Sensitivity-Based Guided Automatic Calibration of Hydrological Models

By:

Mohammad Semnani

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

In partial fulfillment of the requirement for the degree of

MASTER OF SCIENCE

Department of Civil Engineering

University of Manitoba

Winnipeg, Manitoba, Canada

Abstract

A new method for efficient calibration of complex hydrological models that combines Dynamically Dimensioned Search (DDS) global optimization algorithm with Global Sensitivity Analysis (GSA) methods is introduced. This approach, which is called sensitivity-informed DDS, utilizes sensitivity indices to increase the probability of perturbation for the most sensitive parameters, while giving low chance to least sensitive ones. This feature improves the efficiency and effectiveness of optimization by finding good quality solutions in a shorter time. Three different implementations of sensitivity-informed DDS are considered. The first approach is named as GSA↔DDS, in which GSA toolboxes (Morris or Sobol) are performed initially and throughout the optimization process to constantly update the sensitivity information. The second approach is called GSA→DDS. In this method, the GSA methods are only performed initially to include the results of GSA within optimization process. The final implementation is called VARS→DDS. In this method, to enhance the efficiency of sensitivity analysis and optimization, VARS toolbox is performed outside the optimization to provide the sensitivity information. The performances of GSA↔DDS, GSA→DDS and VARS→DDS are compared with original DDS by solving various optimization problems (test functions and model calibration case studies). According to the results, when calibrating complex hydrological models with enough computational budget, VARS→DDS is significantly more efficient and effective than original DDS. However, the results also show that GSA→DDS and GSA↔DDS methods do not substantially improve the convergence rate and the final best solution compared to DDS. Thus, VARS→DDS is the recommended approach for sensitivity-informed DDS in calibration of distributed and semi-distributed models, when enough computational resources are available.

Acknowledgements

I want to mention my deep appreciation to my wonderful supervisor Dr. Masoud Asadzadeh who generously supported me throughout my master's program at the University of Manitoba. His friendly guidance and expert advice brightened the path of success in my studies. I am truly honored of working with him and being a member of his fantastic research group. I would also wish to express my gratitude to Dr. Tricia Stadnyk for her valuable suggestions, which have contributed greatly to the improvement of this work.

I would like to thank my friends, who were of great support in deliberating over my problems and findings, especially my amazing friend Shahram Sahraei who was ready to help me with my thesis and research whenever I needed.

Finally, I want to thank my parents that motivated me to start my master's program, and kept me motivated and supported me throughout my study.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Problem Statement and Motivation

Hydrological models use interrelated, parametric, non-linear equations to represent the underlying real-world watershed system behavior, adequately. More advanced models have more parameters that can be automatically calibrated by means of an optimization algorithm to improve the model performance. The growing complexity of hydrological models has substantially increased the computational burden of automatic model calibration (Tang et al., 2005). For instance, in a study performed by Unduche et al. (2018), complex and fully distributed WATFLOOD model (Kouwen, 1988) with 61 calibration parameters, and less complex, semi-distributed HSPF model (Crawford and Linsley, 1966) with only 13 calibration parameters, were used to simulate runoff at the Shellmouth basin located in Canadian Prairies. The models were automatically calibrated over a ten-year period with considering common statistical metrics such as correlation coefficient (R). With equal model evaluation budget, the reported R value for HSPF was equal to 0.85, while WATFLOOD calibration resulted in lower R value (0.77). Thus, due to the significantly higher number of calibration parameters in WATFLOOD, the calibration should be performed with higher computational budget to improve the performance of WATFLOOD. In another study accomplished by Teshager et al. (2016), Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998) was automatically calibrated for Big Creek River watershed located in Illinois, USA, from 2000 to 2003. The 38 parameters of SWAT related to streamflow and water quality were calibrated. The Nash Sutcliff Efficiency (NSE) metric value for NO_3, TSS and streamflow was reported as 0.11, 0.44 and 0.54 respectively, which illustrates a fragile fit between simulated and measured data. Therefore, the high number of calibration parameters of SWAT requires more computational cost to be automatically calibrated.

Increasing the computational burden of automatic calibration of complex models has led to the development of different approaches to decrease the computational cost of model calibration. According to Razavi et al. (2010) utilization of metamodeling techniques, developing efficient optimization algorithms and utilizing parallel computation are among these approaches. In addition, a common method to efficiently calibrate computationally expensive models is the sensitivity reduction of calibration parameters (Wagener and Kollat, 2007). A brief overview of these approaches is given in the following sections.

Metamodeling approaches can significantly reduce the required computational budget of computationally expensive simulation models by replacing and emulating these models (Razavi et al., 2010). For instance, Khu et al. (2004) utilized artificial neural network as a metamodel and coupled it with genetic algorithm (GA) (Holland, 1975) to calibrate the MIKE11/ NAM rainfall-runoff model. It was reported that in comparison to related previous studies, the proposed calibration method was able to produce same or improved results; however, with only 40 percent of the simulation budget. In addition, Sakata et al. (2003) combined Kriging estimation and neural network (NN) estimation with gradient-based optimization to form an estimated solution space for solving structural optimization problem. The results showed that, the Kriging method has improved the optimum solution by 22% compared to NN estimation method. Although the metamodeling strategy has a benefit of reducing computational cost of complex calibration problems, Razavi et al. (2010) mentioned that the main limitation of metamodeling is the approximation of mathematical model, in which, the quality of this approximation could affect the model outputs. In another study by Shoemaker et al. (2007), the performance of seven global optimization algorithms were compared on calibrating SWAT2000 model applied on Town Brook watershed located in New York, USA. According to the results, among the compared algorithms, the Evolution Strategy with Radial Basis Function approximation (ESRBF) method found the least sum of squared error (SSE) value for simulated and observed streamflow data with the simulation budgets of 500, 1000 and 2500. This illustrates the benefit of ESRBF method in efficient calibration of computationally expensive models with limited budgets. Moreover, Knowles (2006) evaluated the performance of ParEGO and nondominated sorting genetic algorithm II (NSGA-II), multi-objective optimization algorithms on nine difficult test functions. The Mann–Whitney Rank sum test showed that ParEGO significantly outperforms NSGA-II in seven test functions, especially when the budget was limited (i.e. 100 and 250 function evaluations).

Utilizing efficient optimization algorithms such as Dynamically Dimensioned Search (DDS) is another strategy to increase the efficiency of model calibration. DDS is developed by Tolson and Shoemaker (2007) and is claimed to be designed for calibrating highly parameterized hydrological models. Arsenault et al. (2013) made a comparison between ten different optimization algorithms, including DDS, for the calibration of three hydrological models over ten different basins. The results showed that in almost all of the different model calibrations case studies, DDS showed faster convergence, and found significantly better NSE value particularly in models with a large number of parameters.

Parallel computing techniques can also increase the efficiency of hydrological model calibration. In parallel computing, multiple computing resources are simultaneously utilized to solve a computational problem. In these methods, the problem is divided into discrete parts and each part is further divided into a series of instructions. Instructions from each part execute simultaneously on different compute resources (processors). Therefore, the problem can be solved faster in comparison to the serial computation method. Implementing optimization algorithms in parallel format can lead to a considerable reduction in computational time (Barney, 2010; Razavi et al., 2010). For instance, in a study by Tang et al. (2007), two parallelization schemes called Master-Slave (MS) and Multiple-Population (MP) for ε-NSGAII algorithm are introduced and evaluated by applying on SAC-SMA Leaf River model calibration problem. The results showed that compared to serial version of ε-NSGAII, both MS and MP versions of ε-NSGAII with 16 number of processors increased the efficiency of calibration by 50%, as the improved ε-indicator values were being found earlier in time. In another study, Zhang et al. (2013) introduced a parallel version of A Multi-method Genetically Adaptive Multi-objective Optimization Algorithm (AMALGAM) to calibrate SWAT model, which is called PP-SWAT (Python-based parallel computation package). The PP-SWAT has been applied on two watersheds namely, Little River Experimental Watershed (LREW) located in southwest of Georgia, and South Fork Watershed (SFW) located in Iowa, USA. It was reported that, 20000 sequential runs of SWAT with one processor when applied on SFW and LREW, took 644 and 80 hours respectively, while for the same number of iterations, the run time of PP-SWAT was equal to 14 hours for SFW and less than one hour for LREW. This clearly illustrates the higher efficiency of PP-SWAT compared to serial version of SWAT model. In addition, Schutte et al. (2004) developed a parallel version of Particle Swarm Optimization (PSO) algorithm, and compared with the gradient-based optimization method by applying on optimization test functions and biomedical system identification problem. It was reported that in analytical test functions, parallel PSO enhanced the efficiency of optimization by 95%. However, due to the large number of local minima in a biomedical case study, only 25% improvement in the efficiency of parallel PSO was seen, and parallel PSO converged to the same solution in multiple trials, while the gradient-based method found distinct solutions at the end of each trial, thus parallel PSO was reported to be more consistent.

A more common approach to improve the efficiency of model calibration is by applying sensitivity analysis. In complex hydrological models that have a large number of parameters, determining the optimal value for all the parameters is not feasible (Muleta and Nicklow, 2005). Hence, by reducing the calibration parameters to the most sensitive ones, which are the most influential parameters on model outputs, the simulation cost can be reduced substantially (Muleta and Nicklow, 2005; Tang et al., 2007; Wagener and Kollat, 2007; Lu et al., 2015). As an illustration, Muleta and Nicklow (2005) utilized screening, parameterization and sensitivity analysis to reduce the number of parameters in SWAT model calibration. A stepwise regression method (Helton and Davis, 2000) was used to identify parameter's sensitivity, and reduced the calibration parameters to 20 for both streamflow and sediment yield in Big Creek Watershed located in Illinois, USA. To calibrate SWAT, they used GA, and reported 0.744 and 0.461 as the best NSE values for streamflow and sediment yield respectively. In another study carried out by Lu et al. (2015), sensitivity analysis was considered to omit the non-important parameters in calibrating SWAT for Heihe River basin located in Qilian Mountains, China. Seven parameters were identified as the most influential parameters for streamflow simulation. The NSE value for a five-year calibration was reported equal to 0.83. Nevertheless, the main problem with sensitivity-based parameter reduction for model calibration is selecting the appropriate sensitivity threshold, which can affect the model performance (Werkhoven et al., 2009).

Different methods for efficient calibration of complex hydrological models that are explained above have shown serious shortcomings. As an illustration, approximating the mathematical model in metamodeling approach can degrade the model performance, and reduce the quality of model outputs (Razavi et al., 2010). Furthermore, Werkhoven et al. (2009) calibrated SAC-SMA model with different sensitivity reduced parameter sets. The results showed that increasing the sensitivity threshold to extremely reduce calibration parameters can degrade the model performance, as the Runoff Coefficient Error (ROCE) metric was increased compared to calibration with higher number of parameters. Therefore, performing sensitivity analysis prior to model calibration to omit non-sensitive parameters from the calibration process can pose a risk of inaccurate model calibration due to the selection of incorrect sensitivity threshold. Hence, a new approach to efficiently calibrate hydrological models needs to be developed, to overcome to the shortcomings of the previous model calibration techniques.

1.2 Research Objectives

The main objective of this thesis is to increase the efficiency and effectiveness of automatic calibration for complex hydrological models by developing a method that combines the global sensitivity analysis and global optimization. This approach, which is called sensitivity-informed optimization, utilizes sensitivity information to prioritize parameter perturbation in the optimization process. In other words, instead of removing low-sensitivity parameters from calibration process, a low chance for perturbation is given to them based on their sensitivity index, while parameters that are more sensitive are perturbed with higher probability. Sensitivity-informed optimization is intended to increase the efficiency of the hydrological model calibration by identifying a solution that has an equal or better objective function value in a relatively shorter time compared to other algorithms. In addition, this new method can enhance the effectiveness of model calibration by discovering a solution that has a better value of the calibration objective function compared to solutions found algorithms with equal computational budget. However, one should be noted that a better quality solution does not necessarily represent a better calibrated model due to uncertainty in model calibration process. Figure 1 illustrates the overall flowchart for different implementations of sensitivity-informed optimization that are tested in this thesis to reach the defined objectives.

The other objectives of the current thesis are listed below:

- Calculate parameter sensitivity indices at the end of calibration that provides insights in which parameters contribute most to the calibration process (Holvoet et al., 2005). The sensitivity information can be useful in model recalibration when new forcing data is available.
- Investigate the performance of proposed methodology on a wide range of optimization problems from complex hydrological models to simple optimization test functions.
- Compare the efficiency and consistency of different GSA approaches applied to the hydrological model calibration problems.
- Calibrate hydrological models with reduced and full parameter sets to show the advantages and disadvantages of parameter reduction in model calibration.
- Implement the proposed methodology with different GSA tools to identify the most suitable and effective GSA method for sensitivity-informed optimization in model calibration.

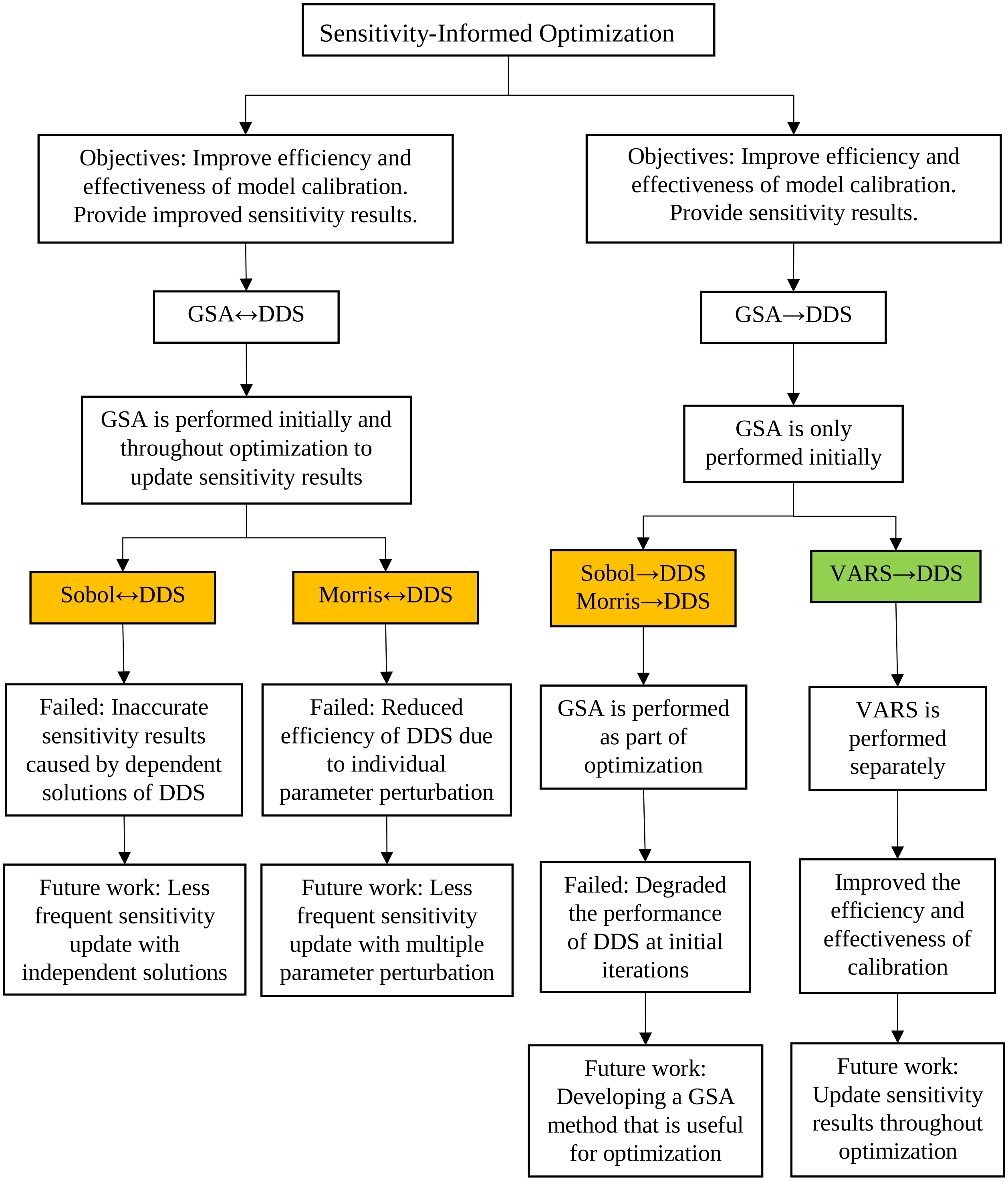

Sensitivity-Informed Optimization
Objectives: Improve efficiency and effectiveness of model calibration. Provide improved sensitivity results.
Objectives: Improve efficiency and effectiveness of model calibration. Provide sensitivity results.
GSA↔DDS
GSA→DDS
GSA is performed initially and throughout optimization to update sensitivity results
GSA is only performed initially
Sobol↔DDS
Morris↔DDS
Sobol→DDS
Morris→DDS
VARS→DDS
Failed: Inaccurate sensitivity results caused by dependent solutions of DDS
Failed: Reduced efficiency of DDS due to individual parameter perturbation
GSA is performed as part of optimization
VARS is performed separately
Future work: Less frequent sensitivity update with independent solutions
Future work: Less frequent sensitivity update with multiple parameter perturbation
Failed: Degraded the performance of DDS at initial iterations
Improved the efficiency and effectiveness of calibration
Future work: Developing a GSA method that is useful for optimization
Future work: Update sensitivity results throughout optimization

Figure 1. The overall flowchart of different approaches for sensitivity-informed optimization tested in this thesis. The approaches colored in orange and green failed and suceeded to meet the objectives respectively.

2 Hydrological Modelling and Model Calibration

2.1 Hydrological Modelling

Hydrological models are mathematical tools that simplify water-related real-world systems such as surface water, groundwater, and wetlands and so on. They help to simulate the underlying processes of the real-world watershed systems and therefore can be used for planning optimal and sustainable use of water resources (Abbaspour et al., 2015). Based on the characteristics of hydrological models, they can be categorized as deterministic or stochastic, physically-based or conceptual and lumped or distributed models (Beven, 2001).

In lumped models, the spatial variability of the studied area is ignored, and the entire area is considered as a single unit. Moreover, in these models, the model response can be evaluated only at the outlet point and not within the study area. Obviously, parameters of lumped models do not allowed to be modified across the basin, and they do not necessarily represent the physical features of hydrological processes. In addition, these models are simple to use and require minimal input data to run (Moradkhani and Sorooshian, 2009; Sahu et al., 2012). Well known lumped models are Sacramento Soil Moisture Accounting (SAC-SMA) Model (Sorooshian et al., 1993) and HEC-HMS developed by US Army Corps of Engineers (USAC-HEC, 2016).

On the other hand, distributed models divide the studied area into small units with the shape of square cells or triangular networks. Therefore, in these models, the basin is modeled at these units, and the model response can be calculated at any location across the basin. Examples of distributed models include WATFLOOD (Kouwen, 1988) and HYDROTEL (Fortin et al., 2001) hydrological models. Unlike the lumped models, in distributed models, parameters can vary spatially within the defined cells, and users can define the resolution of these cells to comply with the available computational resources. Distributed models require a large amount of input data and can produce results not only at the outlet but also at any locations across the basin. If accurate input and measured data are provided, distributed models can produce highly accurate results compared to other rainfall-runoff models (Moradkhani and Sorooshia, 2009; Sahu et al., 2012; Devia et al., 2015).

The major disadvantage of distributed models compared to the lumped models is that usually distributed models require a large amount of distributed input data to accurately generate model outputs. Thus, these models might fail to generate adequate results for regions that suffer from hydrological and meteorological data scarcity. In addition, distributed models need excessive computational resources for simulation and therefore for calibration (Yaduvanshi et al., 2018) compared to the lumped models. On the other hand, distributed models are expected to produce more accurate hydrological outputs (Carpenter and Georgakakos, 2006) given accurate input data are available.

Semi-distributed (SD) hydrological models such as, SWAT and HSPF are somewhere between distributed and lumped models. SD models, lump meteorological and physical parameters into sub basins. Hence, they require less amount of data and computational resources for simulation and are more user-friendly compare to distributed models (Yaduvanshi et al., 2018). Moreover, SD models perform hydrological predictions with significantly higher spatial and temporal resolutions than lumped models (Carpenter and Georgakakos, 2006). However, SD models have their own shortcomings. They have more parameters and require more input data compared to lumped models. Yaduvanshi et al. (2018) compared the parameter uncertainty between SWAT and the lumped probability distribution model (PDM) using GLUE method. The results showed that the reported p factor of Percent Prediction Uncertainty (PPU) band (which represent the parameter uncertainty) for SWAT and PDM were 74% and 64% respectively. Therefore, SD models such as SWAT, demonstrate higher parameter uncertainty compared to lumped models (Yaduvanshi et al., 2018). Furthermore, according to Pina et al. (2014), distributed models are expected to generate more realistic and accurate results at a higher resolution compared to SD models, given accurate input data are available. In conclusion, a modeler has to select lumped, semi-distributed and/or distributed hydrological models according to the characteristics of the studied area, aim of the project, data availability, and available computational resources.

Deterministic versus stochastic is another classification of hydrological models. In deterministic models, the model simulated response is a single realization of the underlying processes. Nevertheless, stochastic models utilize a post-processing procedure that includes model uncertainty. In fact, in deterministic models, model output is fully determined by model parameters, while in the stochastic models, the randomness in the model causes the same parameter set to create different outputs (Devia et al., 2015; Farmer and Vogel, 2016; Sanchez-Vila and Fernandez-Garcia, 2016). In stochastic models, every parameter is represented by a spatial random function. Thus, uncertainty in these models is substantially higher than deterministic models (Sanchez-Vila and Fernandez-Garcia, 2016).

According to Devia et al. (2015), based on the description of hydrological processes used in models, they can also be categorized as conceptual or physically-based. Conceptual hydrological models aim to simplify the complex underlying watershed processes. The main structural components of these models are the empirical equations that are derived based on observation of hydrological processes. Nevertheless, in physically-based models the aim is to understand and incorporate into the model the hydrological phenomena in a watershed system. Therefore, physically-based models are supposed to be able to estimate the effect of watershed modifications (e.g. land cover change and crop rotation) on these phenomena. In these models, the simulation of complex watershed processes is performed based on laws of physics, such as, conservation of mass, momentum and energy with minimal simplifying assumptions compared to conceptual models. Conceptual and physically-based models could also be categorized as lumped, semi-distributed, or distributed (Sahu et. al, 2012; Devia et al., 2015). SWAT and SAC-SMA are among the famous physically-based and conceptual hydrological models, respectively.

2.2 Model Calibration

According to Sorooshian and Gupta (1995), parameters of hydrological models can be categorized as process and physical. Parameters that can be measured directly are called physical parameters, while the process parameters must be inferred by indirect methods and cannot be measured directly. Physical parameters include but are not limited to the watershed area, impervious area in a watershed, areal percentage of water bodies (Sorooshian and Gupta, 1995). On the other hand, parameters such as, soil moisture storage, effective lateral interflow, mean hydraulic conductivity, and surface runoff coefficient are classified under process parameters (Gupta et al., 1998; Moradkhani and Sorooshian, 2009).

The procedure of adjusting process parameters of hydrological models seeking for a better model performance to simulate historical records (e.g. measured streamflow) is called model calibration. In other word, the main objective of hydrological model calibration is to minimize the error between measured and simulated basin responses (Smith et al., 2003). Moreover, calibration of a hydrological model reveals if the model properly represents hydrological processes of the studied area in the absence of equifinality issue (Beven, 2010; Westerberg et al., 2011). Thus, when applying hydrological models to a new basin, model calibration is necessary. As an illustration, in the study by Muleta and Nicklow (2005), the NSE value for the streamflow simulation in the Big Creek Watershed (Illinois, USA) corresponding to default and calibrated parameter values for SWAT model was reported as -0.38 and 0.75 respectively. This shows the significant improvement of model performance caused by calibration. In addition, Holvoet et al. (2005) reported the NSE value for streamflow simulation in Nil basin (Brussel, Belgium) equal to -22.4, when default value for SWAT parameters were used. Nonetheless, the NSE value increased to 0.53, when calibrated parameter values were used.

The two most known approaches for hydrological model calibration are manual and automatic. Manual calibration is a basic method that estimates the value of parameters through a semi-intuitive trial and error process (Boyle et al., 2000). The benefit of manual calibration is that this process is less affected by inaccurate data compared to automatic calibration (Moradkhani and Sorooshian, 2009). However, a high number of model parameters that nonlinearly interact with each other makes the manual calibration process time-consuming. Therefore, to compensate this issue, automatic calibration is introduced (Duan et al., 1993; Gupta et al., 1998; Moradkhani and Sorooshian, 2009). In automatic calibration, computer-based algorithms perform the calibration of hydrological models, and human judgment that is involved with manual calibration is removed (Boyle et al., 2000). Automatic calibration methods consider model calibration as an optimization problem. In these methods, the objective of optimization is minimizing an error function that estimates the overall error between measured and simulated data. Examples of well-known model calibration metrics include Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970), Root Mean Square Error (RMSE), Coefficient of Determination (R^2) (Wright, 1921) and Percent BIAS (PBIAS). When calibrating models as an optimization problem (Figure 2), model parameters (X) are perturbed to find the best value for parameters that produce minimum error.

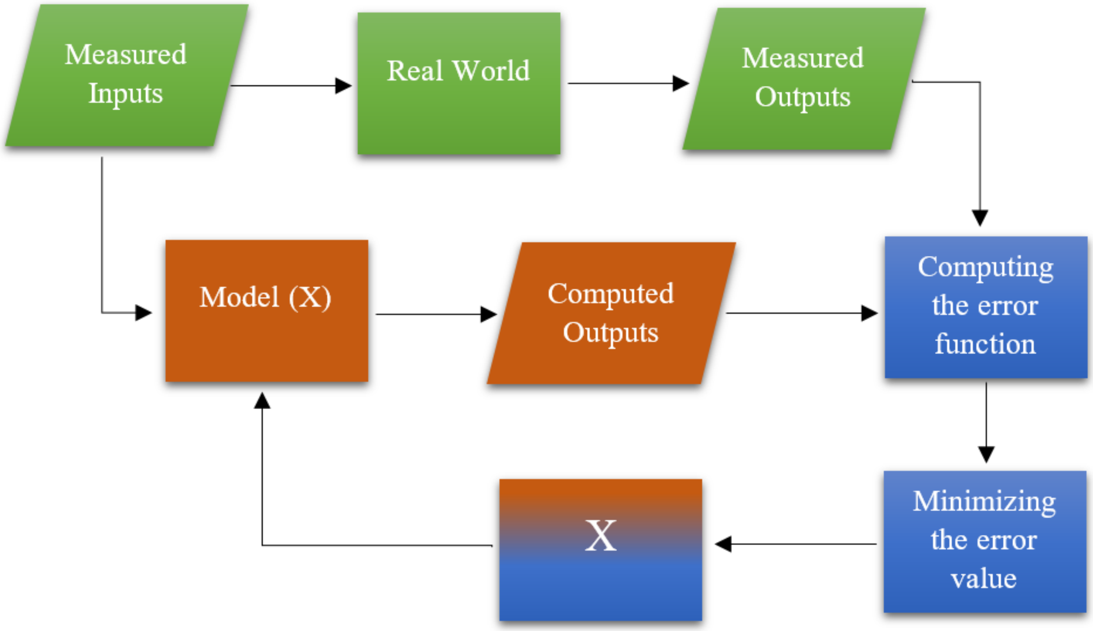


Figure 2. The scheme of automatic model calibration as an optimization problem (Moradkhani and Sorooshian, 2008).

The initial attempts for automatic calibration of models were based on local optimization algorithms (Dawdy and O'Donnell, 1965; Ibbitt, 1970). The local methods start their search at a point in the solution space, and they direct parameter search to improve the objective function value in the vicinity of each new solution, by minimizing or maximizing it (Moradkhani and Sorooshian, 2009). The main classification of local optimization algorithms is gradient-based (derivative-based) and direct search (derivative free) methods. Compared to direct search methods, gradient-based optimization algorithms have more strength in finding optimum objective values as they estimate and use first and/or second order partial derivatives of the objective function (Moradkhani and Sorooshian, 2009). The famous Newton-Raphson method is a gradient-based local optimization algorithm while, Pattern search (Hook and Jeeves, 1961) and downhill simplex developed by Nelder and Mead (1965) are examples of direct methods. These algorithms had been widely used in early model calibration studies (Johnston and Pilgrim, 1976; Pickup, 1977).

As explained by Hendrickson et al. (1988), in local optimization algorithms, the outcome is highly dependent on characteristics of model response in the vicinity of the initial point. This issue could mislead the search toward local optimum points instead of finding the global or near global optimal solutions. Therefore, global optimization algorithms were introduced to increase the efficiency and effectiveness of optimization, and to overcome the mentioned issues of local optimization methods as well as coping with the large number of parameters in hydrological models (Moradkhani and Sorooshian, 2009). For instance, Genetic algorithm or GA is a famous global optimization algorithm developed by Holland (1975) that has been widely used for hydrological model calibration (Wang, 1991; Seibert, 2000). Another global optimization algorithm is the Shuffle Complex Evolution or SCE-UA algorithm (Duan et al., 1992). SCE-UA has also been considered for calibration of hydrological models (Sorooshian et al., 1993; Duan et al., 1994). Conversely, local search algorithms can be used for global optimization if they start their search with different initial solutions to search globally in the solution space. However, this will increase the computational cost of optimization. Accordingly, another advantage of global optimization algorithms over local ones is that they require less computational budget, as they do not need to be performed repetitively (Moradkhani and Sorooshian, 2009).

The global optimization approaches can be categorized as Single-Objective (SO) and Multi-Objective (MO) methods. In SO optimization method, algorithm searches for an optimal value for the objective, while in MO optimization there could be several contradicting objectives that makes it impossible to find a single optimal value but rather a set of optimal values that is called Pareto-optimal solutions (Abraham and Jain, 2005; Deb, 2014). Examples of famous single-objective and multi-objectives global optimization algorithms respectively include GA and NSGA-II algorithms. The MO methods are introduced later than SO methods to better solve real-world optimization problems using several conflicting objectives (Deb, 2014). Unlike SO methods, MO methods generate different optimal solutions that enable users to select the most suitable one based on objectives and decision variable values. However, in this research, in order to develop the sensitivity-informed optimization, a single-objective optimization algorithm is considered simply because SO methods have been widely used in the literature (Pardalos et al., 2017), and it is easier to develop the methodology. Developing the MO version of sensitivity-informed optimization is considered as the future work of this research.

The computational cost of hydrological model calibration mainly depends on the number of calibration parameters. As an illustration, since lumped models ignore the spatial variability of the studied area, they have lower numbers of parameters, which means these models require a lower computational budget for calibration than distributed and semi distributed models. (Muleta and Nicklow, 2005; Moradkhani and Sorooshian, 2009). However, in most watersheds, many factors such as land use, topographical conditions and soil type are varying spatially. Hence, in order to properly simulate the watershed response (e.g. streamflow), semi-distributed and distributed models must be considered (Muleta and Nicklow, 2005), even though, they require higher computational cost for calibration caused by higher number of parameters (Muleta and Nicklow, 2005; Van Griensven et al., 2006).

In order to cope with the curse of dimensionality in hydrological model calibration, more efficient and effective calibration techniques are required. Among the various methods for efficient calibration of hydrological models such as, metamodeling and parallel computing, parameter reduction using sensitivity analysis has been widely used in the literature (Holvoet et al., 2005; Muleta and Nicklow, 2005; White and Chaubey, 2005; Werkhoven et al., 2009; Lu et al., 2015; Gholami et al., 2016). According to Madsen (2003), most sensitive parameters mainly control the reproduction of model response in the process of model calibration. Hence, in sensitivity aided model calibration method, a sensitivity analysis is performed to identify the most sensitive parameters and reduce the calibration parameters to the most sensitive ones. In the following chapter, a comprehensive discussion about different sensitivity analysis methods and their application in model calibration is presented.

3 Sensitivity Aided Model Calibration

3.1 Sensitivity Analysis

The sensitivity analysis is commonly defined as measuring the effect of one unit change in one or more parameter(s) on the output of a model (McCuen, 1973). Sensitivity analysis has been performed in a wide range of disciplines, such as, business, environmental sciences, social sciences, engineering and so on. In mathematical modeling, sensitivity analysis has a crucial role in the way that it can be used to investigate the effect of model parameters on model output (Rakovec et al., 2014).

Sensitivity analysis approaches are categorized under deterministic and statistical methods. Examples of deterministic approaches include Green’s Function Method (GFM) and Forward Sensitivity Analysis Process (FSAP) (Cacuci and Ionescu-Bujor, 2004). Furthermore, the most well-known statistical approaches of sensitivity analysis are the ones introduced by Morris (Morris, 1991) and Sobol (Sobol, 1990). Although deterministic methods can reveal the true sensitivity values for parameters, statistical methods are relatively easier to develop and more user-friendly (Cacuci and Ionescu-Bujor, 2004). In addition, in this thesis, for the purpose of developing sensitivity-informed optimization, an stochastic optimization method is considered. Therefore, to increase the consistency between sensitivity analysis and optimization, and to be able to update the sensitivity results from optimization, statistical approaches for sensitivity analysis have been considered in this thesis.

Sensitivity analysis is also classified into two major categories of local sensitivity analysis (LSA) and global sensitivity analysis (GSA). In LSA methods, the sensitivity indices are calculated either theoretically or numerically based on the derivatives of the model output with respect to the model parameters (Rakovec et al., 2014). To illustrate this concept, consider a model with the output y and input parameters as $x_1, x_2, \dots, x_n$. In LSA methods, the sensitivity index of parameter x_i (S_i) is calculated by the following formula:

$$S_i = \frac{\partial Y}{\partial x_i} \qquad \text{Equation 1}$$

LSA approaches have the advantage of being computationally efficient due to the low number of required simulations (Saltelli et al., 2008). Therefore, as can be seen in the literature, LSA methods have been significantly utilized (Razavi and Gupta, 2015). For instance, Tang et al. (2006) used different sensitivity analysis methods including local analysis using parameter estimation software (PEST) to increase the model calibration efficiency. However, LSA approaches have serious shortcomings. In these methods, sensitivity is performed at a single location in parameter space. Therefore, in nonlinear models the sensitivity results depend on the location that sensitivity is calculated, as the results will change substantially with varying the location. This will cause a limited view of parameter's sensitivity (Razavi and Gupta, 2015). In addition to that, LSA methods neglect the interaction effects between parameters (Kucherenko et al., 2009). Hence, the inaccurate and incorrect results of sensitivity could mislead modelers and users (Rakovec et al., 2014).

While LSA methods calculate the sensitivity at individual point in the parameter domain, GSA approaches provide a more comprehensive assessment of sensitivity at the entire parameter space. Measurement of the parameter sensitivity in GSA methods includes the variation of other parameters as well. Therefore, unlike LSA methods, GSA methods can account for the interaction effects between parameters (Kucherenko et al., 2009). GSA approaches calculate the sensitivity indices at a number of different points across the parameter space to identify parameter sensitivity properly (Razavi and Gupta, 2015). Hence, GSA is an aggregation of several LSAs computed at different points in parameter space. Tang et al. (2006) compared several sensitivity analysis including PEST LSA and Sobol GSA methods applied to the calibration of the SAC-SMA model and reported that compared to PEST, Sobol GSA method showed more consistent results, as the sensitivity rankings were less variable between different trials. In all GSA techniques, there are two different types of sensitivity measurements. One is the main effect or first order sensitivity index, which can be defined as the effect of variation in one parameter on the model output. The other one is the second or higher order sensitivity index, which accounts for the interaction effect between two or more parameters, when they are varied simultaneously. GSA tools are categorized under derivative-based, variance-based and variogram-based methods. The most well-known derivative-based, variance-based and variogram-based approaches are respectively Morris, Sobol and Variogram Analysis of Response Surface (VARS), which are described in the following sections.

3.1.1Derivative-Based Sensitivity Analysis

According to the definition, sensitivity of a parameter can be calculated by the derivative of the model output(s) of interest with respect to that parameter. The most well-known derivative-based sensitivity analysis methods is the elementary effect test (also called Morris method) developed by Morris (1991). In this approach, N number of elementary effect (EE) indices are calculated for each parameter to generate the main and interaction effects. The main effect or the first order sensitivity index of a parameter is identified by factor σ, which is the standard deviation of EE indices, and the interaction effect of each parameter is shown by factor μ, which represents the mean of EE indices. The Morris method uses the famous one parameter at a time (OAT) sampling technique, in which at each step only one parameter is varied. In order to perform sampling, parameters are assumed to be normally distributed with a range between zero and one and then transformed into their actual distributions. The normal distribution is considered to increase the chance of symmetric sampling for each parameter (Salteli et al., 2008). Then, the n dimensional parameter space is divided into equally spaced p spans, and each parameter takes p number of values to form an n by p sampling matrix that is called region of experiment (ω). Thus, the parameter values are randomly selected from the ω matrix to calculate the EE index for each parameter. The EE index of i^{th} parameter (X_i) of a function y with n number of parameters can be calculated by Equation 2.

$$EE_i = \frac{y(X_1, \dots., \ X_{i-1}, X_i + \Delta, \ X_{i+1}, \dots, X_n) - y(\boldsymbol{X})}{\Delta} \qquad \text{Equation 2}$$

Where $\mathit{\Delta}$ is the step size, in which $X_i + \Delta \leq 1$, and $\boldsymbol{X}$ is the vector of selected values for parameters. As recommended by Morris (1991), in order to reduce the number of simulations when more than one elementary effect is computed, p should be considered as an even number and $\mathit{\Delta}$ should be equal to $p/[2(p-1)]$. The p value should be defined based on the value of N. As suggested by Morris (1991), the total number of model evaluations in the Morris method for a model with n parameters is equal to $N(n+1)$. The standard deviation and mean values of EE indices for each parameter is given as the first and higher-order sensitivity index, respectively. Thus, as N increases, the reliability of sensitivity results will also increase (for additional details see Morris, 1991; Campolongo et al., 2007).

The OAT sampling (Morris 1991) uses Monte Carlo (MC) method. The MC sampling methods are notorious for large computational costs due to unorganized random production of samples (Fishman, 2013). Therefore, in order to achieve a reasonable coverage of parameter space, MC method requires many samples, which significantly increase the number of model evaluations (Muleta and Nicklow, 2005). By replacing the MC sampling in OAT method with Latin Hypercube

(LH) design (McKay et al., 1979) the parameter space can be optimally covered without excessive number of simulations (Iman and Conover, 1980; Van Griensven et al., 2006). The new sampling approach is called latin hypercube one parameter at a time (LH-OAT) developed by Van Griensven et al. (2006). The idea of LH sampling is to generate non-overlapping samples of the input parameters. In LH sampling, each parameter range is divided into N intervals with equal probability of $1/N$. Then, parameters are randomly perturbed such that in the direction of each parameter each interval is sampled only once. For instance, in a model with two parameters (illustrated in Figure 3) the LH-OAT sampling utilizes LH design to generate initial points for OAT sampling. After various sample points across parameter space are generated, each of these points are perturbed n times by perturbing each of the n parameters individually. In other words, from each LH point, parameter one and parameter two are perturbed separately based on OAT sampling method (Van Griensven et al., 2006).

The Morris method is a GSA tool, as it explores the entire parameter space and calculates the sensitivity at different points (Saltelli et al., 2008). However, the major shortcoming of Morris approach is that it is a model-based method, due to its reliance on a step size (Δ) to calculate the derivatives. Hence, when applied on non-linear models, its results will significantly change with modifying the Δ value (Saltelli et al., 2008).

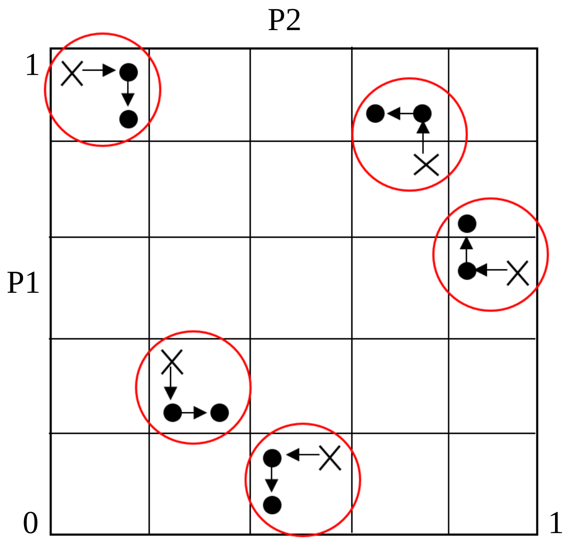


Figure 3. LH-OAT sampling for a two-parameter model. X and black points are represent the LH and OAT sample points respectively (Van Griensven et al., 2006)

3.1.2Variance-Based Sensitivity Analysis

Variance-based sensitivity analysis methods are more complicated and sophisticated than derivative-based tools. As oppose to derivative-based methods such as Morris approach, variance-based techniques are attractive because they produce model-free results. In these methods, the degree of nonlinearity of model output does not affect the sensitivity results, as model-free GSA approaches do not depend on a step size value to generate sensitivity indices. Hence, in general, the variance-based methods are expected to generate more accurate results. (Salteli et al., 2008).

The Sobol method developed by Sobol (1990) is a benchmark variance-based sensitivity analysis method. The method works based on the idea of decomposing the model output variance. Hence, Sobol showed that the total variance of a model output $\left(V(y)\right)$ with n number of parameters can be theoretically decomposed into $2^n - 1$ components as shown below:

$$V(y) = \sum_{i=1}^{n} V_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} V_{ij} + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} V_{ijk} + \dots + V_{1\dots n} \qquad \text{Equation 3}$$

Where V_i is the variance of the model output with respect to the variation of i^{th} parameter, and V_{ij} is representing the variance of the model output with respect to interaction between i^{th} and j^{th} parameters. V_{ijk} is also defined as the model output variance with respect to three parameters interaction. V_i, V_{ij} and V_{ijk} are calculated by the following formulas:

$$V_i = V\big(E(y|X_i)\big) \qquad \text{Equation 4}$$

$$V_{ij} = V\left(E\big(y\big|X_i, X_j\big)\right) - V_i - V_j \qquad \text{Equation 5}$$

$$V_{ijk} = V\left(E\big(y\big|X_i, X_j, X_k\big)\right) - V_i - V_j - V_k - V_{ij} - V_{ik} - V_{jk} \qquad \text{Equation 6}$$

In which, $E(y|X_i)$ is interpreted as the expected value of y given X_i. The same definition is applicable to $E\big(y\big|X_i, X_j\big)$ and $E\big(y\big|X_i, X_j, X_k\big)$. In Sobol method, the sensitivity index for each parameter is considered as the ratio of the model output variance with respect to i^{th} parameter to the total variance of model response ($V(y)$), which is calculated with respect to variation in all the parameters simultaneously. Hence, after identifying the values of V_i, V_{ij} and V_{ijk} the values of first, second and third order sensitivity indices can be identified by the following formulas:

$$S_i = \frac{V_i}{V(y)} \qquad \text{Equation 7}$$

$$S_{ij} = \frac{V_{ij}}{V(y)} \qquad \text{Equation 8}$$

$$S_{ijk} = \frac{V_{ijk}}{V(y)} \qquad \text{Equation 9}$$

Sobol introduced another sensitivity index that is called the total effect that identifies the total contribution of each parameter in the output variation. This index sums over the first order and the higher order effects of each parameter (Saltelli et al., 2008). For a model with three parameters the total effect index of parameter one can be calculated as follow:

$$S_{T1} = S_1 + S_{12} + S_{13} + S_{23} + S_{123} \qquad \text{Equation 10}$$

Where S_1 represents the first order sensitivity index for parameter one, and other indices account for interaction between parameter one and other parameters. As explained by Razavi and Gupta (2015), the general term for calculating the Sobol total effect index for i^{th} parameter of a model with n number of parameters is shown as follow:

$$S_{Ti} = 1 - \frac{V(E(y|X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_n))}{V(y)} \qquad \text{Equation 11}$$

In equation 11, the numerator contains all the terms of any order except the term with i^{th} parameter.

Campolongo and Saltelli (1997) evaluated the performance of Sobol and Morris methods by applying to the GMSK climate model (Gabric et al., 1996). The GMSK climate model is the extended version of dimethylsulphide (DMS) model. It was reported that, both Morris and Sobol methods were able to identify influential parameters, and the sensitivity rankings for influential parameters were similar with the correlation value equal to 0.66. In order to calculate the Morris sensitivity index for each parameter, the total number of 330 simulations were performed associated with ten independent samples for each parameter. However, with Sobol method, the number of samples and total number of simulations were increased to 100 and 3400 respectively (Campolongo and Saltelli, 1997). Thus, the main drawback of the original Sobol approach is the excessive computational cost required to numerically estimate the sensitivity indices. To tackle this issue, Saltelli et al. (2008) proposed an approximation approach to calculate the first order and total order sensitivity indices more efficiently. In this method, two independent N × n matrices, called A and B, of input parameter samples are generated using the LH design (n is the number of input parameters and N is the sample size). In the next stage, a matrix called C_i is generated (for i^{th} parameter) from all columns of B except for the i^{th} column, which is taken from A. After identifying all the matrices of A, B and C_i, the corresponding model output vectors of y_A, y_B and y_{Ci} are generated. To calculate the first order (S_i) and total order (S_{Ti}) sensitivity indices Saltelli et al. (2008) suggested the following formula:

$$S_i = \frac{\left(\frac{1}{N}\right)\sum_{j=1}^{N} y_A^{(j)} y_{Ci}^{(j)} - \left(\frac{1}{N^2}\right)\sum_{j=1}^{N} y_A^{(j)} \sum_{j=1}^{N} y_B^{(j)}}{\left(\frac{1}{N}\right)\sum_{j=1}^{N}(y_A^{(j)})^2 - f_0^2} \qquad \text{Equation 12}$$

$$f_0 = \frac{1}{N}\sum_{j=1}^{N} y_A^{(j)} \qquad \text{Equation 13}$$

$$S_{Ti} = 1 - \frac{\left(\frac{1}{N}\right)\sum_{j=1}^{N} y_B^{(j)} y_{Ci}^{(j)} - f_0^2}{\frac{1}{N}\sum_{j=1}^{N}(y_A^{(j)})^2 - f_0^2} \qquad \text{Equation 14}$$

The total number of model evaluations in this proposed method is $N(n+2)$, which is significantly lower than the N^2 in the original Sobol method, if the number of samples is greater than the number of model parameters (Saltelli et al., 2008). For more information regarding Sobol method and other variance-based methods see Sobol (1993); Saltelli et al. (2008).

The Sobol and Morris approaches have shortcomings in performing GSA. As explained by Razavi and Gupta (2015), in the Morris method, computing the sensitivity indices is highly dependent on the step size (Δ) used for numerical evaluation of local sensitivities. Therefore, selecting different step sizes can significantly change the results. Furthermore, variance-based methods are extremely time consuming, especially when applied on high dimensional models, even when using the approach proposed by Salteli et al. (2008).

3.1.3Variogram-Based

To address the issues of previous GSA approaches, Razavi and Gupta (2016) introduced the Variogram Analysis of Response Surface (VARS) toolbox. VARS is a modern global sensitivity analysis method that works based on the variogram analysis. Variogram is defined as the gradient of the model output variance calculated at a large number of pairs of point across the parameter space (Razavi and Gupta, 2016). In mathematical models, the variogram analysis can be used to illustrate the spatial structure and variability of a model parameter, which is similar to the definition of sensitivity analysis (Razavi and Gupta, 2016). Therefore, the variogram analysis can properly represent sensitivity analysis. As mentioned by Razavi and Gupta (2016), a GSA method is comprehensive when it contains several characteristics of model output simultaneously. Therefore, to be considered as a comprehensive GSA toolbox, VARS includes several features of model output such as, local sensitivity and its global distribution, global distribution of model output and its structural organization. In addition, VARS can generate the results of derivative-based and variance-based methods as its byproducts (Razavi and Gupta, 2016).

In VARS, the sensitivity of each parameter is represented by the directional variogram that is defined as the variogram of model output in the direction of each parameter. For the i^{th} parameter of a model with n input parameters, the directional variogram can be computed using Equation 15:

$$\gamma(h_i) = \frac{1}{2} h_i^2 V\left(\frac{y(x_1, \ldots, x_i + h_i \ldots, x_n) - y(x_1, \ldots, x_i, \ldots, x_n)}{h_i} \right) \qquad \text{Equation 15}$$

Where $h_i = X_i^A - X_i^B$ which represents the difference between i^{th} parameter at different locations (A and B), V is variance and $\gamma(h_i)$ is the directional variogram of i^{th} parameter calculated at a pair of points. The variance is calculated using multiple sample points created by a

STAR sampling procedure, which is explained later in this section. Razavi and Gupta (2016) have illustrated that at any value of h_i, if the value of $\gamma(h_i)$ increases the sensitivity of model output to i^{th} parameter will be increased. This shows that directional variogram represents the sensitivity index for each parameter.

In addition to the directional variogram, VARS generates a useful indicator of parameter sensitivity, which is called integrated variogram across a range of scales (IVARS). According to Razavi and Gupta (2016), a model's nonlinearity can reduce the accuracy of different types of sensitivity measurements. However, IVARS finds a meaningful measure of parameter sensitivity when a particular scale that provides accurate assessment of sensitivity does not exist. IVARS is calculated by integrating directional variogram over different parameter range scales (H_i).

$$\Gamma(H_i) = \int_0^{H_i} \gamma(h_i) dh_i \qquad \text{Equation 16}$$

It is worth mentioning that in calculating directional variogram and IVARS, the parameter ranges are normalized between zero and one. As mentioned by Razavi and Gupta (2016), IVARS can be calculated at different H_i values of 0.2, 0.4, 0.6, 0.8 and 1, which corresponds to 10, 20, 30, 40, and 50 percent of parameter range respectively. The IVARS value corresponded with each scale is shown with an index. For instance, IVARS$_{50}$ represents the IVARS value associated with 0 to 50% scale range ($H_i = 1$). The maximum meaningful value of H_i is equal to one, since IVARS is only meaningful when H_i is smaller than half of the parameter range. In this thesis, since global optimization is performed, global sensitivity analysis metrics is of interest. Hence, in this thesis, instead of directional variogram, integrated variogram or IVARS is considered to represent parameter sensitivity, as it is a summation of different directional variogram values across parameter space.

One of the main features of VARS that makes it an efficient GSA toolbox is the novel Star-Based (STAR) sampling developed by Razavi and Gupta (2016b) and shown in Figure 4. STAR sampling has four major steps. Step 1 is identifying the resolution ($\mathit{\Delta} h$) at which the sampling is performed. Δh is in fact the smallest value of h that is defined by the user. In step 2, the main sample points (called "star centers") are randomly generated across the parameter space by the LH design. The LH method is used to properly cover the entire parameter space. In the next step, from each star center and in the direction of each parameter, new samples with equal distance of $\mathit{\Delta} h$ are generated to cover the entire parameter range. These new points are called cross sections. Each star center and its cross sections form one star. The total number of points generated around each star centers is equal to $n(\left(\frac{1}{\Delta h}\right) - 1)$, where n is the number of dimensions (parameters). In the final stage, for each parameter, all the pairs of points with h value of $\mathit{\Delta} h, 2\mathit{\Delta} h, \ldots, 5\mathit{\Delta} h$ are selected to calculate the corresponding directional variogram and other sensitivity metrics of VARS. The total number of model evaluations (Ti) for each dimension when h is equal to $\mathit{\Delta} h$, is calculated with the following formula, where m is the number of stars.

$$T_i = m(\left(\frac{1}{\Delta h}\right) - 1) \qquad \text{Equation 17}$$

Another characteristic of VARS that increases its efficiency compared to other well-known GSA methods is in fact calculating sensitivity results at pairs of points. This feature results in high number of variogram calculations for each parameter with low number of model evaluations. For instance, if 1000 points are sampled for each parameter, the number of pairs is equal to 499,500. However, when the number of samples are doubled (i.e. 2000 points) the number of pairs are fourfold (1,999,000). Thus, the directional variogram of model output for each parameter is calculated for 1,999,000 times, while the model is only evaluated for 2000 times (Razavi and Gupta, 2016).

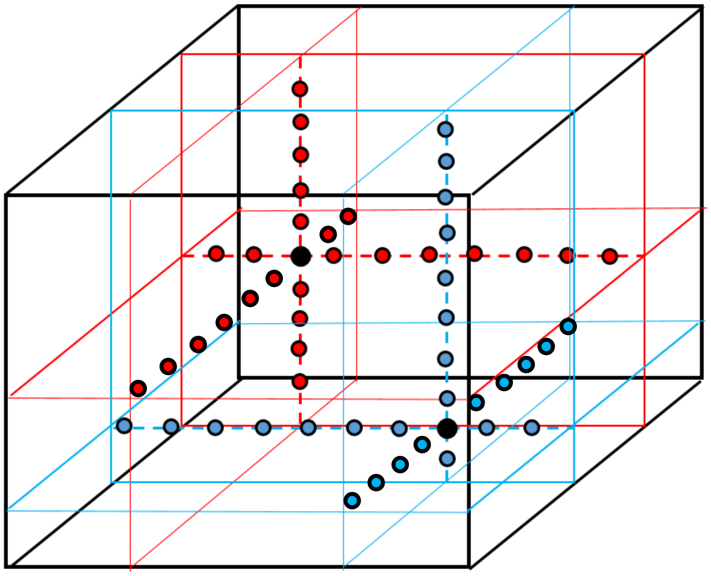


Figure 4. The Star-Based sampling of VARS with two different Stars for a model with three parameters. The two black markers show the star centers generated with LH sampling, and colored markers around each star center are the cross section points (Razavi and Gupta, 2016b).

3.2 Sensitivity Analysis in Model Calibration

The large computational demand due to overparameterization, high dimensional parameter space, and parameter identifiability are the main issues with calibration of modern distributed hydrological models (Box and Jenkins, 1976; Beven, 1989; Carpenter et al., 2001; Tang et al., 2007). One way to reduce the excessive computational cost of calibration of these models is performing a sensitivity analysis before calibration to identify and remove non-sensitive parameters from the calibration process (Tang et al., 2007; Wagener and Kollat, 2007). The main reason for this parameter reduction is that non-sensitive parameters do not have a significant influence on model response (Van Griensven et al., 2006). In addition, if all parameters are carried over to the calibration stage, the calibration becomes significantly complex and time consuming (Muleta and Nicklow, 2005).

Sensitivity analysis has been widely used in hydrological model calibration studies. As an illustration, Holvoet et al. (2005) utilized sensitivity analysis to reduce the number of parameters in calibration of SWAT hydrological model. The streamflow was modeled in the Nil watershed located in Brussel, Belgium. To acquire sensitivity of model parameters, the LH-OAT sensitivity analysis was performed, and six parameters out of 27 were identified as the most sensitive. Afterwards, the manual calibration was performed with considering only the most sensitive parameters. The optimum solution found, generated NSE value equal to 0.53 for streamflow simulation. In addition, White and Chaubey (2005) used the SWAT model to simulate streamflow, sediment yield and total phosphorous (TP) yield in Beaver watershed located in Arkansas, USA. A simple derivative-based sensitivity analysis was used to reduce the total calibration parameters from 51 to 28 most sensitive parameters. After calibrating over a three-year period, it was reported that the model was able to generate results that properly matches the observed data. The R^2 value for streamflow, sediment yield and TP yield was reported equal to 0.89, 0.85 and 0.82 respectively. The high NSE values illustrates that SWAT was able to properly simulate the historical records in the basin.

In another study, Gholami et al. (2016) modeled streamflow using SWAT in the Talar Watershed located in Mazandaran, Iran. A four-year calibration was performed using SUFI-2 algorithm. Parameter sensitivity analysis is utilized to simplify the calibration process by reducing the calibration parameters from 21 to 8 important parameters. Both NSE and R^2 values were reported as 0.93, which represent a very good match between observed and simulated streamflow data. Moreover, Werkhoven et al. (2009) performed sensitivity analysis (Sobol method) using

Runoff Coefficient Error (ROCE) to define the reduced parameter sets for calibration of SAC-SMA model. In addition, four sensitivity threshold values of 0.05, 0.1, 0.2 and 0.3 were utilized to identify the reduced parameter sets. The model was applied on four watersheds (Amite, Spring, Guadalupe and East Fork White) located across USA. For Amite watershed in Louisiana, it was reported that for all the sensitivity thresholds except the highest one (0.3), the performance of SAC-SMA was good and the ROCE value was near-zero. Nevertheless, when considering the 0.3 threshold (reducing parameters to two) the model performance degraded. For the Guadalupe watershed in Texas, the best and worst model performance was reported at the threshold of 0.1 (seven calibration parameters) and 0.2 (four calibration parameters) respectively. In Spring watershed located in Missouri, the best calibration results (ROCE equal to zero) occurred at sensitivity threshold of 0.05 (12 calibration parameters). However, as the threshold value was increased (i.e. lower number of calibration parameters), the performance of SAC-SMA degraded. The last watershed considered in their study was East Fork White located in Indiana, USA. For this watershed they reported that the best calibration result (ROCE near zero) occured when the full parameter set is considered (14 calibration parameters), and for all the reduced parameter sets, ROCE was higher than 0.03. Werkhoven et al. (2009) showed that there is no unique sensitivity threshold value, which can be beneficial in identifying the proper parameter set in model calibration. Therefore, the main problem with the sensitivity-based reduction of parameters, is selecting the appropriate value for sensitivity threshold, which can affect the performance of calibrated models.

To address the issue of sensitivity threshold in sensitivity-aided model calibration, sensitivity analysis should be performed to decrease the computational demand of model calibration without only relying on sensitivity threshold to identify the parameter sets. As an illustration, Chu et al. (2015) performed a parameter reduction using sensitivity information to generate simplified optimization problems. The results of these simplified problems were utilized as initial solutions for solving the reservoir operation problem using ε-NSGAII (epsilon dominance non-dominated sorting genetic algorithm – II). The Dahoufang reservoir operation problem (Liaoning, China) was formulated as a minimization of industrial and agricultural shortage indices. It was illustrated that the sensitivity-informed optimization method dramatically increased the search efficiency of the optimization algorithm in terms of performance metrics compare to using only ε-NSGAII to solve the reservoir operation problem. Moreover, this approach significantly reduced the required computational demand by 51%, and by considering the sensitivity analysis budget, the computational demand was reduced by 47%. In another study, Yang and Becerik-Gerber (2015) performed sensitivity analysis to facilitate the calibration of a building energy simulation model. The Morris approach was selected for sensitivity analysis in order to identify the influential parameters, and to prioritize these parameters in the calibration process. To perform the calibration they have used MBE (non-dimensional bias measure) and CVRMSE (coefficient of variation of root mean squared error) as the objective functions to minimize the error between simulated and measured temperature data. Based on the results, it was reported that robustness of the calibrated model was significantly increased compare to previous related studies. The maximum MBE and CVRMSE was found as 0.085 and 0.135 respectively, while the allowed range for MBE and CVRMSE was reported as ± 5 and ± 20. These values for MBE and CVRMSE shows a good fit between simulated and observed data.

In this thesis, the three sensitivity analysis methods discussed in this chapter: Morris, Sobol, and VARS are utilized individually to inform the optimization search of the Dynamically Dimensioned Search (DDS) optimization algorithm (Tolson and Shoemaker, 2007).

4 Methodology

4.1 Sensitivity-Based Global Optimization

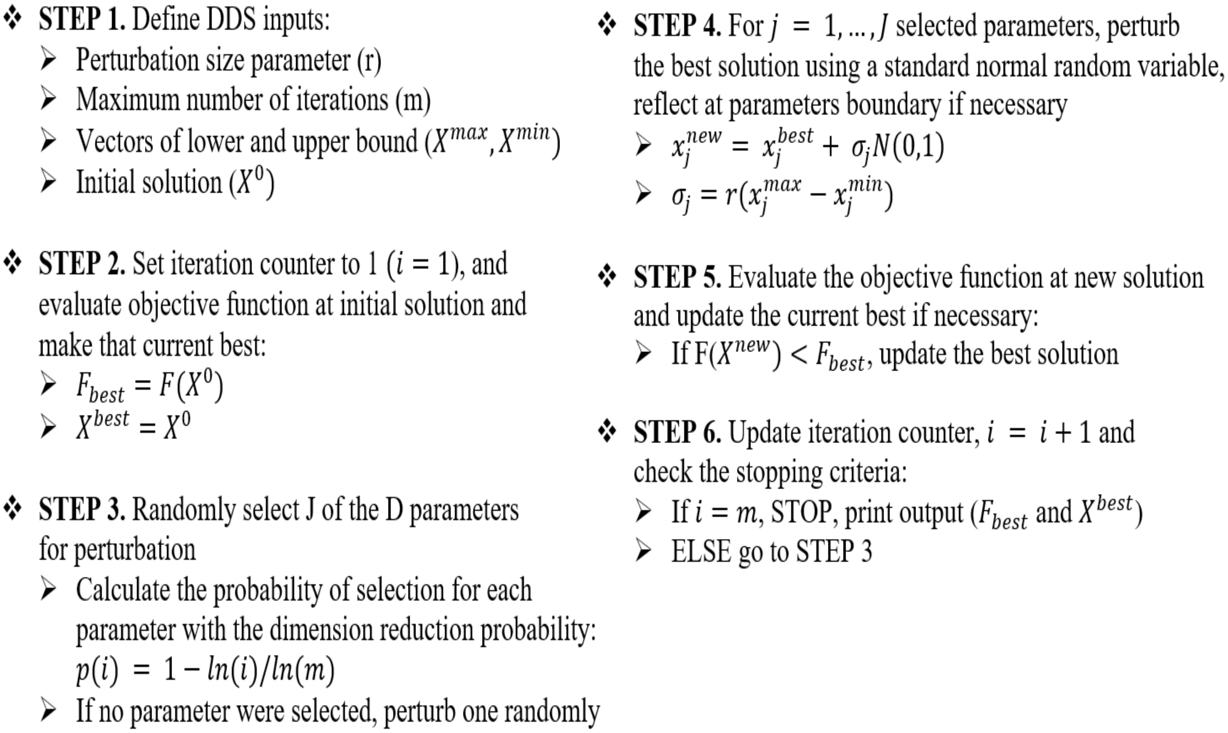
In this chapter, a methodology that includes GSA in global optimization for the purpose of hydrological model calibration is introduced. This approach is intended to increase the computational efficiency of complex hydrological model calibration, by increasing the focus of optimization on sensitive parameters. In this methodology, the Single-objective Dynamically Dimensioned Search (DDS) global optimization algorithm (Tolson and Shoemaker, 2007) performs the optimization part. Furthermore, Morris, Sobol and VARS methods are utilized for sensitivity analysis. Two different implementations of the proposed methodology have been developed and tested in this thesis. In the first approach, which is referred to as GSA↔DDS, the Morris and Sobol methods are coupled with DDS in an interactive approach to feedback to each other during the optimization. In the second approach referred to as GSA→DDS, the initial sensitivity indices are not updated by means of solutions generated during the optimization. In the final proposed approach referred to as VARS→DDS VARS is used to guide DDS. Unlike GSA→DDS, in order to increase the consistency between VARS and DDS, in VARS→DDS, as optimization moves forward, the sensitivity measures are modified from global to local indices, and the VARS is only performed once.

In all the versions of proposed methodology, the first order sensitivity index is utilized, because the sensitivity information provided by this index is enough to guide the optimization process. Thus, parameter interaction and higher order sensitivity index are not considered.

4.2 Dynamically Dimensioned Search Global Optimization

DDS is a single-objective stochastic and heuristic global optimization algorithm that is designed to efficiently calibrate hydrological models with large parameter space (Tolson and Shoemaker, 2007). To illustrate this, Tolson and Shoemaker (2007) compared DDS with SCE-UA algorithm, and it was reported that DDS outperforms SCE-UA algorithm when a large number of parameters needs to be calibrated. As DDS progresses in the optimization, it dynamically and probabilistically reduces the number of parameters for perturbation to a point that it selects only one parameter in each iteration, on average. Therefore, DDS dynamically changes its behavior from a global search to a local search at the end of optimization, This key feature of DDS makes it an efficient optimization algorithm that adjusts its global and local search behavior to the user's computational budget and reduces the number of model evaluations required for finding good quality solution (Tolson and Shoemaker, 2007; Arsenault et al., 2013).

DDS is selected for this thesis because it is claimed to be an efficient tool for automatic model calibration. As pointed out by Tolson et al. (2009), DDS is computationally efficient mainly because it has one algorithmic parameter (i.e. perturbation size) with a robust default value of 0.2. Moreover, Arsenault et al. (2013) made a comparison between ten different optimization algorithms, including DDS, for the calibration of three hydrological models over ten different basins. As a result, it was found that in almost all of the 40 different model calibrations, DDS found better NSE value compare to other methods, especially in models with a large number of parameters.



```
❖ STEP 1. Define DDS inputs:
    ➢ Perturbation size parameter (r)
    ➢ Maximum number of iterations (m)
    ➢ Vectors of lower and upper bound ($X^{max}$,$X^{min}$)
    ➢ Initial solution ($X^0$)

❖ STEP 2. Set iteration counter to 1 ($i = 1$), and
  evaluate objective function at initial solution and
  make that current best:
    ➢ $F_{best} = F(X^0)$
    ➢ $X^{best} = X^0$

❖ STEP 3. Randomly select J of the D parameters
  for perturbation
    ➢ Calculate the probability of selection for each
      parameter with the dimension reduction probability:
      $p(i) = 1 - ln(i)/ln(m)$
    ➢ If no parameter were selected, perturb one randomly

❖ STEP 4. For $j = 1, \dots, J$ selected parameters, perturb
  the best solution using a standard normal random variable,
  reflect at parameters boundary if necessary
    ➢ $x_j^{new} = x_j^{best} + \sigma_j N(0,1)$
    ➢ $\sigma_j = r(x_j^{max} - x_j^{min})$

❖ STEP 5. Evaluate the objective function at new solution
  and update the current best if necessary:
    ➢ If F($X^{new}$) < $F_{best}$, update the best solution

❖ STEP 6. Update iteration counter, $i = i + 1$ and
  check the stopping criteria:
    ➢ If $i = m$, STOP, print output ($F_{best}$ and $X^{best}$)
    ➢ ELSE go to STEP 3
```

Figure 5. The brief systematic pseudo code of DDS algorithm (Tolson and Shoemaker, 2007).

The first step of the DDS algorithm is defining the inputs, which includes the perturbation size (r), maximum number of function evaluations (M), decision variables (parameters) boundary and an initial solution, which is optional, Figure 5. If the user does not define an initial solution, DDS randomly generates initial solutions with the budget of the larger value of 5 solutions or 0.5 percent of total number of solution evaluations (step 2 in Figure 5). The dynamic reduction of dimensions (number of parameters) for perturbation is performed by a uniform probability distribution, step 3 in Figure 4. This probability is a function of the iteration count, thus, it identifies the number of perturbed parameters in each iteration. In step 4, in order to perturb each parameter, the parameter range is multiplied by r and then added to the current value to generate a new value for each parameter of the current best solution. DDS is designed with two different approaches to cope with the parameter values outside the specified range, and each method has a 50% chance to be performed. The first method is called reflection in which, the minimum and maximum values act as a reference point to reflect the perturbed value into the specified range. This reflecting boundary feature of DDS allows parameter values to reach to the minimum or maximum more easily in comparison to ordinary resampling approaches (Tolson and Shoemaker, 2007). However, when parameter values are exactly equal to the boundary, it is difficult to detect them. Therefore, the second approach is introduced that is called absorption in which, parameter values are equal to boundary values (Tolson and Asadzadeh, 2009). As claimed by its developers, instead of finding the exact global optimum point, the main goal of DDS is to converge to the region of the global optimum. Thus, if a better solution is of interest, a local optimization search can be performed from the best solution founded by DDS.

4.3 Sensitivity Analysis for Everybody (SAFE) toolbox

In this thesis, the Morris and Sobol GSA methods are implemented through the well-developed Sensitivity Analysis For Everybody (SAFE) toolbox by Pianosi et al. (2014). Figure 6 illustrates that the SAFE toolbox analyzes sensitivity through three major steps: (i) sampling the input parameter space; (ii) evaluating the model with the provided sample points to create output samples; (iii) computing sensitivity indices by post processing the input and output samples.

SAFE has various sampling methodologies for sampling the input space, including OAT method, LH sampling, all factors at a time (AAT) and a combination of different methods such as, LH-OAT and LH-AAT. In this thesis, in order to increase the efficiency of sampling procedure, the LH-OAT sampling method is used with Morris GSA, and LH-AAT method is used with Sobol GSA. In the post processing step, SAFE is able to calculate sensitivity indices through different methods. These are, Morris method, regional sensitivity analysis (Spear and Hornberger, 1980; Wagener and Kollat, 2007), variance-based sensitivity analysis (VBSA or Sobol method), the fourier amplitude sensitivity test (FAST) by Cukier et al. (1973), dynamic identifiability analysis

(DYNIA) by Wagener et al. (2003) and a novel density-based sensitivity method (PAWN) by Pianosi and Wagener (2015).

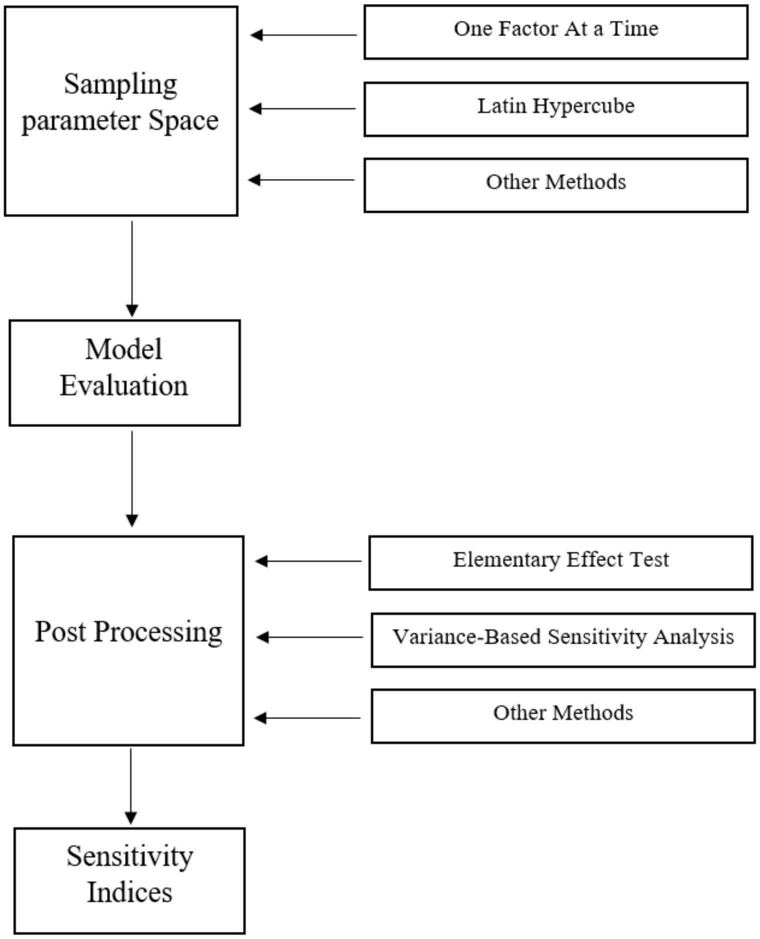


Figure 6. The organization of SAFE toolbox (Pianosi et al., 2015).

Among the different GSA methods provided by the SAFE toolbox, the main reason that the Morris approach is selected is that it is an efficient GSA approach that can fairly compute sensitivity index for a large number of parameters even if the computational budget is very limited (Salteli et al., 2008; Yang and Becerik-Gerber, 2015). Comparison between the results of Morris and Sobol methods showed that the Morris method could properly identify the sensitive parameters but with lower accuracy in terms of sensitivity index compared to Sobol (Campolongo and Saltelli, 1997). Thus, in order to implement the proposed methodology with more robust results of sensitivity and to verify the results of Morris approach, Sobol method is also considered.

4.4 GSA↔DDS

GSA↔DDS is an interactive method that continuously couples DDS with a GSA to update the results of both sensitivity and optimization throughout the optimization process. GSA↔DDS has two major parts. The first part is an initial sensitivity analysis that provides a primary understanding of parameters sensitivity. The sensitivity indices are used to guide the parameter selection of DDS. In this way, the original random parameter selection process of DDS is modified to the roulette wheel procedure (Goldberg, 1989) to select more sensitive parameters, more frequently, while giving a slight chance of perturbation to low sensitive parameters.

Although complex hydrological models have a large set of parameters, a relatively small group of parameters, which are referred to as the most sensitive, mainly controls the model output. Therefore, in GSA↔DDS is expected to finds good quality solutions in a shorter time relative to DDS, by perturbing more sensitive parameters more frequently. Moreover, the sensitivity information of parameters is useful in recalibrating the model when new forcing data is available. Hence, users can recalibrate only more sensitive parameters to increase the efficiency and effectiveness of the model calibration, especially when a limited budget is available.

In the second part of GSA↔DDS, the initial sensitivity indices are updated using the solutions generated by the optimization algorithm. This is the key feature that separates GSA↔DDS from similar methodologies. In previous approaches that combined optimization with sensitivity analysis in hydrological model calibration, usually sensitivity analysis is performed initially with a limited budget to select and only perturb influential parameters in optimization process (Holvoet et al., 2005; Muleta and Nicklow, 2005; White and Chaubey, 2005; Werkhoven et al., 2009; Lu et al., 2015; Gholami et al., 2016). However, performing sensitivity analysis with a low budget can mislead the optimization search. On the other hand, over-emphasizing on sensitivity can weaken the performance of optimization algorithms to find the best solution as the budget of optimization is decreased. One way to tackle this issue is to perform GSA interactively with optimization throughout the calibration process to update the sensitivity indices regularly. Updating the sensitivity indices can reveal the true sensitivity importance of parameters to guide the calibration properly. For instance, initial GSA can identify certain parameters as most sensitive, however, updating sensitivity indices of parameters may results in a different group of most sensitive parameters. In addition, the interaction between DDS and GSA has the benefit of saving computational budget, since the solutions generated through optimization is used to update the results of sensitivity analysis and vice versa. The GSA part in GSA↔DDS is performed using Morris or Sobol methods. The following sections explain the procedure of coupling DDS with the two GSA approaches.

4.4.1Morris↔DDS

In Morris↔DDS, after performing initial Morris, the sensitivity indices are updated in each iteration of automatic calibration by recalculating the numerical derivatives using the new generated solution. After identifying the number of parameters to be selected (J in the DDS algorithm Figure 7) by the original dimension reduction probability of DDS, a parameter is selected by the roulette-wheel process to give priority to more sensitive parameters and perturbed as shown in Figure 7 in a slightly different approach compared to original DDS. A new solution is generated by considering the perturbed value of the parameter while other parameters are kept at their current best solution. The sensitivity index of the selected parameter, and if required, the current best solution is updated.

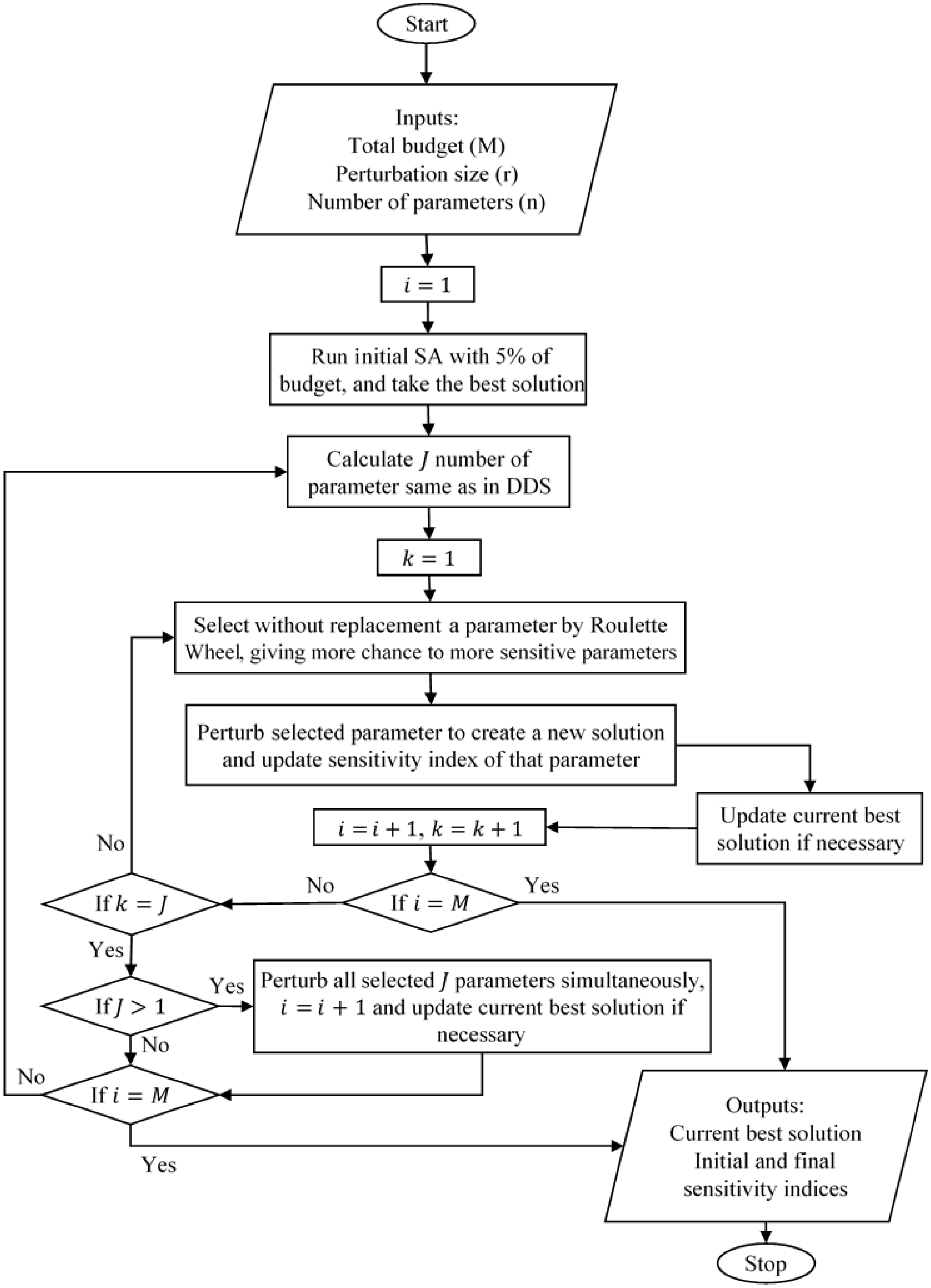

Start
Inputs:
Total budget (M)
Perturbation size (r)
Number of parameters (n)
$i = 1$
Run initial SA with 5% of budget, and take the best solution
Calculate J number of parameter same as in DDS
$k = 1$
Select without replacement a parameter by Roulette Wheel, giving more chance to more sensitive parameters
Perturb selected parameter to create a new solution and update sensitivity index of that parameter
Update current best solution if necessary
$i = i + 1, k = k + 1$
If $i = M$
No
Yes
If $k = J$
No
Yes
If $J > 1$
Yes
No
Perturb all selected J parameters simultaneously, $i = i + 1$ and update current best solution if necessary
If $i = M$
No
Yes
Outputs:
Current best solution
Initial and final sensitivity indices
Stop

Figure 7. The flowchart of Morris↔DDS method

As shown in Figure 8, for calibrating a model with two parameters, in order to recalculate sensitivity indices for each parameter, two points in the solution space are required: 1) a reference point, which is always the initial or current best solution, for example the red point in Figure 8, and 2) a new point. Assuming that the initial solution is called point 1 in the solution space, if parameter 2 of the model is selected for perturbation, it will be perturbed from point 1 to generate point 2, and to update the sensitivity index of parameter 2 using point 1 and point 2. At this stage, if point 2 outperforms point 1, the current best solution and the reference point will be updated to point 2. In addition, if parameter 1 has to be perturbed as well, it will be perturbed from point 2 to produce point 3, and to recalculate the sensitivity index of parameter 1 by considering point 2 and point 3. However, as Figure 8-b shows, if point 1 remains the current best solution, parameter 1 is perturbed from point 1 to generate point 3, and to update the sensitivity index of parameter 1 using point 1 and point 3. At the end of each iteration, considering the entire perturbed parameters simultaneously, a new additional solution is created, and the objective function is evaluated at this solution to update the current best solution if necessary.

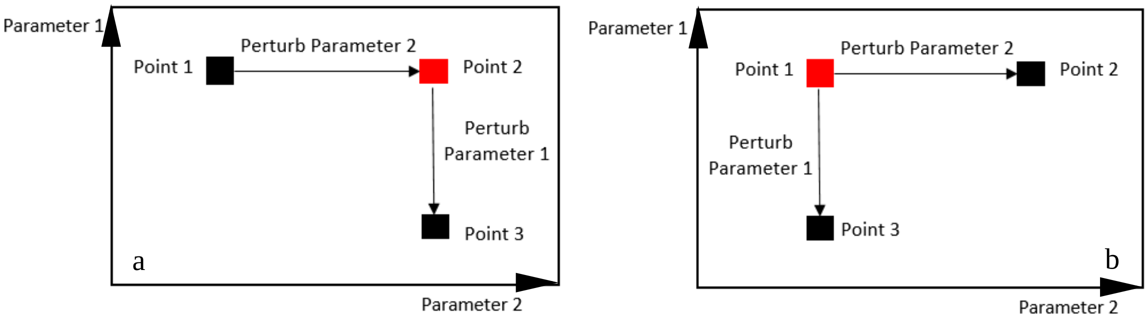


Figure 8. The process of updating sensitivity indices in Morris↔DDS for a model with two parameters. (a) Point 2 is the current best solution and reference point. (b) Point 1 remains the current best solution

4.4.2Sobol↔DDS

Sobol↔DDS follows a similar procedure as Morris↔DDS. However, instead of the Morris GSA method, the approximation of Sobol GSA by Salteli et al. (2008) is used. In addition, in Sobol↔DDS the sensitivity indices are updated in a different way. As shown in Figure 9, for a model with two parameters, if parameter 2 is selected for perturbation, then it is perturbed from its current best value (black square) twice to create two different solutions. In the next step, one solution (black circle) is added to y_C and one solution (black triangle) is added to y_B solution sets. From the solution added to y_C a new solution (red circle) is created by perturbing all parameters except for parameter 2 to update y_A, see section 3.1.2 for the description of y_A, y_B and y_C associated with A, B, and C input sample matrices. Using the updated solution sets (y_A, y_B and y_{Ci}) and applying Equation 12, the sensitivity index for the parameter two (the selected parameter) is updated. The three new solutions are evaluated to update the current best solution if required.

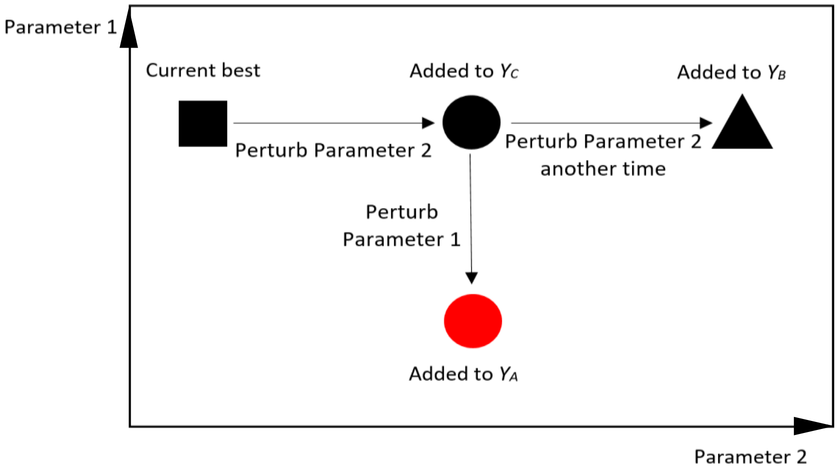


Figure 9. The process of updating sensitivity indices in Sobol↔DDS for a model with two parameters.

4.5 GSA→DDS

Observing the behavior of GSA↔DDS revealed that updating sensitivity results with the DDS generated solutions that did not lead to the expected results (this is further discussed in Chapter 7). Thus, a new approach entitled as GSA→DDS introduced, in which only the initial GSA results are considered in optimization. Three versions of GSA→DDS are considered in this study: Morris→DDS, Sobol→DDS and VARS→DDS. However, in this thesis, only Morris→DDS and Sobol→DDS are referred to as GSA→DDS. In both Morris→DDS and Sobol→DDS approaches, the sensitivity analysis is performed before the optimization process to generate the parameter sensitivity indices. In addition, the solutions derived from GSA are taken as initial solutions for DDS. According to Figure 10, after GSA is performed, the new solutions derived from GSA are evaluated to archive the best one. In the next stage, the parameter sensitivity information is used to guide the parameter selection of DDS by utilizing roulette wheel algorithm to give more chance to more sensitive parameters to be selected and perturbed. Unlike the traditional sensitivity reduction approaches, GSA→DDS does not eliminate the chance for perturbing less sensitive parameters. At the final stage of GSA→DDS, when the calibration budget is spent, the final best solution and initial sensitivity indices are given as the results.

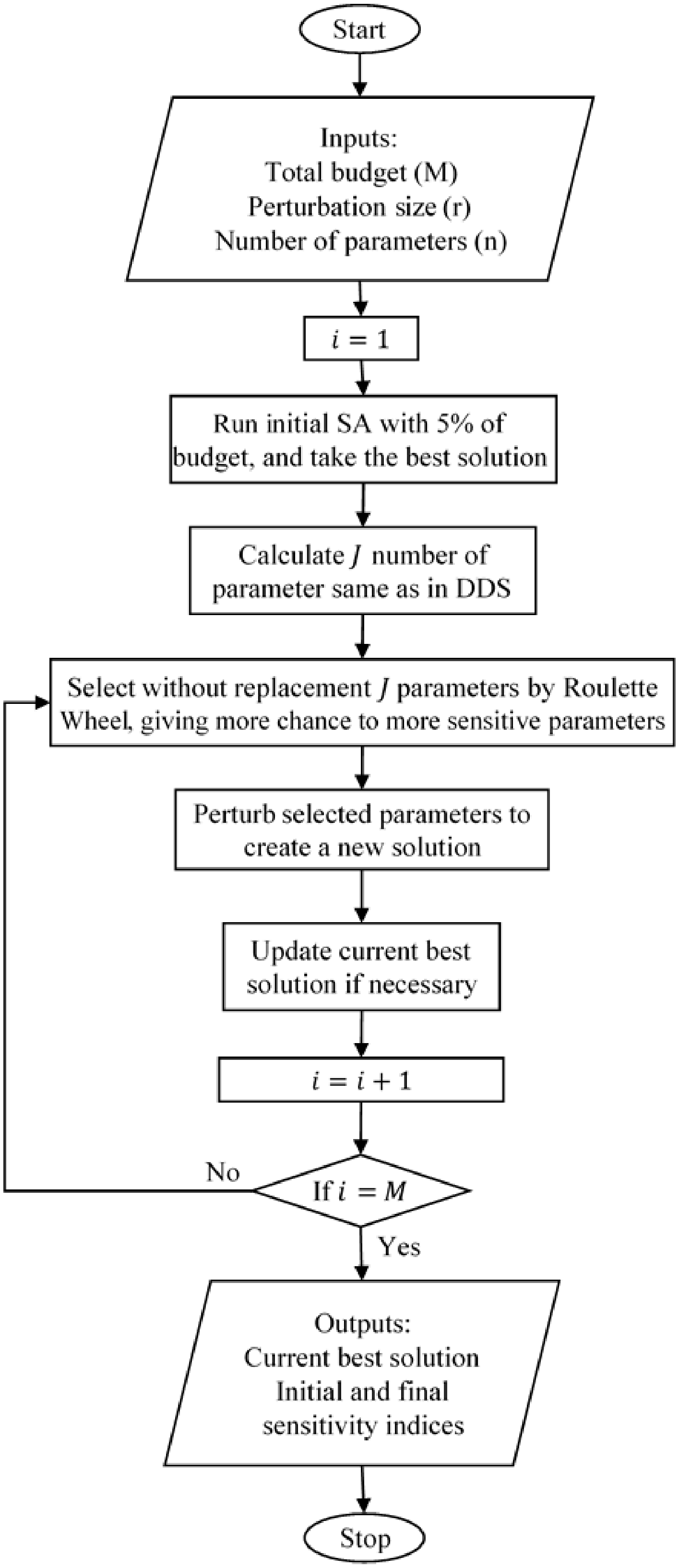

Start
Inputs:
Total budget (M)
Perturbation size (r)
Number of parameters (n)
$i = 1$
Run initial SA with 5% of budget, and take the best solution
Calculate J number of parameter same as in DDS
Select without replacement J parameters by Roulette Wheel, giving more chance to more sensitive parameters
Perturb selected parameters to create a new solution
Update current best solution if necessary
$i = i + 1$
No
If $i = M$
Yes
Outputs:
Current best solution
Initial and final sensitivity indices
Stop

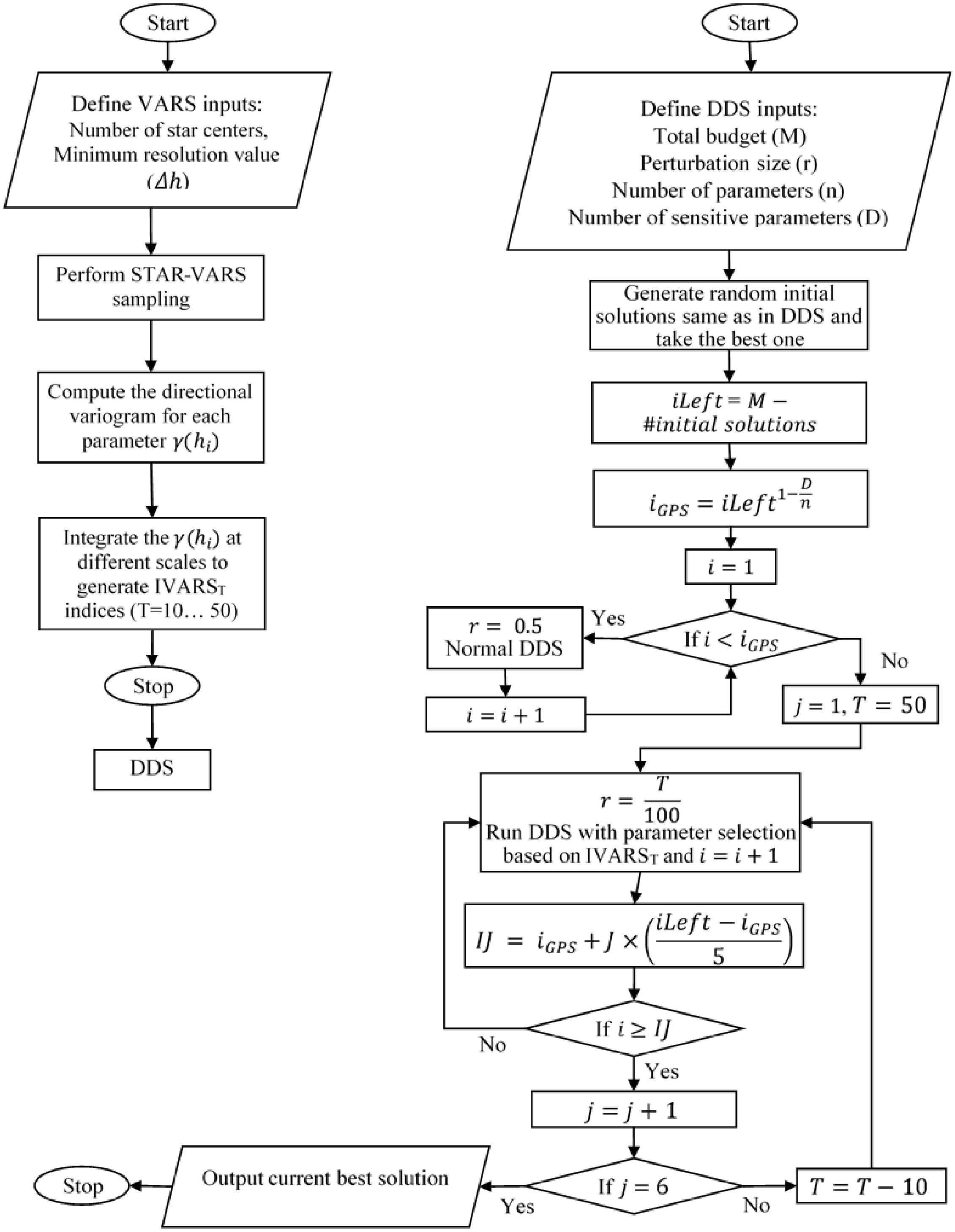
Figure 10. Flowchart of GSA→DDS.

4.6 VARS→DDS

The VARS methodology has some unique characteristics that make it suitable for GSA→DDS. VARS is claimed to be a robust tool for estimating the parameter global sensitivity ranking with a relatively low computational budget, e.g. more than two orders of magnitude fewer solution evaluations compared to Sobol as in Razavi and Gupta (2016). Therefore, one should expect consistent parameter ranking out of multiple independent trials of VARS. This characteristic of VARS is examined in Section 7.3 against Sobol and Morris. The VARS tool is used to calculate IVARS$_{10}$, IVARS$_{20}$, IVARS$_{30,}$ IVARS$_{40,}$ and IVARS$_{50}$ that represent the parameter sensitivity indices respectively at 10%, 20%, 30%, 40% and 50% perturbation sizes. As shown in **Figure 11**, this perturbation size is used to replace the perturbation size of the DDS algorithm in VARS→DDS with a dynamically reducing perturbation size.

As explained in Section 4.2, DDS perturbs all decision variables (parameters of the hydrological model) in the beginning of the search and dynamically reduces the number of perturbed parameters as the optimization progresses toward its finale. By default, the perturbation size (r) is equal to 0.2 and remains constant throughout the optimization. However, in VARS→DDS, the perturbation size is reset to an initial value of 0.5, and parameters are randomly selected without any predefined priority until the average number of perturbed parameters in an iteration becomes equal to the number of most sensitive parameters that are identified by the user based on IVARS$_{50}$. This iteration number is called Guided Parameter Selection iteration (i_{GPS}). The remaining budget ($M - i_{GPS}$ in **Figure 11**) is divided into five equal number of solution evaluations, in each of which an IVARS set of sensitivity indices is used to guide the optimization. IVARS50 indices are used to define the priority of parameters to be selected and perturbed with the perturbation size of $r = 0.5$, in the first 20% of the remaining computational budget right after

i_{GPS}. In the next 20% of the optimization budget, IVARS40 and $r = 0.4$ are used, followed by IVARS30 and $r = 0.3$, then IVARS20 and $r = 0.2$, and finally IVARS10 and $r = 0.1$. This is performed to increase the consistency between the results of VARS and DDS.



Start
Define VARS inputs:
Number of star centers,
Minimum resolution value
(Δh)
Perform STAR-VARS sampling
Compute the directional variogram for each parameter $\gamma(h_i)$
Integrate the $\gamma(h_i)$ at different scales to generate IVARS$_T$ indices (T=10… 50)
Stop
DDS
Start
Define DDS inputs:
Total budget (M)
Perturbation size (r)
Number of parameters (n)
Number of sensitive parameters (D)
Generate random initial solutions same as in DDS and take the best one
$iLeft = M - \#initial\ solutions$
$i_{GPS} = iLeft^{1-\frac{D}{n}}$
$i = 1$
If $i < i_{GPS}$
Yes
$r = 0.5$
Normal DDS
$i = i + 1$
No
$j = 1, T = 50$
$r = \frac{T}{100}$
Run DDS with parameter selection based on IVARS$_T$ and $i = i + 1$
$IJ = i_{GPS} + J \times \left(\frac{iLeft - i_{GPS}}{5}\right)$
If $i \geq IJ$
No
Yes
$j = j + 1$
If $j = 6$
Yes
No
$T = T - 10$
Output current best solution
Stop

Figure 11. Flowchart of VARS→DDS method.

As discussed in Section 4.2, DDS is a stochastic optimization algorithm and has to be run multiple times independently to be able to statistically estimate its performance. However, since VARS is a robust tool for estimating the parameter ranking, it is proposed to run VARS once for each case study at each computational budget and use the same IVARS sensitivity indices for as many trials of VARS→DDS as required.

5 Case Studies

Two mathematical test problems Griewank (Griewank, 1974) and Rastrigin (Rastrigin, 1981) that are well-known for being challenging to solve by global optimization algorithms are used to evaluate and compare the performance of sensitivity-informed DDS methods with original DDS algorithm. In addition, to evaluate the performance of proposed methodology on actual model calibration problems, three hydrological models with different levels of complexity namely, SAC-SMA, SWAT and WATFLOOD are used in this thesis. The lumped SAC-SMA model with 13 calibration parameters is selected to measure the ability of sensitivity-informed DDS on calibrating simple hydrological models with low number of calibration parameters. However, as the benefit of sensitivity-informed DDS is best observed when applied to high-dimensional hydrological model calibrations, the semi-distributed and fully distributed SWAT and WATFLOOD models with respectively 23 and 37 calibration parameters are considered in this thesis.

5.1 Mathematical Test Problems

The Griewank and Rastrigin functions, as shown in Table 1, are continuous unimodal optimization test functions that are scalable in their number of decision variables. These functions are designed to challenge the ability of optimization algorithms in terms of finding the global optimum solution, as they have several local optimum points (Tolson and Shoemaker, 2007). Moreover, Griewank and Rastrigin functions have been commonly used in the literature to measure the optimization algorithms performance (Duan et al., 1993; Jung et al., 2006; Tolson and Shoemaker, 2007; Han et al., 2010). Therefore, they are suitable to test the proposed methodology. Twenty parameters were selected for both functions to resemble the number of parameters that hydrological model calibration problems of this thesis have. The functions range is selected according to Tolson and Shoemaker (2007).

In Figure 12 and Figure 13, various local minimum points of the Griewank and Rastrigin functions with two decision variables are visualized.

Table 1. Characteristics of the mathematical test functions.

Function Name	**Equation**	**Number of Parameters (n)**	**Range**	**Global Minimum**
Griewank (1974)	$f(x) = \sum_{i=1}^{n}\left(\frac{x_i^2}{4000}\right) - \prod_{i=1}^{n}\cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	20	$[-500, 700]^{20}$	0 at $x_i = 0$
Rastrigin (1981)	$f(x) = \sum_{i=1}^{n}[x_i^2 - \cos(2\pi x_i)]$	20	$[-2, 2]^{20}$	-20 at $x_i = 0$

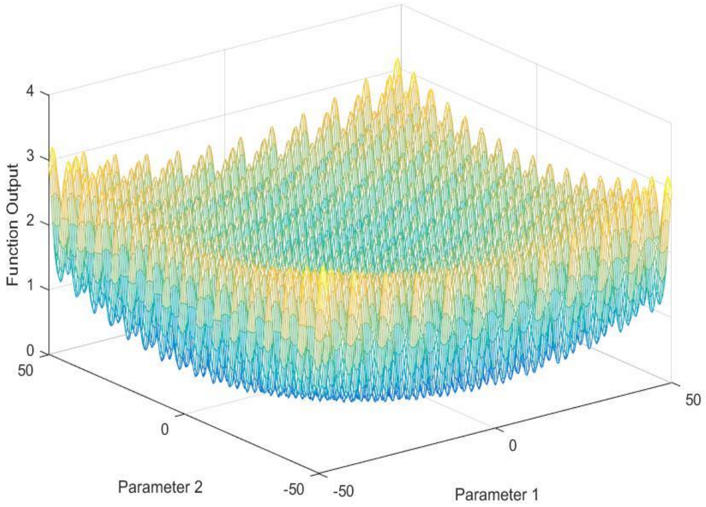


Figure 12. The Griewank function with two decision variables (parameters)

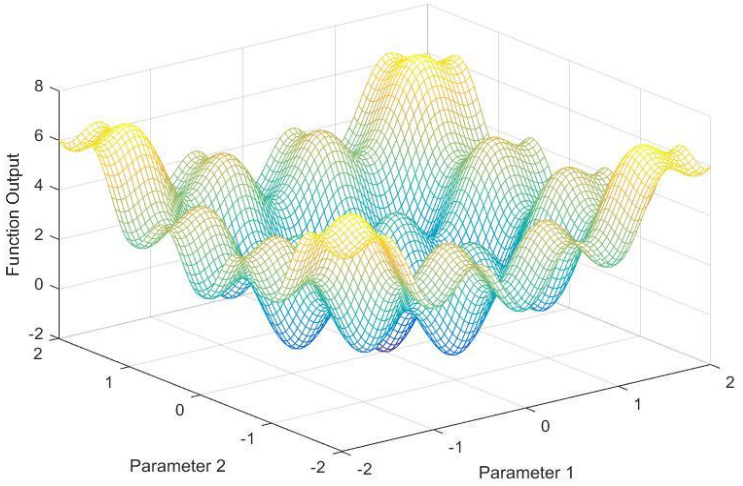


Figure 13. The Rastrigin function with two decision variables (parameters)

5.2 SAC-SMA Leaf River

The Sacramento Soil Moisture Accounting model (SAC-SMA) is a lumped and conceptual rainfall-runoff model developed by Sorooshian et al. (1993). SAC-SMA takes precipitation and potential evapotranspiration as inputs and simulates the streamflow of the modeled watershed. The Leaf River watershed, located north of Collins, Mississippi, United States with the area of 1944 km^2 has been modeled into SAC-SMA by Sorooshian et al. (1993) with 13 calibration parameters (Table 2). This model has lower complexity and lower number of parameters compared to distributed models and is computationally very efficient as a single run of SAC-SMA in Leaf River basin took around 0.093 second using Intel® Core™ i5-6300HQ, with a 2.30 GHz CPU and 8 GB of RAM. Thus, the total computational time required for calibrating SAC-SMA model with 10 trials of 10000 model evaluations was equal to 155 minutes. SAC-SMA has been used in a variety of model calibration studies (Sorooshian et al., 1993; Tang et al., 2006; Asadzadeh et al., 2014). In addition, this model is the main rainfall-runoff model used by River Forecast Center in the United States (Werkhoven et al., 2009). Hence, SAC-SMA Leaf River is a suitable and simple model calibration case study to test the proposed methodology.

Minimizing the $1 - NSE$ value is considered as the objective for calibrating model parameters. Equation 18 is showing the Nash Sutcliff Efficiency (NSE) formula. When the NSE value is equals to 1, it represent the best fit between observed and simulated data. However, NSE equal to 1 represent the ideal situation in which there is no uncertainty in model structures, input data, and observations. Therefore, in model calibration problems, NSE equal to 1 does not necessarily denote that an optimal value for model parameters is found.

$$NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - S_i)^2}{\sum_{i=1}^{n}(O_i - \frac{\sum_{i=1}^{n} O_i}{n})^2} \qquad \text{Equation 18}$$

In Equation 18, o_i is the observed streamflow data, and s_i is the simulated streamflow in day i over the calibration period n. According to Tang et al. (2005), to warm up the model, the period from July 28, 1952 to September 30, 1952 is ignored in the calculation of the objective function value. Moreover, the calibration period is from October 1952 to September 1954. According to Asadzadeh 2012, the best known objective function value for this case study is $1 - NSE = 0.076$.

Table 2. Calibration parameter description and range for the SAC-SMA model of Leaf River

Parameter	Description	Range
UZTWM	Max. capacity of the upper zone tension water storage, mm	[1.00 150.0]
UZFWM	Max. capacity of the upper zone free water storage, mm	[1.00 150.0]
LZTWM	Max. capacity of the lower zone tension water storage, mm	[1.00 500.0]
LZFPM	Max. capacity of the lower zone free water primary storage, mm	[1.00 1000.0]
LZFSM	Max. capacity of the lower zone free water supplemental storage, mm	[1.00 1000.0]
ADIMP	Additional impervious area, decimal fraction	[0.00 0.40]
UZK	Upper zone free water lateral depletion rate, day^{-1}	[0.0001 0.025]
LZPK	lower zone primary free water depletion rate, day-1	[0.010 0.250]
LZSK	lower zone supplemental free water depletion rate, day-1	[0.00 0.10]
PCTIM	Impervious fraction of the watershed area, decimal fraction	[1.00 250.0]
ZPERC	Max. percolation rate, dimensionless	[1.00 150.0]
REXP	Exponent of the percolation equation, dimensionless	[0.00 5.0]
PFREE	Fraction of water percolating from upper zone which goes directly to lower zone free water storage, decimal fraction	[0.00 0.10]

5.3 SWAT2009 Rouge River Watershed

SWAT is a physically-based and semi-distributed watershed scale model that has been under development and improvement for more than three decades. The main inputs of SWAT include weather information, soil properties, topography and land use data as inputs (Neitsch et al., 2011). SWAT has more calibration parameters and requires more computational demand than lumped models such as, SAC-SMA. For instance, White and Chaubey (2005) calibrated 28 parameters of a SWAT model. Furthermore, Gholami et al. (2016) developed a SWAT model with 21 parameters. However, SWAT can continuously and more accurately simulate watershed responses including streamflow and different water quality constituents throughout the basin (Khatun et al., 2018). Compared to distributed models, SWAT is computationally efficient (Lu et al., 2015). Thus, it can be used for modeling large basins including European continental watershed (Abbaspour et al., 2015), West African watershed (Schuol and Abbaspour, 2006), and Scandinavian and Iberian basins (Malagò et al., 2015). Furthermore, SWAT is one of the most common hydrological models used in the literature to study streamflow and water quality indices (Muleta and Nicklow, 2005; Tolson and Shoemaker, 2007; Teshager et al., 2016).

The SWAT version of 2009 model of the Rouge River watershed developed by Asadzadeh et al. (2015) is utilized in this thesis to evaluate the performance of sensitivity-informed DDS in calibrating more complex hydrological models in comparison with original DDS algorithm. The Rouge River watershed is located in Ontario , Canada, with an area of 336 km^2. The watershed consists of two primary branches: the main Rouge River and the little Rouge River with the corresponding area of 222 km^2 and 114 km^2 respectively. The River is originated from northern oak ridges moraine towards the Lake Ontario. The elevation in the watershed ranges from 370 m to 64 m above sea level. The basin experience the continental climate moderated by the Great Lakes. As reported by the Toronto and Region Conservation Authority (TRCA), the land cover types in the Rouge River basin are consist of 40% rural and agricultural uses 35% urban, 24% natural and 1% waterbodies.

SWAT separates basin into different hydrological response unites (HRUs), which are smaller sub-basins that have similar soil and land use characteristics. This model simulates flow and sediment yield in every sub-basin, and aggregate the results of each HRU. According to Asadzadeh et al. (2015), to calibrate the 23 parameters of SWAT2009 model (Table 3) maximizing the weighted average of NSE coefficient for each basin is considered as the objective with a computational time of 21.7 seconds for a single model run, and 36,150 minutes for 10 trials of 10000 model runs. Asadzadeh et al. (2015) reported the corresponding NSE values for Main Rouge and Little Rouge basins equal to 0.65 and 0.69 respectively. Following Asadzadeh et al. (2015), the four-year calibration period from 2006 to 2009 along with a one year spin up period (2005) is considered. The perturbation method in Table 3 shows whether the original parameter value is replaced by or multiplied by a new value.

Table 3. Calibration parameters of SWAT2009 model of Rouge River (Asadzadeh et al., 2015).

Parameter	Description	Perturbation Method	Range
CN	Soil conservation service run-off curve number	Multiply	[0.75, 1.25]
CNCOEF	Plant ET curve number coefficient	Replace	[0.50, 2.00]
SMFMN (mm H2O/°C-day)	Melt factor for snow on December 21st.	Replace	[1.40, 4.50]
SMFMX (mm H2O/°C-day)	Melt factor for snow on June 21st.	Replace	[1.40, 6.90]
TIMP	Snowpack temperature lag factor.	Replace	[0.01, 1.00]
ESCO	Soil evapotranspiration compensation factor.	Replace	[0.01, 1.00]
EPCO	Plant uptake compensation factor.	Replace	[0.01, 1.00]
SURLAG	Surface runoff lag coefficient.	Replace	[0.10, 24.0]
SOL_AWC (mm H2O/mm soil)	Soil available water capacity.	Multiply	[0.10, 2.0]
SOL_K (mm/hr)	Saturated hydraulic conductivity of soil	Multiply	[0.10, 100.0]
SOL_Z (mm)	Depth from soil surface to bottom of layer	Replace	[0.75, 1.25]
GW_DELAY (days)	Groundwater delay time	Replace	[1.00, 500.0]
GW_REVAP	Groundwater coefficient for calculating the amount of water in shallow aquifer that returns to the root zone.	Replace	[0.02, 0.20]
ALPHA_BF (days)	Base-flow alpha factor.	Replace	[0.10, 1.00]
GWQMN	Threshold depth of water in shallow aquifer for return flow to occur.	Replace	[0.00, 5000]
CH_N2	Manning's "n" value for the main channel	Replace	[0.01, 0.10]
CH_K (mm/hr)	Effective hydraulic conductivity in the main channel alluvium.	Replace	[0.00, 100.0]
SFTMP (°C)	Snowfall temperature.	Replace	[-5.0, 5.0]
SMTMP (°C)	Snow melt base temperature	Replace	[-5.0, 5.0]
RCHRG_DP	Deep aquifer percolation fraction.	Replace	[0.0, 1.0]
REVAPMN (mm H2O)	Threshold depth of water in the shallow aquifer to let the water in shallow aquifer return to the root zone	Replace	[0.0, 500.0]
SNOCOVMX (mm H2O)	Minimum snow water content that corresponds to 100% snow cover	Replace	[5.0, 35.0]
SNO50COV	Fraction of snow volume represented by SNOCOVMX that corresponds to 50% of snow cover.	Replace	[0.01, 0.99]

5.4 WATFLOOD Odei River basin

WATFLOOD is a distributed and physically-based hydrologic model developed by Kouwen (1988). It is used for simulating various aspects of water cycle such as, stream flow in a river basin. The model is designed for flood forecasting and long-term hydrologic simulation using distributed precipitation data from radar or numerical weather models. It divides the basin into similar land cover groups, which are called grouped response units (GRUs). GRUs are the main hydrologic computational units in WATFLOOD, and have similar hydrological conditions (Muhammad et al., 2018). WATFLOOD's main strengths are that it is fast, robust, and requires only temperature and precipitation as input data (Kouwen et al., 1993). Furthermore, due to its fully distributed characteristic, WATFLOOD has a large number of calibration parameters, for instance, 61 parameters in the study by Unduche et al. (2018), and 37 parameters in this thesis. Thus, the calibration of this model is expected to reveal the main advantages of the sensitivity-guided calibration approaches proposed in this thesis.

The WATFLOOD model of the Odei River basin developed by Smith et al. (2016) that simulates the streamflow is recalibrated in this thesis to show the benefit of sensitivity-informed DDS in efficient calibration of hydrological models with large number of parameters. The Odei River has a basin relatively larger than the other cases studied in this thesis, 6110 km^2 located in Manitoba, Canada (Smith et al., 2016). This basin has a sub-humid and sub-arctic weather with a low temperature and precipitation rate (550 mm/year) (Smith et al. 2016). About 1/3 of the total annual precipitation is in the form of snowfall. Evapotranspiration accounts for most of the loss of precipitation from the basin (360 mm per year). There are various types of land cover in the Odei basin region. However, as recommended by Holmes (2016) for the purpose of calibration the five main types of land cover are considered in this thesis. These are Coniferous, Mixed wood, Shrub, Wetland and Water. The calibration parameters are considered based on Holmes (2016) recommendation (Table 4). The calibration period was from 1982 to 1987, with a one-year spin up period in 1981. Holmes (2016) performed a manual calibration of WATFLOOD Odei River with a reported NSE value equal to 0.60. A single model run for WATFLOOD Odei River took around 37.23 seconds, and with total computational time of 62050 minutes for 10 trials of 10000 model evaluations. This clearly illustrates higher complexity of this model in terms of computational time compared to SAC-SMA and SWAT that required less time to run. However, in terms of number of unique parameters calibrated in this thesis for SWAT and WATFLOOD models, SWAT model is more complex, as each of the 23 parameters of this model represent a different characteristic of the modeled basin for all land classes or HRUs. However, in WATFLOOD, there are only 13 unique parameters that are repeated for each land class, and increased the total number of calibration parameters to 37.

Table 4. Calibration parameters of WATFLOOD model of the Odei River basin.

Parameter	Land Cover/Class	Description	Range
PWR	N.A.	BF power	[1.50 4.00]
COEFF	N.A.	BF coefficient	[1E-07 0.005]
R2N	N.A.	Channel Roughness	[0.0005 0.05]
THETA	N.A.	Porosity	[0.015 0.90]
KCOND	N.A.	Conductivity	[0.10 0.25]
AK	Coniferous	Infiltration	[1.00 20.00]
AK	Mixed wood	Infiltration	[1.00 20.00]
AK	Shrub	Infiltration	[1.00 20.00]
AK	Wetland	Infiltration	[1.00 20.00]
AKFS	Coniferous	Snow Covered Infiltration	[1.00 5.00]
AKFS	Mixed wood	Snow Covered Infiltration	[1.00 5.00]
AKFS	Shrub	Snow Covered Infiltration	[1.00 5.00]
AKFS	Wetland	Snow Covered Infiltration	[1.00 5.00]
RETN	Coniferous	Soil Retention	[20.00 150.00]
RETN	Mixed wood	Soil Retention	[20.00 150.00]
RETN	Shrub	Soil Retention	[10.00 50.00]
RETN	Wetland	Soil Retention	[10.00 50.00]
R3	Coniferous	Overland flow roughness	[1.00 10.00]
R3	Mixed wood	Overland flow roughness	[1.00 10.00]
R3	Shrub	Overland flow roughness	[1.00 10.00]
R3	Wetland	Overland flow roughness	[1.00 25.00]
R3	Water	Overland flow roughness	[1.00 10.00]
REC	Coniferous	Horizontal Conductivity	[0.05 4.00]
REC	Mixed wood	Horizontal Conductivity	[0.05 4.00]
REC	Shrub	Horizontal Conductivity	[0.05 4.00]
REC	Wetland	Horizontal Conductivity	[0.05 4.00]
FM	Coniferous	Snow Melt Rate	[0.05 0.25]
FM	Mixed wood	Snow Melt Rate	[0.05 0.25]
FM	Shrub	Snow Melt Rate	[0.05 0.25]
FM	Wetland	Snow Melt Rate	[0.05 0.25]
FM	Water	Snow Melt Rate	[0.05 0.25]
FPET	Water	PET Coefficient	[0.90 1.10]
SUB	Coniferous	Sublimation Factor	[0.10 1.50]
SUB	Mixed wood	Sublimation Factor	[0.10 0.30]
SUB	Shrub	Sublimation Factor	[0.10 1.50]
SUB	Wetland	Sublimation Factor	[0.005 0.10]
SUB	Water	Sublimation Factor	[0.10 1.50]

6 Numerical Experiments and Results Analysis Approaches

6.1 Numerical Experiments

Global optimization algorithm such as DDS utilize stochastic processes to find the optimal solution. Therefore, their performance varies when the random seed changes. The performance of these algorithms should be evaluated based on multiple independent trials that represent the distribution of their performance rather than a single trial (Matott et al. 2012). A global optimization algorithm is a robust one if it consistently identifies high quality solutions with its multiple independent trials. Due to stochastic nature of the model calibration methods developed in this thesis, to ensure the robustness of the results, optimizations are performed in 10 independent trials with different random seeds, and the distribution of the results are compared.

DDS and the calibration methods developed in this thesis adjust their optimization behavior to the user-specified computational budget. These algorithms are expected to achieve good quality solutions at different orders of number of solution evaluations. Of course, they are expected to achieve better solutions as the number of solution evaluations increases. In this thesis, in order to assess whether the developed algorithms properly adjust their search behavior to the user-specified computational budgets, the performance of proposed algorithms are evaluated at two levels of the number of solution evaluations: a) a relatively limited budget of 1000 and b) a relatively larger computational budget of 10000 solution evaluations The higher budget of 10000 solution evaluation is set based on the computational limitations available for this research.

The GSA↔DDS, GSA→DDS were tested on minimizing the Rastrigin and Griewank test functions with 20 parameters and calibrating the 13-parameter SAC-SMA model of the Leaf River watershed. VARS→DDS was applied to these optimization problems and to more complex model calibration problems: SWAT with 23 and WATFLOOD with 37 parameters. Due to various shortcomings of GSA↔DDS and GSA→DDS methods (explained in chapter 7) they were not tested on SWAT and WATFLOOD models.

In GSA↔DDS and GSA→DDS, 5% of the budget was used for the initial sensitivity analysis. It was observed in some initial optimization trials that are not reported in this thesis that increasing the budget of initial GSA to more than 5% significantly degraded the algorithm performance. On the other hand, implementing GSA with a very limited budgets (e.g. lower than 5 percent) can generate unreliable sensitivity information that can mislead the calibration. In VARS→DDS, VARS is performed in a separate single trial with the budget of 1000 evaluations to prevent the degradation of DDS performance at initial iterations.

6.2 Results Analysis Methods

This thesis is focused on improving the efficiency and effectiveness of the DDS algorithm on calibrating high dimensional hydrological models. Hence, the proper way to evaluate the performance of proposed toolboxes is by comparing their results with the original DDS algorithm. The performance is compared through convergence and stochastic dominance graphs, statistical significance test. The hydrologic model calibration results are further analyzed by visually comparing and discussing time series of stream flow.

6.2.1Convergence

A common way to compare the functionality of global optimization algorithms is by comparing the average algorithms performance (Ali et al., 2005). Thus, in this thesis, the convergence of DDS and sensitivity-informed DDS methods is compared. The convergence of each algorithm is measured by taking the average over all trials of the current best solution in each iteration. The average convergence graphs can reveal the ability of sensitivity-informed DDS methods in improving the efficiency (i.e. finding a good quality solution in a shorter time) and effectiveness (i.e. discovering a better quality solution with equal budget) of automatic model calibration. Nevertheless, the convergence does not fully represent the algorithms performance, and the distribution of the best objective values found by each algorithm should also be considered (Tolson and Shoemaker, 2007). Hence, other types of result comparison approaches are utilized in this thesis.

6.2.2First-Order Stochastic Dominance

The performance of stochastic optimization algorithms varies between different trials. In order to investigate the distribution of their performance based on the best objective value found at the end of each trial, the first-order stochastic dominance method (levy, 1992) is considered in this thesis. This method has been widely used in the literature for comparing the performance of stochastic search algorithms, for example see, Tang et al. (2007); Hadka and Reed (2012); Asadzadeh and Tolson (2013). The first-order stochastic dominance is based on the empirical cumulative distribution function (CDF) of the best solution found by the optimization algorithm in each trial. The CDF plots demonstrates the probability of achieving equal or better value for the objective function. These plots are visually compared by first-order stochastic dominance concept. According to the definition, when two algorithms A and B are compared by their CDFs $F_A(x)$ and $F_B(x)$ (x is the best solution found in each trial), algorithm A stochastically dominates algorithm B, if and only if $F_A(x) \geq F_B(x)$, when the smaller values of x is of interest and in Figure 14-a. However, in Figure 14-b, the above statement is not true for each of the algorithms, thus, no algorithm dominates the other. The magnitude of difference between compared CDFs can be revealed by the statistical significance tests (Asadzadeh and Tolson, 2013).

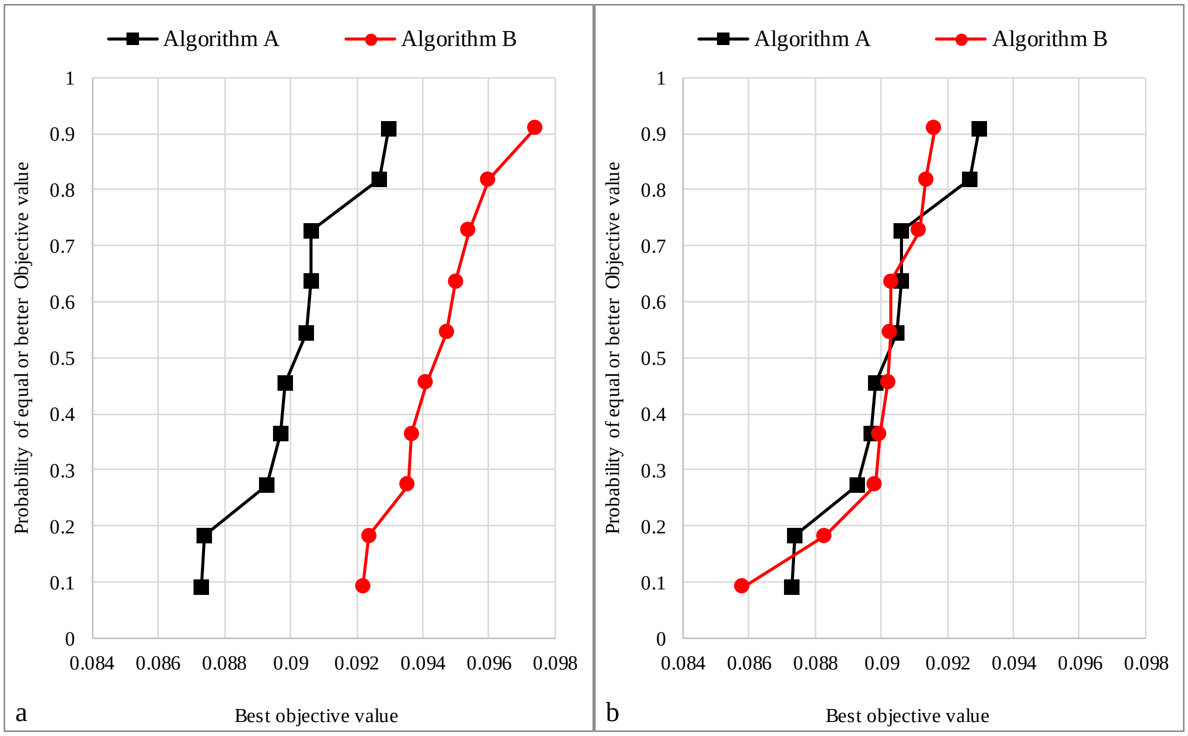


Figure 14. Example CDF plots comparison for algorithms A and B when lower objective values is of interest. (a) Algorithm A stochastically dominates algorithm B. (b) No algorithm dominates each other.

6.2.3Statistical Significance Test

The Wilcoxon rank-sum test (Gibbons and Chakraborti, 2011) is a statistical method to measure the significance of the difference between CDFs of corresponding samples. When comparing the performance of two stochastic optimization algorithms (A and B), this test has been used in the literature (Hadka and Reed, 2012; Asadzadeh and Tolson, 2013) to identify which algorithm has preferable results. In general, the Wilcoxon rank-sum test is usually performed to identify whether two autonomous samples (in this case, the final best solutions found in each trial of compared algorithms) come from identical populations or not. The main characteristic of the Wilcoxon rank-sum test that makes it suitable for comparing optimization algorithm is that it is a nonparametric statistical method, meaning that it does not require the user judgment. The Wilcoxon rank-sum test assumes that data points in the compared sets are independent which holds in this thesis because optimization trials are independent. Thus, this test is considered to compare the results of DDS and sensitivity-informed DDS algorithms. Under the null hypothesis of this test, both samples come from the same distribution of random variables ($F_A(x) = F_B(x)$). However, the alternate hypothesis states that the samples come from different distributions. When comparing algorithms A and B, if the significance level of the null hypothesis (p-value) is smaller than 5 percent, then the null hypothesis is rejected (H-value equals to one) and the alternate hypothesis is approved ($F_A(x) \neq F_B(x)$) with a 95 percent confidence level. Together with the stochastic dominance test, the Wilcoxon rank-sum test can reveal if an optimization algorithm stochastically significantly performs better than the other one (Asadzadeh and Tolson, 2013).

6.2.4Streamflow Time Series

When calibrating a hydrological model, visual comparison of simulated and measured outputs is a basic though effective approach to evaluate the calibration results (Asadzadeh et al., 2015). In hydrological model calibration, although the error function (e.g. NSE) value evaluates the overall fit between simulated and measured data, the model needs to be evaluated across the simulation time horizon to identify potential major issues such overestimating or underestimate streamflow in short but critical time periods such as extremely high-flow periods. Thus, streamflow time series can reveal whether simulated hydrograph matches the peak flows and low flows of measured hydrographs adequately. The calibration results of DDS and sensitivity-informed DDS algorithms generate unique simulated daily hydrographs that belong to specific parameter settings. Hence, in this thesis, the hydrographs derived from the final best solution in the best trial of DDS, GSA↔DDS, GSA→DDS and VARS→DDS are created and compared with the measured hydrograph to evaluate the performance of proposed algorithms in calibrating SAC-SMA, SWAT and WATFLOOD models.

6.2.5Sensitivity Analysis

Parameter sensitivity information is valuable for identifying the most influential parameters on the hydrologic model performance. This information is required when the model has to be recalibrated. Hence, the parameter sensitivity rankings, in which the most sensitive parameter is ranked first, calculated by Morris, Sobol, and VARS methods are presented. In order to generate the sensitivity results in GSA↔DDS and GSA→DDS approaches, the sensitivity analysis is performed for 10 different trials with the budget of 5% of total optimization cost. The average sensitivity rankings over all trials for each parameter derived from Morris and Sobol methods are reported and compared to validate the performance of the Morris and Sobol approaches. In VARS→DDS, however, the sensitivity analysis is only performed once and with the budget of 1000 model evaluations, as VARS is claimed to be a relatively more robust and efficient compared to the Morris and Sobol methods (Razavi and Gupta, 2016). In all three GSA methods, the sensitivity analysis is performed over the exact period as calibration. In order to validate the results of VARS, Morris and Sobol methods are performed with the budget of 1000, and the results are compared. The robustness of the VARS method is revealed by producing the similar sensitivity rankings compared to the Morris and Sobol methods.

7 Results and Discussion

7.1 GSA↔DDS

The performance of Morris↔DDS and Sobol↔DDS methods in minimizing Rastrigin and Griewank test functions and calibrating SAC-SMA model are evaluated. Due to the similarities between the results of Griewank and Rastrigin functions only the results of Griewank are shown in this section, and the results of Rastrigin are presented in the Appendix (Chapter 9). The performance of GSA↔DDS is compared with original DDS algorithm through the results analysis methods explained in Chapter 6. In addition, the average of initial and final parameter sensitivity rankings produced by Morris and Sobol methods are presented and compared.

7.1.1Griewank

According to Table 1, the global optimal objective function value of the Griewank test problem is equal to 0.0 that corresponds to all decision variables equal to 0.0. It is understood from Figure 15 that both Morris↔DDS and Sobol↔DDS show slower convergence over initial iterations. The main reason for this behavior is that GSA methods are not designed to find the optimal solution. Therefore, these toolboxes require more time to find the same quality solutions found by DDS. However, after the initial GSA is performed and the sensitivity indices are provided, Morris↔DDS illustrates faster convergence than DDS when minimizing the Griewank function. According to Figure 15-b, with the total budget of 10000 solution evaluations, after 1000 solution evaluations, on average Morris↔DDS finds very good solutions with the average objective function value equal to 2.84 compared to 32.5 for DDS. Nonetheless, Sobol↔DDS performed the worst after 1000 solution evaluations by scoring an average objective function value of 112.3. The main reason for the slower convergence of Sobol↔DDS is that Sobol could not correctly identify the most sensitive parameters with the relatively low budget used for the initial sensitivity ($0.05 \times 10000 = 500$ solution evaluations) and therefore misled the optimization to converge to a weak solution. Moreover according to Equation 12, Sobol requires independently generated solutions to calculate the sensitivity indices. However, the solutions derived from DDS are not entirely independent throughout the search, because DDS perturbs only some decision variables in each iteration. Thus, updating the original solution sets with the solutions created by DDS can cause miss calculation of parameter sensitivity index, and the inaccurate sensitivity results can misinform the algorithm toward the more frequent perturbation of less-sensitive parameters.

The stochastic dominance graphs in Figure 15-a and -b illustrate that the final best solutions found by Morris↔DDS stochastically dominate the best solutions found by DDS and Sobol↔DDS. When the optimization budget is 1000 solution evaluations, Morris↔DDS dominates DDS, and both Morris↔DDS and DDS dominate Sobol↔DDS. Moreover, the same behavior can be seen at the budget of 10000. This is further confirmed with the results of Wilcoxon rank-sum test in Table 5. The pairwise comparison of DDS, Sobol↔DDS and Morris↔DDS shows that in both optimization budgets, the best solutions found by Morris↔DDS are statistically significantly different from the best solutions found by DDS and Sobol↔DDS (p-value smaller than 0.05). In conclusion, in terms of convergence and the final best solutions found Morris↔DDS is the most promising method in minimizing Griewank function with 20 parameters. Although the average initial sensitivity rankings produced by Sobol and Morris methods are very similar, (Table 6) the significant difference in the average final sensitivity rankings derived from Sobol↔DDS and Morris↔DDS for Griewank function confirms the miss calculation of sensitivity indices.

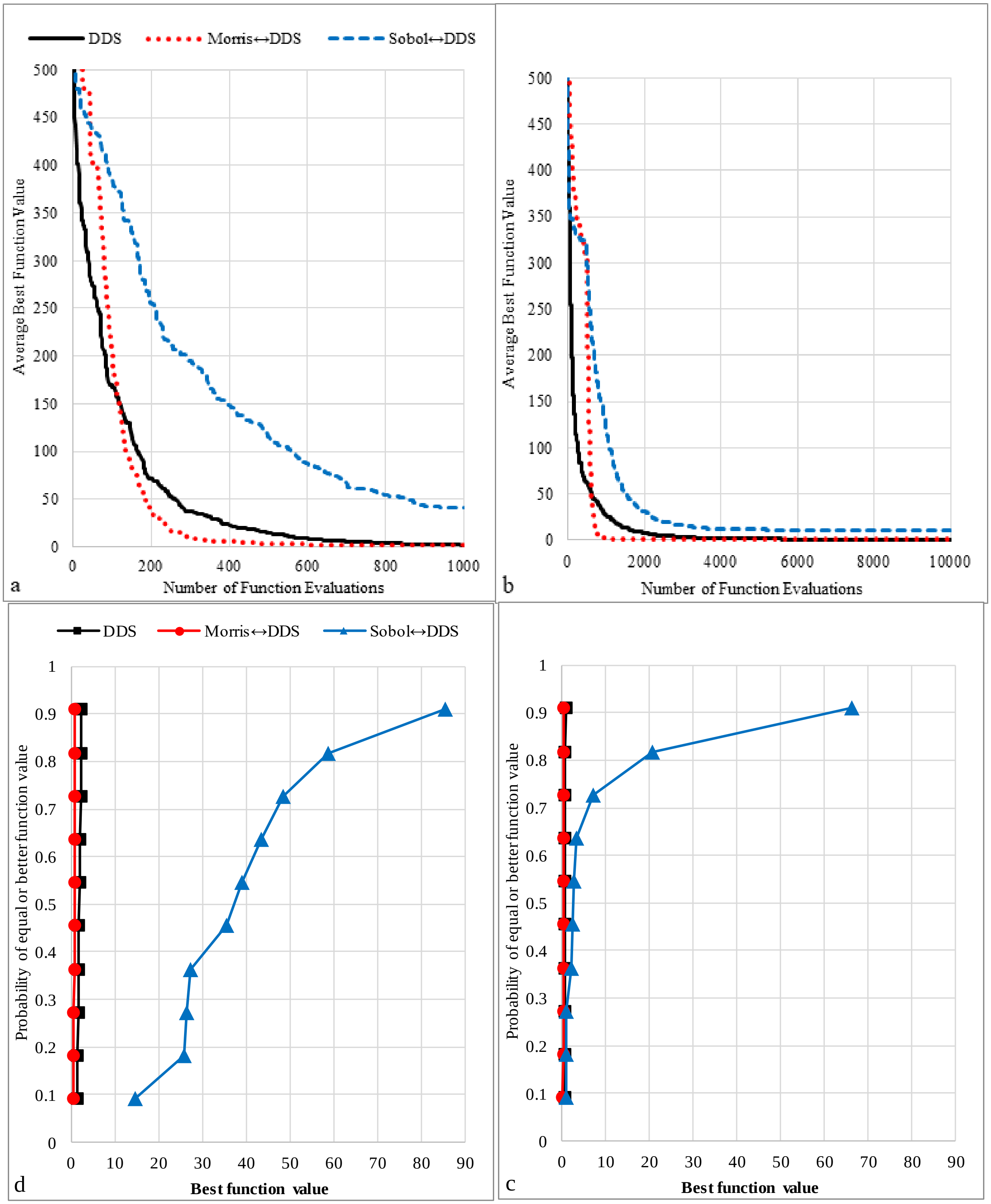


Figure 15. Convergence and first order stochastic dominance graphs of DDS, Sobol↔DDS and Morris↔DDS applied to the Griewank test function with 20 parameters (left side figures show budget of 1000 and right side figures show budget of 10000).

Table 5. Wilcoxon rank-sum test comparing DDS, Sobol↔DDS and Morris↔DDS applied to Griewank.

Computational Budget	Compared Methods	P-value	H-value
1000	DDS versus Morris↔DDS	1.82E-04	1
	DDS versus Sobol↔DDS	1.83E-04	1
10000	DDS versus Morris↔DDS	1.83E-04	1
	DDS versus Sobol↔DDS	1.83E-04	1

Table 6. Average sensitivity ranks for Griewank parameters by Morris and Sobol methods with 5% initial GSA budget (most versus least sensitive parameters are highlighted in red and, blue respectively).

	Budget of 500 (5 percent of 10000)				**Budget of 50 (5 percent of 1000)**			
Parameter Number	**Morris Ranking**		**Sobol Ranking**		**Morris Ranking**		**Sobol Ranking**	
	Initial	**Final**	**Initial**	**Final**	**Initial**	**Final**	**Initial**	**Final**
1	5	12	10	18	19	7	5	8
2	19	10	20	2	16	6	8	3
3	16	18	16	14	10	12	11	2
4	14	16	11	20	2	1	2	13
5	1	5	3	16	9	9	19	1
6	2	20	4	15	6	14	16	17
7	15	1	6	11	8	8	3	15
8	6	2	8	10	11	17	13	20
9	8	14	12	3	17	2	18	10
10	4	17	1	17	1	11	10	12
11	18	9	15	1	4	4	7	7
12	13	19	17	5	5	19	14	19
13	11	3	7	9	14	3	15	5
14	17	7	5	19	3	18	4	4
15	10	13	19	13	7	13	20	16
16	12	4	18	6	13	20	1	11
17	7	15	13	4	12	16	6	18
18	20	11	9	12	18	5	12	9
19	9	6	14	7	20	10	17	14
20	3	8	2	8	15	15	9	6

7.1.2SAC-SMA Model

Calibrating the 13 parameters of the SAC-SMA model of the Leaf River has been a benchmark optimization problem with the best reported objective function value of $1 - NSE = 0.076$ in the literature (Asadzadeh, 2012). According to Figure 16-a and –b, both Sobol↔DDS and Morris↔DDS demonstrate slower convergence rate compare to DDS when calibrating the 13 parameters of this hydrologic model. At the budget of 10000 model evaluations, DDS, Morris↔DDS and Sobol↔DDS found the best 1-NSE value equal to 0.087, 0.091, and 0.092, respectively, which are very close to the best value reported in the literature, 0.076 in Asadzadeh (2012). It should be noted that the calibration of SAC-SMA model in this thesis is started from a random solution to increase the challenge for the algorithms in terms of finding good quality objective values. In addition, due to the incorrect updating of initial sensitivity indices, Sobol↔DDS showed a notably weaker convergence compare to DDS and Morris↔DDS. For instance, when the budget is equal to 1000, at 200th iteration, DDS found 0.092, while Sobol↔DDS found 0.11 as the best $1 - NSE$ value.

The stochastic dominance graphs in Figure 16-c and -d and the corresponding P-values of the Wilcoxon rank-sum test in Table 7 show that DDS is the preferred algorithm for calibrating the SAC-SMA model at both computational budgets of 1000 and 10000 model evaluations. For example, for both optimization budgets, the CDF plots and $P - \text{value} < 0.004$ shows that DDS is statistically significantly preferred over Sobol↔DDS. The CDF plots for DDS and Morris↔DDS are very similar showing that these two algorithms do not stochastically dominate each other and have a comparable performance. Moreover, according to Table 7, the p-value for the pairwise comparison of DDS and Morris↔DDS for both budgets is greater than 0.05 confirming that the two algorithms have statistically similar performance for calibrating the SAC-SMA model of the Leaf River watershed.

Table 8 demonstrates that regardless of the budget of initial GSA, both Morris and Sobol methods similarly define the four most sensitive parameters of SAC-SMA model. To illustrate this, both Morris and Sobol methods initially identify the LZFSM (maximum capacity of the lower zone free water supplemental storage) as the most sensitive parameter of SAC-SMA. In addition, in the final sensitivity results by the Morris method, similar sensitivity rankings are reported for parameters. However, different final parameter rankings produced by Sobol, verifies that the final Sobol sensitivity results are inaccurate in both budgets. The most sensitive parameters identified by Sobol final rankings are in fact among the least sensitive parameters defined by Morris method. In addition, increasing the budget from 1000 to 10000 solution evaluations for the Sobol↔DDS method negatively affects the inaccuracy in the final sensitivity rankings due to lower number of updates in sensitivity indices.

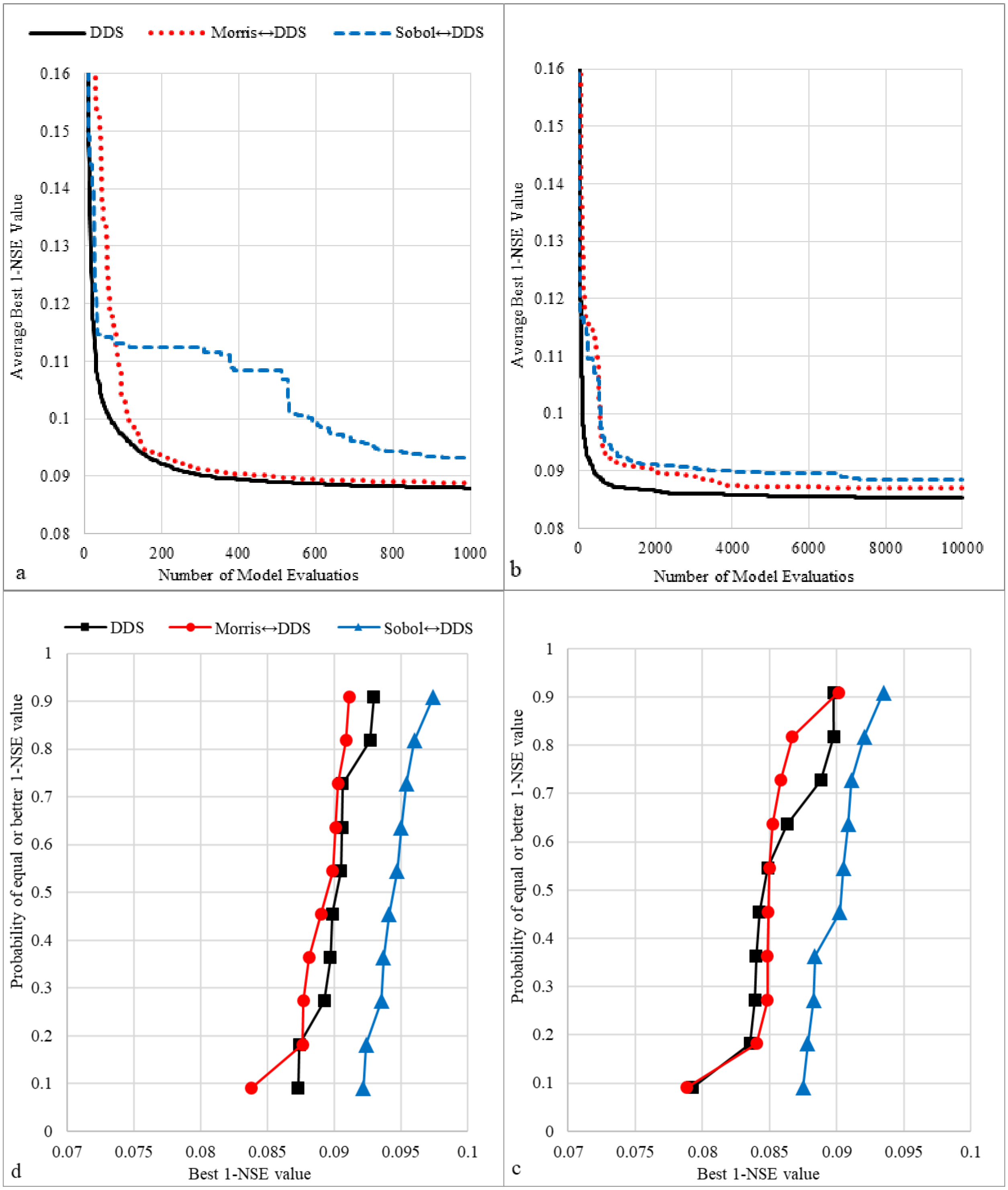


Figure 16. Convergence and stochastic dominance graphs of DDS, Sobol↔DDS and Morris↔DDS for SAC-SMA calibration (left versus right side figures are for 1000 versus 10000 solution evaluations)

Table 7. Wilcoxon rank-sum test comparing DDS, Sobol↔DDS and Morris↔DDS for SAC-SMA calibration

Computational Budget	Compared Methods	P-value	H-value
1000	DDS versus Morris↔DDS	4.27E-01	0
	DDS versus Sobol↔DDS	5.83E-04	1
10000	DDS versus Morris↔DDS	7.33E-01	0
	DDS versus Sobol↔DDS	4.00E-03	1

Table 8. Average sensitivity ranks for SAC-SMA parameters by Morris and Sobol methods with 5% initial GSA budget (most versus least sensitive parameters are highlighted in red and, blue respectively).

	Budget of 500 (5 percent of 10000)				**Budget of 50 (5 percent of 1000)**			
	Morris Ranking		**Sobol Ranking**		**Morris Ranking**		**Sobol Ranking**	
Parameter	**Initial**	**Final**	**Initial**	**Final**	**Initial**	**Final**	**Initial**	**Final**
UZTWM	7	9	9	9	5	9	13	11
UZFWM	3	7	10	10	10	4	4	6
LZTWM	12	12	12	4	12	12	10	12
LZFPM	8	10	3	12	6	10	6	4
LZFSM	1	1	1	7	2	1	1	2
ADIMP	10	6	6	1	8	6	7	7
UZK	11	11	13	13	11	11	5	3
LZPK	2	2	2	6	1	2	2	1
LZSK	6	4	11	5	7	5	11	13
PCTIM	5	5	5	11	4	7	8	8
ZPERC	4	3	4	8	3	3	3	5
REXP	9	8	7	3	9	8	12	10
PFREE	13	13	8	2	13	13	9	9

In Figure 17, the hydrographs are derived from the final best solutions found by DDS, Sobol↔DDS and Morris↔DDS. The hydrographs are generated with respect to the best trial of DDS, Sobol↔DDS and Morris↔DDS algorithms. Although SAC-SMA is calibrated using a two-year period of historical data, the most challenging period, which is from October 1952 to October 1953, is selected to show the hydrographs. As can be seen from this figure, in both optimization budgets, the final NSE value from the best trial of Sobol↔DDS is slightly lower than the NSE value found by DDS and Morris↔DDS. Nevertheless, the corresponding hydrographs of Sobol↔DDS and Morris↔DDS are almost identical to the hydrograph from DDS, and show adequate match with the observed hydrograph in the low-flow and high-flow periods. Hence, based on streamflow time series, it is also concluded that neither of the Sobol and Morris approaches could guide DDS to calibrate this model more efficiently and effectively.

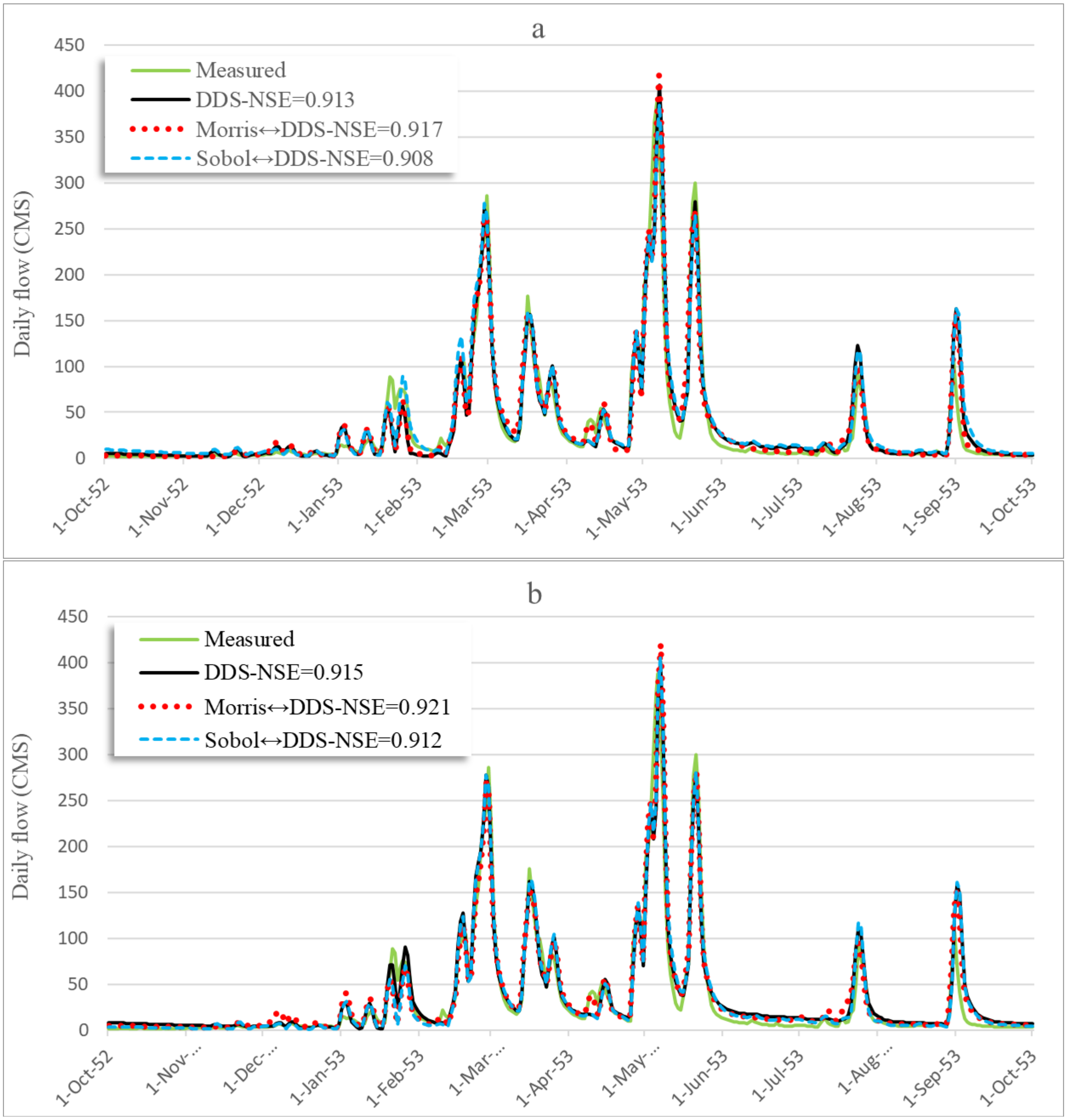


Figure 17. Streamflow time series for SAC-SMA model from best parameter values identified by DDS, DDS-Morris and DDS-Sobol with the budget of (a) 1000 and (b) 10000 model evaluations

7.1.3GSA↔DDS Results Discussion

It was expected that creating an interactive method that continuously couples DDS with a GSA toolbox to update the results of both sensitivity and optimization would increase the efficiency of DDS. However, as the results demonstrated, the performance of Sobol↔DDS was significantly weaker in terms of convergence rate and final best solution found due to the improper updating procedure for sensitivity results. Furthermore, in Morris↔DDS, the performance has not improved significantly compared to original DDS. The main issue is updating the sensitivity indices throughout the optimization process. According to the results of sensitivity analysis, updating sensitivity indices in Sobol↔DDS reduced the accuracy of parameter sensitivity rankings, as the final rankings generated by Sobol method was significantly different from the final rankings generated by Morris method. Moreover, in Morris↔DDS, no significant difference between the initial and final parameter sensitivity rankings is observed. Hence, updating the initial sensitivity indices in both Morris and Sobol approaches showed no benefit. Additionally, in both Sobol↔DDS and Morris↔DDS algorithms, in order to update the sensitivity indices, selected parameters have to be perturbed individually. This will reduce the efficiency of optimization because in DDS, the selected parameters should be perturbed simultaneously.

Hence, to address the shortcomings of GSA↔DDS a new method, which is called GSA→DDS, is introduced. In this approach, the GSAs (Morris and Sobol) are only performed initially to generate the sensitivity indices and initial solutions for optimization. The results of GSA↔DDS illustrated that updating the sensitivity indices of Morris and Sobol methods showed no improvement. Thus, in GSA↔DDS the sensitivity indices are not updated, and the original perturbation method of DDS is respected.

7.2 GSA→DDS

GSA→DDS utilizes Sobol and Morris methods to perform GSA initially with the budget of 5 percent of total optimization cost. The performance of GSA→DDS is compared with DDS for solving the Rastrigin and Griewank functions and SAC-SMA Model. Results for Griewank and SAC-SMA are discussed in this section, and results for Rastrigin are discussed in the appendix.

7.2.1Griewank

According to Figure 18, unlike GSA↔DDS, regardless of the computational budget, both Morris→DDS and Sobol→DDS converge faster than DDS in minimizing the Griewank function. However, the improvement in the convergence rate is more significant when the optimization is performed with higher number of function evaluations. As an illustration, with the budget of 10000 function evaluations, at the 850th iteration, DDS found the function value equal to 35.61, while Morris→DDS and Sobol→DDS found 5.91 and 4.49 as the best function value respectively. This improvement in the convergence is due to the guided search of algorithm toward more frequent perturbation of high sensitivity parameters. This will increase the probability of finding good quality solutions in a shorter time. According to Figure 18-c, GSA→DDS (both Morris and Sobol) outperform DDS when the computational budget is relatively large, 10000 solution evaluations. Moreover, Table 9 shows that in the budget of 10000, the p-value of the Wilcoxon rank-sum test for the pairwise comparison of the corresponding CDFs of GSA→DDS and DDS is smaller than 0.05. Hence, the null hypothesis is rejected (H-value equals to one), and therefore, both Morris→DDS and Sobol→DDS algorithms are statistically significantly preferred over DDS in minimizing Griewank function when the budget is 10000 evaluations. On the other hand, at the budget of 1000 evaluations, Figure 18-d illustrates that Sobol↔DDS stochastically dominates DDS, while Morris↔DDS and DDS do not stochastically dominate each other. Moreover, the p-

value of the pairwise comparison of Sobol→DDS and DDS is less than 0.05. Hence, at lower budget, Sobol→DDS is preferred over DDS and Morris↔DDS.

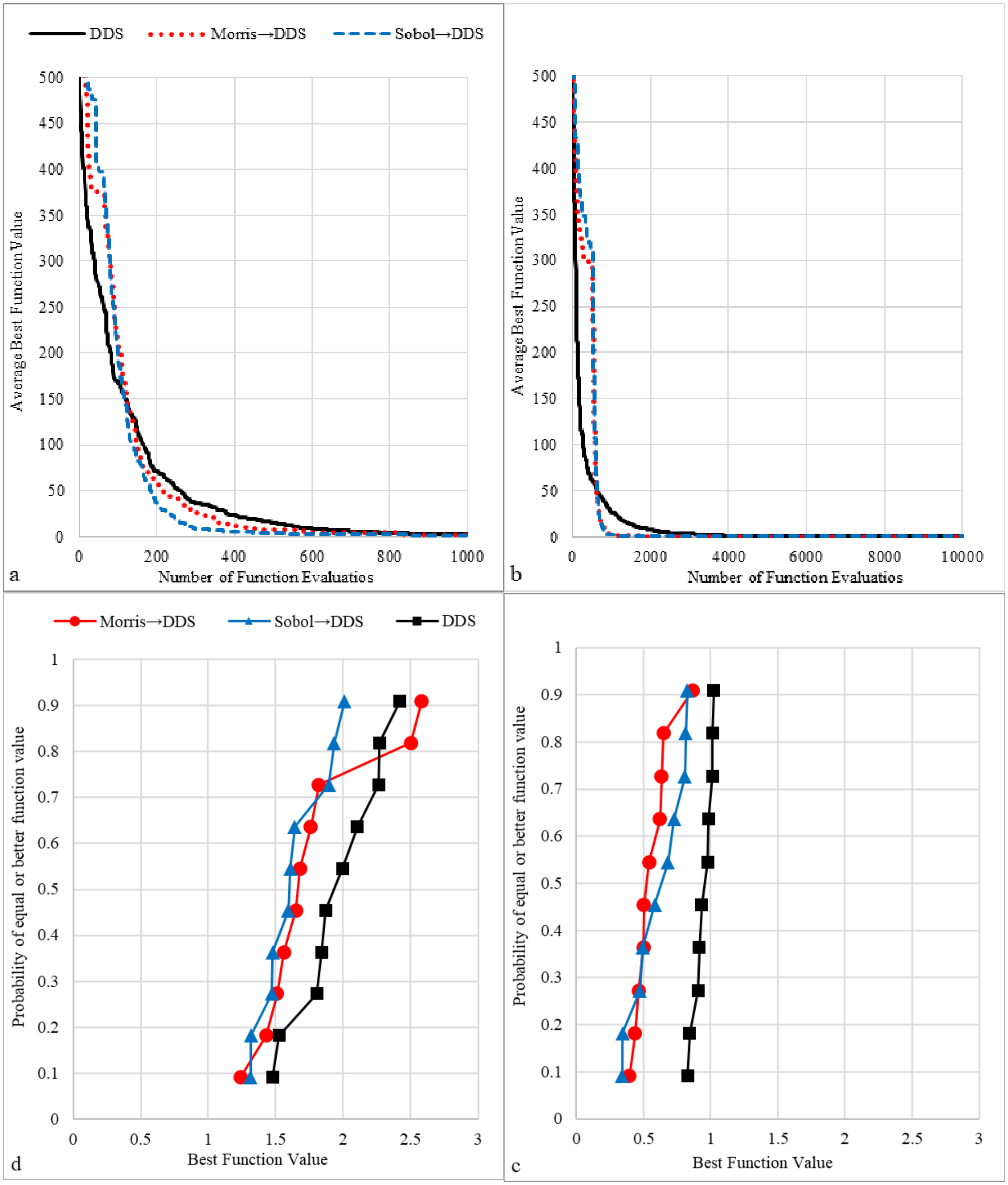


Figure 18. Convergence and first order stochastic dominance DDS, Sobol→DDS and Morris→DDS applied to Griewank with 20 parameters (left versus right side figures for budget of 1000 10000).

Table 9. Wilcoxon rank-sum test comparing DDS, Sobol→DDS and Morris→DDS applied to Griewank.

Computational budgets	Compared Methods	P-value	H-value
1000	DDS versus Morris→DDS	0.185	0
	DDS versus Sobol→DDS	0.037	1
10000	DDS versus Morris→DDS	3.29E-04	1
	DDS versus Sobol→DDS	1.82E-04	1

Table 10. Average sensitivity ranks for Griewank parameters by Morris and Sobol methods with 5% initial GSA budget (most versus least sensitive parameters are highlighted in red and, blue respectively).

	Budget of 500 (5 percent of 10000)		**Budget of 50 (5 percent of 1000)**	
Parameter Number	**Morris**	**Sobol**	**Morris**	**Sobol**
1	5	10	19	5
2	19	20	16	8
3	16	16	10	11
4	14	11	2	2
5	1	3	9	19
6	2	4	6	16
7	15	6	8	3
8	6	8	11	13
9	8	12	17	18
10	4	1	1	10
11	18	15	4	7
12	13	17	5	14
13	11	7	14	15
14	17	5	3	4
15	10	19	7	20
16	12	18	13	1
17	7	13	12	6
18	20	9	18	12
19	9	14	20	17
20	3	2	15	9

The similarities between the results of Morris→DDS and Sobol→DDS is due to the similar sensitivity results generated by Sobol and Morris methods. Table 10 demonstrates that both Sobol and Morris have correctly identified the four most sensitive parameters of Griewank functions. Although, in lower budget (50 evaluations) small discrepancies can be seen between the parameter rankings by Sobol and Morris, as the budget increases, the sensitivity measurements are improved.

However, both Sobol and Morris methods fail to identify the least and moderate sensitive parameters of Griewank function. The main reason is that the Griewank function is not designed for sensitivity analysis, and its value is equally sensitive to all its decision variables. Hence, it is difficult for GSA methods to identify the moderate or least sensitive parameters of Griewank function.

7.2.2SAC-SMA Model

It is illustrated in Figure 19-a and -b that regardless of weaker performance at initial iterations, both Morris→DDS and Sobol→DDS converged faster than DDS for calibrating 13 parameters of the SAC-SMA model of the Leaf River watershed. For instance, with the optimization budget of 1000, the final best solution of DDS is equal to 0.09, while Morris→DDS and Sobol→DDS found a slightly better solution after 766 model evaluations. Moreover, when the optimization budget is an order of magnitude larger (10000), Morris→DDS and Sobol→DDS found similar solutions compared to the final solution of DDS only after 5737 solution evaluations.

However as shown in Figure 19-c and -d, both Morris→DDS and Sobol→DDS methods showed no significant improvement in the final best objective values compared to DDS, because regardless of the computational budget, they did not stochastically dominate of DDS. Additionally, the P-value of Wilcoxon rank-sum test in Table 11 confirms that none of the three algorithms is preferred based on its final optimal solution. Thus, GSA→DDS is not preferred over DDS in calibrating the SAC-SMA model.

According to the results of parameter sensitivity ranking (Table 12), when the GSA budget is equal to 500 model evaluations, both Morris and Sobol methods identify LZPK (lower zone primary free water depletion rate) as the most sensitive parameter, and ZPERC, LZSK, LZFSM as other important parameters. Furthermore, Morris and Sobol showed similar performance in identifying moderate and least sensitive parameters. The analogous rankings by Morris and Sobol methods verifies the robustness of the sensitivity results produced by these methods.

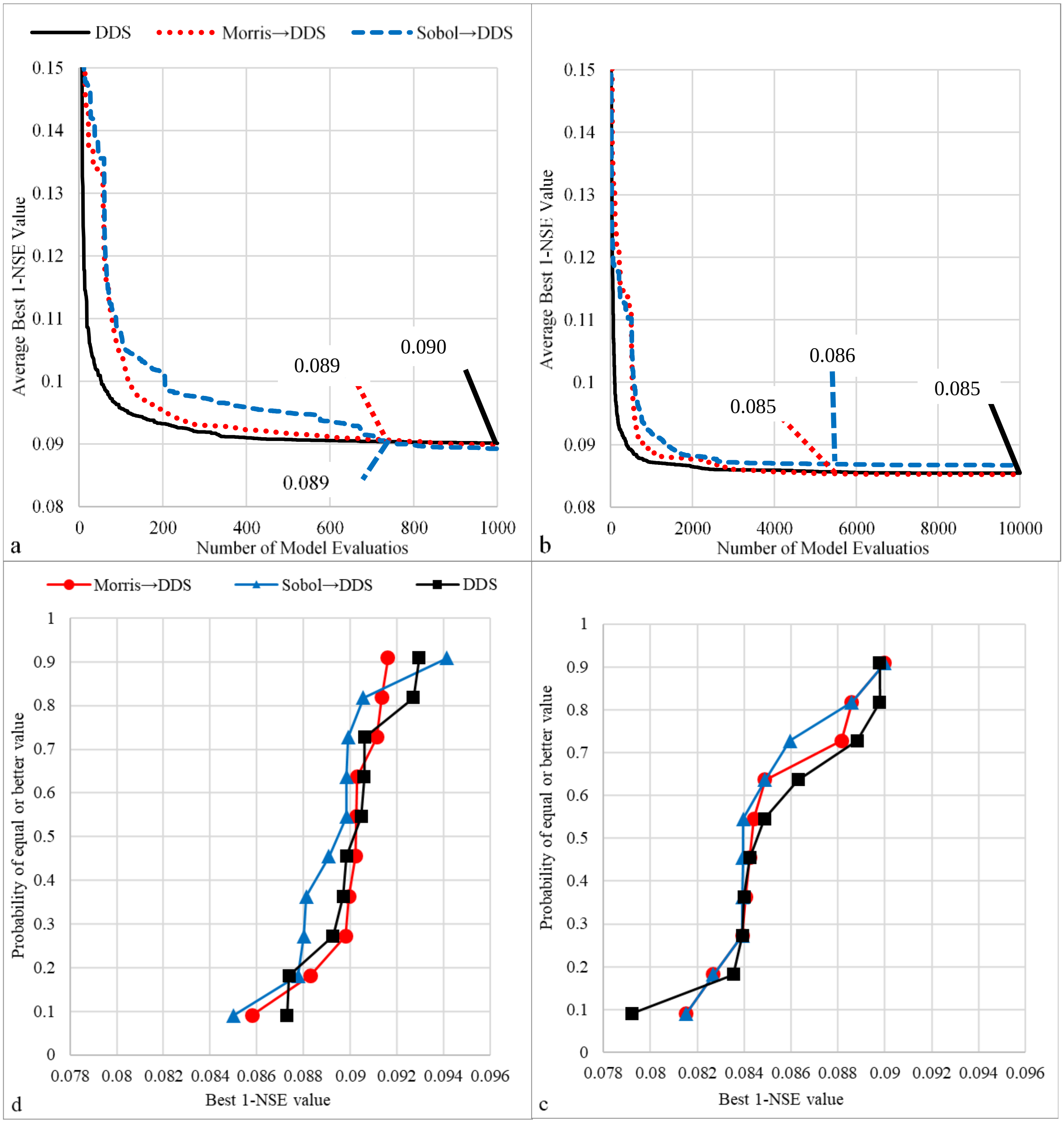


Figure 19. Convergence and first order stochastic dominance graphs of DDS, Sobol→DDS and Morris→DDS applied to the calibration of the SAC-SMA model of the Leaf River watershed (left side figures show budget of 1000 and right side figures show budget of 10000).

In Figure 20, the time series of streamflow simulated by the best solution found by DDS is compared to that of the solution found by Morris→DDS and Sobol→DDS after 5737 iterations for the budget of 10000 and after 766 iterations for the budget of 1000. In order to produce the hydrographs, the best trial of DDS, Morris→DDS and Sobol→DDS algorithms are considered. According to this figure, the hydrographs derived from DDS, Morris→DDS and Sobol→DDS are analogous (due to the similar NSE values), and have a proper match with the observed data. Hence, the hydrographs also validates the faster convergence of GSA→DDS compare to DDS.

Table 11. Wilcoxon rank-sum test comparing DDS, Sobol→DDS and Morris→DDS applied to SAC-SMA.

Computational Budget	Compared Methods	P-value	H-value
1000	DDS versus Morris→DDS	1	0
	DDS versus Sobol→DDS	0.427	0
10000	DDS versus Morris→DDS	0.909	0
	DDS versus Sobol→DDS	0.623	0

Table 12. Average sensitivity ranks for SAC-SMA parameters by Morris and Sobol methods with 5% initial GSA budget (most versus least sensitive parameters are highlighted in red and, blue respectively).

	Budget of 500 (5 percent of 10000)		**Budget of 50 (5 percent of 1000)**	
Parameter	**Morris**	**Sobol**	**Morris**	**Sobol**
UZTWM	6	5	7	2
UZFWM	7	9	9	10
LZTWM	12	6	12	8
LZFPM	8	11	6	13
LZFSM	2	4	1	11
ADIMP	10	12	11	6
UZK	11	10	10	5
LZPK	1	1	3	12
LZSK	4	3	5	4
PCTIM	5	7	4	3
ZPERC	3	2	2	1
REXP	9	13	8	9
PFREE	13	8	13	7

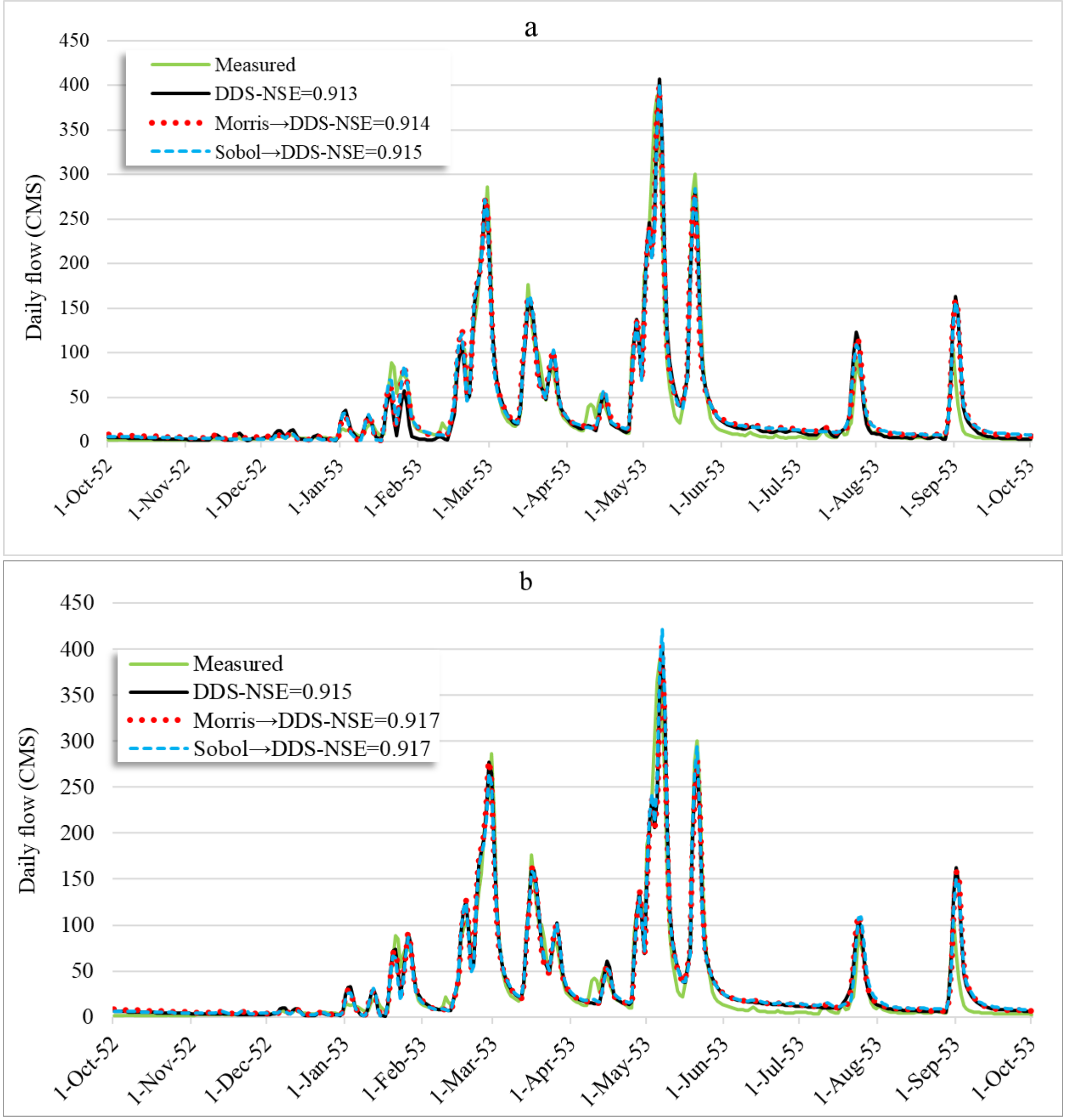


Figure 20. The comparison of streamflow time series for SAC-SMA model from best parameter value generated by DDS, DDS plus Morris and DDS plus Sobol. (a) Calibration budget of 1000. (b) Calibration budget of 10000

7.2.3GSA→DDS Results Discussion

The main objective of this thesis is to increase the efficiency and effectiveness of hydrological model calibration by means of combining the sensitivity and optimization. Results in this section showed that GSA→DDS methods have only improved the efficiency and not the effectiveness of the model calibration. As the results show, Morris→DDS and Sobol→DDS have faster convergence compare to DDS. The improved convergence rate is more significant in minimizing Griewank function than calibrating the SAC-SMA model. The main reason is that the SAC-SMA model has only 13 parameters, while GSA→DDS is more suitable for optimization problems with higher number of parameters. In addition, it is understood from the results that with higher computational budget of optimization, GSA→DDS converge significantly faster than DDS. However, the final solution found by Morris→DDS and Sobol→DDS may not be substantially better than that of DDS. The main issue is applying sensitivity analysis initially within the optimization process, which degrades the performance of optimization algorithm. In fact, initial iterations are the most important part of optimization because DDS has a high convergence rate at those iterations. Thus, performing sensitivity analysis within initial iterations of optimization can reduce the convergence rate and affect the final best solution found, as GSA toolboxes are not designed for optimization. To address this issue, the VARS→DDS approach is proposed in this research and its results are discussed in the next section.

7.3 VARS→DDS

VARS→DDS is applied to solve the Griewank, Rastrigin and SAC-SMA case studies. Furthermore, to validate the usefulness of VARS→DDS, it is used to calibrate two relatively more complex hydrological models, namely SWAT and WATFLOOD that they have a large set of calibration parameters. To demonstrate the benefit of VARS→DDS over calibrating hydrological models with reduced and full parameter sets, the results of VARS→DDS is compared to DDS with full parameter set (referred to as DDS) and DDS with reduced parameter set (referred to as DDS-Reduced). In DDS-Reduced, optimization is performed to calibrate only the most sensitive parameters, which is a common approach in the literature (Muleta and Nicklow, 2005; Tang et al., 2007; Wagener and Kollat, 2007; Lu et al., 2015). Results for the Griewank and Rastrigin test functions are shown in the appendix.

7.3.1SAC-SMA

As listed in Table 14, the IVARS sensitivity ranking derived from VARS, confirms the results of Morris and Sobol methods in identifying the four most sensitive parameters. According to all IVARS levels, the parameter with highest sensitivity is LZPK (lower zone primary free water depletion rate), and parameters LZFSM, PCTIM and ZPERC are ranked as the second, third and fourth most sensitive respectively. Furthermore, the least sensitive parameters identified by all IVARS levels and Morris method is PFREE. Therefore, VARS is also able to identify the least sensitive parameters properly.

The 13 parameters of the SAC-SMA model of the Leaf River watershed are calibrated using VARS→DDS, DDS and DDS-Reduced with respect to the four most sensitive parameters as shown in Table 14. Hence, the Guided Parameter Selection iterations (i_{GPS}) of VARS→DDS when calibrating SAC-SMA model are respectively 120th and 600th iteration for 1000 and 10000 model evaluations.

The convergence graphs in Figure 21 illustrates that VARS→DDS showed no improvement over DDS and DDS-Reduced. In fact, when the optimization budget is at 10000, the convergence rates of all three algorithms are very similar. However, at the budget of 1000, DDS-

Reduced has shown slightly better convergence rate at initial iterations. For instance, at iteration number of 100, both DDS and VARS→DDS found 0.094 as the $1 - NSE$ value, while DDS-Reduced found $1 - NSE$ value equal to 0.080 which is very close to the best known value for this problem (0.076).

The first order stochastic dominance graphs in Figure 22 and the Wilcoxon rank-sum test in Table 13 reveal that at the budget of 1000 model evaluations, VARS→DDS and DDS do not stochastically dominates each other. However, at this budget, DDS-Reduced stochastically significantly dominates both CDFs of DDS and VARS→DDS. On the other hand, at the higher budget of 10000 solution evaluations, VARS→DDS stochastically dominates DDS with the corresponding p-value of 0.02. This clearly illustrates that VARS→DDS is preferable to DDS when calibrating SAC-SMA model with the budget of 10000 model evaluations. In addition, at this budget, although DDS-Reduced does not dominants VARS→DDS, Figure 22-a demonstrates that DDS-Reduce has a high chance for dominating VARS→DDS, as DDS-Reduced generated better objective value in all trials except one, and showed more consistence results between all trials. Furthermore, DDS-Reduced stochastically significantly dominates DDS with the corresponding p-values of 0.00018. Hence, regardless of computational budget, when calibrating the SAC-SMA model, DDS-Reduced is certainly preferable to DDS, and outperforms VARS→DDS in most of the trials.

Figure 23 demonstrates that no significant difference between the corresponding hydrographs and the best NSE values derived from best trial of DDS, DDS-Reduced and VARS→DDS can be seen. This is because calibration of the SAC-SMA model of the Leaf River is not complex enough for VARS→DDS to demonstrate its advantage over DDS.

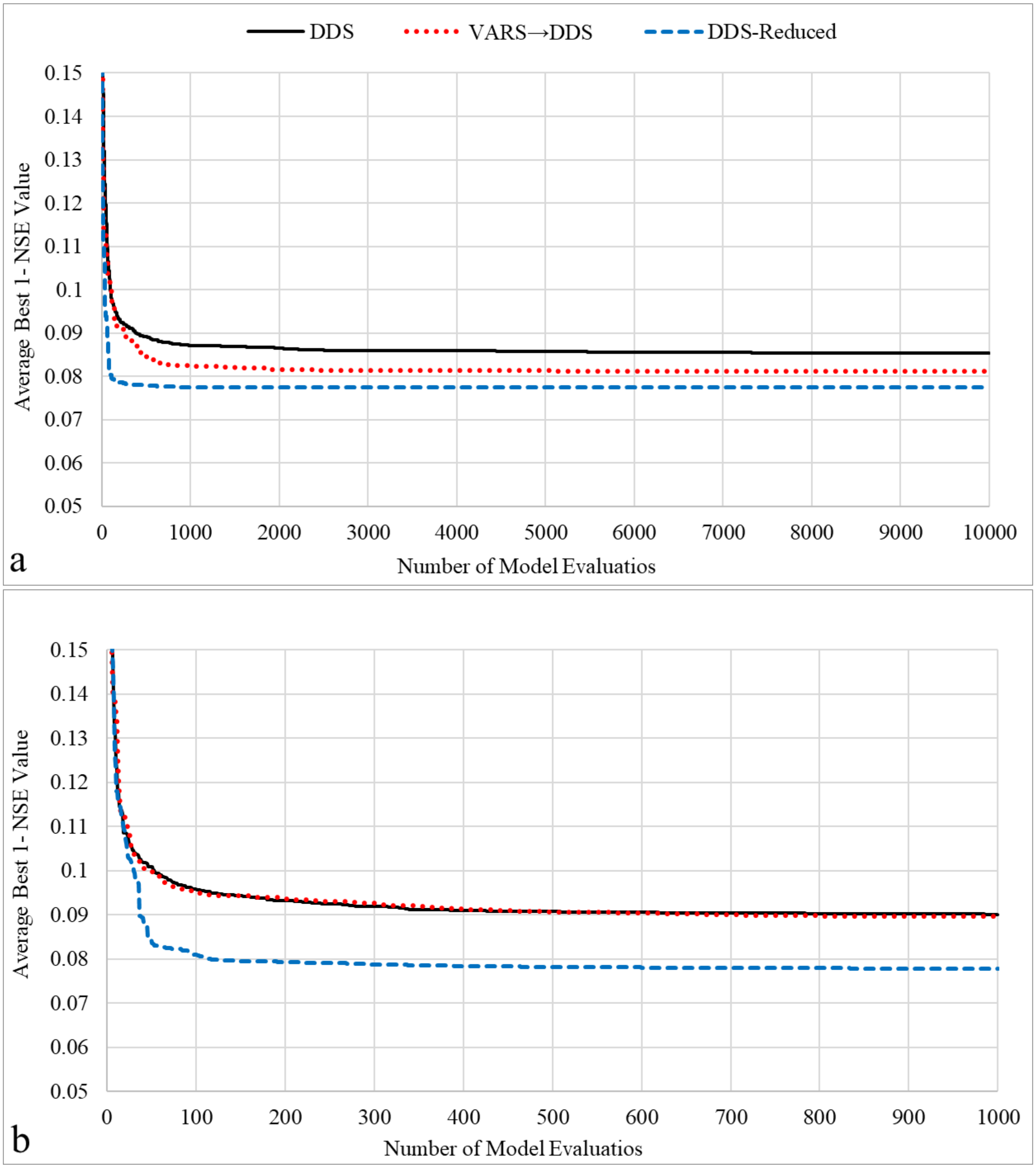


Figure 21. The convergence graphs of DDS, VARS→DDS, and DDS-Reduced applied to the calibration of the SAC-SMA model of the Leaf River watershed (a) Convergence graph for the budget of 10000. (b) Convergence for the budget of 1000.

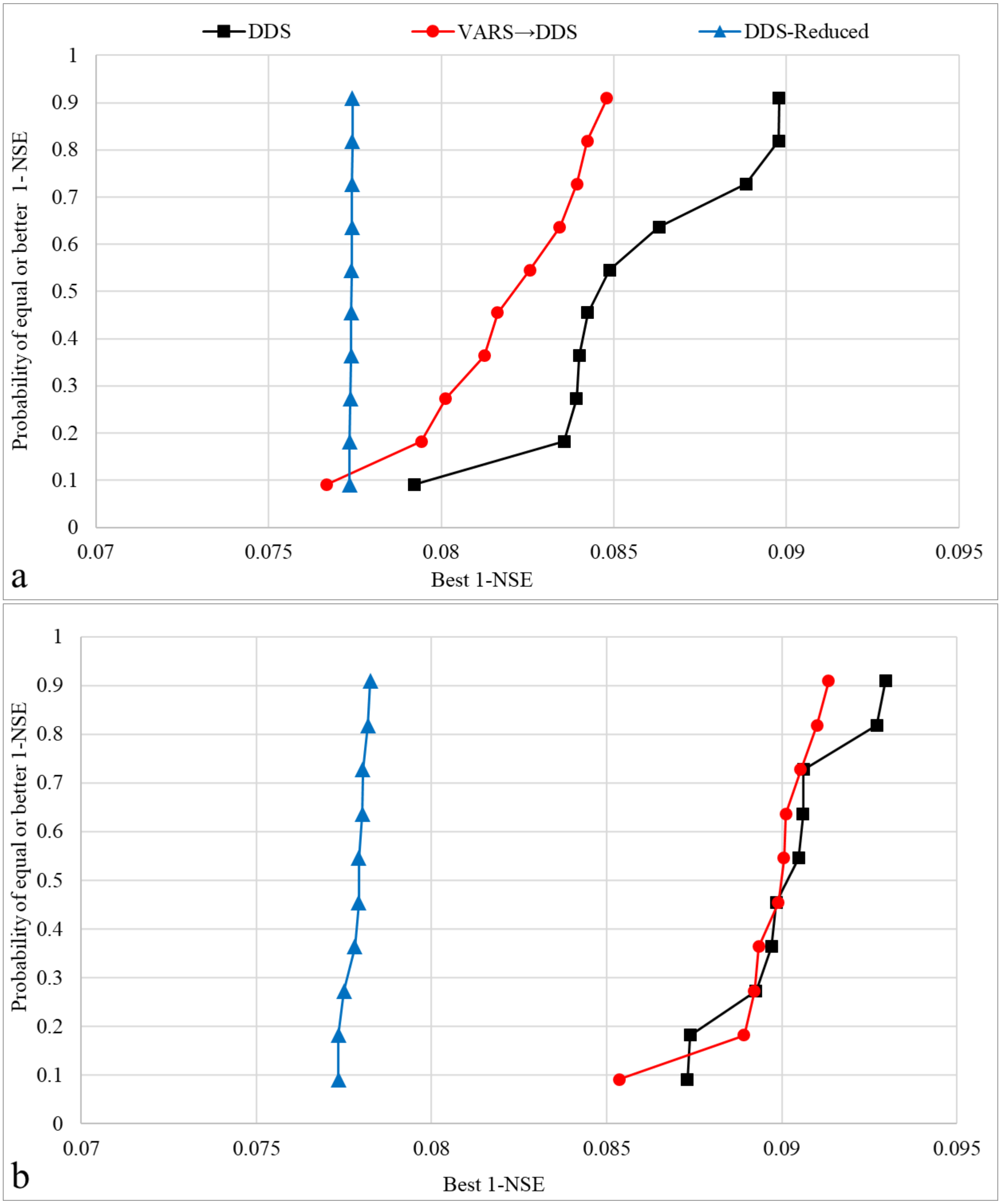


Figure 22. The first order stochastic dominance graphs of DDS, VARS→DDS, and DDS-Reduced applied to the calibration of the SAC-SMA model of the Leaf River watershed. (a) Budget of 10000. (b) Budget of 1000.

Table 13. Wilcoxon rank-sum test comparing DDS, DDS-Reduced and VARS→DDS applied to SAC-SMA.

Computational budgets	Compared Methods	P-value	H-value
1000	DDS versus VARS→DDS	0.6776	0
	DDS versus DDS-Reduced	1.83E-04	1
	DDS-Reduced versus VARS→DDS	1.83E-04	1
10000	DDS versus VARS→DDS	0.0173	1
	DDS versus DDS-Reduced	1.83E-04	1
	DDS-Reduced versus VARS→DDS	2.80E-03	1

Table 14. Average sensitivity ranks for SAC-SMA parameters by VARS, Morris and Sobol with 1000 model evaluations (most versus least sensitive parameters are highlighted in red and, blue respectively).

Parameter	IVARS10	IVARS20	IVARS30	IVARS40	IVARS50	Morris	Sobol
UZTWM	7	7	7	7	7	9	11
UZFWM	9	9	9	9	9	4	6
LZTWM	10	10	10	10	10	12	12
LZFPM	11	11	11	11	11	10	4
LZFSM	2	2	2	2	2	1	2
ADIMP	6	6	6	6	6	6	7
UZK	8	8	8	8	8	11	3
LZPK	1	1	1	1	1	2	1
LZSK	5	5	5	5	5	5	13
PCTIM	3	3	3	3	3	7	8
ZPERC	4	4	4	4	4	3	5
REXP	12	12	12	12	12	8	10
PFREE	13	13	13	13	13	13	9

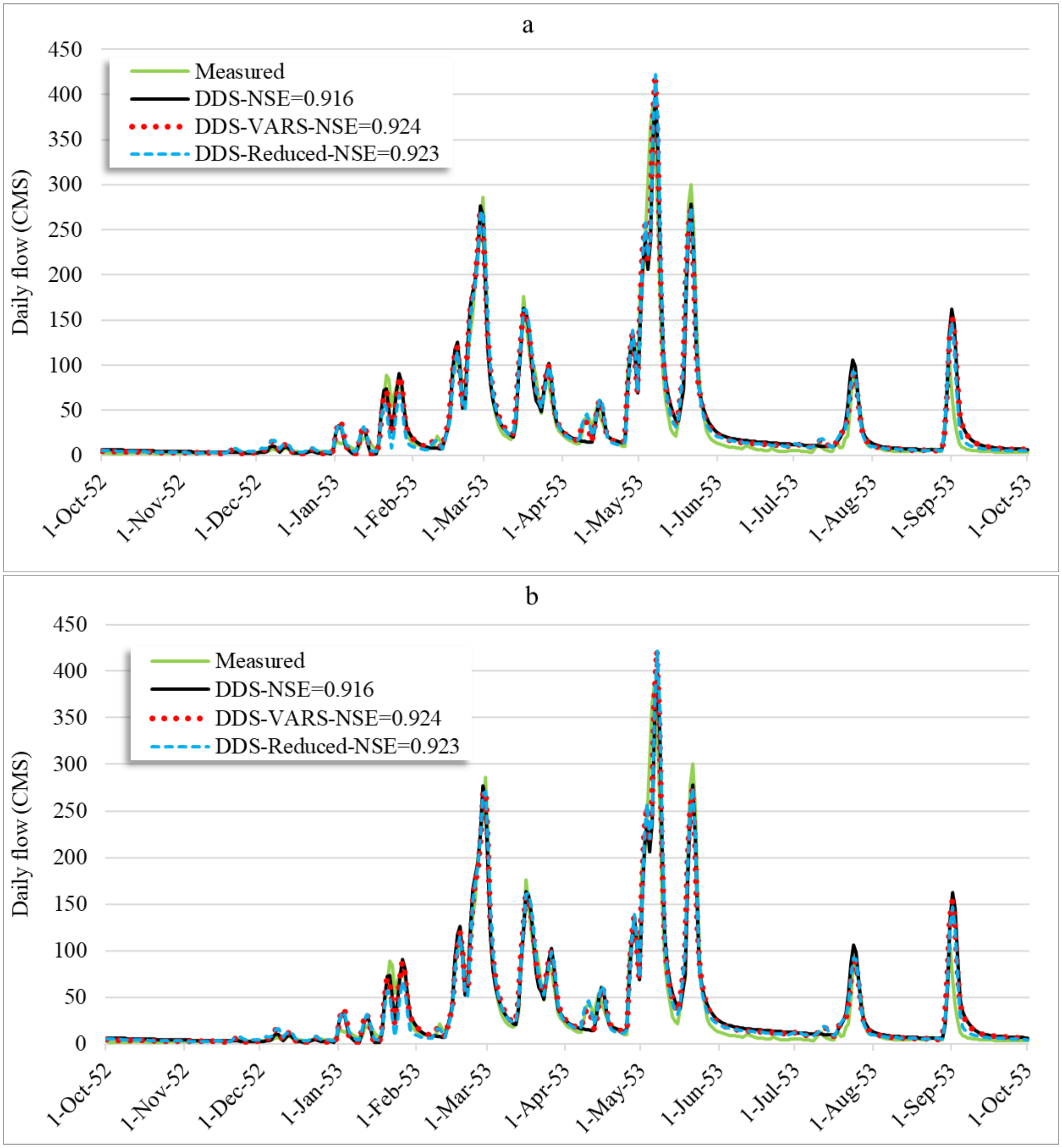


Figure 23. The comparison of streamflow time series for SAC-SMA model from the best parameter values generated by DDS, VARS→DDS and DDS-Reduced. (a) Optimization budget of 10000. (b) Optimization budget of 1000.

7.3.2SWAT Rouge River

The sensitivity rankings of SWAT parameters (Table 16) show that the most sensitive parameters identified by VARS, Morris and Sobol methods in this thesis is SNO50COV, due to the major contribution of snowmelt in streamflow. In addition, the seven most sensitive parameters of SWAT model (SNO50COV, SNOCOVMX, CN, CH_K, TIMP, SOL_Z and ESCO) have been similarly identified by VARS, Sobol and Morris methods in this thesis. Furthermore, the parameters such as, GW_Delay, SMFMN and SMTMP are among the least sensitive parameters identified by all three GSA methods. Thus, the similar sensitivity rankings cross validate the results of VARS, Morris and Sobol methods with the budget of 1000 model evaluations.

According to sensitivity analysis results for SWAT model of the Rouge River basin, seven parameters are selected as the most sensitive. Therefore, when calibrating the 23 parameters of the SWAT model using VARS→DDS, the i_{GPS} iterations are considered at 122th and 600th iterations for the budget of 1000 and 10000 respectively.

As demonstrated by Figure 24-a, DDS, VARS→DDS, and DDS-Reduced have similar convergence rates for calibrating the SWAT model of the Rouge River watershed with 1000 solution evaluations. However as shown in Figure 24-b, when the optimization budget is relatively larger (10000), VARS→DDS outperforms DDS-Reduced and DDS, and finds a solution with $NSE = 0.73$ that is improved respectively by 9% and 6% compared to the best NSE value found by DDS-Reduced (0.67) and DDS (0.69). Moreover, VARS→DDS showed faster convergence than DDS and DDS-Reduced in calibrating 23 parameters of SWAT model. As an illustration, with the optimization budget of 10000, VARS→DDS found the final best solution of DDS-Reduced (0.67) after around 1500 model evaluations. Furthermore, the solution found by

VARS→DDS is better than the best known solution reported for this case study in Asadzadeh et al. (2015).

The stochastic dominance graphs in Figure 25 confirm that when the number of model evaluations is equal to 10000, VARS→DDS stochastically dominates DDS and DDS-Reduced. Moreover, the p-values of Wilcoxon rank-sum test in Table 15 show that VARS→DDS is statistically significantly preferred to the other two approaches for solving this calibration problem. On the other hand, when the optimization budget is relatively limited (1000), DDS is slightly preferred over the other two approaches. These results suggest that, the common approach used in the literature that reduces the calibration problem to only the most sensitive parameters of the model is not necessarily more efficient than calibrating models with full parameter set regardless of the available computational budget.

The time series of streamflow generated by the best solutions found by the three different approaches for this case study are plotted in Figure 26. The corresponding hydrographs are generated using the best trial of DDS, VARS→DDS and DDS-Reduced algorithms. In order to illustrate the details more clearly, the hydrographs are reduced to the most challenging period, which is from December 1, 2008 to July 1, 2009. As can be seen, using the best solution of VARS→DDS1000, SWAT is able to adequately simulate the high flow periods (shown with red circles). In addition, when the budget is 10000, the best NSE values derived from VARS→DDS is better than the best NSE values from DDS and DDS-Reduced. Furthermore, Table 17 shows the similar best NSE values for little Rouge basin. The minor discrepancies between simulated and measured data in the low flow periods is due to the fact that NSE metric is focused on high flow periods. Therefore, the simulated hydrograph derived from VARS→DDS is able to reproduce the historical observed data more accurately, especially for the high flows periods.

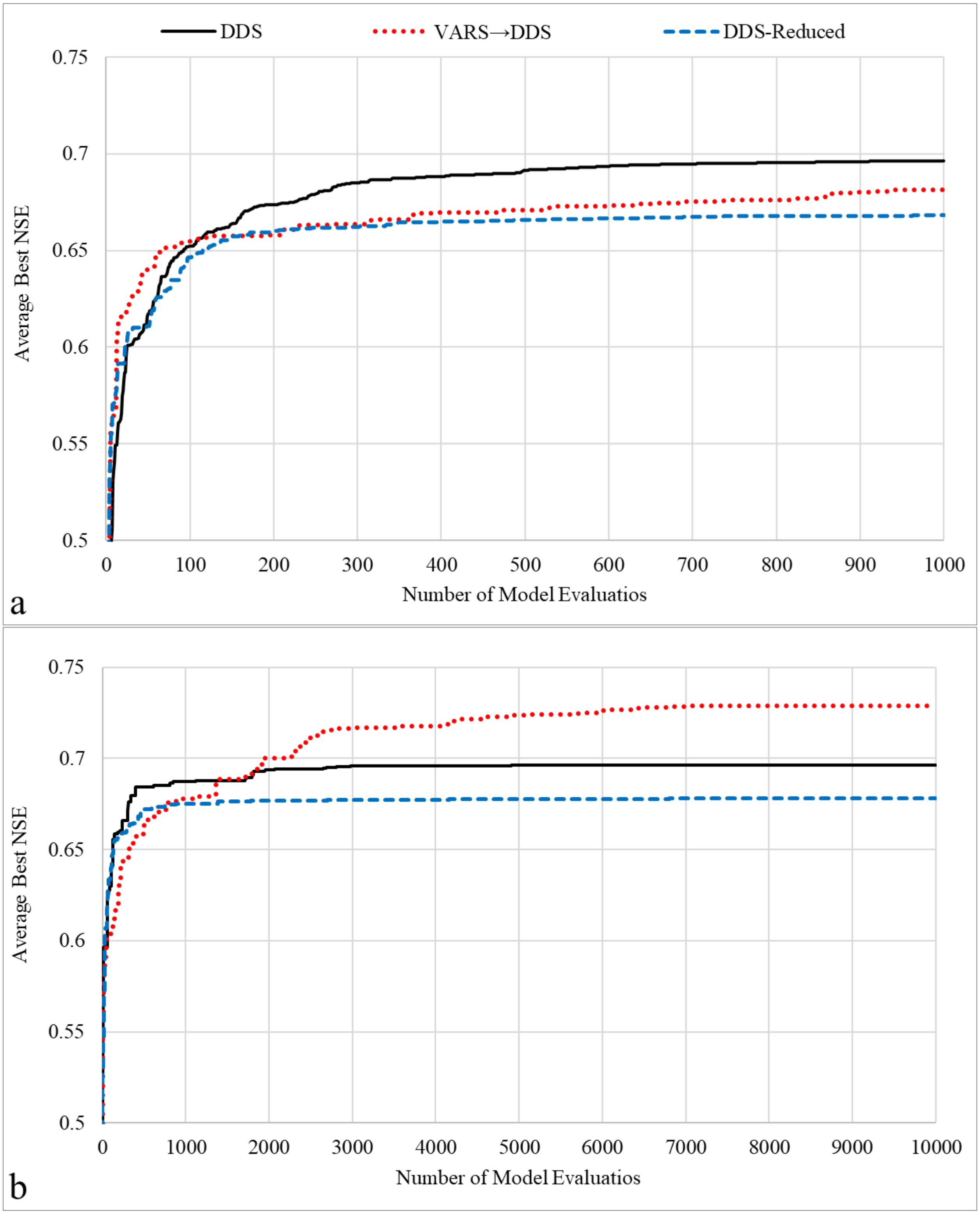


Figure 24. The convergence graphs of DDS, VARS→DDS, and DDS-Reduced applied to the calibration of the SWAT model of the Rouge River watershed (a) Convergence graph for the budget of 1000. (b) Convergence for the budget of 10000.

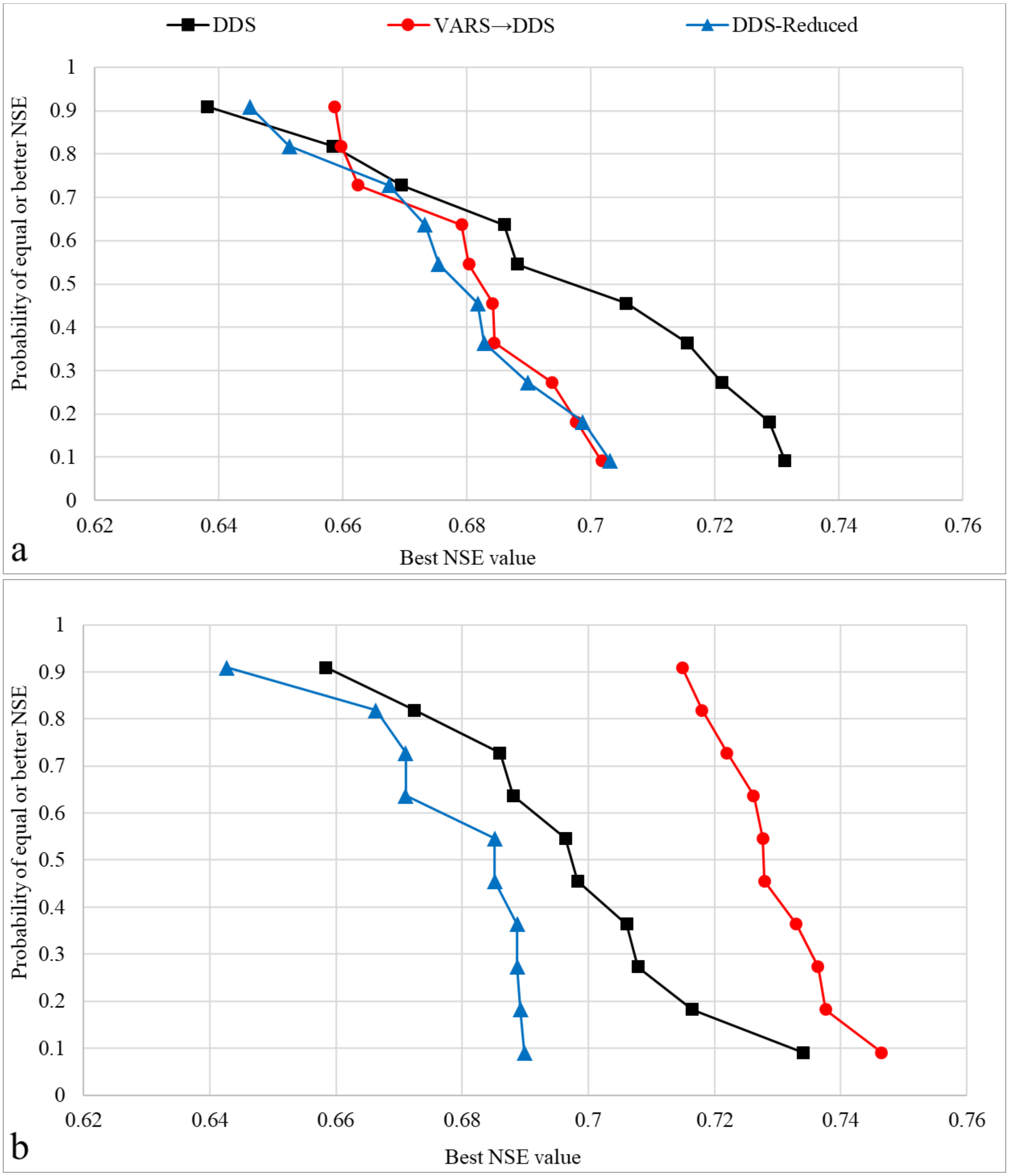


Figure 25. The first order stochastic dominance graphs of DDS, VARS→DDS, and DDS-Reduced applied to the calibration of the SWAT model of the Rouge River watershed (a) Budget of 1000. (b) Budget of 10000.

Table 15. Wilcoxon rank-sum test comparing DDS, DDS-Reduced and VARS→DDS applied to SWAT.

Computational Budgets	Compared Methods	P-value	H-value
1000	DDS versus VARS→DDS	0.212	0
	DDS versus DDS-Reduced	0.162	0
	DDS-Reduced versus DDS-VARS	0.733	0
10000	DDS versus VARS→DDS	0.0017	1
	DDS versus DDS-Reduced	0.045	1
	DDS-Reduced versus DDS-VARS	1.81E-04	1

Table 16. Average sensitivity ranks for SWAT parameters by VARS, Morris and Sobol with 1000 model evaluations (most versus least sensitive parameters are highlighted in red and, blue respectively).

Parameter	IVARS10	IVARS20	IVARS30	IVARS40	IVARS50	Morris	Sobol
CN	3	3	3	3	3	2	3
CNCOEF	21	21	21	21	21	20	22
SMFMN	22	22	22	22	22	21	23
SMFMX	10	10	10	10	10	9	11
TIMP	5	5	5	5	5	6	7
ESCO	7	7	7	7	7	5	5
EPCO	15	15	15	15	15	13	13
SURLAG	18	18	18	18	18	15	8
SOL_AWC	11	11	11	11	11	11	9
SOL_K	16	16	16	16	16	14	16
SOL_Z	6	6	6	6	6	7	6
GW_DELAY	23	23	23	23	23	22	20
GW_REVAP	12	12	12	12	12	12	10
ALPHA_BF	8	8	8	8	8	8	12
GWQMN	9	9	9	9	9	10	14
CH_N2	19	19	19	19	19	23	21
CH_K	4	4	4	4	4	4	4
SFTMP	13	13	13	13	13	19	18
SMTMP	20	20	20	20	20	16	17
RCHRG_DP	17	17	17	17	17	18	15
REVAPMN	14	14	14	14	14	17	19
SNOCOVMX	2	2	2	2	2	3	2
SNO50COV	1	1	1	1	1	1	1

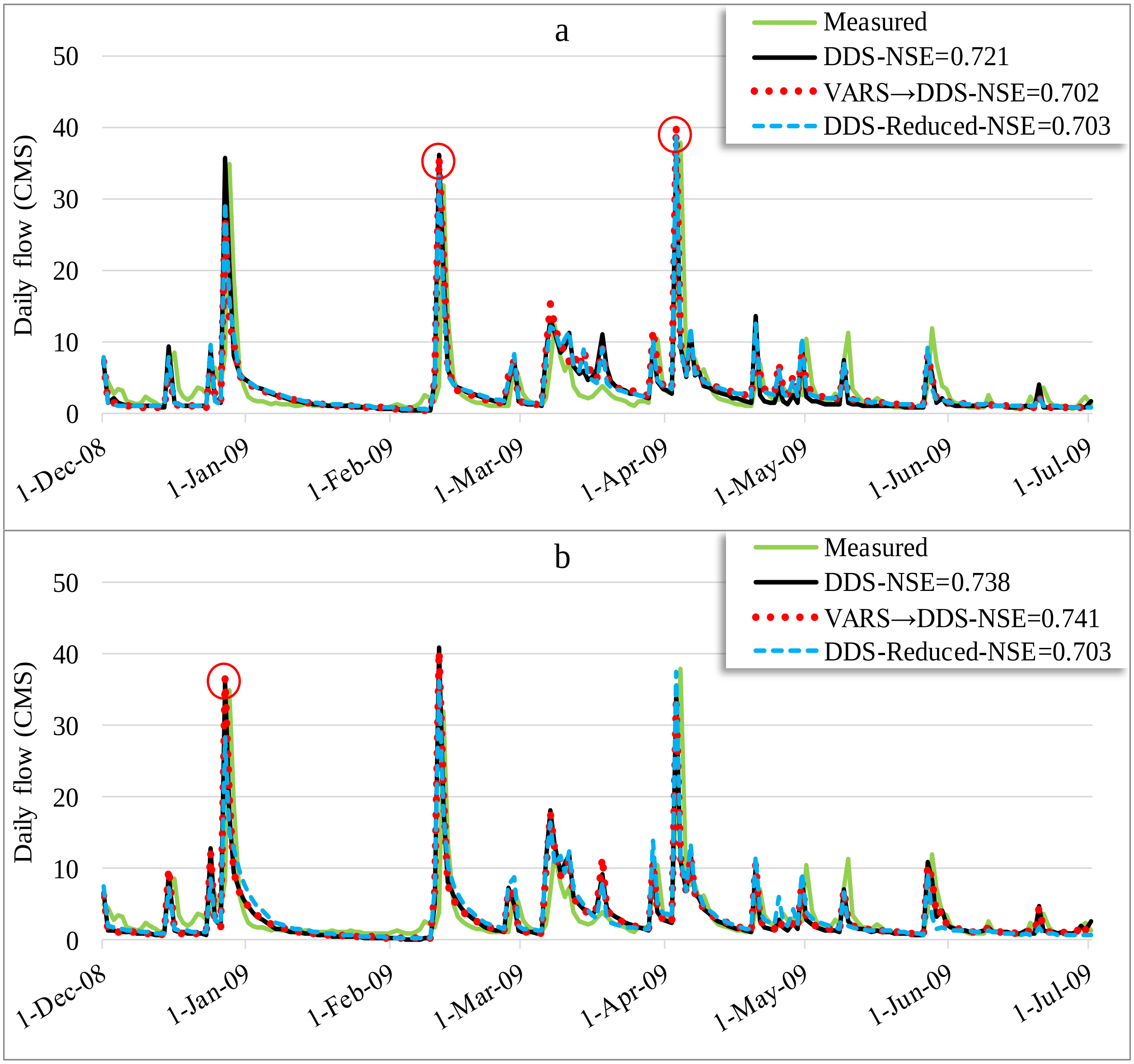


Figure 26. The comparison of streamflow time series for SWAT model for Main Rouge River basin from best parameter values generated by DDS, DDS-VARS and DDS-Reduced. (a) Optimization budget of 1000. (b) Optimization budget of 10000.

Table 17. The corresponding best NSE values for Little Rouge basin

Algorithm	NSE value
DDS-1000	0.720
DDS-Reduced-1000	0.693
VARS→DDS-1000	0.704
DDS-10000	0.728
DDS-Reduced-10000	0.718
VARS→DDS-10000	0.733

7.3.3WATFLOOD Odei River

Similar to SWAT model, VARS, Morris and Sobol GSA tools have generated analogues rankings for most and least sensitive parameters of WATFLOOD. As can be seen from Table 19, since the channel roughness (R2N) is the main parameter that controls streamflow, it is identified as the most sensitive parameter by all the three GSA methods. Moreover, since the studied basin has a sub-arctic climate, the snow related parameters such as, snowmelt rate (FM) and Sublimation factor (SUB) are among the most sensitive parameters. On the other hand, all GSA methods found the infiltration and snow-covered infiltration (AK and AKFS) as none important parameters in streamflow simulation. Hence, the performance of VARS with the limited budget of 1000 model evaluations is verified. The performance of VARS→DDS, DDS-Reduced and DDS algorithms are compared for calibrating the WATFLOOD model of the Odei River that has 37 parameters. According to the results of sensitivity in Table 19, the three GSA methods (VARS, Sobol and Morris) agree on the top 10 sensitive parameters of this model. Hence, when calibrating WATFLOOD, the i_{GPS} iteration of VARS→DDS is respectively at 155th and 828th iterations, for the budgets of 1000 and 10000 model evaluations.

Figure 27b illustrates that with the budget of 10000, VARS→DDS shows significantly faster convergence than DDS and DDS-Reduced. For instance, the average final best NSE value found by DDS and DDS-Reduced is equal to 0.41 and 0.4 respectively, while VARS→DDS found a slightly better solution after 2000 model evaluations. Furthermore, the average final best NSE value found by VARS→DDS is 0.5, which is 25 percent better than the solution found by DDS-Reduced and DDS. The best NSE values found in this thesis are lower from the best known value for this case study (0.60) reported by Holmes (2016). The main reason is that Holmes (2016) performed a manual calibration, while in this thesis, the automatic calibration is performed with random initial solutions to increase the challenge for the algorithms. Moreover, in order to reach higher NSE values for this case study, higher computational budget is required, which is out of scope of this research.

In addition, the stochastic dominance graphs in Figure 28-b confirm that VARS→DDS outperforms DDS and DDS-Reduced. Furthermore, the Wilcoxon rank-sum test results in Table 18 show that this preference is statistically significant. However, according to Figure 27-a, when the optimization budget is relatively limited (1000), VARS→DDS does not show a significant improvement against DDS or DDS-Reduced.

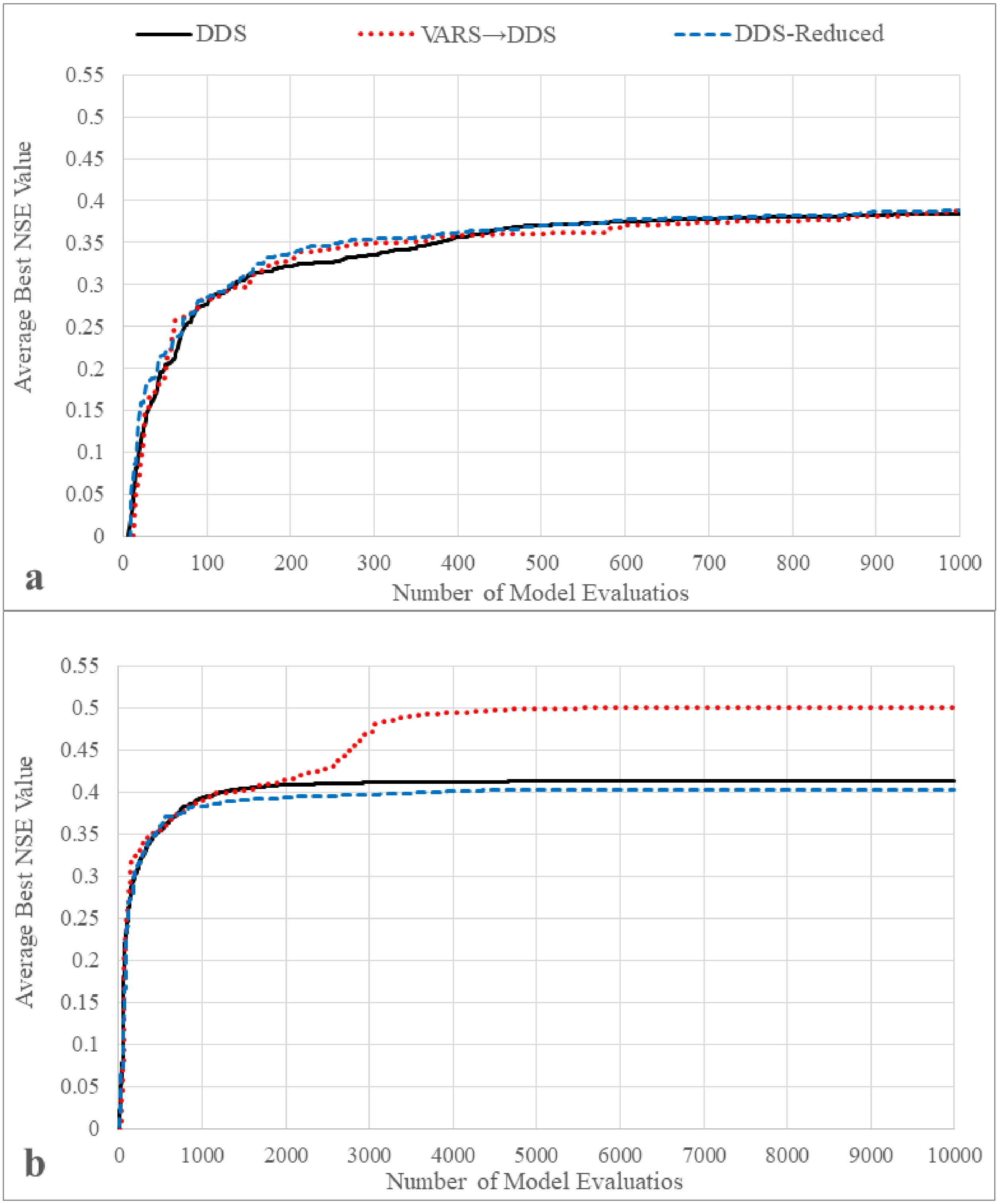


Figure 27. The average convergence graphs of DDS, VARS→DDS, and DDS-Reduced applied to the calibration of the WATFLOOD model of the Odei River watershed (a) Budget of 1000 evaluations. (b) Budget of 10000 evaluations.

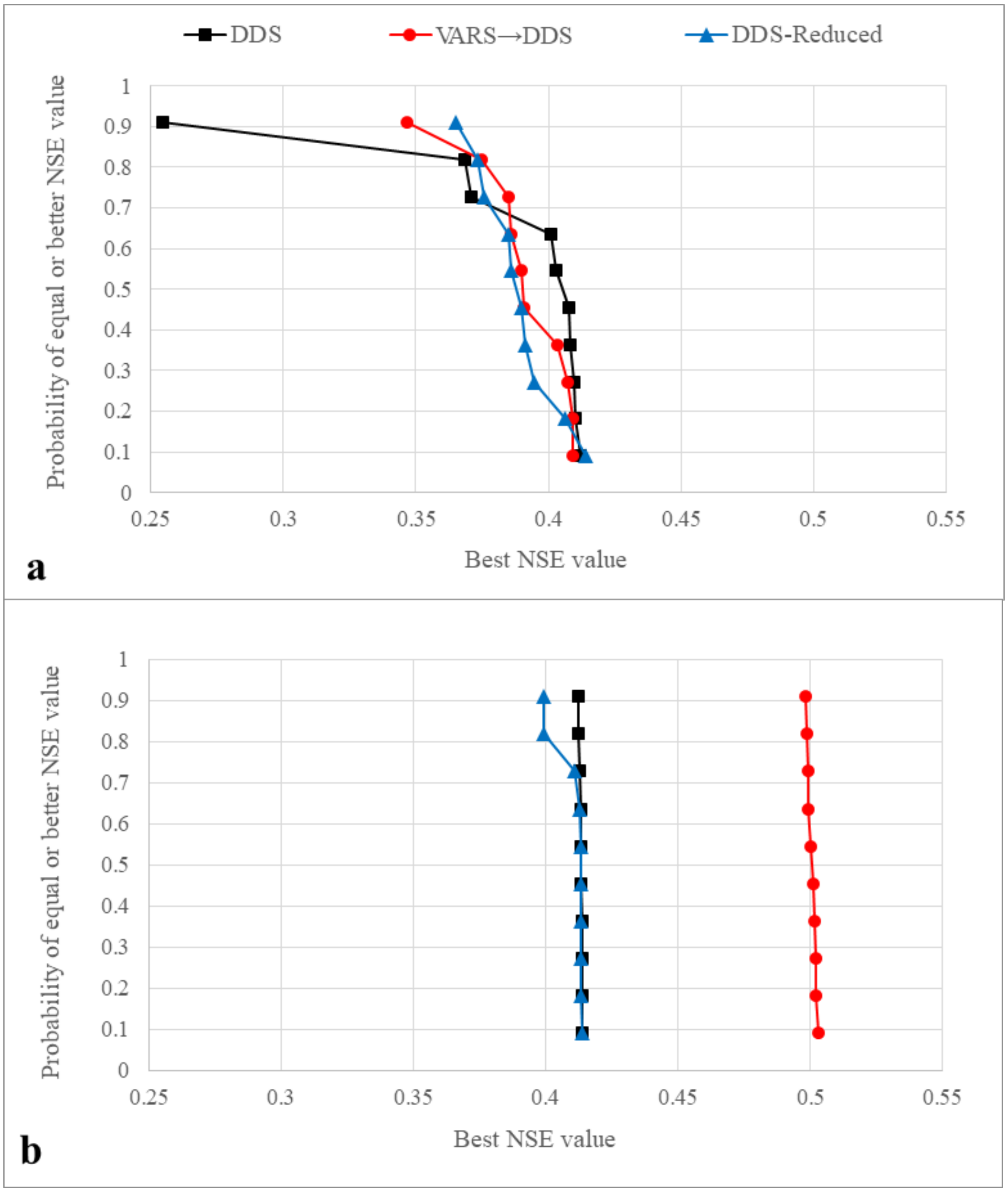


Figure 28. First order stochastic dominance graphs of DDS, VARS→DDS, and DDS-Reduced applied to the calibration of the WATFLOOD model of Odei River, (a) Budget of 1000. (b) Budget of 10000.

Table 18. Wilcoxon rank-sum test for the pairwise comparisons between DDS, VARS→DDS and DDS-Reduced applied to the WATFLOOD model calibration.

Computational Budget	**Compared Methods**	**P-value**	**H-value**
1000	DDS versus VARS→DDS	0.4727	0
	DDS versus DDS-Reduced	0.3447	0
	DDS-Reduced versus VARS→DDS	6.23E-01	0
10000	DDS versus VARS→DDS	1.83E-04	1
	DDS versus DDS-Reduced	0.2123	0
	DDS-Reduced versus VARS→DDS	1.83E-04	1

The comparison between simulated and observed hydrographs for WATFLOOD model in Figure 29 illustrates that when the computational budget is relatively large (10000 in this research), the calibration can benefit from the sensitivity-informed method VARS→DDS. As an illustration, the optimal parameter set found by best trial of VARS→DDS resulted in a hydrograph that has a better match with the observed data especially in the high and low flow periods. Nevertheless, the corresponding hydrographs from best trials of DDS and DDS-Reduced are unable to match with the historical data adequately and respectively underestimate and overestimate the low flow and high flow periods. On the other hand, when calibrating the model with the relative lower budgets (1000 evaluations), the improvement in the hydrograph of VARS→DDS is not significant. In this case, DDS applied to the full calibration problem is recommended.

Table 19. Average sensitivity ranks for WATFLOOD parameters by VARS, Morris and Sobol with 1000 model evaluations (most versus least sensitive parameters are highlighted in red and, blue respectively).

Parameter	Land Cover/Class	IVARS10	IVARS20	IVARS30	IVARS40	IVARS50	Morris	Sobol
PWR	N.A.	3	3	3	3	3	4	3
COEFF	N.A.	5	5	5	5	5	2	2
R2N	N.A.	1	1	1	1	1	1	1
THETA	N.A.	10	10	10	10	10	13	11
KCOND	N.A.	12	12	12	12	12	8	8
AK	Coniferous	33	33	33	33	33	33	24
AK	Mixed wood	32	32	32	32	32	27	25
AK	Shrub	34	34	34	34	34	34	26
AK	Wetland	21	21	21	21	21	18	34
AKFS	Coniferous	31	31	31	31	31	31	27
AKFS	Mixed wood	20	20	20	20	20	25	31
AKFS	Shrub	17	17	17	17	17	16	16
AKFS	Wetland	24	24	24	24	24	12	22
RETN	Coniferous	4	4	4	4	4	7	6
RETN	Mixed wood	22	22	22	22	22	11	29
RETN	Shrub	14	14	14	14	14	17	20
RETN	Wetland	30	30	30	30	30	20	33
R3	Coniferous	6	6	6	6	6	3	5
R3	Mixed wood	25	25	25	25	25	30	14
R3	Shrub	23	23	23	23	23	24	30
R3	Wetland	28	28	28	28	28	32	23
R3	Water	27	27	27	27	27	22	12
REC	Coniferous	11	11	11	11	11	14	18
REC	Mixed wood	19	19	19	19	19	23	15
REC	Shrub	15	15	15	15	15	15	19
REC	Wetland	26	26	26	26	26	28	28
FM	Coniferous	18	18	18	18	18	21	32
FM	Mixed wood	13	13	13	13	13	19	17
FM	Shrub	7	7	7	7	7	10	7
FM	Wetland	35	35	35	35	35	35	21
FM	Water	29	29	29	29	29	29	13
FPET	Water	2	2	2	2	2	5	4
SUB	Coniferous	9	9	9	9	9	6	9
SUB	Mixed wood	16	16	16	16	16	26	35
SUB	Shrub	8	8	8	8	8	9	10
SUB	Wetland	36	36	36	36	36	36	36
SUB	Water	37	37	37	37	37	37	37

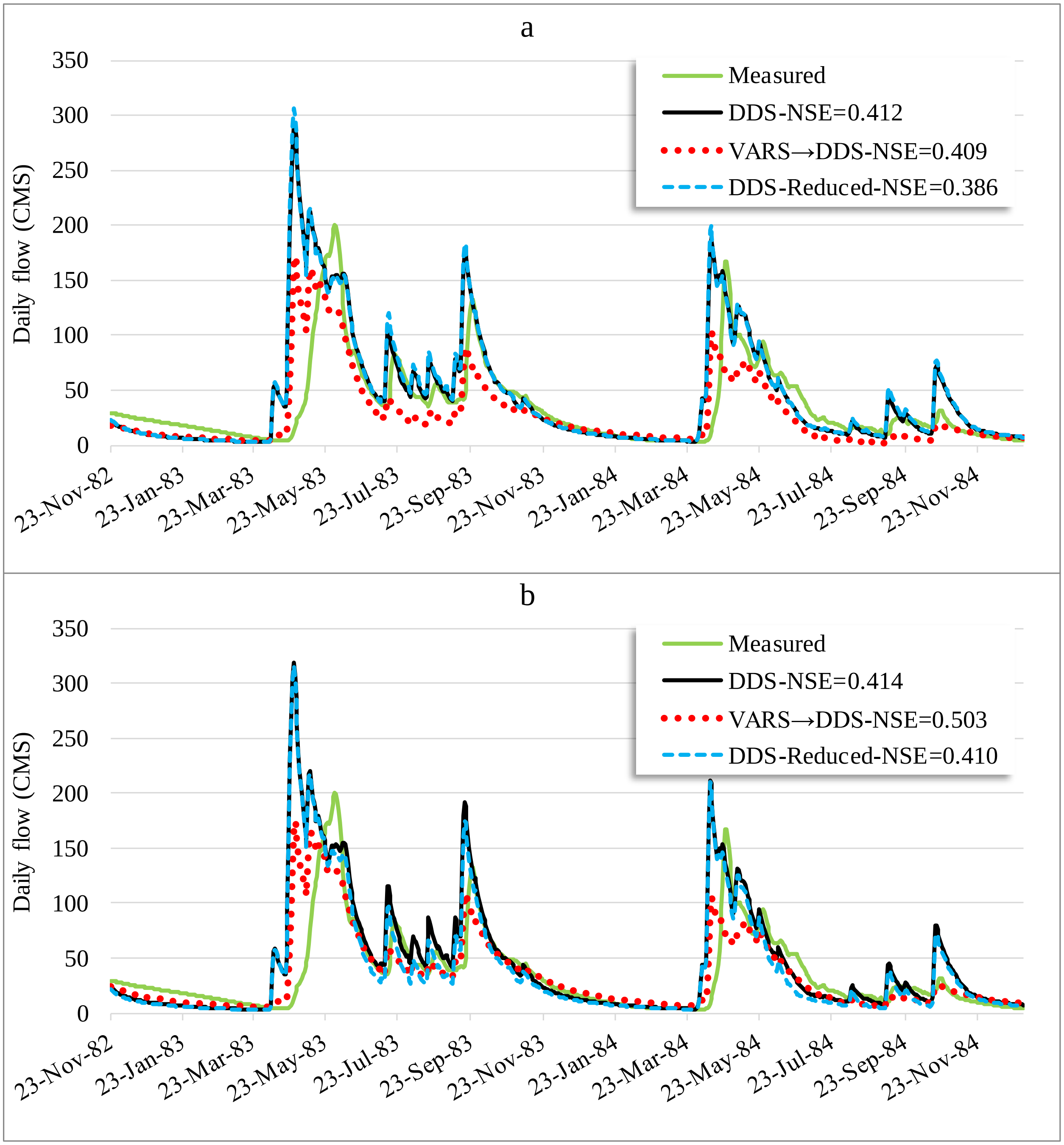


Figure 29. The comparison of simulated and observed hydrographs for WATFLOOD model for Odei River basin from the best parameter values generated by DDS, VARS→DDS and DDS-Reduced. (a) Optimization budget of 1000. (b) Optimization budget of 10000.

7.3.4VARS→DDS Results Discussion

VARS→DDS is successfully applied to different hydrological model calibration problems and test functions. Results show that VARS→DDS has a substantially better convergence and better final solution when the optimization budget is relatively high equal, 10000 model evaluations in this research. Moreover, the benefit of VARS→DDS compared to other automatic calibrations (DDS and DDS reduced) is more significant when applied to more complex problems, WATFLOOD calibration problem in this research. This is due to the higher number of calibration parameters in WATFLOOD compared to other cases of thesis. As the number of calibration parameters increases, DDS is expected to lose its efficiency, because its parameter selection is purely random. The sensitivity-based parameter selection by VARS→DDS increases the chance of the more sensitive parameters to be selected and therefore helps to increase the convergence of the algorithm. Results also suggest that the common practice in the literature that reduces the calibration problem only to the most sensitive parameters is not necessarily beneficial compared to calibrating all parameter using an efficient optimization algorithm such as DDS.

8 Conclusions and Future Works

8.1 Conclusions

Due to the increasing complexity of hydrological models, the automatic calibration of these models is becoming more and more challenging. One of the common way to cope with this issue is applying global sensitivity analysis (GSA) in hydrological model calibration to reduce the number of calibration parameters. However, as shown in this thesis, this technique is expected to result in a sub-optimal calibrated model for distributed and semi-distributed models that have a large number of calibration parameters. In this thesis, a novel and efficient methodology is developed to incorporate parameter sensitivity in automatic calibration without reducing the calibration problem to the most sensitive parameters. The proposed approach gives higher priority to more sensitive parameters to be selected and perturbed in the automatic calibration process; however, it does not eliminate the chance for selecting and perturbing less sensitive parameters.

Among the three implementations of sensitivity-informed model calibration considered in this thesis, VARS→DDS improves both efficiency and effectiveness of model calibration. The results illustrate that VARS→DDS outperforms DDS and DDS-Reduced in calibrating complex hydrological models. As an illustration, compared to DDS and DDS-Reduced, VARS→DDS shows significantly faster convergence, and increases the NSE value respectively by 25% and 9% wen calibrating WATFLOOD and SWAT models. In addition, the first order stochastic dominance and Wilcoxon rank sum test illustrated that VARS→DDS is unambiguously preferred over DDS and DDS-Reduced in calibrating WATFLOOD and SWAT models. Furthermore, the promising performance of VARS→DDS has revealed that if calibration of complex hydrological models such as SWAT, is performed with adequate budget, VARS→DDS can identify better solutions compared to the solutions that had been reported in the literature.

Results also shows that in order to observe the benefit of VARS→DDS, two conditions need to be met. First, the models should have a large number of calibration parameters. Second, the calibration must be performed computational budgets that are deemed large enough, 10000 solution evaluations in this thesis. Nonetheless, if these two criteria are not met, VARS→DDS is still recommended, as the results of this approach are similar or slightly better than original DDS. In addition, VARS→DDS can provide parameter sensitivity information at the end of calibration, which is useful for future model recalibration.

8.2 Lessons Learned

The other implementations of incorporating sensitivity analysis in optimization, GSA↔DDS and GSA→DDS failed to meet the objectives of this research and could not increase the efficiency and effectiveness of the automatic model calibration. For instance, when calibrating SAC-SMA model, Sobol↔DDS decreases the efficiency of optimization by 20% compared to DDS due to the inaccurate sensitivity results, and Morris↔DDS produces similar results to DDS. Additionally, in GSA↔DDS, individual perturbation of selected parameters degrade the efficiency of optimization as in original DDS, the selected parameters are perturbed simultaneously. Moreover, updating the sensitivity results too frequently using the dependent solutions of DDS reduced the accuracy of sensitivity results and performance of DDS in GSA↔DDS method. On the other hand, GSA→DDS approach illustrates faster convergence compared to DDS. As an illustration, when calibrating SAC-SMA model, GSA→DDS improves the convergence rate by approximately 42% compared to DDS. However, performing sensitivity analysis at initial iterations of optimization prevents GSA→DDS to improve the final best solution compared to DDS. The main reason is that DDS has a high convergence rate at initial iterations, and performing

GSA at these iterations reduces the convergence rate of DDS, as GSA methods are not designed for optimization.

8.3 Limitations

Research performed in this thesis faced the following limitations that could be addressed in future research:

- The only optimization algorithm considered in this thesis is DDS. Comparing the performance of VARS→DDS with only DDS algorithm prevents from reaching to a comprehensive statement that sensitivity-informed optimization outperforms normal optimization methods.
- The only performance metric for hydrological models considered in this research is Nash Sutcliff Efficiency. In hydrological model calibration, NSE metric is focused on high flow events. Thus, preforming GSA with respect to NSE will provide a limited measure of parameters sensitivity. In addition, the model may not be effectively calibrated to simulate low flow events.
- The maximum optimization budget in this thesis is limited to 10000 evaluations. As shown in this thesis, even with the sensitivity-informed optimization, complex models such as WATFLOOD requires higher computational budget to be effectively calibrated in terms of finding better NSE value. Thus, with the maximum optimization budget of 10000 the ability of VARS→DDS algorithm on effective calibration of WATFLOOD model cannot be proved.
- Calibrating only a limited number of complex hydrological models in this thesis (only one distributed and one semi-distributed model is considered) prevents from reaching to a comprehensive conclusion in which, sensitivity-informed

optimization is preferred over calibrating complex hydrological models with full and reduced parameter set.

- Too frequent updating of sensitivity results in GSA↔DDS method with the dependent solutions generated by DDS. In DDS, the selected parameters are perturbed simultaneously. However, updating sensitivity results at each iteration requires individual perturbation of selected parameters, in which reduces the performance of DDS algorithm in finding good quality solutions.
- VARS→DDS shows no improvement in the efficiency and effectiveness of model calibration, when the computational budget is limited. The main reason for this behavior is that in VARS→DDS, the real benefit of sensitivity information is revealed after a certain number of iterations, as the process of selecting most sensitive perimeters is probabilistic.

8.4 Recommendation for Future Work

In order to address the mentioned limitations of this research, the following recommendations should be considered.

- It would be advantageous to compare the performance of VARS→DDS with other advanced single-objective global optimization algorithms such as Shuffle Complex Evolutionary (SCE) algorithm (Duan et al., 1993).
- In order to further validate the results of VARS→DDS, it is best to test this approach on calibrating different complex hydrological models such as, VIC (Liang et al., 1994), HBV-96 (Lindström et al., 1997) and MESH (Pietroniro et al., 2007) with applying on various basins across the world.

- The performance of VARS→DDS with the limited computational budget should be improved. This can be achieved by increasing the efficiency of VARS→DDS by implementing it in parallel format, and enhancing the accuracy of sensitivity results.
- Developing a new method that can update parameter sensitivity indices properly in GSA↔DDS approach. For instance, one could rerun initial GSA to update the sensitivity indices after a certain number of iterations. In this way, both accuracy of sensitivity results and performance of GSA↔DDS method can be improved.
- Developing a new GSA method that is designed for both sensitivity analysis and optimization. In this way, when GSA is performed at initial iterations, the performance of DDS is not degraded. For instance, a GSA method that has a sampling procedure similar to DDS can significantly reduce the initial degradation in performance of DDS algorithm, in GSA↔DDS and GSA→DDS approaches.
- Developing a procedure to update VARS sensitivity indices using the solutions generated by DDS, would significantly enhance the efficiency of VARS→DDS method especially when calibrating complex models with limited budget.
- The added complexity of hydrological models requires utilization of multi-objective optimization in model calibration. Thus, GSA toolboxes such as VARS should be incorporated with efficient multi-objective optimization algorithms such as PA-DDS (Asadzadeh, 2012), to perform the calibration with multiple performance metrics and improve the performance of sensitivity-informed model calibration.
- To properly identify most sensitive parameters in hydrological model calibration, VARS should be performed over different time periods (season), as parameters of hydrological models could become important only during a period of time. For

example, the sensitivity of model output to snow melt rate parameter significantly increases during snow melt events. This information can be used by users to properly define the number of most sensitive parameters in VARS→DDS method that can further improve the results of this algorithm, as the sensitivity information is added at i_{GPS} iteration.

- In hydrological models with large number of parameters, the roulette wheel algorithm may mislead VARS→DDS due to the significant difference in parameter sensitivity index. Hence, utilizing a transformation method to decrease the magnitude of difference between parameter sensitivity values is recommended. In addition, among the different transformation methods Box-Cox transformation (Box and Cox, 1964) is desirable, as the degree of transformation can be modified.
- To improve the calibration results, VARS→DDS can be performed with different performance metric or a combination of them to properly increase the focus of calibration on all aspects of hydrograph (i.e. high flow and low flow events).

References

Abbaspour, K. C., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., Yang, H., & Kløve, B. (2015). A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. Journal of Hydrology, 524, 733-752.

Abraham, A., & Jain, L. (2005). Evolutionary multiobjective optimization. In Evolutionary Multiobjective Optimization (pp. 1-6). Springer, London.

Ali, M. M., Khompatraporn, C., & Zabinsky, Z. B. (2005). A numerical evaluation of several stochastic algorithms on selected continuous global optimization test problems. Journal of global optimization, 31(4), 635-672.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment part I: model development 1. JAWRA Journal of the American Water Resources Association, 34(1), 73-89.

Arsenault, R., Poulin, A., Côté, P., & Brissette, F. (2013). Comparison of stochastic optimization algorithms in hydrological model calibration. Journal of Hydrologic Engineering, 19(7), 1374-1384.

Asadzadeh Esfahani, M. (2012). Developing Parsimonious and Efficient Algorithms for Water Resources Optimization Problems.

Asadzadeh, M., & Tolson, B. (2013). Pareto archived dynamically dimensioned search with hypervolume-based selection for multi-objective optimization. Engineering Optimization, 45(12), 1489-1509.

Asadzadeh, M., Tolson, B. A., & Burn, D. H. (2014). A new selection metric for multiobjective hydrologic model calibration. Water Resources Research, 50(9), 7082-7099.

Asadzadeh, M., Leon, L., McCrimmon, C., Yang, W., Liu, Y., Wong, I., ... & Bowen, G. (2015). Watershed derived nutrients for Lake Ontario inflows: Model calibration considering typical land operations in Southern Ontario. Journal of Great Lakes Research, 41(4), 1037-1051.

Barney, B. (2010). Introduction to parallel computing. Lawrence Livermore National Laboratory, 6(13), 10.

Beven, K. J. (2001): Rainfall-runoff modeling – The primer, Wiley: Chichester, UK.

Beven, K. J. (2010). Preferential flows and travel time distributions: defining adequate hypothesis tests for hydrological process models. Hydrological Processes, 24(12), 1537-1547.

Beven, K. J. (1989), Changing ideas in hydrology—The case of physicallybased models, J. Hydrol., 105, 157–172.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. Journal of the Royal Statistical Society: Series B (Methodological), 26(2), 211-243.

Box, G. E., & Jenkins, G. M. (1976). Time series analysis: forecasting and control, revised ed. Holden-Day.

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

Cacuci, D. G., & Ionescu-Bujor, M. (2004). A comparative review of sensitivity and uncertainty analysis of large-scale systems—II: statistical methods. Nuclear science and engineering, 147(3), 204-217.

Campolongo, F., J. Cariboni, and A. Saltelli (2007), An effective screening design for sensitivity analysis of large models, Environ. Modell. Software, 22(10), 1509–1518.

Campolongo, F., & Saltelli, A. (1997). Sensitivity analysis of an environmental model: an application of different analysis methods. Reliability Engineering & System Safety, 57(1), 49-69.

Carpenter, T. M., K. P. Georgakakos, and J. A. Sperfslagea (2001), On the parametric and NEXRAD-radar sensitivities of a distributed hydrologic model suitable for operational use, J. Hydrol., 253, 169–193.

Carpenter, T. M., & Georgakakos, K. P. (2006). Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales. Journal of hydrology, 329(1-2), 174-185.

Cheng, C. T., X. Y. Wu, and K. W. Chau (2005), Multiple criteria rainfall-runoff model calibration using a parallel genetic algorithm in a cluster of computers, Hydrol. Sci. J., 50(6), 1069–1087, doi:10.1623/hysj.2005.50.6.1069.

Christopher Frey, H., & Patil, S. R. (2002). Identification and review of sensitivity analysis methods. Risk analysis, 22(3), 553-578.

Chu, J., Zhang, C., Fu, G., Li, Y., & Zhou, H. (2015). Improving multi-objective reservoir operation optimization with sensitivity-informed dimension reduction. Hydrology and Earth System Sciences, 19, 3557-3570.

Crawford, N. H., & Linsley, R. K. (1966). Digital Simulation in Hydrology'Stanford Watershed Model 4.

Cukier, R. I., C. M. Fortuin, K. E. Shuler, A. G. Petschek, and J. H. Schailby (1973), ‘‘Study of the Sensitivity of the Coupled Reaction Systems to Uncertainties in Rate Coefficients: I. Theory,’’ Journal of Chemical Physics, 59(8):3873–3878.

Dawdy, D. R., & O'Donnell, T. (1965). Mathematical models of catchment behavior. Journal of the Hydraulics Division, 91(4), 123-137.

Deb, K. (2014). Multi-objective optimization. In Search methodologies (pp. 403-449). Springer, Boston, MA.

Devia, G. K., Ganasri, B. P., & Dwarakish, G. S. (2015). A review on hydrological models. Aquatic Procedia, 4, 1001-1007.

Duan, Q. Y., Gupta, V. K., & Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. Journal of optimization theory and applications, 76(3), 501-521.

Duan, Q., Sorooshian, S., & Gupta, V. K. (1994). Optimal use of the SCE-UA global optimization method for calibrating watershed models. Journal of hydrology, 158(3-4), 265-284.

Farmer, W. H., & Vogel, R. M. (2016). On the deterministic and stochastic use of hydrologic models. Water Resources Research, 52(7), 5619-5633.

Fishman, G. (2013). Monte Carlo: concepts, algorithms, and applications. Springer Science & Business Media.

Fortin, J. P., Turcotte, R., Massicotte, S., Moussa, R., Fitzback, J., & Villeneuve, J. P. (2001). Distributed watershed model compatible with remote sensing and GIS data. I: Description of model. Journal of Hydrologic Engineering, 6(2), 91-99.

Feyen, L., J. A. Vrugt, B. O. Nuallain, J. van der Knijff, and A. De Roo (2007), Parameter optimisation and uncertainty assessment for largescale streamflow simulation with the LISFLOOD model, J. Hydrol., 332(3–4), 276–289, doi:10.1016/j.jhydrol.2006.07.004.

Gholami, A., Roshan, M. H., Shahedi, K., Vafakhah, M., & Solaymani, K. (2016). Hydrological stream flow modeling in the Talar catchment (central section of the Alborz Mountains, north of Iran): Parameterization and uncertainty analysis using SWAT-CUP. Journal of Water and Land Development, 30(1), 57-69.

Gabric, A. J., Ayers, G., Murray, C. N., & Parslow, J. (1996). Use of remote sensing and mathematical modelling to predict the flux of dimethylsulfide to the atmosphere in the Southern Ocean. Advances in Space Research, 18(7), 117-128.

Gibbons, J. D., & Chakraborti, S. (2011). Nonparametric statistical inference (pp. 977-979). Springer Berlin Heidelberg.

Goldberg, D. E. (1989). Genetic algorithms in search. Optimization and MachineLearning.

Griewank, A. O. (1981). Generalized descent for global optimization. Journal of optimization theory and applications, 34(1), 11-39.

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

Hadka, D., & Reed, P. (2012). Diagnostic assessment of search controls and failure modes in many-objective evolutionary optimization. Evolutionary Computation, 20(3), 423-452.

Han, W., Yang, P., Ren, H., & Sun, J. (2010, December). Comparison study of several kinds of inertia weights for PSO. In 2010 IEEE International Conference on Progress in Informatics and Computing (Vol. 1, pp. 280-284). IEEE.

Helton, J. C., & Davis, F. J. (2000). Sampling-based methods. Sensitivity analysis, 101-153.

Hendrickson, J. D., Sorooshian, S., & Brazil, L. E. (1988). Comparison of Newton-type and direct search algorithms for calibration of conceptual rainfall-runoff models. Water Resources Research, 24(5), 691-700.

Holmes, T. L. (2016). Assessing the Value of Stable Water Isotopes in Hydrologic Modeling: A Dual-Isotope Approach (Doctoral dissertation, MSc Thesis, Department of Civil Engineering, University of Manitoba, Manitoba).

Holland, J. H. (1975). Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. MIT press.

Holvoet, K., van Griensven, A., Seuntjens, P., & Vanrolleghem, P. A. (2005). Sensitivity analysis for hydrology and pesticide supply towards the river in SWAT. Physics and Chemistry of the Earth, Parts A/B/C, 30(8), 518-526.

Hooke, R. and T. A. Jeeves (1961): Direct search solutions of numerical and statistical problems.

Journal of the Association for Computing Machinery, 8(2), 212–229.

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

Iman, R. L., & Conover, W. J. (1980). Small sample sensitivity analysis techniques for computer models. with an application to risk assessment. Communications in statistics-theory and methods, 9(17), 1749-1842.

Ionescu-Bujor, M., & Cacuci, D. G. (2004). A comparative review of sensitivity and uncertainty analysis of large-scale systems—i: Deterministic methods. Nuclear science and engineering, 147(3), 189-203.

Johnston, P. R., & Pilgrim, D. H. (1976). Parameter optimization for watershed models. Water Resources Research, 12(3), 477-486.

Jung, B. S., Karnev, B. W., & Lambert, M. F. (2006). Benchmark tests of evolutionary algorithms: mathematic evaluation and application to water distribution systems. Journal of Environmental Informatics, 7(1), 24-35.

Khatun, S., Sahana, M., Jain, S. K., & Jain, N. (2018). Simulation of surface runoff using semi distributed hydrological model for a part of Satluj Basin: parameterization and global sensitivity analysis using SWAT CUP. Modeling Earth Systems and Environment, 1-14.

Khu, S. T., D. Savic, Y. Liu, and H. Madsen (2004), A fast evolutionary based metamodelling approach for the calibration of a rainfall-runoff model, paper presented at First Biennial Meeting, Int. Environ. Modell. and Software Soc., Osnabruck, Germany.

Knowles, J. (2006). ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. IEEE Transactions on Evolutionary Computation, 10(1), 50-66.

Kouwen, N. (1988). WATFLOOD: a micro-computer based flood forecasting system based on real-time weather radar. Canadian Water Resources Journal, 13(1), 62-77.

Kucherenko, S., M. Rodriguez-Fernandez, C. Pantelides, and N. Shah (2009), Monte Carlo evaluation of derivative-based global sensitivity measures, Reliab. Eng. Syst. Safety, 94(7), 1135–1148, doi:10.1016/ j.ress.2008.05.006.

Levy, H. (1992). Stochastic dominance and expected utility: survey and analysis. Management science, 38(4), 555-593.

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. Journal of Geophysical Research: Atmospheres, 99(D7), 14415-14428.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., & Bergström, S. (1997). Development and test of the distributed HBV-96 hydrological model. Journal of hydrology, 201(1-4), 272-288.

Lu, Z., Zou, S., Xiao, H., Zheng, C., Yin, Z., & Wang, W. (2015). Comprehensive hydrologic calibration of SWAT and water balance analysis in mountainous watersheds in northwest China. Physics and Chemistry of the Earth, Parts A/B/C, 79, 76-85.

Madsen, H. (2003). Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. Advances in water resources, 26(2), 205-216.

Malagò, A., Pagliero, L., Bouraoui, F., & Franchini, M. (2015). Comparing calibrated parameter sets of the SWAT model for the Scandinavian and Iberian peninsulas. Hydrological Sciences Journal, 60(5), 949-967.

Matott, L. S., Tolson, B. A., & Asadzadeh, M. (2012). A benchmarking framework for simulation-based optimization of environmental models. Environmental Modelling & Software, 35, 19-30.

McCuen, R. H. (1973). The role of sensitivity analysis in hydrologic modeling. Journal of Hydrology, 18(1), 37-53.

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics, 21(2), 239-245.

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments, Technometrics, 33(2), 161–174.

Moradkhani, H., & Sorooshian, S. (2009). General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis. In Hydrological modelling and the water cycle (pp. 1-24). Springer, Berlin, Heidelberg.

Muleta, M. K., & Nicklow, J. W. (2005). Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. Journal of hydrology, 306(1), 127-145.

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(11), 1604.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. Journal of hydrology, 10(3), 282-290.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R., & Williams, J. R. (2011). Soil and water assessment tool theoretical documentation version 2009. Texas Water Resources Institute.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. The computer journal, 7(4), 308-313.

Pardalos, P. M., Žilinskas, A., & Žilinskas, J. (2017). Non-convex multi-objective optimization. Springer International Publishing.

Pianosi, F., Sarrazin, F., & Wagener, T. (2015). A Matlab toolbox for global sensitivity analysis. Environmental Modelling & Software, 70, 80-85.

Pickup, G. (1977). TESTING THE EFFICIENCY OF ALGORITHMS AND STRATEGIES FOR AUTOMATIC CALIBRATION OF RAINFALL-RUNOFF MODELS/Essais de l'efficacité des algorithms et des stratégies pour l'étalonnage des modèles pluie-écoulement. Hydrological Sciences Journal, 22(2), 257-274.

Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B., ... & Pellerin, P. (2007). Development of the MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale. Hydrology and Earth System Sciences Discussions, 11(4), 1279-1294.

Pina, R., Ochoa-Rodriguez, S., Simões, N., Mijic, A., Sa Marques, A., & Maksimovic, Č. (2014). Semi-distributed or fully distributed rainfall-runoff models for urban pluvial flood modelling?

Rastrigin, L. A. (1974). Extremal control systems. Theoretical foundations of engineering cybernetics series, 3.

Razavi, S., Tolson, B. A., Matott, L. S., Thomson, N. R., MacLean, A., & Seglenieks, F. R. (2010). Reducing the computational cost of automatic calibration through model preemption. Water Resources Research, 46(11).

Razavi, S., & Gupta, H. V. (2015). What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in Earth and Environmental systems models. Water Resources Research, 51(5), 3070-3092.

Razavi, S., & Gupta, H. V. (2016). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. Water Resources Research, 52(1), 423-439.

Razavi, S., & Gupta, H. V. (2016b). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application. Water Resources Research, 52(1), 440-455.

Rakovec, O., Hill, M. C., Clark, M. P., Weerts, A. H., Teuling, A. J., & Uijlenhoet, R. (2014). Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models. Water Resources Research, 50(1), 409-426.

Sahu, R. K., Mishra, S. K., & Eldho, T. I. (2012). Performance evaluation of modified versions of SCS curve number method for two watersheds of Maharashtra, India. ISH Journal of Hydraulic Engineering, 18(1), 27-36.

Sakata, S., Ashida, F., & Zako, M. (2003). Structural optimization using Kriging approximation. Computer methods in applied mechanics and engineering, 192(7-8), 923-939.

Saltelli, A., Ratto, M., Andres, T., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). Global sensitivity analysis: the primer. John Wiley & Sons.

Sanchez-Vila, X., & Fernàndez-Garcia, D. (2016). Debates—Stochastic subsurface hydrology from theory to practice: Why stochastic modeling has not yet permeated into practitioners?. Water Resources Research, 52(12), 9246-9258.

Seibert, J. (2000). Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. Hydrology and Earth System Sciences Discussions, 4(2), 215-224.

Schuol, J., & Abbaspour, K. C. (2006). Calibration and uncertainty issues of a hydrological model (SWAT) applied to West Africa. Advances in geosciences, 9, 137-143.

Schutte, J. F., Reinbolt, J. A., Fregly, B. J., Haftka, R. T., & George, A. D. (2004). Parallel global optimization with the particle swarm algorithm. International journal for numerical methods in engineering, 61(13), 2296-2315.

Shoemaker, C. A., Regis, R. G., & Fleming, R. C. (2007). Watershed calibration using multistart local optimization and evolutionary optimization with radial basis function approximation. Hydrological sciences journal, 52(3), 450-465.

Smith, A., Welch, C., and Stadnyk, T., 2016. Assessment of a lumped coupled flow-isotope model in data scarce Boreal catchments. Hydrological Processes, 30, 3871-3884.

Smith, M. B., Laurine, D. P., Koren, V. I., Reed, S. M., & Zhang, Z. (2003). Hydrologic model calibration in the National Weather Service. Calibration of watershed models, 133-152.

Sobol, I. M. (1990). On sensitivity estimation for nonlinear mathematical models, Matematicheskoe Modelirovanie, 2(1), pp. 112–118.

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 resources research, 29(4), 1185-1194.

Sorooshian S. and V. K. Gupta (1995): Model calibration. In: Singh, V. P. (Ed.), Computer
Models of Watershed Hydrology, Water Resources Publications, Highlands Ranch, CO,
pp. 23–67.

Spear, R. C., & Hornberger, G. M. (1980). Eutrophication in Peel Inlet—II. Identification of critical uncertainties via generalized sensitivity analysis. Water Research, 14(1), 43-49.

Tang, Y., Reed, P., & Wagener, T. (2005). How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration?. Hydrology and Earth System Sciences Discussions, 2(6), 2465-2520.

Tang, T., Reed, P., Wagener, T., & Van Werkhoven, K. (2006). Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. Hydrology and Earth System Sciences Discussions, 3(6), 3333-3395.

Tang, Y., Reed, P., Van Werkhoven, K., & Wagener, T. (2007). Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis. Water Resources Research, 43(6).

Teshager, A. D., Gassman, P. W., Secchi, S., Schoof, J. T., & Misgna, G. (2016). Modeling agricultural watersheds with the soil and water assessment tool (swat): Calibration and validation with a novel procedure for spatially explicit hrus. Environmental management, 57(4), 894-911.

Tolson, B. A., & Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resources Research, 43(1).

Tolson, B. A., Asadzadeh, M., Maier, H. R., & Zecchin, A. (2009). Hybrid discrete dynamically dimensioned search (HD-DDS) algorithm for water distribution system design optimization. Water Resources Research, 45(12).

Toronto and Region Conservation Authority (TRCA), 2012. Approved Assessment Report:
Toronto and Region Source Protection Area. Volume 1.

Unduche, F., Tolossa, H., Senbeta, D., & Zhu, E. (2018). Evaluation of four hydrological models for operational flood forecasting in a Canadian Prairie watershed. Hydrological Sciences Journal.

USACE-HEC, 2016. Hydrologic modeling system HEC-HMS, user’s manual, version 4.2. Davis, CA: US Army Corps of Engineers, Hydrologic Engineering Center.

Van Werkhoven, K., Wagener, T., Reed, P., & Tang, Y. (2009). Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. Advances in Water Resources, 32(8), 1154-1169.

Van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., & Srinivasan, R. (2006). A global sensitivity analysis tool for the parameters of multi-variable catchment models. Journal of hydrology, 324(1-4), 10-23.

Wagener, T., & Kollat, J. (2007). Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox. Environmental Modelling & Software, 22(7), 1021-1033.

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

Westerberg, I. K., Guerrero, J. L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., ... & Xu, C. Y. (2011). Calibration of hydrological models using flow-duration curves. Hydrology and Earth System Sciences, 15(7), 2205-2227.

White, K. L., & Chaubey, I. (2005). Sensitivity analysis, calibration, and validations for a multisite and multivariable SWAT model 1. JAWRA Journal of the American Water Resources Association, 41(5), 1077-1089.

Wright, S. (1921). Correlation and causation. Journal of agricultural research, 20(7), 557-585.

Yaduvanshi, A., Srivastava, P., Worqlul, A., & Sinha, A. (2018). Uncertainty in a Lumped and a Semi-Distributed Model for Discharge Prediction in Ghatshila Catchment. Water, 10(4), 381.

Yang, Z., & Becerik-Gerber, B. (2015). A model calibration framework for simultaneous multi-level building energy simulation. Applied Energy, 149, 415-431.

Zhang, X., Beeson, P., Link, R., Manowitz, D., Izaurralde, R. C., Sadeghi, A., ... & Arnold, J. G. (2013). Efficient multi-objective calibration of a computationally intensive hydrologic model with parallel computing software in Python. Environmental modelling & software, 46, 208-218.

9 Appendix

9.1 GSA↔DDS

9.1.1 Rastrigin

According to Table 1, the global optimal objective function value of the Rastrigin test problem with 20 parameters is equal to -20.0 that corresponds to all decision variables equal to 0.0. The performance of Morris↔DDS on Rastrigin function (Figure 30) is similar to Griewank function. Morris↔DDS show substantial faster convergence than DDS. As an illustration, with the optimization budget of 1000, DDS found -0.8 as the best function value at the 200th iteration, while Morris↔DDS found -8.11 at the same iteration. Despite the results of Wilcoxon rank-sum test (Table 20) that suggest Morris↔DDS generated different results than DDS (P-value less than 0.05), stochastic dominance graphs (Figure 30c) demonstrate that Morris↔DDS does not improve the function value substantially, as CDF of DDS and Morris↔DDS almost match each other.

In contrast to Morris↔DDS, Sobol↔DDS shows weaker convergence and final best solution compared to DDS in minimizing Rastrigin function. To illustrate this, at the budget of 1000, Sobol↔DDS found 11 as the best function value at 200th iteration. Moreover, as it is clear from stochastic dominance graphs, in both budgets, the best solution found by DDS in each trial significantly dominates the best solution found by Sobol↔DDS. Furthermore, according to the results of Wilcoxon rank-sum test for the best solution found by Sobol↔ and DDS (Table 20), the P-value for both budget is equal to 0.000183, which represents the significant difference between the results of the two algorithms. In conclusion, although Morris↔DDS shows faster convergence than DDS and Sobol↔DDS, according to Wilcoxon rank-sum test results DDS is unambiguously preferred over GSA↔DDS in minimizing Rastrigin.

As Table 21 illustrates, for both optimization budgets the initial and final parameter sensitivity ranking by Morris method is similar, however, the final (updated) sensitivity ranking by Sobol is significantly different from Morris and initial Sobol ranking. In fact, the performance of Sobol↔DDS on Rastrigin function is more degraded compared to Griewank function. This is due to the higher inaccuracy of updated sensitivity results in Rastrigin.

Table 20. Wilcoxon rank-sum test comparing DDS, Morris↔DDS and Sobol↔DDS applied to Rastrigin.

Computational Budget	Compared Methods	P-value	H-value
1000	DDS versus Morris↔DDS	2.57E-02	1
	DDS versus Sobol↔DDS	1.82E-04	1
10000	DDS versus Morris↔DDS	1.83E-04	1
	DDS versus Sobol↔DDS	1.83E-04	1

Table 21. The average sensitivity rankings for Rastrigin parameters generated by Morris and Sobol methods with the budget of 5 percent of total optimization cost (most sensitive parameters are shown in red and least sensitive parameters are shown in blue).

	Budget of 500 (5 percent of 10000)				**Budget of 50 (5 percent of 1000)**			
	Morris Rankings		**Sobol Rankings**		**Morris Rankings**		**Sobol Rankings**	
Parameter	**Initial**	**Final**	**Initial**	**Final**	**Initial**	**Final**	**Initial**	**Final**
1	9	3	13	20	5	10	2	17
2	4	17	7	2	10	9	17	20
3	6	13	8	8	15	4	12	14
4	13	18	16	1	8	7	1	11
5	8	7	4	15	11	19	15	4
6	16	9	14	18	14	14	9	16
7	7	1	9	16	13	2	6	13
8	5	20	15	14	1	18	14	10
9	19	8	17	6	7	12	3	1
10	15	2	12	13	17	11	10	19
11	11	19	5	5	20	20	7	9
12	17	14	10	4	16	1	5	5
13	1	15	3	19	12	15	11	7
14	3	6	2	11	3	16	19	18
15	2	16	1	9	2	3	16	6
16	20	5	19	10	18	5	20	3
17	10	10	11	3	4	6	8	15
18	12	4	20	12	19	13	13	2
19	18	11	6	17	6	17	4	12
20	14	12	18	7	9	8	18	8

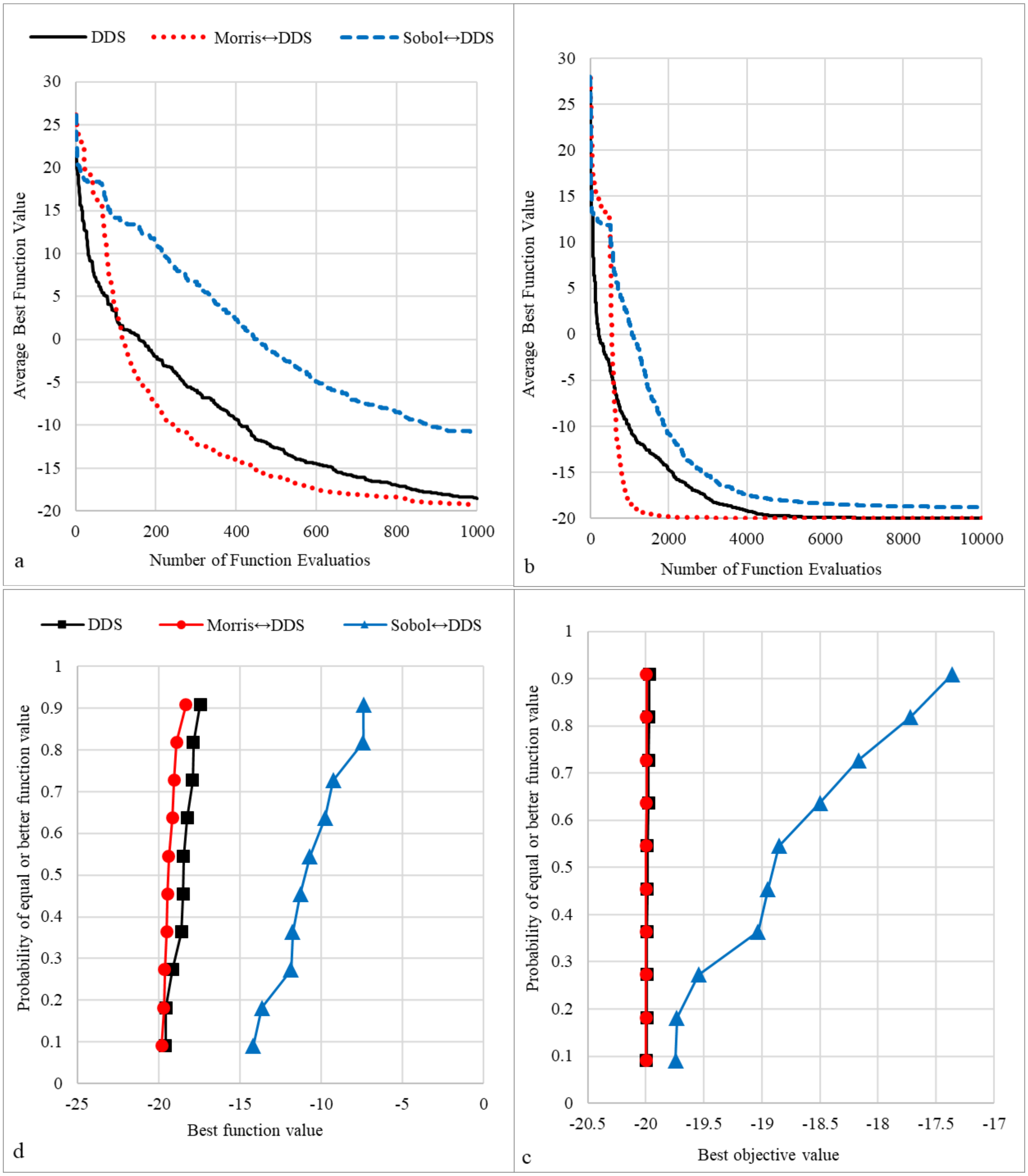


Figure 30. Convergence and first order stochastic dominance graphs of DDS, Sobol↔DDS and Morris↔DDS applied to the calibration of the Rastrigin function (left side figures show budget of 1000 and right side figures show budget of 10000).

9.2 GSA→DDS

9.2.1Rastrigin

As demonstrated in Figure 31, GSA→DDS illustrates similar performance on Rastrigin function compared to Griewank function. Both Morris→DDS and Sobol→DDS showed significantly faster convergence compared to original DDS. The improvement in the convergence rate is more obvious with the budget of 10000 function evaluations. For instance, both GSA→DDS methods found the objective value of -19.5 at 1500th iteration, while DDS found -12 at the same iteration. In addition, Table 22 shows that the CDFs of the final best solutions found in each trial by GSA→DDS significantly dominates the CDF of DDS at computational budgets of 1000 and 10000 evaluations with corresponding p-values smaller than 0.05. Therefore, Rastrigin function also showed that GSA→DDS is preferable over original DDS, and can significantly reduce the convergence time especially when the optimization budget is in higher order of magnitude.

Similar to Griewank function, the Morris and Sobol methods have generated analogous sensitivity rankings for most and least sensitive parameters of Rastrigin test function (Table 23). Thus, Morris→DDS and Sobol→DDS have similar performances on Rastrigin function. Nonetheless, due to insufficient GSA budget in 1000 function evaluations, small variation can be seen between Sobol and Morris rankings. Hence, the improvement in the performance of GSA→DDS in budget of 1000 is not significant due to less accurate sensitivity results.

Table 22. Wilcoxon rank-sum test comparing DDS, Morris→DDS and Sobol→DDS applied to Rastrigin

Computational Budget	Compared Methods	P-value	H-value
1000	DDS versus Morris→DDS	0.014	1
	DDS versus Sobol→DDS	0.031	1
10000	DDS versus Morris→DDS	5.80E-04	1
	DDS versus Sobol→DDS	1.82E-04	1

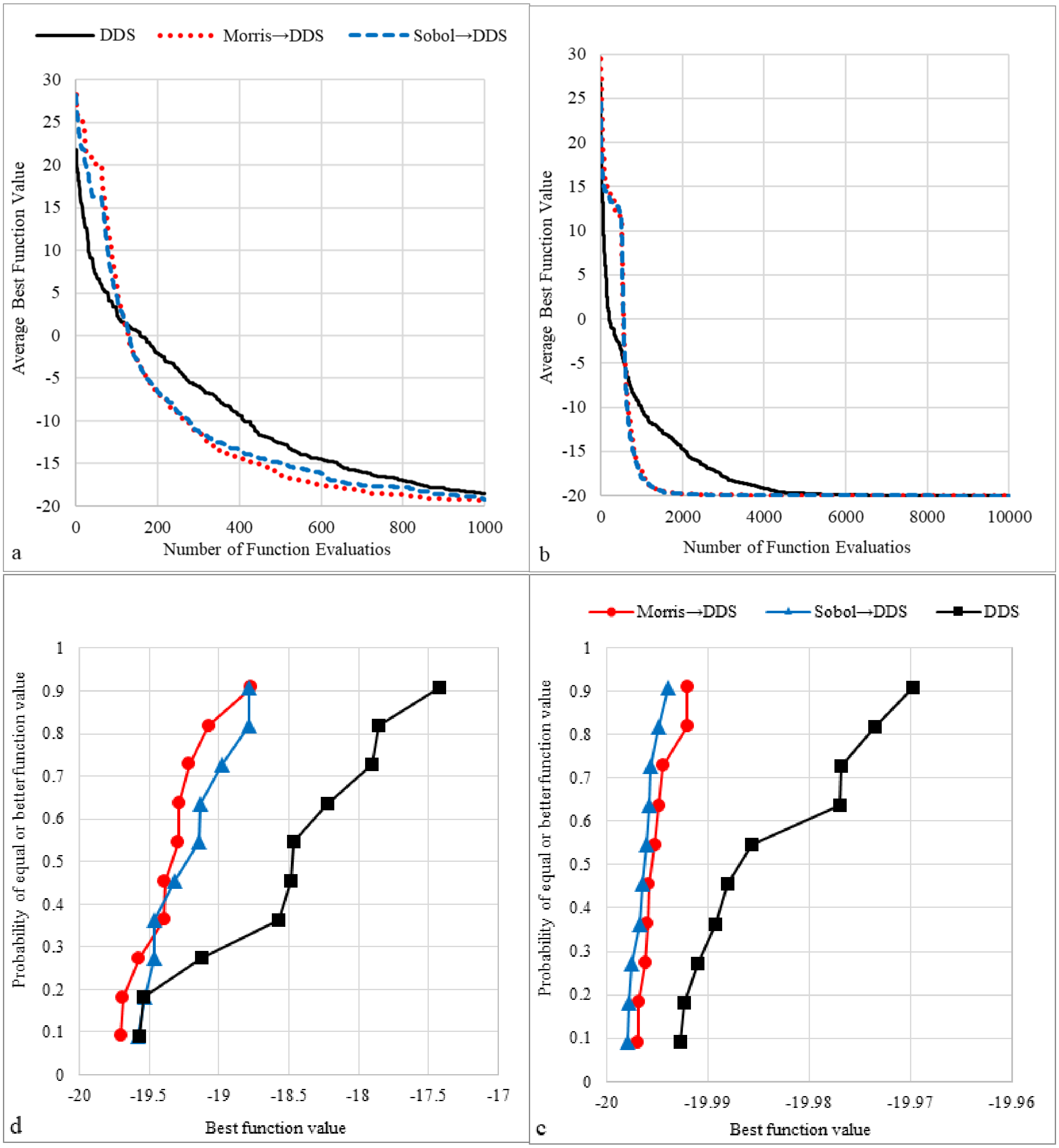


Figure 31. Convergence and first order stochastic dominance graphs of DDS, Sobol→DDS and Morris→DDS applied to the optimization of the Rastrigin function (left side figures show budget of 1000 and right side figures show budget of 10000).

Table 23. The average sensitivity rankings for Rasstrigin parameters generated by Morris and Sobol methods with the budget of 5 percent of total optimization cost (most sensitive parameters are shown in red and least sensitive parameters are shown in blue).

Budget of 500 (5 percent of 10000)		
Parameter	**Morris-Ranking**	**Sobol-Ranking**
1	9	13
2	4	7
3	6	8
4	13	16
5	8	4
6	16	14
7	7	9
8	5	15
9	19	17
10	15	12
11	11	5
12	17	10
13	1	3
14	3	2
15	2	1
16	20	19
17	10	11
18	12	20
19	18	6
20	14	18

Budget of 500 (5 percent of 10000)		
Parameter	**Morris-Ranking**	**Sobol-Ranking**
1	5	2
2	10	17
3	15	12
4	8	1
5	11	15
6	14	9
7	13	6
8	1	14
9	7	3
10	17	10
11	20	7
12	16	5
13	12	11
14	3	19
15	2	16
16	18	20
17	4	8
18	19	13
19	6	4
20	9	18

9.3 VARS→DDS

9.3.1Griewank and Rastrigin

To verify the sensitivity results produced by VARS toolbox, Morris and Sobol GSA methods have been applied on Griewank and Rastrigin functions and the sensitivity rankings are compared. As Table 25 and Table 27 demonstrate, the IVARS, Morris and Sobol rankings for both Griewank and Rastrigin functions have similarly identified the four most sensitive parameters.

Thus, the analogous ranking of VARS, Morris and Sobol verify their performance in calculating parameter sensitivity of Griewank and Rastrigin functions.

The performance of VARS→DDS, DDS, and DDS-Reduced in minimizing Griewank and Rastrigin functions with respect to four most sensitive parameters are evaluated. In DDS-Reduced, the four most sensitive parameters identified by VARS are considered for perturbation, while other parameters were kept constant at their mean values. According to the number of most sensitive parameters, in VARS→DDS, the i_{GPS} iterations are respectively equal to 250 and 1580 for the budget of 1000 and 10000 function evaluations.

In general, the performance of VARS→DDS on both Griewank and Rastrigin functions is similar. According to Figure 32 and Figure 33, no significant improvement in the results of VARS→DDS in terms of convergence and final best solutions can be seen. In Griewank, when the budget is 10000, slight improvement in the convergence rate is visible. However, it cannot be concluded from stochastic dominance graph (Figure 32c) that the CDF of VARS→DDS is able to stochastically dominate the CDF of DDS. Moreover, in Rastrigin function, (Figure 33) at the computational budget of 10000, VARS→DDS stochastically dominate DDS.

Nevertheless, it is clear from the results that regardless of computational budget, in both Griewank and Rastrigin functions, DDS-Reduced showed substantially better convergence compared to VARS→DDS and DDS. The CDF graphs for both Rastrigin and Griewank functions illustrate that DDS-Reduced stochastically dominants DDS and VARS→DDS. The main reason for the weaker performance of VARS→DDS compared to DDS-Reduced is that both Griewank and Rastrigin functions are not designed for sensitivity analysis, thus, the actual sensitivity index of parameters are very similar to each other. The similarity in the sensitivity indices prevents VARS→DDS to improve the convergence rate and final best solution.

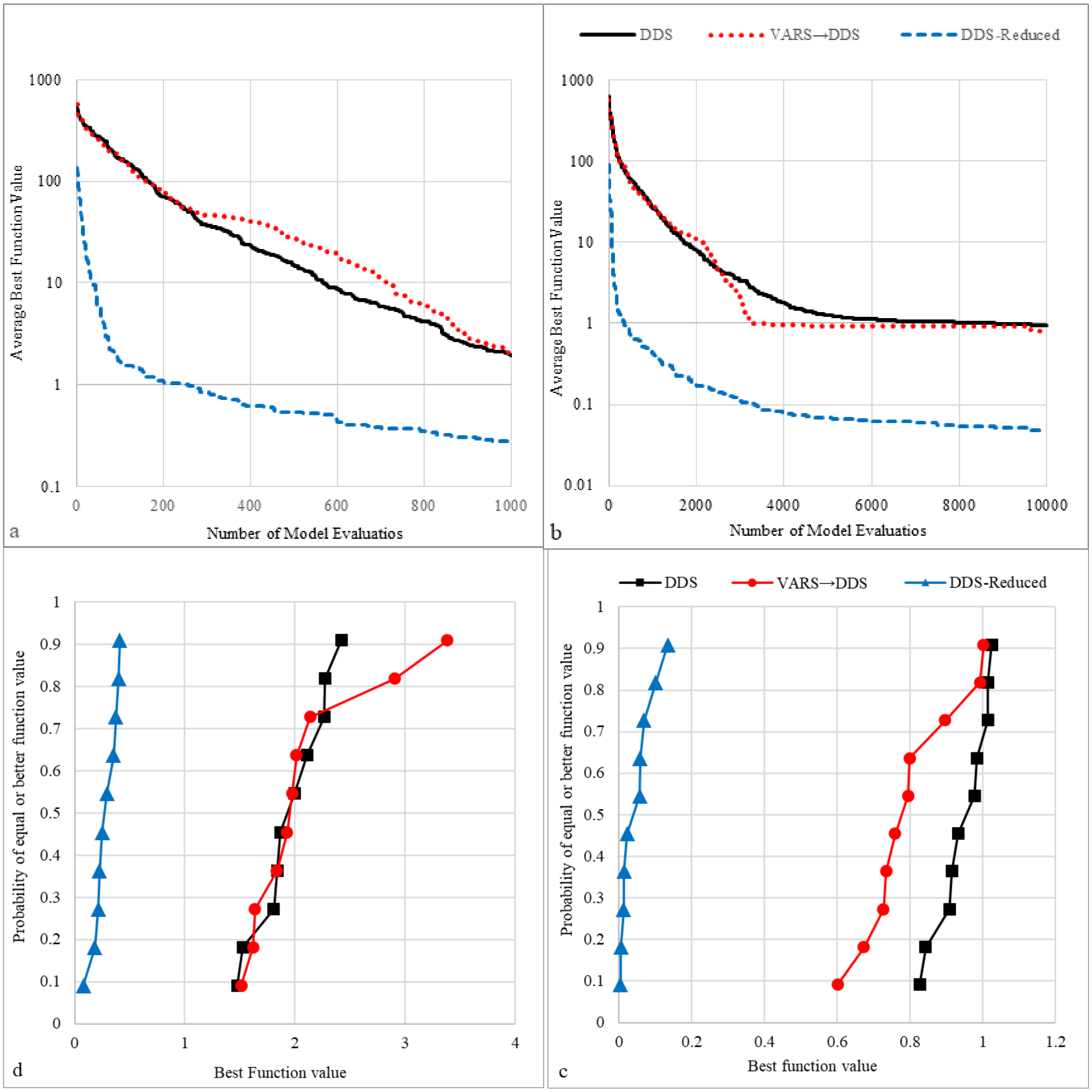


Figure 32. Convergence and first order stochastic dominance graphs of DDS, DDS-Reduced and VARS→DDS applied to the optimization of Griewank test function (left side figures show budget of 1000 and right side figures show budget of 10000).

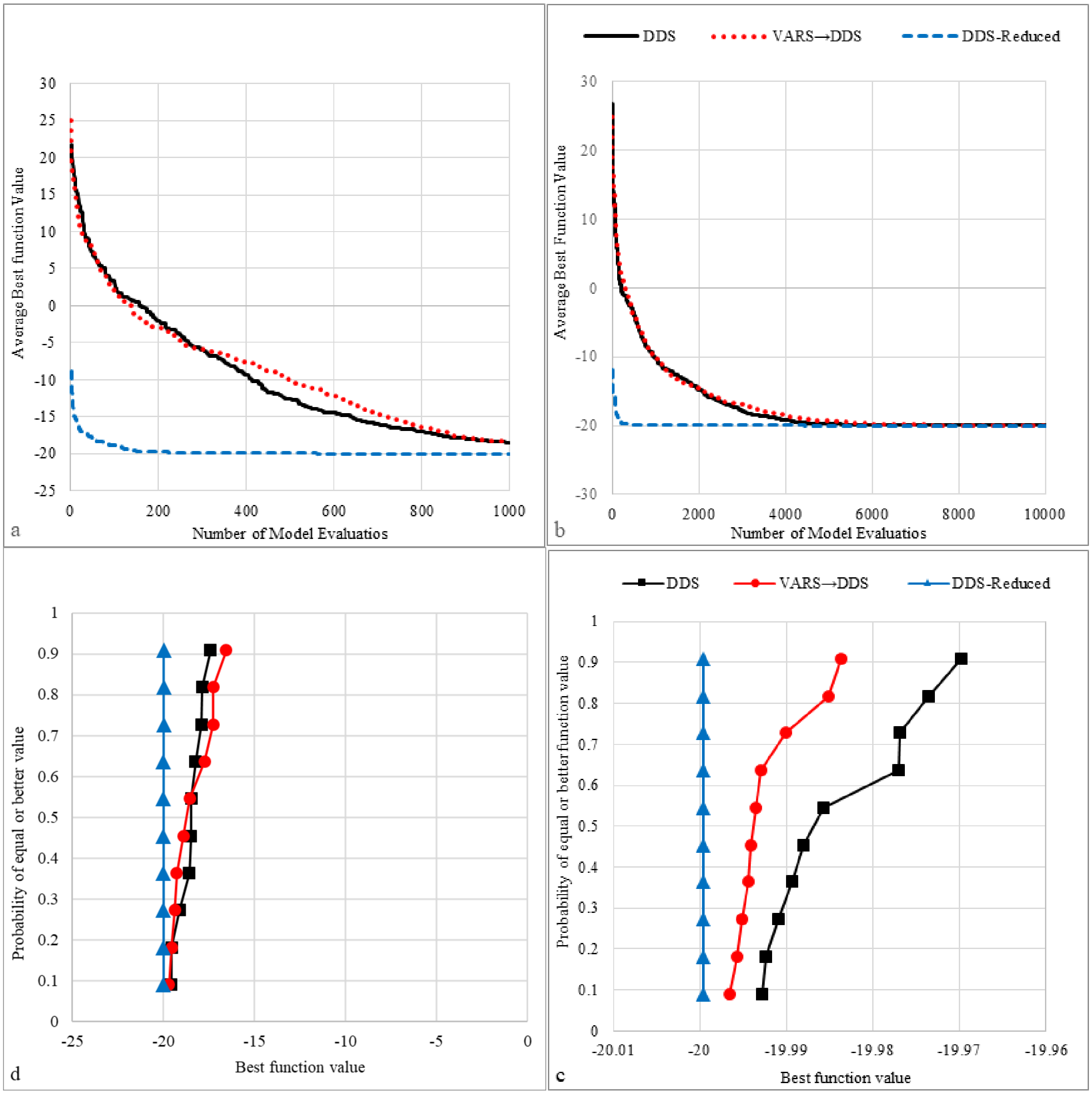


Figure 33. Convergence and first order stochastic dominance graphs of DDS, DDS-Reduced and VARS→DDS applied to the optimization of Rastrigin test function (left side figures show budget of 1000 and right side figures show budget of 10000).

The Wilcoxon rank-sum test for Griewank and Rastrigin functions (Table 24 and

Table **25**) confirm the results of stochastic dominance graphs. For Griewank function at the budget of 10000, the pairwise comparison of the best solutions generated by VARS→DDS, DDS and DDS-Reduced are significantly different as the *P*-values are less than 0.05. However, since VARS→DDS does not stochastically dominate DDS and DDS-Reduced, it is not preferred in optimizing Griewank.

On the other hand, when minimizing Rastrigin with the budget of 10000, the CDF of VARS→DDS stochastically dominates the CDF of DDS, and the corresponding *P*-value is less than 0.05. Thus, VARS→DDS is preferred over DDS in minimizing Rastrigin function. Nevertheless, it is worth mentioning that in both computational cost (i.e. 1000 and 10000 evaluations) DDS-Reduced is significantly better than DDS and VARS→DDS in optimizing Rastrigin and Griewank functions as the *P*-values for the pairwise comparison of the CDFs are substantially lower than 0.05. Hence, according to stochastic dominance graphs, since DDS-Reduced dominates both DDS and VARS→DDS, it is unambiguously preferred.

Table 24. Wilcoxon rank-sum test comparing DDS, DDS-Reduced and VARS→DDS for Griewank.

Computational Budget	Compared Method	P-value	H-value
10000	DDS versus VARS→DSS	0.0113	1
	DDS versus DDS-Reduced	1.83E-04	1
	DDS-Reduced versus VARS→DSS	1.83E-04	1
1000	DDS versus VARS→DSS	0.9698	0
	DDS versus DDS-Reduced	1.83E-04	1
	DDS-Reduced versus VARS→DSS	1.83E-04	1

Table 25. Wilcoxon rank-sum test comparing DDS, DDS-Reduced and VARS→DDS for Rastrigin.

Computational Budget	Compared Methods	P-value	H-value
10000	DDS versus DDS-VARS	0.0091	1
	DDS versus DDS-Reduced	1.72E-04	1
	DDS-Reduced versus DDS-VARS	1.72E-04	1
1000	DDS versus DDS-VARS	0.9698	0
	DDS versus DDS-Reduced	1.83E-04	1
	DDS-Reduced versus DDS-VARS	1.83E-04	1

Table 26. Average sensitivity rank of Griewank parameters by VARS, Morris and Sobol methods with 1000 model evaluations (most and least sensitive parameters are highlight in red and blue, respectively).

Parameter	IVARS10	IVARS20	IVARS30	IVARS40	IVARS50	Morris	Sobol
1	15	15	15	15	15	7	8
2	2	2	2	2	2	6	3
3	1	1	1	1	1	12	2
4	3	3	3	3	3	1	1
5	9	9	9	9	9	9	13
6	10	10	10	10	10	14	17
7	18	18	18	18	18	8	15
8	6	6	6	6	6	17	20
9	4	4	4	4	4	2	10
10	11	11	11	11	11	11	12
11	12	12	12	12	12	4	4
12	16	16	16	16	16	19	19
13	5	5	5	5	5	3	5
14	20	20	20	20	20	18	7
15	8	8	8	8	8	13	16
16	13	13	13	13	13	20	11
17	17	17	17	17	17	16	18
18	14	14	14	14	14	5	9
19	19	19	19	19	19	10	14
20	7	7	7	7	7	15	6

Table 27. Average sensitivity rank of Rastrigin parameters by VARS, Morris and Sobol methods with 1000 model evaluations (most and least sensitive parameters are highlight in red and blue, respectively).

Parameter	IVARS10	IVARS20	IVARS30	IVARS40	IVARS50	Morris	Sobol
1	20	20	20	20	20	10	17
2	15	15	15	15	15	9	20
3	18	18	18	18	18	4	14
4	12	12	12	12	12	7	11
5	2	2	2	2	2	19	4
6	17	17	17	17	17	14	16
7	6	6	6	6	6	2	13
8	19	19	19	19	19	18	10
9	1	1	1	1	1	12	1
10	14	14	14	14	14	11	19
11	10	10	10	10	10	20	9
12	4	4	4	4	4	1	5
13	5	5	5	5	5	15	7
14	11	11	11	11	11	16	18
15	13	13	13	13	13	3	6
16	16	16	16	16	16	5	3
17	8	8	8	8	8	6	15
18	3	3	3	3	3	13	2
19	7	7	7	7	7	17	12
20	9	9	9	9	9	8	8