the feasible parameter space have been sampled.

ACKNOWLEDGEMENTS

The authors wish to thank the following colleagues, who provided valuable feedback early in the recalibration method development process: D. Hutchinson and A. Cannon (Environment Canada); D. Roche (formerly UBC Civil Engineering under M. Quick, now at Kerr Wood Leidal); Y.-H. Lan (formerly UBC Civil Engineering under M. Quick, now at GVRD); M. Schnorbus (formerly BC MOE River Forecast Centre, now at Pacific Climate Impacts Consortium); C. Walford (Alberta Environment River Forecast Centre); C. Mathieu and N. Evora (Hydro-Québec); and P. Larsson (Vattenfall). The following BC Hydro personnel also contributed valuable perspectives: V. Plesa, D. Nixon, J. McNaughton, M. Jasek, K. Meidal, D. Druce, S. Smith, Z. Micovic, and G. Sreckovic. Additional helpful feedback was provided by K. Beven (Lancaster University), and M. Demuth (Geological Survey of Canada) provided us with time series of seasonal and annual glacier mass balance.

We also thank J. Boyd (BC Hydro) for contributing some BC Hydro background material appearing in the text. Finally, we particularly thank Kim Dargent (Accenture) for providing technical support at some key stages of the software development process.

REFERENCES

Babu, B.V., Chakole, P.G., Mubeen, J.H.S., 2005: Multiobjective differential evolution (MODE) for optimization of adiabatic styrene reactor. *Chemical Engineering Science*, 60, 4822-4837.

Beldring, S., 2002: Multi-criteria validation of a precipitation-runoff model. *Journal of Hydrology*, 257, 189-211.

Beven, K., 2006: A manifesto for the equifinality thesis. *Journal of Hydrology*, 320, 18-36.

Beven, K., Binley, A., 1992: The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes*, 6, 279-298.

Binley, A., Buckley, K., Calore, C., Parodi, U., La Barbera, P., 1997: Modelling uncertainty in estimates of recharge to a shallow coastal aquifer. *Hydrological Sciences Journal*, 42, 155-168.

Block, P.J., Filho, F.S., Sun, L., Kwon, H., 2007: *Accounting for Uncertainty Propagation: A Streamflow Forecasting Framework using Multiple Climate and Hydrological Models*. AGU Fall Meeting, Dec 10-14, San Francisco CA.

Chu, W., Gao, X., Sorooshian, S., 2007: *A Hierarchical Approach to Distributed Parameter Estimation in Rainfall-Runoff Modeling.* AGU Fall Meeting, Dec 10-14, San Francisco CA.

Daly, C., Gibson, W.P., Taylor, G.H., Johnson, G.L., Pasteris, P., 2002: A knowledge-based approach to the statistical mapping of climate. *Climate Research*, 22, 99-113.

Duan, Q., 2003: Global optimization for watershed model calibration. In: *Calibration of Watershed Models*, Duan Q, Gupta HV, Sorooshian S, Rousseau AN, Turcotte R (eds). American Geophysical Union, Washington DC, p. 89-104.

Duan, Q., Sorooshian, S., Gupta, V., 1992: Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28, 1015-1031.

Duan, Q., Sorooshian, S., Gupta, V., 1993: Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology*, 158, 265-284.

Duffy, C.J., Al-Hassan, S., 1988: The time and frequency response of tracer experiments. *Journal of Hydrology*, 97, 59-73.

Fleming, S.W., 2007: Artificial neural network forecasting of nonlinear Markov processes. *Canadian Journal of Physics*, 85, 279-294.

Fleming, S.W., Haggerty, R., 2001: Modeling solute diffusion in the presence of pore-scale heterogeneity: method development and an application to the Culebra dolomite Member of the Rustler Formation, New Mexico, USA. *Journal of Contaminant Hydrology*, 48, 253-276.

Fleming, S.W., Lavenue, A.M., Aly, A.H., Adams, A., 2002: Practical applications of spectral analysis to hydrologic time series. *Hydrological Processes*, 16, 565-574.

Fleming, S.W., Trehu, A.M., 1999: Crustal structure beneath the central Oregon convergent margin from potential-field modeling: evidence for a buried basement ridge in local contact with a seaward dipping backstop. *Journal of Geophysical Research*, 104, 20431-20447.

Fleming, S.W., Weber, F.A., 2008: *Parameter uncertainty estimation for an operational river forecast model using a modified GLUE architecture.* HEPEX Sub-Group Meeting on Uncertainty Post-Processing for Hydrologic Ensemble Prediction, June 23-25, Delft, Netherlands.

Gallagher, M., Doherty, J., 2007: Parameter estimation and uncertainty analysis for a watershed model. *Environmental Modelling and Software*, 22, 1000-1020.

Garen, D.C., 1992: Improved techniques in regression-based streamflow volume forecasting. *ASCE Journal of Water Resources Planning and Management*, 118, 654-670.

Georgakakos, K.P., Seo, D.-J., Gupta, H., Schaake, J., Butts, M.B., 2004: Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. *Journal of Hydrology*, 298, 222-241.

Gill, M., Wigmosta, M., Coleman, A., Prasad, R., Vail, L., 2007: *Ensemble Streamflow Forecasting via Data Assimilation*. AGU Fall Meeting, Dec 10-14, San Francisco CA.

Goddard, L., Barnston, A.G., DeWitt, D.G., Mason, S.J., Roberston, A.W., Tippett, M.K., 2007: *Multi-Model Ensembling for Seasonal-to-Interannual Climate Forecasting at the IRI*. AGU Fall Meeting, Dec 10-14, San Francisco CA.

Goswami, M., O'Connor, K.M., 2007: Comparative assessment of six automatic optimization techniques for calibration of a conceptual rainfall-runoff model. *Hydrological Sciences Journal*, 52, 432-449.

Gupta, H.V., Bastidas, L.A., Vrugt, J.A., Sorooshian, S., 2003b: Multiple criteria global optimization for watershed model calibration. In: *Calibration of Watershed Models*, Duan Q, Gupta HV, Sorooshian S, Rousseau AN, Turcotte R (eds). American Geophysical Union, Washington DC, p. 125-132.

Gupta, H.V., Sorooshian, S., Hogue, T.S., Boyle, D.P., 2003a: Advances in automatic calibration of watershed models. In: *Calibration of Watershed Models*, Duan Q, Gupta HV, Sorooshian S, Rousseau AN, Turcotte R (eds). American Geophysical Union, Washington DC, p. 9-28.

Haggerty, R., Fleming, S.W., Meigs, L.C., McKenna, S.A., 2001: Tracer tests in a fractured dolomite, 2, analysis of mass transfer in single-well injection-withdrawal tests. *Water Resources Research*, 37, 1129-1142.

Hay, L.E., Umemoto, M., 2006: *Multiple-Objective Stepwise Calibration using Luca*. USGS Open-File Report 2006-1323. US Geological Survey, Reston VA.

Hogue, T.S., Gupta, H.V., Sorooshian, S., Tomkins, C.D., 2003: A multi-step automatic calibration scheme for watershed models. In: *Calibration of Watershed Models*, Duan Q, Gupta HV, Sorooshian S, Rousseau AN, Turcotte R (eds). American Geophysical Union, Washington DC, p. 165-174.

Hsieh, W.W., 2001: Nonlinear canonical correlation analysis of the tropical Pacific climate variability using a neural network approach. *Journal of Climate*, 14, 2528-2539.

Jia, Y., Culver, T.B., 2008: Uncertainty analysis for watershed modelling using generalized likelihood uncertainty estimation with multiple calibration measures. *ASCE Journal of Water Resources Planning and Management*, 134, 97-106.

Kling, H., Gupta, H.V., 2007: *Relationship between Spatial Discretization and the Parameters and Model Performance of Precipitation-Runoff and Water Balance Models*. AGU Fall Meeting, Dec 10-14, San Francisco CA.

Kuczera, G., Parent, E., 1998: Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *Journal of Hydrology*, 211, 69-85.

Lan, Y.-H., 2001: *A Genetic-Algorithm Based Automatic Model Calibrator for the UBC Watershed Model*. M.A.Sc. Thesis. University of British Columbia, Vancouver BC.

Li, J., 2005: *A Neural Network Model for Flood Forecasting for Small Hydro Plants*. M.A.Sc. Thesis. University of British Columbia, Vancouver BC.

Lindström, G., 1997: A simple automatic calibration scheme for the HBV model. *Nordic Hydrology*, 28, 153-168.

Liu, Y., Gupta, H., Durcik, M., Wagener, T., 2007: *A Novel Approach to Parameter Identification for Distributed Hydrologic Modelling*. AGU Fall Meeting, Dec 10-14, San Francisco CA.

McMillan, H.K., Clark, M.P., Ibbitt, R.P., 2007: *Calibrating a Distributed Hydrological Model in a Basin with Heterogeneous Geology*. AGU Fall Meeting, Dec 10-14, San Francisco CA.

Mendes, B.S., Draper, D., 2007: *Estimating Different Regimes in a Tracer Breakthrough Curve with Bayesian Statistics*. AGU Fall Meeting, Dec 10-14, San Francisco CA.

Pokrhel, P., Gupta, H.V., Wagener, T., 2007: *Distributed Parameter Estimation Using a Regularization Approach*. AGU Fall Meeting, Dec 10-14, San Francisco CA.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992: *Numerical Recipes in Fortran 77: The Art of Scientific Computing, 2nd Ed*. Cambridge University Press, Cambridge UK.

Quick, M.C., 1995: *UBC Watershed Model Manual, Version 4.0*. University of British Columbia, Vancouver BC.

Roche, A.D., 2005: *Effects of Uncertainty in Hydrologic Model Calibration on Extreme Event Simulation*. M.A.Sc. Thesis. University of British Columbia, Vancouver BC.

Samaniego, L., Kumar, R., Attinger, S., 2007: *Adaptive Parameter Optimization of a Grid-Based Conceptual Hydrologic Model*. AGU Fall Meeting, Dec 10-14, San Francisco CA.

Seibert, J., 1997: Estimation of parameter uncertainty in the HBV model. *Nordic Hydrology*, 28, 247-262.

Seibert, J., 2000: Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. *Hydrology and Earth System Sciences*, 4, 215-224.

Seibert, J., Uhlenbrook, S., Leibundgut, C., Halldin, S., 2000: Multiscale calibration and validation of a conceptual rainfall-runoff model. *Physics and Chemistry of the Earth*, 25, 59-64.

Shi, X., Wood, A.W., Lettenmaier, D.P., 2007: *The Effect of Hydrologic Model Calibration on Seasonal Streamflow Forecasts for the Western US*. AGU Fall Meeting, Dec 10-14, San Francisco CA.

Smith, M.B., Laurine, D.P., Koren, V.I., Reed, S.M., Zhang, Z., 2003: Hydrologic model calibration in the National Weather Service. In: *Calibration of Watershed Models*, Duan Q, Gupta HV, Sorooshian S, Rousseau AN, Turcotte R (eds). American Geophysical Union, Washington DC, p. 133-152.

Smith, T., Marshall, L., 2007: *A Bayesian Uncertainty Framework for Conceptual Snowmelt and Hydrologic Models Applied to the Tenderfoot Creek Experimental Forest*. AGU Fall Meeting, Dec 10-14, San Francisco CA.

Storn, R., Price, K., 1997: Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11, 341-359.

Tolson, B.A., Craig, J.R., Esfahani, M.A., 2007: *Incorporating Search History into the Dynamically Dimensioned Search (DDS) Optimization Algorithm*. AGU Fall Meeting, San Francisco CA.

Tolson, B.A., Shoemaker, C.A., 2007: Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, 43, doi:0.1029/2005WR004723.

Unkrich, C., Yatheendradas, S., Gupta, H., Wagener, T., Goodrich, D., Schaffner, M., Stewart, A., 2007: *Forecast Uncertainty in Semi-Arid Flash Flood Modeling Using Radar Rain Input*. AGU Fall Meeting, Dec 10-14, San Francisco CA.

Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003: A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, 39, doi: 10.1029/2002WR001642.

Vrugt, J.A., ter Braak, C.F., Robinson, B.A., Hyman, J.M., 2007: *Speed-Up of Markov Chain Monte Carlo Simulation Using Self-Adaptive Different Evolution with Subspace Sampling*. AGU Fall Meeting, Dec 10-14, San Francisco CA.

Wagener, T., Bhushan, R., Zhang, Z., Reed, P., 2007: *A New Approach to the Calibration of Complex Watershed Models in Ungauged Basins Using Regionalized Hydrologic Indices*. AGU Fall Meeting, Dec 10-14, San Francisco CA.

Wagener, T., Kollat, J., 2007: Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox. *Environmental Modelling and Software*, 22, 1021-1033.

Wagener, T., Wheater, H.S., Gupta, H.V., 2004: *Rainfall-Runoff Modelling in Gauged and Ungauged Basins*. Imperial College Press, London UK.

Wang, Q.J., 1991: The genetic algorithm and its applications to calibrating conceptual rainfall-runoff models. *Water Resources Research*, 27, 2467-2471.

Wang, T., Hamann, A., Spittlehouse, D., Aitken, S.N., 2006: Development of scale-free climate data for western Canada for use in resource management. *International Journal of Climatology*, 26, 383-397.

Weber, F., 2001: *Quality Control of Hydrologic Data for water use planning (WUP) applications*. CWRA – BC Branch Conference: Changing Water Environments: Research and Practice. May 8-11, 2001, Whistler, BC*.*

Weiss, E., 2001: *Integrating the UBC watershed model into a river forecast system.* CWRA – BC Branch Conference: Changing Water Environments: Research and Practice. May 8-11, 2001, Whistler, BC.

Xiong, L., Wan, M., Wei, X., O'Connor, K.M., 2009: Indices for assessing the prediction bounds of hydrological models and application by generalised likelihood uncertainty estimation. *Hydrological Sciences Journal*, 54, 852-781.

Yilmaz, K.K., Gupta, H.V., Wegener, T., 2007: *A Signature Index Approach to Diagnostic Evaluation and Parameter Estimation of Watershed Models*. AGU Fall Meeting, Dec 10-14, San Francisco CA.

Zhang, G., Fenicia, F., Savenije, H.H., 2007: *Parameter Estimation for a Physically Based Model Using Multi-Objective Approach Constrained with Internal States*. AGU Fall Meeting, Dec 10-14, San Francisco CA.

Table 1 Preliminary comparison of model performance, existing (manual) and Absynthe calibrations.

	Current RFS (manual calibration)	Best (by global likelihood function) Absynthe ensemble member	Best (by daily Nash-Sutcliffe) Absynthe ensemble member
Nash-Sutcliffe efficiency, daily mean flow rate	0.73	0.70	0.74
Maximum monthly volume bias	128%	47%	80%
Nash-Sutcliffe efficiency, annual flow volume	0.81	0.79	0.78

Table 2 Benchtesting Absynthe optimization for Alouette, with and without moving (rank-based) constraints

	Screening on	Screening off
Run time (hrs)	14.9	15.7
% of generated sets that are non-pathological	2.4	n/a
% of generated sets accepted	0.2	7.2
% of accepted sets that are pathological	n/a	96.0

Table 3 Results from production-line calibration of BC Hydro forecast watersheds to date. Results are shown for both the current (2009) calibration, performed using Absynthe in conjunction with updated data sources and composite meteorological stations, as well as the previous (2001) calibration which served as the basis of the operational RFS and which the new results are intended to replace.

Basin	Calibration	Total Period	Daily Nash-Sutcliffe	Monthly % Volume Error												Yearly % Volume Error	Yearly R^2
				Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		
Alouette	2001	1984-1998	0.73	-20	-7	-9	-1	3	5	18	58	127	27	-6	-10	1.0	0.82
	2009	1987-2007	0.80	-18	-1	-3	8	22	8	11	36	26	7	-3	-15	0.4	0.81
Arrow	2001	1984-1994*	0.76	-36	-29	2	-11	-7	8	-5	9	22	8	0	-13	-0.7	0.17
	2009	1970-2007	0.89	9	14	12	7	6	4	1	-3	6	7	7	4	4.1	0.65
Donald	2001	1965-1998*	0.95	20	9	-18	-29	-4	3	-4	4	1	9	21	27	0.9	0.74
	2009	1989-2007	0.95	-4	-6	-9	6	-1	3	1	-6	-5	2	13	6	-0.4	0.79
Duncan	2001	1987-1999*	0.87	-21	0	-14	-24	-13	9	2	-15	-7	12	5	-10	-3.7	0.74
	2009	1967-2007	0.92	-2	9	-3	5	4	1	1	0	1	14	8	-1	2.1	0.70
Mica Local	2001	1965-1998*	0.92	17	9	7	-16	-5	4	-3	1	-11	8	25	35	0.4	0.68
	2009	1972-1989	0.92	3	-1	6	7	7	6	2	-9	-10	10	13	15	1.9	0.74
Parsnip	2001	1981-1997*	0.82	-14	-5	-18	-12	-1	0	-7	15	28	1	-25	-22	-1.6	0.07
	2009	1980-2004	0.82	1	25	18	13	5	-3	2	9	8	-5	-11	-21	2.0	0.31
Revelstoke	2001	1984-1994*	0.91	21	45	0	-24	-1	3	-6	7	4	4	-4	23	0.6	0.37
	2009	1984-2007	0.92	3	2	7	-2	-1	4	-3	2	7	6	-9	-5	0.2	0.80
Sugar	2001	1985-1998	0.88	20	43	27	-12	-1	2	2	15	23	12	0	-1	4.3	0.66
	2009	1974-2007	0.86	13	4	2	0	7	5	2	17	18	13	7	2	6.0	0.50

Table 4 Inclusion statistics for Absynthe envelope, defined as the minimum and maximum predicted values of daily flow for a given day drawing from 100 GLUE ensemble members.

Basin	% observations within envelope	% observations below envelope	% observations above envelope
Alouette	65.8	23.0	11.2
Mica Local	69.9	13.5	16.7
Donald	71.3	8.9	19.8
Revelstoke	73.2	13.0	13.8
Duncan	71.9	8.4	19.7
Sugar	70.0	13.0	16.9
Parsnip	69.7	16.5	13.8
Arrow	68.3	16.2	15.5

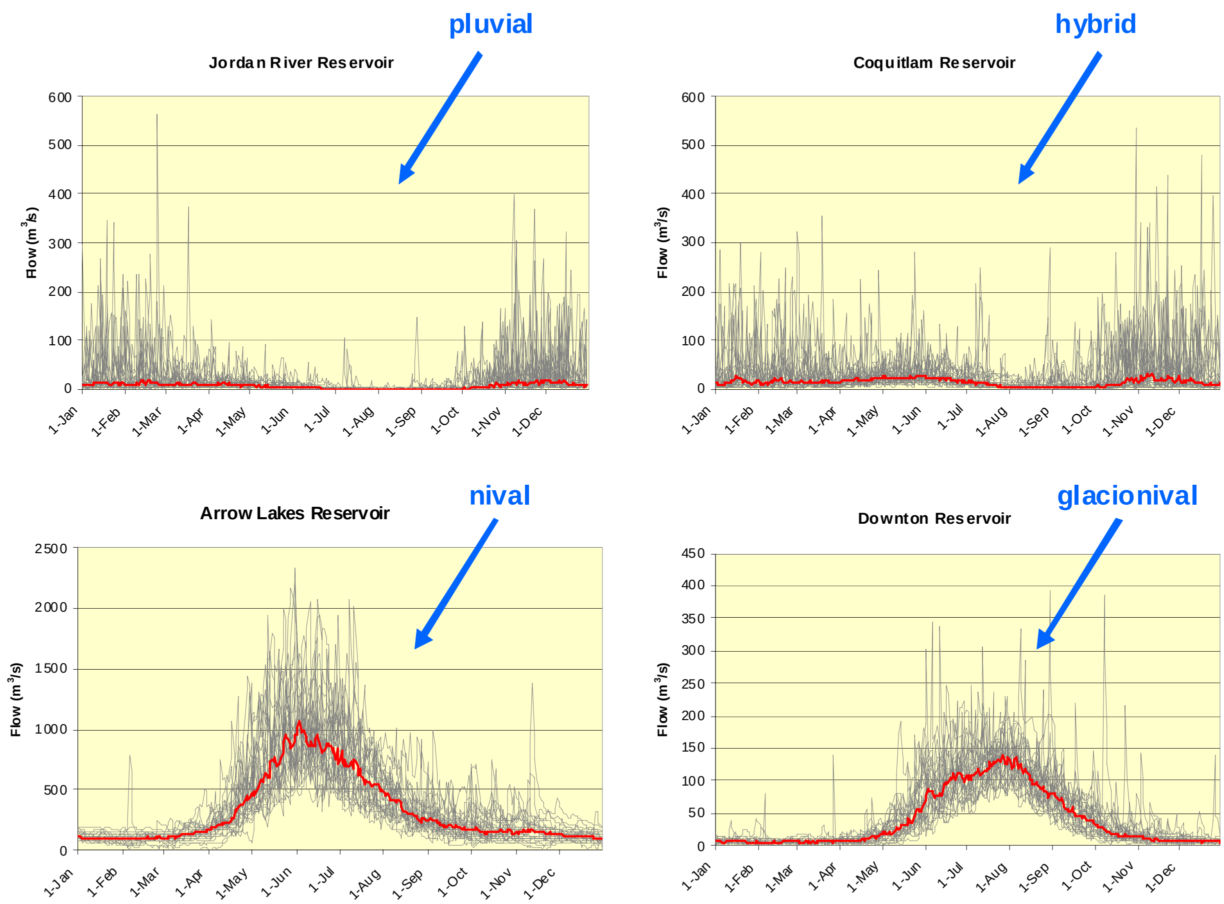

Figure 1 Streamflow regimes in British Columbia

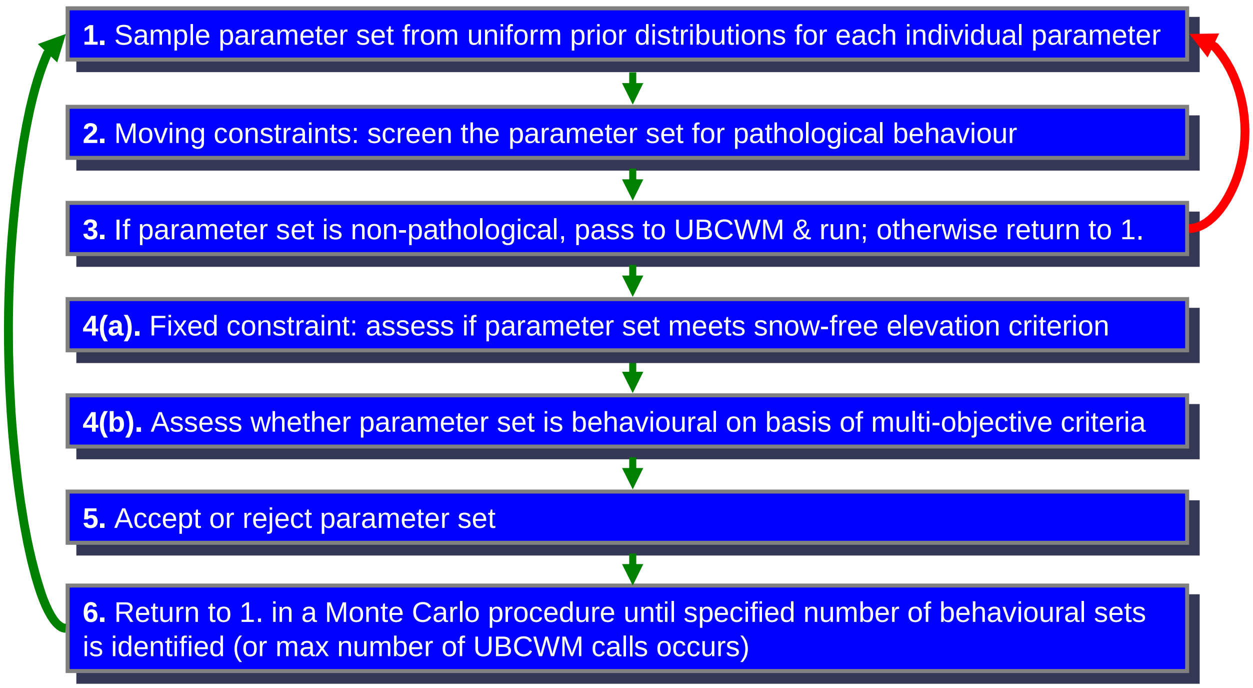

Figure 2 Flow chart for second-generation Abysnthe parameter optimization and uncertainty estimation algorithm

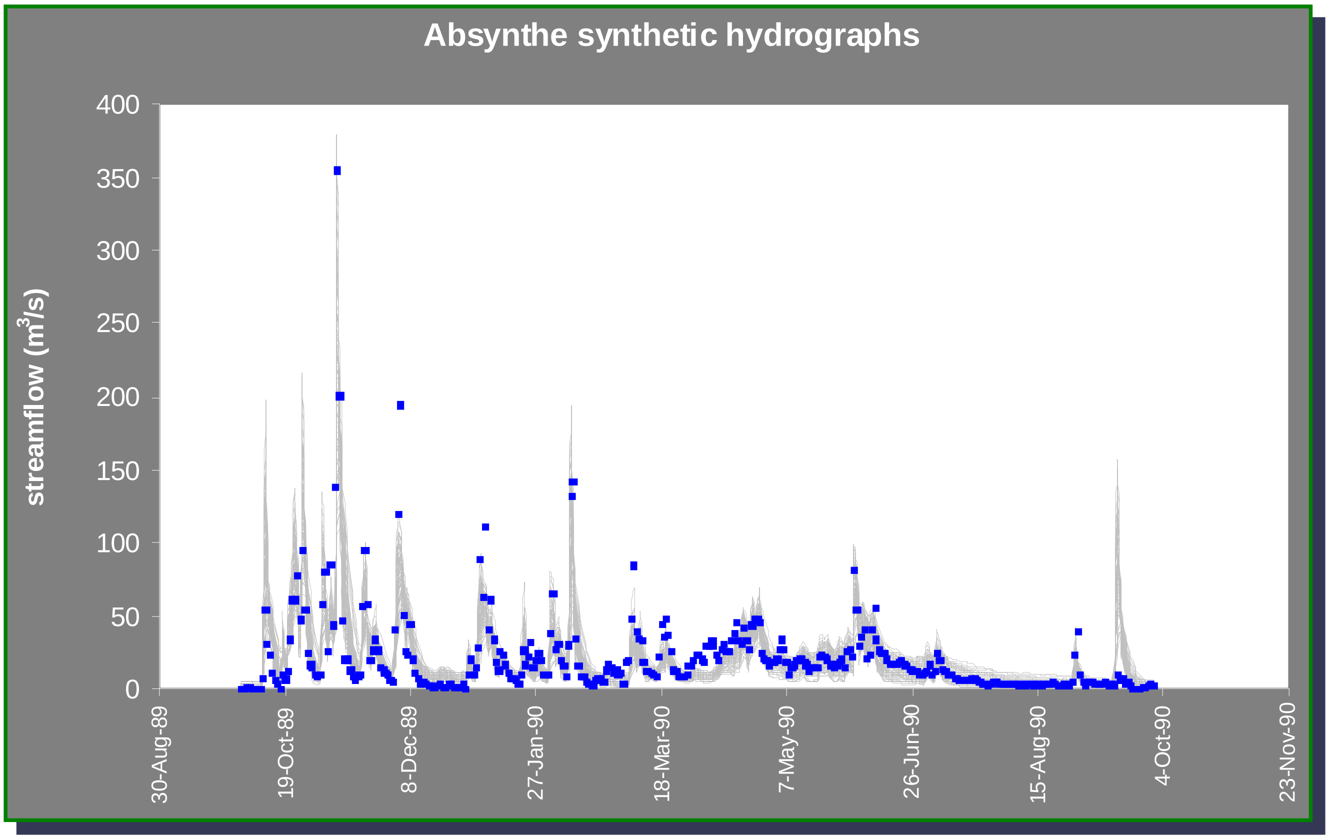

Figure 3 Alouette, 1989-90 water year, data (blue) and Abysnthe ensemble hydrographs (grey).

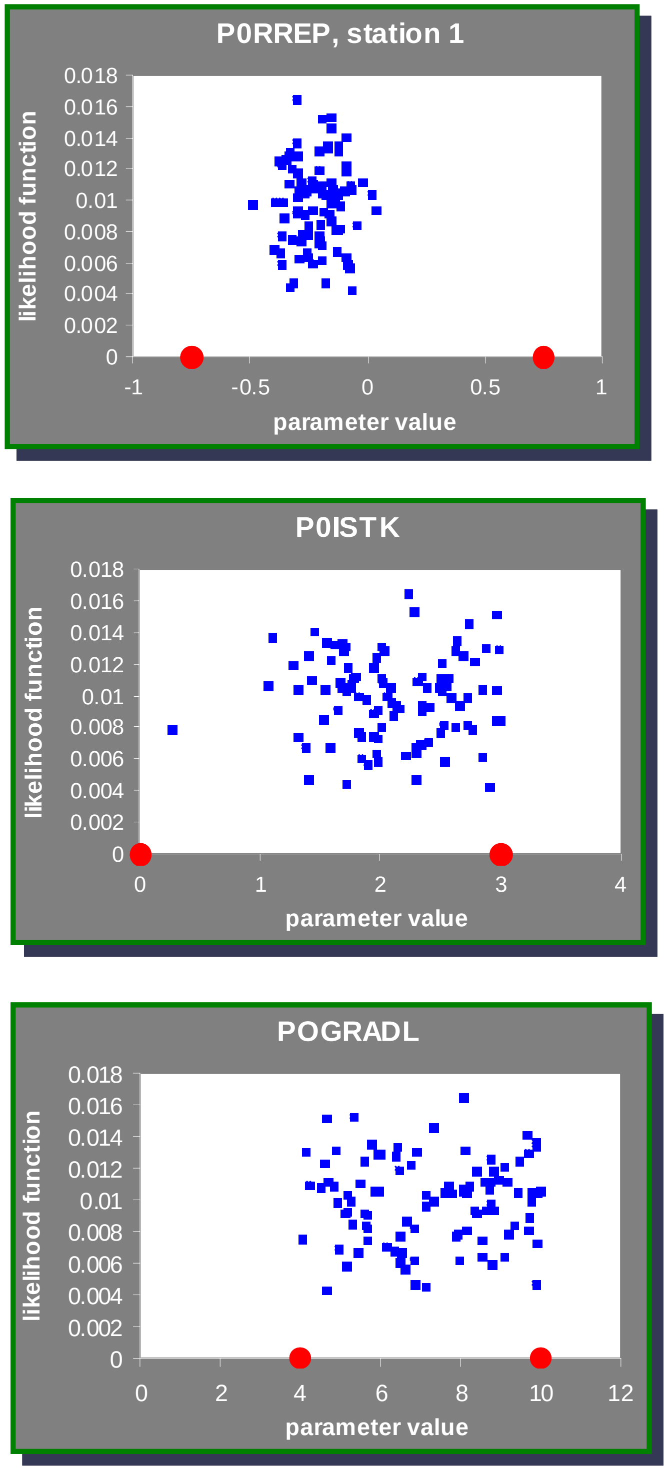

Figure 4 Example dotty plots for benchtesting Alouette optimization. Red dots indicate the specified limits of the noninformative Bayesian priors.

	A0TLZP	A0TLXM	A0TLNM	P0TEDL	P0GRADL	P0GRADM	P0GRADU	E0LMID	V0FLAS	P0PERC	P0DZSH	P0FRTK	P0FSTK	P0GLTK	P0IRTK	P0ISTK	P0UGTK	P0DZTK	P0SREP (1)	P0SREP (2)	P0RREP (1)	P0RREP (2)
A0TLZP																						
A0TLXM	-0.08																					
A0TLNM	-0.16	-0.03																				
P0TEDL	0.19	-0.03	-0.11																			
P0GRADL	-0.12	-0.16	0.06	-0.04																		
P0GRADM	-0.03	0.07	0.06	0.15	0.04																	
P0GRADU	-0.07	0.01	-0.06	-0.10	0.02	-0.18																
E0LMID	-0.19	0.34	-0.09	-0.03	0.07	0.13	0.12															
V0FLAS	-0.05	0.03	0.18	-0.10	0.01	0.00	-0.27	-0.17														
P0PERC	-0.10	-0.13	-0.06	0.01	-0.09	-0.16	-0.13	-0.12	0.08													
P0DZSH	-0.24	-0.05	-0.07	0.05	0.09	0.11	0.00	0.15	-0.08	-0.25												
P0FRTK	0.12	-0.01	0.12	0.13	0.07	0.18	-0.08	-0.09	0.00	-0.12	0.05											
P0FSTK	0.17	0.01	0.03	0.05	0.07	0.05	-0.18	-0.13	0.04	0.07	-0.05	0.64										
P0GLTK	0.30	0.06	0.01	-0.02	0.19	0.16	-0.10	0.06	-0.03	0.03	-0.22	0.45	0.72									
P0IRTK	0.15	0.07	-0.02	-0.08	0.11	0.14	-0.01	0.23	-0.04	0.07	-0.18	0.25	0.42	0.63								
P0ISTK	0.19	0.03	0.16	0.13	-0.12	-0.09	-0.07	-0.10	0.01	-0.01	-0.23	0.17	0.36	0.35	0.13							
P0UGTK	-0.20	-0.10	-0.06	-0.16	0.02	0.06	0.00	0.10	-0.07	0.03	0.04	-0.01	0.07	0.12	0.16	-0.09						
P0DZTK	-0.06	0.03	0.00	-0.21	-0.09	-0.19	0.01	0.01	0.02	0.11	-0.22	-0.07	-0.13	0.02	0.17	0.06	-0.08					
P0SREP (1)	-0.18	-0.18	-0.08	-0.02	-0.12	-0.04	-0.07	0.08	0.08	0.07	-0.05	0.10	0.02	-0.10	0.02	0.14	0.06	0.12				
P0SREP (2)	-0.06	0.10	0.07	-0.30	0.04	-0.06	0.10	-0.16	0.03	-0.04	-0.02	-0.09	-0.14	-0.02	0.01	-0.03	0.04	0.16	-0.12			
P0RREP (1)	0.17	-0.02	0.09	-0.03	-0.11	-0.52	0.05	-0.49	0.23	0.18	-0.26	-0.15	-0.02	-0.04	-0.13	0.03	-0.07	-0.03	-0.45	0.12		
P0RREP (2)	0.17	-0.15	0.05	-0.02	0.09	-0.11	-0.06	-0.09	0.09	0.03	-0.08	0.14	0.02	0.06	0.03	0.06	0.01	-0.13	-0.02	0.04	0.17	

Figure 5 Correlation matrix between 21 free UBC Watershed Model parameters across 100 behavioural and non-pathological parameter sets identified in preliminary Absynthe optimization using Alouette as a proving ground. Correlation coefficients ≥ 0.30 are highlighted in red.

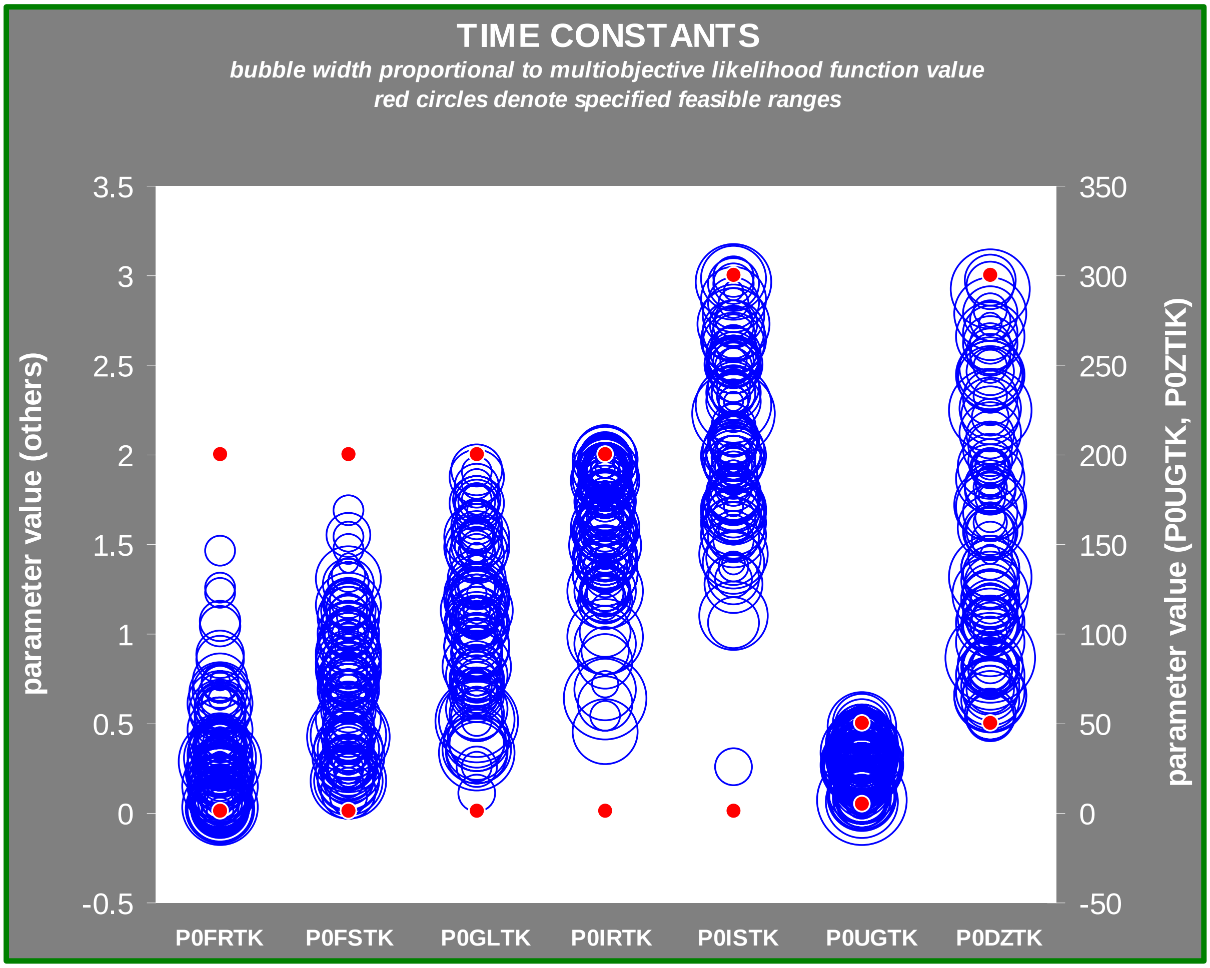

Figure 6 Bubble plots for terrestrial hydrologic time constants as estimated in Alouette benchtesting of the first-generation Absynthe software. Red dots indicate the pre-specified feasible ranges for each of the parameters shown.