Toward Improved Global Multi-Objective Optimization for Solving Water Resources Engineering and Decision-Making Problems

By

Shahram Sahraei

A thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in conformity with the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Civil Engineering
University of Manitoba
Winnipeg

Abstract

Water resources engineering problems that are solved in a simulation-optimization framework often have multiple conflicting objectives that form a tradeoff. Global multi-objective (MO) optimization has been successful in solving complex water resources problems. Naturally, these problems often have more than three objectives, making MO algorithms inefficient. These problems are also characterized by multi-modal properties meaning that multiple optimal and near-optimal tradeoffs exist. To deal with multi-modal MO problems, a modern MO global optimization algorithm should be capable of identifying a well-distributed set of solutions both in the decision and objective spaces, but minor attention has been given to the solution diversity in the decision space. Diversity maintenance in the decision space can help the algorithm archive optimal and near-optimal solutions for flexible and well-informed *a posteriori* decision-making.

The first contribution of this research was developing and testing a novel though simple approach to tackle the dominance resistance for solving many-objective water resources problems, i.e. problems with more than three objectives. The proposed approach, called rounded archiving, does not need any changes in the structure of the MO algorithm because it rounds the value of the objectives to the user-specified precision level prior to the dominance check. The method is implemented and assessed for multiple MO algorithms. The results show that the rounded archiving used for solving many-objective applications effectively reduces the archive size by up to 87% for algorithms with unbounded archive structure compared to the original archiving strategy in these algorithms. It effectively tackles the deterioration issue in algorithms with bounded archive size by eliminating similar solutions in the objective space and preventing them from populating the bounded archive.

The necessity to tackle the multi-modal characteristic of water resources problems led this research to develop a novel cluster-based solution archiving strategy to preserve a diverse set of solutions for MO algorithms. Solutions are dynamically clustered in the decision space and alternative solutions that are distant from each other are kept in the archive of good solutions, even

if they are globally dominated by solutions from outside their cluster. Results show that, the proposed method helps the MO algorithm identify good quality solutions that belong to significantly different parts of the decision space, provides a larger archive size, and detects optimal and near-optimal tradeoffs with distinguished cluster labels. Such a diverse set of solutions (policies) give decision-makers high flexibility in refining their preferences for robust decision-making.

The cluster-based optimization developed in this research is also used for approximating parameter uncertainty of hydrologic models in a MO fashion. This algorithm finds distinct parameter sets that satisfy the so-called acceptance thresholds in each optimization trial. The proposed method is found to be either equally effective to or perform better than popular single-objective uncertainty methods for deriving 95% prediction bounds. Furthermore, it is found that the acceptance threshold defined for the aggregated objective formulation in a single-objective uncertainty approximation method introduces more uncertainty to behavioral solution set identification, and these thresholds need to be individually specified for each objective in an MO uncertainty approximation method. A multi-threshold assignment for behavioral solution identification shrinks the behavioral solution space and reduces the size of the behavioral solution set by up to 98% compared to a single-threshold assignment for the aggregated objective formulation.

Further research on uncertainty analysis motivated the development of a new multi-modelling framework. Multi-modelling aims to reduce the structural uncertainty by using the strengths of single hydrologic models for improving the accuracy of the simulations of the system behavior. Prior to multi-modelling, each model should be individually calibrated based on hydrological signature metrics for better parameter identifiability. The calibrated models are then aggregated using a novel weighted average model-wrapper based on flow duration curve segmentation. The proposed model-wrapper is developed for four different operational flood forecasting models of the Upper Assiniboine River Basin. The model-wrapper is found to perform better than the individual models, and the weights associated with each model indicate its contribution rate to the

model-wrapper in all ranges of streamflow simulation. The proposed framework aids in model selection strategy, especially when a hydrologic model has minimal contribution to the model-wrapper performance.

Acknowledgements

I would first like to express my sincere gratitude to Dr. Masoud Asadzadeh for his patience and continuous guidance throughout my doctoral program. I genuinely appreciate his unstoppable support and encouraging comments, and the time he dedicated to the successful conclusion of this research.

Also, I would like to thank my internal advisory committee members, Dr. Tricia Stadnyk from the Department of Geography, University of Calgary and Dr. Ian Jeffrey from the Department of Electrical and Computer Engineering, the University of Manitoba for their encouragements and mindful comments at different stages of this thesis.

I would like to acknowledge the computational resource and financial supports that I received from the Hydrologic Forecast Centre of Manitoba Infrastructure. Without their support, this thesis would not have been possible.

Finally, thank you to my parents Nasrat and Abdullah, who supported their children for the best education while under tremendous financial burden. Reaching this stage in my life without their sacrifice was impossible. Mom and Dad, I hope I have made you proud.

Table of Contents

List of Figures

List of Tables

1 INTRODUCTION

1.1 Problem Statement

Water resources engineering problems are closely associated with design, operation, rehabilitation, ecological balance, and forecast. These problems are often cast as mathematical computer-based simulations that also involve optimization. Traditionally, a large number of studies formulated and solved water resources engineering problems including, but not limited to, hydrologic model calibration (Brazil and Krajewski, 1987; Duan et al., 1994; Jiang et al., 2015; Sorooshian et al., 1993; Wang, 1991), reservoir operation (Murray and Yakowitz, 1979; Wardlaw and Sharif, 1999), water distribution network design (Savic and Walters, 1997; Simpson et al., 1994; Tolson et al., 2004; Wu and Simpson, 2002), and pump operations (Lansey and Awumah, 1994; McCormick and Powell, 2003; Sakarya and Mays, 2000; Yu et al., 1994; Zessler and Shamir, 1989) as single-objective optimization problems. These problems, nonetheless, have multiple criteria that are important for informed decision-making but do not have a single solution that simultaneously optimizes all of them. Instead, a set of alternatives that is often called tradeoff, non-dominated front, or Pareto front after Vilfredo Pareto (1848-1923), exists since an improvement in the quantity of one objective function requires sacrificing at least one other objective function. Multi-objective optimization has been successfully applied to a wide range of water resources engineering problems including hydrologic model calibration (Budhathoki et al., 2020; Gupta et al., 1998; Holmes et al., 2020; Madsen, 2003; Pfannerstill et al., 2017; Yapo et al., 1998) , river-reservoir system operation (Asadzadeh et al., 2014a; Kim et al., 2008; Labadie et al., 2012; Tsoukalas and Makropoulos, 2015; Wan et al., 2018), water distribution network design (Farmani et al., 2005; Huang et al., 2020; Prasad and Park, 2004; Yazdi, 2016; Zhang et al., 2019), and

groundwater pumping operation (Barán et al., 2005; Carpitella et al., 2019; Kougias and Theodossiou, 2013; Sadatiyan A. and Miller, 2017).

Often, water resources engineering problems that are formulated with many (more than three) objectives are multi-modal, in that they have multiple solutions that have the same performance (Reed et al., 2013). Multi-modality arises from the wicked nature of water resources simulation-optimization problems due to insufficient understanding of the system, incomplete data, diverse stakeholders' interests, implicit uncertainties and/or simplification in these problems (Kasprzyk et al., 2013; Rittel and Webber, 1973; Zechman et al., 2013). For instance, a classical least-cost optimization problem assumes that an analyst has a perfect knowledge and problem formulation of the system under study that perfectly emulates future system states and responses, which is called an "omnipotent" system analysis (Liebman, 1976; Reed and Kasprzyk, 2009). For this reason, a single-criterion formulation, hence dimensionality reduction of objectives through *a priori* subjective weighted-sum aggregation of some or all objectives, narrows the mathematical definition of many-objective water resources optimization problems, which is not very attractive to an analyst for practical design and decision-making (Reed and Kasprzyk, 2009).

Hydrologic model calibration is an example of a water resources optimization problem that is mostly used in this thesis. A hydrologic model is a similar but simpler structure of the water cycle that involves a number of natural components including precipitation, snowmelt, interception, evapotranspiration, infiltration, sub-surface flow, surface flow, etc., and their interactions. Figure 1-1 demonstrates the components of a typical hydrologic model and shows how its parameters is calibrated. In fact, model calibration does not necessarily denote the perfect match between a simulated dataset to its corresponding measurement mainly due to the imperfect model structure, simplifying assumptions in simulation, uncertainty in measurements as well as spatial and temporal variations in

model parameters (Efstratiadis and Koutsoyiannis, 2010). More importantly, no single simulation error metric as the objective function of the model calibration can measure every important aspect of the fit between simulated and measured data (Gupta et al., 1998). Hence, a multi-objective model calibration using an efficient global optimization algorithm is necessary to identify a solution that is a tradeoff between all error metrics.

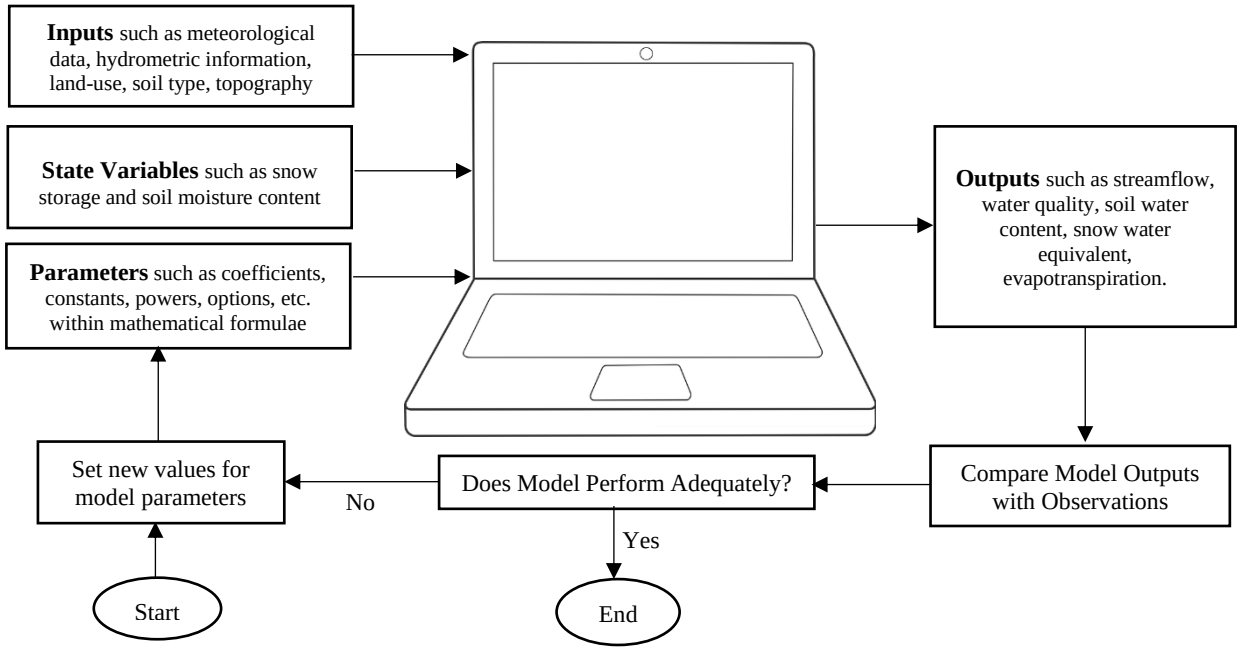


Figure 1-1: Schematic of a typical hydrologic model and model parameter calibration

The tradeoff solutions in multi-objective optimization are obtained solely based on the objective functions in the problem formulation, which may be socially undesirable or unable to reflect unmodelled or intangible priorities in water resources systems. Providing decision-makers with diverse and distinct alternatives (design options) that might not be optimal with respect to the modeled objectives but are close enough to the optimal alternatives can give decision-makers a great deal of flexibility to identify their most *preferred* solution. As a result, innovations in optimization and modelling are necessary to give a range of probable water futures and help

diversification of decision-making and our understanding of the system under study (Brill et al., 1990).

1.2 Research Objectives

This thesis is concentrated on the development of tools and modelling strategies that efficiently and effectively solve water resources problems formulated with many-objectives that can be multi-modal. The following sub-sections outline four inter-connected research goals of this thesis.

1.2.1 Many-Objective Water Resources Optimization

In the mid-1980's, researchers showed interest in optimizing problems that have multiple (more than one but no more than three) conflicting criteria using global optimization algorithms (Schaffer, 1985). Since then, multi-objective global optimization has been successfully applied in different fields including, but not limited to, medicine, chemistry, finance, computer science, environmental, and water resources (Coello et al., 2007). However, these algorithms have performed fairly poorly when applied to problems that have more than three objectives, because they normally return a large number of non-dominated solutions (Farina and Amato, 2002). Many-objective nature of water resources systems are demonstrated in the literature. For instance, Reed et al. (2007) solved a four-objective monitoring network design for a cost-effective contamination management and quality control in groundwater resources through long-term sampling of observation well networks. Kasprzyk et al. (2009) introduced a many-objective risk-based water utility portfolio planning problem over a 10-year planning horizon that allows purchasing water supplies using three market mechanisms of permanent rights, leases, and options with six conflicting objectives including supply cost, cost variability, reliability, average volume of surplus water, frequency of using leases, and dropped water transfers after one year of nonuse. Water

distribution network design and hydrologic model calibration problems are other examples that are set up respectively with five and fifteen objectives in Fu et al. (2013) and Shafii et al. (2017).

Figure 1-2 demonstrates the general procedure for solving a multi-objective problem by a multi-objective global optimization algorithm. With this approach, the Pareto front is estimated by each trial of the optimization algorithm, while the single-objective reformulation of the problem results in a single solution in each trial. The dominance check and selection strategies constitute the principal parts of a multi-objective global optimization algorithm to guide the search toward a faster convergence and a well-distributed Pareto front (Coello et al., 2007; Deb, 2001). An increase in the number of objectives leads to an increase in the non-dominated portion of the objective space, which in turn aggravates the similarity of solutions and exponentially expands the number of archived solutions in a multi-objective problem (Farina and Amato, 2004). This leads to dominance resistance in many-objective optimization meaning that finding any new dominating (better) solution becomes challenging due to high similarity in the objective space. Therefore, the structure of multi-objective optimization algorithms need to be modified by a specialized solution archiving strategy like epsilon-dominance (Laumanns et al., 2002) that grids the objective space to dimensionally meaningful boxes and preserves only the best non-dominated solution in each box for a sparse Pareto front. In this thesis, a rounded-archiving method is suggested that rounds the value of each objective to a pre-determined precision recommended by decision-makers without restructuring the optimization algorithms for solving many-objective problems to deal with the archive size and better distribution of solutions along the Pareto front.

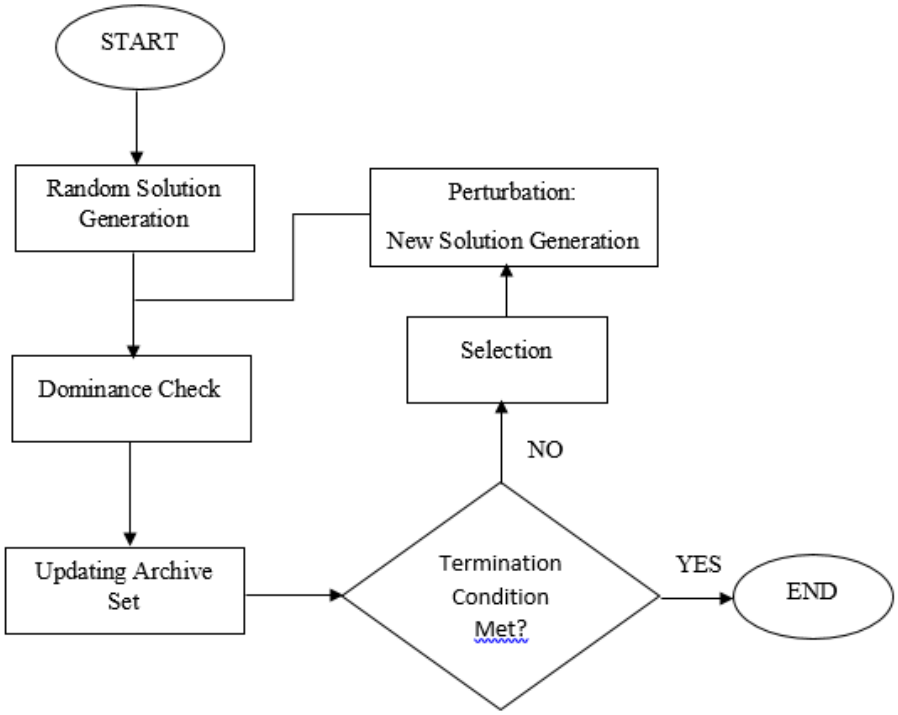


Figure 1-2: General structure of heuristic MO algorithms

1.2.2 Diversity Preservation in Water Resources Optimization

A robust multi-objective global optimization algorithm must converge to a diverse set of solutions (Deb, 2001). Solution diversity is divided into two categories, diversity in the objective space and diversity in the decision space. A large number of studies proposed novel methods to maintain a diverse set of solutions in the objective space. Nonetheless, minor attention is given to the importance of the solution diversity in the decision space. The diversity of solutions in the decision space is very important for two main reasons:

1. Distinct design options are required in the decision-making process, particularly when some stakeholders' expectations cannot be directly formulated mathematically.

2. The presence of distinct solutions allows decision-makers to review the available design options and flexibly change their decisions owing to alterations in the system in terms of policy, economy, geography, climate, social or cultural environments.

The current modern multi-objective algorithms in the field of water resources do not consider solution diversity in the decision space during the optimization and make highly subjective judgement about diversity measurement based solely on objective function values. Moreover, the obtained optimal solutions usually have highly similar characteristics in the decision space limiting the ability of decision-makers for finding a preferable option. This thesis argues that multi-objective optimization algorithms do not retain near-optimal design options that are meaningfully distant (dissimilar) from the optimal solutions in the decision space as they are dominated by the archived optimal Pareto solutions. In this thesis, in order to take the dissimilarity of solutions in the decision space into consideration, an approach is developed that links the density-based spatial clustering to the optimization. Unlike the majority of other optimization algorithms that tend to progressively move the search toward a globally optimal region, the proposed cluster-based optimization allows forming and populating multiple clusters of solutions in distinct regions of the decision space by giving a higher expansion opportunity to a cluster in the less crowded region of the decision space but closer to the reference or ideal objective function values.

1.2.3 Uncertainty Estimation in Hydrologic Model Calibration

Hydrologic model calibration is an example water resources engineering problem that is inherently a multi-objective optimization problem (Gupta et al., 1998). Hydrologic models are mathematical simulation tools that model our understanding of the hydrological processes in a watershed system. Uncertainty due to errors in input and output data, simplified model structure, initial model condition, model parameters, and subjective decisions may lead to over conditioning in model

calibration problems, i.e. an optimal parameter set may not be desirable and/or necessary. Therefore, an approximation of uncertainty associated with response time series in hydrologic modelling is crucial for an effective decision-making.

Available methods for uncertainty approximation broadly fall into two categories: formal Bayesian (Krzysztofowicz, 1999; Schoups and Vrugt, 2010; Thiemann et al., 2001; Vrugt et al., 2009) and "informal" approaches (Beven and Freer, 2001; Blasone et al., 2008b; Ragab et al., 2020; Uhlenbrook and Sieber, 2005) for deriving *posterior* distribution of parameters and model responses. The former approach explicitly separates the individual sources of uncertainty using strong statistical assumptions and evolves solutions into an initially assumed *posterior* probability distribution of model residuals, e.g. Gaussian output error structure (Beven et al., 2008; Schoups and Vrugt, 2010; Vrugt et al., 2009). The latter approach, on the other hand, rejects the concept of a correct model, acknowledges model *equifinality*, is structurally simple, and propagates all interacting sources of uncertainty into model parameter estimates by defining a subjective model evaluation metric, e.g. Nash-Sutcliffe Efficiency, for deriving the *posterior* total error distribution (Beven, 2006; Beven and Freer, 2001; Beven and Smith, 2015; Stedinger et al., 2008). The Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Freer, 2001) for parameter uncertainty is the most well-known uncertainty method that uses simple Monte Carlo sampling for approximating total uncertainty. Due to its computationally inefficient sampling, more specialized sampling strategies have been coupled with GLUE, for example Latin Hypercube Sampling (Uhlenbrook and Sieber, 2005), sensitivity-informed GLUE (Ratto et al., 2001), and gradient-based GLUE (Abbaspour et al., 2004). Other alternatives include low-budget single-objective optimization methods with many independent trials (in the order of hundreds of trials) is also suggested by Tolson and Shoemaker (2008) that can reduce the total computational budget to one

order of magnitude less than random sampling. These methods, however, can handle a single objective function and therefore should use one or aggregate multiple calibration objectives into one. Ideally, multiple calibration objectives must be analyzed in a multi-objective approach to be able to learn about different aspects of the model performance, for example in high-flow versus low-flow periods. Therefore, in this thesis, a multi-objective algorithm, primarily developed for maintaining diverse sets of optimal and near-optimal solutions through cluster-based archiving and localized dominance strategies, is proposed and used for approximating parameter uncertainty in hydrologic modelling in a multi-objective context.

1.2.4 Signature-Based Calibration and Multi-Modelling in Hydrologic Systems

Calibration of hydrologic model parameters is a water resources systems analysis problem that does not necessarily need a *global* optimal parameter set for a reliable forecasting management as its simulation is subject to the model initial condition, simplified model structure, errors in forcing data, and uncertainty inherent in physically unmeasurable model parameters. Many techniques are used to reduce uncertainty in hydrologic modelling such as data assimilation strategies for reducing uncertainty in model initial state and boundary condition (Beck, 1987; Evensen, 1992; Kitanidis and Bras, 1980a, 1980b; Moradkhani et al., 2005), suggesting precipitation multipliers for input uncertainty reduction (Kavetski et al., 2003; Vrugt et al., 2009), model combination techniques for reducing structural uncertainty (Abrahart and See, 2002; Ajami et al., 2006; Shamseldin et al., 1997), and adopting hydrologic signatures to reduce parameter uncertainty (McMillan, 2021)

The application of ensemble modelling is claimed to enhance modelling robustness and reduce prediction uncertainty (Li and Sankarasubramanian, 2012; Razavi and Coulibaly, 2016). The ensemble prediction systems can be created by multiple meteorological forcing (He et al., 2009), multiple sets of model parameters (McIntyre et al., 2005; Seibert and Beven, 2009), or simulations

of multiple models with distinct structures (Muhammad et al., 2018; Velázquez et al., 2010). The evaluation of the ensemble predictions is commonly performed by the weighted average (Shamseldin et al., 1997), quantile mapping (Hashino et al., 2007; Piani et al., 2010), and Bayesian model averaging (Duan et al., 2007; Parrish et al., 2012) of ensemble members. The weighted average is an optimal linear combination of the simulations by assigning weight to each ensemble member, with weights summing up to one (Ajami et al., 2006). The Quantile mapping is a bias correction method that adjusts the quantiles of the cumulative distribution function of the simulations according to the cumulative distribution function of observed response time series (Hashino et al., 2007; Wood and Schaake, 2008). Bayesian model averaging is essentially a weighted average of the probability density function associated with a set of individual model predictions concentrated around their forecasts (Raftery et al., 2005). An optimal multi-model ensemble at least partially mitigates model structural uncertainty by compensating for a particular model weakness with the help of other models (Ajami et al., 2007; Georgakakos et al., 2004; Li and Sankarasubramanian, 2012). Hence, it produces more reliable forecasts compared to the best calibrated individual models and a modeller can dig up and extract further information from a group of existing models (Shamseldin et al., 1997).

Another major source of uncertainty is the unknown value of the hydrologic model parameters. According to the concept of *equifinality* introduced by (Beven, 2006; Beven and Freer, 2001), model calibration using a traditional performance metric like sum of squared errors can lead to multiple *behavioral* parameter sets. Hence, these metrics do not supply sufficient information for diagnostic assessment of hydrologic models (Arnold et al., 2015; Guzman et al., 2015) or distinguishing different model structural forms, unless they are accompanied with hydrologic signatures, e.g. the flow duration curve (Westerberg et al., 2011; Yadav et al., 2007; Yilmaz et al.,

2008), soil moisture information (Rajib et al., 2016), isotope tracer data (Bergström et al., 2002; Holmes et al., 2020), and any information about hydrological components including, but not limited to, base flow (Arnold and Allen, 1999; Vrugt and Sadegh, 2013), annual evaporation and excess precipitation (Asadzadeh et al., 2015). Using the hydrologic signature metrics aids in reducing parameter uncertainty and achieving a more hydrologically consistent simulated system response.

In this thesis, a novel signature-based multi-modelling method is developed that combines the simulated response of multiple, structurally different models based on the flow duration curve quantiles of the measured streamflow to confront both model structural and parameter uncertainties for a more reliable operational streamflow forecasting, for example at the Hydrologic Forecast Centre of Manitoba Infrastructure.

1.3 Thesis Content, Research Contributions, and Scope

The remainder of this thesis is comprised of five chapters. The next four chapters closely mirror three peer-reviewed research publications and one article under the peer-review process, each with independent literature reviews, methods, results presentation and discussion. Despite ending with a brief conclusion in each chapter, extensive concluding remarks followed by recommendations for future work are also provided in Chapter 6.

The extensive results, discussions, and conclusions in this thesis are mostly focused on the hydrologic model calibration case studies. However, similar outcomes and conclusions, specifically for Chapter 2, Chapter 3, and Chapter 4 are expected in other water resources optimization problems for multi-criteria design, operation, and management. The rounded archiving method discussed in Chapter 2 is applicable to other multi-objective optimization

algorithms for solving many-objective problems. Likewise, the cluster-based archiving and dominance localization in Chapter 3 can be applied to other optimization algorithms for diversity maintenance in the decision space. The overall goal of this research is thus to enhance the performance of the multi-objective global optimization algorithms with a focus on hydrologic model calibration, uncertainty approximation, uncertainty reduction, and therefore aiding the decision-making process. The specific thesis achievements are listed as follows:

1) A novel solution archiving approach is developed and compared with existing archiving strategies for solving many-objective model calibration problems. The proposed approach rounds the objective values based on a pre-set required resolution for each objective to guarantee only meaningful discrepancies between solutions during the optimization. This approach significantly decreases the final number of the archived solutions for a less complicated decision-making process. The proposed archiving strategy is readily available for any optimization algorithm solving many-objective applications. In-depth information about the rounded archiving and comparison results are reported in Chapter 2. Chapter 2 is published in the Journal of Environmental Modelling and Software.

   Sahraei, S., Asadzadeh, M., & Shafii, M. (2019). Toward effective many-objective optimization: Rounded-archiving. *Environmental Modelling & Software*, 122, 104535.

2) A novel optimization algorithm called Cluster-based Pareto Archive Dynamically Dimensioned Search (CB-PA-DDS) is developed that considers distribution of candidate solutions in the decision space (also called parameter space in model calibration or design space in a design problem) through a density-based solution clustering. Similar solutions

are grouped and dissimilar solutions reside in distinct clusters in the decision space. Each cluster is evolved based on its distance to other clusters in the decision space and its contribution to the convergence. The quality of solutions in terms of convergence is judged locally in each cluster and solutions residing in two distinct clusters are not compared for convergence. This helps each cluster evolve independently and converge to a locally optimal tradeoff in the decision space. The proposed algorithm is successfully applied to two benchmark multi-modal mathematical problems and two benchmark environmental design and management problems. It offers highly distinct solutions on the entire feasible range of decision variables and gives a higher flexibility to decision-makers in the post-processing stage to refine their preferences. Details are further discussed in Chapter 3 published in the Journal of Environment Modelling and Software.

Sahraei, S., & Asadzadeh, M. (2021). Cluster-based multi-objective optimization for identifying diverse design options: Application to water resources problems. *Environmental Modelling & Software*, 135, 104902.

3) The Developed CB-PA-DDS algorithm is further studied to investigate its application in estimating hydrologic model parameter uncertainty in a multi-objective context. The available widely used methods such as Generalized Likelihood Uncertainty Estimation (GLUE) and Dynamically Dimensioned Search-Approximation of Uncertainty (DDS-AU) mostly use an aggregate approach to convert multiple objectives into a single objective function by pre-setting weight to each objective, while the proposed algorithm is able to consider all important objectives simultaneously without aggregation and identify multiple tradeoffs, each with distinct parameter space characteristics. Different subsets of multiple metrics including correlation coefficient, variability index, bias in

long-term average of response time series, flow-duration curve partitioning metrics, Nash-Sutcliffe Efficiency metric, and water balance (Asadzadeh et al., 2016; Gupta et al., 2009; Yilmaz et al., 2008) are used for multi-objective uncertainty estimation using CB-PA-DDS in three hydrologic model calibration problems. All these three uncertainty methods need to define a rejection threshold to separate acceptable (or *behavioral*) from unacceptable (or *non-behavioral*) solutions. To this end, multiple cutoff threshold values are considered and the performance of these methods are evaluated in the calibration and evaluation periods by comparing the response time series' length of coverage, size of the behavioral set, and thickness of the uncertainty bound. Similar to DDS-AU, the proposed CB-PA-DDS approach is found to be one order of magnitude more efficient than the GLUE method. The uncertainty bound produced by CB-PA-DDS is however more reliable than DDS-AU as the latter is prone to overfitting to the response time series in the calibration period, which can result in unacceptable simulations in the evaluation period. Details are reported in Chapter 4 and the outcomes are recently submitted to the Journal of Water Resources Research.

4) A novel weighted average multi-modelling based on the flow duration curve partitions is designed to linearly combine the simulation results of four structurally different hydrologic models used at Hydrologic Forecast Centre of Manitoba Infrastructure for operational flood forecasting downstream of Upper Assiniboine River Basin. The multi-modelling approach assigns a weight to each model component in each segment of the flow duration curve signifying its contribution to high-flow, mid-flow, and low-flow partitions. Grouping the models significantly improves simulation accuracy in both calibration and evaluation periods and reduces structural uncertainty in hydrologic

modelling. The proposed multi-modelling method can be used as a diagnostic tool to identify less effective or redundant models in a multi-modelling problem based on their weight factors for a better model selection strategy. Details about this research contribution are published in the Journal of Hydrology and also reported in Chapter 5.

Sahraei, S., Asadzadeh, M., & Unduche, F. (2020). Signature-Based Multi-Modelling and Multi-Objective Calibration of Hydrologic Models: Application in Flood Forecasting for Canadian Prairies. *Journal of Hydrology*, 125095.

2 TOWARD EFFECTIVE MANY-OBJECTIVE OPTIMIZATION: ROUNDED-ARCHIVING

Shahram Sahraei[1], Masoud Asadzadeh[2], Mahyar Shafii[3]

[1] PhD Candidate, Department of Civil Engineering, University of Manitoba, Canada

E-mail: sahraeis@myumanitoba.ca

[2] Assistant Professor, Department of Civil Engineering, University of Manitoba, EITC E1-332, 15 Gillson Street, Winnipeg, MB, Canada, R3T 5V6, Ph: +1 (204) 474 9535

E-mail: masoud.asadzadeh@umanitoba.ca

[3] Research Scientist, Dept. of Earth and Environmental Science, University of Waterloo

E-mail: mshafiih@uwaterloo.ca

2.1 Abstract

Heuristic multi-objective optimization (MO) algorithms lose their efficiency and performance as the number of objectives increases, due to the so-called dominance resistance, unless they are equipped with a specialized solution archiving strategy, like epsilon (ε-) archiving. This study introduces an alternative approach to tackle dominance resistance for solving environmental and water resources engineering problems with more than three objectives. In the proposed approach, objectives are rounded to user-defined precision levels before checking the dominance. Rounded-archiving is developed and assessed for PA-DDS and Borg MOEA and verified for AMALGAM applied to hydrologic model calibration problems with more than three objectives. Results show that rounded-archiving significantly improves the performance of MO algorithms and is at least as effective as (if not better than) the ε-archiving for solving many-objective optimization problems

without the need to restructure the algorithm, which is the requirement for the implementation of the ε-archiving approach.

Keywords: Many-objective optimization, Rounded-Archiving, Epsilon-Archiving, Dominance Resistance, Model Calibration.

Software Availability

Name of Software: PA-DDS with rounded-archiving method

Developer and Contact Information: Shahram Sahraei & Masoud Asadzadeh; Department of Civil Engineering, University of Manitoba, EITC E1-332, 15 Gillson Street, Winnipeg, MB, Canada, R3T 5V6

Email address: sahraeis@myumanitoba.ca & masoud.asadzadeh@umanitoba.ca

Year First Available: 2019

Hardware Required: Desktop Computer

Availability: go to ZDT_round folder in http://home.cc.umanitoba.ca/~asadzadm/software.html to see rounded-archiving method in two-objective ZDT1 test problem

Software Required: MATLAB

Program Size: <1MB

2.2 Introduction

Many-objective optimization, i.e. multi-objective (MO) optimization with four or more conflicting objectives, has become an attractive tool in a broad range of environmental and water resources problems such as calibration of environmental models (e.g. Shafii and Tolson, 2015; Zhang et al., 2010), water distribution network design (Fu et al., 2013), reservoir operation (Geressu and Harou, 2019; Giuliani et al., 2014), stormwater management (Di Matteo et al., 2019), robust decision-

making (Eker and Kwakkel, 2018; Kasprzyk et al., 2013; Singh et al., 2015), and crop/food production (Lautenbach et al., 2013). For instance, the hydrologic model calibration is well-known to have a MO nature for which no single solution simultaneously optimizes all the objective functions (model performance metrics), see for example Boyle et al. (2000), Gupta et al. (2005), Maier et al. (2019), and Wagener and Gupta (2005). Instead, a set of alternatives that is often called tradeoff or Pareto, after Vilfredo Pareto (1848-1923), exists since improving one objective degrades some other objectives. Traditionally, hydrologic models had been calibrated with two or three objective functions (Asadzadeh et al., 2015; Asadzadeh and Tolson, 2013; Gupta et al., 1998; Shafii and De Smedt, 2009), but recent developments in the hydrologic models, computational resources and MO algorithms is changing the trend toward the many-objective model calibration (Ercan and Goodall, 2016; Kollat et al., 2012; Reed et al., 2013; Wang and Brubaker, 2015).

The number of objectives can easily grow beyond three when the calibration problem is formulated to represent internal hydrological processes besides optimizing traditional model performance metrics such as the sum of squared errors or its variants. For instance, Shafii and Tolson (2015) formulated and solved a 15-objective model calibration problem and showed that incorporating the objective functions pertaining to the dominant hydrologic processes of a watershed aids in selecting the most reliable and hydrologically consistent model parameter values with respect to the selected dominant hydrologic processes.

On the other hand, the size of the non-dominated space grows exponentially as the number of objective functions increases (Farina and Amato, 2002) and leads to a two-fold difficulty in solving many-objective optimization problems by a heuristic MO algorithm, unless it is equipped with a specialized archiving approach. First, identifying any new dominating (better) solution is very challenging in many-objective optimization due to the high similarity in the objective space. This

is often referred to as the dominance resistance in the literature (Ikeda et al., 2001; Purshouse and Fleming, 2007). Second, heuristic MO algorithms that have a bounded archive size, e.g. non-dominated sorting genetic algorithm-II (NSGA-II) (Deb et al., 2002b), have to discard some good quality solutions from the current archive and therefore are subject to the oscillating issue that makes them inefficient for solving many-objective optimization problems. On the other hand, MO algorithms that have an unbounded archive size such as the Pareto Archived Dynamically Dimensioned Search (PA-DDS in Asadzadeh and Tolson, 2013) might archive a large number of Pareto solutions. Because solutions for further generations are generated based on the archived Pareto set, such algorithms assign a very low selection probability to high-quality solutions that are scattered among a huge number of Pareto solutions (Maier et al., 2014; Reed et al., 2013; Teytaud, 2007). MO algorithms that have a bounded archive keep at most an *a priori* specified number of non-dominated solutions by the end of optimization. The bounded archive set was first suggested in non-dominated sorting genetic algorithm (NSGA) (Srinivas and Deb, 1994) in order to reduce the computational time for checking the dominance rank, accelerating the decision-making process, and concentrating solely on appealing parts of the Pareto front. However, tightly bounded archive size of MO algorithms results in an oscillating behavior, in that, some currently non-dominated solutions have to be discarded from the archive in the current generation and be replaced by some inferior ones in subsequent generations (Hanne, 1999).

To increase the efficiency of MO algorithms in many-objective applications, several approaches have been developed in the literature, including fuzzy-optimality (Farina and Amato, 2004), order of efficiency (di Pierro et al., 2007), preferability (Drechsler et al., 2001), and epsilon-preferability (Sülflow et al., 2007). These methods are conceptually similar. Preferability is defined as choosing non-dominated solutions that are dominating a larger number of objectives when comparing two

non-dominated solutions (Sülflow et al., 2007). Fuzzy-optimality is a fuzzy-based concept of optimality that considers the number of improved objectives and the order of improvement between two solutions. Order of efficiency is the minimum size of non-dominated objective sub-space for a solution compared with the archived ones. However, these methods must be used with caution as the search may be misled into undesired regions of the decision space, especially when the objective functions are not equally important to decision-makers (Coello et al., 2007).

Laumanns et al. (2002) suggested the cell/epsilon (ε-) archiving approach to reduce the size of non-dominated archived solutions and therefore improve the diversity in the objective space. Deb et al. (2003) successfully applied the ε-archiving technique to a multi-objective evolutionary algorithm (MOEA) called ε-MOEA that outperformed five other MOEAs for solving mathematical MO test problems with two, three, and four objective functions, in terms of diversity and convergence. Reed and Devireddy (2004) introduced a variant of NSGA-II called ε-NSGAII that utilizes the ε-archiving, applied it to solve a bi-objective groundwater monitoring problem, and concluded that ε-archiving significantly increased the efficiency of the optimization algorithm. Kollat and Reed (2006) applied ε-NSGAII to a long-term groundwater monitoring design problem with four objective functions and showed a superior performance of ε-NSGAII against NSGA-II as well as other modern MO algorithms. Hadka and Reed (2013) combined a group of MO operators including the ε-archiving and introduced the Borg MOEA that showed a superior performance in comparison with six well-known MO algorithms for solving benchmark suites of test problems with 2 to 8 objectives. Many modern MO algorithms do not have the ε-archiving strategy, e.g. PA-DDS, A Multi-ALgorithm Genetically Adaptive Multi-objective (AMALGAM) (Vrugt and Robinson, 2007), strength pareto evolutionary algorithm 2 (Zitzler et al., 2001), NSGA-

III (Deb and Jain, 2014) and therefore are anticipated to lose their performance in terms of convergence and solution diversity when applied to many-objective problems.

In this study, an alternative approach to ε-archiving is introduced. This approach that is called rounded-archiving is technically the point-based dominance archiving of the rounded value of objective functions. The main advantage of the proposed approach over ε-archiving is that it does not require the user to restructure the MO algorithm computer code. The rounded-archiving approach is compared to the ε-archiving approach for solving hydrologic model calibration problems that are formulated and solved as many-objective optimization problems, in the literature. The proposed approach is not algorithm- or case-specific and is readily applicable to other MO algorithms such as PA-DDS and AMALGAM as shown in this study. The remainder of this paper is organized as follows. The Methods section describes the proposed rounded-archiving method applied to MO algorithms including Borg MOEA, PA-DDS, and AMALGAM, case studies including the mathematical and hydrologic model calibration problems, numerical experiment setup, and the results analysis approach. Next, results are presented and a general discussion is given, followed by the concluding remarks.

2.3 Methods

MO algorithms typically start the optimization with a random solution generation and archive solutions based on the dominance check in the objective space. Then, one or a population of archived solutions are selected for perturbation or recombination to create new solution(s) for subsequent evaluations. This process continues until a termination condition is met.

Point-based archiving and selection strategies therefore constitute the principal parts of a MO algorithm to guide the search toward a better proximity (convergence) and diversity of solutions

to approximate the Pareto front (Coello et al., 2007; Deb, 2001). Pareto approximate front is referred to as the archived non-dominated solutions found by a MO algorithm at the end of optimization. In point-based archiving, solution *A* is said to be non-dominated if and only if there is no other solution *B* that dominates it (Van Veldhuizen and Lamont, 2000). Equation (1) shows the mathematical representation of point-based archiving concept of solution *A* dominating *B* for a minimization problem that has *n* conflicting objective functions. In the case of maximization, the corresponding objective function needs to be multiplied by *-1* to change it to a minimization problem.

$$\boldsymbol{A} \preccurlyeq \boldsymbol{B} \Leftrightarrow \begin{cases} f_i(\boldsymbol{A}) \leq f_i(\boldsymbol{B}) & \forall\, i \in \{1, 2, \dots, n\} \\ f_j(\boldsymbol{A}) < f_j(\boldsymbol{B}) & \exists\, j \in \{1, 2, \dots, n\} \end{cases} \quad \textbf{(1)}$$

In general, the regular, point-based dominance check and archiving strategy become ineffective as the number of objective functions grows to more than three. Farina and Amato (2002) mathematically showed by Equation (2) that the proportion of non-dominated space (*e*) exponentially increases as the number of objectives (*n*) increases. This issue increases the similarity of the archived solutions in the objective space and causes the dominance resistance (Ikeda et al., 2001; Purshouse and Fleming, 2007) that degrades the performance of MO optimization algorithms.

$$e = \frac{2^n - 2}{2^n} \quad \textbf{(2)}$$

The cell-based dominance check and the corresponding ε-archiving introduced by Laumanns et al. (2002), which is based on the ε-dominance, addresses the dominance resistance issue and improves the performance of MO algorithms for solving many-objective optimization problems.

By definition, in a minimization problem with n objectives, solution $\boldsymbol{A}$ ε-dominates solution $\boldsymbol{B}$ if and only if $\left\lfloor \frac{f_i(\boldsymbol{A})}{\varepsilon_i} \right\rfloor \leq \left\lfloor \frac{f_i(\boldsymbol{B})}{\varepsilon_i} \right\rfloor$ for all objectives ($i=1,\ldots,n$) with $\left\lfloor \frac{f_j(\boldsymbol{A})}{\varepsilon_j} \right\rfloor < \left\lfloor \frac{f_j(\boldsymbol{B})}{\varepsilon_j} \right\rfloor$ for at least one objective j. The symbol $\lfloor\ \rfloor$ returns the floor of a real value. ε-archiving uses this inequality to develop a mesh grid in the objective space and archive at most one solution in each grid cell. If two solutions $\boldsymbol{A}$ and $\boldsymbol{B}$ reside in one cell, i.e. $\left\lfloor \frac{f_i(\boldsymbol{A})}{\varepsilon_i} \right\rfloor = \left\lfloor \frac{f_i(\boldsymbol{B})}{\varepsilon_i} \right\rfloor$ for all objectives i, the solution that is closer to the dominating (bottom-left in minimization problems) corner of the cell is archived and the other solution is discarded. Green points in Figure 2-1 are ε-dominated by the red points demonstrating both conditions of ε-archiving relation. The size of each grid-cell is equal to $\boldsymbol{\varepsilon}$, which should be determined *a priori* by the decision-maker or systems analyst (Reed and Devireddy, 2004) for each objective function based on the desired precision level. Equation (3) shows the mesh grid formulation for the ε-archiving method in the objective space.

$$\left\{\ldots, \boldsymbol{\varepsilon}\left(\left\lfloor \frac{\boldsymbol{f}}{\boldsymbol{\varepsilon}} \right\rfloor - 1\right), \boldsymbol{\varepsilon}\left(\left\lfloor \frac{\boldsymbol{f}}{\boldsymbol{\varepsilon}} \right\rfloor\right), \boldsymbol{\varepsilon}\left(\left\lfloor \frac{\boldsymbol{f}}{\boldsymbol{\varepsilon}} \right\rfloor + 1\right), \ldots\right\},\ \boldsymbol{f} = \{f_1, \ldots, f_n\},\ \ \boldsymbol{\varepsilon} = \{\varepsilon_1, \ldots, \varepsilon_n\},\qquad n\text{: No. of objectives} \tag{3}$$

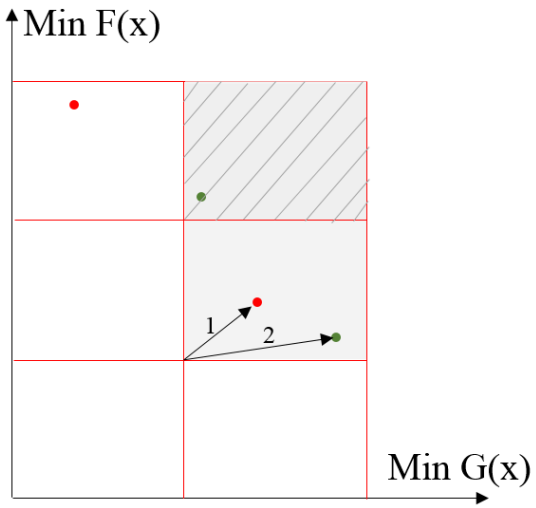


Figure 2-1: Schematic of ε-archiving concept. Green points are ε-dominated by red points

As discussed in Reed et al. (2007), ε is not a parameter of the MO algorithm. It is rather a case-dependent value determined either by the decision-maker or by the systems analyst prior to the optimization based on the required precision in each objective function to ensure that only meaningful differences between non-dominated solutions are considered in solution archiving. ε-MOEA, ε-NSGAII, and Borg MOEA are among algorithms that are equipped with ε-archiving.

2.3.1 Rounded-Archiving Method

In this study, an alternative solution archiving approach to ε-archiving is proposed and called the rounded-archiving in which the objective values are rounded to their desired precision levels before performing the point-based dominance check. The idea of rounded-archiving evolved after observing a significant decrease in the number of non-dominated solutions for a hypothetical example where 10,000 points are sampled from a uniformly random distribution in the $[0,1]^n$ n-dimensional space representing the objective space of an MO problem with n independent objective functions, Figure 2-2.

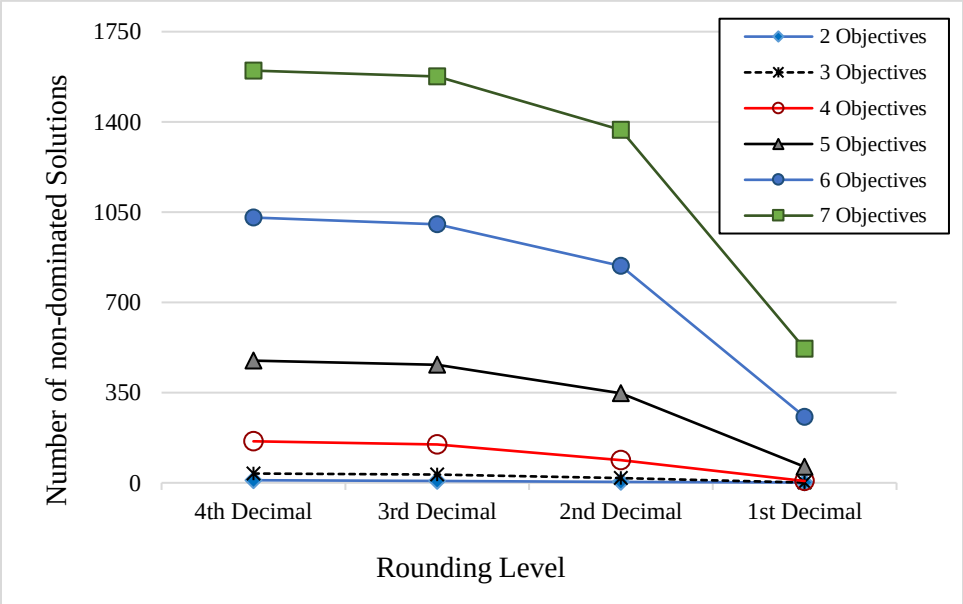


Figure 2-2: The impact of the rounding level and number of objectives on the number of non-dominated solutions for a hypothetical example of the objective space

By definition, in a minimization problem with *n* objectives, solution ***A*** dominates solution ***B*** in terms of rounded-archiving technique if and only if $\varepsilon_i \times round\left(\frac{f_i(\boldsymbol{A})}{\varepsilon_i}\right) \leq \varepsilon_i \times round\left(\frac{f_i(\boldsymbol{B})}{\varepsilon_i}\right)$ for all objectives (*i=1,…, n*), and there exists at least one objective *j* that the inequality $\varepsilon_i \times round\left(\frac{f_i(\boldsymbol{A})}{\varepsilon_i}\right) < \varepsilon_i \times round\left(\frac{f_i(\boldsymbol{B})}{\varepsilon_i}\right)$ holds. The symbol $round(\,)$ rounds a real value to the nearest integer. If $\varepsilon_i \times round\left(\frac{f_i(\boldsymbol{A})}{\varepsilon_i}\right) = \varepsilon_i \times round\left(\frac{f_i(\boldsymbol{B})}{\varepsilon_i}\right)$ for all objectives *i*, both solutions are equally important and only one of them is retained in the archive.

Technically, the rounded-archiving method archives the solution that has an objective vector value corresponding to the black point $\left(\text{i.e. } \boldsymbol{\varepsilon} \times round\left(\frac{\boldsymbol{f}}{\boldsymbol{\varepsilon}}\right)\right)$ in Figure 2-3. This means that the rounded-archiving maintains the dominance relation only if it shows a difference that is meaningful at the rounded level of objectives. To be more specific, if solution ***A*** dominates solution ***B*** at the full precision of objective values, rounded archiving will prefer ***A*** over ***B*** if ***A*** still dominates ***B*** after its objective values are rounded. However, if ***A*** and ***B*** have the same rounded objective values, one of them will be archived by the rounded archiving. Similar to the ε-archiving, the size of the grids for the rounded-archiving method is equal to $\boldsymbol{\varepsilon}$ but its location is not the same as ε-archiving. The mesh grid formulation for the rounded-archiving is shown in Equation (4). The symbols ⌈ ⌉ and ⌊ ⌋ in Equation (4) return the ceiling and floor of a real value, respectively.

$$\left\{\ldots, \boldsymbol{\varepsilon}\left(\frac{\left\lfloor\frac{\boldsymbol{f}}{\boldsymbol{\varepsilon}}\right\rfloor + \left\lceil\frac{\boldsymbol{f}}{\boldsymbol{\varepsilon}}\right\rceil}{2} - 1\right), \boldsymbol{\varepsilon}\left(\frac{\left\lfloor\frac{\boldsymbol{f}}{\boldsymbol{\varepsilon}}\right\rfloor + \left\lceil\frac{\boldsymbol{f}}{\boldsymbol{\varepsilon}}\right\rceil}{2}\right), \boldsymbol{\varepsilon}\left(\frac{\left\lfloor\frac{\boldsymbol{f}}{\boldsymbol{\varepsilon}}\right\rfloor + \left\lceil\frac{\boldsymbol{f}}{\boldsymbol{\varepsilon}}\right\rceil}{2} + 1\right), \ldots\right\} \quad \textbf{(4)}$$

$$\boldsymbol{f} = \{f_1, \ldots, f_n\}, \quad \boldsymbol{\varepsilon} = \{\varepsilon_1, \ldots, \varepsilon_n\}, \qquad n\text{: } No.of\ objectives$$

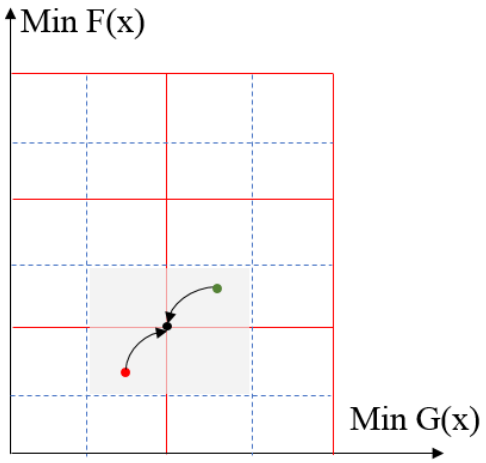


Figure 2-3: Schematic of rounded-archiving method. Both green and red points change to the black point in rounded-archiving method. Red line and blue dash line respectively represent the grids for the ε-archiving and rounded-archiving methods.

As noted in Reed and Devireddy (2004), the resolution (precision level in this paper) is case-dependent and should reflect decision-makers' opinion about the meaningful difference between different options (solutions). Rounding should be performed in the scaled objective space by the precision/resolution level, $\boldsymbol{\varepsilon}$ in $\boldsymbol{\varepsilon} \times round\left(\frac{f}{\boldsymbol{\varepsilon}}\right)$. For example, for a cost function that ranges up to billions of dollars, solutions need to be rounded to the nearest 1 million dollar if it is the deal-breaker for the decision makers. In the case of hydrologic model calibration, the systems analyst might want to round the percent bias error metric to the nearest 1% if this resolution defines the meaningful difference between two solutions.

Script 1 shows a function coded in MATLAB that rounds a set of objective function values to the precision level that is given/set by the user. This shows how the modeller needs to implement the rounded-archiving strategy. The rounded-archiving method is formulated for the optimization problem formulation and not the MO algorithm; therefore, the optimization algorithm does not need to be altered to take advantage of the rounded archiving approach for solving many-objective optimization problems.

Script 1: The pseudo-code function for rounding the exact value of objective function to the user-specified precision levels

```
Y = RoundObjectives (Y_Exact, rounding_lvl)
   % User defines rounding_lvl that is a vector variable with the desired rounding level, e.g. [0.1, 0.01, 10, 1000] for a four-objective optimization problem.
   % This function takes as input the vector variable Y_Exact that has the exact value of the objective functions and returns in Y their rounded value according to the rounding_lvl.
   % Length of input vectors should be the same
  if length(Y_Exact) ~= length(rounding_lvl)
    error('Error: input vector variables Y_Exact & rounding_lvl should have the same size');
  else
   % Round each objective function to desired precision levels.
    for i = 1 : length(Y_Exact)
      Y(i) = rounding_lvl(i) * round (Y_Exact(i) / rounding_lvl(i));
    end
  end
```

2.3.2 Optimization Algorithms

In order to evaluate and compare the effect of the point-based, rounded-, and ε-archiving methods on proximity and diversity, three MO algorithms that are frequently used for solving environmental and water resources MO problem are considered: Borg MOEA, PA-DDS, and AMALGAM. These algorithms have completely different structures in terms of search, selection, and solution generation, and therefore represent a wide range of different MO algorithms that have similar structures.

2.3.2.1 BORG MOEA

Borg MOEA was developed by Hadka and Reed (2013) for solving many-objective optimization problems. It is a robust variant of the ε-MOEA algorithm (Deb et al., 2003) and couples multiple components of different MO algorithms including the ε-archiving (Laumanns et al., 2002), adaptive population sizing (Tang et al., 2006), and adaptive tournament sizing (Hadka and Reed, 2013) for preserving the selection probability at a constant rate, and multiple recombination

operators. This algorithm automatically adjusts the population size and selection probability when no meaningful improvement in the objective space is measured after a specified number of solution evaluations. The ε-archiving feature of Borg MOEA improves its convergence and preserve the diversity of solutions in the objective space during optimization by diminishing the effect of dominance resistance and reducing the number of archived solutions. Borg MOEA has been successfully applied to many water resources problems such as a constrained six-objective urban water portfolio planning (Hadka and Reed, 2015), a four-objective lake pollution control (Quinn et al., 2017), and a constrained three-objective engineered injection and extraction for enhanced groundwater remediation (Piscopo et al., 2015).

2.3.2.2 PA-DDS

PA-DDS (Asadzadeh and Tolson, 2013) is an efficient MO algorithm for solving environmental and water resources MO problems. It is not population-based and generates one solution in each iteration by selecting one currently non-dominated solution and perturbing it to search for better solutions. Asadzadeh and Tolson (2013) recommended hyper-volume contribution as the most effective selection metric for PA-DDS solving general MO problems. Convex hull contribution (CHC) is another selection metric developed by Asadzadeh et al. (2014b) for PA-DDS applied to MO problems with a convex Pareto front. However, CHC is not used in this study since it has not been tested on problems with more than three objectives. PA-DDS has only one parameter, which is known as the solution perturbation size with a robust suggested value of 0.2 (Asadzadeh and Tolson, 2013). PA-DDS has an unbounded archive set and uses the point-based archiving to archive all non-dominated solutions throughout the search; therefore, it does not suffer from the oscillating behavior, but it is expected to suffer from dominance resistance when applied to many-objecive optimization problems. In this research, ε-archiving is implemented for PA-DDS to be

able to test its performance against the point-based and rounded-archiving approaches. Outstanding applications of PA-DDS in water resources include reservoir operation (Asadzadeh et al., 2014a; Razavi et al., 2014), model calibration (Asadzadeh et al., 2016, 2015), and water distribution network design (Asadzadeh et al., 2012).

2.3.2.3 *AMALGAM*

AMALGAM introduced by (Vrugt and Robinson, 2007) is a multi-algorithm evolutionary optimization strategy that combines four well-known and widely used heuristic optimization algorithms including NSGA-II (Deb et al., 2002b), particle swarm optimization (Kennedy, 2006), adaptive metropolis search (Haario et al., 2001), and differential evolution (Storn and Price, 1997). The population size and the number of generations of AMALGAM are set based on the computational budget of this study and all its other parameters are set to their default values (Vrugt and Robinson, 2007). AMALGAM generates an initial population of solutions by latin hypercube sampling and gives each of its sub-algorithms a pre-determined portion of the whole population for generating new solutions. Then, the parent and offspring solutions are combined and sorted using the fast non-dominated sorting approach introduced in (Deb et al., 2002a, 2002b). If necessary, the crowding distance metric is utilized for archiving solutions that have an equal dominance rank and the ones that have higher distance are retained in the archive. The sub-algorithm that contributed more to the set of archived solutions is given a higher portion of population size for generating new solutions in subsequent generation. AMALGAM has a point-based archiving and a bounded archive; therefore it is not reinforced against the dominance resistance and oscillating effect. The effectiveness of the proposed rounded-archiving is validated on AMALGAM. Outstanding recent applications of AMALGAM include the many-objective

signature-based hydrologic model calibration in Shafii et al. (2017) and the multi-site hydrologic model calibration in Zhang et al. (2010).

2.3.3 Case Studies

2.3.3.1 Mathematical Test Problems

Benchmark mathematical test problems are designed to challenge the optimization algorithms. Often times, they are quick-to-evaluate and the closed form of their true Pareto front is known. Therefore, they can be used to efficiently and effectively evaluate the performance of the optimization algorithms.

DTLZ suite of test problems are scalable in terms of the number of decision variables and objective functions (Deb et al., 2005). In this paper, the bi-objective and five-objective versions of DTLZ1 and R2-DTLZ2 are solved to compare the effect of the archiving methods on the archive size at different number of objectives.

2.3.3.1.1 DTLZ1

DTLZ1 has *m* variables with the range [0, 1] and *n* objectives that are designed to trap MO algorithms in local fronts. The ten-variable bi-objective and five-objective DTLZ1 solved in this research are proven to have respectively (11^{10-2+1}-1) and (11^{10-5+1}-1) local fronts with the true Pareto front where $\sum_{i=1}^{n} f_i(\boldsymbol{X}) = 0.5$.

2.3.3.1.2 DTLZ2

The five-objective R2-DTLZ2 test problem is a more complicated version of DTLZ2 with a concave true Pareto front with 30 decision variables. According to Zhang et al. (2009), the decision variable space of R2-DTLZ2 is mapped by a linear transformation orthogonal matrix and the

objective space is extended by a stretching function to make it more challenging to solve. The true Pareto front of R2-DTLZ2 is where $\sum_{i=1}^{n}(f_i - 1)^2 = 1$, and the range of its variables is [0, 1].

2.3.3.2 Watershed model calibration

2.3.3.2.1 RAVEN Model of Grand River Watershed

RAVEN is a semi-distributed watershed model introduced by Craig (2015) capable of simulating short-term rainfall-runoff events and long-term synthesis of hydrologic processes in a basin for resource management and water quality assessment. It divides a watershed into several sub-watersheds that are further partitioned into multiple hydrologic response units (HRUs). Each HRU is lumped areas with a unique combination of topography, geometry/geography, land use/type, and aquifer/soil. Meteorological conditions such as rainfall, temperature, and wind velocity are then assigned to HRUs. The vertical water balance and energy balance are used for simulating and then assembling relevant hydrological processes in each HRU. The flow is then routed downstream and laterally by reconnecting HRUs. The interesting fact is that RAVEN's level of complexity can change from a lumped model to a distributed model with a myriad of HRUs depending on data availability or analyst's desire.

The RAVEN model of an upstream sub-watershed (~274 km^2) of the greater Grand River Watershed in south-western Ontario, Canada, is re-calibrated in this study. Following Shafii et al. (2017), the hydrometric data for time span of 2009-2014 is considered to calibrate 20 tunable parameters of the model, using three quarter of the year 2009 as the warm-up period and the rest of the data for calibration.

The preliminary analysis of this case study showed that some of the 15 objectives calibrated by Shafii and Tolson (2015) (see Table 2-1) are highly correlated (none conflicting) and therefore can

safely be removed from the set of calibration objectives. The proportion of non-dominated space unnecessarily increases by considering highly correlated objectives in optimization and MO algorithms encounter difficulty in directing the search toward the desired non-dominated front. In addition, according to the definition, objective functions of a MO problem should display relatively conflicting behaviors. Giuliani et al. (2014a), for example, incorporated the principal component analysis in a many-objective reservoir operation problem to reduce the objective functions to a few uncorrelated principal objectives (vectors) that describe high percentile variance of the original MO problem to reach a more consistent and diverse Pareto approximate front.

In this study, the calibration objectives in Shafii and Tolson (2015) are compared in a pairwise manner and a strong linear correlation higher than 0.8 is observed between eight of these to the other seven objectives, see Table 2-1. For instance, the overall runoff ratio (Obj 3) is perfectly correlated with the mean of streamflow data (Obj 8) as they both represent the capability of the model to simulate the water balance. The seven objective functions that are used for the calibration of RAVEN in this paper are highlighted in Table 2-1 and briefly introduced next.

It is well-known that an optimal parameter set for a single calibration metric such as *NSE* (Nash and Sutcliffe, 1970) does not necessarily guarantee the model to emulate the detailed processes of the real system (Gupta et al., 1998). This also holds true in the case of multi- and many-objective calibration to a lesser extent (Maier et al., 2019), which is referred to as overfitting due to incorporating certain characteristics of error distribution in calibration, imperfect model structure, simplifying assumptions in simulation, uncertainty in measurements as well as spatiotemporal variations in model parameters (Efstratiadis and Koutsoyiannis, 2010). *NSE* as shown in Equation (5) is more sensitive to larger errors that often happen in high-flow periods and to the timing and shape of the measured stream flux data due to the presence of three components including mean,

variance, and the correlation coefficient (Gupta et al, 2009). *MSE* denote the mean of squared errors between the observations and simulations and $\sigma_0{}^2$ represents the variance of the observations.

Table 2-1: Pairwise linear correlation coefficient between 15 calibration objectives in Shafii and Tolson (2015). Cells shaded in yellow show the highly correlated pairs. Calibration objectives shaded in green are considered for this study.

OBJECTIVES	Obj 1	Obj 2	Obj 3	Obj 4	Obj 5	Obj 6	Obj 7	Obj 8	Obj 9	Obj 10	Obj 11	Obj 12	Obj 13	Obj 14	Obj 15
Obj 1-NSE(Q)	-	-0.25	-0.26	-0.24	0.23	-0.23	0.25	-0.26	-0.12	-0.05	0.00	0.80	0.14	0.31	0.07
Obj 2-NSE(logQ)		-	0.46	0.40	-0.96	0.55	-0.47	0.46	0.71	-0.27	0.52	-0.16	-0.89	-0.97	0.34
Obj 3-Overall Runoff Ratio(Q)			-	0.98	-0.38	0.94	-0.08	1.00	0.64	0.18	0.48	0.04	-0.18	-0.62	0.59
Obj 4-Overall Runoff Ratio(logQ)				-	-0.31	0.93	-0.05	0.98	0.62	0.20	0.49	0.08	-0.13	-0.56	0.63
Obj5-FDC Mid-segment Slope					-	-0.48	0.42	-0.38	-0.68	0.24	-0.48	0.19	0.89	0.90	-0.27
Obj 6-FDC High Segment Vol.						-	-0.12	0.94	0.81	0.11	0.66	0.05	-0.39	-0.69	0.71
Obj 7-FDC Low Segment Vol.							-	-0.08	-0.34	0.12	-0.31	0.35	0.42	0.45	-0.02
Obj 8-Mean(Q)								-	0.64	0.18	0.48	0.04	-0.18	-0.62	0.59
Obj 9-Variance(Q)									-	-0.05	0.87	-0.05	-0.72	-0.77	0.70
Obj 10-Median(Q)										-	-0.02	0.01	0.37	0.17	0.00
Obj 11-Peak(Q)											-	-0.07	-0.59	-0.59	0.68
Obj 12-One-day-lag Autocorrelation												-	0.17	0.20	0.16
Obj 13-Mean(LogQ)													-	0.84	-0.34
Obj 14-Variance(logQ)														-	-0.44
Obj 15-Max_Monthly(Q)															-

$$NSE = 1 - \frac{MSE}{\sigma_0^2} \quad \textbf{(5)}$$

Hydrologic signatures represented by Equations (6) to (11) are reformulated in the form of absolute bias in simulated versus measured values to be minimized with an ideal value of zero. Q and $\bar{Q}$ in Equations (6) to (11) signify the streamflow and the mean of streamflow, P is the probability of exceedence, and k is the number of time steps. The metrics in Equations (6) to (8),

evaluate the model performance in emulating the flow-duration curve (FDC). FDC is a sorted logarithmic flow rate curve plotted versus cumulative frequency of exceedence. Following Shafii and Tolson (2015), FDC is divided into three segments of mid-segment slope (25% to 75%), high flow volume (0% to 2%), and low flow volume (75% to 100%), as shown in Figure 2-4. The respective partitioning of FDC represents soil storage capacity, quick runoff (due to snow-melt and/or rainfall), and base flow components of the streamflow (Yilmaz et al., 2008). Median in Equation (9) and peak in Equation (10) evaluate the performance in simulating mid-flow and max-flow (Bennett et al., 2013). The one-day lag auto-correlation coefficient (AC_{lag1}) in Equation (11) captures the repeated periodic patterns in streamflow time series (Bennett et al., 2013).

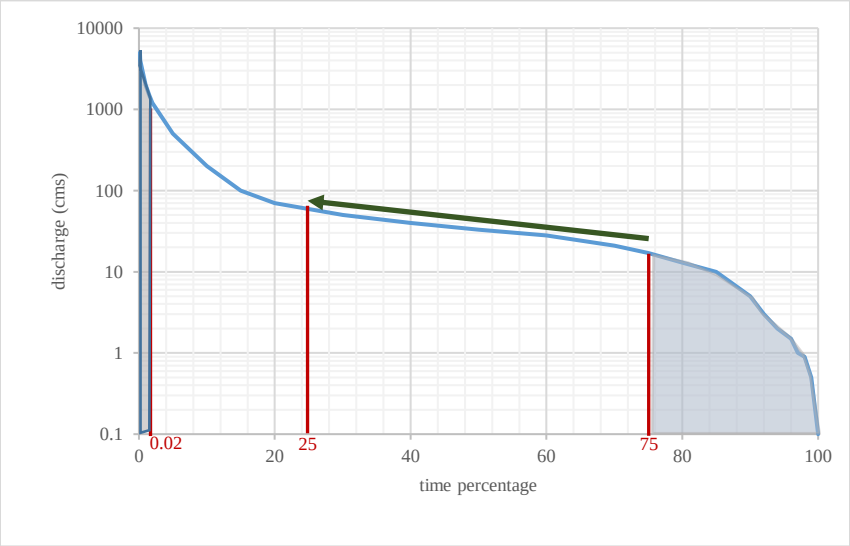


Figure 2-4: Flow duration curve segmentation

$$Mid\_Slope = log_{10}(Q_{25}) - log_{10}(Q_{75}) \quad \ni \quad Q_{25} = \{Q|P(Q) = 0.25\} \ \& \ Q_{75} = \{Q|P(Q) = 0.75\} \tag{6}$$

$$High_{Flow} = \sum(Q_h) \quad \ni \quad Q_h \in \{Q|P(Q) < 0.02\} \tag{7}$$

$$Low_{Flow} = -\sum(log_{10}(Q_l)) \quad \ni \quad Q_l \in \{Q|P(Q) > 0.75\} \tag{8}$$

$$Median = \{Q|P(Q) = 0.5\} \quad \textbf{(9)}$$

$$Peak = \{Q|Q = \max(Q_i), \ni i = 1, 2, \dots, k\} \quad \textbf{(10)}$$

$$AC_{lag1} = \frac{\sum_{j=1}^{k-1}(Q_j - \bar{Q})(Q_{j+1} - \bar{Q})}{\sum_{j=1}^{k}(Q_j - \bar{Q})^2} \quad \textbf{(11)}$$

2.3.3.2.2 SWAT Model of the Rouge River Watershed

SWAT stands for soil and water assessment tool introduced by Arnold et al. (1990) for long-term basin-scale simulations of the hydrologic cycle. SWAT combines features of different models to enable estimation of runoff, sediment transport rate, and water quality constituents such as suspended solids, nitrogen and phosphate components at the outlet of each sub-watershed, Neitsch et al. (2011). SWAT is a semi-distributed watershed model that divides the watershed into sub-watersheds that can have multiple HRUs. However, HRUs do not carry any geographical location inside their sub-watershed. For each sub-watershed, SWAT aggregates the simulation components that are contiguously pyramided en route from sub-watersheds to the streams and thus to the outlet of the watershed. In this paper, the SWAT model of the Rouge River watershed (331 km^2), Ontario, Canada is used, which was developed by Asadzadeh et al. (2015). The watershed encompasses four main land-use classes, agricultural, urban, natural, and water bodies. Agricultural land management in terms of crop rotation, cultivation, planting, tillage and fertilizer applications are simulated by the model.

Daily hydrometric and climatic data from 2006 to 2009 are used to automatically calibrate the 13 most sensitive parameters of the model, considering four objectives: maximizing NSE for daily streamflow, and minimizing absolute bias in the volume of high flow segment [less than 2%

probability of exceedance], the volume of low flow segment [more than 70% probability of exceedance] of FDC, and the slope of the mid-segment [probability of exceedance between 20% to 70%] of FDC. Asadzadeh et al. (2015) formulated the SWAT calibration as a constrained optimization problem to satisfy the estimated values of evapotranspiration (62% of the annual average precipitation) and surface flow contribution to streamflow (60%). In this study, these two constraints are handled by adding a fifth objective function that minimizes the total constraint violation. As discussed in Asadzadeh et al. (2016), Rouge river watershed has a sub-daily hydrologic response to precpitation events and the daily simulation mode of SWAT cannot precisely model the timing of peak flows, especially for late calendar day precipitation events. As a result, Asadzadeh et al. (2016) observed a one-day offset issue in simulating some of the peak flow rates of Rouge River watershed and added a feature to the calibration of SWAT to shift the simulated hydrograph for fixing the timing error. The SWAT model calibration in this paper automatically handles the timing error.

2.3.4 Numerical Experiment Setup

The three approaches of point-based, rounded-, and ε-archiving are implemented and tested in PA-DDS and Borg MOEA for solving problems explained in the previous section. Table 2-2 provides details on these problems, as well as the ε value or rounding level for each objective. The effectiveness of the rounded-archiving as an alternative of the ε-archiving approach is subsequently validated for the AMALGAM algorithm. The algorithms are compared based on their performance in a multi-trial optimization to capture the variation in their performance due to their stochastic nature. We considered 10 and 50 trials for hydrologic model calibration and mathematical test problems, respectively. Moreover, the performance of the MO algorithms is assessed at a relatively low computational budget of 1000 solution evaluations and a relatively

larger budget of 10000 solution evaluations per trial. The low budget case is considered to assess the effect of dominance relations in low budget situations and how they help the performance of the algorithm concerning proximity and diversity. The high budget case is suggested to minimize the effect of initialization of the algorithms and find good quality solutions in each trial of calibration (Asadzadeh et al., 2015; Shafii et al., 2017).

Table 2-2: Number of objectives and parameters and ε or rounding level for MO case studies

Test Case	Number of Parameters	Number of Objectives	Precision Level						
			obj1	obj2	obj3	obj4	obj5	obj6	obj7
SWAT	13	5	0.01	1	1	1	1	-	-
RAVEN	20	7	0.05	5	5	5	5	5	5
DTLZ1-2D & R2-DTLZ2-2D	10 & 30	2	0.01	0.01	-	-	-	-	-
DTLZ1-5D & R2-DTLZ2-5D	10 & 30	5	0.01	0.01	0.01	0.01	0.01	-	-

The parameterization of PA-DDS, Borg MOEA, and AMALGAM has been done according to the default settings recommended by their developers. The initial population for Borg MOEA and AMALGAM is set to 100 solutions and PA-DDS automatically commences the search with at least five solutions or 0.5% of the total evaluations, whichever is higher, as the size of initial solutions. PA-DDS uses the hyper-volume contribution selection metric. ε-archiving is an inseparable part of Borg MOEA structure. Therefore, in order to assess its performance with rounded-archiving and point-based archiving, an extremely small ε value (10^{-6} for all objectives) is used in this study. AMALGAM has a bounded archive size equal to its population size. As a result, archive truncation is highly likely to occur in many-objective optimization and may lead to oscillating effect in AMALGAM.

2.3.5 MO algorithm Performance Metrics

Performance metrics are used to quantify the convergence, diversity, and consistency of the Pareto approximate fronts obtained by stochastic optimization algorithms. A number of performance

indicators exist in the literature including normalized hyper-volume (Fonseca et al., 2006), additive epsilon (Zitzler et al., 2003), and generational distance (Van Veldhuizen and Lamont, 1998) that evaluate different characteristics of the Pareto approximate front against the true Pareto front or a reference Pareto front. These metrics have been widely used for evaluating MO algorithms since they can quantify the algorithms' performance in terms of proximity and solution diversity in the objective space (Ward et al., 2015; Yuan et al., 2016).

2.3.5.1 Normalized Hyper-Volume (NHV)

NHV normalizes the objective space to the unit hypercube using a Nadir point, an ever-dominated point, and a Utopia point, an absolutely non-dominated point for a minimization problem (Fonseca et al., 2006). Basically *NHV* for a front is the portion of this unit hypercube that is dominated by the front. Obviously, a higher value of *NHV* shows better proximity and diversity of the Pareto approximate front. The computational cost of the *NHV* calculation increases exponentially as the number of objective functions increases. Therefore, in the many-objective cases of this study, the approximation version of *NHV* (Bader and Zitzler, 2011) is used.

2.3.5.2 Additive Epsilon ($\boldsymbol{\varepsilon^+}$) Indicator

The ε^+ indicator (Zitzler et al., 2003) calculates the minimum distance required for shifting a Pareto approximate front to dominate a reference set of points as a whole. The ε^+ indicator has a minimum value of zero that is ideal when the reference set is the true Pareto front and is very sensitive to the gaps in the Pareto approximate front; therefore, it shows if outliers with poor proximity exist in the Pareto approximate front, the so called inconsistency (Hadka and Reed, 2012).

2.3.5.3 Generational Distance (GD)

GD is the average Euclidean distance between each point on the Pareto approximate front and its closest point on a reference set (the true or best known Pareto) with minimum value of zero that is ideal when the Pareto approximate front perfectly matches the reference set (Van Veldhuizen and Lamont, 1998). *GD* purely emphasizes on the proximity rather than the diversity preservation and gaps in the Pareto approximate front have minimal effect on it.

2.3.5.4 Performance Metrics Calculation Procedure

The following steps are taken in this study to calculate the performance metrics:

1) In order to have a fair comparison between the MO algorithms, the rounded objective value of non-dominated solutions are converted back to their full precision by re-running the case studies for the corresponding solutions.
2) The true Pareto front is used as the reference set for mathematical test problems. The best non-dominated points are collected from all calibration runs, disregarding the type of algorithm and archiving method, and used as reference set for these problems.
3) Instead of exact *NHV,* the Monte-Carlo approximation *NHV* (Bader and Zitzler, 2011) is calculated for the seven-objective RAVEN model calibration problem. 10000 random points are generated from the uniform distribution in the $[0, 1]^7$ space. The portion of these points that are dominated by each Pareto approximate front is calculated. To reduce the effect of random number generator on the results, this process is repeated ten times, independently, and the average of the portion of dominated points is calculated as the approximate *NHV* for each algorithm.

4) Borg MOEA automatically terminates the optimization when no improvement is measured in objective values after a certain number of iterations. Results are post-processed to make sure that premature convergance did not occur.

2.3.6 Optimization Algorithm Performance Comparison Approach

The MO optimization algorithms in this paper result in different performance metric values in each trial due to their stochastic search behavoiur. The MO algorithms are compared based on the empirical cumulative distribution function (CDFs) that shows the probability of equal to or better performance than the argument of the function. The CDF plots of the algorithms are compared by the first degree stochastic dominance (Levy, 1992). The concept of stochastic domination is referred to as better performance at all levels of probability. With regard to ε^+ and *GD* indicators, algorithm *A* stochastically dominates algorithm *B* if and only if the CDF of *A* is less than or equal to that of *B* at each level of probability (Carrano et al., 2011). Whereas, a better performance in terms of *NHV* corresponds to an equal or higher metric value of *A* compared to *B*. The statistical significance of the stochastic dominance is examined using the two-sided Wilcoxon rank sum test with the 95% confidence level. The null hypothesis of this test states no significant difference between algorithms *A* and *B*. A p-value smaller than the significant level (5% in this paper) in the two-sided Wilcoxon rank sum test is preferred to firmly reject the null hypothesis and confirm the difference between *A* and *B*. Similar performance comparison studies are used in Asadzadeh and Tolson (2013) and Hadka and Reed (2012).

2.4 Results and Discussion

2.4.1 Method Development: Borg MOEA and PA-DDS

2.4.1.1 Mathematical Test Problems

According to the observations on the DTLZ1-2D and R2-DTLZ2-2D experiments with low computational budget, the three archiving approaches make PA-DDS and Borg MOEA store a similar and relatively small number of non-dominated solutions. CDF plots of the performance metrics (not shown in this paper) show that, altering the archiving approach does not alter the MO algorithm performance. In the case of 10000 solution evaluations, point-based archiving makes MO algorithms archive a high number of non-dominated solutions relative to the rounded- and ε-archiving approaches (see Table 2-3), while CDF plots (not shown in this paper) do not reveal any significant preference in the archiving approach. Figure 2-5 shows the Pareto approximate fronts obtained by the median performing trial of these algorithms based on ε^+ indicator and confirms that rounded- and ε-archiving reproduce the same quality Pareto approximate front as the point-based archiving but with fewer solutions.

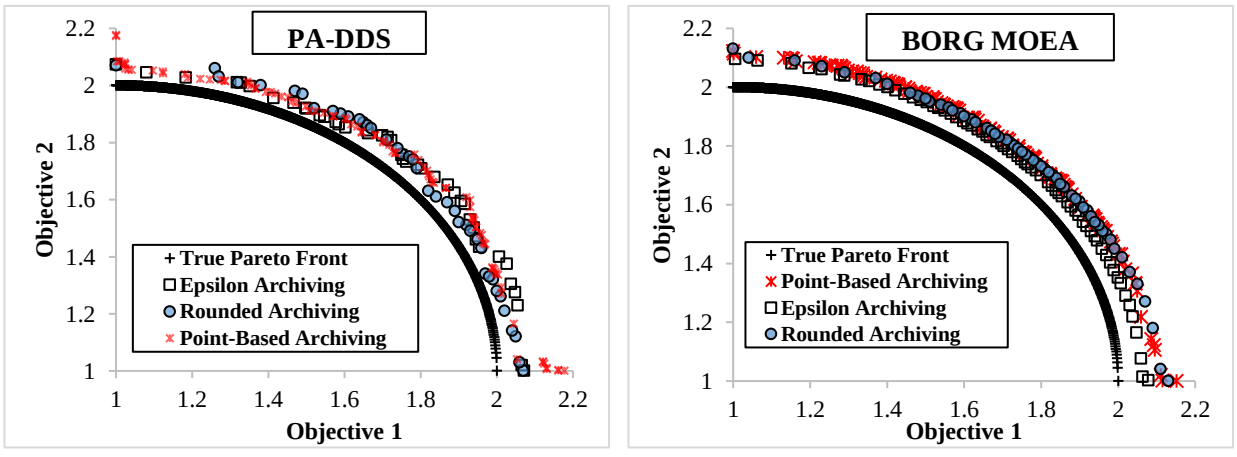


Figure 2-5: Pareto approximate fronts identified by PA-DDS and Borg MOEA using alternative archiving approaches solving R2-DTLZ2-2D with the budget of 10000 solution evaluations. The trial with median ε^+ indicator value is represented for each algorithm.

According to Figure 2-6-a, rounded-archiving is preferred for PA-DDS for solving DTLZ1-5D with 10000 solution evaluations, even though it helps PA-DDS archive fewer solutions on average, 383 versus 585 and 608 as in Table 2-3. Other CDF plots that are not provided in this paper show crossing behavior similar to Figure 2-6-b meaning that changing the archiving technique results in a similar performance of PA-DDS and Borg MOEA.

Table 2-3: The average number of solutions archived by different solution archiving techniques for solving mathematical test problems with a budget of 10000 solution evaluations

MO Problem	MO Algorithm	Archiving Method		
		Rounded-Archiving	ε-Archiving	Point-Based Archiving
DTLZ1_2D	PA-DDS	10	9	80
	Borg MOEA	20	16	41
DTLZ1_5D	PA-DDS	383	585	608
	Borg MOEA	354	296	622
R2_DTLZ2_2D	PA-DDS	41	43	85
	Borg MOEA	58	68	177
R2_DTLZ2_5D	PA-DDS	589	561	730
	Borg MOEA	760	780	780

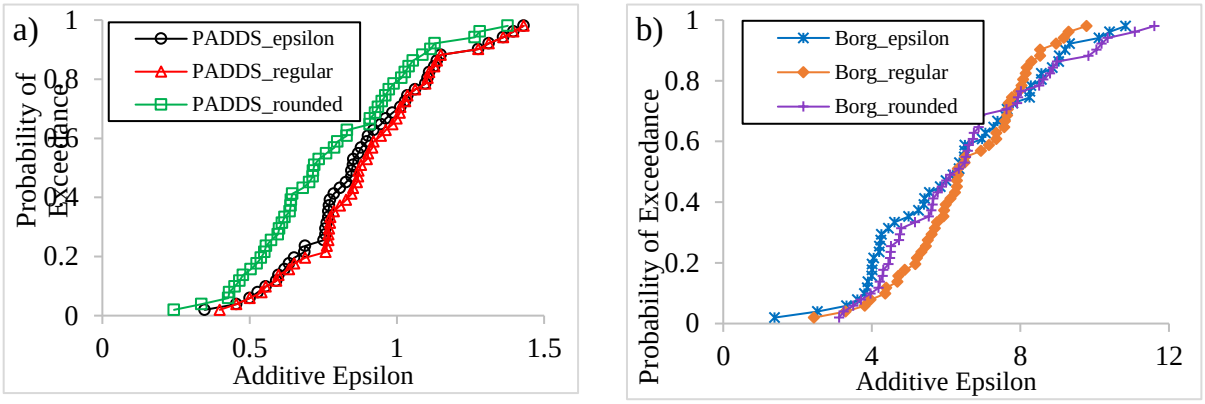


Figure 2-6: Empirical CDF plots of ε^+ indicator comparing alternative archiving approaches for a) PA-DDS and b) Borg MOEA for solving DTLZ1-5D with 10000 solution evaluaitons. A vertical line at zero is ideal.

2.4.1.2 Hydrologic Models Calibrations

According to Table 2-4, the number of archived solutions decreases significantly when the rounded- or ε-archiving technique is applied. Results show that the average archive size reduces

to about 22% and 50% in PA-DDS and Borg MOEA, respectively, when rounded- or ε-archiving is applied for solving the five-objective SWAT calibrated with 10000 solution evaluations. Moreover, the rounded-based Borg MOEA stochastically dominates the CDF plot of the point-based Borg MOEA based on ε^+, as illustrated in Figure 2-7. In the high computational budget calibration case, the two-sided Willcoxon rank-sum tests confirm the preference toward rounded- over point-based archiving for Borg MOEA in the SWAT calibration problem (p-values of 0.0022, 0.000769, and 0.0046 for the ε^+ indicator, *GD*, and *NHV*, respectively). The preference toward rounded- against ε-archiving Borg MOEA is only statistically significant based on the *GD* (p-value of 0.0312). Results did not show any statistially significant difference between alternative archiving approaches for Borg MOEA in the low computational budget case (i.e., 1000-simulation runs). For PA-DDS used for SWAT calibration, no statistically significant preference is observed among the three approaches, except for the point-based archiving that is preferred over the other two archiving approaches based only on *GD* for 10000 simulations with a p-value of 0.000583. All the three archiving approaches statistically preserve the same level of diversity in high computational budget for PA-DDS in terms of ε^+ indicator and *NHV* indicators with considerably smaller archive size of 504 for the rounded- and ε-based PA-DDS against 2264.

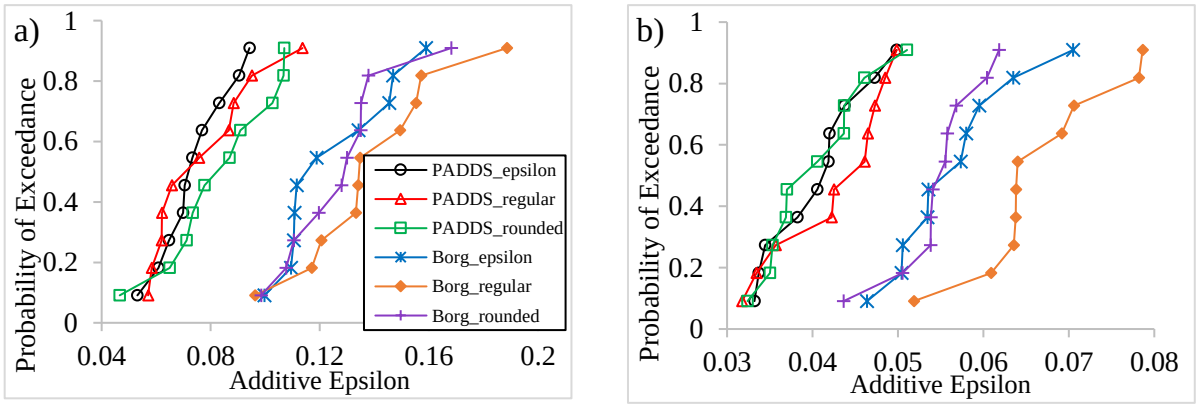


Figure 2-7: Empirical CDF plot of ε^+ indicator comparing alternative archiving approaches for PA-DDS and Borg MOEA for calibrating five-objective SWAT with a) 1000 and b) 10000 solution evaluations

Table 2-4: The average number of solutions archived by different solution archiving techniques for model calibration problems

MO Problem	Number of Evaluations	MO Algorithm	Type of Archiving Method		
			Rounded-Archiving	ε-Archiving	Point-Based Archiving
SWAT	1000	PA-DDS	136	135	320
		Borg MOEA	130	104	166
	10000	PA-DDS	504	504	2264
		Borg MOEA	483	437	916
RAVEN	1000	PA-DDS	207	213	680
		Borg MOEA	309	326	787
	10000	PA-DDS	812	689	6097
		Borg MOEA	1160	1104	3749

Table 2-5 demonstrates the proximity toward the utopia or the ideal point by finding the solution that has the shortest distance (knee point) from the utopia point. In light of the fact that the extreme events are of greater importance compared with low flow rates, a higher weight (0.4) is given to the *NSE* that is more sensitive to larger errors that often occur in high-flow periods. The rest of the objective functions are equally weighted. In general, the distance to the utopia point is shorter when rounded-archiving is used for SWAT, except for PA-DDS that performed better with ε-archiving with the budget of 1000 solution evaluations. This means that the Pareto approximate front is bent more towards the ideal point in the case of rounded-archiving.

According to Table 2-4, the point-based PA-DDS archives more than 60% of generated solutions in the case of the seven-objective RAVEN model calibration for 10000 function evaluations. While, the rounded- and $\boldsymbol{\varepsilon}$-based PA-DDS respectively archive only about 8 and 7 percents of all generated solutions in the same number of simulations. In the 1000-evaluation experiment on PA-DDS, the rounded- and $\boldsymbol{\varepsilon}$-based techniques store only 207 and 213 solutions in the archive significantly less than 680 solutions that are archived using the original PA-DDS. This means the traditional archiving method retains all the relatively similar solutions that have even

slightly better value in only one out of seven objectives that is challenging for the calculation of one final preferred solution.

Figure 2-8 shows that despite the significant decrease in the number of archived solutions, both rounded- and ε-archiving approaches are preferred over the point-based archiving version of the Borg MOEA for 10000 simulations with a p-value of 0.0211. According to Table 2-5, rounded-archiving is preferred for both PA-DDS and Borg MOEA for calibrating the seven-objective RAVEN model because it helps the MO algorithms find the Pareto approximate front with a knee point closer to the ideal point. The only exception is the point-based archiving that is preferred for Borg MOEA at the limited budget of 1000 solution evaluations.

Table 2-5: Knee points distance from the utopia point for the SWAT and RAVEN models calibration using Borg MOEA and PA-DDS

MO Algorithm	Number of evaluations	Archiving method	Distance	
			SWAT	RAVEN
Borg MOEA	1000	Point-Based Archiving	**0.135**	**0.162**
		ε-Archiving	0.141	0.178
		Rounded-Archiving	**0.135**	0.170
	10000	Point-Based Archiving	0.110	0.155
		ε-Archiving	0.107	0.155
		Rounded-Archiving	**0.105**	**0.152**
PA-DDS	1000	Point-Based Archiving	0.118	0.163
		ε-Archiving	**0.111**	0.159
		Rounded-Archiving	0.114	**0.158**
	10000	Point-Based Archiving	0.097	0.153
		ε-Archiving	0.099	0.149
		Rounded-Archiving	**0.096**	**0.147**

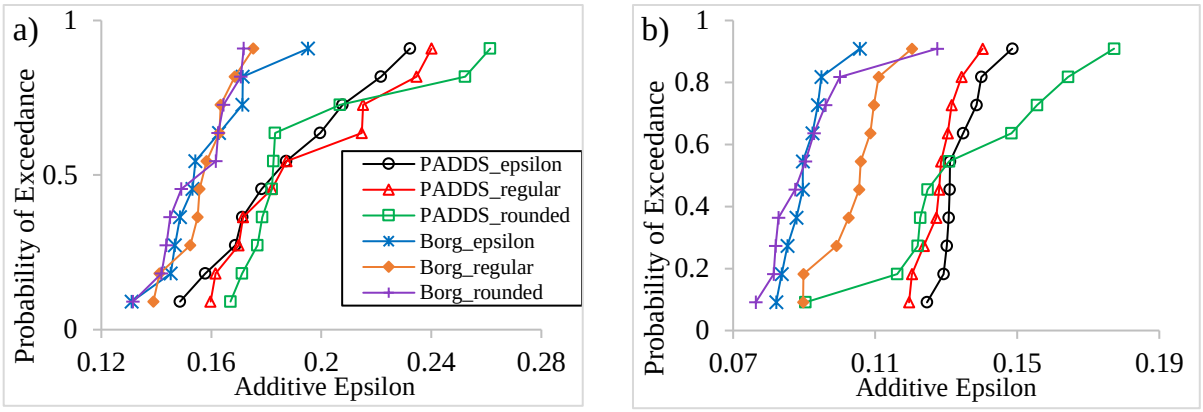


Figure 2-8: Empirical CDF plot of ε^+ indicator comparing alternative archiving approaches for PA-DDS and Borg MOEA for calibrating seven-objective RAVEN with a) 1000 and b) 10000 solution evaluations

2.4.2 Method Validation: AMALGAM

The rounded-archiving is tested on AMALGAM for solving the many-objective hydrologic model calibrations. AMALGAM uses the point-based archiving technique and has a bounded archive size that is defined *a priori*. Therefore, it is expected to suffer from both deterioration and dominance resistance issues when applied to many-objective optimization problems.

According to Figure 2-9, rounded-archiving improves the performance of AMALGAM for calibrating both of the five-objective SWAT and seven-objective RAVEN models with the budget of 10000 solution evaluations. This improvement is statistically meaningful with 95% confidence level based on the Wilcoxon rank-sum test results in Table 2-6, except for the case of RAVEN based on the ε^+ indicator. At the lower budget of 1000 solution evaluations, the preference of rounded-archiving over the point-based archiving for AMALGAM is statistically significant in case of the RAVEN model calibration but not for the SWAT model calibration, see Figure 8-1 and Table 2-6.

Table 2-6: The p-value of Wilcoxon rank-sum by testing the hypothesis of existing significant difference between the optimization results of rounded- versus point-based archiving for AMALGAM

Hydrologic Model	1000			10000		
	ε^+	*GD*	*NHV*	ε^+	*GD*	*NHV*
SWAT	0.7913	0.2413	0.0757	**0.014**	**0.000583**	**0.000183**
RAVEN	**0.0211**	**0.0022**	**0.0073**	0.4727	**0.0004374**	**0.0036**

It is understood that, rounded-archiving decreased the oscillating issue due to the bounded archive of AMALGAM. The bounded archive makes the MO algorithm eliminate some currently non-dominated parameter sets and retain inferior ones in future iterations confirming the published study of Hanne (1999). In addition, the archive is rapidly filled with the initially generated low-quality non-dominated solutions and the subsequent solution production are based on the perturbation of the low-quality archived ones. While, the rounded-archiving method gradually stores the limited archive set by producing higher quality solutions during the optimization and it properly represents the entire range of the reference set if the precision level of each objective function is properly quantified.

A further experiment on the impact of the rounding level of the five-objective R2-DTLZ2 mathematical test problem in Figure 2-10 demonstrate that consideration of the low-resolution desired precision level results in a significant improvement on the ε^+ value. A similar result to Figure 2-10 is expected for the computationally intensive problems such as the model calibration problems of this study. However, the modeller should be aware that a high rounding level increases the chance of multi-modality in optimization.

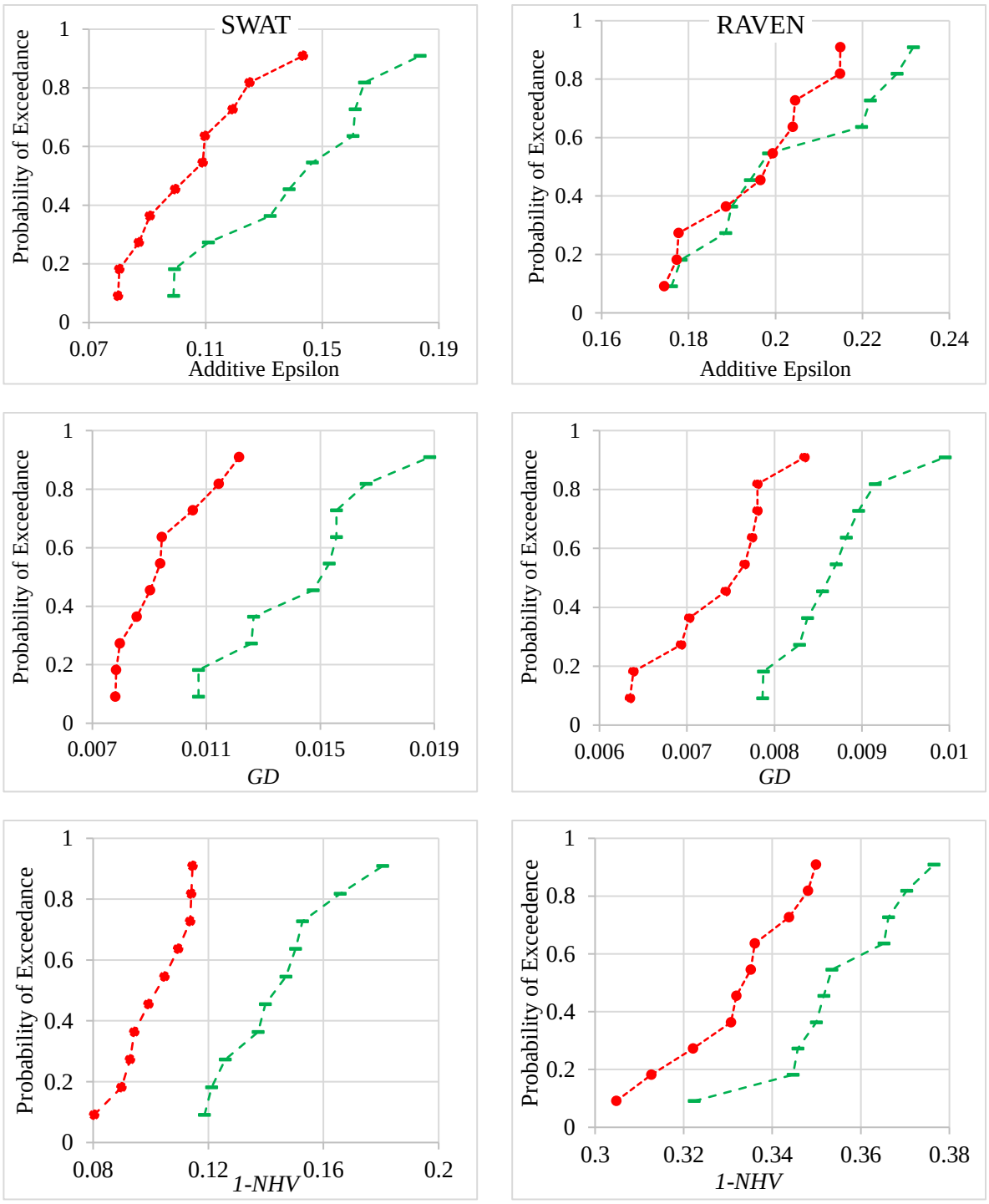


Figure 2-9: Empirical CDF plots comparing point-based (green series) versus rounded- (red series) archiving for AMALGAM for calibrating five-objective SWAT and seven-objective RAVEN models with 10000 solution evaluations

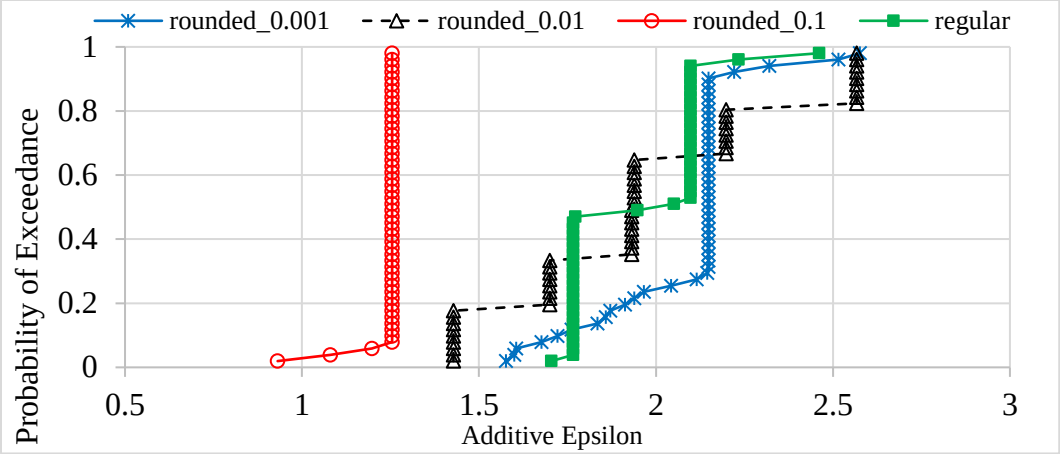


Figure 2-10: The effect of rounding level on the empirical ε^{+}-CDF of AMALGAM solving five-objective R2-DTLZ2 mathematical problem

2.5 Conclusions

A comparative study of an alternative rounded-archiving method to ε-archiving was conducted. The proposed archiving method of this study can carry out the task of ε-archiving for algorithms that are not equiped with ε-archiving. The rounded-archiving is user-friendly, in that it does not require altering the MO algorithm computer code. This method is not algorithm- or case-specific and is much needed in situations where (1) the algorithm is not designed for solving many-objective case studies and (2) the algorithm has a bounded archive structure. The rounded-based Pareto approximate front is expected to properly represent the entire range of the true Pareto front if the objectives' resolutions are sufficiently quantified, similar to ε-archiving in Reed et al. (2007).

At the methodology development stage of this study, the rounded-archiving resulted in significantly smaller archive size especially in high computational budgets for the five-objective SWAT and seven-objective RAVEN model calibrations using PA-DDS and Borg MOEA while maintaining a well-diverse set of solutions in the objective space. The rounded-archiving has at least the same convergence level as ε-archiving with no significant difference in the Pareto front

diversity for solving mathematical and model calibration problems using PA-DDS and Borg MOEA. The calculated knee points resulted from the proposed method have shorter distance to the ideal or utopia point than those obtained by the point-based and ε-archiving in the majority of numerical experiments. It is concluded that, the MO optimization algorithms need to be equipped with more specialized archiving strategies such as the ε- or the rounded-archiving to solve many-objective optimizaton problems since the traditional point-based archiving method leads to exponential increase in the number of non-dominated solutions, especially when MO algorithms have an unbounded archive.

The proposed rounded-archiving method is tested on AMALGAM that has a bounded archive and is not equipped with ε-archiving. Rounded-archiving reamarkably improves the level of proximity and diversity of the Pareto approximate front compared to the original AMALGAM. Results show that rounded-archiving made AMALGAM perform better than the original AMALGAM in eight out of 12 hypothesis testings. The rounded-archiving can be easily generalized to other multi-objective algorithms for solving many-objective optimization problems.

Reducing the number of archived solutions using either rounded- or ε-archiving assists decision-makers and model analysts to find their desired solution or parameter set based on the status-quo of their projects with less confusion. It should be noted that the resolution/rounding level of each objective needs to be consulted with decision-makers prior to optimization setting. It is shown in this study that using a coarse rounding level leads to a smaller number of archived solutions, which confirms findings by Reed et al. (2007). The main advantage of rounded-archiving over ε-archiving, is that it is readily available for any MO algorithm that is not equipped with any specialized archiving, such as PA-DDS and AMALGAM in this study. Moreover, rounded-archiving is computationally more efficient, but this advantage becomes trivial when

solving problems with a time-consuming simulation and/or evaluation models such as the model calibration case studies of this paper.

The ε-archiving approach retains the exact values of objective vectors while rounded-archiving will increase the probability of multi-modality if a very low resolution is set for objectives. Rounded-archiving needs to be applied to other MO algorithms and other water resources case studies but owing to the highly time-demanding process of this comparison study, three MO algorithms are considered as sufficient for generalizing the concluding remarks of this study.

2.6 Acknowledgment

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant of the second author, [RGPIN-2016-05896].

3 CLUSTER-BASED MULTI-OBJECTIVE OPTIMIZATION FOR IDENTIFYING DIVERSE DESIGN OPTIONS: APPLICATION TO WATER RESOURCES PROBLEMS

Shahram Sahraei[1], Masoud Asadzadeh[2]

[1] Ph.D. Candidate (Corresponding Author), Department of Civil Engineering, University of Manitoba, Canada

E-mail: sahraeis@myumanitoba.ca

[2] Assistant Professor, Department of Civil Engineering, University of Manitoba, EITC E1-332, 15 Gillson Street, Winnipeg, MB, Canada, R3T 5V6, Ph: +1 (204) 474 9535

E-mail: masoud.asadzadeh@umanitoba.ca

3.1 Abstract

In this study, a novel density-based spatial clustering method is developed to maintain a diverse set of solutions for stochastic multi-objective optimization algorithms. This method dynamically clusters solutions in the decision space after solution evaluations. Dominance check is localized to maintain solutions that are globally dominated but locally non-dominated in their cluster. Unlike the original solution archiving, the proposed method implemented for Pareto Archived-Dynamically Dimensioned Search successfully finds optimal and near-optimal fronts with different cluster labels in two mathematical case studies. Two environmental benchmark problems are also solved and a three-stage screening process is applied to their archive sets to identify the number of dissimilar options. The dissimilarity index devised for this study shows a significantly higher distinction level and archive size for the cluster-based solution archiving, which allows

decision-makers to have higher flexibility in refining their preferences for robust decision-making in the environmental problems, compared with the original archiving.

Keywords: Multi-objective optimization, decision-space diversity, dynamic clustering, DBSCAN, water resources, engineering design

3.2 Introduction

Engineering Design Problems are perceived as multi-objective decision-making problems to mitigate the needs of all parties by enhancing the transparency, auditability and reliability of these decisions (Dunning et al., 2000; Hajkowicz and Collins, 2007). Explicit consideration of all objectives simultaneously can help stakeholders to avoid decision biases, particularly when planners inadvertently neglect aspects of the problem by concentrating on a narrow definition of problem optimality (Kasprzyk et al., 2013; Reed et al., 2013). However, due to the system complexity and various sources of uncertainty in the decision-making and future condition in water resources engineering systems such as reservoir operation (Asadzadeh et al., 2014a; Erfani et al., 2020; Geressu and Harou, 2019; Zatarain Salazar et al., 2017), hydrologic models (Ahmadi et al., 2014; Asadzadeh et al., 2016; Koppa et al., 2019; Sahraei et al., 2020; Sikorska and Renard, 2017), and system design (Asadzadeh and Tolson, 2012; Bode et al., 2019; Marques et al., 2015; Zheng et al., 2015), the multi-objective techniques require a tool that provides a flexibile environment for stakeholders to select among a large number of distinct options.

Access to multiple distinct optimal and/or near-optimal alternatives in environmental systems optimization is important in many real-world circumstances as these types of optimizations are mathematically modelled in a simplified fashion considering highly important objectives, only. These problems are under modelling uncertainties due to data unavailability, unmodelled

objectives, objective prioritization, and other modelling limitations, which changes their behavior to a multi-modal optimization (Brill et al., 1982; Burton et al., 1987; Chang et al., 1982; Harrington and Gidley, 1985; Rogers and Fiering, 1986; Rosenberg, 2015; Voll et al., 2015; Zechman et al., 2013) and/or interval multi-objective optimization (Gong et al., 2020, 2013; Sun et al., 2020). Multi-modality refers to a situation that different solutions perform similarly in the objective space. Interval multi-objective optimization refers to a problem that at least one objective or constraint depends on an uncertain coefficient whose upper and lower bounds are known (or can be identified) a priori with a high confidence level making the objective function have an interval characteristic for each point in the decision space (Gong et al., 2020, 2013; Sun et al., 2020). Therefore, other criteria such as stakeholders' perspectives about social, environmental, and economical issues that cannot be mathematically formulated are involved in the post-processing stage for selecting most desirable solution(s) through subjective judgements (Liebman, 1976). Examples are synthesis of distributed energy supply systems (Voll et al., 2015), pollution control (Rosenberg, 2015), reservoir operation (Liu et al., 2011), and aircraft engine design problems (Zadorojniy et al., 2012). This paper proposes a novel approach for identifying distinct solutions of multi-objective optimization (MO) problems, including the ones on the best Pareto front and the ones on the near-optimal front. Moreover, the proposed approach can identify and archive distinct solutions of multi-modal problems.

The past twenty years have seen increasingly rapid advancements in the performance of MO algorithms to efficiently identify multiple optimal solutions (Pareto set), which form an optimal set of points (tradeoff or Pareto front) in the objective space (Asadzadeh and Tolson, 2013; Deb et al., 2002b; Macro et al., 2019; Reed et al., 2013; Sahinidis, 2004; Vrugt and Robinson, 2007; Zhang and Li, 2007). Mainstream MO algorithms are equipped with a dominance check and

selection strategies that often look at the distribution of points in the objective space to approximate the entire Pareto front but they overlook the position of the solutions in the decision space. As a result, they are not able to preserve near-optimal solutions that have completely distinct decision variables compared to Pareto solutions since they will be dominated by optimal solutions. Figure 3-1 demonstrates a hypothetical bi-objective example with 2-dimensional decision space containing a cluster of solutions that is distant from the Pareto set but form a near-optimal front in the objective space. A MO algorithm, as *a posteriori* decision-making approach, that is able to identify distant local Pareto sets and preserve near-optimal solutions with distinct design or modelling characteristics from the globally optimal Pareto set(s) has a higher chance to locate the most desirable solution.

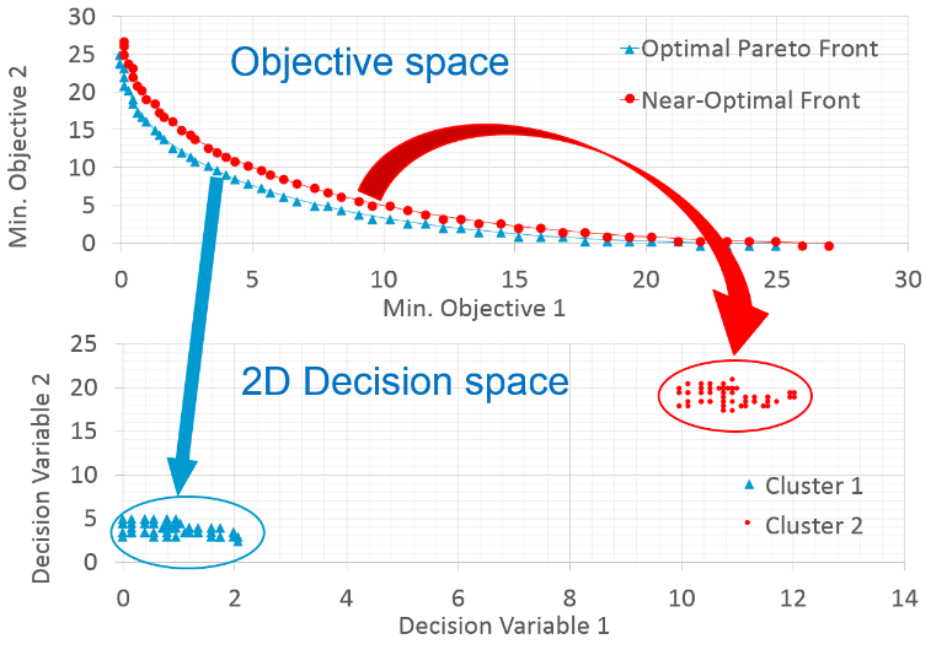


Figure 3-1: an illustrative example of near-optimal and optimal fronts with entirely distinct clusters in 2D decision space.

This study focuses on decision-space diversity maintenance for finding distinct decision/design options to have a flexible and robust decision making analysis. To this end, a novel dynamic clustering approach coupled with a localized dominance relation is introduced to retain near-optimal fronts along with the optimal front for MO algorithms with an unbounded archive set. Solutions residing in one cluster are mutually compared in terms of dominance, and those that have different cluster labels are not compared for domination. The proposed methodology is applied to a Pareto Archived-Dynamically Dimensioned Search (PA-DDS) algorithm (Asadzadeh and Tolson, 2013) and its performance is compared to PA-DDS with original archiving and the omni-optimizer (Deb and Tiwari, 2008) as a reference algorithm that considers decision-space diversity in multi-modal situation.

3.3 Related Work

The stochastic MO algorithms in the literature are classified into three categories in terms of diversity maintenance in the decision space. These algorithms are capable of solving multi-modal MO problems, effectively but not designed to preserve near-optimal solutions in most cases.

3.3.1 Decision Space Diversity as an Optimization Criterion

A group of indicator-based MO algorithms use a diversity metric calculated in the decision space to guide their search. The genetic diversity evolutionary algorithm by Toffolo and Benini (2003) is equipped with two metrics used as optimization criteria. One metric is maximizing the shortest distance to neighbor solutions to encourage exploration of the decision space. The second metric is maximizing the dominance rank of each solution with respect to the original objectives of the MO problem to exploit the promising regions of the decision space. The former metric is designed to keep a diverse set of solutions while the latter emphasizes the optimality or convergence. Ulrich et al. (2010) developed an indicator-based evolutionary MO algorithm that integrates the decision-

space diversity metric into the objective-space hypervolume metric. The diversity metric sums the distance of solutions from the median of the non-dominated set in the decision space that monotonically increases as new solutions are added to the non-dominated solution set. The modified metric is a weighted hypervolume measure that divides the dominated portion of the objective space into hypercubical segments, where each segment is dominated by a specific subset of the entire population. The hypervolume partition for each segment is thus weighted by the diversity of its dominating solutions and their summation is maximized.

The aforementioned methods solely aim at enhancing relative distribution of solutions with each other in the decision space, but they are not designed to retain solutions that are distant from the optimal solution in the decision space but have a similar quality in the objective space. Zadorojniy et al. (2012) suggested two algorithms for product design problems to find near-optimal options by considering a degree of compromise of the known global optima. The first algorithm maximizes the diversity of solutions in the decision space constrained by a maximum allowable compromise (for example, 2%-5% optimality violation) from the Pareto optimal front in the objective space. In the second algorithm, violation from the Pareto optimal front is minimized subject to a required decision space diversity. This method, however, requires a prior knowledge of the Pareto optimal front to commence the search for finding solutions with maximum decision space diversity, and it requires a careful setting of optimality violation and decision-space diversity thresholds, which are case-specific.

3.3.2 Decision Space Diversity as Selection Operator

The selection operators in mainstream MO optimization algorithms focus on the diversity of points in the objective space and/or the convergence toward the Pareto optimal front; therefore, it is a challenge for them to maintain different solutions that have similar objective vectors. To address

this challenge, Deb and Tiwari (2008) developed a toolkit known as Omni-optimizer, a type of generational genetic algorithm (Vavak and Fogarty, 1996), that uses the crowding distance calculated in both objective and decision spaces (Deb et al., 2002b) for diversity preservation in multi-modal problems. Crowding distance is a measure of the solution density around a particular solution, and a higher importance should be given to solutions in a less crowded region, i.e. higher crowding distance value. If two solutions are identical in the objective space but are distant in the decision space, the Omni-optimizer retains both of them, since they are two different local optima. This algorithm, nevertheless, gives a superior weight to the diversity in the objective space by considering the crowding distance in the decision space only if there is a tie in the crowding distance in the objective space.

Chmielewski (2013) introduced a diversity ranking evolutionary MO algorithm (DREMA) with a similar non-dominated sorting method to Deb et al. (2002) that uses the hypervolume contribution as the objective space fitness metric for sorting solutions in each non-dominated front. DREMA sorts solutions based on three diversity metrics calculated in the decision space to find distinct solutions: a) dispersion that is the sum of Euclidean distances of a solution from two neighboring solutions in the decision space; b) remoteness that is the distance to the nearest solution in the decision space; and c) alternate ranking that is the Euclidean distance in the decision space between two nearest solutions in the objective space. Solutions situated in the above-average portion of solution ranking list based upon at least one out of three decision-space diversity metrics and above-average hypervolume contribution are given more opportunity to be chosen for generating new solutions. Nonetheless, this method is algorithm-specific and applicable only to population-based optimization algorithms that use solution-sorting strategies.

The concern about multi-modality led Cuate and Schütze (2019) to define the (inverse) variation rate that assigns each non-dominated solution a combined measure of proximity in the objective space (such as crowding distance, hypervolume contribution, or weighted aggregation of objectives) divided by the average distance from other solutions in the decision space. A solution with lower variation rate (or higher value for its inverse version) has a higher chance to be selected for generating new solutions, since it is distant from other solutions in the decision space while having the same proximity in the objective space. However, this metric cannot preserve potentially useful solutions that are nearly optimal in terms of objectives but completely distinct in the decision space.

3.3.3 Niching Method

The concept of niching that was first introduced by Cavicchio (1970) formed the basis for developing different conventional approaches such as crowding factor (De Jong, 1975), fitness sharing (Goldberg and Richardson, 1987), and speciation (Li et al., 2002; Petrowski, 1996) for solving multi-modal, single-objective problems by giving a higher importance to solutions in less crowded regions of the decision space. In light of the fact that the fitness sharing gives a higher chance of selecting solutions in smaller niches, their preservation is not guaranteed (Li et al., 2002). Each niche in speciation method is called a species. The dominating solution in each species is called species seed, and all the solutions fall within a pre-defined neighborhood radius from the species seed belong to the same species. The seeds belonging to different species that are locally non-dominated solutions are copied into the next generation of solutions to maintain elite solutions. Defining a suitable species radius as a measure of dissimilarity requires a great knowledge of the optimization problem at hand and the relationship between decision variables and the objective function (Li et al., 2002).

The niching method was first developed for single-objective optimization algorithms. It was first utilized in a MO problem by (Horn et al., 1994). In an early study of MO optimization, Zitzler and Thiele (1998) used a distance-based niching approach for decreasing the size of archive set in Strength Pareto Evolutionary Algorithm (SPEA) based on the objective-space diversity. However, they stated that niching could be based on the distribution of solutions in the decision space (Deb, 2001).

Shir et al. (2010) proposed a dynamic niching framework for the covariance matrix adaptation evolution strategy. The niching framework decreases the contribution of the domination ranking in the selection process that uses a dynamically adjustable niching radius. In order to form a pre-specified number of niches, the solutions are checked in their neighborhood after non-dominated sorting, and the ones belonging to the same niche fall within a hyper-sphere. The Euclidean distance between solutions is calculated in the aggregated space, i.e. the decision and objective spaces. The fittest solution in each niche is the representative of that niche, and they are retained to guarantee elitism. Zechman et al. (2013) proposed a MO niching co-evolutionary algorithm that creates independent multi-sets of solutions in parallel in each generation assuming that each set of solutions represents distinct non-dominated sets in the decision space. The original procedure in non-dominated sorting genetic algorithm (NSGA-II) (Deb et al., 2002b) is applied to the primary set to detect non-dominated front as a reference for other parallel sets and ensure convergence. The algorithm then combines all the solution sets and groups solutions using the *k*-means clustering approach (Macqueen, 1967) in the objective space. Therefore, solutions with the same cluster label in the same solution set reside in the same niche in the decision space. The algorithm prefers selecting a solution distant from its niche center in less crowded regions within the *T* percent optimality of the non-dominated front of the primary set. Despite finding near-optimal

fronts, some niches in different sets happen to reside in the same region of the decision space, since independent solution sets independently form and evolve niches, which wastes the computational budget.

Kramer and Danielsiek (2010) developed an evolutionary optimization strategy that uses reference lines for attaining uniform distribution of non-dominated solutions in the objective space (Kramer and Koch, 2009) and uses a density-based clustering approach (Ester et al., 1996) to preserve diversity in the decision space. Niches are identified in the decision space and are evolved independently for a specific number of generations, until two neighboring solutions in the objective space belonging to one niche have a distance higher than a user-specified threshold in the decision space, then all the niches are combined and re-clustered. The re-clustering threshold is problem-dependent and one cluster may be excessively expanded during evolvement leading to a merge with other niches in re-clustering stage and losing useful local optimal fronts. For the sake of preservation of local optimal solutions along with global optimal Pareto front, Pajares et al. (2018) introduced a new concept of domination for the MO genetic algorithm that considers closeness of solutions in screening process, i.e. dominance localization. Two archives are provided for this algorithm. Near-optimal solutions are preserved in a separate archive set and the archived solutions are mutually distant. If a dominated solution is far enough from the non-dominated ones based on a pre-defined dissimilarity vector for decision-variable vector, it moves into the second archive for dominance and closeness check with near-optimal solutions. As a result, a dominated solution that is not nearby any archived solution is retained. The emphasis is given to less crowded regions of decision space for selecting from the archive by an assignment of sharing fitness and niche count (Sareni and Krähenbühl, 1998) as a measure of neighborhood density. This algorithm, however, has a bounded archive and requires a careful tuning of a user-specified vector for all

dimensions of the decision space as a measure of dissimilarity for removing similar solutions if decision variables have no physical meaning. A complete survey on niching-based optimization can be found in Li et al. (2017), Cheng et al. (2018), and Tanabe and Ishibuchi (2019).

Liu et al. (2019) devised an evolutionary MO algorithm with two bounded archives and a recombination method for multi-modal problems. It can also find near-optimal solutions that are distant from the optimal solutions in the decision space. Decision variables that contribute to convergence only are identified with an analytical technique and separated from convergence-independent decision variables. One archive is assigned to convergence-related decision sub-space and a second archive is used to retain convergence independent decision variables for diversity maintenance. Parent solutions are then chosen from both archives using a tournament selection for reproducing offspring solutions and updating the archives accordingly. Solutions in the convergence archive are ranked based on a convergence indicator. Solutions nearby another solution in the convergence-related decision sub-space are de-emphasized from the archive if they have the same convergence rank. For the sake of diversity preservation in the objective and decision spaces, solutions nearby another solution in the convergence-independent decision sub-space are de-emphasized and given a lower chance of selection if they are clustered around a reference vector in the objective space. After the termination of optimization, two archives are recombined to obtain a final set for a posteriori decision-making. They stated that the basic clearing technique that de-emphasizes the neighbor solutions may not find all local optima regions in the decision space if the spacing between these regions are different.

The remainder of this paper is organized as follows. Next section begins with a description of the density-based spatial clustering method and its application for a cluster-based solution archiving in MO algorithms with an implementation example on PA-DDS. The benchmark

optimization problems and numerical experiment settings are explained in section 3.4.5 and section 3.4.6, respectively. The results of the proposed methodology are presented in section 3.5 and a comparison discussion is made with original version of PA-DDS for each problem, followed by the concluding remarks (section 3.6).

3.4 Materials and Methods

The density-based spatial clustering is restructured to dynamically revise the clusters as new solutions are archived by the MO algorithm. The dominance-check is decentralized to eliminate solutions from the archive only if they are dominated within their own cluster. The advantage of the introduced clustering approach against *k*-mean, hierarchical method, and distribution-based methods is that it does not require *a priori* information about the number of clusters or the relationship between the objective space and decision space.

3.4.1 Density-Based Spatial Clustering Of Applications with Noise (DBSCAN)

DBSCAN introduced by Ester et al. (1996) is an unsupervised clustering method for data mining that looks at the neighboring density of each data point in large spatial databases with minimal domain knowledge requirements. Each cluster identified by DBSCAN contains at least one core point (blue doughnuts in Figure 3-2) that has to be within a user-defined neighboring distance (ε), i.e. the adjacency from a pre-defined minimum number of points including itself (*MinPts*) (Ester et al., 1996). The value of ε in Figure 3-2 is equal to the radius of the circles and *MinPts* is set to four or five. Each member of a cluster (blue and red doughnuts in Figure 3-2) is therefore density-reachable from at least one core member. Two members of a cluster are called density-connected if they are neighbors of (density-reachable from) a core. A point that is not a core member of a cluster but is density-reachable from a core is called a border member (red doughnuts). An

unreachable point from any core is an unclassified data point and called noise in DBSCAN (green doughnut in Figure 3-2).

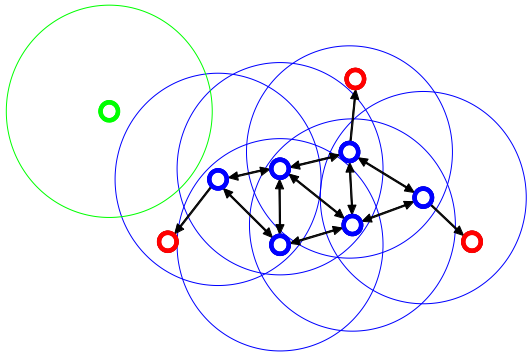


Figure 3-2: schematic of the DBSCAN clustering method. Blue, red, and green doughnuts are respectively core, border, and noise points with respect to *MinPts*=4 and a unit radius (ε) in 2D space

3.4.2 Cluster-Based Solution Archiving

Figure 3-3 illustrates the general structure of a MO algorithm equipped with the cluster-based solution archiving. In order to eliminate the scaling effect of decision variables for clustering, the decision space is normalized to a hypercube of size one. Coordinates of each solution $\boldsymbol{X} = (x_1, \dots, x_n)$ in the normalized decision space is calculated by Equation (12) based on the preset lower bounds, $\boldsymbol{X_{min}}$, and upper bounds, $\boldsymbol{X_{max}}$, for decision variables. The Euclidean distance is used for clustering but other measures such as Manhattan and Minkowski distances are applicable (Xu and Wunsch, 2008). After clustering update in every iteration, solutions are converted back to their original ranges for model simulation and solution archiving.

$$\boldsymbol{X_{normalized}} = \frac{\boldsymbol{X} - \boldsymbol{X_{min}}}{\boldsymbol{X_{max}} - \boldsymbol{X_{min}}} \qquad \textbf{(12)}$$
$$\boldsymbol{X_{min}} = (x_{1,min}, \dots, x_{n,min})$$

$$\boldsymbol{X_{max}} = (x_{1,max}, \dots, x_{n,max})$$

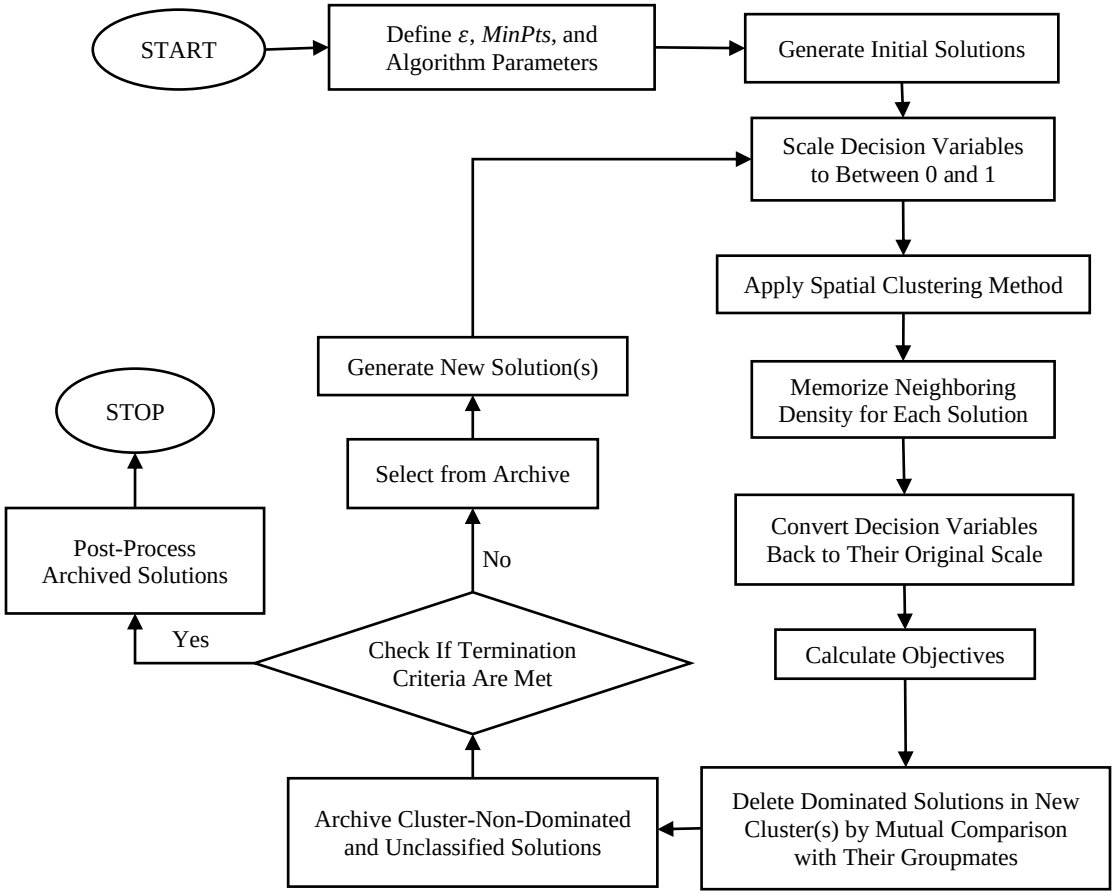


Figure 3-3: General structure of a MO algorithm with cluster-based archiving

The original DBSCAN is used to cluster initial solutions generated by the MO algorithm in the normalized decision space. However, it is modified to dynamically evolve clusters as new solutions are introduced by the optimization algorithm. Upon generating new solutions, their neighborhood is checked for forming a new cluster or expanding the so-far-clustered solutions. To prevent the formation of only one Pareto set in the decision space, the dominance and archiving strategy of the algorithm are decentralized from the global to a local dominance check. Equation (13) shows that solution X_A cluster-dominates solution X_B ($X_A \preccurlyeq_c X_B$) if and only if three

conditions are met. The first condition is extra to the regular dominance relation (second and third conditions) and ensures that both solutions reside in the same cluster. To this end, solutions with the same cluster label are mutually compared and solutions belonging to different groups are not compared by the dominance check. If a solution is not classified yet, its cluster label is assumed to be zero and a dominance check is not applied to it. The cluster-dominated solutions are eliminated from the cluster and the cluster-dominating and cluster-non-dominated solutions are archived as representative of the cluster.

$$X_A \preccurlyeq_c X_B \; iff$$

$$\begin{aligned} &1)\;\; Cluster\_ID(X_A) = Cluster\_ID(X_B) \neq 0 \\ &2)\;\; f_i(X_A) \leq f_i(X_B)\;\; \forall i \epsilon \{1, \dots, m\} \\ &3)\;\; f_j(X_A) < f_j(X_B)\;\; \exists j \epsilon \{1, \dots, m\} \end{aligned} \qquad \textbf{(13)}$$

Archived solutions are assigned a memory to store their coverage history for subsequent cluster expansions. Coverage history is the cumulative adjacency density that represents the number of reachable solutions to a cluster-non-dominated solution from the beginning of optimization. This memory helps the border points of a cluster to increase their coverage history and turn to a core point, which aids in cluster expansion by connecting nearby unclassified solutions, if any, to the cluster and their involvement in cluster-dominance. Each cluster forms a tradeoff in the objective space. Figure 3-4 demonstrates a schematic example with two decision variables and two objective functions that are minimized. Assume that, STEP1 shows the current set of archived solutions, two of which have not yet been clustered and the rest are classified into one cluster (shown as Cluster1 in Figure 3-4). Solutions in the cluster are called cluster-non-dominated because they are not

dominated by any other member of the cluster. In STEP2, a new solution is added to the cluster. This new solution is reachable from the core of the cluster and one border point, therefore the coverage values of the core and border point are updated to 5 and 3, respectively. However, one of the cluster members (the blue point in STEP2) is dominated by the new solution; therefore, it is omitted from the cluster and the archive. The coverage values of other solutions in the cluster that are cluster-non-dominated, remain unchanged in order to keep track of the coverage history. In other words, there remain four members inside the cluster in STEP3 but the core solution keeps its coverage history of 5.

Solutions that have not been clustered yet are retained in the archive disregarding their objective values and are called noises to be consistent with the DBSCAN terminologies (see noises in Figure 3-4). The reason for preserving unclassified solutions is to give an opportunity to the optimization algorithm to produce more solutions nearby the unclassified ones to subsequently join previous clusters or form a new cluster. A new cluster forms when a density-unreachable noise from the core(s) of other clusters becomes a core. Two or multiple nearby clusters complementing one local Pareto set are merged if they are sufficiently populated during the optimization.

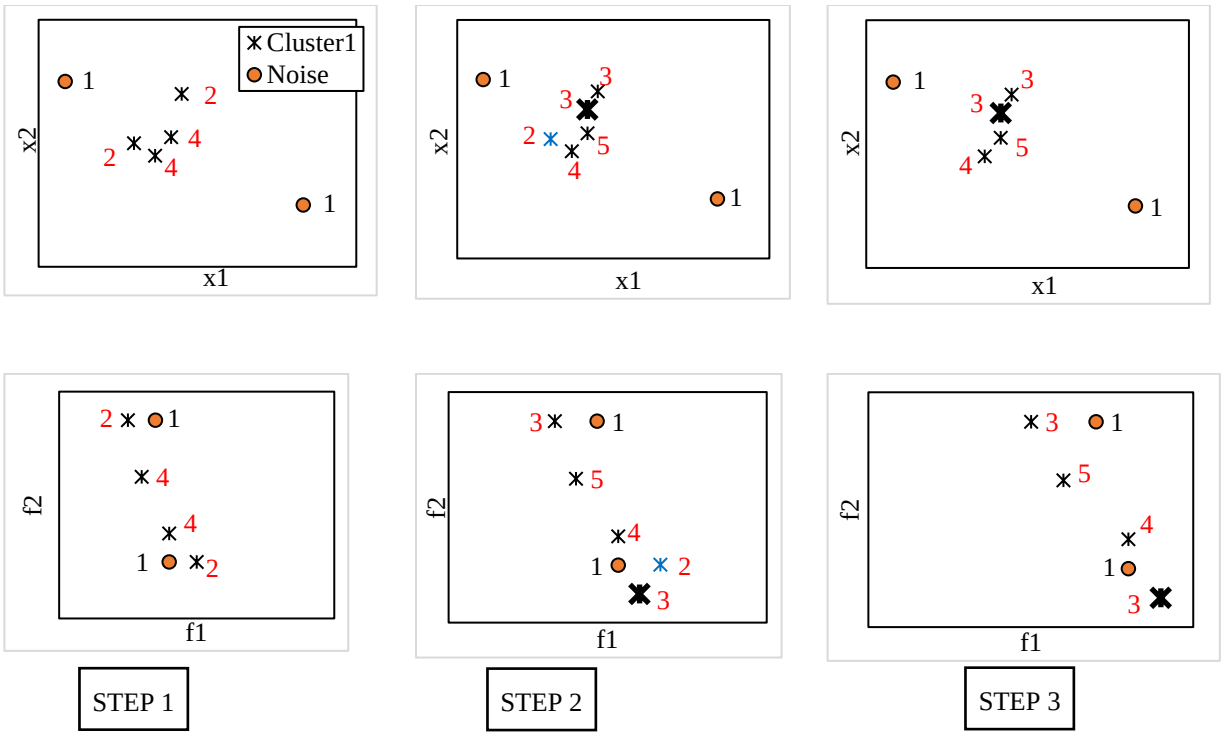


Figure 3-4: Illustration of cluster evolution and coverage history (red and black labels) upon introduction of a new solution to a cluster during optimization for a bi-objective hypothetical problem with two decision variables. Top row shows the decision space and the bottom row displays the objective space. *MinPts* is assumed equal to 4 for a cluster formation. Solutions with labels equal to 4 in Step 1 are the core points of their clusters. The instance with bold marker (Steps 2 and 3) is now generated and added to the cluster. The cluster-dominated solution that is a border point with a coverage of 2 and colored in blue is eliminated from the archive but the coverage history of its groupmates remain unchanged upon solution omission.

The proposed cluster archiving inherits the two parameters of DBSCAN, cluster radius (ε) and minimum solutions coverage (*MinPts*). Assuming a constant value for *MinPts*, defining a small radius for the initial formation of clusters in a limited computational budget may result in appearing many small clusters at the end of the optimization, which could be merged at some point if the number of function evaluations was not low. Ester et al. (1996) suggested that the results of clustering do not significantly change for *MinPts* higher than or equal to four and increasing *MinPts* increases the computational time. *MinPts* is recommended to be set to four or five (Ester et al., 1996). However, ε is a case-specific parameter that has to be defined based on the decision space dimension and the number of function evaluations. A very high ε value results in one cluster covering the entire or a large portion of decision space and preserving only the globally non-

dominated solutions, since all distant and nearby solutions lie in one cluster. By contrast, a very low ε leads to the identification and maintenance of all or the majority of generated solutions as unclassified solutions with no opportunity for new cluster formations and dominance-check. According to our experience, it is recommended to consult with decision-makers prior to the optimization process about the least meaningful discrepancy percentage (D_i) for each decision variable and approximate ε as in Equation (14). The reason for defining a range for ε in Equation (14) is that the available computational budget and the number of dimensions in the decision and objective spaces affect the value of ε for an efficient and effective optimization using the cluster-based solutions archiving.

$$\varepsilon \in [0.8z, 1.2z\,], \quad z = \frac{1}{100}\sqrt{D_1^2 + D_2^2 + \cdots + D_n^2} \tag{14}$$

3.4.3 Cluster-Based Multi-Objective Optimization

The proposed solution archiving strategy is implemented for the Pareto Archiving Dynamically Dimensioned Search (PA-DDS algorithm to find high quality solutions that are distinct in terms of their design characteristics for MO design problems.

3.4.3.1 PA-DDS Algorithm

PA-DDS stochastic MO algorithm generates solutions one-at-a-time. The optimization is initialized with a set of randomly generated solutions within the decision variable boundaries with a budget of the higher value of five solutions and 0.5 percent of the total number of evaluations. The dominance check is then applied to identify and archive the non-dominated solutions in an unbound archive. One solution from the archive is selected for generating one new solution. PA-

DDS commences the heuristic optimization by perturbing all decision variables sampled from Normal distributions centered at the current value of each decision variables and dynamically reduces the number of perturbed decision variables to transform from a global search to a local search near the end of the computational budget. If the new solution is non-dominated or dominating, PA-DDS archives it and selects it for generating the next solution; otherwise, it selects one of the archived solutions. Asadzadeh and Tolson (2013) and Asadzadeh et al. (2014) recommended the hypervolume contribution as the selection metric for solving general MO problems and the convex hull contribution for solving MO problems with expected convex Pareto front. This process continues until the maximum number of function evaluation condition is met.

The dominance relation and selection metrics constitute the principal components of the PA-DDS structure. The process of the domination check and selection strategies are based on the objective space proximity of solutions. Therefore, a near-optimal solution is discarded by PA-DDS even if it is far distant from the dominating solutions.

3.4.3.2 Cluster-Based Archiving in PA-DDS Algorithm

In order to incorporate the proposed cluster-based solution archiving in PA-DDS, the dominance check from a global comparison is reduced to local comparison of solutions with equal cluster tags. Once a new solution is generated, its neighboring solutions are identified based on the minimum adjacency radius and their coverage history are inspected for forming a new or joining a previous cluster or linking two or more cores with different labels and merging the associated clusters. If none of the above cases occurs, the generated solution will be stored as noise or unclassified solution in the archive set for possibly subsequent cluster expansions.

3.4.3.3 Selection Operator

To improve the diversity in the decision space, dissimilar solutions should be retained and selected for generating new solutions, even if they are dominated by other dissimilar solutions. Moreover, the selection operator should consider the convergence in the objective space. In this study, a selection indicator is introduced to promote the solution diversity in the decision space and the convergence in the objective space. The proposed selection indicator is a summation of two metrics that range between zero and one, resulting in a total value of the indicator between zero and two. A higher value gives a higher chance of selection for subsequent solution generation.

The first term of the proposed selection indicator describes the proximity of a cluster of solutions and is calculated based on the distance of its closest solution (called knee point hereafter) to the Utopia point, which has the best value of each objective function. The distance from each noise (unclassified solution) to the Utopia is also calculated. The groupmates of a knee point have equal normalized convergence and equal chance of selection. Solutions belonging to a cluster with higher normalized convergence indicator are given a higher chance to be selected for generating new solutions.

The decision-space dissimilarity index is the second term in the proposed selection indicator. The dissimilarity index value for clusters is a measure of Euclidean distance between knee points with different cluster labels. The dissimilarity of a cluster is calculated as the summation of pairwise distances from its knee point to the knee points of other clusters and to unclassified solutions in the decision space. This metric is scaled to between zero and one based on its maximum and minimum value in each step of optimization. If some of clusters and/or unclassified solutions are packed in one region of the decision space, they will get a low dissimilarity index.

Solutions in clusters and those that are unclassified in less crowded regions receive a higher normalized dissimilarity index of close to one.

Figure 3-5 illustrates the process of calculating the proposed selection metric for a hypothetical situation where there are two clusters of locally non-dominated solutions along with two unclassified solutions in a two-dimensional decision and objective spaces for a minimization problem. In each step of optimization, the Nadir and Utopia points are determined based on the currently generated and archived solutions. The Nadir point correspond to an ever dominated point whose objectives equal to the maximum objective values of the solutions for a minimization problem. The objective space is then normalized to zero and one based on the Nadir and Utopia points and the knee point for each cluster is found, shown with larger marker sizes in Figure 3-5-b. The selection metric for each cluster and each unclassified solution is calculated as Equation (15). The first term is the convergence term that is scaled by diagonal of the normalized objective space, i.e. $\sqrt{2}$ in a bi-objective space and $\sqrt{m}$ in an m-dimensional space, for a convergence between zero and one. The shorter the distance to the origin, the better convergence term and the better quality solution. Once the knee point for each cluster is identified, sum of distance of each knee point from other knee points and from unclassified solutions is calculated in the normalized decision space and it is scaled by the maximum distance summation to have a dissimilarity index between zero and one. Knee points represent their cluster and their groupmates are not involved in the calculation of the proposed selection metric. After computing the selection metric value for a knee point, its groupmates will get the same metric value and one of them in a cluster is randomly selected for generating the subsequent solution.

$$S_{cluster1} = \left(1 - \frac{A}{\sqrt{2}}\right) + \frac{\sum_{i=1}^{3} a_i}{\max\{\sum a_i,\ \sum b_i,\ \sum c_i,\ \sum d_i\}}$$

$$S_{cluster2} = \left(1 - \frac{B}{\sqrt{2}}\right) + \frac{\sum_{i=1}^{3} b_i}{\max\{\sum a_i,\ \sum b_i,\ \sum c_i,\ \sum d_i\}} \quad \textbf{(15)}$$

$$S_{Noise1} = \left(1 - \frac{C}{\sqrt{2}}\right) + \frac{\sum_{i=1}^{3} c_i}{\max\{\sum a_i,\ \sum b_i,\ \sum c_i,\ \sum d_i\}}$$

$$S_{Noise2} = \left(1 - \frac{D}{\sqrt{2}}\right) + \frac{\sum_{i=1}^{3} d_i}{\max\{\sum a_i,\ \sum b_i,\ \sum c_i,\ \sum d_i\}}$$

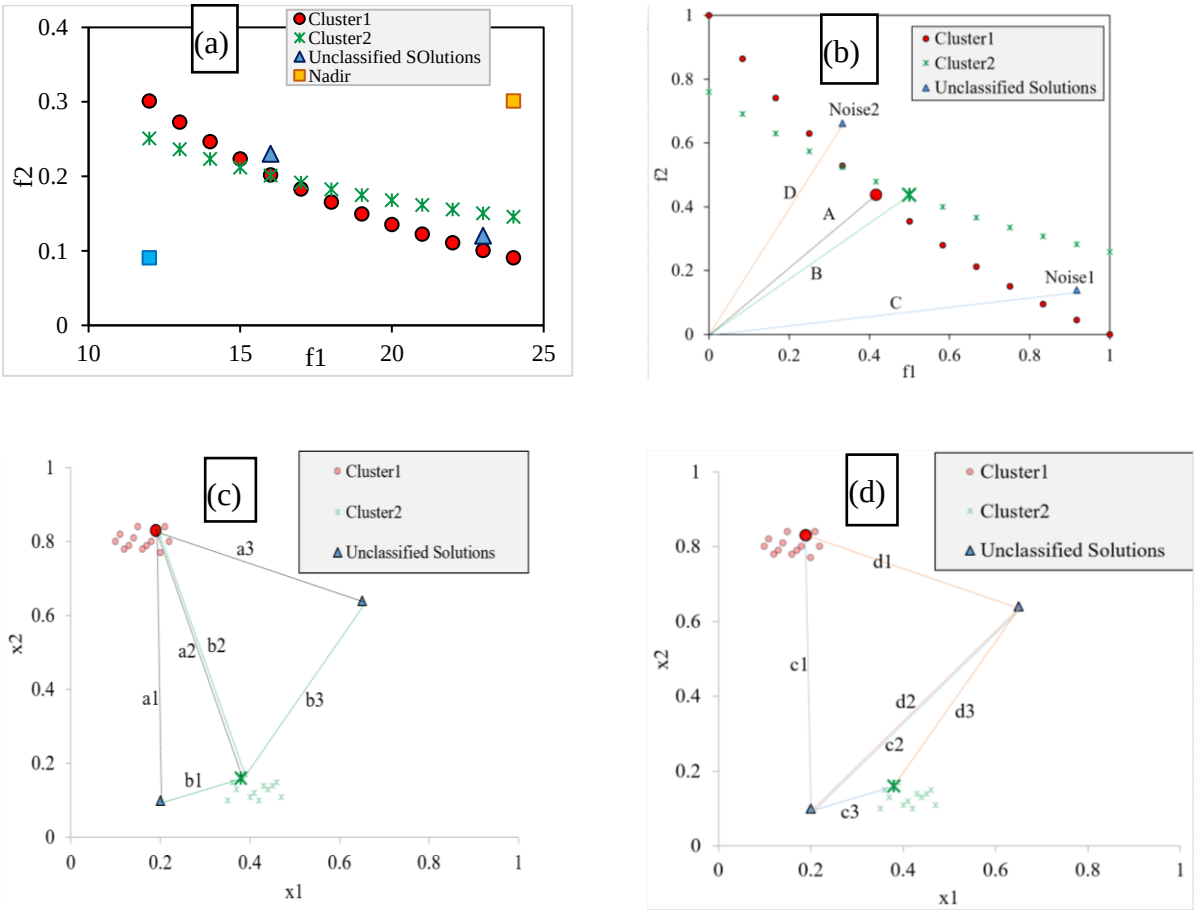


Figure 3-5: Illustration of selection metric calculation process is a particular step of optimization for a hypothetical problem with two dimensional decision and objective spaces. Panels (a) and (b) show solutions in the original and normalized objective space. Panels (c) and (d) show solutions in the normalized decision space.

3.4.4 Omni-Optimizer Algorithm

Omni-optimizer is a population-based evolutionary optimization algorithm with a bounded archive developed by Deb and Tiwari (2008). The structure of omni-optimizer is similar to that of NSGA-II algorithm with additional operators to help the algorithm solve multi-objective, multi-modal problems. It uses a hypercube sampling to generate the initial population and constructs a bigger set by two random ordering of the current population in each iteration. It chooses four solutions from the bigger set using the nearest neighbor based method in the objective space, for encouraging convergence, to determine two parent solutions using binary tournament selection. The parent solutions are then recombined and mutated by simulated binary crossover and polynomial mutation for generating offspring solutions. The parent and offspring solutions are then combined and ranked based on non-dominated sorting and crowding distance. The omni-optimizer favors non-dominated over dominated solutions, and less-crowded solutions to more crowded solutions. An important feature of the omni-optimizer is that in multi-modal situation where two solutions have identical objective vectors, the crowding distance metric is calculated in the decision space instead of the objective space. The latter becomes zero while the former has a non-zero value.

Omni-optimizer was not designed to identify and maintain near-optimal solutions; however, it is used in this study as a reference algorithm for results comparison since it considers crowding distance calculation in the decision space among solutions with identical objective vectors. Readers are referred to Deb and Tiwari (2008) for more detail about omni-optimizer's structure and its performance.

3.4.5 Optimization Problems

The cluster-based PA-DDS algorithm is compared to the original version of PA-DDS for solving two bi-objective mathematical test problems, a bi-objective sorptive barrier design problem, and a lake pollution control problem introduced in this section for the identification of distinct optimal and near-optimal solutions.

3.4.5.1 Modified SYM-PART Test Problem

The SYM-PART bi-objective, bi-variable problem was first introduced by Rudolph et al. (2007) as a multi-modal test problem. Schütze et al. (2011) modified SYM-PART to make it have one Pareto optimal front and eight near-optimal fronts, shown in problem formulation (16) where x_1 and x_2 are real-valued decision variables ranging in [-8, 8], f_1 and f_2 are the objective functions, and t_1 and δ_t are auxiliary variables.

$$f_1(\boldsymbol{x}) = (x_1 - t_1(c+2) + a)^2 + (x_2 - t_2 b)^2 + \delta_t$$

$$f_1(\boldsymbol{x}) = (x_1 - t_1(c+2a) - a)^2 + (x_2 - t_2 b)^2 + \delta_t$$

$$t_1 = sgn(x_1) min\left(\left\lceil \frac{|x_1| - a - 0.5c}{2a + c} \right\rceil, 1\right) \quad \textbf{(16)}$$

$$t_2 = sgn(x_2) min\left(\left\lceil \frac{|x_2| - 0.5b}{b} \right\rceil, 1\right)$$

$$\delta_t = \begin{cases} 0 & if\ t_i = 0, \quad i = 1,2 \\ 0.1 & Otherwise \end{cases}$$

As shown in Figure 3-6, a, b, and c are constant and if they are respectively set to 0.5, 5, and 5, the global (continuous red line) and local optimum (dashed black lines) Pareto sets and fronts are formed.

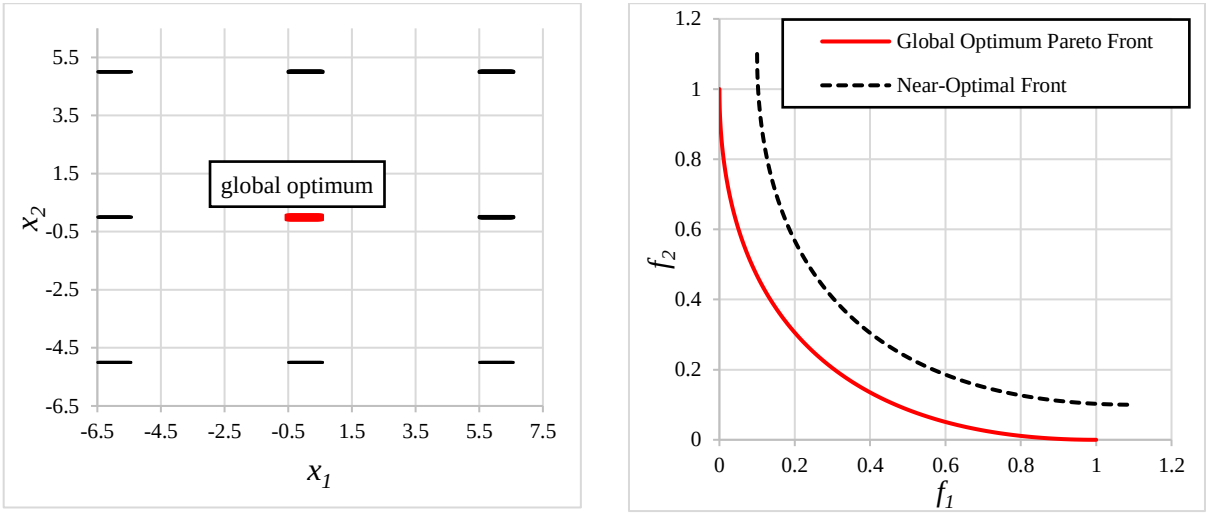


Figure 3-6: optimal and near-optimal subsets in decision (left) and objective (right) spaces for modified SYM-PART Simple problem

3.4.5.2 Modified Omni-Test

The scalable Omni-Test problem in problem formulation (17) is adopted from Deb and Tiwari (2008) with a slight modification. This minimization MO problem is slightly modified in this paper to a test with one global optimum and multiple local Pareto subsets by introducing a new constant δ. Increasing the range or the number of decision variables increases the multi-modality of the problem (Deb and Tiwari, 2008). Three decision variables with a range between 0 and 4.9 are defined in this paper that contains eight near-optimal subsets.

$$f_1(\boldsymbol{x}) = \sum_{i=1}^{3} \sin(\pi x_i) + \delta \qquad \textbf{(17)}$$

$$f_2(\boldsymbol{x}) = \sum_{i=1}^{3} \cos(\pi x_i) + \delta$$

$$\delta = \begin{cases} 0 & if \ \ x_i < 2 \\ 0.1 & Otherwise \end{cases}$$

The global and local Pareto sets occur where decision variables are between 1 and 1.5 or between 3 and 3.5. If all decision variables range from 1 to 1.5, the Pareto optimal front (continuous red line in Figure 3-7) is created and the local minimal front (dashed black line in Figure 3-7) is produced with other combinations of the mentioned extremum intervals.

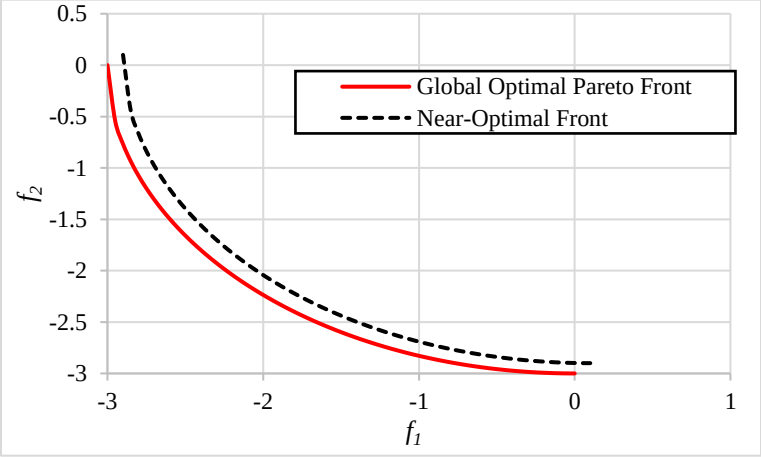


Figure 3-7: Illustration of Pareto optimal and near-optimal fronts for modified Omni-Test function

3.4.5.3 Lake Pollution Control Problem

Ward et al. (2015) developed a scalable four-objective benchmark optimization problem from a modelling study conducted by Carpenter et al. (1999) for the management of eutrophication of a shallow lake. This problem, also known as the lake Problem, aims to maximize the economic profits of a town by finding the amount of yearly anthropogenic phosphorous release (a_t) while maintaining the reliability and control policy inertia. The dimensionless total mass concentration of phosphorous (P_t) at annual time step *t* is calculated using Equation (18). Besides the annual

phosphorous release from the town, P_t in each time step also depends on P_{t-1} and the uncertain non-point natural sources of pollution flowing into the lake (ε_t) that is emulated by a random number sampled from a log-Normal distribution with a mean of 0.02 and a log10 variance of -5.5 as reported in Ward et al. (2015). This problem is a Monte Carlo simulation-based function evaluation due to the uncertainty in the uncontrolled pollution term, ε_t. Unlike its original variant in Ward et al. (2015) that considered 100-year pollution management, twenty decision variables (20-year management policy, $a_1, a_2, \dots, a_{20}$) varying from 0 to 0.1 (dimensionless) are considered in this problem. There are two parameters, b and q, that are associated with phosphorous recycling and decaying rates in the lake. These parameters are respectively fixed to 0.42 and 2 (Ward et al., 2015) to impose an irreversible eutrophic state on the lake if P_t exceeds a pre-defined threshold, P_{crtcl} that is a function of b and q.

$$P_t = P_{t-1} + a_t - bP_{t-1} + \frac{P_{t-1}^q}{1 + P_{t-1}^q} + \varepsilon_t \quad \textbf{(18)}$$

$$f_1(\boldsymbol{a}) = \frac{1}{k}\sum_{i=1}^{k}\sum_{t=0}^{T-1}\left(\alpha a_{t,i}\delta^t\right), \quad T = 20, \quad k = 100 \quad \textbf{(19)}$$

As shown in Equation (19), the first objective is to maximize the average economic benefit across *k* simulations of *T* years of random ε_t. As in Ward et al. (2015), α and δ are dimensionless parameters that are set to 0.4 and 0.98 respectively representing the town's desire to pay for pollution control and the discount factor to convert future profits to present utilities. The second objective, Equation (20), is to minimize the highest total phosphorous concentration (P_t) averaged across *k* simulations. Stability in anthropogenic phosphorous rate over time is another important criterion that needs to be taken into consideration, since rapid reduction in phosphorous in the lake

requires large infrastructural investments, and it is best to preserve policy inertia. For this reason, the difference between two consecutive release rates should be less than a pre-specified threshold, $I_{crtcl} = 0.02$ as in Ward et al. (2015), in Equation (21). The second term in Equation (21) is designed to find a reliable management policy to prevent from permanent eutrophication in the lake by keeping P_t below a critical value, i.e. irreversible threshold $P_{crtcl} = 0.5$.

$$f_2(\boldsymbol{a}) = \max_{t\in\{1,2,\ldots,T\}} \frac{1}{k}\sum_{i=1}^{k} P_{t,i} \tag{20}$$

$$f_3(\boldsymbol{a}) = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{1}{T-1}\sum_{t=1}^{T-1}\theta_{t,i}\right) + \left(1 - \frac{1}{kT}\sum_{i=1}^{k}\sum_{t=1}^{T}\vartheta_{t,i}\right)$$

$$\theta_{t,i} = \begin{cases} 0 & if\ a_{t-1,i} - a_{t,i} \le I_{crtcl} \\ 1 & if\ a_{t-1,i} - a_{t,i} > I_{crtcl} \end{cases} \tag{21}$$

$$\vartheta_{t,i} = \begin{cases} 0 & if \quad P_{t,i} \ge P_{crtcl} \\ 1 & if \quad P_{t,i} < P_{crtcl} \end{cases}$$

3.4.5.4 Sorptive Barrier Design

The sorptive barrier design problem introduced by Bartelt-Hunt et al. (2006) is a combinatorial simulation-optimization problem that seeks the cheapest option(s) for waste management while mitigating the migration of contaminants from organic wastes through a multi-layer sorptive liner. The landfill liner design is converted to a single-objective constrained optimization benchmark problem by Matott et al. (2012) consisting of six integer-valued decision variables. Thirteen alternative *15 cm* layers with coded values from one to thirteen are available for each decision variable that is made of variable mixture of sand, bentonite, benzyltriethylammonium-bentonite,

hexadecyltrimethylammonium-bentonite, shale, and granular activated carbon. The fifth and sixth decision variables can take a fourteenth option, which is a no-layer option, to allow for a design with variable number of layers (l). The first objective function, f_1 in problem formulation (22), is the design cost that has to be minimized ($\$/m^2$). If the number of layers is higher than four, the opportunity cost ($cost_1$) will be added to the material cost ($cost_2$). This problem also uses a one-dimensional numerical model to simulate the cumulative amount of 1,2-dichlorobenzene (1,2-DCB) contaminants $A(\boldsymbol{x})$ infiltrated into the ground from the liner over time. The second objective function used in this study, in problem formulation (22) that was a constraint in Matott et al. (2012) is to keep cumulative amount of contaminants below a pre-defined allowable amount of 5.0 $\mu g/m^2$ over 100 years (design lifetime). Readers are referred to Matott et al. (2012) regarding layer compositions and the associated costs.

$$f_1(\boldsymbol{x}) = cost_1(l) + cost_2(\boldsymbol{x})$$

$$cost_1(l) = 5.625(l - 4)$$

$$cost_2(\boldsymbol{x}) = \sum_{i=1}^{6} layer\ cost\ (x_i) \quad \textbf{(22)}$$

$$f_2(\boldsymbol{x}) = A(\boldsymbol{x}) - 5$$

3.4.6 Numerical Experiment Setup and Results Comparison Approach

PA-DDS is a stochastic MO optimization algorithm, in that its solution differs in different trials. Therefore, in order to compare the distribution and proximity of the original PA-DDS algorithm with hypervolume contribution metric to its cluster-based variant introduced in this study, ten independent trial runs are conducted on each of the MO optimization problems. The number of

function evaluations is different for each type of problem mentioned in Table 3-1 depending on the computational complexity and search/objective space dimensionality of the problem. Moreover, determining the value of minimum adjacency density is highly dependent on the available computational budget, and the results will be highly different for different values of ε. Due to their simple structure, the value of ε is set to 0.09 for the benchmark test problems, which is equivalent to 5 to 6.5 percent discrepancy for each decision variable. It is assumed if the difference between two solutions in the lake problem is at least 15 percent for each of their decision variables, they are two dissimilar options and the value of ε is equal to 0.67 for a decision space normalized to between zero and one. The decision space in the sorptive barrier design problem is integer, and if the values of four out of six decision variables are one step higher or lower between two solutions, they are considered as distinct design options, resulting in ε equal to 0.16 when the design space is scaled to between zero and one for each decision variable.

The omni-optimizer algorithm is compared with the cluster-based PA-DDS in terms of decision-space diversity and near-optimal solution preservation. The same computational budget and post-processing procedure is used for both of these optimization algorithms. The omni-optimizer, however, has multiple parameters that influence its performance. They are the population size, number of generation, probability of crossover, probability of mutation, distribution indices for crossover and mutation that can take different values for different optimization problems. Following Deb and Tiwari (2008), the crossover probability and index for mutation are respectively set on 0.9 and 20 for all optimization cases. The distribution index for crossover is subjectively set to 15 for the SYM-PART and sorptive barrier design problem and 10 for the modified Omni-Test and lake problems. A higher distribution index value aids in escaping local optima and producing an offspring that is far away from its parent while a lower value helps

fine-tuning and convergence (Deb and Beyer, 2001). The mutation probability is set to *1/n* where *n* is the number of decision variables based on the recommendations by Deb and Tiwari (2008). The population size and number of generations are shown in Table 3-2 based on the set computational budget in this study.

The resulted archive in the case of the modified SYM-PART and Omni-Test problems are analyzed and processed based on the closeness to their mathematically known global and local optimal regions. The optimization results for the modified SYM-PART and Omni-Test problems are visualized and compared in their 2D and 3D decision spaces.

Unlike the mathematical test problems, there is no prior information about the global or local optimal solution sets in the lake problem and the sorptive barrier design problem. Therefore, a three-stage post-processing analysis is performed on their archives to identify the desirable solution(s) based on decision-makers' preference. The first stage is to (re-)cluster archived solutions using DBSCAN in the decision space to distinguish similar solutions from dissimilar solutions based on their cluster labels. The same adjacency density parametrizations in Table 3-1 are considered for (re-)clustering the archives in each type of problem. The second stage is to identify and retain dissimilar solutions that include the unclassified archived solutions and only one solution in each cluster, which is closest to a reference point in the objective space among its groupmates. If there are also constraints, the closest solution to the reference point must meet the constraints. The reference point can be an ideal objective vector (Utopia) or any desirable values of objectives defined by a decision-maker. The third stage is to define an acceptable threshold for each objective function and identify desirable solutions among dissimilar solutions in the second stage that have better objective values than the thresholds. In the end, the performance of two versions of PA-DDS algorithms are compared based on the solutions that passed the screening

process in the third stage by showing their decision variable vectors on a parallel coordinate plot and a dissimilarity index shown in Equation (23).

$$I = \frac{size\left(screen(\sum_{i=1}^{10} archive(i))\right)}{\sum_{i=1}^{10} size(screen(archive\ (i))} \quad \textbf{(23)}$$

In order to calculate the dissimilarity index in Equation (23), the archives of all optimization trial runs are combined and the three-stage screening process is applied to the combined archive. The dissimilarity index value is equal to the number of screened dissimilar solutions after combining the archive sets of all individual trials divided by sum of number of dissimilar screened solutions for each trial. The higher the value of dissimilarity index, the better the performance of the optimization algorithm in terms of identifying distinct desirable options.

Table 3-1: The numerical experiment specifications for each type of optimization problem considering their computational complexity

Problem	Number of Decision Variables/Objective Functions	# of Evaluations	Average Computational Time per Evaluation (sec)	Minimum Adjacency Density ($MinPts/\ \varepsilon$)	Reference Point
Modified SYM-PART	2/2	2000	0.02	4/0.09	-
Modified Omni-test	3/2	5000	0.07	4/0.09	-
Sorptive Barrier Design	6/2	3000	22	4/0.16	(0, log10(5e-6))
Lake Problem	20/3	20000	1.73	4/0.67	(0, 0.784, 1, 1)

Table 3-2: The population size and number of generations for the omni-optimizer algorithm applied to each optimization problem of this study

Problem	Population Size	# of Generation
Modified SYM-PART	40	50
Modified Omni-Test	100	50
Sorptive Barrier Design	60	50
Lake Problem	100	200

3.5 Results and Discussion

3.5.1 Mathematical Test Functions

The cluster-based PA-DDS identifies all optimal and near-optimal regions in each trial in the modified SYM-PART and Omni-Test problems while the original PA-DDS and omni-optimizer algorithms only find the optimal Pareto front in each trial, and they are unable to detect distinct near-optimal solution sets. Figure 3-8 and Figure 3-9 demonstrate the results for one trial respectively for the modified SYM-PART and Omni-Test problems and the cluster-based PA-DDS is able to find and assign a unique cluster tag to each locally optimal region. The archive set also contains other low-quality clusters, each of which form a unique tradeoff, along with unclassified solutions in other regions of decision space but they are filtered out after post-processing based on the information we have about the location of the near-optimal and optimal regions in the decision space for these test problems. The post-processed results for other trial runs are similar to Figure 3-8 and Figure 3-9 but with different cluster tags.

The original PA-DDS concentrates on global convergence and uniform distribution of non-dominated solutions in the objective space. Figure 3-8 and Figure 3-9 show that a large number of solutions are produced and retained by the original PA-DDS that are well-diversified along the optimal Pareto front with no solution representing the near-optimal front. This does not mean the PA-DDS algorithm does not explore the near-optimal regions in the decision space. It may find and preserve distant near-optimal solutions in the archive set. However, since a global dominance check is carried out based on the objective values in the traditional solution archiving, they are eliminated from the archive when a new dominating solution is produced. Therefore, the original PA-DDS does not consider the decision-space diversity in the optimization process and therefore does not guarantee maintenance of distinct near-optimal solutions.

The result of the omni-optimizer algorithm is similar to that of original PA-DDS with the difference that the omni-optimizer has a bounded archive, i.e. 40 and 100 solutions respectively for the SYM-PART and Omni-Test Problems. In light of the fact that the multi-modality of the mathematical test problems are modified to a global and multi-local optimal regions, the decision space crowding measure in the omni-optimizer cannot help retain solutions representing local optima. In fact, that the decision-space crowding measure is activated when there are at least two solutions with identical objective vectors, that never happened in problems solved in this study.

On the contrary, most of the searching power in the cluster-based PA-DDS is used for the identification of local optimal regions and dissimilarity maintenance, but distribution of solutions along the optimal or near-optimal fronts in the objective space is not in priority. Performing a local gradient-based strategy on the desirable clusters or solutions in the archive of the cluster-based PA-DDS after the termination of optimization may help to sufficiently populate each cluster for having a uniform distribution in the objective space.

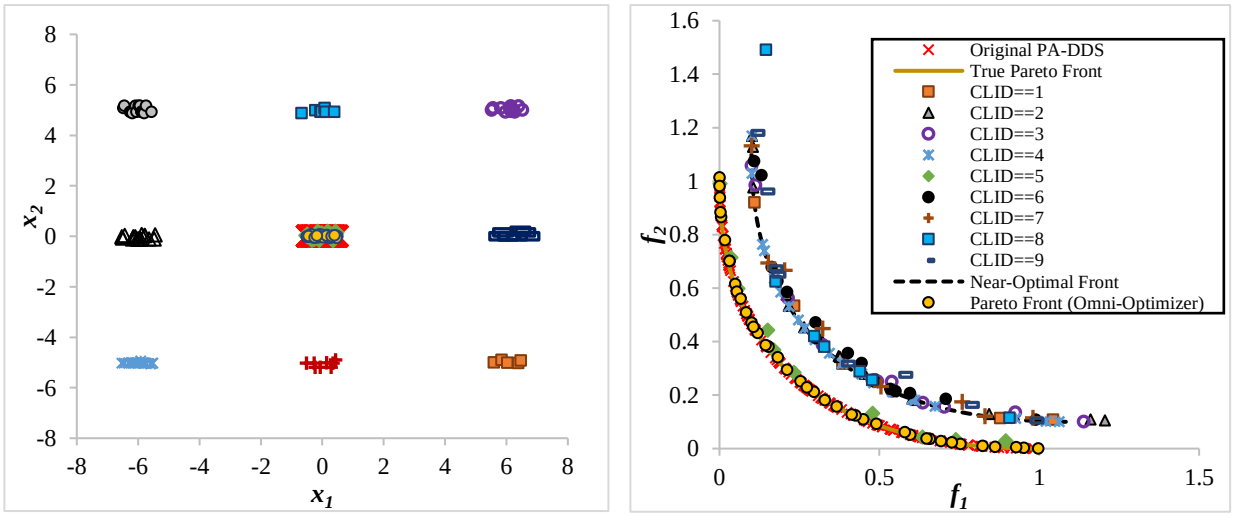


Figure 3-8: Distribution of the post-processed archived solutions found by cluster-based PA-DDS for solving modified SYM-PART test problem compared to the original PA-DDS and the omni-optimizer algorithms in the decision (left) and objective space (right). CLID shows the cluster tag.

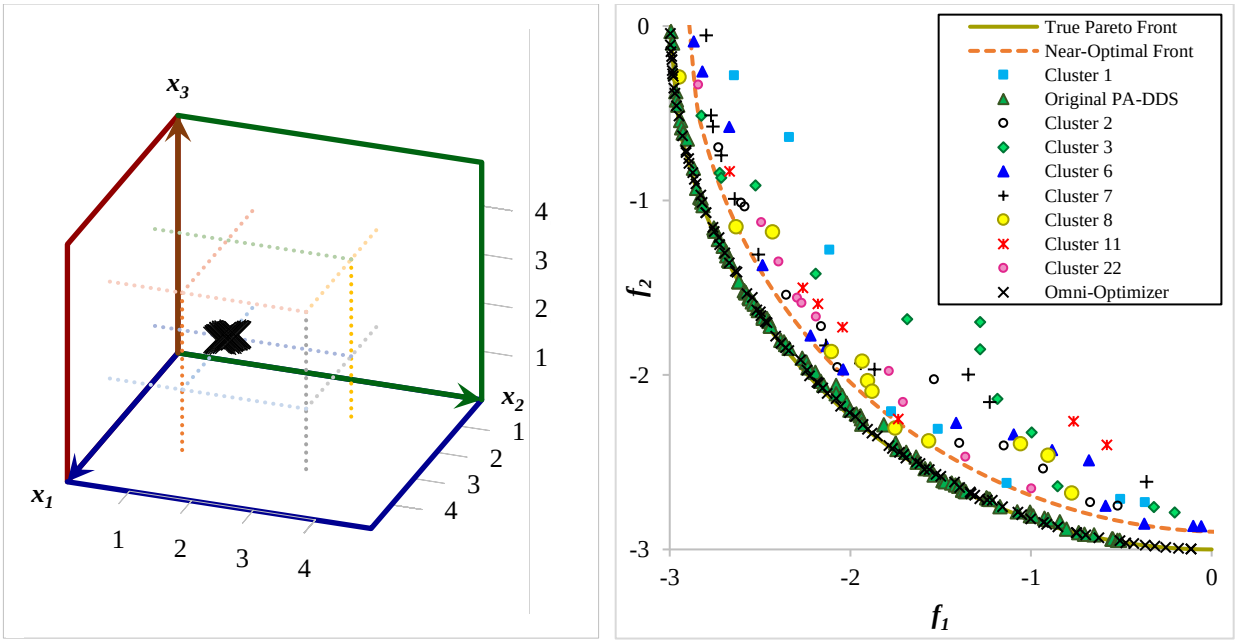


Figure 3-9: Distribution of the screened archived solutions found by cluster-based PA-DDS for solving modified Omni-Test problem compared to the original PA-DDS and omni-optimizer algorithms in the decision (left) and objective space (right).

3.5.2 Pollution Control Problem

Table 3-3 presents the performance of two variants of the PA-DDS algorithm along with the omni-optimizer algorithm in terms of distinction level of their archives for the lake problem. The archive set for the original PA-DDS algorithm has a very small size ranging from 36 to 87 solutions out of 20000 generated solutions compared to the cluster-based version and it has a constant size of 100 in the omni-optimizer algorithm for each optimization trial. The size of archive in the cluster-based PA-DDS is close to 3000 solutions.

Seeking a desirable solution in an archive by comparing the values of decision variables and the corresponding objectives is laborious if not impossible. As a result, a systematic post-processing strategy is utilized in this study to further screen the archived solutions based on a series of subjective judgements that often require consulting with decision-makers. The first screening stage is to re-cluster archived solutions using DBSCAN with the same adjacency density

parameters used by the cluster-based PA-DDS. The nearest solution to the reference point in each cluster is maintained along with unclassified solutions and the rest of the archived solutions are filtered out. In the second screening stage, a subjective threshold is set for each objective function. This threshold should be defined based on the decision-makers' desirable range for each objective function. For example, the lake problem is assumed to have an irreversible eutrophic condition, meaning that the reliability index should be 1.0; otherwise the corresponding 20-year pollution management policy leads to a permanent eutrophic state in the lake with no possible water quality restoration by solely reducing phosphorous loading. Therefore, solutions that have a reliability index of less than 1.0 are removed from the archive. The numbers highlighted by the bold font in columns five to seven of Table 3-3 show the number of dissimilar design options (solutions) that do not tip into irreversible polluted state according to the adjacency density radius in Table 3-1 for the lake problem. The higher number of archived solutions by the cluster-based PA-DDS suggests that each of its trial identified a significantly larger number of distinct design options that are considered reliable in the lake problem.

In the third screening stage, it is assumed that only design options (solutions) that score more than 85% inertia index require acceptable infrastructural investments for phosphorous reduction in the lake. Therefore, solutions not meeting this screening criterion are filtered out and the archive size is reduced to 2 to 8 solutions for trials of the original PA-DDS and 1 to 4 solutions for the cluster-based PA-DDS depending on the trial number, which is displayed inside parenthesis in Table 3-3. It is interesting that only a few distinct archived solutions with reliability value of 1.0 are identified by the original PA-DDS (3 to 10 solutions), and the majority or all of them have inertial maintenance index of higher than 85%. Further analysis of the results shows that solutions identified by all trials of the original PA-DDS are highly similar. This means that most of the

computational budget is consumed by all trials of PA-DDS to converge to a common location in the decision space.

Despite the fact that the original PA-DDS is equipped with an unbounded archive that is expected to archive a high number of solutions, the omni-optimizer has a higher number of archived solutions and offers higher number of solutions with reliability index of one in each optimization trial. The ratio of solutions with inertia index higher than 85 percent among reliable solutions is lower for the omni-optimizer compared to that for the original PA-DDS in all trials, for example 12/28 versus 7/7 in the first trial, see Table 3-3. This is due to considering decision-space crowding distance in the structure of omni-optimizer. This ratio is the lowest for the cluster-based PA-DDS in all trial runs since the algorithm does not solely focus on a specific region in the search space.

The archive sets of all trials for the omni-optimizer algorithm and each variant of PA-DDS are combined, resulting in 1000 solutions for the omni-optimizer, 625 solutions for the original PA-DDS, and 29808 solutions for the cluster-based PA-DDS. The aforementioned three-stage screening procedure is applied to these three archive sets, separately. Only five reliable, dissimilar solutions are identified by original PA-DDS and only three of these solutions have policy inertia index of higher than 85%. Therefore, most of the high-quality solutions identified by different trials of the original PA-DDS are similar and only 9% percent of them are dissimilar, giving different design options. This number is 149 reliable, dissimilar solutions for the omni-optimizer algorithm, 32 of which having inertia index of higher than 85%. As a result, the dissimilarity index of the reliable solutions in the omni-optimizer algorithm is 60%, considerably higher than the original PA-DDS. On the other hand, the cluster-based PA-DDS identified 1358 dissimilar solutions with reliability of 1.0, 23 of which have policy inertia higher than 85%. These 23

solutions are the union of solutions identified by all trials of the cluster-based PA-DDS and none of them have similar cluster tags. Interestingly, the dissimilarity index increases from 98% to 100% with fine-filtering the archived solutions for the proposed PA-DDS structure, while it decreases from 60% and 9% to 45% and 6% respectively for the omni-optimizer and the original PA-DDS algorithms. Therefore, it is concluded that the proposed solutions archiving helps the optimization algorithm identify dissimilar solutions.

Identifying higher number of dissimilar solutions after combining the archives of the cluster-based PA-DDS gives a higher flexibility to decision-makers to choose among higher number of solutions with a wider range of benefit (f_2) and maximum annual pollution (f_1) compared to the solutions identified by the conventional algorithm. The range of the first and the second objectives for the omni-optimizer algorithm is wider than the proposed PA-DDS structure among reliable solutions with inertia maintenance index of higher than 85%. This is mainly due to the structure of the omni-optimizer that is a population-based algorithm while PA-DDS is a single-solution-based method. For a fair comparison about the range of the objectives, the omni-optimizer needs to be compared to its cluster-based variant with an adaptive archive size, which is outside the scope of this paper. Figure 3-10 demonstrates the parallel coordinate plot of the remaining solutions ($f_4=1.0$, and $f_3>0.85$) after screening the combined archives for the omni-optimizer, original, and cluster-based PA-DDS, respectively. The cluster-based PA-DDS provides different values of annual phosphorous release from the town that are scattered all over their defined range, while if the decision-maker relies on the original PA-DDS, they are limited to options with annual release values in the lower half of their range, especially from the fourth year to the sixteenth year. Similarly, the majority of the post-processed solutions in the omni-optimizer are clustered in the lower half of the range in a sub-space containing decision variables from four to sixteen. Therefore,

the original PA-DDS discards many solutions that could be interesting for the decision maker, because those solutions are dominated but near-optimal in the objective space.

Table 3-3: Archive size and number of dissimilar solutions for the original and cluster-based PA-DDS after performing DBSCAN and further screening based on the reliability index and inertial maintenance in each optimization trial and the combined archive of all trials.

Optimization Trial	Archive Size			Number of Distinct Archived Solutions with f_4 = **1.0** (& f_3 >0.85) [Range of f_1]/[Range of f_2] when f_4=1.0 and f_3>0.85		
	Omni-Optimizer	Original PA-DDS	Cluster-Based PA-DDS	Omni-Optimizer	Original PA-DDS	Cluster-Based PA-DDS
1	100	36	2977	**28** (12) [0.17, 0.45]/ [0.14, 0.29]	**7** (7) [0.22, 0.43]/ [0.24, 0.30]	**111** (3) [0.29, 0.43]/ [0.23, 0.26]
2	100	46	3017	**27** (6) [0.20, 0.42]/ [0.21, 0.29]	**3** (3) [0.25, 0.35]/ [0.25, 0.28]	**142** (1) 0.27/0.23
3	100	59	3171	**30** (8) [0.13, 0.46]/ [0.12, 0.30]	**3** (3) [0.15, 0.27]/ [0.15, 0.26]	**128** (1) 0.26/0.23
4	100	77	3020	**30** (12) []0.13, 0.44]/ [0.12, 0.29]	**6** (3) [0.21, 0.32]/ [0.23, 0.27]	**133** (4) [0.28, 0.37]/ [0.22, 0.28]
5	100	42	2871	**28** (6) [0.14, 0.34]/ [0.14, 0.25]	**10** (8) [0.14, 0.43]/ [0.16, 0.30]	**164** (3) [0.35, 0.40]/ [0.23, 0.26]
6	100	87	3051	**27** (6) [0.18, 0.45]/ [0.15, 0.29]	**4** (4) [0.25, 0.40]/ [0.23, 0.29]	**136** (1) 0.38/0.23
7	100	76	2931	**15** (6) [0.24, 0.44]/ [0.24, 0.30]	**6** (6) [0.19, 0.43]/ [0.21, 0.30]	**149** (4) [0.25, 0.36]/ [0.19, 0.25]
8	100	64	2942	**15** (3) [0.22, 0.41]/ [0.21, 0.29]	**10** (10) [0.14, 0.43]/ [0.16, 0.30]	**146** (2) [0.28, 0.41]/ [0.22, 0.25]
9	100	64	2933	**21** (5) [0.32, 0.47]/ [0.24, 0.30]	**3** (2) [0.26, 0.35]/ [0.25, 0.28]	**130** (2) [0.31, 0.34]/ [0.21, 0.22]
10	100	74	2895	**26** (7) [0.14, 0.36]/ [0.13, 0.27]	**4** (4) [0.16, 0.31]/ [0.18, 0.27]	**147** (2) [0.33, 0.47]/ [0.24, 0.24]
All Trials Combined	1000	625	29808	**149** (32) [0.12, 0.43]/ [0.1, 0.28]	**5** (3) [0.24, 0.39]/ [0.25, 0.29]	**1358** (23) [0.25, 0.47]/ [0.19, 0.28]
Dissimilarity Index Value	-	-	-	**149/247=0.60** (32/71=0.45)	**5/56=0.09** (3/50=0.06)	**1358/1386=0.98** (23/23=1.00)

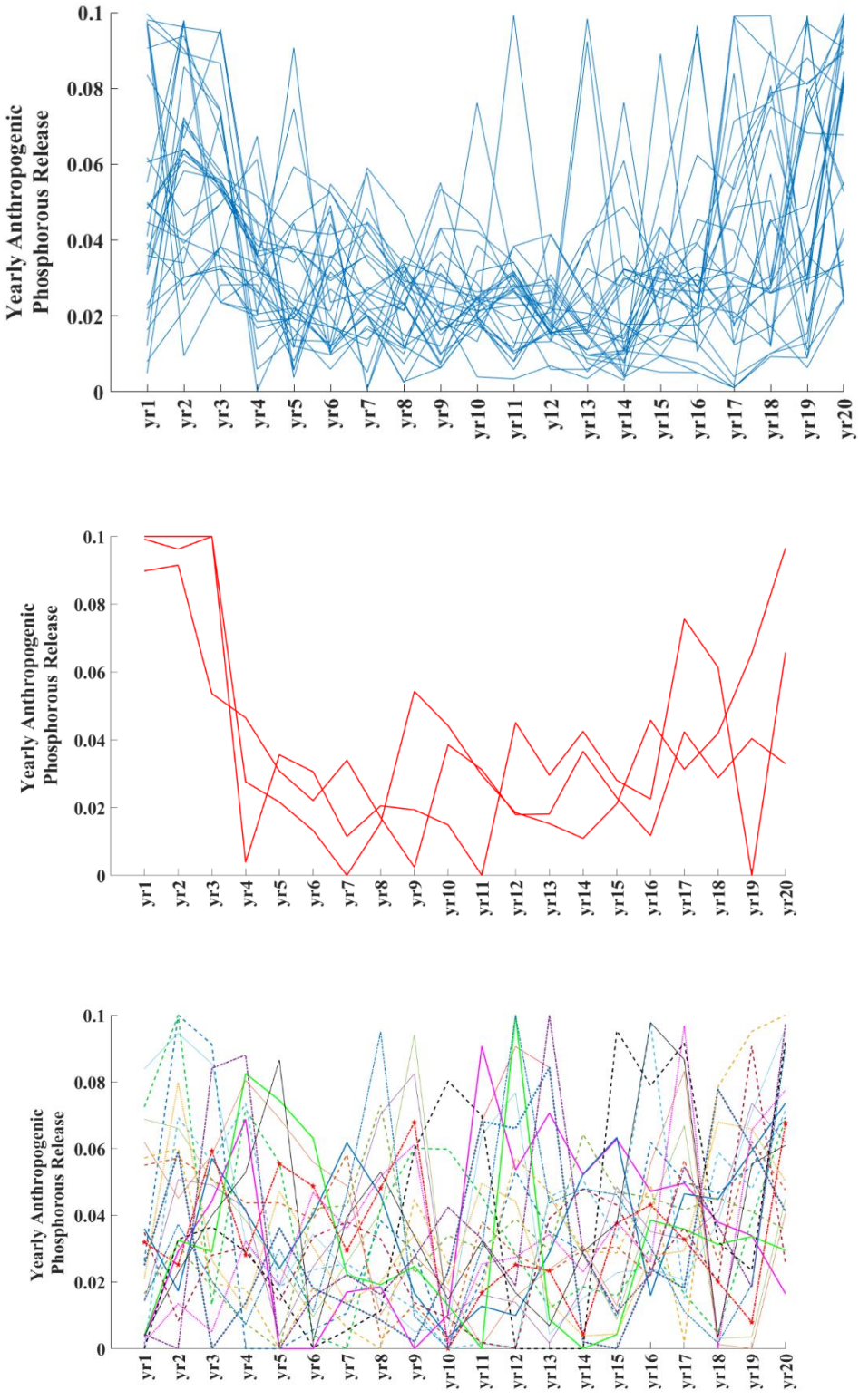


Figure 3-10: Parallel coordinate plot of decision variables for solutions with the reliability index of equal to one and inertial maintenance index of higher than 0.85 after combining the archives of ten optimization trial runs for the omni-optimizer (top), original PA-DDS (middle) and the cluster-based PA-DDS (bottom)

3.5.3 Sorptive Barrier Design Problem

A similar filtering process to the lake Problem is conducted for the results of the sorptive barrier design problem. The first stage is to employ DBSCAN for clustering and re-clustering archived solutions of the omni-optimizer algorithm and PA-DDS with the traditional and cluster-based archiving. The second stage is to screen the solutions based on their Euclidean distance to the reference point (see Table 3-1) in a semi-logarithmic objective space for this problem, which is an ideal zero cost and allowable contaminant rate transported to the soil in the barrier design lifetime. The last three columns of Table 3-4 shows the number of dissimilar solutions closest to the reference point in the objective space that have a cost (f_1) of less than 50 and satisfy the contaminant rate (f_2) of 5.0 $\mu g/m^2$ for each optimization trial and the combined archive of all trials. Apparently, the original PA-DDS is able to identify multiple distinct design options after the filtering process and offer more number of distinct solutions compared to the omni-optimizer algorithm. The reason can be attributed to the discontinuity of the decision space. However, the dissimilarity level of these algorithms is still less than the cluster-based version when the archive sets are combined. These algorithms result in dissimilarity index values of 0.53 and 0.61 versus 0.76 for the cost values less than 50. Similar to the lake problem, fine-filtering of the archives by considering a lower threshold for the maximum acceptable cost increases the dissimilarity ratio for the cluster-based PA-DDS to 88%. On the contrary, the dissimilarity ratio for the omni-optimizer and original PA-DDS decreases to 32% and 51%.

The high dissimilarity index in the results of the sorptive barrier design problem compared to the lake problem has increased when the original PA-DDS is used and reduced when the omni-optimizer is used. This is mainly due to two reasons. One reason is that the archive size resulted from the original PA-DDS for each trial of this type of problem with discrete decision variables is

so small that DBSCAN identifies some solutions as noise since they do not meet the minimum adjacency density requirement for cluster formation despite having high similarity to other archived solutions, which should not be considered as dissimilar options. The second reason is that the discrete nature of this problem does not let the original PA-DDS algorithm fine-tune decision variable values to decimal places. The integer-valued decision variables also aid in small archive size that exacerbates the situation for DBSCAN method to find really dissimilar design options in the post-processing stage as it needs a sufficiently populated archive for more accurate clustering. The integer-valued decision space also causes low distinction level in the archive of the omni-optimizer since this type of problem does not have a multi-modal characteristic to make the algorithm activate its decision-space crowding measure, since there exists no solution to have an identical objective vector with another solution in the decision space.

Figure 3-11 demonstrates the joint parallel coordinate plot of the decision variable vectors and the corresponding scaled objective vectors $\left(\frac{f_1}{10}, \frac{\log(f_2)}{\log(5e-6)}\right)$ for solutions with cost of less than 50 and the contamination of less than 5.0 $\mu g/m^2$. The red lines represent solutions with cost of less than 25. According to Figure 3-11, all the solutions found by the omni-optimizer and original PA-DDS algorithms have decision variable values of equal to or higher than seven providing no option with decision variable values of one to six. However, the cluster-based PA-DDS allows decision-makers to choose from design options that have decision variable of one to six, when they are interested to include these layers for the barrier design provided they are flexible with a higher design cost. A more detailed inspection of Figure 3-11 reveals that the cluster-based PA-DDS shows a higher distinction among solutions costing less than 25 compared to its original counterpart and the omni-optimizer algorithm. For instance, the original PA-DDS cannot identify alternate designs having option No. 13 for the fifth layer. Figure 3-11 also illustrates that, among

less costly solutions, the proposed PA-DDS structure offers five options for the next layer if layer No. 13 is used for a layer, while the omni-optimizer offers one to maximum three options for the next layer.

Table 3-4: Archive size and number of dissimilar solutions for the original and cluster-based PA-DDS after performing DBSCAN and further screening based on the cost (f_1) and contaminant rate (f_2) in each optimization trial and the combined archive of all trials.

Optimization Run	Archive Size			Number of Distinct Archived Solutions w/ f_1<50 & f_2< 5.0e-6 (**f_1<25 &f_2<5.0e-6**)		
	Omni-Optimizer	Original PA-DDS	Cluster-Based PA-DDS	Omni-Optimizer	Original PA-DDS	Cluster-Based PA-DDS
1	60	47	2533	20 (**9**)	21 (**11**)	62 (**5**)
2	60	38	2588	14 (**9**)	19 (**11**)	88 (**10**)
3	60	38	2595	18 (**11**)	16 (**6**)	80 (**9**)
4	60	33	2602	16 (**10**)	13 (**5**)	104 (**11**)
5	60	36	2563	16 (**10**)	15 (**7**)	107 (**13**)
6	60	41	2548	13 (**8**)	22 (**13**)	68 (**3**)
7	60	32	2546	21 (**12**)	15 (**7**)	113 (**13**)
8	60	43	2532	21 (**12**)	22 (**11**)	61 (**3**)
9	60	43	2530	18 (**11**)	14 (**6**)	47 (**4**)
10	60	39	2579	15 (**9**)	16 (**10**)	97 (**4**)
All Trials Combined	600	390	25608	91 (**32**)	106 (**45**)	628 (**66**)
Dissimilarity Index	-	-	-	91/172=0.53 (**32/101=0.32**) dissimilar	106/173=0.61 **(45/87= 0.51)** dissimilar	628/827=0.76 **(66/75=0.88)** dissimilar

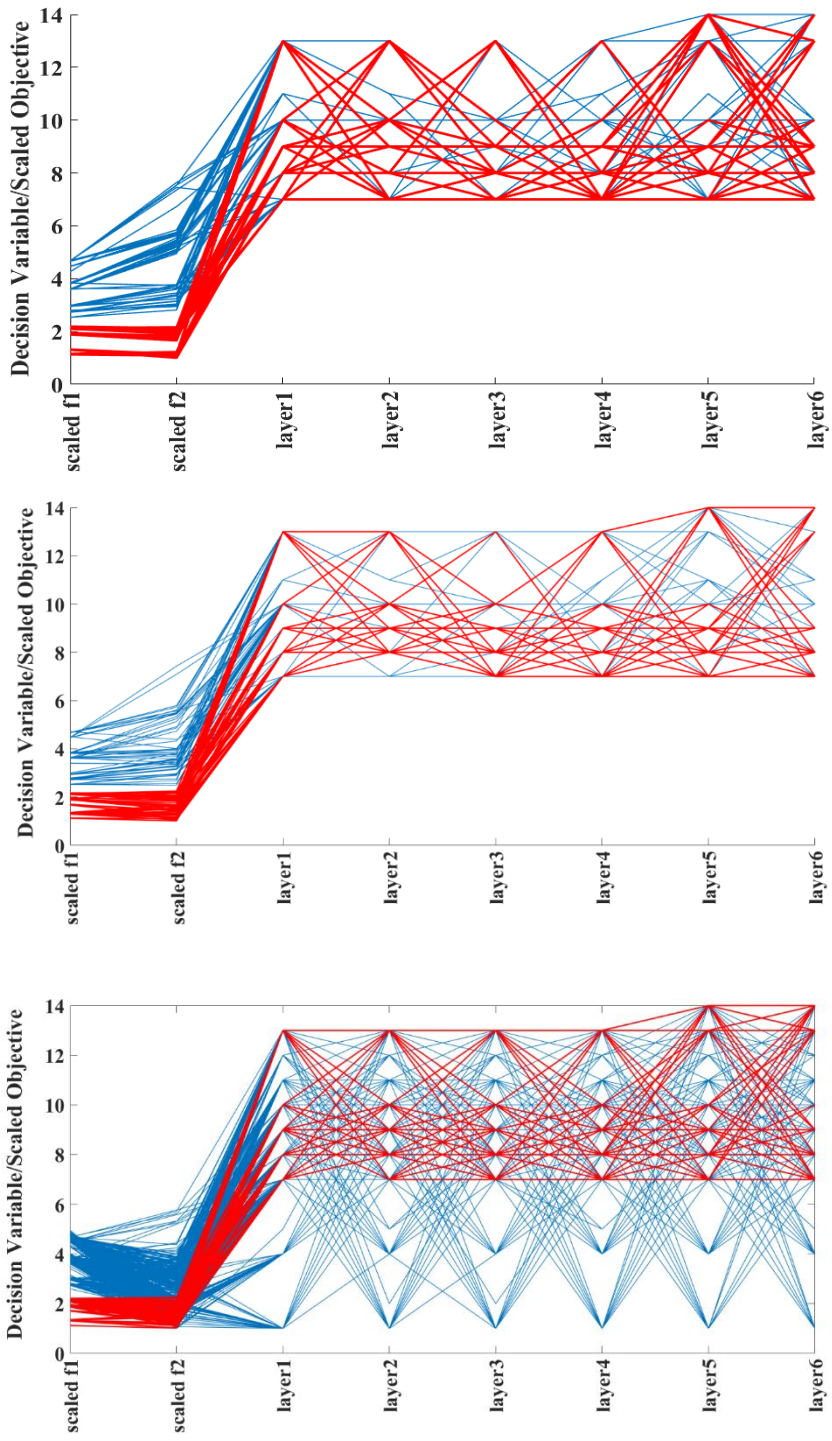


Figure 3-11: Parallel coordinate plot of decision variables for solutions with the cost of less than 50 and contaminant rate of less than 5e-6 after combining the archives of ten optimization trial runs for the omni-optimizer (top), original PA-DDS (middle), and the cluster-based PA-DDS (bottom). Red lines highlights archived solutions with cost of less than 25.

3.6 Conclusion

This work has laid out a foundation for incorporating decision-space diversity maintenance within stochastic MO algorithms. A novel cluster-based solution archiving approach is introduced by restructuring the DBSCAN clustering strategy to provide multiple distinct optimal and near-optimal options to decision-makers for a robust decision-making. Similar solutions are clustered in the decision space if they meet the pre-defined solution density requirement for initial cluster formation. The dominance-check is decentralized such that solutions are compared only within their own cluster. Dominated clustered solutions are discarded from the archive after a mutual comparison with their groupmates. A selection metric is also devised that gives higher chance of selection to distant clusters with higher normalized hypervolume. This solution archiving method adds two parameters to the MO algorithms that represent the minimum adjacency density for cluster formation and recommendations has been given with regard to pre-setting their values.

The proposed cluster-based PA-DDS was successfully applied to two bi-objective mathematical tests, a three-objective lake pollution control, and a bi-objective integer-valued barrier design problems and compared its performance to the original version and the omni-optimizer as a reference algorithm that considers decision-space diversity in multi-modal optimization. The original PA-DDS and the omni-optimizer algorithms provided better distributed optimal fronts in the objective space with much lower archive size in a single trial, but they were highly similar and clustered in one region in the decision space. The omni-optimizer and PA-DDS with the traditional archiving strategy were also unable to preserve dissimilar near-optimal solutions in their archive with the progress of optimization due to global domination check among solutions, while the cluster-based PA-DDS performed better in terms of diversity maintenance in the decision space.

The proposed archiving approach was developed for the single solution-based PA-DDS algorithm in this study, but it can be also developed for population-based MO algorithms provided they are equipped with an unbounded archive or an adaptive archive size, such as BORG MOEA (Hadka and Reed, 2013). In a single solution-based MO algorithm such as PA-DDS, clustering should be performed as soon as a solution is generated, i.e. 10000 times clustering update for 10000 function evaluations. However, in a population-based optimization algorithm, clustering needs to be performed only after a new generation of solutions are evaluated, e.g. 100 times clustering update is required for 10000 solution evaluations when the population size is 100. In addition, selection metrics such as the objective space crowding distance, hypervolume contribution, or convex hull contributions are not immediately applicable when solutions are clustered. Instead, a simple random selection metric or more advanced effective metrics need to be introduced to jointly consider the decision space and objective space diversities, such as the one used for the cluster-based PA-DDS in this work, or variation rate in Cuate and Schütze (2019) is suggested for the cluster-based solution archiving.

The cluster-based PA-DDS provides a large archive of solutions that gives more flexibility to the decision-maker in refining their preferences. There are also objectives in the real world that are not quantifiable and cannot be formulated definitively. These types of objectives are evaluated through subjective judgements in the post-processing stage (Liebman, 1976). In addition, decision-makers may change their preferences in the post-processing stage after interacting with stakeholders. For example, if a decision-maker is interested in benefit values higher than a specified threshold instead of an inertial maintenance index of higher than 0.85 in the lake problem, other new distinct or dissimilar solutions will be found. Therefore, we recommend archiving and presenting all clustered solutions for the decision makers. Advanced visualization techniques such

as VIDEO (Kollat and Reed, 2007) and moGrams (Trawiński et al., 2018) should be used for presenting the large number of design options for the decision makers and helping them find their most desirable option.

3.7 Acknowledgement

This work was supported by Dr. Asadzadeh's Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant [RGPIN-2016-05896]; and Graduate Enhancement of Tri-Council Stipends (GETS), University of Manitoba, Winnipeg, MB.

4 PARAMETER UNCERTAINTY ESTIMATION USING CLUSTER-BASED MULTI-OBJECTIVE OPTIMIZATION

Shahram Sahraei[1], Masoud Asadzadeh[2]

[1] Ph.D. Candidate (Corresponding Author), Department of Civil Engineering, University of Manitoba, Canada
E-mail: sahraeis@myumanitoba.ca

[2] Associate Professor, Department of Civil Engineering, University of Manitoba, EITC E1-332, 15 Gillson Street, Winnipeg, MB, Canada, R3T 5V6, Ph: +1 (204) 474 9535
E-mail: masoud.asadzadeh@umanitoba.ca

4.1 Abstract

A reliable hydrologic model simulation requires an effective quantification of the uncertainty in model parameter values and predictions. This paper introduces a novel method to estimate model prediction bounds due to the uncertainty in model parameters considering multiple model performance metrics. The cluster-based multi-objective optimization is utilized to identify distinct parameter sets that meet the so-called behavioral thresholds in a single trial of the optimization algorithm. The proposed method is applied to estimate the parameter uncertainty for three hydrologic models with different parameterizations and is compared to GLUE and the many-trial, low-budget, dynamically dimensioned search (DDS-AU), based on reliability (percent coverage of output time series), sharpness (width of the prediction uncertainty), and the number of behavioral solutions. Results indicate that the proposed method is at least as effective as (if not better than) DDS-AU and GLUE for deriving 95% prediction bounds. Moreover, it is found that

the behavioral threshold identification based on the aggregated model performance metrics can mislead DDS-AU and GLUE; therefore, it is concluded that these thresholds should be defined in a multi-objective uncertainty quantification approach such as the proposed cluster-based multi-objective optimization.

Keywords: Hydrologic Models, Model Calibration, Parameter Uncertainty Estimation, Multi-Objective Uncertainty, Cluster-Based Optimization

4.2 Introduction

Precipitation-runoff models are essential tools for a myriad of applications in water resources systems. These models are mathematical representations of the water cycle and the simplification of complex non-linear hydrologic processes. Model calibration can increase the accuracy of simulating a desired hydrologic response, typically measured streamflow, in a deterministic fashion. Calibration of hydrologic model parameters is perceived as a wicked problem as models suffer from multiple sources of uncertainty, including imperfect model structure, observational data, model parameters, the initial state of the hydrologic system, model output, modeller's subjective judgements, and natural complexity in the hydrologic system (Beven and Freer, 2001; Hoffman and Hammonds, 1994; Nearing et al., 2016; Reed and Kasprzyk, 2009; Uusitalo et al., 2015). The first three aforementioned sources are considered as fundamental sources of the simulation uncertainty (Matott et al., 2009; Tolson and Shoemaker, 2008). Various techniques developed for quantifying the model parameter uncertainty and the model simulation reliability assessment can be categorized under fuzzy set or possibilistic theory[1] (Ahmadi et al., 2019; Franks et al., 1998; Freer et al., 2004; Jacquin and Shamseldin, 2007; Wang et al., 2016), probabilistic

Bayesian framework (Jin et al., 2010; Schoups and Vrugt, 2010; Vrugt et al., 2009), and rejectionist methods (Beven, 2006; Beven and Binley, 1992; Tolson and Shoemaker, 2008; van Griensven and Meixner, 2006). Fuzzy set or possibilistic theory originates from the concept that uncertainty is due to insufficient or inaccurate information (Jacquin and Shamseldin, 2007; Zadeh, 1978). There are many variants of fuzzy methods in the literature that have been successfully used for characterizing uncertainty including standard and modified fuzzy mathematics and incremental modified fuzzy extension (Alvisi and Franchini, 2013; Faybishenko, 2010; Huang et al., 2010; Maskey et al., 2004; Nasseri et al., 2013). This thesis mainly focuses on probabilistic and rejectionist views of uncertainty analysis in hydrologic modelling.

Formal uncertainty quantification techniques attempt to separate various sources of uncertainty from the model parameters uncertainty. This technique explicitly assumes *a priori* statistical definition for the residuals between measurements and simulated model responses, i.e. the probability density function of residuals, to derive *a posteriori* likelihood and verify the validity of assumptions (Box and Tiao, 1992; Dotto et al., 2012; Freni and Mannina, 2010; Hutton et al., 2014; Schoups and Vrugt, 2010; Stedinger et al., 2008; Yang et al., 2007). For example, Vrugt et al. (2009) considered three assumptions to separate model structural and forcing data uncertainty from model parameter uncertainty. First, it was assumed that residuals (discrepancy between simulated and measured values) take on a Gaussian distribution. Second, it was assumed residuals are auto-correlated. Third, the error in forcing data was corrected by unknown multipliers. Beven et al. (2008) argued that, the formal Bayesian uncertainty quantification methods introduce new unknown variables whose joint inference alongside model parameters may lead to new sources of uncertainty implicitly affecting the parameter uncertainty quantification, which is inconsistent with the explicit formal inference concept. In addition, the shape of the parameter probability

distribution is over-conditioned by *a priori* definition of *a posteriori* likelihood and other *a priori* assumptions for the residuals; therefore, a wrong definition/assumption can result in bias in the parameter probability distribution and can mislead model prediction bounds (Beven and Smith, 2015; Beven et al., 2008).

Approximate Bayesian Computation (ABC), originally developed by Pritchard et al. (1999), is a likelihood-free method to deal with impracticality or difficulty of calculating a formal likelihood function since it has no general definite form. This technique relies upon a summary statistic that compares simulated to observed time series (Turner and Van Zandt, 2012). The summary statistic should sufficiently explain the parameter sets by encapsulating entire information in the data (Csilléry et al., 2010; Turner and Van Zandt, 2012). For example, mean and variance are sufficient for a Gaussian model (Kavetski et al., 2018). However, it is practically infeasible to derive a sufficient summary statistic for more general probability models (Fenicia et al., 2018; Kavetski et al., 2018). The simplest form of ABC is called ABC rejection sampling that eliminates a sampled parameter set if its summary statistic value is worse than a pre-defined tolerance (Beaumont et al., 2002; Pritchard et al., 1999). This version is often inefficient since the rejection rate becomes exponentially high if a small tolerance value is chosen (Turner and Van Zandt, 2012). On the other side, a set of acceptable parameters resulting from very loose tolerance value cannot represent approximate posterior distribution (Kavetski et al., 2018; Turner and Van Zandt, 2012). Many algorithmic enhancements have been made for ABC technique for computational efficiency such as Markov Chain Monte Carlo sampling (Marjoram et al., 2003), Population Monte Carlo Sampling (Cappé et al., 2004), Sequential Monte Carlo Sampling (Toni et al., 2008), and adaptive acceptance tolerance tightening (Albert et al., 2014; Lenormand et al., 2013). The studies by

Fenicia et al. (2018), Kavetski et al. (2018), Nott et al. (2012), and Vrugt and Sadegh (2013) are examples of ABC application in hydrology for uncertainty assessment.

Beven and Freer (2001) introduced a Monte Carlo uncertainty quantification approach that does not require pre-conditioning and is based on the concept of model equifinality, i.e. different model configurations have similar performance. The generalized likelihood uncertainty estimation (GLUE), proposed by Beven and Binley (1992), is arguably the most well-known approach that uniformly samples from the parameter space and obtains a likelihood by subjectively evaluating parameter sets based on commonly used error metrics including, but not limited to, Nash-Sutcliffe Efficiency (*NSE*) (Nash and Sutcliffe, 1970), Kling-Gupta Efficiency (*KGE*) (Gupta et al., 2009), or sum of squared errors. The modeller also requires to subjectively specify a threshold to select behavioral solutions for quantifying uncertainty in the model prediction results. Numerous studies have used GLUE for quantifying parameter uncertainty in the context of hydrologic model calibration (Kan et al., 2020; Ragab et al., 2020; Tegegne et al., 2019; Teweldebrhan et al., 2018; Xie et al., 2019; Xue et al., 2018). Despite its simple structure, GLUE is criticized for its inefficient Monte Carlo sampling method, especially in computationally intensive models, and subjective screening analysis to identify behavioral parameter sets (Blasone et al., 2008a; McMichael et al., 2006; Tolson and Shoemaker, 2008; Vrugt et al., 2009). It is believed that "GLUE does not adequately address the quest for rigorous evaluation of hydrological hypotheses" (Clark et al., 2011). Interested readers are referred to review comments and discussions raised by scholars with different schools of thought about the philosophies and working hypotheses in uncertainty assessments of hydrologic models in (Beven, 2006; Clark et al., 2012, 2011; Nearing et al., 2016).

For the sake of efficient sampling and reliable uncertainty quantification, a number of alternate methods have been proposed. Christiaens and Feyen (2002) and Uhlenbrook and Sieber (2005)

used Latin Hypercube Sampling (LHS) (Iman et al., 1980) within GLUE for a more uniformly distributed and stratified parameter sampling. van Griensven and Meixner (2006) proposed the parameter solution (ParaSol) method that utilizes a modified structure of Shuffled Complex Evolution-University of Arizona (SCE-UA) algorithm (Duan et al., 1992) with an increased value of perturbation rate for parameter space coverage enhancement. ParaSol retains all simulations generated during the optimization, classifies them to well-performed and ill-performed simulations based on a threshold defined by χ^2–statistics. All of the well-performed simulations are equally important for creating uncertainty bounds in the ParaSol method. Abbaspour et al. (2004) developed Sequential Uncertainty Fitting-2 (SUFI-2) that couples parameter sensitivity with LHS for predictive uncertainty analysis. The calibration parameter ranges in SUFI-2 are initially set as wide as possible while maintaining their physical feasibility. SUFI-2 recursively narrows down the calibration parameter space to the range that sufficiently replicates the hydrological processes using the sensitivity matrix and parameter covariance matrix. In addition, SUFI-2 evaluates the prediction uncertainty performance based on the percent coverage of response time series (with an ideal value of 100%) and the average width of the uncertainty (with an ideal value of zero). If the value of the uncertainty performance metrics in SUFI-2 are satisfactory, the posterior parameter distribution can be derived by uniform sampling from the narrowed parameter range. McMillan and Clark (2009) used a Markov Chain Monte Carlo method called Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm based on a modified *NSE* metric that accounts for the timing and magnitude errors. They concluded that the SCEM-UA method provides a more reliable total simulation uncertainty bound than its formal Bayesian counterpart due to a better investigation of the behavioral region. Tolson and Shoemaker (2008) devised a Dynamically Dimensioned Search algorithm for the Approximation of Uncertainty (DDS-AU) that divides the total computational

budget into a large number of low-budget independent optimization trial runs. DDS starts with a global search and ends with a local search to converge to an effective parameter set. After testing DDS-AU for the calibration of multiple hydrologic model cases with 13 to 30 parameters, Tolson and Shoemaker (2008) asserted that this algorithm is two to three orders of magnitude more efficient than GLUE for identifying behavioral solutions with sharper and more reliable uncertainty bounds.

Model calibration is inherently multi-objective (Gupta et al.,1998), and a behavioral model should acceptably pass all the objectives used for model evaluation. For example, a hydrologist estimating flood would be concerned about the accuracy of the model simulating the timing, magnitude, and volume of the flood. The methods above, however, are not designed to consider multiple model performance metrics. Therefore, the modeller must lump the model performance metrics to use these methods for uncertainty analysis. Balin (2004), Lamb et al. (1998), Muleta and Nicklow (2005) used a combined likelihood measure that weights behavioral parameter sets based on the aggregation of performance metrics for multiple responses of their watershed model case studies. Blazkova et al. (2002) identified behavioral parameter sets through combined evaluation of streamflow time series and simulation of the measured saturated area. Definition of a likelihood that accounts for simulating the saturated area strongly constrained the transmissivity parameter of their model case study, while it had minimal impact on streamflow prediction uncertainty bounds. Gallart et al. (2007) used more internal watershed information for updating the likelihood of behavioral parameters, which accounts for water table measured data besides the saturated area for conditioning streamflow prediction uncertainty. Other related studies include (Dean et al., 2009; Mitchell et al., 2009; Sun et al., 2016). An aggregated formulation of multiple performance metrics for uncertainty analysis often increases uncertainty and expands the behavioral solution

set due to subjective weighting of performance criteria. Also, updating the likelihood of already selected behavioral solutions does not check the acceptability of solutions based on behavioral threshold specification of new measured data. A multi-criteria uncertainty analysis should simultaneously account for multiple pre-determined acceptable thresholds for all likelihood measures without combining likelihoods or aggregating the performance metrics. For example, Pang et al. (2020, 2019) adopted a multi-criteria decision analysis within GLUE by multiple threshold assignments to four performance metrics (representing different aspects of streamflow error distribution) for parameter identification and reducing prediction uncertainty of streamflow time series.

This study introduces a multi-objective parameter uncertainty approach that uses the sampling strategy of Cluster-Based Pareto Archived-Dynamically Dimensioned Search (CB-PA-DDS) for quantifying parameter uncertainty. The proposed approach is compared against GLUE and DDS-AU based on the number of identified behavioral solutions, length of measured response time series covered by the estimated uncertainty band in percent, and the average width of the estimated uncertainty band. Moreover, the effect of aggregation of objectives in GLUE and DDS-AU on the process of identifying behavioral parameter sets is studied in comparison with the proposed CB-PA-DDS that simultaneously considers all conflicting objectives without aggregating them. The remainder of this paper is organized as follows. Section 4.3 introduces the hydrologic model case studies and available data and describes the structure of GLUE, DDS-AU, and the proposed cluster-based algorithm for uncertainty estimation of the hydrologic models. The performance of the proposed approach is compared with GLUE and DDS-AU in section 4.4, followed by concluding remarks in section 4.5.

4.3 Materials and Methods

4.3.1 GLUE

GLUE was developed by Beven and Binley (1992) based on the equifinality paradigm that rejects the idea of finding a single best parameter set as it is an overfit to a specific aspect of modelling error distribution. Equifinality stems from insufficient information of the hydrologic system and assumes that many model configurations can be considered equally likely (Beven, 1993). The GLUE uncertainty quantification method encompasses several steps. First, the feasible parameter space is sampled with a Uniform Distribution. Second, a pre-specified likelihood function is calculated for each sampled parameter set. Specification of the likelihood function should be coherent with the simulation performance and the behavior of the hydrologic system (Beven et al., 2008). Third, the sampled parameter sets that do not meet a pre-specified likelihood function value are discarded and the retained samples represent behavioral models. Each behavioral sample is associated with a likelihood value used to weigh its model predictions in obtaining a cumulative distribution function of the model prediction variable over all behavioral samples. Then, the predictive uncertainty limits are derived from the quantiles of this distribution (for example 2.5% and 97.5% percentiles in this paper).

4.3.2 DDS-AU

As proposed by Tolson and Shoemaker (2008), DDS-AU estimates the model prediction bounds using a single-objective stochastic optimization algorithm executed multiple independent times with a relatively small computational budget. Similar to GLUE, DDS-AU uses a subjective summary statistic and behavioral threshold, but its sampling strategy is not uniform. The

optimization commences with a global search that dynamically and probabilistically turns into a local search by decreasing the number of dimensions for perturbing the so-far best solution for generating a new candidate solution. DDS perturbs each of the selected model parameters of the current best solution using a Normal Distribution centered at the current best value of the parameter. Each trial of DDS-AU has a different random initialization leading to a different final solution for each trial. For instance, if the computational budget allows for 10,000 solution evaluations, DDS-AU can be set to 100 independent trials/searches, each with 100 samples from the parameter space, resulting in 100 archived solutions for behavioral solution assessment and uncertainty quantification. The low computational budget of each trial prevents the algorithm from converging to a good quality solution in all trials. An archived solution that has a likelihood worse than the pre-determined threshold value is identified as a non-behavioral solution and is excluded for uncertainty approximation. Unlike GLUE, DDS-AU does not estimate the cumulative distribution function of model predictions. Instead, it approximates prediction bounds without probability assignment to the prediction variable. DDS-AU has only one parameter called perturbation rate that is suggested to be fixed at 0.2 by the developers of the algorithm. Readers are referred to Tolson and Shoemaker (2008) for more details about the DDS-AU structure.

4.3.3 Cluster-Based Pareto Archived-Dynamically Dimensioned Search (CB-PA-DDS)

This algorithm, developed by Sahraei and Asadzadeh (2021), is a restructured PA-DDS algorithm (Asadzadeh and Tolson, 2013) that uses a cluster-based archiving strategy to find optimal and near-optimal diverse design tradeoffs by optimizing multiple conflicting objectives simultaneously. Upon generating a solution in each optimization iteration, the solution density around the new solution is calculated by CB-PA-DDS. Suppose the density is equal to or higher

than a pre-specified population density threshold. In that case, the new solution will either form a new cluster along with the neighboring solutions with a unique cluster tag assignment or joins a previously formed cluster. On occasions that the neighboring density of a new solution is below the pre-specified threshold, it is considered as an unclassified solution. The CB-PA-DDS is equipped with a localized dominance-check strategy meaning that solutions with different cluster labels are not compared, and solutions dominated by their groupmates are discarded from the archive. The archive contains only cluster-non-dominated and so-far unclassified solutions. The currently unclassified solutions are retained and may form new clusters in subsequent iterations. The non-dominated members are assigned a memory of solution coverage density for tracking the history of the number of solutions in a cluster from the beginning of the optimization and help the algorithm to comply with the definition of cluster formation once a cluster becomes less crowded after the local dominance strategy. Each cluster forms a tradeoff in the objective space. CB-PA-DDS utilizes a selection metric that jointly considers dissimilarity in the decision-space and convergence in the objective space for selecting one solution from the archive to generate a new solution for the next iteration. The selection metric assigns a value between zero and two to each cluster and unclassified solutions. An unclassified solution or a cluster with the shortest distance to the ideal point in the normalized objective space (scaled to between zero and one) that is located far away from other clusters and unclassified solutions in the normalized decision space (scaled to between zero and one based on the feasible parameter range) has a higher selection metric value, hence a higher chance of selection. Once a cluster is chosen, one of its members is randomly selected for generating new solution.

CB-PA-DDS has three parameters: (1) the perturbation rate parameter that belongs to the original PA-DDS algorithm and, similar to DDS-AU, it is set to a recommended value of 0.2, (2) the

minimum number of solutions (*MinPts*) for initial cluster formation that is recommended to be four, (3) the cluster radius or neighboring distance (ε) to calculate neighboring density around each solution for cluster formation. The value of ε depends on the available computational budget, the dimensions of the decision-space and objective-space, and the multi-modality nature of the problem. For this reason, Sahraei and Asadzadeh (2021) recommended the range in equation (14) for ε that reflects its dependency to the computational budget and objective space dimensionality. In order to determine the upper and lower bounds for ε, the user needs to identify a meaningful distinction level between two decision vectors. For example, if two solutions with at least a 7% discrepancy in each decision variable are deemed distinct/dissimilar, ε would range from 0.125 to 0.188. For a higher computational budget and a low dimensional decision and objective space, a lower value of ε in this range is chosen and vice versa. D_i in equation (14) is the minimum meaningful discrepancy percentage for the decision variable *i*, and *n* is the decision space dimension.

$$\varepsilon \in [0.8z, 1.2z\,], \quad z = \frac{1}{100}\sqrt{D_1^2 + D_2^2 + \cdots + D_n^2} \tag{24}$$

In general, the sampling strategy in optimization algorithms is designed to rather focus on exploitation of the parameter space than exploration. Exploitation is defined as rapid convergence to the optimal region, while exploration is important to emphasize coverage of a large proportion of the parameter space without leaving unsearched areas (Khu and Werner, 2003). The sampling method within GLUE is purely random and focuses only on exploration, hence not efficient. Each trial of DDS-AU, on the other hand, progressively converges to only one region of interest assuming that the behavioral space is composed of multiple distinct regions of interest in the domain of the parameter space. For this reason, each region of interest is represented by one

solution (or sometimes more than one solution if multiple trials of DDS-AU converge to the same region) due to its single-objective characteristic (Tolson and Shoemaker, 2008).

The clustering strategy and the localized dominance implemented for PA-DDS is to put more emphasis on exploration and identifying distinct parameter sub-spaces that contain good quality solutions. Each of these sub-spaces is represented with a cluster of solutions that form a tradeoff with respect to multiple objectives considered for uncertainty estimation using CB-PA-DDS. As a result, unlike DDS-AU, each region of interest is not represented by one solution. Instead, all cluster-non-dominated solutions are equally important for a multi-objective uncertainty estimation using CB-PA-DDS. For example, Zhou et al. (2016) used a multi-objective optimization-based sampling strategy based on the non-dominated sorting genetic algorithm (ε-NSGAII) and included all non-dominated archived solutions that meet the behavioral thresholds assigned to each objective for multi-objective uncertainty estimation. This is called a Pareto-based uncertainty approximation in Shafii et al. (2015) and refer to this citation for more details.

4.3.4 Hydrologic Model Case Studies

The calibration of two hydrologic models, HBV-SASK and RAVEN, is considered in this study to develop and fine-tune the cluster-based optimization approach for uncertainty estimation against GLUE DDS-AU. The developed approach is tested for the calibration of a third hydrologic model, SWAT, in comparison with GLUE and DDS-AU. These models were previously configured, and their calibration problem was published in the literature.

4.3.4.1 HBV-SASK Model of Bow River

This hydrologic model is developed by Saman Razavi based on the Hydrologisca Byråns Vattenbalansavdelning (HBV) conceptual model (Lindström et al., 1997) at the University of

Saskatchewan for educational purposes and research applications (Gupta and Razavi, 2018; Razavi et al., 2019). The model demands at least four classes of input data, including long-term monthly temperature and potential evapotranspiration to calculate daily potential evapotranspiration, daily temperature, and daily precipitation, along with the specification of 12 uncertain parameters for successful modelling. The list of model parameters and their ranges is displayed in Table 4-1. The snowmelt and snow accumulation in the HBV-SASK model is computed based on the degree-day method. A fraction of the released water either from the melted snow or direct rainfall or their mixture evaporates and the remaining infiltrates into the soil storage. A proportion of the stored moisture in the soil percolates into deep groundwater linear storage (base flow generation), and the rest runs off quickly as fast flux. After routing based on the unit hydrograph, the fast discharge is added to the base flow, or slow discharge resulted from deep percolation to obtain the total watershed daily hydrograph.

The study area is Bow River Basin at Banff, Canada, with an area of 2178.53 km^2 that ultimately discharges into the South Saskatchewan River Basin. The historical forcing and response observations span from 1950 to 2011, the first three years of which are considered for model spin-up and the following 15 years are used for model parameter uncertainty analysis. The remainder of data is used as evaluation period to compare the performance of the uncertainty techniques for a situation that the model has not experienced. The multi-objective uncertainty of the HBV-SASK model parameters is estimated by considering three metrics shown in equation (25) that are the decomposed terms of the *KGE* metric, i.e. correlation coefficient (r), long-term mean flow bias (β), and the variability index (α) (Gupta et al., 2009). σ_s, σ_o, μ_s, and μ_o represent the standard deviation and mean of simulations and observations. The correlation coefficient metric accounts for streamflow magnitude and timing performance of the response time series with respect to

precipitation events whereas the bias and variability respectively seek a better fit for the first and second moments of the response time series (Gupta et al., 2009). The aggregated *KGE* metric in equation (26) is used for uncertainty estimation in GLUE and DDS-AU since they can only handle one model performance metric.

$$f_1 = (r-1)^2$$
$$f_2 = (\alpha-1)^2 = (\sigma_s/\sigma_o - 1)^2 \quad \textbf{(25)}$$
$$f_3 = (\beta-1)^2 = (\mu_s/\mu_o - 1)^2$$

$$KGE = 1 - \sqrt{f_1 + f_2 + f_2} \quad \textbf{(26)}$$

Table 4-1: name, range and description of HBV-SASK model parameters

Name	Range	Parameter Description
TT	[-4, 4]	Threshold for air temperature in *°C* for distinguishing snowfall from rainfall or melting/freezing
C0	[0, 10]	Daily base melt factor, in *mm/°C*
ETF	[0, 1]	Correction factor in *1/°C* to account for temperature anomaly in potential evapotranspiration
LP	[0, 1]	A multiplier to FC that limits evapotranspiration from soil moisture
FC	[50, 500]	Maximum soil water containment, i.e. field capacity of soil, in *mm*.
Beta	[1, 3]	Exponent of soil water release equation, i.e. shape parameter (No unit)
FRAC	[0.1, 0.9]	A fraction of released water from soil moisture that enters fast storage (No unit)
K1	[0.05, 1]	A regulating factor for daily release of water from fast storage (No unit)
Alpha	[1, 3]	Exponent of fast storage equation (No unit)
K2	[0, 0.05]	A regulating factor for daily release of water from slow storage (No unit)
UBAS	[1, 3]	Base value of unit hydrograph for routing daily flux produced from fast storage
PM	[0.5, 2]	A multiplier for handling uncertainty in daily precipitation (No unit)

4.3.4.2 RAVEN Model of Grand River

The semi-distributed Raven model developed by Craig (2015) partitions a basin into multiple sub-basins. Each sub-basin is sub-divided into multiple smaller components called Hydrologic Response Units (HRUs). The areas that have similar combination of land type/use, geometry, geography, topography, and soil/aquifer are lumped into one HRU. The meteorological

assignments to each HRU in the model include the precipitation, temperature, and wind velocity. The Raven model simulates the vertical energy and water balance and assembles the simulated hydrological components in each HRU. The HRUs are reconnected to laterally route simulated flow from each HRU toward main river reach in each sub-basin. The sub-basin elements are then reconnected to route streamflow to toward the outlet.

The Raven model used in this study simulates the outlet streamflow of an upstream sub-basin of the greater Grand River with an area of 274 km^2 in Ontario, Canada. The model calibration and validation were set up as an optimization problem for tuning 20 parameters from 2003 to 2014 by Shafii et al. (2017). Table 4-2 lists the name, range, and description of model parameters. In this study, the year 2003 is used for the model spin-up period when the model performance is not measured for model calibration. Daily measured streamflow in years 2004 to 2009 is used to quantify parameter uncertainty, and daily measured streamflow in the years 2010 to 2014 are used for model evaluation. The same criteria used for the HBV-SASK model, i.e. equations (25) and (26), are used for evaluating the performance of the RAVEN model. For details regarding the calculation of hydrologic compartments in the RAVEN modelling framework, readers are referred to Craig (2015) and Shafii et al. (2017).

Table 4-2: name, range, and description of the RAVEN model parameters

Name	Range	Parameter Description
T_{RS}	[-1, 3]	Temperature at which a mixture of snow and rain occurs (°*C*)
ΔT	[0, 1]	Range of temperature with T_{RS} as a midpoint to consider precipitation as a mixture of snow and rain (°*C*). Out of this range, precipitation is either rain or snow.
β_1	[1, 3]	Exponent of HBV method developed by Bergström (1995) to calculate infiltration for 3 out of 7 soil groups
β_2	[0.1, 2]	Same as above for the rest of soil groups
F_{sparse}	[0.75, 0.99]	Vegetation sparseness index for calculating canopy interception and evaporation
P_{max}	[10, 50]	Maximum percolation rate when soil water storage in upper layer exceeds field capacity (*mm/day*)
BF_{max}	[1, 5]	Baseflow equal to maximum baseflow rate in *mm/day* for top layer of soil classes assuming top layer is always at saturation
k	[0.0001, 0.01]	Baseflow coefficient (*1/day*) in non-linear storage method to calculate the contribution of lower layer of soil classes to baseflow

n	[1, 3]	Exponent of non-linear storage method to calculate the contribution of lower layer of soil classes to baseflow
S_{canopy}	[1, 10]	Maximum canopy storage (*mm*)
α	[1, 50]	Abstraction rate as a proportion of accumulated ponded water due to excess precipitation (%)
$CPET_{lake}$	[0.1, 1]	Potential evapotranspiration correction coefficient for lakes and open water area
$M_{a,inc}$	[0.1, 5]	Potential melt rate added to $M_{a,min}$ (*mm/d/°C*)
$M_{a,min}$	[1, 10]	Minimum potential melt rate in *mm/d/°C* in HBV method
C_a	[0.1, 0.99]	Aspect melt correction coefficient for the potential melt calculation using HBV method
$CPET_{winter}$	[0.5, 1.5]	Correction factor for the potential evapotranspiration from November to April
$CPET_{spring}$	[1, 1.2]	Correction factor for the potential evapotranspiration for May and June
$CPET_{summer}$	[1.2, 1.5]	Correction factor for the potential evapotranspiration for July and August
$CPET_{fall}$	[1.2, 1.5]	Correction factor for the potential evapotranspiration for September and October
$C_{raingauge}$	[0.85, 1.15]	A correction factor near 1.0 to estimate proper rainfall volumes

4.3.4.3 SWAT Model of the Rouge River

The first version of the semi-distributed soil and water assessment tool (SWAT) was developed by Arnold et al. (1990). SWAT divides a basin into sub-basins to simulate long-term hydrological processes on a daily time scale. SWAT lumps areas with homogeneous land use and soil characteristics within each sub-basin into HRUs. Unlike RAVEN, the HRUs in SWAT are not assigned any geographical location meaning that the simulated hydrological sub-components in each HRU are aggregated for each sub-basin. The resulting streamflow at the outlet of each sub-basin is routed downstream via the drainage network towards the outlet. The SWAT model of this study is adopted from Asadzadeh et al. (2016) that utilizes the Soil Conservation Service (SCS) method for surface runoff modelling of the 331-km^2 Rouge River Basin to study the effect of agricultural land operations on the streamflow rate and quality (water quality simulation/evaluation is not included in this study). The SCS method assigns different values recommended by Neitsch

et al. (2011) to the curve number (SCS parameter) based on the land use and slope, soil type and permeability, and antecedent moisture. Four main land use classes, including urban, agricultural, natural, and water bodies, are defined within the model. Asadzadeh et al. (2015) used this model to simulate agricultural operations in terms of cultivation, tillage, planting, crop rotation, and fertilizers. Asadzadeh et al. (2016) demonstrated that daily streamflow simulation of the Rouge River Basin suffers from a one-day offset issue mainly because of the long concentration time between most of the peaks in the hyetograph and the corresponding hydrograph. The SWAT model calibration in this paper accounts for the effect of the one-day offset issue.

The historically observed data, including the hydrometric and climatic data, are available for multiple stations from 2005 to 2011. See Asadzadeh et al. (2016) for details about gauges and their spatial distributions in the basin. The year 2005 is used for model spin-up, years 2006 to 2009 are used for the uncertainty analysis of the model parameters listed in Table 4-3, and years 2010 and 2011 are used to evaluate the effectiveness of the parameter uncertainty bounds. Two hydrometric stations are used for parameter uncertainty quantification of SWAT by maximizing the lower *NSE* value and minimizing the higher simulation error in each segment of the flow-duration curve between two hydrometric stations after partitioning their curves into three segments of high-flow volume (<2% occurrence time), logarithmic low-flow volume (>70% occurrence time), and slope in the mid-flow portion (from 20% to 70% occurrence time). The model is also constrained to account for evapotranspiration of equal to 62% of annual precipitation in the basin and 60% of the remaining precipitation as surface flow, according to Asadzadeh et al. (2015). The constraint is considered as the fifth objective to penalize the parameter uncertainty estimation. The SWAT model uncertainty quantification uses the formulations in equation (27) for calculating objective functions. $MSE, O, S, E_{HF}, E_{MF}, E_{LF}, ET_S, AP, SF_S, TY_S$ in equation (27) respectively denote mean

of squared errors, measured streamflow, simulated streamflow, simulated high-flow volume error, simulated mid-flow slope error, simulated low-flow volume error, simulated evapotranspiration, annual precipitation, simulated surface flow, and simulated total water yield. A weighted aggregation is employed for GLUE and DDS-AU since they are single-criterion uncertainty methods. As shown in equation (28), a high weight of 0.6 is specified for *NSE* and the remainder are given an equal weight of 0.1 to have a more accurate coverage around peak flows. The subscripts and superscripts 1 and 2 in equation (28) represent the hydrometric stations number one and two (With respective station IDs of 02HC022 and 02HC028 in https://wateroffice.ec.gc.ca/search/historical_e.html). The hydrometric station ID of 02HC022 is located downstream of 02HC028.

$$NSE = 1 - \frac{MSE}{\sigma_O^2}$$

$$E_{HF} = 100\left(\frac{\sum_{t\cong 0\%}^{2\%}|O_t - S_t|}{\sum_{t\cong 0\%}^{2\%} O_t}\right)$$

$$E_{MF} = 100\left(\frac{\left|log_{10}^{\frac{S_{20\%}}{S_{70\%}}} - log_{10}^{\frac{O_{20\%}}{O_{70\%}}}\right|}{log_{10}^{\frac{O_{20\%}}{O_{70\%}}}}\right) \quad \textbf{(27)}$$

$$E_{LF} = 100\left(\frac{\sum_{t=70\%}^{100\%}|O_t - S_t|}{\sum_{t=70\%}^{100\%} O_t}\right)$$

$$f = \frac{|ET_S/AP - 0.62| + |SF_S/TY_S - 0.6|}{2}$$

$$F = 0.6\,(max\{-NSE_1, -NSE_2\} + 1) + 0.1(max\{E_{HF}^1, E_{HF}^2\} + max\{E_{MF}^1, E_{MF}^2\} + max\{E_{LF}^1, E_{LF}^2\} + f) \quad \textbf{(28)}$$

Table 4-3: name, range, and description of the SWAT model parameters

Name	Range	Parameter Description

CN	[0.75, 1.25]	Curve number multiplier
SMFMN	[1.4, 4.5]	Snowmelt factor in $mm\ H_2O$/day/°C on the last day of the fall season, December 21.
TIMP	[0.01, 1]	Snowmelt factor representing lag in snow temperature
ESCO	[0.01, 1]	Soil depth distribution modifier to compensate for evaporative demand of the soil
EPCO	[0.01, 1]	A factor compensating for plant uptake from deeper layers when soil moisture content is low
SURLAG	[0.1, 24]	A parameter that accounts for lag in surface runoff
SOL_AWC	[0.1, 2]	Available soil water capacity
SOL_K	[0.1, 100]	Hydraulic conductivity when soil is saturated (mm/hr)
GW_DELAY	[1, 500]	Groundwater recharge delay time in days
ALPHA_BF	[0.1, 1]	Baseflow recession constant
GWQMN	[0, 5000]	Water level threshold in shallow aquifer to contribute to the main channel streamflow ($mm\ H_2O$)
CH_N2	[0.01, 0.1]	Channel Roughness (Manning's n)
CH_K2	[0, 100]	Effective hydraulic conductivity of channel in mm/hr

4.3.5 Numerical Experiment Setup

The uncertainty estimation techniques in this paper use a stochastic sampling process for seeking behavioral solutions, and their performance should be evaluated through multiple independent runs for uncertainty analysis. 10 independent runs with 10000 solution evaluations are considered for each uncertainty technique. In each run, DDS-AU is restarted 100 times with a constant number of solution evaluations to sample from different regions of the parameter space and archive at most 100 behavioral solutions, while GLUE continues with no restart until the maximum number of function evaluations, i.e. 10000 per trial, is satisfied. CB-PA-DDS is run for multiple settings of ε and the algorithm for each ε value is run for 10 independent trials, each with 10000 solution evaluations. According to the recommended range in equation (14) for ε, at least 9 to 11 percent difference (D_i) between two solutions in each of their parametric dimensions is assumed in this paper to have a meaningful level of distinction between them. As a result, different ranges are obtained for ε for each model as the models have a different number of unknown parameters. The HBV-SASK and RAVEN hydrologic models are used for fine-tuning the proposed methodology and determine recommended values for ε. Assuming that 9% difference results in a lower bound ε of 0.249 and 0.322 respectively for the HBV-SASK and RAVEN models. The upper bound of ε

will become 0.457 and 0.590 when an 11% difference is assumed for the HBV-SASK and RAVEN models, respectively. Table 4-4 shows the ε values tested within the recommended ranges.

An ε value of 0.5 can be interpreted as the distance between two *m*-dimensional solutions with $0.5/\sqrt{m}$ difference in each of their parameters, for example. Alternatively, these two solutions can also particularly have equal values in *m-1* parameters and a 50% difference in one parameter. Both instances can occur, but the latter is similar to dividing each dimension of the normalized parameter space by two, resulting in 1048576 (i.e. 2^{20}) hypothetical boxes in 20-parameter RAVEN, 4096 (i.e. 2^{12}) boxes in 12-parameter HBV-SASK, or 8192 (i.e. 2^{13}) boxes in 13-parameter SWAT model. The first example can occur within a box if we assume one of the two solutions is located on one corner of the box. According to this description, an ε value of 0.5 is equivalent to gridding the normalized parameter space through dividing each dimension by two to stochastically allow cluster formation in the majority of these hypothetical boxes in the normalized parameter space. A higher dimensional model calibration needs a higher computational budget (such as RAVEN) to explore more boxes in the normalized parameter space. If assigning a higher computational budget is not possible in RAVEN, for example, the value of ε can increase to more than 0.5 to merge the neighboring hypothetical grids for achieving more number of behavioral solutions.

Table 4-4: ε values and behavioral thresholds for the uncertainty analysis of the HBV-SASK, RAVEN and SWAT model parameters using CB-PA-DDS algorithm

Model Name	Recommended Range of ε	ε values	Behavioral Thresholds
HBV-SASK	[0.249, 0.457]	0.24, 0.31, 0.38, 0.45	Threshold1: $KGE > 0.67016$ Threshold2: $KGE > 0.80$
RAVEN	[0.322, 0.590]	0.35, 0.42, 0.49, 0.55	Threshold1: $KGE > 0.6529$ Threshold2: $KGE > 0.70$
SWAT	[0.26, 0.47]	0.47	Threshold1: $[1 - NSE, E_{HF}, E_{LF}, E_{MF}, f] < [0.5, 30, 30, 30, 30]$

4.3.6 Behavioral Thresholds, Performance Metrics, and Comparison Strategy

The impact of behavioral threshold on the quantity of uncertainty in Monte Carlo and, more generally, stochastic methods is studied in Beven and Freer (2001), Jin et al. (2010), Montanari (2005), Yang et al. (2008), stating that a choice of high threshold value leads to a small number of behavioral solutions, a narrow uncertainty interval, and hence a low coverage of observational response time series. Lamb et al. (1998) demonstrated that relaxation of the acceptability threshold to get a higher number of behavioral simulations would lead to only minor changes in the range of simulation uncertainty bounds as the added solutions have a poorer performance and lower likelihood values. Low and high behavioral threshold values are chosen (shown in Table 4-4) for the HBV-SASK and RAVEN hydrologic models to examine the effect of high and low thresholds on the uncertainty prediction results.

Multiple performance metrics, including reliability, number of behavioral solutions, and sharpness, are considered to mutually compare the uncertainty estimation strategies for low and high threshold values in the calibration period. Reliability is defined by Yadav et al. (2007) as the number of covered time steps on the response time series divided by the length of the simulation period. When all measured data points are within the simulated bound for the whole length of the simulation period, the reliability index would become 100%. Sharpness, defined by Yadav et al. (2007), on the other hand, is the average contraction level of the prediction interval relative to a posterior prediction bound of a large number of samples uniformly distributed in the feasible range of parameters. The sharpness of only one behavioral solution is 100% since it has no width and the value of sharpness for a group of behavioral solutions becomes less than 100%.

Relaxing behavioral threshold is expected to add more solutions to the behavioral set, increase the reliability index, and decrease the sharpness indicator since the estimated uncertainty interval gets

wider and more time steps on the response time series are likely to be covered. Hence, reliability and sharpness are two conflicting criteria and form a tradeoff if the behavioral threshold is moved from a strict to a generous cutoff value. The size of behavioral solution set is also crucial for a robust estimate of prediction uncertainty bounds (Blasone et al., 2008b).

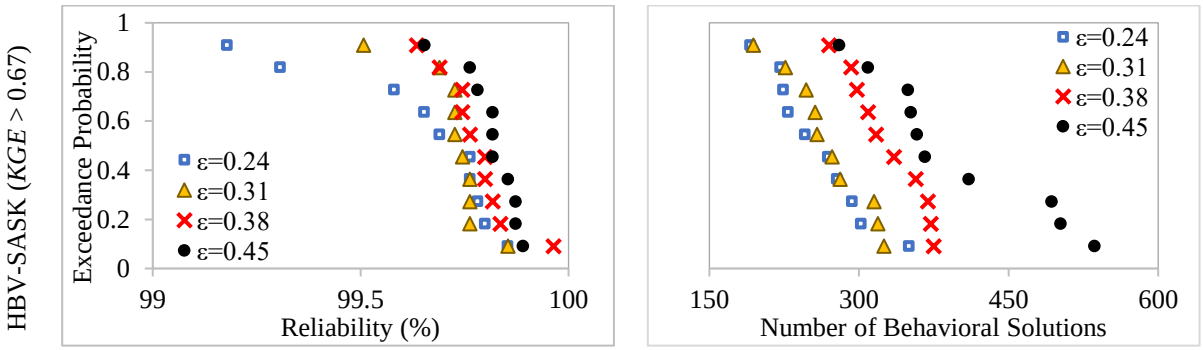
The performance of the estimated uncertainty interval by each method for the evaluation period is assessed by the lower bound *KGE* in the HBV-SASK and RAVEN models and lower bound *NSE* in the SWAT model for the behavioral solutions. Since each prediction uncertainty technique is run for 10 trials, an empirical cumulative distribution function (CDF) plot is constructed such that the best performing trial has the lowest probability of exceedance and the worst performing trial has the highest probability of exceedance. The probability of exceedance is empirically calculated by $R/(N+1)$, where R is the trial rank number in a list of N trials sorted from the best to worst performance metric value. The CDF plots are compared based on the first-degree stochastic dominance strategy, which is defined as better performance in each probability level (Levy, 1992). The statistical significance level of the stochastic dominance is evaluated through the two-sided Wilcoxon rank-sum test (Asadzadeh and Tolson, 2013; Hadka and Reed, 2012). If the p-value is less than 0.05, the null hypothesis of the equal median of two CDFs is rejected, and the stochastic dominance is considered significant.

4.4 Results and Discussion

4.4.1 Methodology Development and Tuning

The 12-parameter HBV-SASK and 20-parameter RAVEN model calibration problems are used to find the best value for ε to effectively and efficiently use the CB-PA-DDS algorithm for model parameter uncertainty approximation. Figure 4-1 compares four different ε settings of CB-PA-

DDS shown in Table 4-4 based on the CDF plots of reliability and size of behavioral set after subjectively specifying the *KGE* values of 0.67 and 0.65 as behavioral thresholds for the HBV-SASK and RAVEN models, respectively. Unlike RAVEN, HBV-SASK is a well-behaved model since the CDF of reliability covers more than 99 percent of the outflow time series for all ε values. Figure 4-1 demonstrate that the reliability index and size of behavioral set tend to increase for both models if a higher value of ε is selected for CB-PA-DDS. The empirical CDF plots for the sharpness metric are not shown as this metric decreases when the number of behavioral solutions increases by increasing ε values. However, the aggregate performance of sharpness and reliability, which is their Euclidean distance from their ideal values of 100%-100% (Shafii et al., 2015), is either similar for different ε values or better for higher ε values (not shown here). Therefore, high ε values of 0.45 and 0.55 are decided for the CB-PA-DDS technique to estimate parameter uncertainty in the HBV-SASK and RAVEN models since higher reliability and number of behavioral solutions are of interest.



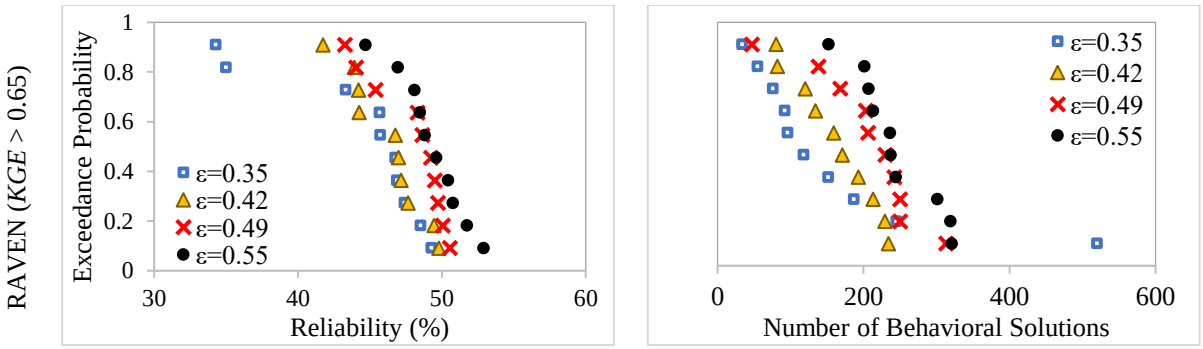


Figure 4-1: empirical CDF plot comparison of different spatial clustering resolutions of CB-PA-DDS for uncertainty estimation of HBV-SASK (top row) and RAVEN (bottom row) model calibration case studies, respectively, with *KGE* behavioral threshold values of 0.67 and 0.65

4.4.1.1 HBV-SASK Model

CB-PA-DDS with ε equal to 0.45 is compared to DDS-AU and GLUE in Figure 4-2 for two behavioral thresholds in the calibration and evaluation periods. Table 4-5 presents the p-values comparing CB-PA-DDS and GLUE (P_G), and p-values comparing CB-PA-DDS and DDS-AU (P_D). P_G and P_D values of less than 0.05 reject the null hypothesis, assuming that the median of a metric for CB-PA-DDS is equal to the median of the same metric for GLUE and DDS-AU, respectively.

A low *KGE* value of 0.67 is first considered for identifying behavioral solutions in each uncertainty method. This threshold is chosen based on the maximum value among the lowest *KGE* values in 10 runs of DDS-AU since this method archives at most one solution per trial, i.e. 100 solutions per run in this research. According to their CDF plots in calibration and evaluation periods, all tested uncertainty estimation methods result in relatively high reliability. DDS-AU results in a wider reliability range of (96%-100%) compared to GLUE and CB-PA-DDS that have a higher than 99.6% reliability in all their trials. CB-PA-DDS stochastically dominates DDS-AU in terms of reliability and size of the behavioral set with P_D values less than 0.05. Despite having very similar reliability, CB-PA-DDS offers a higher number of behavioral solutions with higher sharpness than

GLUE in an equal computational budget, i.e. 280-536 solutions versus 214-266 solutions and a sharpness index of 77.1%-79.3% versus 74.85%-77.04%, respectively (see Figure 8-2 for sharpness CDFs). P_G value of less than 0.05 in Table 4-5 for the size of behavioral set and sharpness plots presented in Figure 4-2 and Figure 8-2 confirms the superior performance of CB-PA-DDS over GLUE in the calibration period.

When the uncertainty estimation methods are compared for the evaluation period, CB-PA-DDS and GLUE with 10000 solution evaluations hold their reliability index values above 99.6% but a sharper uncertainty band for the CB-PA-DDS method, a P_G value of 0.0017. The lowest *KGE* values of the evaluation period among all behavioral sets identified by each method are also compared in Figure 4-2-*f* and Figure 4-2-*h*. The behavioral sets identified by DDS-AU contain solutions that have negative *KGE* in the evaluation period. A negative *KGE* in DDS-AU may be due to conditioning the simulation to overfit the calibration period. This can be resolved by increasing the number of optimization trials and decreasing the number of solution evaluations per trial, e.g. 200 independent searches, each with 50 solution evaluations resulting in up to 200 archived solutions. Therefore, *a priori* setting of the DDS-AU structure requires a profound understanding of the hydrologic model to prevent overconditioning in a well-behaved model like HBV-SASK.

When the behavioral threshold is increased to a very high *KGE* value, the number of behavioral solutions and reliability become more important than sharpness. A low number of behavioral solutions results in a less robust estimate of prediction uncertainty bounds (Blasone et al., 2008b). Moreover, sharpness is significantly affected by the number of behavioral solutions. For example, when the behavioral threshold is set to the *KGE* value of 0.8, GLUE results in a sharper uncertainty band calculated based on 22-31 behavioral solutions compared to the band identified by 47-158

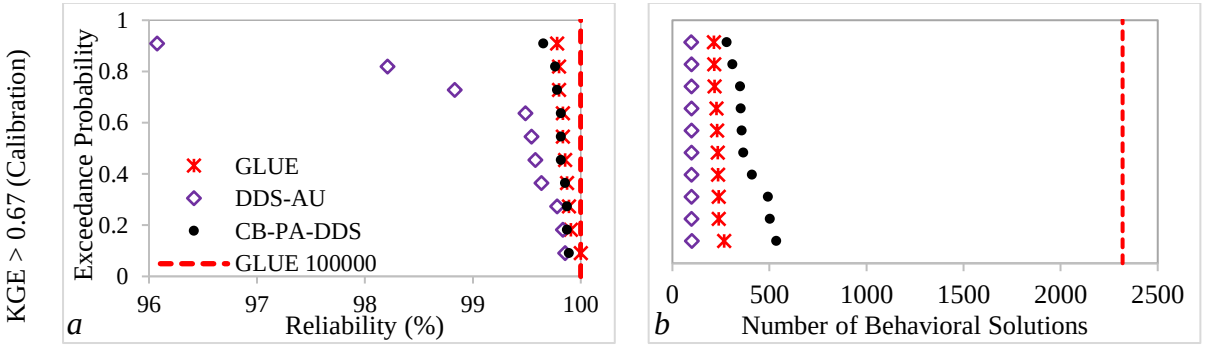
behavioral solutions in CB-PA-DDS. Therefore, a better sharpness in GLUE due to a low number of behavioral solutions does not mean that it performs better. Besides a larger behavioral set, CB-PA-DDS stochastically dominates GLUE in the calibration and evaluation periods, with respective p-values of 0.0452 and 0.0211 (see Table 4-5), in terms of reliability when these methods have an equal computational budget, see Figure 4-2-*c* and Figure 4-2-*g*.

More than 75% of archived solutions in DDS-AU have *KGE* values higher than 0.8. The reliability CDFs associated with DDS-AU stochastically dominate those of CB-PA-DDS for the calibration and evaluation periods. However, it comes at the cost of a significantly reduced sharpness (see Figure 8-2) in the entire simulation period and negative *KGE* values in the evaluation period. A negative *KGE* value means that the prediction band generated by DDS-AU for the uncertainty estimation in the HBV-SASK model is misleading. Hence, high-reliability indices are affected by solutions that overcondition the model simulations for the calibration but perform poorly in the evaluation period.

The 10 behavioral sets associated with 10 trials of GLUE with 10000 solution evaluations are combined in another experiment that is equivalent to one trial of GLUE with 100000 solution evaluations. The uncertainty performance metrics for the GLUE method with 100000 solutions are displayed as a vertical red dashed line. The combined behavioral set contains 2319 and 270 solutions for the *KGE* behavioral thresholds of 0.67 and 0.80, respectively. This GLUE run achieves more behavioral solutions in a relatively high *KGE* threshold of 0.8 compared to CB-PA-DDS, DDS-AU, and GLUE with an order of magnitude less computational budget, 10000 solution evaluations.

4.4.1.2 RAVEN Model

Based on the results, two behavioral likelihood thresholds of *KGE* equal to 0.65 and 0.70 are considered. GLUE achieves the lowest number of behavioral solutions at the lower threshold of 0.65 compared to the other two uncertainty estimation methods. At this behavioral threshold, DDS-AU achieved almost twice as many solutions as GLUE achieved, but Figure 4-3-*a* and Figure 4-3-*e* show that their reliability CDF plots are intertwined, respectively for the calibration and evaluation periods. However, the sharpness in Figure 8-2-*d* in the appendix shows a dominating performance for GLUE for the evaluation period meaning that the higher number of behavioral solutions in the DDS-AU only widens the uncertainty band with no added coverage of streamflow time series compared to GLUE. The reason could be due to a higher similarity among behavioral solutions found by DDS-AU. On the other hand, CB-PA-DDS generates a significantly high number of behavioral solutions, i.e. 152-321 solutions in 10 trials. As shown in Figure 4-3-*a* and Figure 4-3-*e*, CB-PA-DDS dominates GLUE and DDS-AU in terms of reliability with p-values significantly less than 0.05 (shown in Table 4-5) in calibration and evaluation periods when the *KGE* threshold is set to 0.65.



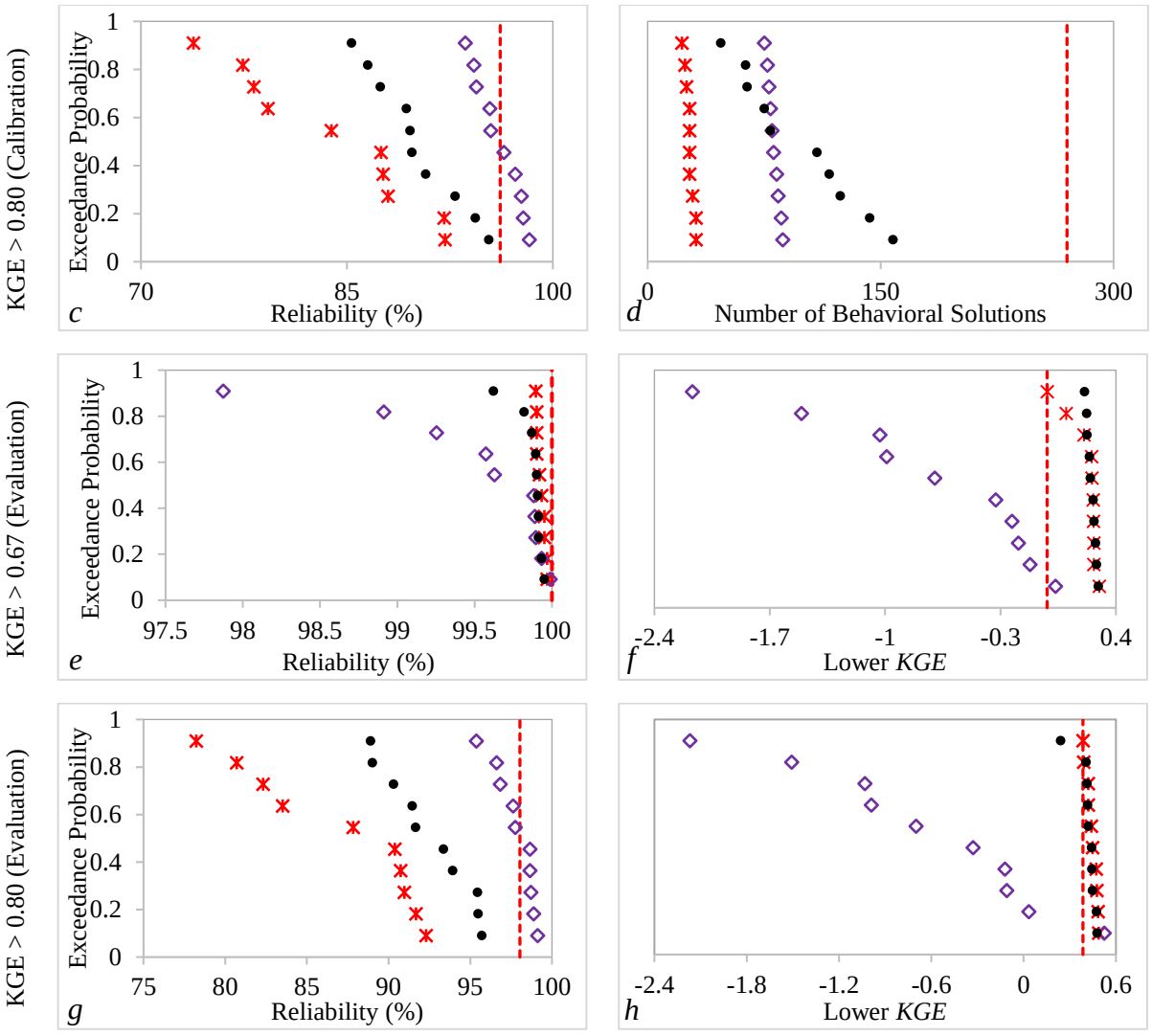
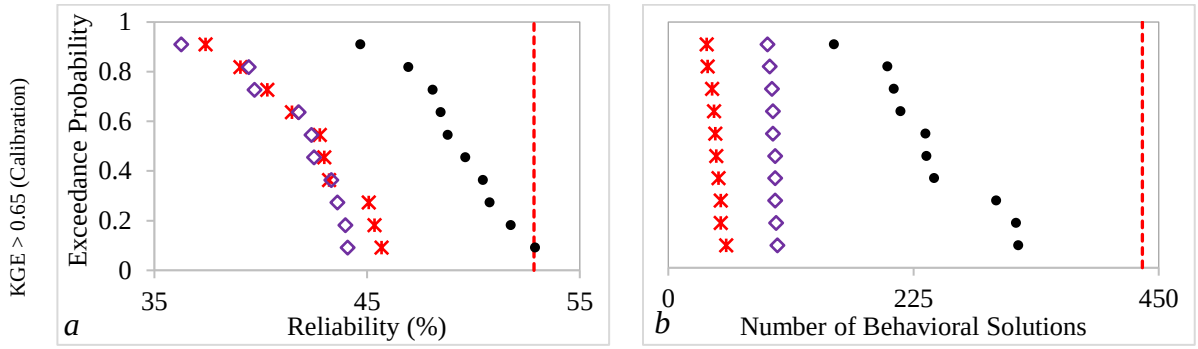


Figure 4-2: CDF plot comparison of CB-PA-DDS ($\varepsilon = 0.45$) with DDS-AU, GLUE (10000), and GLUE (100000) for uncertainty estimation with two behavioral thresholds of 0.67 and 0.80 in calibration and evaluation periods for the 12-parameter HBV-SASK model. The higher the value of a metric, the better the performance.

CB-PA-DDS and DDS-AU become comparable in a higher behavioral threshold and their CDF plots exhibit some crossing behavior, see panels resulting from the *KGE* threshold of equal to 0.70 in Figure 4-3. However, a slight relaxation of 0.05 in the behavioral threshold significantly shifts (improves) the empirical CDF of reliability and size of behavioral set for CB-PA-DDS to the right side of DDS-AU, compare panels *c* and *d* respectively with panels *a* and *b* in Figure 4-3. Therefore, CB-PA-DDS is preferred over DDS-AU. GLUE with 100000 computational budget (shown as red

dashed line in Figure 4-3) at the relaxed behavioral threshold of 0.65 outperforms DDS-AU and CB-PA-DDS. Still, the CDFs associated with CB-PA-DDS have the shortest distance to it.

Selecting a tight behavioral threshold of *KGE*=0.70 deteriorates the GLUE performance because it identifies between 2 and 11 behavioral solutions out of 10000 solutions evaluated in each trial. In this situation, the uncertainty interval is very narrow (see sharpness CDFs in panels *f* and *h* in Figure 8-2), and the reliability of solutions found by GLUE in Figure 4-3-*c* and Figure 4-3-*g* approaches zero. Therefore, the prediction uncertainty band of the RAVEN model parameters using the small behavioral set of GLUE is not recommended (Blasone et al., 2008b). At this tight behavioral threshold, when the archives of 10 trials of GLUE are combined to theoretically form one trial of GLUE with 100000 solution evaluations, its performance becomes similar to one trial of DDS-AU or CB-PA-DDS with one order of magnitude less computational budget, see red dashed lines in Figure 4-3 for the calibration period in panels *c* and *d* and the evaluation period in panels *g* and *h*.



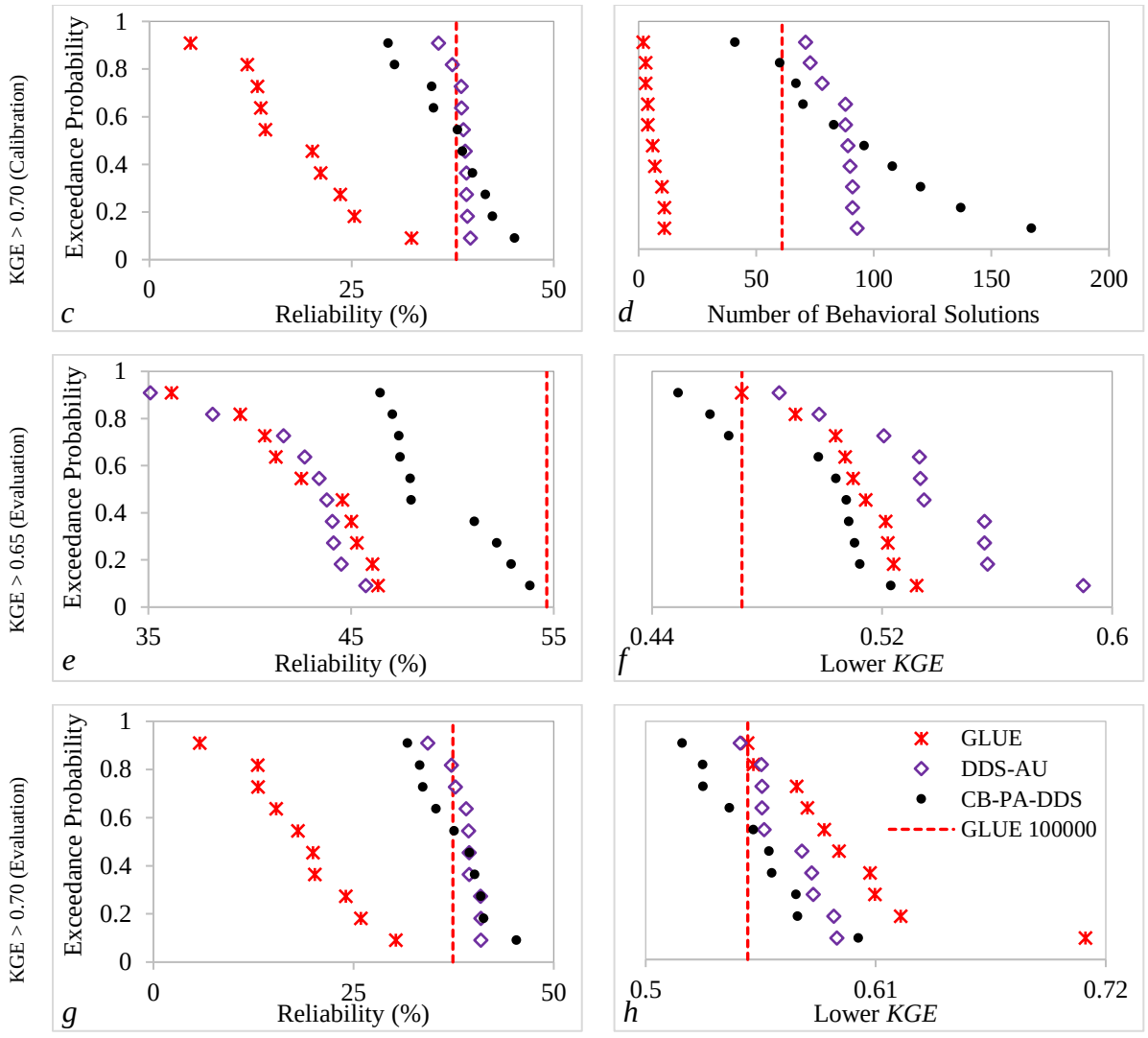


Figure 4-3: CDF plot comparison of CB-PA-DDS ($\varepsilon = 0.55$) with DDS-AU, GLUE (10000), and GLUE (100000) for uncertainty estimation with two behavioral thresholds of 0.65 and 0.70 in calibration and evaluation periods for the 20-parameter RAVEN model. The higher the value of a metric, the better the performance.

Table 4-5: P_G and P_D, respectively representing p-value for the HBV-SASK and RAVEN model calibration comparing CB-PA-DDS to GLUE and DDS-AU based on the median of four metrics: reliability, number of behavioral solutions, lower bound KGE, and sharpness. P-values less than 0.05 are bolded and show one algorithm is statistically significantly preferred over the other algorithm.

			P_G / P_D			
Model	**Threshold**	**Period**	**Reliability**	**Number of Behavioral Solutions**	**Lower Bound KGE**	**Sharpness**
HBV-SASK	KGE ≥ 0.67	Calibration	0.2876 / **0.0089**	**1.83e-4 / 1.59e-4**	-	**1.83e-4** / 0.3847
		Evaluation	0.0940 / 0.1208	-	0.6776 / **1.83e-4**	**0.0017** / 0.3075
	KGE ≥ 0.80	Calibration	**0.0452 / 7.69e-4**	**1.72e-4** / 0.7912	-	**0.0091 / 3.30e-4**

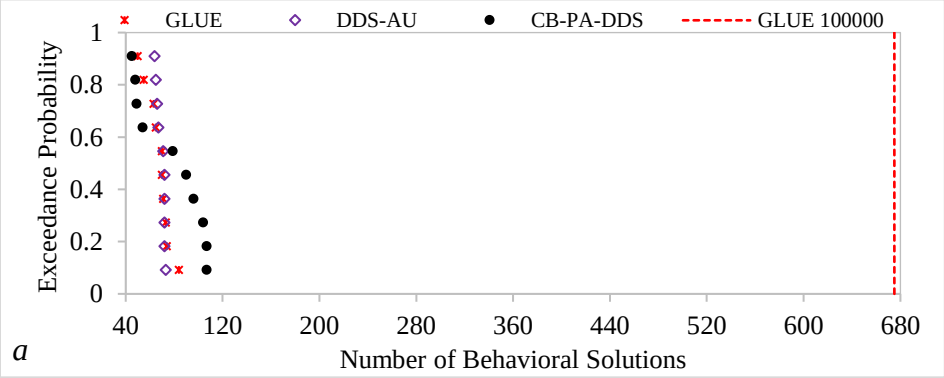
		Evaluation	**0.0211 / 4.40e-4**	-	0.5205 / **0.0028**	**0.0046 / 5.83e-4**
RAVEN	KGE ≥ 0.65	Calibration	**1.83e-4 / 3.30e-4**	**1.82e-4 / 1.78e-4**	-	**1.83e-4 / 3.30e-4**
		Evaluation	**1.83e-4 / 1.83e-4**	-	0.1620 / **0.0073**	**1.83e-4 / 2.46e-4**
	KGE ≥ 0.70	Calibration	**3.30e-4 /** 0.7913	**1.80e-4 /** 0.850	-	**1.83e-4 /** 1.0
		Evaluation	**1.83e-4 /** 0.7910	-	**0.0140 /** 0.1618	**1.83e-4 /** 0.2730

4.4.2 Methodology Validation

Selecting a proper ε value for a computationally intensive case study like the SWAT model of the Rouge River is crucial. According to the methodology testing of CB-PA-DDS, a value of 0.47 is used for ε as this case study has the same computational budget and a similar number of parameters with the HBV-SASK model. In this study, a solution is flagged as behavioral when the *NSE* is higher than 0.5, the error in different segments of the flow-duration curve is less than 30%, and a water balance constraint violation is less than 30% at both hydrometric stations (see Table 4-4). Figure 4-4-*a* shows that, on average, all algorithms have a similar performance in this case study. Although CDF plots of the number of behavioral solutions identified by GLUE and CB-PA-DDS cross each other (Figure 4-4-*a*), CB-PA-DDS shows a dominating performance in terms of reliability (coverage of streamflow time series) with a p-value of 0.0091 in Figure 4-4-*b* for the calibration period at station 02HC022. A similar dominating behavior (but not statistically significant) is observed for CB-PA-DDS compared to GLUE at station 02HC028 in Figure 4-4-*c*. Figure 4-4-*b* and Figure 4-4-*c* also show a better coverage by CB-PA-DDS results than DDS-AU for the calibration period at both stations, but this difference is not statistically significant according to the p-values in Table 4-6. DDS-AU possesses a sharper uncertainty interval (Figure 8-3-*a* and Figure 8-3-*c* in the appendix) for calibration. Despite having similar reliability (panels *b* and *d* of Figure 4-5) and sharpness (see panels *b* and *d* in Figure 8-3) in the evaluation period,

DDS-AU produces solutions that have a higher range of *NSE* (see panels *a* and *c* of Figure 4-5) compared to CB-PA-DDS.

Figure 4-6 describes the distribution of behavioral solutions in the parameter space for a trial whose number of behavioral solutions is closest to the average size of behavioral set in 10 trials for each algorithm, 79, 71, and 70 solutions respectively for CB-PA-DDS, DDS-AU, and GLUE out of 10000 solution evaluations. Overall, all methods found a fairly diverse set of solutions that are spread across the parameter space. Figure 4-7 displays the 95% uncertainty band for a representative portion of the evaluation period (the year 2011) at station 02HC022, downstream of station 02HC028, for the sets of behavioral solutions shown in Figure 4-6. The 95% uncertainty band in this figure is obtained from the quantiles of the estimated cumulative likelihood distribution. The likelihood value for each behavioral solution is estimated based on the aggregated objective function in equation (28) divided by the sum of this objective value for all behavioral solutions. The cumulative distribution function for each time step is then calculated by sorting the weights based on the simulated streamflow values in each time step over all behavioral solutions. Then, the 95% predictive uncertainty limit for the SWAT model is derived from the 2.5 and 97.5 percentiles of this distribution for each time step.



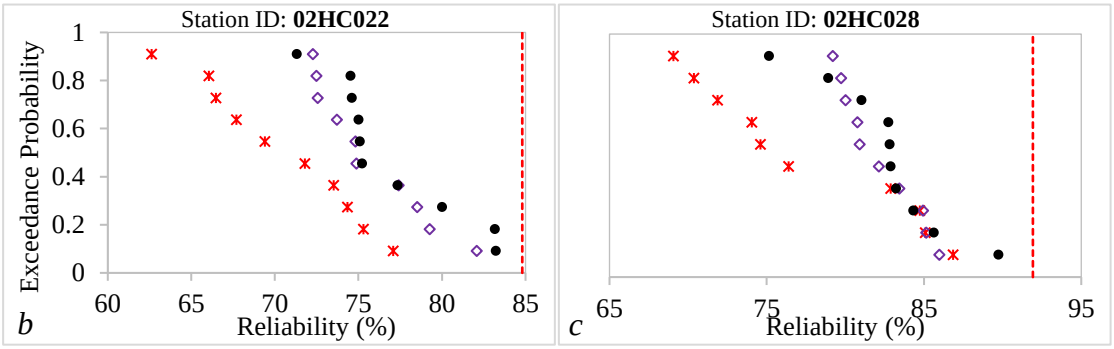
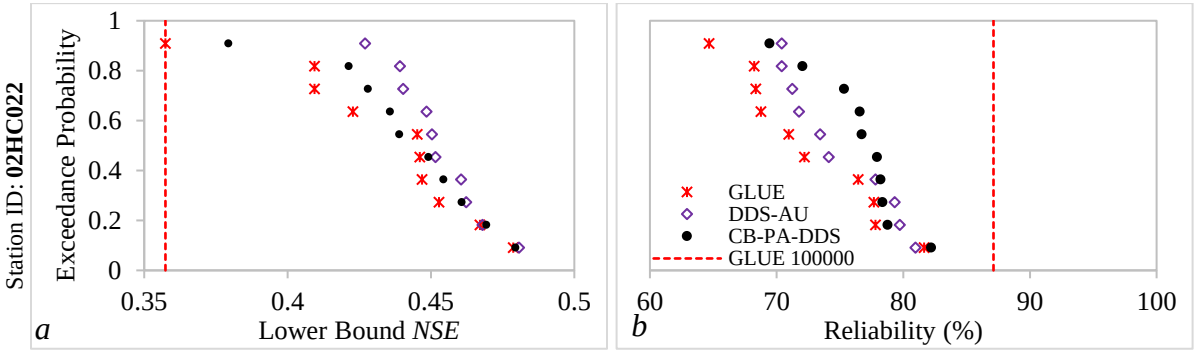


Figure 4-4: CDF plot comparison of CB-PA-DDS ($\varepsilon = 0.47$) with DDS-AU, GLUE (10000), and GLUE (100000) for uncertainty estimation in calibration period for the 13-parameter SWAT model. The higher the value of a metric, the better the performance.

Table 4-6: P_G and P_D, respectively representing p-value for the SWAT model calibration comparing CB-PA-DDS to GLUE and DDS-AU based on the median of four metrics: reliability, number of behavioral solutions, lower bound *NSE*, and sharpness. P-values less than 0.05 are bolded and show one algorithm is statistically significantly preferred over the other algorithm.

Hydrometric Station	Period	P_G / P_D Reliability	Number of Behavioral Solutions	Lower Bound *NSE*	Sharpness
02HC022	Calibration	**0.0091** / 0.4274	0.4723 / 0.4708	-	1.0 / **0.0173**
	Evaluation	0.0640 / 0.4725		0.5706 / 0.3847	0.6232 / 0.1859
02HC028	Calibration	0.1735 / 0.7337	0.4723 / 0.4708	-	0.8501 / **0.0376**
	Evaluation	0.0889 / 0.5706		0.3845 / **0.0073**	0.9097 / 0.0640



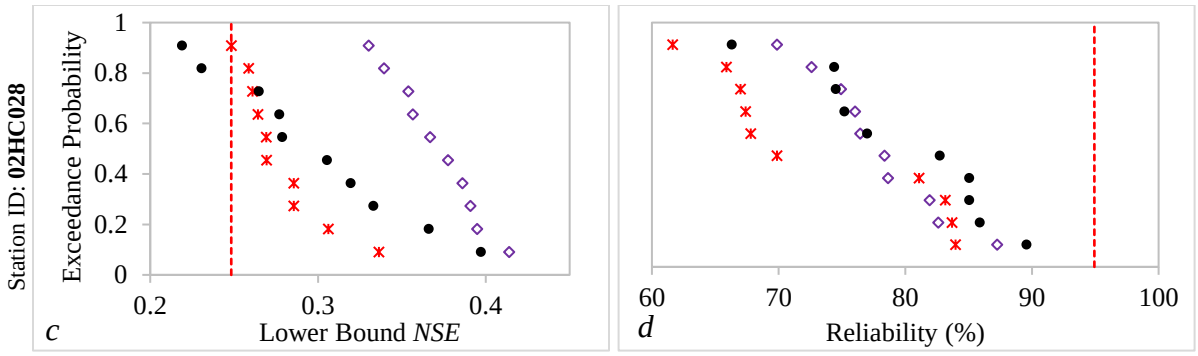


Figure 4-5: CDF plot comparison of CB-PA-DDS ($\varepsilon = 0.47$) with DDS-AU, GLUE (10000), and GLUE (100000) for evaluating their performance in the evaluation period for the 13-parameter SWAT model. The higher the value of a metric, the better the performance.

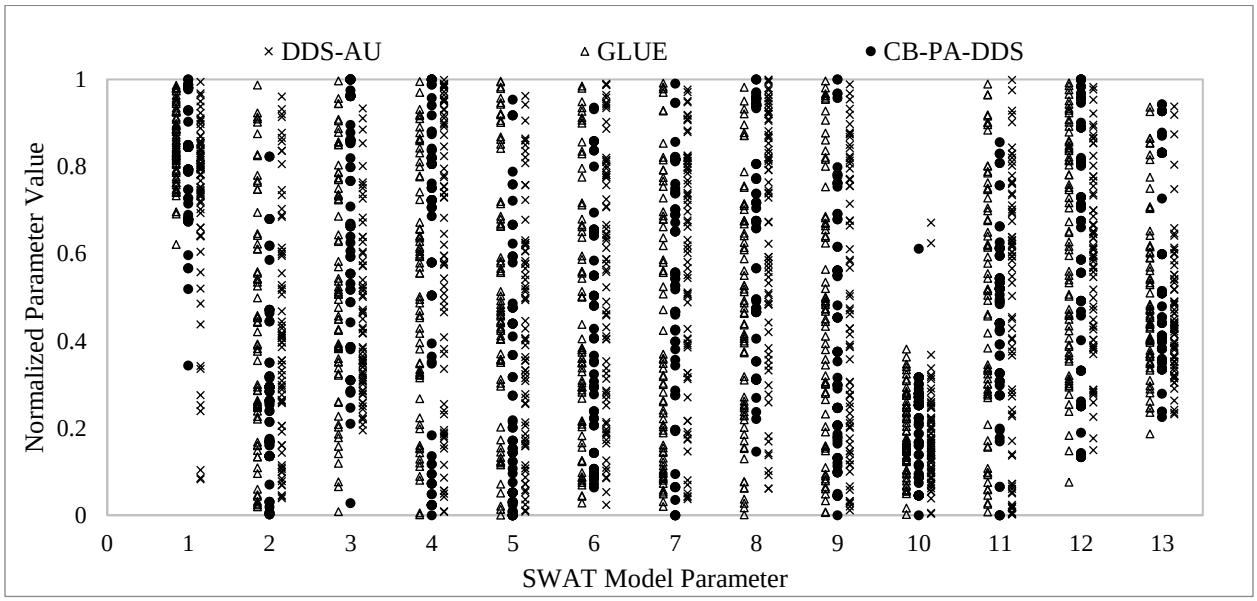


Figure 4-6: Distribution of behavioral solutions identified by each algorithm in the normalized parameter space. The mid-performing trial in terms of the identified number of behavioral solutions is used for this figure.

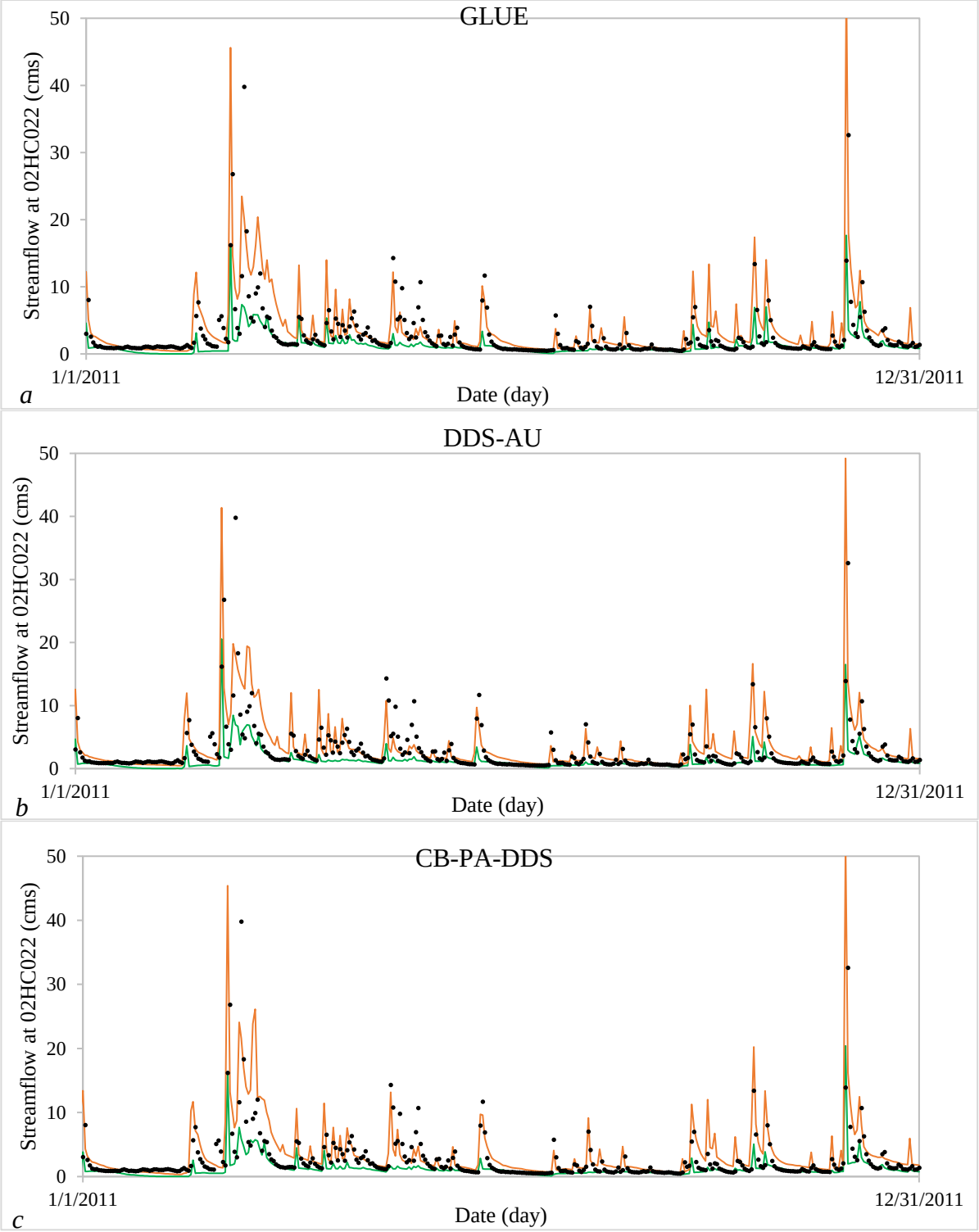


Figure 4-7: Simulated 95% uncertainty interval of the SWAT model parameters for 2011 (last year of the evaluation period) at station 02HC022. The bands in panels *a*, *b*, and *c* respectively belong to a mid-performing trial in terms of the identified size of behavioral set in GLUE, DDS-AU, and CB-PA-DDS. The continuous lines show the upper bound and lower bound of the 95% uncertainty interval.

4.4.3 Multi-Objectivity and Behavioral Solution Identifiability

Adding more criteria to the model evaluation process adds more number of behavioral thresholds and, therefore, shrinks the acceptability region and decreases the expected number of behavioral solutions. Figure 4-8 exhibits the effect of the number of calibration objective functions that dictate the number of behavioral thresholds on the size of behavioral set using a Box and Whisker plots. Considering a single criterion and threshold for identifying behavioral solutions such as $f_1 < 0.036$ for HBV-SASK, $f_1 < 0.040$ for RAVEN, or $\boldsymbol{A} = \{NSE_i > 0.5\}$ for SWAT in the first column of Figure 4-8 shows that GLUE finds more behavioral solutions than CB-PA-DDS and DDS-AU. However, adding a second criterion and the corresponding behavioral threshold, i.e. $\{f_1 < 0.036 \,\&\, f_2 < 0.036\}$ for HBV-SASK, $\{f_1 < 0.040 \,\&\, f_2 < 0.040\}$ for RAVEN, and $\boldsymbol{B}$: $\{NSE_i > 0.5 \,\&\, E_{HF,i} < 30\%\}$ for SWAT, removes more than 50% of the solutions that were considered behavioral in the single-objective analysis for both GLUE and CB-PA-DDS that do not have a cap on the number of behavioral solutions they can archive. Adding a third criterion or threshold further shrinks the region of acceptability (behavioral space) for all hydrologic models in the case of GLUE and CB-PA-DDS, but adding the fourth and fifth behavioral threshold in the SWAT model did not reduce the number of archived behavioral solutions noticeably.

The aggregation-based behavioral solution identification is likely to expand the region of acceptability. The last column in Figure 4-8 shows the number of archived solutions when all important criteria are aggregated into one. When the behavioral solution identification resulting from simultaneous consideration of all criteria are compared to the aggregation-based identification in the case of GLUE, for example, the size of behavioral solution set decreases significantly by 65%-75% in HBV-SASK, 80%-98% in RAVEN models (compare the last two columns in Figure 4-8). CB-PA-DDS has similar behavior to GLUE when the criteria are

aggregated for behavioral solution identification. In this circumstance, solutions that are behavioral with respect to one or two error metrics but do not meet the other threshold may still meet the threshold for the aggregated formulation, which is not ideal for multi-objective uncertainty analysis.

As more calibration objectives are aggregated, the size of the behavioral set is expected to increase relative to multi-objective uncertainty estimation. For example, the difference in the size of behavioral sets for the five-objective SWAT model uncertainty estimation with and without aggregation becomes more significant (compare the last two columns in Figure 4-8-*c*). DDS-AU is less sensitive to the number of thresholds. The main reason is that its archive size is limited to a maximum that is the number of trials, 100 in this study. Therefore, all generated behavioral solutions in one trial of DDS-AU are not retained, and only the best performing behavioral solution, if found any, is archived for uncertainty estimation.

Since the behavioral space shrinks as more behavioral thresholds are considered, the multi-objective uncertainty estimation is expected to require a higher computational budget to find more behavioral solutions. Unlike DDS-AU, CB-PA-DDS is a less subjective approach that tends to identify the entire range of optimal and near-optimal tradeoffs clustered at different regions of the parameter space with no discrimination between the objectives. Lumping objectives in DDS-AU is equivalent to converging to only one point of these distinct optimal and near-optimal tradeoffs in each search. Extracting the entire range of optimal and near-optimal tradeoffs for the five-objective calibration problem in CB-PA-DDS demands higher computational budgets than DDS-AU. If a higher budget is not feasible, the analyst can reduce the objective-space dimension to improve the performance or expand the behavioral (i.e. acceptable or feasible) space by relaxing the behavioral threshold to include more solutions to the behavioral set. Thus, despite being more

computationally intensive, we recommend multi-objective approximation of uncertainty in model calibration and assigning the corresponding behavioral threshold to each objective as it is more consistent with the multi-objective nature of model calibration, Gupta et al. (1998).

4.5 Conclusion

Uncertainty quantification tools can assist decision-makers in assessing the credibility of hydrologic models and improve flood management reliability. Flood management systems have different forecasting requirements that cannot be represented by one simple objective function. In this paper, CB-PA-DDS is proposed as a multi-objective parameter uncertainty method and compared with popular single-objective DDS-AU and GLUE algorithms for uncertainty analysis of three hydrologic models. CB-PA-DDS is equipped with dynamic clustering and localized dominance strategy to help search and retain distinct groups of solutions in different regions of the parameter space, each of which form a tradeoff in the objective space. The performance of the proposed method depends on the adjacency density parameter (ε) for clustering. The 12-parameter HBV-SASK and 20-parameter RAVEN models were used for the methodology development by evaluating the performance of multiple settings of CB-PA-DDS by changing the ε value. The results indicated that setting a higher value of ε leads to better performance. Therefore, the upper bound value of ε was tested for comparing CB-PA-DDS with DDS-AU and GLUE in HBV-SASK and RAVEN, then validated its performance on 13-parameter five-objective SWAT model.

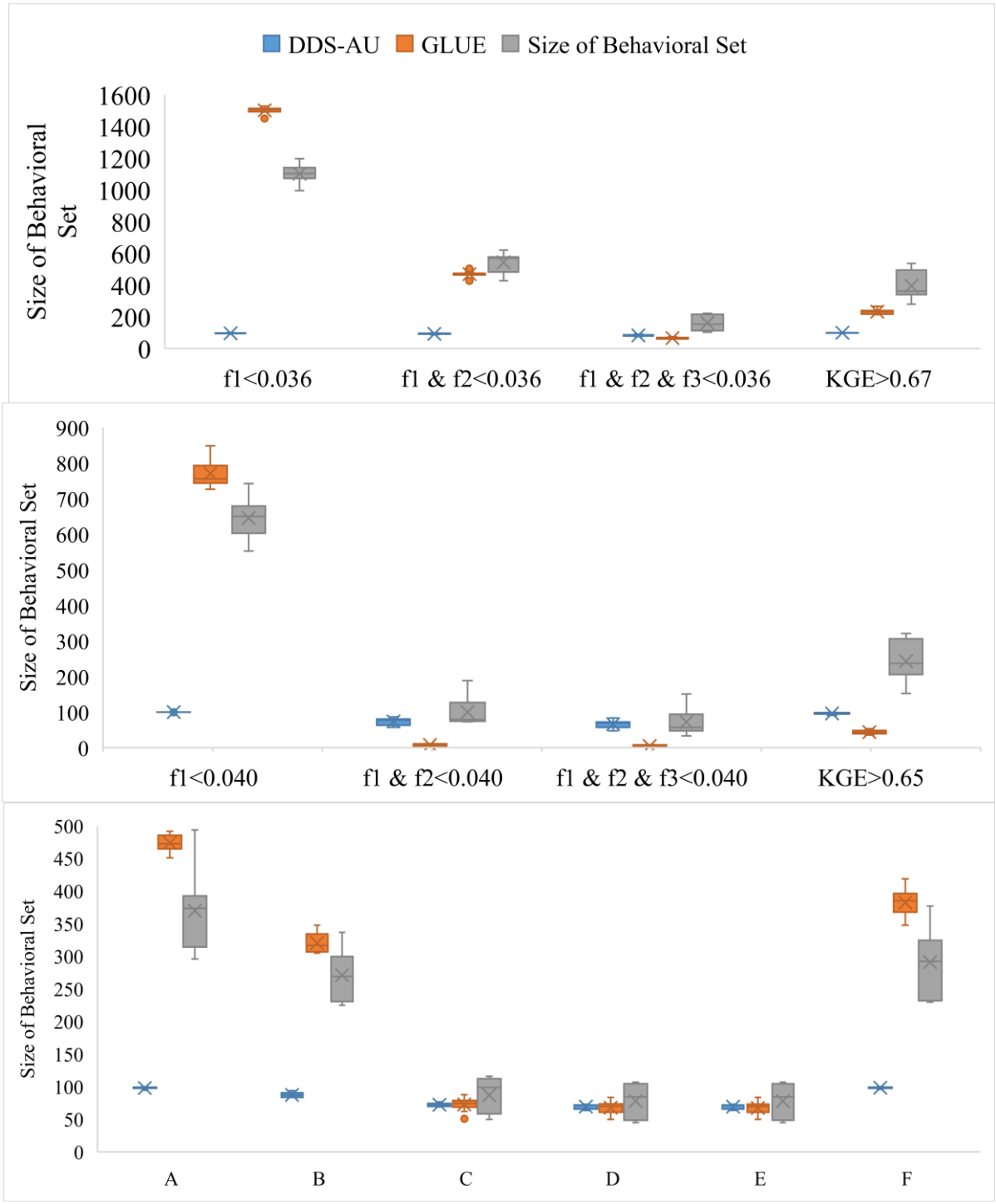


Figure 4-8: The effect of an increase in the number of objectives on the identification process and number of behavioral solutions. The box and whisker plot on the top, middle, and bottom respectively belong to HBV-SASK, RAVEN, and SWAT models. f_1, f_2, f_3 are the decomposition terms of *KGE*, as defined in equation (25). A, B, C, D, E, and F on the bottom figure are defined as follows. A: $\{NSE_i > 0.5\}$, B: $\{NSE_i > 0.5$ & $E_{HF,i} < 30\%\}$, C: $\{NSE_i > 0.5$ & $E_{HF,i} < 30\%$ & $E_{MF,i} < 30\%\}$, D: $\{NSE_i > 0.5$ & $E_{HF,i} < 30\%$ & $E_{MF,i} < 30\%$ & $E_{LF,i} < 30\%\}$, E: $\{NSE_i > 0.5$ & $E_{HF,i} < 30\%$ & $E_{MF,i} < 30\%$ & $E_{LF,i} < 30\%$ & $f < 30\%\}$, F: Aggregated Objective as in Equation (28).

Multiple performance metrics, including reliability, sharpness, and size of behavioral solution set, were used for comparison between the uncertainty methods in each behavioral threshold. CB-PA-DDS has a similar (or better) performance to DDS-AU, but both methods are likely to outperform GLUE as they are equipped with specialized sampling strategies. One issue with DDS-AU for uncertainty analysis is that it only retains at most one behavioral solution per search, and other potential behavioral solutions, if any, are removed in each search. Therefore, its performance depends on the prior specifications of DDS-AU archive size.

The aggregation-based behavioral threshold identification expands the region of acceptability, identifies more number of solutions as behavioral, and overestimates the prediction uncertainty compared to multiple-threshold behavioral solution identification. GLUE and DDS-AU aggregate the objectives in a multi-objective problem and the value of each objective for the archived solutions in these methods is not available after their termination. Therefore, the behavioral solutions in these methods are identified through specifying a threshold for the aggregated objective function. This problem is prevented in CB-PA-DDS as this method simultaneously evaluates all objectives without aggregation.

We acknowledge that a multiple-threshold behavioral solution identification is not fair for the single-objective DDS-AU compared to GLUE and CB-PA-DDS since the perturbation and sampling processes in DDS-AU depend on the definition of the objective function.

The value of ε in CB-PA-DDS can theoretically go beyond one. For instance, two solutions in a hydrologic model with a 100 dimensional parameter space, which is normalized to between zero and one based on the feasible ranges of parameters, can have a distance up to $\sqrt{100}$ (or 10) that is equivalent to the diagonal distance of the unit hypercubic parameter space. Suppose a modeller

deals with a model with very high number of parameters. In that case, first, it is preferred to fix some of the less uncertain, physically-based parameters or group the correlated parameters logically to decrease the dimensionality of the hydrologic model. Second, it is preferred to increase the computational budget. For example, a 30-parameter model needs more than a billion function evaluations to effectively place at least one solution in each box in a gridded parameter space that divides each dimension by two. Since assigning a high computational budget in the order of millions or billions is not normally possible, the accuracy of all uncertainty estimation techniques is deteriorated. In this situation, the value of ε can be increased to beyond 0.5 to mitigate a low-computational budget in a time-consuming hydrologic simulation if a model has more than 15 parameters. It is also recommended to apply the uncertainty estimation techniques to models that have at most 30 uncertain parameters for a more accurate prediction uncertainty interval.

4.6 Acknowledgement

This work was supported by Dr. Asadzadeh's Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant [RGPIN-2016-05896]; and Graduate Enhancement of Tri-Council Stipends (GETS), University of Manitoba, Winnipeg, MB.

Conflict of Interest

The authors declare no real or perceived financial or conflict of interest with respect to the results and publication of this paper.

Data Availability Statement

The hydrologic models and the corresponding hydrometeorological data used in this study are adopted from Razavi et al. (2019) for HBV-SASK, Shafii et al. (2017) for RAVEN, and Asadzadeh

et al. (2016, 2015) for SWAT. DDS-AU and GLUE uncertainty methods used in this study are publicly available and archived by its code developer Dr. Shawn Matott on https://www.eng.buffalo.edu/~lsmatott/Ostrich/OstrichMain.html.

The Cluster-Based PA-DDS algorithm is a freeware available on http://home.cc.umanitoba.ca/~asadzadm/files/ClusterPADDS.zip.

5 SIGNATURE-BASED MULTI-MODELLING AND MULTI-OBJECTIVE CALIBRATION OF HYDROLOGIC MODELS: APPLICATION IN FLOOD FORECASTING FOR CANADIAN PRAIRIES

Shahram Sahraei[1], Masoud Asadzadeh[2], Fisaha Unduche[3]

[1] PhD Candidate, Department of Civil Engineering, University of Manitoba, Canada (Corresponding Author)

E-mail: sahraeis@myumanitoba.ca

[2] Assistant Professor, Department of Civil Engineering, University of Manitoba, EITC E1-332, 15 Gillson Street, Winnipeg, MB, Canada, R3T 5V6, Ph: +1 (204) 474 9535

E-mail: masoud.asadzadeh@umanitoba.ca

[3] Hydrologic Forecast Centre, Manitoba Infrastructure, Winnipeg, Canada

E-mail: Fisaha.unduche@gov.mb.ca

5.1 Abstract

Multi-modelling aims to make use of the strengths of single hydrologic models to improve the accuracy of simulating the watershed system behavior. Considering hydrological signatures such as the flow duration curve segmentation in the calibration of each hydrologic model leads to a better parameter identifiability. In this study, a novel weighted average model-wrapper based on flow duration curve segmentation is introduced to aggregate the calibrated models into a multi-model. The proposed framework is applied to develop a model-wrapper of the Upper Assiniboine

River Basin for flood forecasting upstream of the Shellmouth reservoir in the Prairie region of Canada. The HEC-HMS, HBV-EC, HSPF, and WATFLOOD hydrologic models that are being used at the Hydrologic Forecast Centre of Manitoba Infrastructure for operational inflow forecasting are calibrated using signature-based multi-objective optimization. These models have significantly different structural complexities. The calibration of each of these models is set up as three simulation-optimization problems with different objective functions to balance the model capability in simulating multiple important hydrological signatures. Results show that the model-wrapper outperforms each of the single calibrated models that are of operational use at Manitoba Infrastructure, e.g. *NSE* improved from 0.44 for the best individual model to 0.76 for the model-wrapper in the calibration period. Moreover, the weights associated with each hydrologic model component indicate the contribution rate of the individual models to the model-wrapper in high-flow, mid-flow, and low-flow portions of streamflow time series. Quantifying the contribution of each model component provides a deeper insight into model selection strategy, especially when a component has minimal or no contribution, e.g. HEC-HMS and HBV-EC in this paper, to the model-wrapper performance in all ranges of streamflow simulation compared to other model components.

Keywords: Multi-Modelling, Model-Wrapper, Multi-Objective Optimization, Hydrological Signature, Canadian Prairie, Flood Forecasting

5.2 Introduction

Hydrologic models are systems analysis tools that simulate our understanding of the hydrological processes and can help us reanalyze the historical events and estimate the watershed system responses to future hydrological events. These models can be categorized based on their structural complexity in terms of the distribution of the geographical representation of the watershed system

from lumped to distributed models and in terms of the base of their governing equations from conceptual to physically-based models. It is well-known that a single model structure cannot adequately represent all governing processes of a watershed system response to hydrological events (Ajami et al., 2007). Therefore, in order to address the model structural uncertainty in the model prediction, one can combine results of multiple models.

In principle, multi-modelling aims to take advantage of strengths of a group of models that have different structural complexities to improve the overall prediction accuracy, because it is expected that simulation error from different models can compensate each other (Ajami et al., 2006). Multi-modelling has its roots in the study by Bates and Granger (1969) who showed that the weighted average of two sets of forecasted airline passenger datasets outperforms each individual dataset for forecasting the number of passenger miles flown. It took about three decades for the community of hydrologic modelling to recognize multi-modelling as a tool to improve the model prediction accuracy, when Shamseldin et al. (1997) showed that even a simple or a weighted linear model averaging approach can outperform the individual models, a conclusion that was confirmed by Ajami et al. (2006) and Georgakakos et al. (2004). To date, several studies have investigated the superior performance of multi-model ensembles to individual hydrologic models (Arsenault et al., 2015; Chen et al., 2015; Goswami et al., 2007; Jeong and Kim, 2009; Kumar et al., 2015; Xiong et al., 2001; X. Zhang et al., 2009). Shamseldin et al. (1997) recommended a neural network approach to build a non-linear relationship between the models to improve the prediction accuracy. Ajami et al. (2006) compared four different multi-modelling approaches, simple model average, weighted model average (WMA), multi-model super-ensemble (MMSE), and modified multi-model super-ensemble (M3SE) methods and recommended MMSE and M3SE that implement bias correction. However, Madadgar et al. (2014) argued that the bias correction through quantile

mapping (used in M3SE for example) that uses the cumulative distribution function of measured and simulated events does not maintain the sequence of events and therefore is not hydrologically expressive. Bayesian model averaging (BMA) introduced by Hoeting et al. (1999) is a more sophisticated multi-modelling approach that uses the Bayesian statistical inference to calculate the weight of each model based on its performance during the training period (Ajami et al., 2007). In this paper, a novel model-wrapper approach is introduced for multi-modelling based on the weighted model averaging and flow duration curve partitioning. This method allows the models for having different weights for predicting high-, average-, and low-flow rates. The optimal value of weights shows whether every model adds to the accuracy of flood forecasting. The optimal value of weights also show the relative contribution of these models to the flow prediction for three mutually exclusive ranges of high-, mid-range, and low-flow. Moreover, the comparison between individual models and the model-wrapper shows whether multi-modelling could fix any model structural inadequacy.

The hydrologic models need to be calibrated individually before inserting their simulated streamflow time series into the model-wrapper. The performance of a calibrated hydrologic model depends on many factors including its structural complexity, parametrization, computational budget and problem formulation. In principle, due to the model parameter uncertainty, more than one setting of a hydrologic model can adequately simulate the watershed system behavior, the concept of equifinality (Beven and Freer, 2001). Therefore, in practice, different settings of a hydrologic model are used to predict a range of watershed responses to future events, either through a formal Bayesian uncertainty approach by DREAM (Vrugt et al., 2009), an informal uncertainty approach by Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and

Freer, 2001), or a multi-trial calibration with different calibration objectives such as Tolson and Shoemaker (2008).

Hydrologic model calibration is inherently multi-objective, because no single metric can comprehensively represent the distribution of simulation error (residual between the simulated and measured data) (Gupta et al., 1998). Over the past decade, novel model performance metrics are developed and used in the literature as model calibration metrics that can guide the optimization to identify solutions that are more representative of hydrological signatures, see for example Pfannerstill et al. (2014), Schaefli (2016), Shafii and Tolson (2015), Yilmaz et al. (2008). Yilmaz et al. (2008) advised that the automatic calibration of hydrologic models should consider objective functions that represent four major hydrological signatures: the overall water balance, vertical, horizontal, and temporal redistribution of water. They introduced three model performance metrics using the flow duration curve partitioning. Multi-objective optimization identifies the tradeoff among the conflicting objectives, the so-called Pareto front. Solutions corresponding to the tails of the Pareto front individually optimize a calibration objective; however, as discussed in Asadzadeh et al. (2014b), Kollat et al. (2012), the multi-objective optimization can balance between the conflicting calibration objectives and identify a solution corresponding to the so-called knee point of the Pareto front that is closest to the ideal point and therefore can be selected as the best single tradeoff solution.

This study expresses the importance of using the signature-based, multi-objective calibration approach along with multi-modelling by combining the strong points of individual models to improve the calibration of the models for accurate forecasting of the streamflow with a focus on peak flow estimation. The paper is structured as follows. Section 5.3 describes the case study watershed and the hydrologic models that are used for the streamflow simulations. Three different

types of formulations for the multi-objective calibration of the models along with the proposed model combination approach are then explained. The calibration results for each model and the effect of the model-wrapper on simulation accuracy are presented and discussed in Section 5.4, followed by study limitations and concluding remarks in Section 5.5 and Section 5.6, respectively.

5.3 Materials and Methods

5.3.1 Importance of Flood Forecasting in Canadian Prairies

Flood forecasting is one of the major applications of hydrologic models in the Canadian Prairie Provinces (Ahmari et al., 2016; Blais et al., 2016b, 2016a) where flood is one of the most costly natural disasters. Over the past 100 years, this region has experienced significant floods. The Assiniboine River in Manitoba alone, for example, has experienced at least six major floods in the past 100 years (Manitoba Infrastructure and Transportation, 2016). Flooding has been worse in the past 20 years due to the wet-cycle going through the provinces (Unduche et al., 2018) causing billions of dollars infrastructural damages (Buttle et al., 2016; Jakob and Church, 2011). The 2013 flood in Alberta, with estimated damage and recovery cost of CAD 6 billion (Burn and Whitfield, 2016; Pomeroy et al., 2016), and the 2011 flood in Manitoba, with estimated damage and recovery cost of CAD 1.2 billion (Burn and Whitfield, 2016; Manitoba Infrastructure and Transportation, 2016) are few of the recent examples to mention. In order to better prepare for future flood events and mitigate the damages, the provincial and federal governments have initiated reviews of past major floods. In Manitoba, the 2011 flood review task force was established to assess the cause of the 2011 flood, flood recovery procedures and flood forecasting methodologies (Manitoba Infrastructure and Transportation, 2013). A similar task force was established in Alberta to review the 2013 historic flood (Alberta WaterSMART, 2013; Flood Recovery Task Force, 2013). Both of these reviews emphasized the need for accurate and timely dissemination of flood forecasts to the

public and policy and decision makers. Accurate flood forecasting can minimize infrastructural damages and the economic and social impacts of floods to communities and can result in measures that will enhance ecological conditions (Carsell et al., 2004; Penning-Rowsell et al., 2000).

5.3.2 Region of Study

The Upper Assiniboine River Basin at Kamsack in the Prairie Region of Canada is selected for this study, see Figure 5-1. It has a total drainage area of 13,000 km*2* that flows into the Shellmouth Reservoir, also known as Lake of the Prairies. The Shellmouth reservoir, which is retained by the Shellmouth dam, is used for multiple purposes including flood control, municipal and irrigation water supplies, and recreation. The operation of the dam control structure at the outlet of the reservoir is assisted by inflow forecasts that are obtained by simulating multiple hydrologic models (Unduche et al., 2018). The Upper Assiniboine River Watershed is dominated by abundant potholes, sloughs, and wetlands. These potholes and wetlands significantly increase the surface storage capacity through the fill-and-spill process in potholes during snowmelt and/or rainfall periods and, consecutively affect the ratio of contributing versus non-contributing area to streamflow (Fang et al., 2007; Pomeroy et al., 2005). In addition to the potholes and wetlands, the upper Assiniboine River watershed is also characterized by the presence of poorly developed stream networks that considerably affect the intensity and timing of surface runoff (Fang et al., 2007). The other complexity of the upper Assiniboine River watershed rises from the frozen ground effect during spring melt. The frozen ground condition reduces infiltration rate during the freshet period and remarkably increases the runoff intensity at the outlet (Hayashi et al., 2003).

The Canadian prairie is a semi-arid region with long, cold winters and short, warm summers with perennial flow regime in the Upper Assiniboine River (Upper Assiniboine River Basin Study, 2000). The continental polar air mass that basically resides over this region creates relatively low

amounts of precipitation (Mcginn, 2010). The precipitation amount in this region is highly variable and according to Shepherd and McGinn (2003), the annual precipitation over the entire Canadian Prairie region ranges from an annual low of 300 mm to a high of 550 mm with a mean value of 486 mm. Around 70–80% of precipitation is received as rain in the Canadian prairies (Wheaton, 1998). Within the simulation period used in this study (that is between 1994 and 2017), the average, lowest, and highest annual precipitation amounts are 520 mm, 395.1 mm, and 668.1 mm, respectively.

The input data quality check is not a part of this study and the HFC of MI were previously conducted a rigorous quality control mechanism on the raw historical meteorological data obtained from the Environment and Climate Change Canada weather stations. The data quality check tasks included filling missing values (about three percent) in daily temperature and daily precipitation time series by interpolating data from nearby stations and using the area ratio method. The precipitation data were not corrected for rain-gauge undercatch due to plugging, a considerable source of error for precipitation data in snow-dominated regions as discussed in Rasmussen et al. (2012), and need to be corrected in future research.

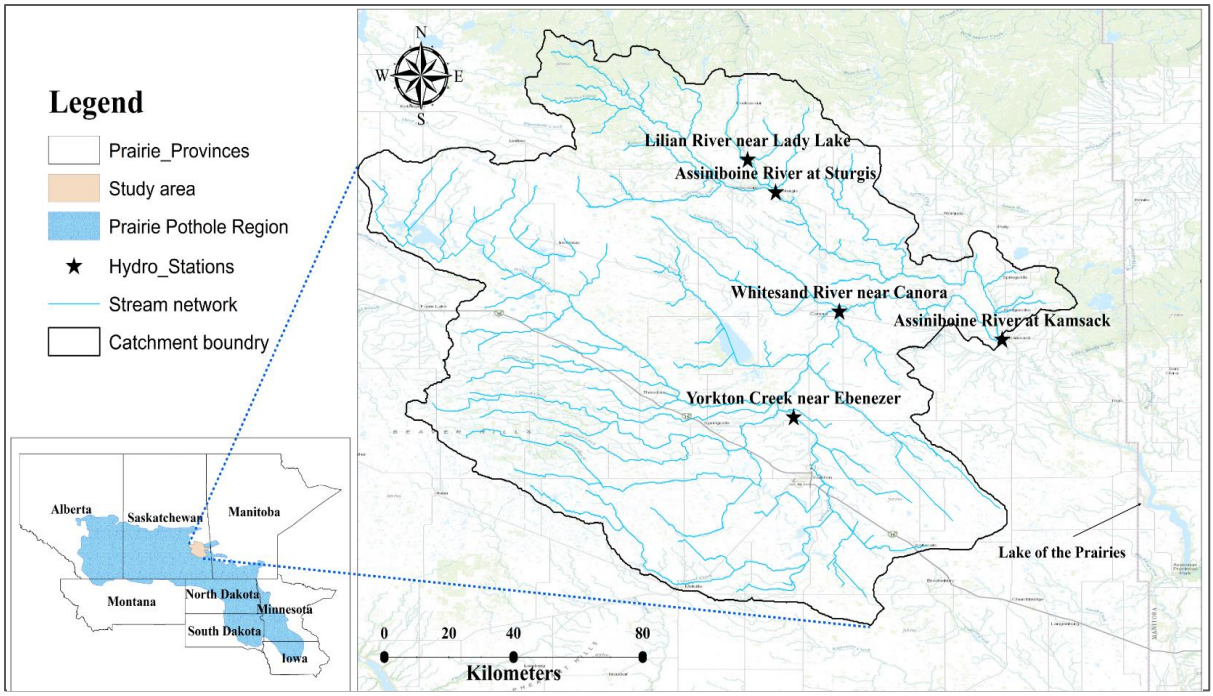


Figure 5-1: Map of the study area in Canada. Adopted from (Muhammad et al., 2018)

5.3.3 Hydrologic Models

The list of hydrologic models that have been applied for flood forecasting and historical reanalysis at HFC of MI include the WATFLOOD™ model (Kouwen, 2010), the Hydrologiska Byråns Vattenbalansavdelning (HBV) model (Lindström et al., 1997), the Hydrologic Modelling System (HEC-HMS) developed by the US Army Corps of Engineers (USACE, 2017), and the Hydrologic Simulation Program-Fortran (HSPF) developed by Bicknell et al. (2001).

These models have significantly different levels of complexity. WATFLOOD and HBV-EC models are distributed but HEC-HMS and HSPF are respectively lumped and semi-distributed models. The winter processes simulation is another main difference between these models. The winter processes have a major contribution to shaping the hydrograph from the beginning of freeze-up to the freshet period. All the models simulate snow accumulation and melting processes but only the WATFLOOD model is able to simulate the winter processes such as ice formation

and channel/lake freeze-up. This study is solely focused on the hydrologic models that are in operation at Hydrologic Forecast Centre (HFC) of MI and use conceptual methods for the hydrological processes. However, more advanced physically-based hydrologic models need to be set up at HFC to help them have a better simulation. Information regarding the model structure and main characteristics of each model in terms of input requirements and process representation of each module are found in the following and readers are referred to Unduche et al. (2018) for more detail about the Upper Assiniboine River Basin models such as digital elevation model and land cover map.

5.3.3.1 WATFLOOD Model (Kouwen, 2016)

The WATFLOOD hydrologic model was first developed in 1973 for long-term hydrological simulation and flood forecasting. This model uses conceptual approaches for simulating the hydrological processes, but the streamflow routing is based on physical methods. WATFLOOD has been under continuous development in the past four decades, for example Stadnyk et al. (2013) made it capable of simulating stable water isotopes to make the model applicable to remote basins that are minimally gauged and Kouwen (2010) added the reservoir operation simulation component to it. The model maintains both energy and water balances to respectively calculate the losses through evapotranspiration and outflow from the basin. It uses a simple conceptual approach for the winter processes simulation that cuts off evaporation, starts sublimation, and stops direct precipitation from being added to the streamflow. Land cover, precipitation, and temperature data are the minimum required datasets for WATFLOOD; however, it can incorporate information about snow depth, solar radiation, wind speed, humidity, and soil moisture as spatially distributed input variables. WATFLOOD includes data pre-processing tools to convert the input data points into gridded data sets. The model works based on the grouped response unit (GRU) as the smallest

computational component that assumes regions with similar land cover have similar soil type and topography. Therefore, the hydrologically significant land cover classes are placed in one GRU in each grid cell if the land cover resolution is higher than the cell size. This way, the parameters are associated with the land cover classes rather than the grid cells, that is, each grid cell has the same parametrization. The contribution of each land cover class to the runoff is calculated based on their coverage in each cell. The total hydrological response for a cell is equal to the summation of GRU responses for the cell. The hydrological responses are routed from the upstream to downstream cells connected by the drainage network.

The drainage of the study basin is modelled with 980 cells at 0.09° longitude and 0.06° latitude (7 × 6 km2). Each cell is further sub-divided into eight general GRUs based on land classes that are croplands, grasslands, deciduous forest, coniferous forest, mixed forest, wetlands, water, and impervious lands. Six river classes are also identified for streamflow routing purposes based on natural boundaries and the nature of streamflow. Further details about WATFLOOD can be found in Kouwen (2016).

5.3.3.2 HEC-HMS Model (Feldman, 2000)

The conceptual Hydrologic Modelling System (HEC-HMS) has long been used for flood forecasting and control due to its advanced graphical user interface and flexibility of methods selection for the hydrological process calculations. HEC-HMS is a semi-distributed hydrologic model that is developed for the spatiotemporal hydrologic simulation of a river basin with stream channels branching to form a tree-like pattern (USACE, 2017). It models the entire basin by multiple sub-basin and reach elements. Lump models are primarily used for the simulation of hydrological processes in each sub-basin element, and the resulted outflow in each sub-basin is

routed downstream through reach elements. User is able to choose interactively among different methods of precipitation distribution across the watershed, loss models, direct runoff, base flow, and streamflow channel routing. Given precipitation, the outflow from a sub-basin element, as a representative of the physical watershed, is computed by the summation of the overland flow, interflow and base flow components after subtracting losses due to deep percolation and evapotranspiration. A portion of excess precipitation turns directly into the overland flow and another portion infiltrates into the soil layer. A portion of the infiltrated water then moves vertically to the land surface via capillary effect and moves horizontally as interflow to the channel. The other portion of the infiltrated water also percolates to the groundwater aquifer and, after losing to the deep groundwater, moves horizontally until it is discharged as base flow to the stream. The outflow from a sub-basin is routed downstream if it enters a reach element to account for river channel flow attenuation.

Three lumped sub-basin elements are defined for the study area, two of which with outlets at Whitesand and Sturgis (see Figure 5-1) and one element for the remaining portion of the Upper Assiniboine River Basin from the outlet of the other sub-basin elements to the outlet at Kamsack. The outflow from the sub-basin elements are routed downstream through the reach elements and then aggregated by defining junction elements. The HEC-HMS model of this study uses the soil moisture accounting model, SCS unit hydrograph, linear reservoir, and Muskingum-Cunge methods respectively for simulating runoff volume, direct runoff, base flow, and channel routing (Unduche et al., 2018). The hyetograph of mean areal precipitation over a watershed is derived by applying the inverse-distance-squared weighting method to precipitation gauges. The Priestley Taylor method is selected for evapotranspiration and the temperature index method is the only

option for snowmelt calculation in HEC-HMS, which represents a conceptual snowpack energy balance.

5.3.3.3 HBV-EC Model (Moore, 1993)

The conceptual HBV-EC (Hydrologiska Byråns Vattenbalansavdelning-Environment Canada) hydrologic model was developed by Moore (1993) that accounts for a glacier routine in the model structure. HBV-EC is used for forecasting as it proved to be effective and efficient in simulating streamflow in a wide variety of climatic and physiographic conditions, e.g., Beck et al. (2016), Te Linde et al. (2008), Vetter et al. (2015), Zhang and Lindström (1996). HBV-EC model works based on the grouped response unit concept. Each GRU contains regions with similar elevation, slope, aspect, and land cover. The model accepts only four land cover type, that are forest (densely vegetated), open (sparsely vegetated or non-vegetated), glacier, and water (e.g., lake and other types of wetlands). HBV-EC also divides a watershed into climate zones with unique parametrization to account for lateral climatic gradients across a watershed. The snowmelt rate varies as a function of terrain slope or aspect within an area of a watershed with the same climate zone, elevation band, and land use class.

HBV-EC requires daily precipitation and temperature in conjunction with long-term monthly potential evaporation for a successful hydrologic modelling (Bergström, 1995; Hamilton et al., 2000; Zegre, 2008). The model then uses the degree-day method (Hamilton et al., 2000) for snowmelt and snow accumulation to calculate the volume of liquid released water. The released water from the snowmelt, if any, is added to the rainwater after subtracting evaporation as a function of soil moisture. A portion of the released water runs off quickly as fast discharge and the

remainder percolates into a lower linear reservoir that generates slow discharge or base flow (Grillakis et al., 2010; Zegre, 2008).

In this study, the HBV-EC model of the Upper Assiniboine River Basin encompasses four major climate zones and it is developed and run in Green-Kenue (Canadian Hydraulic Centre, 2010), a freely available data processing tool. Readers are referred to Bergström (1995), Lindström et al. (1997) for further details about HBV-EC structure.

5.3.3.4 HSPF Model

Hydrologic Simulation Program-Fortran (HSPF) is a conceptual model developed by Bicknell et al. (1996) for continuous modelling of the hydrological processes, point, and non-point source pollutant loadings in agricultural basins based on water balance. Modellers are empowered to operate the model in different temporal scales from minute to day. Weather data and basin morphology characteristics such as land use, drainage system, soil type, and topography are required for the runoff response quantification. The HSPF model partitions the watershed into sub-basins with similar topographic characteristics. Each sub-basin is further sub-divided into multiple smaller units to account for land use variations inside the sub-basin elements. Each unit is called hydrologic response unit that groups land uses with similar hydrological response. Each sub-basin spills into a reach element and the overland flow (pervious and impervious lands), base flow (pervious lands), and interflow (pervious lands) from the local drainage is summed up with the inflows from upstream reaches and routed downstream by using the stage-discharge-volume relationship estimated by the user for the element. The HSPF model uses the kinematic wave assumption for the channel routing. The air temperature varies with altitude. The model simulates a wide range of hydrological processes, including snowmelt and snow accumulation, infiltration,

percolation, soil moisture, surface retention, interflow, base flow, evapotranspiration, and surface flow. For more in-depth information regarding the model structure, concepts, and applications readers are referred to Bicknell et al. (2001), Duda et al. (2012), Kim et al. (2007), Xie and Lian (2013).

The spatial breakdown of the basin is user-specific and the Upper Assiniboine River Basin in this paper is divided into 88 reach elements based on the topography and drainage system. Six land use type including upland and lowland glacial, glaciolacustrine, and glaciofluvial are assigned to the reach elements. Each element is given the portion of drainage area covered by each land cover type. The degree-day method is used for snowmelt and snow accumulation processes.

5.3.4 Model Calibration

5.3.4.1 Model Warm-up, Calibration and Validation Periods

Daily time series are used for hydrologic modelling of the Upper Assiniboine River Basin with simulation period starting from 1994 and ending in 2017. In order to determine the best out of the simulation period for the model calibration, a pre-processing analysis is carried out on the precipitation, temperature, and runoff data. Figure 5-2 shows the annual mean flow rate to distinguish the wet and dry years from the normal years. The average flow rate in this period is equal to 15.023 m^3/s and all years falling outside the range [75%, 125%] of the average flow rate are considered as dry or wet years. This range is case-dependent and is selected based on visual inspection of the time series of streamflow to determine the streamflow variability from year to year and include years with relatively low and relatively high water volume in the calibration. According to this range and Figure 8-4, 1994, 1998 to 2005, 2008, and 2009 are considered as dry

years, and 1995, 2007, 2011, 2012, 2014 and 2016 are considered as wet years. Major hydrographic events in the precipitation-runoff-temperature graph (Figure 8-4) is also checked for each year to identify years with similar dominant hydrological processes. For instance, a peak flow in the freshet period can be assumed to be due to a major snowmelt (e.g. 2009) or rain on snow (e.g. 2015) while a peak flow in the mid-summer can be assumed to be caused by major rainfall events only (e.g. 2010). Each year is therefore classified into one of the categories in Table 5-1 and 14 consecutive years from 2001 to 2014 are accordingly used for the calibration as this period contains both dry and wet years and encompasses at least one year from each category in Table 5-1. According to Kouwen (2016), at least one year, preferably dry year, is assumed sufficient for WATFLOOD model warm-up. However, 1994 and 1995 that are respectively dry and wet years are used for the warm-up to minimize the effect of wet and dry situations on the model initialization. 1996 to 2000 and 2015 to 2017 are used for the evaluation of the calibrated models. Category 6 in Table 5-1 is between categories 4 and 5. If the hydrologic models are capable of simulating years in categories 4 and 5 adequately, they are expected to simulate 2016 as well. That is why the only year in category 6 is left for model evaluation.

Table 5-1: Classification of years based on the dominant hydrological processes. Italic font is used for model warm-up, bold font is used for model evaluation years, and normal font is used for calibration years.

Dominant Hydrological Process	**Years**
1) Base flow dominated	*1994*, **2000**, 2002
2) Snowmelt Dominated + High Precipitation in Spring	*1995*, **1996**, **1997**, **1999**, 2001, 2003, 2006, 2007, 2013, **2015**
3) Snowmelt Dominated + Low Precipitation in Spring	**1998**, 2004, 2005, 2008, 2009, **2017**
4) Snowmelt Dominated + High Rainfall in Summer	2011, 2014
5) Rainfall Dominated	2010, 2012
6) Snowmelt Dominated + Low Rainfall in Summer	**2016**

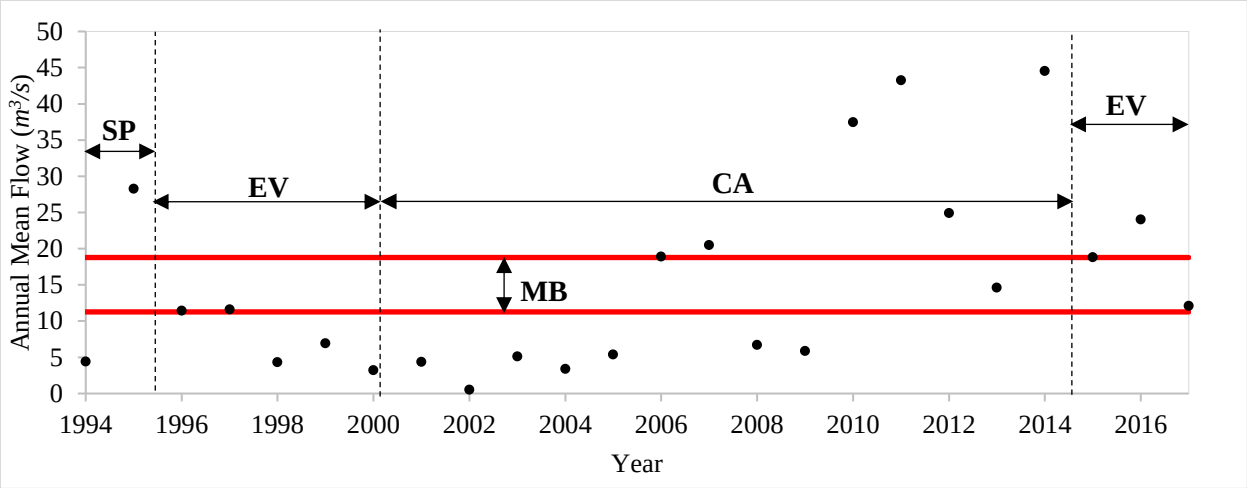


Figure 5-2: identification of the dry and wet years by comparing the mean annual flow with the long-term mean flow rate band (MB) $\left[0.75\overline{Q}_{long-term}, 1.25\overline{Q}_{long-term}\right]$. SP, CA, and EV respectively stand for spin-up, calibration, and evaluation periods.

5.3.4.2 Flow-Duration Curve Segmentation

Flow duration curve (FDC) is an effective tool to evaluate multiple aspects of a hydrologic model performance against the actual watershed system behavior. Yilmaz et al. (2008) suggested that the automatic model calibration should incorporate the partitioning of the flow-duration curve (FDC), the sorted logarithmic flow rate plotted against cumulative probability of exceedance, as in Figure 5-3. They suggested dividing FDC into three segments of high-, low-, and mid-range that represent important hydrological signatures. The mid-range flow often forms a straight line in the logarithmic scale, (20% to 60% probability of exceedance in this study as in Figure 5-3, and its slope shows how flashy the rainfall-runoff relationship is. This is a signature of a watershed because it is a function of the soil water retaining capacity. A watershed with low field capacity is expected to have a steep mid-segment slope and a higher chance of surface flow, while a flatter mid-segment slope signifies slower and more sustained groundwater response (Yilmaz et al., 2008). The volume of high-flow (0% to 5% probability of exceedance in this study) shows the quick flow runoff due to the snowmelt or major rainfall and the volume of low-flow (60% to 100%

probability of exceedance in this study) represents the vertical redistribution of water and the base-flow contribution to the streamflow. FDC does not preserve the sequence of events; therefore, metrics defined based on FDC must be accompanied with other metrics such as Nash-Sutcliffe Efficiency (*NSE*) or root mean squared error for model calibration proposes.

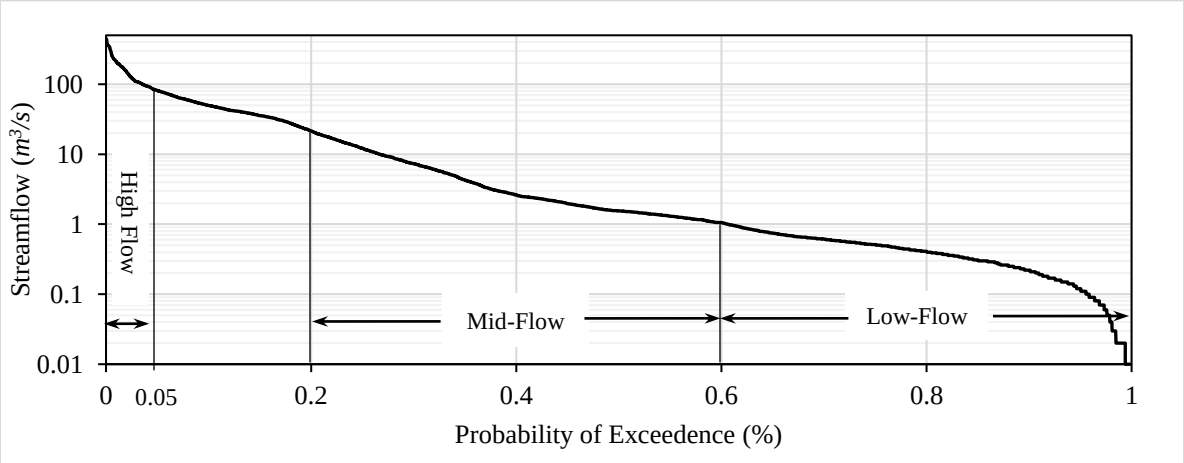


Figure 5-3: Segmentation of Flow-Duration Curve at Kamsack Hydrometric Station. The exceedance probability ranges [0, 0.05], [0.2, 0.6], and [0.6, 1] respectively denote high-, mid-, and low-flow rate partitions.

5.3.4.3 NSE Decomposition

Nash-Sutcliffe Efficiency is arguably one of the most widely used metrics in the literature of hydrologic model calibration. It specifically stresses the high-flow portion of hydrograph since large residuals, i.e. absolute difference between simulation and measured values, usually occur in high-flow periods and the fit between the simulation and observation in low-flow periods has a minimal effect on the value of *NSE*. Gupta et al. (2009) showed that *NSE* is composed of the variability index α, the correlation coefficient r, and the normalized water balance bias β_n as shown in Equation (29) where, M, S, t, N, $\bar{M}$, $\bar{S}$, σ_M, and σ_S respectively denote measured streamflow, simulated streamflow, time in daily scale, total number of time steps, mean of the

measured values, mean of simulations, standard deviation of measured values, and standard deviation of simulations.

$$NSE = 2\alpha r - \alpha^2 - {\beta_n}^2$$

$$\alpha = \frac{\sigma_S}{\sigma_M}, \beta_n = \frac{(\bar{M} - \bar{S})}{\sigma_M}, r = \frac{N\sum(S_t M_t) - (\sum S_t)(\sum M_t)}{\sqrt{(N\sum {M_t}^2 - (\sum M_t)^2)(N\sum {S_t}^2 - (\sum S_t)^2)}} \quad (29)$$

The variability index and normalized water balance bias aim to fit the model results to the first and second moments of the distribution of the measured streamflow, and the correlation coefficient evaluates the model performance in terms of the shape and timing of the hydrograph response to rainfall and snowmelt dominated events. Gupta et al. (2009) showed that the optimal *NSE* value of 1.0 will be achieved when the variability index is equal to correlation coefficient and the bias term is equal to zero. This means that α is constrained by the value of the correlation coefficient and cannot reach its optimal value of one due to the correlation coefficient being always less than or equal one. The water balance bias has also a minimal effect on the value of *NSE* since the discrepancy between the mean values is normalized by standard deviation of measured streamflow (Gupta et al., 2009). Hence, Gupta et al. (2009) recommended a revised aggregation of these three metrics called the Kling-Gupta Efficiency (*KGE*) or a segregated multi-objective optimization of α, r, and $\beta_n$instead of the single-objective optimization of *NSE*. Each of these components is used as one calibration objective, in this study.

5.3.5 Model Calibration Problem Formulations

Three different multi-objective problems are formulated to automatically calibrate the four hydrologic models of the Upper Assiniboine River Basin. These problems use metrics that are common for automatic calibration of hydrologic models. Due to the lack of other information in

the region such as evapotranspiration, water quality records, or groundwater data, which increase parameter identifiability, the model calibration objectives are error metrics and hydrological signatures that are based on streamflow. The comparison between the results of these three problems show whether one formulation can guide the optimization algorithm toward better solutions. The solution of a multi-objective optimization problem with conflicting objectives is a tradeoff or Pareto front between the objectives. The true Pareto front is the global optimal solution and any sub-optimal tradeoff such as the result of multi-objective model calibration is often called the Pareto approximate front.

5.3.5.1 P1. Max.Min. Formulation

The first problem formulation, P1 has three objective functions to 1) maximize the model performance over the whole calibration period as in Equation (30), 2) maximize the model performance for the most challenging wet year as in Equation (31), and 3) minimize the bias for the most challenging dry year as in Equation (32), simultaneously. Since flood forecasting is the main goal of models at HFC of MI, *NSE* that is sensitive to residuals in high-flow periods is selected for objectives 1 and 2. If the lowest *NSE* value based on Equation (30) is maximized, the *NSE* values for other wet years are definitely higher. *NSE* is not a relevant metric for dry years since they are expected to have low or no peak flow rate and there is not a large variability in the annual hydrograph. As a result, absolute bias in Equation (32) is selected for the third objective to evaluate the model performance for simulating persistent low-flows in dry years.

$$f_1 = NSE_{Overall} = 1 - \sum_{t=t_0}^{T}(S_t - M_t)^2 / \sum_{t=t_0}^{T}(M_t - \overline{M})^2, \qquad \overline{M} = \frac{1}{N}\sum_{t=t_0}^{T} M_t \tag{30}$$

$$f_2 = max\left\{min\{NSE_i | i \in \{2001, \dots, 2014\}, \overline{M_i} > 1.25\overline{M}\}\right\} \tag{31}$$

$$f_3 = min\left\{max\left\{\left|\frac{\bar{S}_i - \bar{M}_i}{\bar{M}_i}\right| | i \in \{2001, \dots, 2014\}, \overline{M_i} < 0.75\bar{M}\right\}\right\} \quad (32)$$

5.3.5.2 P2. FDC Segmentation

Calibration objectives for the second problem formulation, P2 are based on the hydrological signatures recommended by Yilmaz et al. (2008). It has four objectives to maximize the overall *NSE* (f_1) as in Equation (30) and find the best fit in all three segments of the simulated and measured FDC. Three segments of FDC are identified based on the measured flow at the Kamsack hydrometric station. Minimizing the error in estimating the slope of the mid-segment of FDC, Equation (33), error in the volume of high-flow, Equation (34), and low-flow rates, Equation (35), are three objectives in this problem formulation.

$$f_2 = 100 \times \frac{|Slope_S - Slope_M|}{Slope_M} \qquad Slope = \frac{1}{N}\sum \frac{\log(Q_t) - \log(Q_{t'})}{t - t'}, \qquad \{t, t'\} = \{\{t, t'\} \in T | \ 20\% \le T \le 60\%\} \quad (33)$$

N: total number of slopes calculated, Q: flow rate, T: percent time of exceedance

$$f_3 = 100 \times \frac{|High_{Vol,S} - High_{Vol,M}|}{High_{Vol,M}} \qquad High_{Vol} = \sum(Q_t), \ t = \{t \in T | \ T \le 5\%\} \quad (34)$$

Q: flow rate

$$f_4 = 100 \times \frac{|Low_{Vol,S} - Low_{Vol,M}|}{Low_{Vol,M}}, \qquad Low_{Vol} = \sum \log(Q_t), \qquad t = \{t \in T | \ 60\% < T \le 100\%\} \quad (35)$$

5.3.5.3 P3. NSE Decomposition

Calibration objectives for the third problem, P3, are three components of the *KGE* metric shown in Equation (36) recommended by Gupta et al. (2009).

$$f_1 = (r-1)^2, f_2 = (\alpha - 1)^2, f_3 = (\beta - 1)^2 = \left(\frac{\bar{S}}{\bar{M}} - 1\right)^2 \quad (36)$$

5.3.6 Automatic Calibration Approach

Rounded archiving-based PA-DDS (Sahraei et al., 2019), a more effective version of Pareto Archived-Dynamically Dimensioned Search (Asadzadeh and Tolson, 2013) for solving many-objective optimization problems, is utilized for the automatic calibration of the hydrologic model cases in this paper. PA-DDS commences the search with a random solution generation. If there are good quality behavioral solutions (for example, with overall *NSE* of above 0.5) available prior to the optimization, they can be injected by the user as the initial solutions. The PA-DDS structure has three key elements, namely dominance relation, selection metric, and perturbation to guide the search toward the optimal tradeoff. PA-DDS generates only one solution at a time by perturbing the recently archived solution and the corresponding objective functions are evaluated. The dominance relation is then applied to assess whether or not the objective values of the generated solution surpasses those of the previously archived solutions. If the solution is non-dominated, it would be retained in the archive. The archive set contains only non-dominated solutions that means none is dominating (or dominated by) any other archived solutions. A solution (parameter set) is dominated by another solution if all of its objective functions are worse than the other. A solution is dominating another solution if all of its objective function values are better than the other. Two solutions are mutually non-dominated if one has at least one objective function better and at least one objective function worse than the other solution.

The hyper-volume contribution selection metric of PA-DDS is used in this paper as it is reported the best selection metric for solving general multi-objective optimization problems (Asadzadeh and Tolson, 2013). An archived solution that has higher hyper-volume contribution value has more chance to be selected for perturbation and subsequent solution generation. (Asadzadeh et al., 2014b) introduced the convex-hull contribution as the preferred selection metric for model

calibration problems with up to three objectives; however, it is not applied in this study because P2 has four objectives. The perturbation size is the only parameter of the PA-DDS that has a recommended robust value of 0.2 (Asadzadeh and Tolson, 2013).

The multi-objective calibration of the hydrologic models of this study are not readily available. WATFLOOD is equipped with two types of single-objective automatic calibration methods, namely, the Pattern Search (Hooke and Jeeves, 1961) and the Dynamically Dimensioned Search (DDS) optimization tools (Tolson and Shoemaker, 2007). The user is restricted to only select among the statistical metrics defined in WATFLOOD for single-objective calibration of the model. The HEC-HMS model uses traditional optimization methods such as univariate gradient and Nelder Mead (or downhill simplex method) techniques for fine-tuning the model parameters based on minimizing the statistical error metrics. The modeller is restricted to use the Monte-Carlo method for the automatic calibration of HBV-EC model and HSPF model is limited to using manual calibration. Thus, the need for linking a modern multi-objective optimization tool like PA-DDS to the hydrologic models for a better model parameter identifiability is undeniable.

Table 5-2 shows the computational budgets and number of trials for each model. Due to the model complexity, simulation time of each hydrologic model, and available computing resources, different budgets and trials are considered for their calibration. To calibrate WATFLOOD, three independent trials with 5000 iterations per trial (4-5 minutes per solution evaluation) are executed. Five, Ten, and Five independent calibration trials are performed respectively for HBV-EC (1-1.5 minutes per solution evaluation), HEC-HMS (35-50 seconds per solution evaluation), and HSPF (40-80 seconds per solution evaluation) models with 10000 solution evaluations for HBV-EC and HEC-HMS and 5000 solution evaluations for HSPF. The results of multiple trials for each calibration problem are aggregated and normalized to between zero and one after removing

solutions that produced negative overall *NSE* values. Then, the solution with the smallest weighted Euclidean distance to known ideal objective values is selected as the preferred solution for each problem formulation based on the weight factors presented in Table 5-2. The objectives are normalized to between zero and one to minimize the scaling effect of the objective functions on the Euclidean distance calculation. A higher weight is given to the overall *NSE*, correlation coefficient, and variability index to emphasize the high-flow rates since the flood forecast and control is in higher priority for HFC (see Table 5-2). The preferred solution for each model is used to create the best calibrated model for the model-wrapper.

Table 5-2: Hydrologic model parameters and experimental setups

Hydrologic Model	# of Parameters	# of Trials per Objective suite	# of simulations per Trial	P1 weights	P2 Weights	P3 Weights
HBV-EC	19	5	10000	(0.4,0.4,0.2)	(0.4,0.3,0.2,0.1)	(0.4,0.4,0.2)
WATFLOOD	114	3	5000			
HEC-HMS	14	5	10000			
HSPF	19	5	5000			

5.3.7 Linking PA-DDS to the Hydrologic Models

In order to use an external optimization algorithm such as PA-DDS to calibrate the hydrologic models of this study, a MATLAB ® function is developed to receive the set of parameter values from the optimization algorithm, update the model input files, run the stand-alone model application, load the results and calculate the objective functions. The structure of this code is shown Figure 5-4.

5.3.8 Segmentation-Based Model-Wrapper

The novel multi-modelling approach developed in this study is based on the hydrologic modelling practice at HFC of MI. The forecast results from different hydrologic models are quantitatively and visually compared for more accurate prediction of the inflow rate to the Shellmouth reservoir in high-flow and low-flow seasons in order to control flood and supply water to downstream communities.

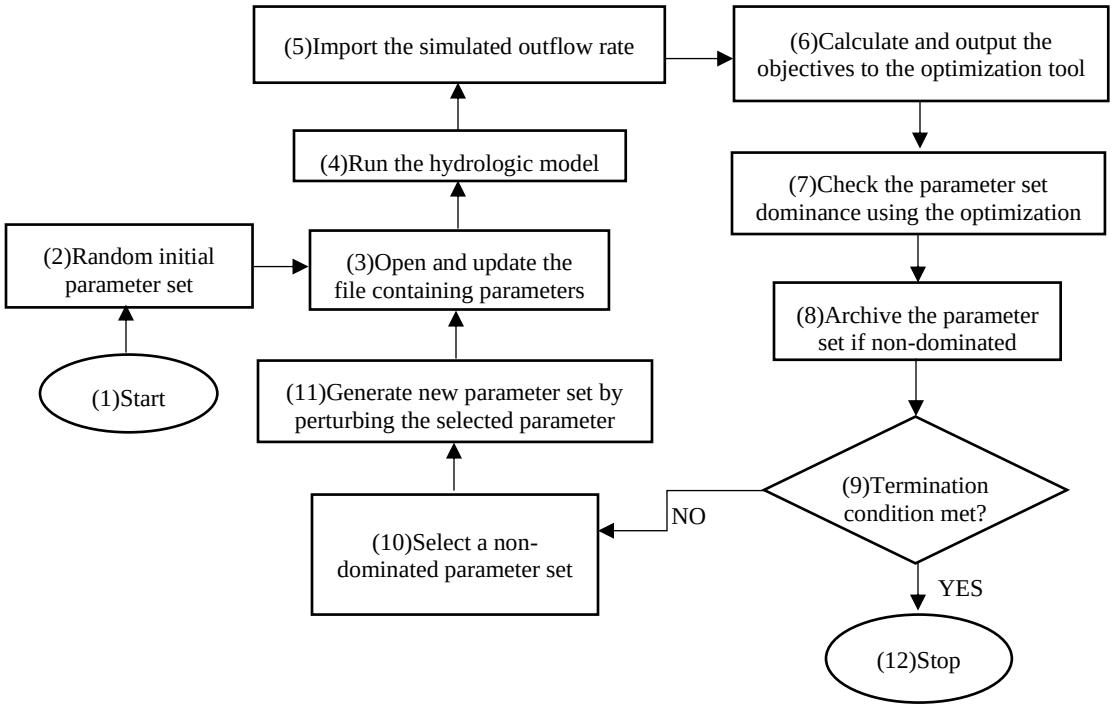


Figure 5-4: HBV-EC, WATFLOOD, HEC-HMS, and HSPF model calibration process using an external optimization tool

The multi-modelling approach proposed in this study is based on the weighted average of the simulations of the calibrated HBV-EC, WATFLOOD, HEC-HMS, and HSPF models. In practice, one model surpasses another depending on the flow rate and season. This is due to the structure of the models and the simulation performance of a model for high-flow and low-flow may be different. As a result, the predictability of these models are separately evaluated for high-, mid-,

and low-flow periods, which is also in line with the practical operations at HFC of MI. Equation (37) shows the weighted average multi-modelling for each time step (Q^t_{Multi}) as a system of linear equations. *w* is the weight between zero and one and the summation of the weights for each linear equation should be equal to one. According to the FDC partitioning of the calibration period described in Figure 5-3, 1.7 m^3/s and 15.7 m^3/s are used for the mid-flow range with relatively constant slope from 0.6 to 0.2 probability of exceedance, and beyond this range is considered as either high-flow or low-flow condition for the model-wrapper. The optimal value of the weights is achieved by minimizing the *NSE* metric between the multi-modelling simulations and measured streamflow data for the calibration period, shown in Equation (38) as constrained minimization.

$$S^t_{Multi} = \begin{cases} w_1 Q^t_{HBVEC} + w_2 Q^t_{WATFLOOD} + w_3 Q^t_{HECHMS} + w_4 Q^t_{HSPF} & M_t \geq 15.7 \\ w_5 Q^t_{HBVEC} + w_6 Q^t_{WATFLOOD} + w_7 Q^t_{HECHMS} + w_8 Q^t_{HSPF} & 1.7 < M_t < 15.7 \\ w_9 Q^t_{HBVEC} + w_{10} Q^t_{WATFLOOD} + w_{11} Q^t_{HECHMS} + w_{12} Q^t_{HSPF} & M_t \leq 1.7 \end{cases} \quad (37)$$

$$Min. \left\{ 1 - \sum_{t=t_0}^{T} (S^t_{Multi} - M_t)^2 \Big/ \sum_{t=t_0}^{T} (M_t - \overline{M})^2 \right\} \quad (38)$$

$$\text{Subject to} \begin{cases} 0 \leq w_i \leq 1 \\ w_1 + w_2 + w_3 + w_4 = 1 \\ w_5 + w_6 + w_7 + w_8 = 1 \\ w_9 + w_{10} + w_{11} + w_{12} = 1 \end{cases}$$

5.4 Results and Discussion

Two comparisons are made for each model: (1) between the calibration problem formulations and (2) between the calibrated models and the model-wrapper. The comparisons are mainly based on the heat map visualization of daily *NSE* values for each year in Figure 5-5 and annual streamflow hydrograph for the calibration and evaluation periods in Figure 8-5. The main reason for

considering *NSE* as the main criteria for the model performance is that it is highly sensitive to the model performance in estimating the timing and volume of high-flow events that are of particular interest for flood forecasting at HFC of MI. Moreover, *NSE* has been a very popular metric that is normalized for evaluating the hydrologic model performance; therefore, guidelines had been developed for interpreting the model performance based on its value. For example Moriasi et al. (2007) noted that, negative values of daily *NSE* can be interpreted as inadequate simulation, while *NSE* values higher than 0.5 can be interpreted as adequate model simulation. A negative *NSE* is colored in red and a positive *NSE* of higher than 0.5 is colored in blue in Figure 5-5. The annual mean flow rate in Figure 5-5 shows that the simulation period begins with a relatively dry hydrological cycle from 1996 to 2005 and transitions into a wet cycle from 2006 to 2017. The annual precipitation and annual mean temperature as the most important input variables are also visualized in Figure 5-5 to distinguish between all combinations of humid, cool, dry, and hot years. According to the annual precipitation, 2017 and 2010 are the years with extremely dry and extremely humid weather conditions, respectively. 2012 and 2014 are the hottest (red) and coolest years (blue) in the entire simulation period.

5.4.1 Single Model Performance Comparison

Calibrated models in this study are compared to the same models that were calibrated by Unduche et al. (2018) and are currently being used at HFC of MI for flood forecasting. This comparison highlights the advantages of the multi-objective calibration conducted in this research. According to Figure 5-5, the parameter settings found by the multi-objective automatic calibration using any optimization problem formulation (P1, P2, or P3) outperforms the previously used settings. The number of years with negative *NSE* values is reduced and the model performance in years with positive *NSE* values is improved. The previous settings that are used for the HBV-EC,

WATFLOOD, HEC-HMS, and HSPF models fail to simulate the annual peak streamflow in 8, 6, 8, and 7 years out of the 14-year calibration period (2001-2014) and 5, 3, 5, and 5 years out of 8-year evaluation period, respectively, by giving negative annual *NSE* values. According to Figure 8-5, the hydrologic models with the old setting tend to generally overestimate the peak flow rate except for WATFLOOD, which had been previously calibrated with the single-objective automatic calibration of overall *NSE* metric. Moreover, the calibrated models in this study generally give a more accurate simulation of the peak flows in the wet and normal years.

All hydrologic models calibrated in this study or by Unduche et al. (2018) failed to adequately simulate the extremely and persistently dry years 2002 to 2005 (e.g. year 2002 with no peak throughout the year). All the models overestimated streamflow in this period. Therefore, it is concluded that all models that are operational at HFC of MI lack hydrological processes that are active in persistent low-flow periods. These processes should be related to the interaction between groundwater and surface water.

5.4.2 Problem Formulation Comparison

In general, none of the three problem formulations is superior for calibrating all the models. HEC-HMS and HSPF calibrated using any of the three problem formulations can adequately simulate peak flows, see Figure 8-5. The simulated peak flow by P1 as a result of snowmelt is generally lower than the simulated peak flow by P2 and P3 for all hydrologic models. P1 tends to find a setting that decreases the variance for annual *NSE* and annual bias during the calibration period by improving these metrics for the most challenging years toward their ideal values. Using P1 formulation for the calibration of HEC-HMS and HBV-EC results in better estimation of peak

flow in freshet period than P2 and P3. However, HBV-EC and WATFLOOD underestimate the high-flow rates when they are calibrated with P1.

According to Figure 5-5, the HBV-EC model calibrated with P1 captures more number of years with positive *NSE* values (16 years) compared to the same model calibrated using P2 or P3 (14 years). Despite higher overall *NSE* value of 0.54, HBV-EC calibrated with P1 generally underestimates the daily average annual peak flow in Figure 8-5, and P2 and P3 formulations with respective *NSE* values of 0.50 and 0.48 better represent the streamflow variability for the HBV-EC model. Therefore, the preferred parameter values found by P3 is selected for multi-modelling since the major peak flow rates from 2010 to 2015 are more accurately simulated (not shown in this paper), compared to P1 and P2, and HFC is more concerned about major flood forecast accuracy.

Likewise, the setting identified with P1 formulation for WATFLOOD excels those with P2 and P3 as its heat map column in Figure 5-5 is more blue-dominated, 17 versus 13 and 15 years, respectively. However, the peak flow rates for P1 are generally underestimated based on the annual streamflow hydrograph in Figure 8-5. The WATFLOOD setting identified by P3 is therefore preferred for the multi-modelling since there is a balance between dry year and wet year simulation accuracies compared to P2.

As regards the HEC-HMS and HSPF models, the parameter set found by P1 has the best performance among other settings and its match to the snowmelt-dominated peak flow is preferred to P2 and P3 (see Figure 8-5). The preferred solution of P1 is thus selected for the HEC-HMS and HSPF models for the multi-modelling analysis.

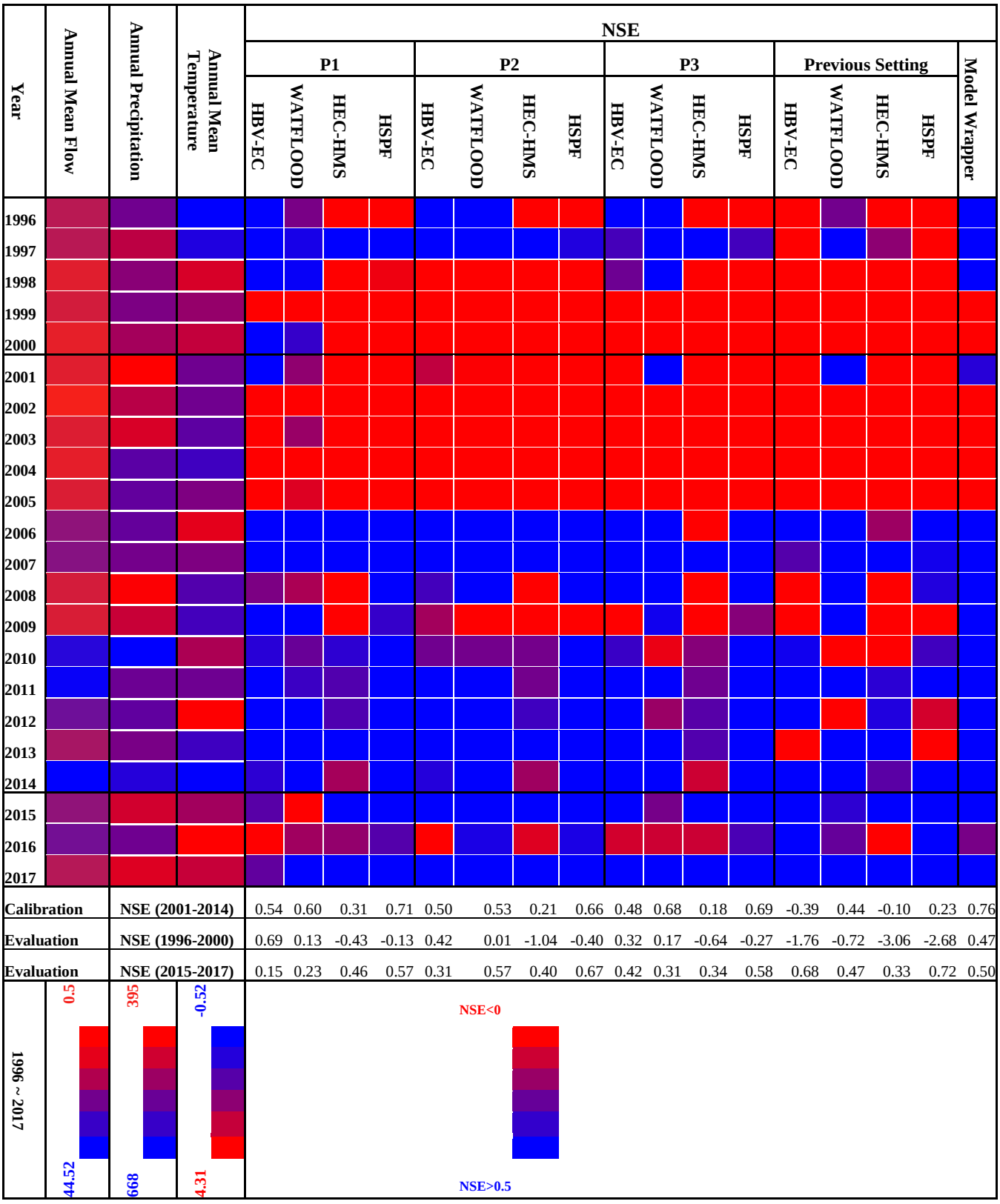


Figure 5-5: Performance comparison for HBV-EC, WATFLOOD, HEC-HMS, and HSPF with previous setting and model-wrapper based on annual *NSE* values for a simulation period from 1996 to 2017.

5.4.3 Multi-Model Simulation

Figure 5-5 shows that the proposed multi-model that assigns weight to each model based on the flow duration curve segmentation results in higher simulation accuracy and better fit in FDC value compared to individual model simulation. As shown in Equation (39), the model-wrapper weight factors are obtained by maximizing the overall *NSE* value for the calibration period. Each model has a different contribution to the multi-model flow rate simulation. HSPF has the highest contribution (68%) to the multi-model for simulating high-flows. WATFLOOD contributes to the simulation of high-flows as well with a weight equal to 32%, but its most significant contribution is to the simulation of mid-range and low-flow periods with weights equal to 90% and 99%, respectively. The high contribution and simulation accuracy of WATFLOOD is also evident from Figure 5-7, especially for flow rates with frequency of exceedance higher than 35%. The low-flow condition mostly occurs in the winter that is base flow or groundwater flow dominated. The main reason for a more accurate simulation of low-flows is that WATFLOOD is capable of simulating channel and lake ice formation, freeze-up, and sublimation processes in the cold months of the year. WATFLOOD automatically stops the evapotranspiration and activates sublimation in the winter. However, other models only consider the snow accumulation and melting processes and the winter processes are not simulated. For this reason, the low flow rate is mostly overestimated, an example of which is HSPF model in Figure 5-6.

Figure 5-6 displays the worst (1999) and best (2017) performing hydrograph simulated by the model-wrapper in the evaluation period compared to the hydrograph simulated by its individual components that are HSPF and WATFLOOD. Since model-wrapper is the optimal weighted average version of WATFLOOD and HSPF, its hydrograph falls between its individual components. Peak flow as a result of snowmelt is simulated but WATFLOOD starts an early melt

process in March due to experiencing a temperature warmer than the calibrated base temperature for different land cover classes while soil is still frozen and infiltration is low. The snowmelt process is conceptualized and it does not consider the temporal changes such as early or late melt. As a result, the calibrated WATFLOOD model looks for the best base temperature and other snowmelt parameters to give the overall best fit in the calibration period. In addition, WATFLOOD gives an overestimated response to the summer rainfall in June in 1999 and similar years owing to experiencing high summer peaks in calibration period. One reason is that WATFLOOD relates soil characteristics to land cover classes and it assumes that areas with similar land cover have similar soil properties. However, HSPF that is distributed based on soil properties, gives a more reasonable estimation of streamflow in June, 1999 compared to WATFLOOD.

$$S^t_{Multi} = \begin{cases} 0.3206Q^t_{WATFLOOD} + 0.6792Q^t_{HSPF} & M_t \geq 15.7\ cms \\ 0.897Q^t_{WATFLOOD} + 0.103Q^t_{HSPF} & 1.7\ cms < M_t < 15.7\ cms \\ 0.9922Q^t_{WATFLOOD} + 0.0035Q^t_{HSPF} & M_t \leq 1.7\ cms \end{cases} \quad (39)$$

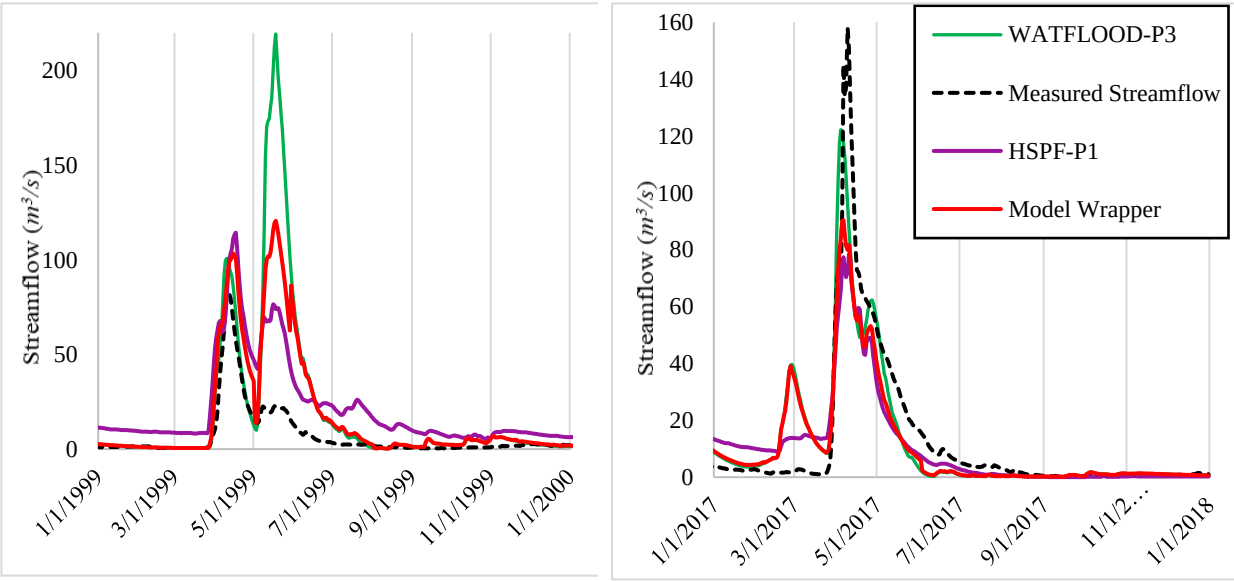


Figure 5-6: simulated hydrographs of individual hydrologic models compared with model-wrapper. 1999 and 2017 are respectively the worst and the best performing simulations based on model-wrapper in the evaluation period.

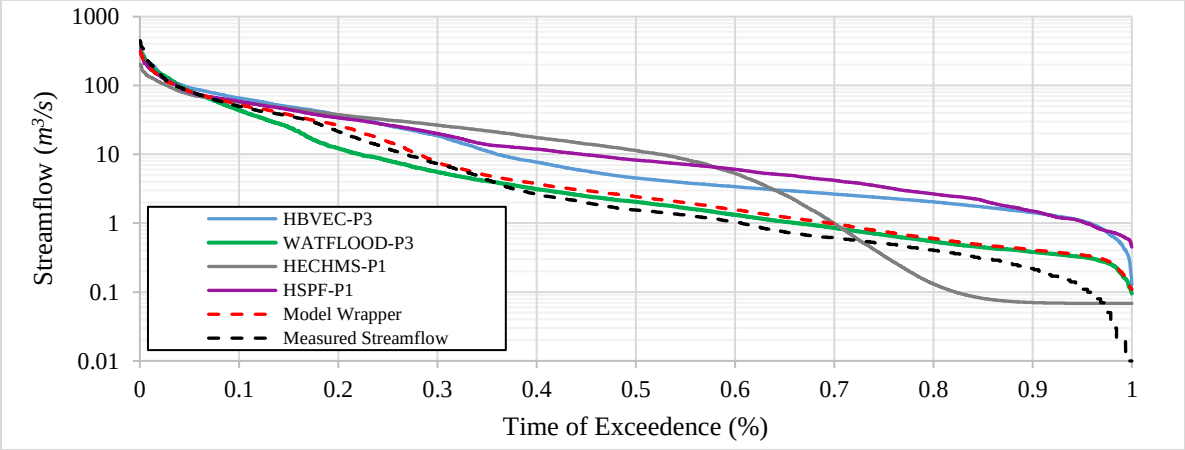


Figure 5-7: Flow-duration curve comparison of the model-wrapper against its hydrologic model components in calibration period

Multi-model combination technique proposed in this study improves the overall and year-to-year daily *NSE* metric in the calibration and evaluation periods. As is evident from the heat map in Figure 5-5, 17 out of 22 years of simulation have positive annual *NSE* values, 14 of which are greater than 0.5 significantly improving the performance of its individual hydrologic model components. Figure 5-8 also compares the overall performance of multi-modelling against single hydrologic modelling based on *NSE* and its log-transformed *NSE* (*Log-NSE*) to check the goodness-of-fit for high- and low-flow situations, respectively. The red continuous line with circle marker in Figure 5-8 represents the model-wrapper performance that is obtained by considering all the hydrologic models. Figure 5-8 demonstrates that the model-wrapper performance surpasses its individual model components for the entire simulation period except for 2015 to 2017 when it performs slightly worse than HSPF based on *NSE* and HBV-EC based on *Log-NSE*. For instance, HSPF alone fails to simulate the evaluation period 1996 to 2000 by giving negative *NSE* while its combination with other models increases the *NSE* value to 0.43 in this period. This improvement is also demonstrated in Figure 5-9 in both the calibration and evaluation stages that daily average annual streamflow hydrograph of the model-wrapper better represents its measured counterpart in

the peak, and rising and falling limbs. Moreover, the model-wrapper gives out a simulation range and a simulation mean that are consistent with the range and mean of daily streamflow in both the calibration and evaluation periods.

Optimal weights of 0.0 for HBV-EC and HEC-HMS show that these two models do not contribute to the multi-model. According to Figure 5-7, FDC of HEC-HMS is significantly different compared to that for the measured streamflow in all three segments of FDC; therefore, the multi-model does not benefit from HEC-HMS simulation in high-, mid-range, or low-flow periods. The optimal weight of HBV-EC is 0.0 as well, because as shown in Figure 5-7, it is very similar to HSPF; however, HSPF performs significantly better in high-flow periods that have the highest influence on the overall *NSE*. When HSPF or WATFLOOD is removed from the pool of models used for multi-modelling, HEC-HMS and HBV-EC contribute to the multi-model. If HSPF is removed, the HBV-EC and HEC-HMS receive optimal weighs of 27.3% and 5.9% respectively for simulating high-flows but close to zero for simulating mid- and low-flow partitions. On the other hand, by eliminating WATFLOOD and keeping HSPF in the pool of models, HBV-EC contributes to the simulation of low-flow rates with a weight of 22.5% but still does not contribute to the simulation of high- and mid-range flows. In this case, HEC-HMS would only contribute 9.5% to the simulation of mid-range flows. The performance of the aforementioned two variants of the multi-model is represented by the dashed and dot lines, respectively in Figure 5-8, showing that including all hydrologic models in the multi-model combination increases the reliability of the multi-model by producing more consistently positive and relatively high *NSE* and log-transformed *NSE* values due to decreasing the structural uncertainty in the hydrologic models.

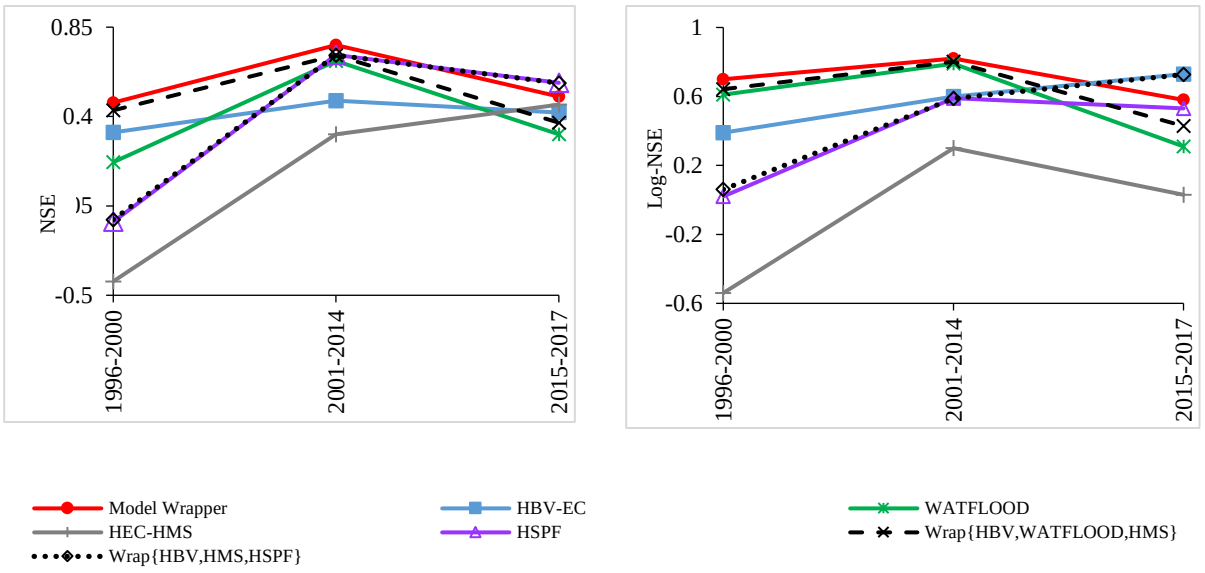


Figure 5-8: Overall performance metric comparison between the model-wrapper technique and its hydrologic model components for calibration (2001-2014) and evaluation periods (1996-2000 and 2015-2017)

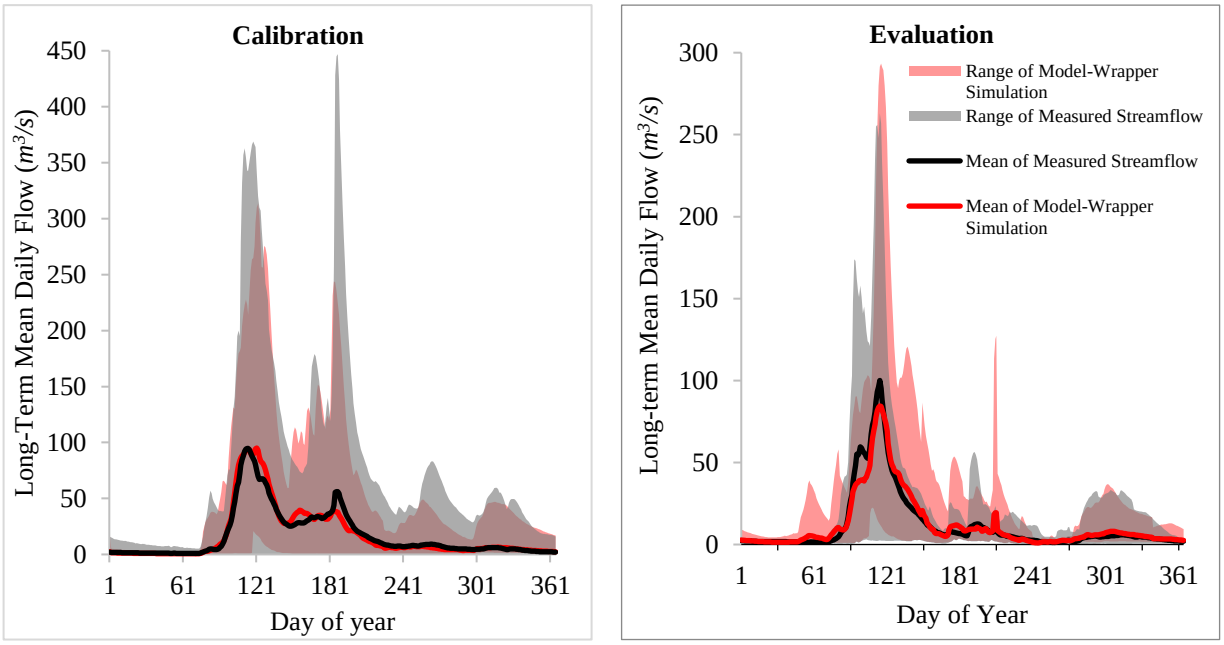


Figure 5-9: Daily average annual streamflow hydrograph of the model-wrapper technique for the calibration and evaluation periods at Kamsack outlet station compared to the measured streamflow data. The Day number (1-365) on the horizontal axis starts from January 1st and ends on December 31st. The shaded area in grey and pink respectively demonstrate the historical range and multi-model simulation range of daily streamflow for the calibration and evaluation periods.

5.5 Study Limitations

In this study, we limited the hydrologic models to those that are currently operational at HFC of MI. This does not mean that these models are the best ones for flood forecasting upstream of the Shellmouth reservoir. One would expect that including other models that perform better than the models used at HFC of MI individually would improve the overall performance of the multi-model as well. Moreover, each hydrologic model in this study has a significantly different number of parameters, with WATFLOOD having the highest number of parameters (114) and HEC-HMS having the lowest number of parameters (14). In general, calibrating more parameters require more computational budget; therefore, WATFLOOD model should have been calibrated with a high number of solution evaluations; however, such computational resources were not available for this study. It should be also noted that more solution evaluations can only improve the calibration results. Therefore, we believe that this limitation does not have any negative impact on the results and conclusion. However, we believe that the WATFLOOD model used at HFC of MI has too many parameters to be identifiable. Therefore, we suggest that parameters of this model be grouped for calibration, in an approach similar to the calibration of curve number parameter of SWAT in Abbaspour (2013).

The parameters of the HEC-HMS model are assumed spatially constant. Automatically calibrated HEC-HMS model performs weakly for the streamflow simulation of Upper Assiniboine River Basin at Kamsack with an overall *NSE* of 0.31 (problem type P1) for the calibration period. In an attempt to increase the model simulation accuracy, each sub-basin element was considered to have a unique parameterization for HEC-HMS. However, this increases the number of parameters and leads to ill-conditioning in the model. Ill-conditioning means that it has a high condition number, and it is very sensitive to the number of decimals in its parameter values. A

small change in the parameter values changes the simulation results drastically. During the optimization, PA-DDS generates a new parameter set and archives it, if non-dominated, with full-precision level but it is evaluated with lower precision since the HEC-HMS model, first, rounds up or down the full-precision parameter values in the parameter file to lower-decimal values, then it simulates the flow rate. The parameter distribution in the model domain would be highly non-smooth and consist of micro-scale nonlinearities if it is assumed that parameters are spatially independent. By taking spatial independency assumption into consideration, inappropriate time stepping scheme and internal model thresholds lead to an ill-conditioned model (Kavetski et al., 2006). This issue is resolvable by implementing the regularization methods on model parameters such as Tikhonov regularization (Tikhonov and Arsenin, 1977) or constraining the calibration problem by adding additional information about the parameter interdependency and their relations with the watershed characteristics (Pokhrel et al., 2008). The implementation of the regularization methods is out of the scope of this study and the readers are referred to the mentioned citations to better understand the concept of ill-conditioning and regularization.

5.6 Conclusions

This study demonstrated a segmentation-based model blending strategy for improving the performance of individual hydrologic models. The proposed model-wrapper was successfully applied to the models used for flood forecasting at HFC of MI. The previously developed hydrologic models at HFC of MI were linked to the PA-DDS multi-objective optimization tool to re-calibrate the models using three suites of signature-based calibration objective functions. The final best parameter set or preferred solution, found by post-processing the archive set of PA-DDS, outperformed the previous versions of the hydrologic models meaning that signature-based multi-objective calibration increases the simulation accuracy and parameter identifiability.

After calibrating the models, a segmentation-based weighted average combination technique was built to let each model have a different contribution to the simulation of high-, mid-, and low-flow ranges. The multi-model had a significantly improved performance compared to each individual model. Results showed that HBV-EC and HEC-HMS do not partake in multi-model simulation. HSPF is the main contributor to simulation of the flow rates higher than 15.7 m^3/s (time of exceedance less than 20%) but the WATFLOOD model takes over the simulation of mid- and low-flow partitions (less than 15.7 m^3/s). Assuming the situation when either HSPF or WATFLOOD is not available, the HBV-EC and HEC-HMS models join the multi-model combination helping to enhance the model-wrapper simulation in comparison with its individual hydrologic model components. This means that none of the hydrologic models is able to individually capture all the aspects of simulation error distribution. Therefore, grouping the models increases the performance and decreases the model structural uncertainty. The segmentation-based model-wrapper technique is in line with the practical operations at the HFC of MI for predicting inflow to the Shellmouth Reservoir at downstream during low-flow and high-flow seasons. It is highly encouraged to include more physically-based hydrologic models that take the temporal variations such as early or late melt and spatial distribution of soil properties into consideration to increase the timing and magnitude accuracy of peak flows resulted from snowmelt and summer rainfall.

The Upper Assiniboine River Basin in the prairie region of Canada is chosen for the method implementation, results demonstration and discussions in this study. However, the signature-based multi-objective calibration and segmentation-based model combination techniques are not limited to the Upper Assiniboine River Basin and its hydrologic models in this paper. The model-wrapper is expected to perform better than its individual model components regardless of region of study

and type of hydrologic modelling. If one model, outperform other individual models in all aspects of simulation error distribution, the weight factor in the proposed model-wrapper converges to one in all ranges of measured streamflow. More broadly, research is needed to study the impact of more physically-based hydrologic models on flood forecasting performance of the model blending method in regions with different dominant hydrological processes.

The signature-based calibration metrics in this study were defined solely based on streamflow measurements that were in line with the purpose of the models for flood forecasting. However, in general, modellers are highly encouraged to include other calibration metrics in the optimization that explain other components of hydrological processes, such as the interaction between surface water and groundwater, evapotranspiration and sublimation, snow pack, upon existence of information and/or measured data. The models in this study were evaluated temporally but the spatial evaluation would help increase parameter identifiability if hydrometric observations exist at multiple locations across the basin.

Future work can also focus on considering the ill-conditioning issues for the multi-objective calibration of the hydrologic models of this study via implementation of the regularization methods on the parameters for a better-conditioned and more identifiable watershed system (Kavetski et al., 2006). This is facilitated by exploiting additional information from the system such as spatial dependency of the model parameters on the watershed properties such as curve number, soil properties, and land cover to decrease the parameter space dimensionality and constrain the parameters not to take on arbitrary values in their feasible ranges (Pokhrel et al., 2012, 2008).

5.7 Acknowledgement

This work was supported in part by Dr. Asadzadeh's Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant [RGPIN-2016-05896]; Graduate Enhancement of Tri-Council Stipends (GETS), University of Manitoba, Winnipeg, MB; and Manitoba Hydrologic Forecast Centre of the Government of Manitoba.

6 RESEARCH SUMMARY, RESEARCH SIGNIFICANCE, LIMITATIONS, AND RECOMMENDATIONS FOR FUTURE WORK

6.1 Summary of Thesis Findings

Multi-objective global optimization algorithms have pushed the boundaries of our knowledge about water resources engineering optimization. Water resources engineering and design problems essentially have many objectives with nonlinear, multi-modal characteristics, and their optimization demands modern multi-objective algorithms that efficiently converge to a diverse set of solutions as close as possible to the Pareto optimal set. This dissertation aims at addressing the dominance resistance issue and preserving solutions diversity for water resources engineering problems.

In the first phase of this thesis, the effect of a high number of objectives on the optimization process was studied. As the number of objective function increases, the fraction of non-dominated space in the objective space exponentially increases and mainstream multi-objective global optimization algorithms are not able to effectively solve many-objective problems. An increased objective space dimensionality results in a high number of similar alternatives that would quickly populate the archive set, which is more deteriorative in an algorithm with a bounded-archive structure. Therefore, a rounded-archiving strategy is proposed to tackle dominance resistance and solution similarity in the objective space. The proposed method discretizes the objective space based on the required resolution for each objective function and prevents retaining multiple solutions with finer resolution discrepancies. As opposed to the ε-archiving, the proposed

archiving method is ready to use and requires no modification or adding extension to the structure of the multi-objective global algorithms to be able to solve many-objective optimization problems.

The multi-modal property of water resources engineering problems, particularly hydrologic model calibration, directed the research towards development of a diversity-based multi-objective optimization algorithm for retaining optimal and near-optimal tradeoffs. For this purpose, a density-based clustering strategy is utilized to identify dissimilar options by placing similar solutions in a cluster. The dominance archiving strategy only compares solutions in the same group and dominated solutions in a cluster are eliminated. The designed selection metric for this algorithm gives more evolving opportunities to clusters that are distant from other clusters in the decision space but are closer to the ideal point in the objective space. This way both the decision-space diversity and the convergence are encouraged at the same time. The developed algorithm adds two parameters that represent minimum spatial density for clustering solutions and recommendations are provided for a proper setting of these parameters. The cluster-based archiving strategy was developed on PA-DDS algorithm and applied to mathematical and environmental benchmark problems with multi-modal characteristics. The cluster-based PA-DDS algorithm could offer diverse options required for an informed and flexible decision-making.

The clustering operation imposes extra computational cost to cluster-based PA-DDS. The average computational cost of original DBSCAN is equal to $O(N.Log(N))$ for N points of a database (Ester et al., 1996). In the proposed cluster-based archiving approach, unlike original DBSCAN, clusters are dynamically formed and expanded, and clustered solutions that are dominated by their groupmates are eliminated from the archive. Therefore, the additional computational cost due to clustering is less than the original DBSCAN in each iteration of cluster-based PA-DDS, because the dominated solutions are eliminated from each cluster and cluster-non-

dominated solutions are assigned a memory of solution coverage. Moreover, the significance of this extra computational cost is different for solving different case studies. For example, each WATFLOOD model simulation of the Upper Assiniboine River Basin in the previous chapter approximately takes 4 to 5 minutes on a laptop with Intel ® Core ™ i5-6300HQ CPU @ 2.3 GHz with 8.00 GB of RAM. The 1000 simulation of this case study takes 2 days, 18 hours and 40 minutes on this machine while the cluster-based PA-DDS auto-calibration with the budget of 1000 solution evaluations approximately takes 2 days, 18 hours and 42 minutes. Therefore, the computational cost of the cluster-based optimization for automatic calibration of the WATFLOOD model can be considered negligible when solving computationally intensive optimization problems.

The developed cluster-based PA-DDS algorithm was applied to 12-parameter HBV-SASK, 20-parameter RAVEN, and 13-parameter SWAT models to quantify parameter uncertainty in these models in a multi-objective perspective. The 12-parameter and 20-parameter models were used for testing the method in order to identify the best setting for the adjacency density parameter of the algorithm and the 13-parameter SWAT model was used for validating the algorithm's modus operandi. The proposed multi-objective algorithm was run for multiple independent trials and compared to 10-trial GLUE and 10-trial DDS-AU that are popular single-objective tools for approximation of parameter uncertainty among hydrologic science community. The uncertainty approximation methods are compared based on several performance metrics including reliability, sharpness, and number of behavioral solutions for a specified behavioral threshold. The thresholds were mainly determined based on the aggregated version of objectives in each hydrologic model. It was found that the aggregation of objectives and behavioral threshold specification based on the aggregated objective function led to misidentifying behavioral solutions compared to a situation

where the behavioral threshold were defined separately for each objective function in a multi-objective uncertainty analysis.

In order to increase parameter identifiability, different hydrologic signature-based objectives were incorporated into hydrologic model calibration. A signature-based multi-objective calibration resulted in more hydrologically consistent simulations but it was not able to account for errors induced by model structural deficiencies. One scheme to correct this deficiency was to combine the predictions of multiple models to compensate for a specific model's simulation error by other model components. The under-study models included HEC-HMS, HSPF, WATFLOOD, and HBV-EC for the Upper Assiniboine River Basin at Kamsack Hydrometric Station. This research was conducted on these models at Hydrologic Forecast Centre of Manitoba Infrastructure and developed a multi-model platform for the sake of more accurate predictions. This platform, also called flow-duration curve segmentation-based model-wrapper, was based on the weighted average of individual model simulations. The weight factor assigned to each model component was different for different streamflow rates as one model might perform differently in high-, mid-, and low-flow conditions. The proposed model blending method outperformed the individual model components in terms of simulation accuracy, and it was able to compute the contribution percentage of each model to each flow partition. Therefore, it could identify redundant model components and helped modellers to review their model selection presumptions.

6.2 Research significance

This dissertation has made the following contributions to the research and development of solution approaches for multi-objective water resources engineering problems.

- Improving multi-objective algorithms to effectively solve many-objective problems

- Introducing a novel approach for improving decision-space diversity in multi-objective algorithms through clustering so-far generated solutions and dominance localization to increase the probability of dissimilar alternatives in the archive set for a more flexible and well-informed decision making
- Introducing a novel approach for estimating the parameter uncertainty of hydrologic models in a multi-objective context
- Developing a new method via a flow-duration curve partitioning scheme for multi-modelling to reduce structural uncertainty in hydrologic models

The diversity-based archiving strategies in the objective and decision spaces, i.e. the rounded-objective and cluster-based archiving, developed in this dissertation were successfully applied to the PA-DDS algorithm for solving a number of multi-objective simulation-optimization water resources problems. However, these archiving methods are neither algorithm specific nor problem specific and can be implemented on other multi-objective global optimization tools.

The objective space discretization using the rounded-based archiving controls the dominance resistance issue, significantly decreases the number of alternatives, prevents from quickly populating the pre-set archive of an algorithm with a bounded-archive structure, e.g. AMALGAM, and provides a better distribution of the approximate tradeoff in many-objective applications. Having applied to the PA-DDS, Borg-MOEA, and AMALGAM algorithms, the rounded-objective archiving outperforms the traditional point-based dominance archiving and it is as effective as the ε-dominance archiving. The difference between the latter and the proposed method is in its implementation. Contrary to the ε-dominance archiving that requires changing the internal structure of the algorithm, the rounded-objective archiving is implemented within the problem formulation. The rounding level of each objective should be set *a priori* by consulting with

decision-makers in order to prevent an unnecessarily very coarse resolution, which increases the probability of multi-modality in the optimization problem.

The cluster-based archiving and decentralized dominance strategies implemented on PA-DDS is able to characterize not only the optimal front but also a number of near-optimal tradeoffs that are locally non-dominated. The cluster-based PA-DDS is compared to the original PA-DDS and the Omni-Optimizer algorithm to solve two mathematical problems and two environmental design and management problems. The original PA-DDS and the Omni-Optimizer are unable to retain near-optimal alternatives, while the cluster-based PA-DDS significantly increases diversity in the decision space and provides a number of distinct alternatives that have identical or near identical performance to optimal solutions. This scheme is advisable when a multi-objective problem is featured with multi-modality, an intrinsic property in water resources simulation-optimization problems due to the simplified modelling of the system complexity or absence of objectives and constraints not incorporated into the optimization problem formulation. The cluster-based archiving and local domination methods are neither algorithm nor problem specific and are applicable to other multi-objective global optimization algorithms with unbounded archives or with adaptive archive size. This approach can be used simultaneously with the proposed objective-space discretization providing that the rounding resolution of the objectives accords with the instructions given in Chapter 2. Implementation of the cluster-based archiving approach for diversity preservation in the decision-space can extend the application of the multi-objective optimization tools for an efficient uncertainty of hydrologic model parameters.

The application of cluster-based PA-DDS for approximating parameter uncertainty of 12-, 20-, and 13-parameter hydrologic models demonstrates that it is at least as efficient as other uncertainty approximation methods such as DDS-AU and GLUE, while it can handle multiple

calibration objectives simultaneously. The multi-objective parameter uncertainty is the main advantage of the proposed algorithm over GLUE and DDS-AU that aggregate all model performance metrics into one objective by assigning weight to each performance metric. Aggregation of performance metrics apparently expands the behavioral set due to misidentifying some of archived solutions as behavioral samples. Incorporating more number of performance metrics into a multi-objective uncertainty analysis of hydrologic models decreases the number of behavioral solutions.

Furthermore, in this Ph.D. thesis, four operational hydrologic models at Hydrologic Forecast Centre of Manitoba Infrastructure were calibrated using three suites of signature-based metrics. The calibrated models using signatures improves parameter identifiability, against the older version of the same models calibrated at the center. The flow duration curve segmentation-based model-wrapper resulting from the weighted-average combination of the calibrated individual models outperforms each of the best calibrated single models in the calibration and evaluation periods. Therefore, it is concluded that the proposed method is a reliable framework that enumerates the strengths and weaknesses of each model for simulating historically low-, mid-, and high-flow ranges. The model-wrapper decreases the unpredictable simulation noise by optimal combination of multiple hydrologic models and assign a higher weight to best performing individual model. The proposed framework is readily available for real-time operational applications and flexible to include more hydrologic models.

6.3 Limitations and Recommendations for Future Work

This study introduced methods and developed tools for water resources systems with emphases on many-objective optimization, diversity preservation for robust decision-making, multi-objective

uncertainty analysis, informed calibration of hydrologic models, and a more reliable forecasting system. Throughout this Ph.D. research, some research gaps are identified but could not be addressed, mainly because of the limited amount of time. These gaps can be explored further in the future for further advancement of the results presented in this thesis.

- The rounded archiving approach was evaluated on three multi-objective global optimization algorithms including Borg-MOEA, PA-DDS, and AMALGAM for solving many-objective optimization of hydrologic model calibration. Future work can implement the same archiving approach for other well-known multi-objective optimization algorithms, e.g. NSGA-II, SPEA2, and other problems such as many-objective reservoir operation and water distribution network design problems.

- The cluster-based archiving and local domination techniques were developed for the single-solution PA-DDS algorithm. However, these techniques can be implemented on population-based multi-objective global optimization algorithms with unbounded or adaptive archive size, e.g. NSGA-II. The clustering technique in this case would be updated once in each generation. However, the non-dominated sorting method needs to be replaced with local dominance method keeping only cluster non-dominated solutions in their adaptive archive.

- The cluster-based archiving can be developed and tested for single-objective global optimization algorithms. In this case, only one solution is retained in each cluster. However, the effectiveness of this method for finding optimal and near-optimal solutions in the case of single-objective optimization remains unclear.

- This thesis did not evaluate the effect of different selection metrics on the performance of the Cluster-Based PA-DDS. The selection metric used in Chapter 3 can be replaced with any metric that jointly considers diversity in the decision and objective space such as the variation rate metric in Cuate and Schütze (2019). The performance of CB-PA-DDS with the variation rate as a selection metric can be compared, in future, with the version developed in this thesis.

- The application of the cluster-based PA-DDS for approximating parameter uncertainty was evaluated for 10000 solutions evaluations and the performance of this method for higher or lower computational budgets is required to be assessed. Specifically, the behavior of the adjacency density parameter needs to be investigated for high and low budgets. A very high value of adjacency density parameter (ε) places all generated solutions in the decision space into one cluster representing one Pareto front after the optimization process. The cluster-based archiving can be computationally intensive if clusters contain very high number of solutions that are locally non-dominated in their clusters.

- The multi-objective model parameter uncertainty analysis in this dissertation was conducted on a single-response hydrologic model, i.e. streamflow time series as the output. In cases where the parameter uncertainty is estimated for multi-response hydrologic models, it is recommended to define at least one objective function for each model response provided that field measurements for these responses are available.

- The archive size of CB-PA-DDS algorithm can be large as it may contain less desirable solutions besides the optimal and near-optimal clustered solutions. Identifying and screening less desirable solutions from the archive of CB-PA-DDS was performed

separately after the optimization. The post-processing stage can be coupled with the CB-PA-DDS algorithm to automatically screen the archives and keep only optimal and near-optimal solutions provided that there is an access to the preferred range of objectives by decision-makers. Therefore, future research can focus on developing an interactive advanced visualization technique for post-processing and truncating the archive set for stakeholders. VIDEO (Kollat and Reed, 2007) and moGrams (Trawiński et al., 2018) are two examples of visualization techniques.

- The WATFLOOD model in Chapter 5 was calibrated with 114 parameters. Model calibration with a large number of parameters generally requires a very high number of solution evaluations. Large model parameterization also limits the identifiability of parameters. Therefore, grouping parameters is suggested for a model like WATFLOOD, similar to the approach used for the curve number parameter of SWAT in Abbaspour (2013). Future works can focus on decreasing the number of parameters by implementing regularization methods on the parameters that exploits additional information from the hydro-system or constrains the feasible range of parameters not to take on arbitrary values (Kavetski et al., 2006). For example, modellers can investigate the spatial dependency of model parameters on the properties of a watershed such as soil or land cover (Pokhrel et al., 2012, 2008).

- The multi-objective calibration of models in Chapter 5 was based on the hydrologic signatures that were derived from the measured streamflow time series. Future research can include other signatures that describe other components of hydrological processes, such as groundwater, evapotranspiration, isotopes, and/or snowmelt for the sake of a better-posed set of parameters. The model calibrations were also based on temporal evaluation of

the simulations. However, spatial evaluation of the calibrated parameter sets is highly encouraged if multiple hydrometric gauges exist across the watershed.

- The robustness of the proposed model-wrapper is limited to the methodological choices in this thesis. One choice is utilizing only four hydrologic models for only one catchment. Future research on multi-modelling can concentrate on drawing a more comprehensive conclusion by incorporating a larger number of hydrologic models into model-wrapper for several catchments with significantly different catchment sizes and dominant hydrological processes.

7 LIST OF REFERENCES

Abbaspour, K.C., 2013. SWAT-CUP 2012 SWAT Calibration and Uncertainty Programs. Duebendorf, Switzerland.

Abbaspour, K.C., Johnson, C.A., van Genuchten, M.T., 2004. Estimating Uncertain Flow and Transport Parameters Using a Sequential Uncertainty Fitting Procedure. Vadose Zo. J. 3, 1340–1352. https://doi.org/10.2136/vzj2004.1340

Abrahart, R.J., See, L., 2002. Multi-model data fusion for river flow forecasting: an evaluation of six alternative methods based on two contrasting catchments. Hydrol. Earth Syst. Sci. Discuss. Eur. Geosci. Union 6, 655–670.

Ahmadi, A., Nasseri, M., Solomatine, D.P., 2019. Parametric uncertainty assessment of hydrological models: coupling UNEEC-P and a fuzzy general regression neural network. Hydrol. Sci. J. 64, 1080–1094. https://doi.org/10.1080/02626667.2019.1610565

Ahmadi, M., Arabi, M., Ascough, J.C., Fontane, D.G., Engel, B.A., 2014. Toward improved calibration of watershed models: Multisite multiobjective measures of information. Environ. Model. Softw. 59, 135–145. https://doi.org/10.1016/j.envsoft.2014.05.012

Ahmari, H., Blais, E.L., Greshuk, J., 2016. The 2014 flood event in the Assiniboine River Basin: Causes, assessment and damage. Can. Water Resour. J. 41, 85–93. https://doi.org/10.1080/07011784.2015.1070695

Ajami, N.K., Duan, Q., Gao, X., Sorooshian, S., 2006. Multimodel combination techniques for analysis of hydrological simulations: Application to distributed model intercomparison project results. J. Hydrometeorol. 7, 755–768. https://doi.org/10.1175/JHM519.1

Ajami, N.K., Duan, Q., Sorooshian, S., 2007. An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resour. Res. 43. https://doi.org/10.1029/2005WR004745

Albert, C., Künsch, H.R., Scheidegger, A., 2014. A simulated annealing approach to approximate Bayes computations. Stat. Comput. 2014 256 25, 1217–1232. https://doi.org/10.1007/S11222-014-9507-8

Alberta WaterSMART, 2013. The 2013 Great Alberta Flood: Actions to Mitigate, Manage and Control Future Floods. Edmonton,AB.

Alvisi, S., Franchini, M., 2013. A grey-based method for evaluating the effects of rating curve uncertainty on frequency analysis of annual maxima. J. Hydroinformatics 15, 194–210. https://doi.org/10.2166/HYDRO.2012.127

Arnold, J.G., Allen, P.M., 1999. Automated methods for estimating baseflow and ground water recharge from streamflow records. J. Am. Water Resour. Assoc. 35, 411–424. https://doi.org/10.1111/j.1752-1688.1999.tb03599.x

Arnold, J.G., Williams, J.R., Nicks, A.D., Sammons, N.B., 1990. SWRRB; a basin scale simulation model for soil and water resources management., SWRRB; a basin scale simulation model for soil and water resources management. Texas A & M University Press, College Station, Texas, USA.

Arnold, J.G., Youssef, M.A., Yen, H., White, M.J., Sheshukov, A.Y., Sadeghi, A.M., Moriasi, D.N., Steiner, J.L., Amatya, D.M., Skaggs, R.W., Haney, E.B., Jeong, J., Arabi, M., Gowda, P.H., 2015. Hydrological Processes and Model Representation: Impact of Soft Data on

Calibration. Trans. ASABE 58, 1637–1660. https://doi.org/10.13031/trans.58.10726

Arsenault, R., Gatien, P., Renaud, B., Brissette, F., Martel, J.L., 2015. A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation. J. Hydrol. 529, 754–767. https://doi.org/10.1016/j.jhydrol.2015.09.001

Asadzadeh, M., Leon, L., McCrimmon, C., Yang, W., Liu, Y., Wong, I., Fong, P., Bowen, G., 2015. Watershed derived nutrients for Lake Ontario inflows: Model calibration considering typical land operations in Southern Ontario. J. Great Lakes Res. 41, 1037–1051. https://doi.org/10.1016/j.jglr.2015.09.002

Asadzadeh, M., Leon, L., Yang, W., Bosch, D., 2016. One-day offset in daily hydrologic modeling: An exploration of the issue in automatic model calibration. J. Hydrol. 534, 164–177. https://doi.org/10.1016/j.jhydrol.2015.12.056

Asadzadeh, M., Razavi, S., Tolson, B.A., Fay, D., 2014a. Pre-emption strategies for efficient multi-objective optimization: Application to the development of Lake Superior regulation plan. Environ. Model. Softw. 54, 128–141. https://doi.org/10.1016/j.envsoft.2014.01.005

Asadzadeh, M., Tolson, B., 2013. Pareto archived dynamically dimensioned search with hypervolume-based selection for multi-objective optimization. Eng. Optim. 45, 1489–1509. https://doi.org/10.1080/0305215X.2012.748046

Asadzadeh, M., Tolson, B., 2012. Hybrid Pareto archived dynamically dimensioned search for multi-objective combinatorial optimization: Application to water distribution network design. J. Hydroinformatics 14, 192–205. https://doi.org/10.2166/hydro.2011.098

Asadzadeh, M., Tolson, B.A., Burn, D.H., 2014b. A new selection metric for multiobjective

hydrologic model calibration. Water Resour. Res. 50, 7082–7099. https://doi.org/10.1002/2013WR014970

Asadzadeh, M., Tolson, B.A., McKillop, R., 2012. A two stage Optimization approach for calibrating water distribution systems, in: Water Distribution Systems Analysis 2010 - Proceedings of the 12th International Conference, WDSA 2010. American Society of Civil Engineers, pp. 1682–1694. https://doi.org/10.1061/41203(425)148

Bader, J., Zitzler, E., 2011. HypE: An algorithm for fast hypervolume-based many-objective optimization. Evol. Comput. 19, 45–76. https://doi.org/10.1162/EVCO_a_00009

Balin, D., 2004. Hydrological behaviour through experimental and modelling approaches. Application to the Haute-Mentue catchment. Thèse. ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE (EPFL).

Barán, B., Von Lücken, C., Sotelo, A., 2005. Multi-objective pump scheduling optimisation using evolutionary strategies, in: Advances in Engineering Software. Elsevier Ltd, pp. 39–47. https://doi.org/10.1016/j.advengsoft.2004.03.012

Bartelt-Hunt, S.L., Culver, T.B., Smith, J.A., Matott, L.S., Rabideau, A.J., 2006. Optimal design of a compacted soil liner containing sorptive amendments. J. Environ. Eng. 132, 769–776. https://doi.org/10.1061/(ASCE)0733-9372(2006)132:7(769)

Bates, J.M., Granger, C.W.J., 1969. The Combination of Forecasts. J. Oper. Res. Soc. 20, 451–468. https://doi.org/10.1057/jors.1969.103

Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian Computation in Population Genetics. Genetics 162, 2025–2035.

https://doi.org/10.1093/GENETICS/162.4.2025

Beck, H.E., van Dijk, A.I.J.M., de Roo, A., Miralles, D.G., McVicar, T.R., Schellekens, J., Bruijnzeel, L.A., 2016. Global-scale regionalization of hydrologic model parameters. Water Resour. Res. 52, 3599–3622. https://doi.org/10.1002/2015WR018247

Beck, M.B., 1987. Water quality modeling: A review of the analysis of uncertainty. Water Resour. Res. 23, 1393–1442. https://doi.org/10.1029/WR023i008p01393

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environ. Model. Softw. 40, 1–20. https://doi.org/10.1016/j.envsoft.2012.09.011

Bergström, S., 1995. The HBV Model, in: Singh, V.P. (Ed.), Computer Models of Watershed Hydrology. Rev. Ed. Water Resouces Publications, Highlands Ranch, Colorado, pp. 443–476.

Bergström, S., Lindström, G., Pettersson, A., 2002. Multi-variable parameter estimation to increase confidence in hydrological modelling. Hydrol. Process. 16, 413–421. https://doi.org/10.1002/hyp.332

Beven, K., 2006. A manifesto for the equifinality thesis, in: Journal of Hydrology. Elsevier, pp. 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Adv. Water Resour. 16, 41–51. https://doi.org/10.1016/0309-1708(93)90028-E

Beven, K., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrol. Process. 6, 279–298. https://doi.org/10.1002/hyp.3360060305

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J. Hydrol. 249, 11–29.

Beven, K., Smith, P., 2015. Concepts of information content and likelihood in parameter calibration for hydrological simulation models. J. Hydrol. Eng. 20, A4014010 (1-15). https://doi.org/10.1061/(ASCE)HE.1943-5584.0000991

Beven, K.J., Smith, P.J., Freer, J.E., 2008. So just why would a modeller choose to be incoherent? J. Hydrol. 354, 15–32. https://doi.org/10.1016/j.jhydrol.2008.02.007

Bicknell, B.R., Imhoff, J.C., Kittle Jr, J.L., Donigian Jr, A.S., Johanson, R.C., 1996. HYDROLOGICAL SIMULATION PROGRAM-FORTRAN USER'S MANUAL FOR RELEASE 11. Athens, GA.

Bicknell, B.R., Imhoff, J.C., Kittle Jr, J.L., Jobes, T.H., Donigian Jr, A.S., 2001. Hydrological Simulation Program-Fortran (HSPF). User's Manual for Release 12. Athens, GA.

Blais, E.L., Clark, S., Dow, K., Rannie, B., Stadnyk, T., Wazney, L., 2016a. Background to flood control measures in the Red and Assiniboine River Basins. Can. Water Resour. J. 41, 31–44. https://doi.org/10.1080/07011784.2015.1036123

Blais, E.L., Greshuk, J., Stadnyk, T., 2016b. The 2011 flood event in the Assiniboine River Basin: causes, assessment and damages. Can. Water Resour. J. https://doi.org/10.1080/07011784.2015.1046139

Blasone, R.S., Madsen, H., Rosbjerg, D., 2008a. Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling. J. Hydrol. 353, 18–32. https://doi.org/10.1016/j.jhydrol.2007.12.026

Blasone, R.S., Vrugt, J.A., Madsen, H., Rosbjerg, D., Robinson, B.A., Zyvoloski, G.A., 2008b. Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling. Adv. Water Resour. 31, 630–648. https://doi.org/10.1016/j.advwatres.2007.12.003

Blazkova, S., Beven, K.J., Kulasova, A., 2002. On constraining TOPMODEL hydrograph simulations using partial saturated area information. Hydrol. Process. 16, 441–458. https://doi.org/10.1002/hyp.331

Bode, F., Reed, P., Reuschen, S., Nowak, W., 2019. Search Space Representation and Reduction Methods to Enhance Multiobjective Water Supply Monitoring Design. Water Resour. Res. 55, 2257–2278. https://doi.org/10.1029/2018WR023133

Box, G.E.P., Tiao, G.C., 1992. BAYESIAN INFERENCE IN STATISTICAL ANALYSIS. John Wiley & Sons.

Boyle, D.P., Gupta, H. V., Sorooshian, S., 2000. Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. Water Resour. Res. 36, 3663–3674. https://doi.org/10.1029/2000WR900207

Brazil, L., Krajewski, W., 1987. Optimization of complex hydrologic models using random search methods, in: Paper Presented at Conference on Engineering Hydrology, Hydraulics Division, Am. Soc. of Civ. Eng. Williamsburg, Va.

Brill, E.D., Chang, S.Y., Hopkins, L.D., 1982. MODELING TO GENERATE ALTERNATIVES: THE HSJ APPROACH AND AN ILLUSTRATION USING A PROBLEM IN LAND USE PLANNING. Manage. Sci. 28, 221–235. https://doi.org/10.1287/mnsc.28.3.221

Brill, E.D., Flach, J.J.M., Lewis D., H., Ranjithan, S., 1990. MGA: A Decision Support System for Complex, Incompletely Defined Problems. IEEE Trans. Syst. MAN, C'YBLRNFTIC 20, 745–757.

Budhathoki, S., Rokaya, P., Lindenschmidt, K.E., Davison, B., 2020. A multi-objective calibration approach using in-situ soil moisture data for improved hydrological simulation of the Prairies. Hydrol. Sci. J. 65, 638–649. https://doi.org/10.1080/02626667.2020.1715982

Burn, D.H., Whitfield, P.H., 2016. Changes in floods and flood regimes in Canada. Can. Water Resour. J. 41, 139–150. https://doi.org/10.1080/07011784.2015.1026844

Burton, R.O., Gidley, J.S., Baker, B.S., Reda-Wilson, K.J., 1987. Nearly Optimal Linear Programming Solutions: Some Conceptual Issues and a Farm Management Application. Am. J. Agric. Econ. 69, 813–818. https://doi.org/10.2307/1242192

Buttle, J.M., Allen, D.M., Caissie, D., Davison, B., Hayashi, M., Peters, D.L., Pomeroy, J.W., Simonovic, S., St-Hilaire, A., Whitfield, P.H., 2016. Flood processes in Canada: Regional and special aspects. Can. Water Resour. J. 41, 7–30. https://doi.org/10.1080/07011784.2015.1131629

Canadian Hydraulic Centre, 2010. Green Kenue Reference Manual. National Research Council, Ottawa, Ontario, CA.

Cappé, O., Guillin, A., Marin, J.M., Robert, C.P., 2004. Population Monte Carlo. J. Comput.

Graph. Stat. 13, 907–929. https://doi.org/10.1198/106186004X12803

Carpenter, S.R., Ludwig, ; D, Brock, W.A., 1999. Management of Eutrophication for Lakes Subject to Potentially Irreversible Change. Ecol. Appl. 9, 751–771.

Carpitella, S., Brentan, B., Montalvo, I., Izquierdo, J., Certa, A., 2019. Multi-criteria analysis applied to multi-objective optimal pump scheduling in water systems. Water Sci. Technol. Water Supply 19, 2338–2346. https://doi.org/10.2166/ws.2019.115

Carrano, E.G., Wanner, E.F., Takahashi, R.H.C., 2011. A multicriteria statistical based comparison methodology for evaluating evolutionary algorithms. IEEE Trans. Evol. Comput. 15, 848–870. https://doi.org/10.1109/TEVC.2010.2069567

Carsell, K.M., Pingel, N.D., Ford, D.T., 2004. Quantifying the Benefit of a Flood Warning System. Nat. Hazards Rev. 5, 131–140. https://doi.org/10.1061/(ASCE)1527-6988(2004)5:3(131)

Cavicchio, D., 1970. Adaptive search using simulated evolution. University of Michigan.

Chang, S. -Y, Brill, E.D., Hopkins, L.D., 1982. Use of mathematical models to generate alternative solutions to water resources planning problems. Water Resour. Res. 18, 58–64. https://doi.org/10.1029/WR018i001p00058

Chen, L., Zhang, Y., Zhou, J., Singh, V.P., Guo, S., Zhang, J., 2015. Real-time error correction method combined with combination flood forecasting technique for improving the accuracy of flood forecasting. J. Hydrol. 521, 157–169. https://doi.org/10.1016/j.jhydrol.2014.11.053

Cheng, R., Li, M., Li, K., Yao, X., 2018. Evolutionary Multiobjective Optimization-Based Multimodal Optimization: Fitness Landscape Approximation and Peak Detection. IEEE Trans. Evol. Comput. 22, 692–706. https://doi.org/10.1109/TEVC.2017.2744328

Chmielewski, H., 2013. Satisfying Multiple Priorities with a Diversity Preserving Evolutionary Algorithm. North Carolina University, Raleigh, North Carolina.

Christiaens, K., Feyen, J., 2002. Constraining soil hydraulic parameter and output uncertainty of the distributed hydrological MIKE SHE model using the GLUE framework. Hydrol. Process. 16, 373–391. https://doi.org/10.1002/hyp.335

Clark, M.P., Kavetski, D., Fenicia, F., 2012. Reply to comment by K. Beven et al. on "'Pursuing the method of multiple working hypotheses for hydrological modeling.'" Water Resour. Res. 48, W11802. https://doi.org/10.1029/2012WR012547

Clark, M.P., Kavetski, D., Fenicia, F., 2011. Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resour. Res. 47, 9301. https://doi.org/10.1029/2010WR009827

Coello, C., Lamont, G., Van Veldhuizen, D., 2007. Evolutionary algorithms for solving multi-objective problems. Springer, New York.

Craig, J., 2015. RAVEN's User Manual. Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, Ont., CA.

Csilléry, K., Blum, M.G.B., Gaggiotti, O.E., François, O., 2010. Approximate Bayesian Computation (ABC) in practice. Trends Ecol. Evol. 25, 410–418. https://doi.org/10.1016/J.TREE.2010.04.001

Cuate, O., Schütze, O., 2019. Variation Rate to Maintain Diversity in Decision Space within Multi-Objective Evolutionary Algorithms. Math. Comput. Appl. 24, 82. https://doi.org/10.3390/mca24030082

De Jong, K., 1975. Analysis of the behavior of a class of genetic adaptive systems. University of Michigan.

Dean, S., Freer, J., Beven, K., Wade, A.J., Butterfield, D., 2009. Uncertainty assessment of a process-based integrated catchment model of phosphorus. Stoch. Environ. Res. Risk Assess. 23, 991–1010. https://doi.org/10.1007/s00477-008-0273-z

Deb, K., 2001. Multi-Objective Optimization using Evolutionary Algorithms. John Wiley & Sons.

Deb, K., Beyer, H.G., 2001. Self-adaptive genetic algorithms with simulated binary crossover. Evol. Comput. 9, 197–221. https://doi.org/10.1162/106365601750190406

Deb, K., Jain, H., 2014. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, Part I: Solving problems with box constraints. IEEE Trans. Evol. Comput. 18, 577–601. https://doi.org/10.1109/TEVC.2013.2281535

Deb, K., Joshi, D., Anand, A., 2002a. Real-coded evolutionary algorithms with parent-centric recombination, in: Proceedings of the 2002 Congress on Evolutionary Computation, CEC 2002. IEEE Computer Society, pp. 61–66. https://doi.org/10.1109/CEC.2002.1006210

Deb, K., Mohan, M., Mishra, S., 2003. Towards a quick computation of well-spread Pareto-optimal solutions, in: Fonseca C.M., Fleming P.J., Zitzler E., Thiele L., D.K. (Ed.), Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, Berlin, pp. 222–236. https://doi.org/10.1007/3-540-36970-8_16

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002b. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. IEEE Trans. Evol. Comput. 6, 182–197.

Deb, K., Thiele, L., Laumanns, M., Zitzler, E., 2005. Scalable Test Problems for Evolutionary Multiobjective Optimization, in: Abraham, A.., Jain, L.., Goldberg, R. (Eds.), Evolutionary Multiobjective Optimization. Advanced Information and Knowledge Processing. Springer-Verlag, London, pp. 105–145. https://doi.org/10.1007/1-84628-137-7_6

Deb, K., Tiwari, S., 2008. Omni-optimizer: A generic evolutionary algorithm for single and multi-objective optimization. Eur. J. Oper. Res. 185, 1062–1087. https://doi.org/10.1016/j.ejor.2006.06.042

Di Matteo, M., Maier, H.R., Dandy, G.C., 2019. Many-objective portfolio optimization approach for stormwater management project selection encouraging decision maker buy-in. Environ. Model. Softw. 111, 340–355. https://doi.org/10.1016/j.envsoft.2018.09.008

di Pierro, F., Khu, S.T., Savić, D.A., 2007. An investigation on preference order ranking scheme for multiobjective evolutionary optimization. IEEE Trans. Evol. Comput. 11, 17–45. https://doi.org/10.1109/TEVC.2006.876362

Dotto, C.B.S., Mannina, G., Kleidorfer, M., Vezzaro, L., Henrichs, M., McCarthy, D.T., Freni, G., Rauch, W., Deletic, A., 2012. Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. Water Res. 46, 2545–2558. https://doi.org/10.1016/j.watres.2012.02.009

Drechsler, N., Drechsler, R., Becker, B., 2001. Multi-objective optimisation based on relation favour, in: In International Conference on Evolutionary Multi-Criterion Optimization. Springer Verlag, Berlin Heidelberg, pp. 154–166. https://doi.org/10.1007/3-540-44719-9_11

Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S., 2007. Multi-model ensemble hydrologic

prediction using Bayesian model averaging. Adv. Water Resour. 30, 1371–1386. https://doi.org/10.1016/j.advwatres.2006.11.014

Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour. Res. 28, 1015–1031. https://doi.org/10.1029/91WR02985

Duan, Q., Sorooshian, S., Gupta, V.K., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. J. Hydrol. 158, 265–284. https://doi.org/10.1016/0022-1694(94)90057-4

Duda, P.B., Hummel, P.R., Donigian, A.S., Imhoff, J.C., 2012. BASINS/HSPF: MODEL USE, CALIBRATION, AND VALIDATION. Trans. ASABE 55, 1523–1547.

Dunning, D., Ross, Q., Policy, M.M.-E.S., 2000, U., 2000. Multiattribute utility analysis; best technology available; adverse environmental impact; Clean Water Act; Section 316 (b). Environ. Sci. Policy 3, 7–14.

Efstratiadis, A., Koutsoyiannis, D., 2010. One decade of multi-objective calibration approaches in hydrological modelling: a review. Hydrol. Sci. J. 55, 58–78. https://doi.org/10.1080/02626660903526292

Eker, S., Kwakkel, J.H., 2018. Including robustness considerations in the search phase of Many-Objective Robust Decision Making. Environ. Model. Softw. 105, 201–216. https://doi.org/10.1016/j.envsoft.2018.03.029

Ercan, M.B., Goodall, J.L., 2016. Design and implementation of a general software library for using NSGA-II with SWAT for multi-objective model calibration. Environ. Model. Softw.

84, 112–120. https://doi.org/10.1016/j.envsoft.2016.06.017

Erfani, T., Pachos, K., Harou, J.J., 2020. Decision-dependent uncertainty in adaptive real-options water resource planning. Adv. Water Resour. 136, 103490. https://doi.org/10.1016/j.advwatres.2019.103490

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, in: Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press, pp. 226–231.

Evensen, G., 1992. Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. J. Geophys. Res. 97. https://doi.org/10.1029/92jc01972

Fang, X., Minke, A., Pomeroy, J., Brown, T., Westbrook, C., Guo, X., Guangul, S., 2007. A Review of Canadian Prairie Hydrology: Principles, Modelling and Response to Land Use and Drainage Change. Saskatoon, SK.

Farina, M., Amato, P., 2004. A fuzzy definition of "optimality" for many-criteria optimization problems. IEEE Trans. Syst. Man, Cybern. Part ASystems Humans. 34, 315–326. https://doi.org/10.1109/TSMCA.2004.824873

Farina, M., Amato, P., 2002. On the optimal solution definition for many-criteria optimization problems. Annu. Conf. North Am. Fuzzy Inf. Process. Soc. - NAFIPS 2002-Janua, 233–238. https://doi.org/10.1109/NAFIPS.2002.1018061

Farmani, R., Savic, D.A., Walters, G.A., 2005. Evolutionary multi-objective optimization in water distribution network design. Eng. Optim. 37, 167–183.

https://doi.org/10.1080/03052150512331303436

Faybishenko, B., 2010. Fuzzy-probabilistic calculations of water-balance uncertainty. Stoch. Environ. Res. Risk Assess. 2010 246 24, 939–952. https://doi.org/10.1007/S00477-010-0379-Y

Feldman, A.D., 2000. Hydrologic Modeling System HEC-HMS: Technical Reference Manual. US Army Corps of Engineers, Hydrologic Forecast Center, Davis, California, USA.

Fenicia, F., Kavetski, D., Reichert, P., Albert, C., 2018. Signature-Domain Calibration of Hydrological Models Using Approximate Bayesian Computation: Empirical Analysis of Fundamental Properties. Water Resour. Res. 54, 3958–3987. https://doi.org/10.1002/2017WR021616

Flood Recovery Task Force, 2013. Southern Alberta 2013 Floods: The Provincial Recovery Framework. Edmonton,AB.

Fonseca, C.M., Paquete, L., López-Ibáñez, M., 2006. An improved dimension-sweep algorithm for the hypervolume indicator, in: 2006 IEEE Congress on Evolutionary Computation, CEC 2006. pp. 1157–1163. https://doi.org/10.1109/cec.2006.1688440

Franks, S.W., Gineste, P., Beven, K.J., Merot, P., 1998. On constraining the predictions of a distributed model: The incorporation of fuzzy estimates of saturated areas into the calibration process. WATER Resour. Res. 34, 787–797. https://doi.org/10.1029/97WR03041

Freer, J.E., McMillan, H., McDonnell, J.J., Beven, K.J., 2004. Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures. J. Hydrol. 291, 254–277. https://doi.org/10.1016/J.JHYDROL.2003.12.037

Freni, G., Mannina, G., 2010. Bayesian approach for uncertainty quantification in water quality modelling: The influence of prior distribution. J. Hydrol. 392, 31–39. https://doi.org/10.1016/j.jhydrol.2010.07.043

Fu, G., Kapelan, Z., Kasprzyk, J.R., Reed, P., 2013. Optimal Design of Water Distribution Systems Using Many-Objective Visual Analytics. J. Water Resour. Plan. Manag. 139, 624–633. https://doi.org/10.1061/(asce)wr.1943-5452.0000311

Gallart, F., Latron, J., Llorens, P., Beven, K., 2007. Using internal catchment information to reduce the uncertainty of discharge and baseflow predictions. Adv. Water Resour. 30, 808–823. https://doi.org/10.1016/j.advwatres.2006.06.005

Georgakakos, K.P., Seo, D.J., Gupta, H., Schaake, J., Butts, M.B., 2004. Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, in: Journal of Hydrology. Elsevier, pp. 222–241. https://doi.org/10.1016/j.jhydrol.2004.03.037

Geressu, R.T., Harou, J.J., 2019. Reservoir system expansion scheduling under conflicting interests. Environ. Model. Softw. 118, 201–210. https://doi.org/10.1016/j.envsoft.2019.04.002

Giuliani, M., Galelli, S., Soncini-Sessa, R., 2014a. A dimensionality reduction approach for many-objective Markov Decision Processes: Application to a water reservoir operation problem. Environ. Model. Softw. 57, 101–114. https://doi.org/10.1016/j.envsoft.2014.02.011

Giuliani, M., Herman, J.D., Castelletti, A., Reed, P., 2014b. Many-objective reservoir policy identification and refinement to reduce policy inertia and myopia in water management. Water Resour. Res. 50, 3355–3377. https://doi.org/10.1002/2013WR014700

Goldberg, D., Richardson, J., 1987. Genetic algorithms with sharing for multimodal function optimization, in: Genetic Algorithms and Their Applications: Proceedings of the Second International Conference on Genetic Algorithms. Lawrence Erlbaum, Hillsdale, NJ, pp. 41–49.

Gong, D., Sun, J., Ji, X., 2013. Evolutionary algorithms with preference polyhedron for interval multi-objective optimization problems. Inf. Sci. (Ny). 233, 141–161. https://doi.org/10.1016/j.ins.2013.01.020

Gong, D., Xu, B., Zhang, Y., Guo, Y., Yang, S., 2020. A Similarity-Based Cooperative Co-Evolutionary Algorithm for Dynamic Interval Multiobjective Optimization Problems. IEEE Trans. Evol. Comput. 24, 142–156. https://doi.org/10.1109/TEVC.2019.2912204

Goswami, M., O'Connor, K.M., Bhattarai, K.P., 2007. Development of regionalisation procedures using a multi-model approach for flow simulation in an ungauged catchment. J. Hydrol. 333, 517–531. https://doi.org/10.1016/j.jhydrol.2006.09.018

Grillakis, M.G., Tsanis, I.K., Koutroulis, A.G., 2010. Application of the HBV hydrological model in a flash flood case in Slovenia. Nat. Hazards Earth Syst. Sci. 10, 2713–2725. https://doi.org/10.5194/nhess-10-2713-2010

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. Water Resour. Res. 34, 751–763. https://doi.org/10.1029/97WR03495

Gupta, H. V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J.

Hydrol. 377, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Gupta, H. V., Razavi, S., 2018. Revisiting the Basis of Sensitivity Analysis for Dynamical Earth System Models. Water Resour. Res. 54, 8692–8717. https://doi.org/10.1029/2018WR022668

Gupta, H. V, Beven, K.J., Wagener, T., 2005. Model Calibration and Uncertainty Estimation, in: Encyclopedia of Hydrological Sciences. John Wiley & Sons, Ltd, Chichester, UK. https://doi.org/10.1002/0470848944.hsa138

Guzman, J.A., Shirmohammadi, A., Sadeghi, A.M., Wang, X., Chu, M.L., Jha, M.K., Parajuli, P.B., Harmel, R.D., Khare, Y.P., Hernandez, J.E., 2015. Uncertainty Considerations in Calibration and Validation of Hydrologic and Water Quality Models. Trans. ASABE 58, 1745–1762. https://doi.org/10.13031/trans.58.10710

Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive Metropolis algorithm. Bernoulli 7, 223–242.

Hadka, D., Reed, P., 2015. Large-scale parallelization of the Borg multiobjective evolutionary algorithm to enhance the management of complex environmental systems. Environ. Model. Softw. 69, 353–369. https://doi.org/10.1016/j.envsoft.2014.10.014

Hadka, D., Reed, P., 2013. Borg: An auto-adaptive many-objective evolutionary computing framework. Evol. Comput. 21, 231–259. https://doi.org/10.1162/EVCO_a_00075

Hadka, D., Reed, P., 2012. Diagnostic assessment of search controls and failure modes in many-objective evolutionary optimization. Evol. Comput. 20, 423–452. https://doi.org/10.1162/EVCO_a_00053

Hajkowicz, S., Collins, K., 2007. A review of multiple criteria analysis for water resource planning

and management. Water Resour. Manag. https://doi.org/10.1007/s11269-006-9112-5

Hamilton, A.S., Hutchinson, D.G., Moore, R.D., 2000. Estimating Winter Streamflow Using Conceptual Streamflow Model. J. Cold Reg. Eng. 14, 158–175. https://doi.org/10.1061/(ASCE)0887-381X(2000)14:4(158)

Hanne, T., 1999. On the convergence of multiobjective evolutionary algorithms. Eur. J. Oper. Res. 117, 553–564. https://doi.org/10.1016/S0377-2217(98)00262-8

Harrington, J.J., Gidley, J.S., 1985. The Variability of Alternative Decisions in a Water Resources Planning Problem. Water Resour. Res. 21, 1831–1840. https://doi.org/10.1029/WR021i012p01831

Hashino, T., Bradley, A.A., Schwartz, S.S., 2007. Evaluation of bias-correction methods for ensemble streamflow volume forecasts. Hydrol. Earth Syst. Sci. 11, 939–950. https://doi.org/10.5194/hess-11-939-2007

Hayashi, M., Van Der Kamp, G., Schmidt, R., 2003. Focused infiltration of snowmelt water in partially frozen soil under small depressions. J. Hydrol. 270, 214–229. https://doi.org/10.1016/S0022-1694(02)00287-1

He, Y., Wetterhall, F., Cloke, H.L., Pappenberger, F., Wilson, M., Freer, J., McGregor, G., 2009. Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions. Meteorol. Appl. 16, 91–101. https://doi.org/10.1002/met.132

Hoeting, J.A.., Madigan, D., Raftery, A.E.., Volinsky, C.T.;, 1999. Bayesian Model Averaging: A Tutorial. Stat. Sci. 14, 382–401.

Hoffman, F., Hammonds, J., 1994. Propagation of uncertainty in risk assessments: the need to

distinguish between uncertainty due to lack of knowledge and uncertainty due to variabi... - PubMed - NCBI. Risk Anal. 14, 707–712.

Holmes, T., Stadnyk, T.A., Kim, S.J., Asadzadeh, M., 2020. Regional Calibration With Isotope Tracers Using a Spatially Distributed Model: A Comparison of Methods. Water Resour. Res. 56. https://doi.org/10.1029/2020WR027447

Hooke, R., Jeeves, T.A., 1961. "Direct Search" Solution of Numerical and Statistical Problems. J. ACM 8, 212–229. https://doi.org/10.1145/321062.321069

Horn, J., Nafpliotis, N., Goldberg PREPRINT, D.E., Goldberg, D.E., 1994. A Niched Pareto Genetic Algorithm for Multiobjective Optimization, IEEE World Congress on Computational Intelligence. IEEE Service Center.

Huang, Y., Chen, X., Li, Y.P., Huang, G.H., Liu, T., 2010. A fuzzy-based simulation method for modelling hydrological processes under uncertainty. Hydrol. Process. 24, 3718–3732. https://doi.org/10.1002/HYP.7790

Huang, Y., Zheng, F., Duan, H.F., Zhang, Q., 2020. Multi-Objective Optimal Design of Water Distribution Networks Accounting for Transient Impacts. Water Resour. Manag. 34, 1517–1534. https://doi.org/10.1007/s11269-020-02517-4

Hutton, C.J., Kapelan, Z., Vamvakeridou-Lyroudia, L., Savić, D., 2014. Application of formal and informal Bayesian methods for water distribution hydraulic model calibration. J. Water Resour. Plan. Manag. 140, 04014030 (1–10). https://doi.org/10.1061/(ASCE)WR.1943-5452.0000412

Ikeda, K., Kita, H., Kobayashi, S., 2001. Failure of Pareto-based MOEAs: Does non-dominated

really mean near to optimal?, in: Proceedings of the IEEE Conference on Evolutionary Computation, ICEC. pp. 957–962. https://doi.org/10.1109/cec.2001.934293

Iman, R., Davenport, J., Zeigler, D., 1980. Latin hypercube sampling (program user's guide).[LHC, in FORTRAN] (No. SAND-79-1473). Albuquerque, NM (USA).

Jacquin, A.P., Shamseldin, A.Y., 2007. Development of a possibilistic method for the evaluation of predictive uncertainty in rainfall-runoff modeling. Water Resour. Res. 43, 4425. https://doi.org/10.1029/2006WR005072

Jakob, M., Church, M., 2011. The Trouble with Floods. Can. Water Resour. J. / Rev. Can. des ressources hydriques 36, 287–292. https://doi.org/10.4296/cwrj3604928

Jeong, D. Il, Kim, Y.O., 2009. Combining single-value streamflow forecasts - A review and guidelines for selecting techniques. J. Hydrol. 377, 284–299. https://doi.org/10.1016/j.jhydrol.2009.08.028

Jiang, Y., Liu, C., Li, X., Liu, L., Wang, H., 2015. Rainfall-runoff modeling, parameter estimation and sensitivity analysis in a semiarid catchment. Environ. Model. Softw. 67, 72–88. https://doi.org/10.1016/j.envsoft.2015.01.008

Jin, X., Xu, C.Y., Zhang, Q., Singh, V.P., 2010. Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. J. Hydrol. 383, 147–155. https://doi.org/10.1016/j.jhydrol.2009.12.028

Kan, G., He, X., Ding, L., Li, J., Hong, Y., Liang, K., 2020. Heterogeneous parallel computing accelerated generalized likelihood uncertainty estimation (GLUE) method for fast hydrological model uncertainty analysis purpose. Eng. Comput. 36, 75–96.

https://doi.org/10.1007/s00366-018-0685-4

Kasprzyk, J.R., Nataraj, S., Reed, P.M., Lempert, R.J., 2013. Many objective robust decision making for complex environmental systems undergoing change. Environ. Model. Softw. 42, 55–71. https://doi.org/10.1016/j.envsoft.2012.12.007

Kasprzyk, J.R., Reed, P.M., Kirsch, B.R., Characklis, G.W., 2009. Managing population and drought risks using many-objective water portfolio planning under uncertainty. Water Resour. Res. 45. https://doi.org/10.1029/2009WR008121

Kavetski, D., Fenicia, F., Reichert, P., Albert, C., 2018. Signature-Domain Calibration of Hydrological Models Using Approximate Bayesian Computation: Theory and Comparison to Existing Applications. Water Resour. Res. 54, 4059–4083. https://doi.org/10.1002/2017WR020528

Kavetski, D., Franks, S.W., Kuczera, G., 2003. Confronting input uncertainty in environmental modelling, in: Duan et al, Q.. (Ed.), Calibration of Watershed Models, Water Sci. Appl. Ser. American Geophysical Union (AGU), pp. 49–68. https://doi.org/10.1029/ws006p0049

Kavetski, D., Kuczera, G., Franks, S.W., 2006. Calibration of conceptual hydrological models revisited: 2. Improving optimisation and analysis, in: Journal of Hydrology. Elsevier, pp. 187–201. https://doi.org/10.1016/j.jhydrol.2005.07.013

Kennedy, J., 2006. Swarm Intelligence, in: Zomaya, A.Y. (Ed.), Handbook of Nature-Inspired and Innovative Computing. Springer, Boston, MA, pp. 187–219. https://doi.org/10.1007/0-387-27705-6_6

Khu, S.T., Werner, M.G.F., 2003. Reduction of Monte-Carlo simulation runs for uncertainty

estimation in hydrological modelling. Hydrol. Earth Syst. Sci. 7, 680–692.

Kim, S.M., Benham, B.L., Brannan, K.M., Zeckoski, R.W., Doherty, J., 2007. Comparison of hydrologic calibration of HSPF using automatic and manual methods. Water Resour. Res. 43, 1402. https://doi.org/10.1029/2006WR004883

Kim, T., Heo, J.H., Bae, D.H., Kim, J.H., 2008. Single-reservoir operating rules for a year using multiobjective genetic algorithm. J. Hydroinformatics 10, 163–179. https://doi.org/10.2166/hydro.2008.019

Kitanidis, P.K., Bras, R.L., 1980a. Real-time forecasting with a conceptual hydrologic model: 1. Analysis of uncertainty. Water Resour. Res. 16, 1025–1033. https://doi.org/10.1029/WR016i006p01025

Kitanidis, P.K., Bras, R.L., 1980b. Adaptive filtering through detection of isolated transient errors in rainfall-runoff models. Water Resour. Res. 16, 740–748. https://doi.org/10.1029/WR016i004p00740

Kollat, J.B., Reed, P., 2007. A framework for Visually Interactive Decision-making and Design using Evolutionary Multi-objective Optimization (VIDEO). Environ. Model. Softw. 22, 1691–1704. https://doi.org/10.1016/j.envsoft.2007.02.001

Kollat, J.B., Reed, P.M., 2006. Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design. Adv. Water Resour. 29, 792–807. https://doi.org/10.1016/j.advwatres.2005.07.010

Kollat, J.B., Reed, P.M., Wagener, T., 2012. When are multiobjective calibration trade-offs in hydrologic models meaningful? Water Resour. Res. 48, 3520.

https://doi.org/10.1029/2011WR011534

Koppa, A., Gebremichael, M., Yeh, W.W.G., 2019. Multivariate calibration of large scale hydrologic models: The necessity and value of a Pareto optimal approach. Adv. Water Resour. 130, 129–146. https://doi.org/10.1016/j.advwatres.2019.06.005

Kougias, I.P., Theodossiou, N.P., 2013. Multiobjective Pump Scheduling Optimization Using Harmony Search Algorithm (HSA) and Polyphonic HSA. Water Resour. Manag. 27, 1249–1261. https://doi.org/10.1007/s11269-012-0236-5

Kouwen, N., 2016. WATFLOOD™/CHARM Canadian Hydrological and Routing Model, User's Manual. University of Waterloo, Waterloo, ON.

Kouwen, N., 2010. WATFLOOD/WATROUTE Hydrological Model Routing & Flow Forecasting SYSTEM. Department of Civil Engineering, University of Waterloo, Waterloo, ON.

Kramer, O., Danielsiek, H., 2010. DBSCAN-based multi-objective niching to approximate equivalent pareto-subsets, in: Proceedings of the 12th Annual Genetic and Evolutionary Computation Conference, GECCO '10. pp. 503–510. https://doi.org/10.1145/1830483.1830575

Kramer, O., Koch, P., 2009. Rake selection: A novel evolutionary multi-objective optimization algorithm, in: Mertsching B., Hund M., A.Z. (Ed.), Annual Conference on Artificial Intelligence; KI 2009: Advances in Artificial Intelligence. Springer Verlag, Berlin Heidelberg, pp. 177–184. https://doi.org/10.1007/978-3-642-04617-9_23

Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. Water Resour. Res. 35, 2739–2750.

https://doi.org/10.1029/1999WR900099

Kumar, A., Singh, R., Jena, P.P., Chatterjee, C., Mishra, A., 2015. Identification of the best multi-model combination for simulating river discharge. J. Hydrol. 525, 313–325. https://doi.org/10.1016/j.jhydrol.2015.03.060

Labadie, J.W., Zheng, F., Wan, Y., 2012. Optimal integrated operation of reservoir-assisted stormwater treatment areas for estuarine habitat restoration. Environ. Model. Softw. 38, 271–282. https://doi.org/10.1016/j.envsoft.2012.07.005

Lamb, R., Beven, K., Myrabø, S., 1998. Use of spatially distributed water table observations to constrain uncertainty in a rainfall-runoff model. Adv. Water Resour. 22, 305–317. https://doi.org/10.1016/S0309-1708(98)00020-7

Lansey, K.E., Awumah, K., 1994. Optimal Pump Operations Considering Pump Switches. J. Water Resour. Plan. Manag. 120, 17–35. https://doi.org/10.1061/(ASCE)0733-9496(1994)120:1(17)

Laumanns, M., Thiele, L., Deb, K., Zitzler, E., 2002. Combining convergence and diversity in evolutionary multiobjective optimization. Evol. Comput. 10, 263–282. https://doi.org/10.1162/106365602760234108

Lautenbach, S., Volk, M., Strauch, M., Whittaker, G., Seppelt, R., 2013. Optimization-based trade-off analysis of biodiesel crop production for managing an agricultural catchment. Environ. Model. Softw. 48, 98–112. https://doi.org/10.1016/j.envsoft.2013.06.006

Lenormand, M., Jabot, F., Deffuant, G., 2013. Adaptive approximate Bayesian computation for complex models. Comput. Stat. 2013 286 28, 2777–2796. https://doi.org/10.1007/S00180-

013-0428-3

Levy, H., 1992. Stochastic Dominance and Expected Utility: Survey and Analysis. Manage. Sci. 38, 555–593. https://doi.org/10.1287/mnsc.38.4.555

Li, J.P., Balazs, M.E., Parks, G.T., Clarkson, P.J., 2002. A species conserving genetic algorithm for multimodal function optimization. Evol. Comput. 10, 207–234. https://doi.org/10.1162/106365602760234081

Li, W., Sankarasubramanian, A., 2012. Reducing hydrologic model uncertainty in monthly streamflow predictions using multimodel combination. Water Resour. Res. 48. https://doi.org/10.1029/2011WR011380

Li, X., Epitropakis, M.G., Deb, K., Engelbrecht, A., 2017. Seeking Multiple Solutions: An Updated Survey on Niching Methods and Their Applications. IEEE Trans. Evol. Comput. https://doi.org/10.1109/TEVC.2016.2638437

Liebman, J.C., 1976. Some Simple-Minded Observations on the Role of Optimization in Public Systems Decision-Making. Interfaces (Providence). 6, 102–108. https://doi.org/10.1287/inte.6.4.102

Lindström, G., Johansson, B., Persson, M., Gardelin, M., Bergström, S., 1997. Development and test of the distributed HBV-96 hydrological model. J. Hydrol. 201, 272–288. https://doi.org/10.1016/S0022-1694(97)00041-3

Liu, P., Cai, X., Guo, S., 2011. Deriving multiple near-optimal solutions to deterministic reservoir operation problems. Water Resour. Res. 47. https://doi.org/10.1029/2011WR010998

Liu, Y., Yen, G.G., Gong, D., 2019. A Multimodal Multiobjective Evolutionary Algorithm Using

Two-Archive and Recombination Strategies. IEEE Trans. Evol. Comput. 23, 660–674. https://doi.org/10.1109/TEVC.2018.2879406

Macqueen, J., 1967. SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS, in: Lecam, L.., Neyman, J.. (Eds.), Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1, No. 14). pp. 281–297.

Macro, K., Matott, L.S., Rabideau, A., Ghodsi, S.H., Zhu, Z., 2019. OSTRICH-SWMM: A new multi-objective optimization tool for green infrastructure planning with SWMM. Environ. Model. Softw. 113, 42–47. https://doi.org/10.1016/j.envsoft.2018.12.004

Madadgar, S., Moradkhani, H., Garen, D., 2014. Towards improved post-processing of hydrologic forecast ensembles. Hydrol. Process. 28, 104–122. https://doi.org/10.1002/hyp.9562

Madsen, H., 2003. Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. Adv. Water Resour. 26, 205–216. https://doi.org/10.1016/S0309-1708(02)00092-1

Maier, H.R., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L.S., Cunha, M.C., Dandy, G.C., Gibbs, M.S., Keedwell, E., Marchi, A., Ostfeld, A., Savic, D., Solomatine, D.P., Vrugt, J.A., Zecchin, A.C., Minsker, B.S., Barbour, E.J., Kuczera, G., Pasha, F., Castelletti, A., Giuliani, M., Reed, P.M., 2014. Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions. Environ. Model. Softw. 62, 271–299. https://doi.org/10.1016/j.envsoft.2014.09.013

Maier, H.R., Razavi, S., Kapelan, Z., Matott, L.S., Kasprzyk, J., Tolson, B.A., 2019. Introductory

overview: Optimization using evolutionary algorithms and other metaheuristics. Environ. Model. Softw. https://doi.org/10.1016/j.envsoft.2018.11.018

Manitoba Infrastructure and Transportation, 2016. Assiniboine River & Lake Manitoba basins flood mitigation study. Winnipeg, MB.

Manitoba Infrastructure and Transportation, 2013. Manitoba 2011 Flood Review Task Force Report. Winnipeg, MB.

Marjoram, P., Molitor, J., Plagnol, V., Tavaré, S., 2003. Markov chain Monte Carlo without likelihoods. Proc. Natl. Acad. Sci. 100, 15324–15328. https://doi.org/10.1073/PNAS.0306899100

Marques, J., Cunha, M., Savić, D.A., 2015. Multi-objective optimization of water distribution systems based on a real options approach. Environ. Model. Softw. 63, 1–13. https://doi.org/10.1016/j.envsoft.2014.09.014

Maskey, S., Guinot, V., Price, R.K., 2004. Treatment of precipitation uncertainty in rainfall-runoff modelling: a fuzzy set approach. Adv. Water Resour. 27, 889–898. https://doi.org/10.1016/J.ADVWATRES.2004.07.001

Matott, L.S., Babendreier, J.E., Purucker, S.T., 2009. Evaluating uncertainty in integrated environmental models: A review of concepts and tools. Water Resour. Res. 45, W06421(1-14). https://doi.org/10.1029/2008WR007301

Matott, L.S., Tolson, B.A., Asadzadeh, M., 2012. A benchmarking framework for simulation-based optimization of environmental models. Environ. Model. Softw. 35, 19–30. https://doi.org/10.1016/j.envsoft.2012.02.002

McCormick, G., Powell, R.S., 2003. Optimal Pump Scheduling in Water Supply Systems with Maximum Demand Charges. J. Water Resour. Plan. Manag. 129, 372–379. https://doi.org/10.1061/(ASCE)0733-9496(2003)129:5(372)

Mcginn, S.M., 2010. Weather and Climate Patterns in Canada's Prairie Grasslands. Arthropods Can. Grasslands 1, 105–119. https://doi.org/10.3752/9780968932148.ch5

McIntyre, N., Lee, H., Wheater, H., Young, A., Wagener, T., 2005. Ensemble predictions of runoff in ungauged catchments. Water Resour. Res. 41, 1–14. https://doi.org/10.1029/2005WR004289

McMichael, C.E., Hope, A.S., Loaiciga, H.A., 2006. Distributed hydrological modelling in California semi-arid shrublands: MIKE SHE model calibration and uncertainty estimation. J. Hydrol. 317, 307–324. https://doi.org/10.1016/j.jhydrol.2005.05.023

McMillan, H., Clark, M., 2009. Rainfall-runoff model calibration using informal likelihood measures within a Markov chain Monte Carlo sampling scheme. Water Resour. Res. 45. https://doi.org/10.1029/2008WR007288

McMillan, H.K., 2021. A review of hydrologic signatures and their applications. WIREs Water 8, e1499. https://doi.org/10.1002/wat2.1499

Mitchell, S., Beven, K., Freer, J., 2009. Multiple sources of predictive uncertainty in modeled estimates of net ecosystem CO2 exchange. Ecol. Modell. 220, 3259–3270. https://doi.org/10.1016/j.ecolmodel.2009.08.021

Montanari, A., 2005. Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations. Water Resour. Res. 41,

1–13. https://doi.org/10.1029/2004WR003826

Moore, R.D., 1993. Application of a conceptual streamflow model in a glacierized drainage basin. J. Hydrol. 150, 151–168. https://doi.org/10.1016/0022-1694(93)90159-7

Moradkhani, H., Hsu, K.L., Gupta, H., Sorooshian, S., 2005. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. Water Resour. Res. 41, 1–17. https://doi.org/10.1029/2004WR003604

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Trans. ASABE 50, 885–900. https://doi.org/10.13031/2013.23153

Muhammad, A., Stadnyk, T., Unduche, F., Coulibaly, P., 2018. Multi-Model Approaches for Improving Seasonal Ensemble Streamflow Prediction Scheme with Various Statistical Post-Processing Techniques in the Canadian Prairie Region. Water 10, 1604. https://doi.org/10.3390/w10111604

Muleta, M.K., Nicklow, J.W., 2005. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. J. Hydrol. 306, 127–145. https://doi.org/10.1016/j.jhydrol.2004.09.005

Murray, D.M., Yakowitz, S.J., 1979. Constrained differential dynamic programming and its application to multireservoir control. Water Resour. Res. 15, 1017–1027. https://doi.org/10.1029/WR015i005p01017

Nash, J.E., Sutcliffe, J. V., 1970. River flow forecasting through conceptual models part I - A discussion of principles. J. Hydrol. 10, 282–290. https://doi.org/10.1016/0022-

1694(70)90255-6

Nasseri, M., Zahraie, B., Ansari, A., Solomatine, D.P., 2013. Uncertainty assessment of monthly water balance models based on Incremental Modified Fuzzy Extension Principle method. J. Hydroinformatics 15, 1340–1360. https://doi.org/10.2166/HYDRO.2013.159

Nearing, G.S., Tian, Y., Gupta, H. V, Clark, M.P., Harrison, K.W., Weijs, S. V, 2016. A philosophical basis for hydrological uncertainty. Hydrol. Sci. J. 61, 1666–1678. https://doi.org/10.1080/02626667.2016.1183009

Neitsch, S.., Arnold, J.., Kiniry, J.., Williams, J.., 2011. Soil & Water Assessment Tool Theoretical Documentation Version 2009. College Station, Texas. https://doi.org/10.1016/j.scitotenv.2015.11.063

Nott, D.J., Marshall, L., Brown, J., 2012. Generalized likelihood uncertainty estimation (GLUE) and approximate Bayesian computation: What's the connection? Water Resour. Res. 48, 12602. https://doi.org/10.1029/2011WR011128

Pajares, A., Blasco, X., Herrero, J.M., Reynoso-Meza, G., 2018. A Multiobjective Genetic Algorithm for the Localization of Optimal and Nearly Optimal Solutions Which Are Potentially Useful: nevMOGA. Complexity. https://doi.org/10.1155/2018/1792420

Pang, B., Shi, S., Zhao, G., Shi, R., Peng, D., Zhu, Z., 2020. Uncertainty assessment of urban hydrological modelling from a multiple objective perspective. Water (Switzerland) 12. https://doi.org/10.3390/W12051393

Pang, B., Yue, J., Huang, Z., Zhang, R., 2019. Parameter uncertainty assessment of a flood forecasting model using multiple objectives. J. Flood Risk Manag. 12, 1–12.

https://doi.org/10.1111/jfr3.12493

Parrish, M.A., Moradkhani, H., DeChant, C.M., 2012. Toward reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation. Water Resour. Res. 48. https://doi.org/10.1029/2011WR011116

Penning-Rowsell, E.C.., Tunstall, S.M.., Tapsell, S.M.., Parker, D.J.., 2000. The benefits of flood warnings: real but elusive, and politically significant. Water Environ. J. 14, 7–14.

Petrowski, A., 1996. Clearing procedure as a niching method for genetic algorithms, in: Proceedings of the IEEE Conference on Evolutionary Computation. IEEE, pp. 798–803. https://doi.org/10.1109/icec.1996.542703

Pfannerstill, M., Bieger, K., Guse, B., Bosch, D.D., Fohrer, N., Arnold, J.G., 2017. How to Constrain Multi-Objective Calibrations of the SWAT Model Using Water Balance Components. JAWRA J. Am. Water Resour. Assoc. 53, 532–546. https://doi.org/10.1111/1752-1688.12524@10.1111/(ISSN)17521688/SWAT

Pfannerstill, M., Guse, B., Fohrer, N., 2014. Smart low flow signature metrics for an improved overall performance evaluation of hydrological models. J. Hydrol. 510, 447–458. https://doi.org/10.1016/j.jhydrol.2013.12.044

Piani, C., Haerter, J.O., Coppola, E., 2010. Statistical bias correction for daily precipitation in regional climate models over Europe. Theor. Appl. Climatol. 99, 187–192. https://doi.org/10.1007/s00704-009-0134-9

Piscopo, A.N., Kasprzyk, J.R., Neupauer, R.M., 2015. An iterative approach to multi-objective engineering design: Optimization of engineered injection and extraction for enhanced

groundwater remediation. Environ. Model. Softw. 69, 253–261. https://doi.org/10.1016/j.envsoft.2014.08.030

Pokhrel, P., Gupta, H. V., Wagener, T., 2008. A spatial regularization approach to parameter estimation for a distributed watershed model. Water Resour. Res. 44, 12419. https://doi.org/10.1029/2007WR006615

Pokhrel, P., Yilmaz, K.K., Gupta, H. V., 2012. Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures. J. Hydrol. 418–419, 49–60. https://doi.org/10.1016/j.jhydrol.2008.12.004

Pomeroy, J.W., De Boer, D, Martz, L.W., Pomeroy, J., De Boer, Dirk, Martz, L., 2005. Hydrology and Water Resources of Saskatchewan. Saskatoon, SK.

Pomeroy, J.W., Stewart, R.E., Whitfield, P.H., 2016. The 2013 flood event in the South Saskatchewan and Elk River basins: Causes, assessment and damages. Can. Water Resour. J. 41, 105–117. https://doi.org/10.1080/07011784.2015.1089190

Prasad, T.D., Park, N.-S., 2004. Multiobjective Genetic Algorithms for Design of Water Distribution Networks. J. Water Resour. Plan. Manag. 130, 73–82. https://doi.org/10.1061/(asce)0733-9496(2004)130:1(73)

Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A., Feldman, M.W., 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol. Biol. Evol. 16, 1791–1798. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A026091

Purshouse, R.C., Fleming, P.J., 2007. On the evolutionary optimization of many conflicting objectives. IEEE Trans. Evol. Comput. 11, 770–784.

https://doi.org/10.1109/TEVC.2007.910138

Quinn, J.D., Reed, P.M., Keller, K., 2017. Direct policy search for robust multi-objective management of deeply uncertain socio-ecological tipping points. Environ. Model. Softw. 92, 125–141. https://doi.org/10.1016/j.envsoft.2017.02.017

Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. Mon. Weather Rev. 133, 1155–1174. https://doi.org/10.1175/MWR2906.1

Ragab, R., Kaelin, A., Afzal, M., Panagea, I., 2020. Application of Generalized Likelihood Uncertainty Estimation (GLUE) at different temporal scales to reduce the uncertainty level in modelled river flows. Hydrol. Sci. J. 65, 1856–1871. https://doi.org/10.1080/02626667.2020.1764961

Rajib, M.A., Merwade, V., Yu, Z., 2016. Multi-objective calibration of a hydrologic model using spatially distributed remotely sensed/in-situ soil moisture. J. Hydrol. 536, 192–207. https://doi.org/10.1016/j.jhydrol.2016.02.037

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A.P., Black, J., Thériault, J.M., Kucera, P., Gochis, D., Smith, C., Nitu, R., Hall, M., Ikeda, K., Gutmann, E., 2012. How Well Are We Measuring Snow: The NOAA/FAA/NCAR Winter Precipitation Test Bed. Bull. Am. Meteorol. Soc. 93, 811–829. https://doi.org/10.1175/BAMS-D-11-00052.1

Ratto, M., Tarantola, S., Saltelli, A., 2001. Sensitivity analysis in model calibration: GSA-GLUE approach. Comput. Phys. Commun. 136, 212–224. https://doi.org/10.1016/S0010-

4655(01)00159-X

Razavi, S., Asadzadeh, M., Tolson, B., Fay, D., Moin, S., Bruxer, J., Fan, Y., 2014. Evaluation of New Control Structures for Regulating the Great Lakes System: Multiscenario, Multireservoir Optimization Approach. J. Water Resour. Plan. Manag. 140, 04014018. https://doi.org/10.1061/(asce)wr.1943-5452.0000375

Razavi, S., Sheikholeslami, R., Gupta, H. V., Haghnegahdar, A., 2019. VARS-TOOL: A toolbox for comprehensive, efficient, and robust sensitivity and uncertainty analysis. Environ. Model. Softw. 112, 95–107. https://doi.org/10.1016/j.envsoft.2018.10.005

Razavi, T., Coulibaly, P., 2016. Improving streamflow estimation in ungauged basins using a multi-modelling approach. Hydrol. Sci. J. 61, 2668–2679. https://doi.org/10.1080/02626667.2016.1154558

Reed, P., Devireddy, V., 2004. GROUNDWATER MONITORING DESIGN: A CASE STUDY COMBINING EPSILON DOMINANCE ARCHIVING AND AUTOMATIC PARAMETERIZATION FOR THE NSGA-II, in: Coello-Coello, C. (Ed.), Applications of Multi-Objective Evolutionary Algorithms. Advances in Natural Computation Series. World Scientific, New York, pp. 79–100. https://doi.org/10.1142/9789812567796_0004

Reed, P., Kollat, J.B., Devireddy, V.K., 2007. Using interactive archives in evolutionary multiobjective optimization: A case study for long-term groundwater monitoring design. Environ. Model. Softw. 22, 683–692. https://doi.org/10.1016/j.envsoft.2005.12.021

Reed, P.M., Hadka, D., Herman, J.D., Kasprzyk, J.R., Kollat, J.B., 2013. Evolutionary Multiobjective Optimization in Water Resources: The Past, Present, and Future. Adv. Water

Resour. 51, 438–456.

Reed, P.M., Kasprzyk, J., 2009. Water Resources Management: The Myth, the Wicked, and the Future. J. Water Resour. Plan. Manag. 135, 411–413. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000047

Rittel, H.W.J., Webber, M.M., 1973. Dilemmas in a general theory of planning. Policy Sci. 4, 155–169. https://doi.org/10.1007/BF01405730

Rogers, P.P., Fiering, M.B., 1986. Use of systems analysis in water management. Water Resour. Res. 22, 146S-158S. https://doi.org/10.1029/WR022i09Sp0146S

Rosenberg, D.E., 2015. Blended near-optimal alternative generation, visualization, and interaction for water resources decision making. Water Resour. Res. 51, 2047–2063. https://doi.org/10.1002/2013WR014667

Rudolph, G., Naujoks, B., Preuss, M., 2007. Capabilities of EMOA to detect and preserve equivalent pareto subsets, in: International Conference on Evolutionary Multi-Criterion Optimization. Springer Verlag, Berlin, Heidelberg, pp. 36–50. https://doi.org/10.1007/978-3-540-70928-2_7

Sadatiyan A., S., Miller, C., 2017. PEPSO: Reducing Electricity Usage and Associated Pollution Emissions of Water Pumps. Water 9, 640. https://doi.org/10.3390/w9090640

Sahinidis, N. V, 2004. Optimization under uncertainty: state-of-the-art and opportunities. Comput. Chem. Eng. 28, 971–983. https://doi.org/10.1016/j.compchemeng.2003.09.017

Sahraei, S., Asadzadeh, M., 2021. Cluster-based multi-objective optimization for identifying diverse design options: Application to water resources problems. Environ. Model. Softw. 135,

104902. https://doi.org/10.1016/j.envsoft.2020.104902

Sahraei, S., Asadzadeh, M., Shafii, M., 2019. Toward effective many-objective optimization: Rounded-archiving. Environ. Model. Softw. 122. https://doi.org/10.1016/j.envsoft.2019.104535

Sahraei, S., Asadzadeh, M., Unduche, F., 2020. Signature-based multi-modelling and multi-objective calibration of hydrologic models: Application in flood forecasting for Canadian Prairies. J. Hydrol. 588, 125095. https://doi.org/10.1016/j.jhydrol.2020.125095

Sakarya, A.B.A., Mays, L.W., 2000. Optimal Operation of Water Distribution Pumps Considering Water Quality. J. Water Resour. Plan. Manag. 126, 210–220. https://doi.org/10.1061/(ASCE)0733-9496(2000)126:4(210)

Sareni, B., Krähenbühl, L., 1998. Fitness sharing and niching methods revisited. IEEE Trans. Evol. Comput. 2, 97–106. https://doi.org/10.1109/4235.735432ï

Savic, D.A., Walters, G.A., 1997. Genetic Algorithms for Least-Cost Design of Water Distribution Networks. J. Water Resour. Plan. Manag. 123, 67–77. https://doi.org/10.1061/(ASCE)0733-9496(1997)123:2(67)

Schaefli, B., 2016. Snow hydrology signatures for model identification within a limits-of-acceptability approach. Hydrol. Process. 30, 4019–4035. https://doi.org/10.1002/hyp.10972

Schaffer, J.D., 1985. Multiple objective optimization with vector evaluated genetic algorithms, in: In Proceedings of the First International Conference on Genetic Algorithms and Their Applications. Lawrence Erlbaum Associates. Inc., Publishers.

Schoups, G., Vrugt, J.A., 2010. A formal likelihood function for parameter and predictive

inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Resour. Res. 46. https://doi.org/10.1029/2009WR008933

Schütze, O., Vasile, M., Coello Coello, C.A., 2011. Computing the set of epsilon-efficient solutions in multiobjective space mission design. J. Aerosp. Comput. Inf. Commun. 8, 53–70. https://doi.org/10.2514/1.46478

Seibert, J., Beven, K.J., 2009. Gauging the ungauged basin: how many discharge measurements are needed? Hydrol. Earth Syst. Sci. 13, 883–892. https://doi.org/10.5194/hess-13-883-2009

Shafii, M., Basu, N., Craig, J.R., Schiff, S.L., Cappellen, P. Van, 2017. A diagnostic approach to constraining flow partitioning in hydrologic models using a multiobjective optimization framework. Water Resour. Res. 53, 3279–3301. https://doi.org/10.1002/2016WR019736.Received

Shafii, M., De Smedt, F., 2009. Multi-objective calibration of a distributed hydrological model (WetSpa) using a genetic algorithm. Hydrol. Earth Syst. Sci. 13, 2137–2149. https://doi.org/10.5194/hess-13-2137-2009

Shafii, M., Tolson, B., Shawn Matott, L., 2015. Addressing subjective decision-making inherent in GLUE-based multi-criteria rainfall-runoff model calibration. J. Hydrol. 523, 693–705. https://doi.org/10.1016/j.jhydrol.2015.01.051

Shafii, M., Tolson, B.A., 2015. Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. Water Resour. Res. 51, 3796–3814. https://doi.org/10.1002/2014WR016520

Shamseldin, A.Y., O'Connor, K.M., Liang, G.C., 1997. Methods for combining the outputs of

different rainfall-runoff models. J. Hydrol. 197, 203–229. https://doi.org/10.1016/S0022-1694(96)03259-3

Shepherd, A., McGinn, S.M., 2003. Assessment of climate change on the Canadian Prairies from downscaled GCM data. Atmos. - Ocean 41, 301–316. https://doi.org/10.3137/ao.410404

Shir, O.M., Preuss, M., Naujoks, B., Emmerich, M., 2010. Enhancing decision space diversity in evolutionary multiobjective algorithms, in: International Conference on Evolutionary Multi-Criterion Optimization. Springer, Berlin, Heidelberg, pp. 95–109. https://doi.org/10.1007/978-3-642-01020-0_12

Sikorska, A.E., Renard, B., 2017. Calibrating a hydrological model in stage space to account for rating curve uncertainties: general framework and key challenges. Adv. Water Resour. 105, 51–66. https://doi.org/10.1016/j.advwatres.2017.04.011

Simpson, A.R., Dandy, G.C., Murphy, L.J., 1994. Genetic Algorithms Compared to Other Techniques for Pipe Optimization. J. Water Resour. Plan. Manag. 120, 423–443. https://doi.org/10.1061/(asce)0733-9496(1994)120:4(423)

Singh, R., Reed, P.M., Keller, K., 2015. Many-objective robust decision making for managing an ecosystem with a deeply uncertain threshold response. Ecol. Soc. 20, 12. https://doi.org/10.5751/ES-07687-200312

Sorooshian, S., Duan, Q., Gupta, V.K., 1993. Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model. Water Resour. Res. 29, 1185–1194. https://doi.org/10.1029/92WR02617

Srinivas, N., Deb, K., 1994. Muiltiobjective Optimization Using Nondominated Sorting in Genetic

Algorithms. Evol. Comput. 2, 221–248. https://doi.org/10.1162/evco.1994.2.3.221

Stadnyk, T.A., Delavau, C., Kouwen, N., Edwards, T.W.D., 2013. Towards hydrological model calibration and validation: simulation of stable water isotopes using the isoWATFLOOD model. Hydrol. Process. 27, 3791–3810. https://doi.org/10.1002/hyp.9695

Stedinger, J.R., Vogel, R.M., Lee, S.U., Batchelder, R., 2008. Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. Water Resour. Res. 44, W00B06 (1-17). https://doi.org/10.1029/2008wr006822

Storn, R., Price, K., 1997. Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. J. Glob. Optim. 11, 341–359. https://doi.org/10.1023/A:1008202821328

Sülflow, A., Drechsler, N., Drechsler, R., 2007. Robust multi-objective optimization in high dimensional spaces, in: Obayashi, S., Deb, K., Poloni, C., Hiroyasu, T., Murata, T. (Eds.), Evolutionary Multi-Criterion Optimization. EMO 2007. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 715–726. https://doi.org/10.1007/978-3-540-70928-2_54

Sun, J., Miao, Z., Gong, D., Zeng, X.J., Li, J., Wang, G., 2020. Interval Multiobjective Optimization with Memetic Algorithms. IEEE Trans. Cybern. 50, 3444–3457. https://doi.org/10.1109/TCYB.2019.2908485

Sun, M., Zhang, X., Huo, Z., Feng, S., Huang, G., Mao, X., 2016. Uncertainty and sensitivity assessments of an agricultural-hydrological model (RZWQM2) using the GLUE method. J. Hydrol. 534, 19–30. https://doi.org/10.1016/j.jhydrol.2015.12.045

Tanabe, R., Ishibuchi, H., 2019. A Review of Evolutionary Multi-modal Multi-objective

Optimization. IEEE Trans. Evol. Comput. https://doi.org/10.1109/TEVC.2019.2909744

Tang, Y., Reed, P., Wagener, T., 2006. How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration? Hydrol. Earth Syst. Sci. 10, 289–307. https://doi.org/10.5194/hess-10-289-2006

Te Linde, A.H., Aerts, J.C.J.H., Hurkmans, R.T.W.L., Eberle, M., 2008. Comparing model performance of two rainfall-runoff models in the Rhine basin using different atmospheric forcing data sets. Hydrol. Earth Syst. Sci. 12, 943–957. https://doi.org/10.5194/hess-12-943-2008

Tegegne, G., Kim, Y.O., Seo, S.B., Kim, Y., 2019. Hydrological modelling uncertainty analysis for different flow quantiles: a case study in two hydro-geographically different watersheds. Hydrol. Sci. J. 64, 473–489. https://doi.org/10.1080/02626667.2019.1587562

Teweldebrhan, A.T., Burkhart, J.F., Schuler, T. V., 2018. Parameter uncertainty analysis for an operational hydrological model using residual-based and limits of acceptability approaches. Hydrol. Earth Syst. Sci. 22, 5021–5039. https://doi.org/10.5194/he-22-5021-2018

Teytaud, O., 2007. On the hardness of offline multi-objective optimization. Evol. Comput. 15, 475–491. https://doi.org/10.1162/evco.2007.15.4.475

Thiemann, M., Trosset, M., Gupta, H., Sorooshian, S., 2001. Bayesian recursive parameter estimation for hydrologic models. Water Resour. Res. 37, 2521–2535. https://doi.org/10.1029/2000WR900405

Tikhonov, A.N., Arsenin, V.Y., 1977. Solutions of ill posed problems. Winston and Sons, New York.

Toffolo, A., Benini, E., 2003. Genetic diversity as an objective in multi-objective evolutionary algorithms. Evol. Comput. 11, 151–167. https://doi.org/10.1162/106365603766646816

Tolson, B.A., Maier, H.R., Simpson, A.R., Lence, B.J., 2004. Genetic Algorithms for Reliability-Based Optimization of Water Distribution Systems. J. Water Resour. Plan. Manag. 130, 63–72. https://doi.org/10.1061/(asce)0733-9496(2004)130:1(63)

Tolson, B.A., Shoemaker, C.A., 2008. Efficient prediction uncertainty approximation in the calibration of environmental simulation models. Water Resour. Res. 44, 1–19. https://doi.org/10.1029/2007WR005869

Tolson, B.A., Shoemaker, C.A., 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resour. Res. 43, 1413. https://doi.org/10.1029/2005WR004723

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., Stumpf, M.P.., 2008. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J. R. Soc. Interface 6, 187–202. https://doi.org/10.1098/RSIF.2008.0172

Trawiński, K., Chica, M., Pancho, D.P., Damas, S., Córdon, O., 2018. MoGrams: A Network-Based Methodology for Visualizing the Set of Nondominated Solutions in Multiobjective Optimization. IEEE Trans. Cybern. 48, 474–485. https://doi.org/10.1109/TCYB.2016.2642886

Tsoukalas, I., Makropoulos, C., 2015. Multiobjective optimisation on a budget: Exploring surrogate modelling for robust multi-reservoir rules generation under hydrological uncertainty. Environ. Model. Softw. 69, 396–413.

https://doi.org/10.1016/j.envsoft.2014.09.023

Turner, B.M., Van Zandt, T., 2012. A tutorial on approximate Bayesian computation. J. Math. Psychol. 56, 69–85. https://doi.org/10.1016/J.JMP.2012.02.005

Uhlenbrook, S., Sieber, A., 2005. On the value of experimental data to reduce the prediction uncertainty of a process-oriented catchment model. Environ. Model. Softw. 20, 19–32. https://doi.org/10.1016/j.envsoft.2003.12.006

Ulrich, T., Bader, J., Zitzler, E., 2010. Integrating decision space diversity into hypervolume-based multiobjective search, in: Proceedings of the 12th Annual Genetic and Evolutionary Computation Conference, GECCO '10. pp. 455–462. https://doi.org/10.1145/1830483.1830569

Unduche, F., Tolossa, H., Senbeta, D., Zhu, E., 2018. Evaluation of four hydrological models for operational flood forecasting in a Canadian Prairie watershed. Hydrol. Sci. J. 63, 1133–1149. https://doi.org/10.1080/02626667.2018.1474219

Upper Assiniboine River Basin Study, 2000. Study Report Prepared at the Conclusion of the Upper Assiniboine River Basin Study Agreement Signed in October 1996 by the Governments of Saskatchewan. Manitoba, CA.

USACE, 2017. Hydrologic Modeling System (HEC-HMS) Applications Guide: Version 4.2.1. Institute for Water Resources, Hydrologic Engineering Cente, Davis, California.

Uusitalo, L., Lehikoinen, A., Helle, I., Myrberg, K., 2015. An overview of methods to evaluate uncertainty of deterministic models in decision support. Environ. Model. Softw. 63, 24–31. https://doi.org/10.1016/j.envsoft.2014.09.017

van Griensven, A., Meixner, T., 2006. Methods to quantify and identify the sources of uncertainty for river basin water quality models. Water Sci. Technol. 53, 51–59. https://doi.org/10.2166/wst.2006.007

Van Veldhuizen, D.A., Lamont, G.B., 2000. Multiobjective evolutionary algorithms: analyzing the state-of-the-art. Evol. Comput. https://doi.org/10.1162/106365600568158

Van Veldhuizen, D.A., Lamont, G.B., 1998. Evolutionary Computation and Convergence to a Pareto Front, in: Late Breaking Papers at the Genetic Programming 1998 Conference. pp. 221–228.

Vavak, F., Fogarty, T.C., 1996. Comparison of steady state and generational genetic algorithms for use in nonstationary environments, in: Proceedings of the IEEE Conference on Evolutionary Computation. pp. 192–195. https://doi.org/10.1109/icec.1996.542359

Velázquez, J.A., Anctil, F., Perrin, C., 2010. Hydrology and Earth System Sciences Performance and reliability of multimodel hydrological ensemble simulations based on seventeen lumped models and a thousand catchments. Hydrol. Earth Syst. Sci 14, 2303–2317. https://doi.org/10.5194/hess-14-2303-2010

Vetter, T., Huang, S., Aich, V., Yang, T., Wang, X., Krysanova, V., Hattermann, F., 2015. Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents. Earth Syst. Dyn. 6, 17–43. https://doi.org/10.5194/esd-6-17-2015

Voll, P., Jennings, M., Hennen, M., Shah, N., Bardow, A., 2015. The optimum is not enough: A near-optimal solution paradigm for energy systems synthesis. Energy 82, 446–456. https://doi.org/10.1016/j.energy.2015.01.055

Vrugt, J.A., Robinson, B.A., 2007. Improved evolutionary optimization from genetically adaptive multimethod search. Proc. Natl. Acad. Sci. U. S. A. 104, 708–711. https://doi.org/10.1073/pnas.0610471104

Vrugt, J.A., Sadegh, M., 2013. Toward diagnostic model calibration and evaluation: Approximate Bayesian computation. Water Resour. Res. 49, 4335–4345. https://doi.org/10.1002/WRCR.20354@10.1002/(ISSN)1944-7973.ACMIH1

Vrugt, J.A., ter Braak, C.J.F., Gupta, H. V., Robinson, B.A., 2009. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? Stoch. Environ. Res. Risk Assess. 23, 1011–1026. https://doi.org/10.1007/s00477-008-0274-y

Wagener, T., Gupta, H. V., 2005. Model identification for hydrological forecasting under uncertainty. Stoch. Environ. Res. Risk Assess. 19, 378–387. https://doi.org/10.1007/s00477-005-0006-5

Wan, W., Guo, X., Lei, X., Jiang, Y., Wang, H., 2018. A Novel Optimization Method for Multi-Reservoir Operation Policy Derivation in Complex Inter-Basin Water Transfer System. Water Resour. Manag. 32, 31–51. https://doi.org/10.1007/s11269-017-1735-1

Wang, Q.J., 1991. The Genetic Algorithm and Its Application to Calibrating Conceptual Rainfall-Runoff Models. Water Resour. Res. 27, 2467–2471. https://doi.org/10.1029/91WR01305

Wang, S., Huang, G.H., Baetz, B.W., Huang, W., 2016. Probabilistic Inference Coupled with Possibilistic Reasoning for Robust Estimation of Hydrologic Parameters and Piecewise Characterization of Interactive Uncertainties. J. Hydrometeorol. 17, 1243–1260. https://doi.org/10.1175/JHM-D-15-0131.1

Wang, Y., Brubaker, K., 2015. Multi-objective model auto-calibration and reduced parameterization: Exploiting gradient-based optimization tool for ahydrologic model. Environ. Model. Softw. 70, 1–15. https://doi.org/10.1016/j.envsoft.2015.04.001

Ward, V.L., Singh, R., Reed, P.M., Keller, K., 2015. Confronting tipping points: Can multi-objective evolutionary algorithms discover pollution control tradeoffs given environmental thresholds? Environ. Model. Softw. 73, 27–43. https://doi.org/10.1016/j.envsoft.2015.07.020

Wardlaw, R., Sharif, M., 1999. Evaluation of Genetic Algorithms for Optimal Reservoir System Operation. J. Water Resour. Plan. Manag. 125, 25–33. https://doi.org/10.1061/(ASCE)0733-9496(1999)125:1(25)

Westerberg, I.K., Guerrero, J., Younger, P.M., Beven, K.J., Seibert, J., Halldin, S., Freer, J.E., Xu, C., 2011. Calibration of hydrological models using flow-duration curves. Hydrol. Earth Syst. Sci. 15, 2205–2227. https://doi.org/10.5194/hess-15-2205-2011

Wheaton, 1998. But it's a dry cold: weathering the Canadian Prairies. Fifth House Publishers, Calgary, AB.

Wood, A.W., Schaake, J.C., 2008. Correcting errors in streamflow forecast ensemble mean and spread. J. Hydrometeorol. 9, 132–148. https://doi.org/10.1175/2007JHM862.1

Wu, Z.Y., Simpson, A.R., 2002. A self-adaptive boundary search genetic algorithm and its application to water distribution systems. J. Hydraul. Res. 40, 191–203. https://doi.org/10.1080/00221680209499862

Xie, H., Lian, Y., 2013. Uncertainty-based evaluation and comparison of SWAT and HSPF applications to the Illinois River Basin. J. Hydrol. 481, 119–131.

https://doi.org/10.1016/j.jhydrol.2012.12.027

Xie, H., Shen, Z., Chen, L., Lai, X., Qiu, J., Wei, G., Dong, J., Peng, Y., Chen, X., 2019. Parameter estimation and uncertainty analysis: A comparison between continuous and event-based modeling of streamflow based on the Hydrological Simulation Program-Fortran (HSPF) model. Water (Switzerland) 11, 171. https://doi.org/10.3390/w11010171

Xiong, L., Shamseldin, A.Y., O'Connor, K.M., 2001. A non-linear combination of the forecasts of rainfall-runoff models by the first-order Takagi-Sugeno fuzzy system. J. Hydrol. 245, 196–217. https://doi.org/10.1016/S0022-1694(01)00349-3

Xu, R., Wunsch, D., 2008. Clustering. John Wiley & Sons.

Xue, L., Yang, F., Yang, C., Wei, G., Li, W., He, X., 2018. Hydrological simulation and uncertainty analysis using the improved TOPMODEL in the arid Manas River basin, China. Sci. Rep. 8, 1–12. https://doi.org/10.1038/s41598-017-18982-8

Yadav, M., Wagener, T., Gupta, H., 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. Adv. Water Resour. 30, 1756–1774. https://doi.org/10.1016/j.advwatres.2007.01.005

Yang, J., Reichert, P., Abbaspour, K.C., 2007. Bayesian uncertainty analysis in distributed hydrologic modeling: A case study in the Thur River basin (Switzerland). Water Resour. Res. 43, W10401 (1-13). https://doi.org/10.1029/2006WR005497

Yang, J., Reichert, P., Abbaspour, K.C., Xia, J., Yang, H., 2008. Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. J. Hydrol. 358, 1–23. https://doi.org/10.1016/j.jhydrol.2008.05.012

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. J. Hydrol. 204, 83–97. https://doi.org/10.1016/S0022-1694(97)00107-8

Yazdi, J., 2016. Decomposition based Multi Objective Evolutionary Algorithms for Design of Large-Scale Water Distribution Networks. Water Resour. Manag. 30, 2749–2766. https://doi.org/10.1007/s11269-016-1320-z

Yilmaz, K.K., Gupta, H. V., Wagener, T., 2008. A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. Water Resour. Res. 44. https://doi.org/10.1029/2007WR006716

Yu, G., Powell, R.S., Sterling, M.J.H., 1994. Optimized pump scheduling in water distribution systems. J. Optim. Theory Appl. 83, 463–488. https://doi.org/10.1007/BF02207638

Yuan, Y., Xu, H., Wang, B., Zhang, B., Yao, X., 2016. Balancing Convergence and Diversity in Decomposition-Based Many-Objective Optimizers. IEEE Trans. Evol. Comput. 20, 180–198. https://doi.org/10.1109/TEVC.2015.2443001

Zadeh, L.A., 1978. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst. 1, 3–28. https://doi.org/10.1016/0165-0114(78)90029-5

Zadorojniy, A., Masin, M., Greenberg, L., Shir, O.M., Zeidner, L., 2012. Algorithms for finding maximum diversity of design variables in multi-objective optimization, in: Procedia Computer Science. Elsevier B.V., pp. 171–176. https://doi.org/10.1016/j.procs.2012.01.035

Zatarain Salazar, J., Reed, P.M., Quinn, J.D., Giuliani, M., Castelletti, A., 2017. Balancing exploration, uncertainty and computational demands in many objective reservoir optimization. Adv. Water Resour. 109, 196–210.

https://doi.org/10.1016/j.advwatres.2017.09.014

Zechman, E.M., Giacomoni, M.H., Shafiee, M.E., 2013. An evolutionary algorithm approach to generate distinct sets of non-dominated solutions for wicked problems. Eng. Appl. Artif. Intell. 26, 1442–1457. https://doi.org/10.1016/j.engappai.2013.03.004

Zegre, N.P., 2008. Local and downstream effects of contemporary forest harvesting on streamflow and sediment yield. Oregon State University.

Zessler, U., Shamir, U., 1989. Optimal Operation of Water Distribution Systems. J. Water Resour. Plan. Manag. 115, 735–752. https://doi.org/10.1061/(asce)0733-9496(1989)115:6(735)

Zhang, K., Yan, H., Zeng, H., Xin, K., Tao, T., 2019. A practical multi-objective optimization sectorization method for water distribution network. Sci. Total Environ. 656, 1401–1412. https://doi.org/10.1016/j.scitotenv.2018.11.273

Zhang, Q., Li, H., 2007. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. IEEE Trans. Evol. Comput. 11. https://doi.org/10.1109/TEVC.2007.892759

Zhang, Q., Zhou, A., Suganthan, Ponnuthurai N, Zhao, S., Suganthan, Ponnuthurai Nagaratnam, Liu, W., Tiwari, S., 2009. Rescheduling View project Evolutionary algorithms for power system optimisation View project Multiobjective optimization Test Instances for the CEC 2009 Special Session and Competition.

Zhang, X., Lindström, G., 1996. A COMPARATIVE STUDY OF A SWEDISH AND A CHINESE HYDROLOGICAL MODEL. J. Am. Water Resour. Assoc. 32, 985–994. https://doi.org/10.1111/j.1752-1688.1996.tb04067.x

Zhang, X., Srinivasan, R., Bosch, D., 2009. Calibration and uncertainty analysis of the SWAT

model using Genetic Algorithms and Bayesian Model Averaging. J. Hydrol. 374, 307–317. https://doi.org/10.1016/j.jhydrol.2009.06.023

Zhang, X., Srinivasan, R., Liew, M. Van, 2010. On the use of multi-algorithm, genetically adaptive multi-objective method for multi-site calibration of the SWAT model. Hydrol. Process. 24, 955–969. https://doi.org/10.1002/hyp.7528

Zheng, F., Simpson, A., Zecchin, A., 2015. Improving the efficiency of multi-objective evolutionary algorithms through decomposition: An application to water distribution network design. Environ. Model. Softw. 69, 240–252. https://doi.org/10.1016/j.envsoft.2014.08.022

Zhou, R., Li, Y., Lu, D., Liu, H., Zhou, H., 2016. An optimization based sampling approach for multiple metrics uncertainty analysis using generalized likelihood uncertainty estimation. J. Hydrol. 540, 274–286. https://doi.org/10.1016/j.jhydrol.2016.06.030

Zitzler, E.;, Laumanns, M.;, Thiele, L., Zitzler, E., Laumanns, M., 2001. SPEA2: Improving the strength pareto evolutionary algorithm. https://doi.org/10.3929/ethz-a-004284029

Zitzler, E.;, Laumanns, M., Thiele, L., 1998. An evolutionary algorithm for multiobjective optimization: the strength Pareto approach ETH Library An Evolutionary Algorithm for Multiobjective Optimization: The Strength Pareto Approach, TIK-report, 103, research-collection.ethz.ch. Zurich. https://doi.org/10.3929/ethz-a-004288833

Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., Da Fonseca, V.G., 2003. Performance assessment of multiobjective optimizers: An analysis and review. IEEE Trans. Evol. Comput. https://doi.org/10.1109/TEVC.2003.810758

8 SUPPLEMENTARY MATERIAL

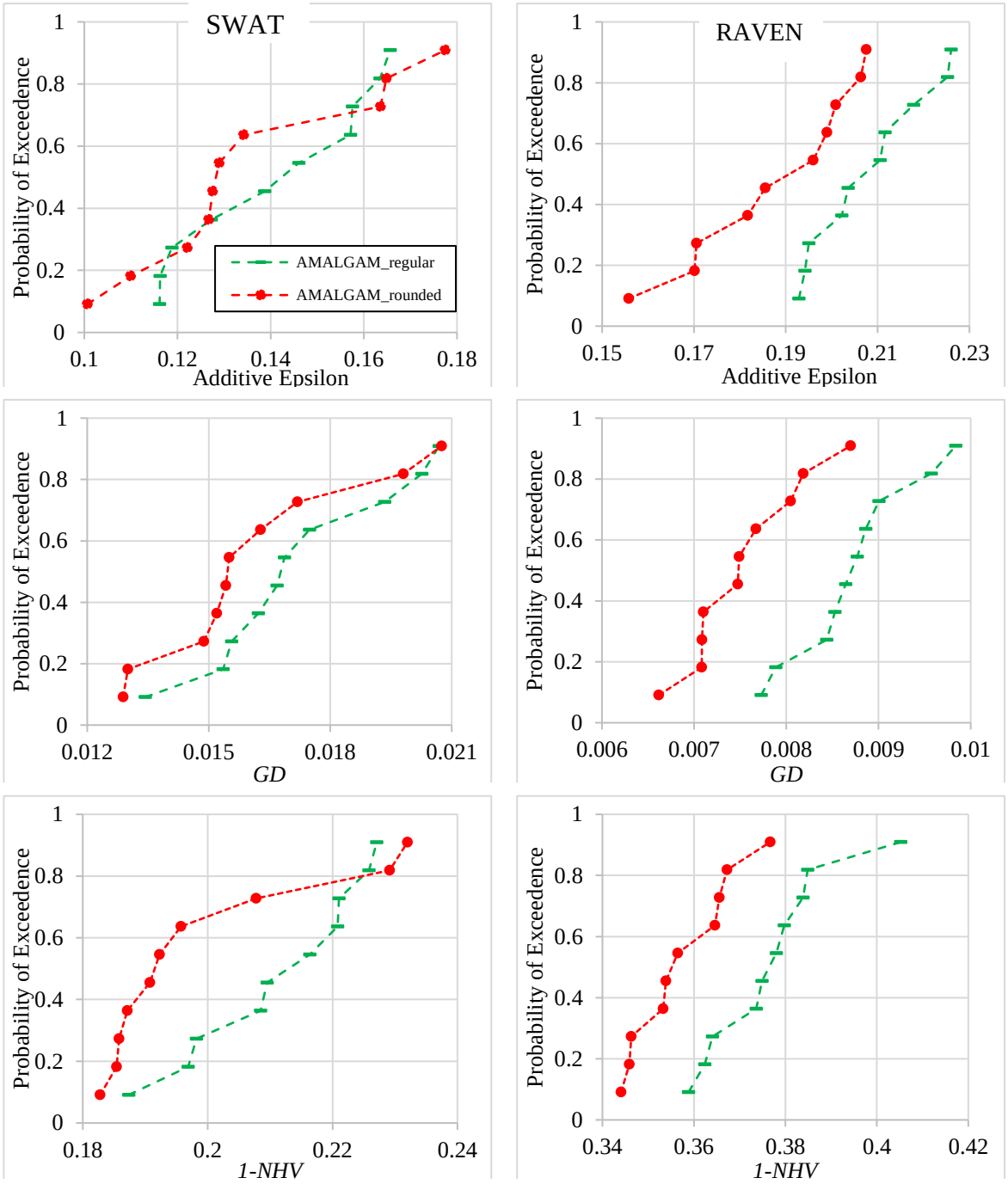


Figure 8-1: Empirical CDF plots comparing point-based (green series) versus rounded- (red series) archiving for AMALGAM for calibrating five-objective SWAT and seven-objective RAVEN models with 1000 solution evaluations

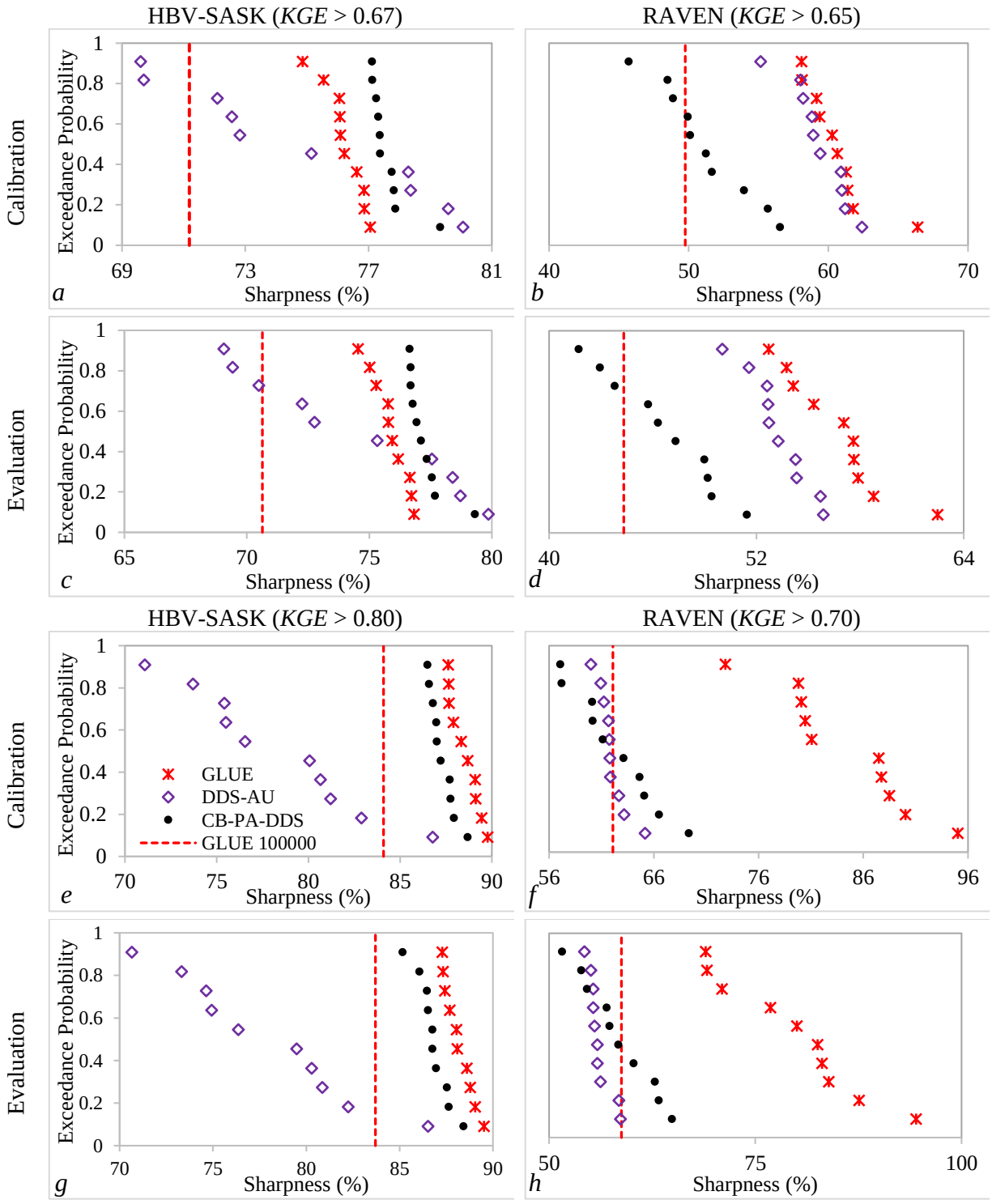


Figure 8-2: empirical CDF plot comparison of CB-PA-DDS with DDS-AU, GLUE (10000), and GLUE (100000) for evaluating their estimated uncertainty in the calibration and evaluation periods in terms of sharpness for the 12-parameter HBV-SASK (left panels) and 20-parameter RAVEN (right panels) models. The methods are compared for two *KGE* thresholds. The CB-PA-DDS parameter (ε) is equal to 0.45 and 0.55 respectively for the HBV-SASK and RAVEN models. Vertical sharpness at 100% is the ideal performance.

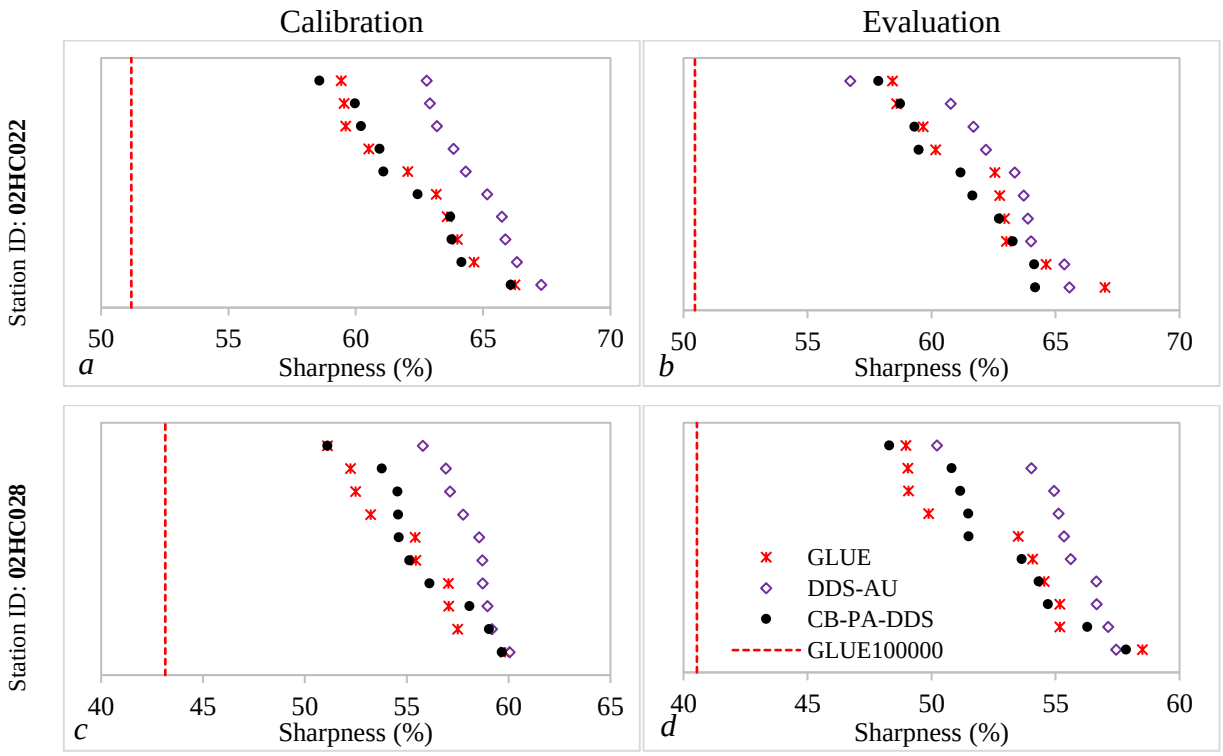


Figure 8-3: empirical CDF plot comparison of CB-PA-DDS ($\varepsilon = 0.47$) with DDS-AU, GLUE (10000), and GLUE (100000) for evaluating their estimated uncertainty in the calibration and evaluation periods in terms of sharpness metric for the 13-parameter SWAT model. The higher the value of sharpness, the better the performance.

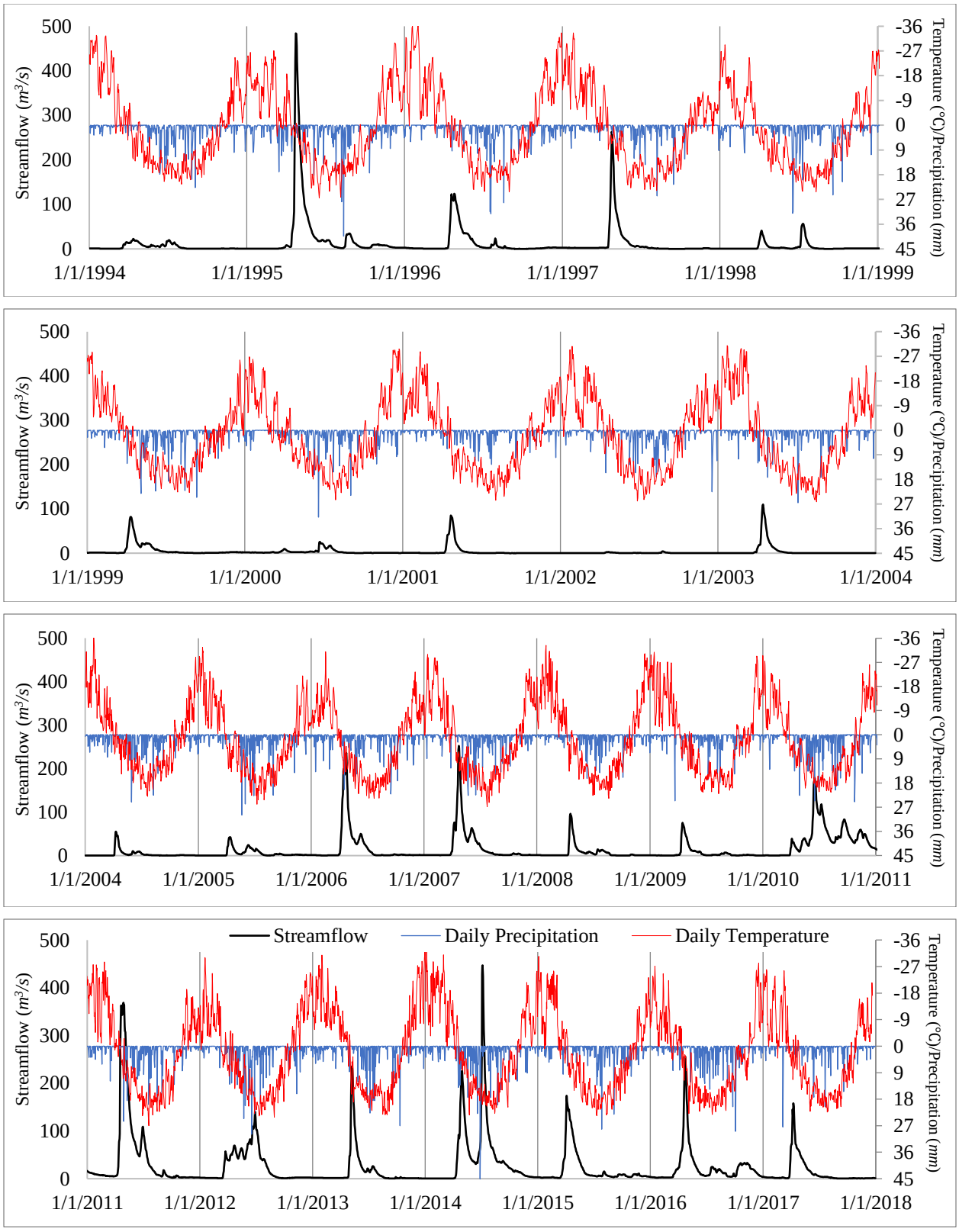


Figure 8-4: daily precipitation-runoff-temperature graph used for visual inspection and hydrological process-based categorization of the years throughout the simulation period.

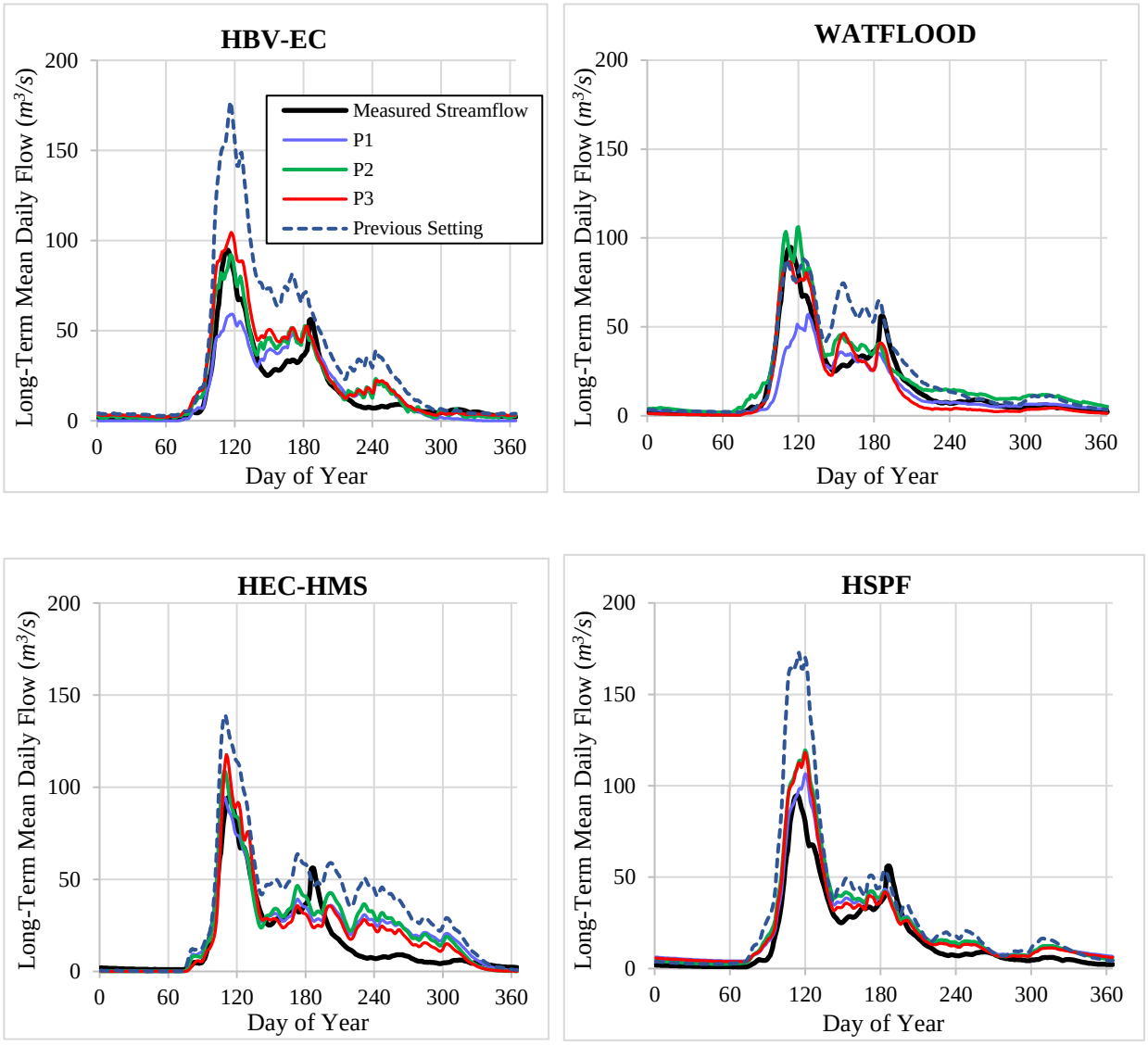


Figure 8-5: Measured versus simulated daily mean annual streamflow hydrographs for the calibration period (2001-2014) at Kamsack outlet station considering different preferred parameter sets found by each P1, P2, and P3 calibration problems.