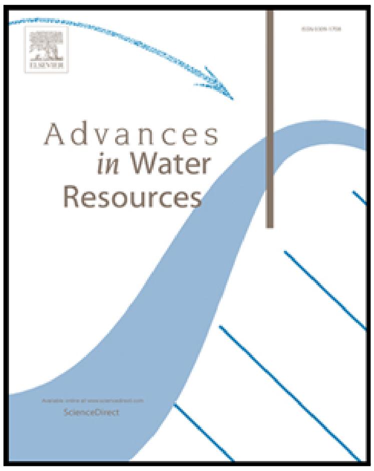
Accepted Manuscript



Event-based model calibration approaches for selecting representative distributed parameters in semi-urban watersheds

Frezer Seid Awol , Paulin Coulibaly , Bryan A. Tolson

PII: S0309-1708(17)31082-5
DOI: 10.1016/j.advwatres.2018.05.013
Reference: ADWR 3146

To appear in: *Advances in Water Resources*

Received date: 24 November 2017
Revised date: 16 May 2018
Accepted date: 20 May 2018

Please cite this article as: Frezer Seid Awol , Paulin Coulibaly , Bryan A. Tolson , Event-based model calibration approaches for selecting representative distributed parameters in semi-urban watersheds, *Advances in Water Resources* (2018), doi: 10.1016/j.advwatres.2018.05.013

Manuscript

Highlights

- A proposed event-based calibration process integrating multi-site, and single and multi-objective optimizations was used to select representative SWMM5 model parameter sets in a semi-urban watershed.
- Four calibration approaches (Multi-site simultaneous (MS-S), Multi-site average objective function (MS-S), Multi-event multi-site (ME-MS) and a benchmark At-catchment outlet (OU)) were compared for their performances at different gauging stations.
- Using the single objective DDS algorithm in MS-A approach to find the best average performance of five gauging stations in the catchment area is found to be more efficient than using the multi-objective PA-DDS algorithm in MS-S to find non-dominated Pareto-front of five individual performances.
- The study discovered that combination of efficient optimization tools with a series of calibration approaches and steps is important in finding candidate parameters sets and representing distributed catchments by event-based hydrological models.

Manuscript

Event-based model calibration approaches for selecting representative distributed parameters in semi-urban watersheds

Frezer Seid Awol[1], Paulin Coulibaly[2] and Bryan A. Tolson[3]

[1]Ph. D. Candidate, Department of Civil Engineering, McMaster University. 1280 Main Street West, Hamilton, Ontario, L8S4K1, Canada

[2]Professor, Jointly in School of Geography and Earth Sciences and Department of Civil Engineering, McMaster University. 1280 Main Street West, Hamilton, Ontario, L8S4K1, Canada

[3]Associate Professor, Department of Civil and Environmental Engineering, University of Waterloo, 200 University Ave W, Waterloo, ON, N2L 3G1, Canada

Corresponding author: Frezer Seid Awol (awolf@mcmaster.ca)

Manuscript

Abstract

The objective of this study is to propose an event-based calibration approach for selecting representative semi-distributed hydrologic model parameters and to enhance peak flow prediction at multiple sites of a semi-urban catchment. The performance of three multi-site calibration approaches (multi-site simultaneous (MS-S), multi-site average objective function (MS-A) and multi-event multi-site (ME-MS)) and a benchmark at-catchment outlet (OU) calibration method, are compared in this study. Additional insightful contributions include assessing the nature of the spatio-temporal parameter variability among calibration events and developing an advanced event-based calibration approach to identify skillful model parameter-sets. This study used a SWMM5 hydrologic model in the Humber River Watershed located in Southern Ontario, Canada. For MS-S and OU calibration methods, the multi-objective calibration formulation is solved with the Pareto Archived Dynamically Dimensioned Search (PA-DDS) algorithm. For the MS-A and ME-MS methods, the single objective calibration formulation is solved with the Dynamically Dimensioned Search (DDS) algorithm.

The results indicate that the MS-A calibration approach achieved better performance than other considered methods. Comparison between optimized model parameter sets showed that the DDS optimization in MS-A approach improved the model performance at multiple sites. The spatial and temporal variability analysis indicates a presence of uncertainty on sensitive parameters and most importantly on peak flow responses in an event-based calibration process. This finding implied the need to evaluate potential model parameters sets with a series of calibration steps as proposed herein. The proposed calibration and optimization formulation successfully identified representative model parameter set, which is more skillful than what is attainable when using simultaneous multi-site (MS-S), multi-event multi-site (MS-ME) or at basin outlet (OU) approach.

Manuscript

1. Introduction

Hydrological prediction in semi-urban watersheds requires a thorough understanding of the physical processes and the integrated response to storm events in partly urbanized and rural watersheds. In the last couple of decades, there have been research advances in understanding the urban and semi-urban hydrology with new emerging modelling tools. However, challenges remain due to the complex rainfall-runoff responses of combined urban, rural and urbanizing areas. Such mixed responses could result in multiple peak flows, which increase prediction uncertainty (Fletcher et al., 2013). Consideration of the gradual loss of pervious surfaces in semi-urban areas within hydrological models is non-trivial because this transformation could lead to increased peak flows, and reduced flood duration and response time (Miller et al., 2014). Impervious surfaces, on the other hand, amplify irregular and periodic flows (Ackerman et al., 2005). Although the research interest grows, there are only a few guidelines mentioned in calibrating urbanizing catchments. One possible reason is due to the challenges in transferring calibrated land cover parameters between catchments (Jacobson, 2011).

Despite their limitation in setting realistic initial conditions, event-based models are conservative in nature in simulating individual flood hydrographs and peak flows and provide better flood prediction when compared to continuous hydrological models (Tramblay et al., 2012; WMO, 2011). Several event-based models have been used for urban and semi-urban catchments. For example, El-Hassan et al., (2013) compared the performances of a conceptual HEC-HMS model and physically based distributed Gridded Surface Subsurface Hydrologic Analysis (GSSHA) model in simulating flood events of a semi-urban watershed and showed that the latter performed better. To identify the dominant peak flow mechanisms, Kennedy et al., 2013 used the Kinematic Runoff and Erosion Model (KINEROS2) in a semi-arid urban environment, whilst Zhang et al., (2013) applied Dynamic Watershed Simulation Model (DWSM) in semi-urban landscape. The effect of urbanization on hydrological responses is well studied by using several models, such as Catchment hydrological cycle Assessment Tool (CAT) (Miller et al., 2014), Distributed Hydrology–Soil–Vegetation Model (DHSVM) (Cuo et al. , 2008), a coupled Conversion of

Land Use and its Effect at Small regional extent (CLUE-E) and Soil and Water Assessment Tool (SWAT) (Arnold et al.,1998; Zhou et al., 2013) model. Event-based models were also used to assess their ability to reproduce past extreme, catastrophic flood events (Furl et al., 2015; Ogden et al., 2000; Sharif et al., 2013; Sharif et al., 2010).

The most widely used model for simulating extreme events in urban and semi-urban areas is the Environmental Protection Agency's Storm Water Management Model (SWMM) (Huber & Dickinson, 1988; Rossman, 2010). Gironás et al., (2010) studied the effects of various urban terrain morphologies on peak flow simulation by the SWMM model. Sun et al., (2014) compares two levels of SWMM catchment discretization (macro and micro-scale) to examine the degree of parameterizations and uncertainties using GLUE. Some advances were made on the calibration strategies of the SWMM. Krebs et al., (2013) and Zhang et al., (2013) employed Non-dominated Sorted Genetic Algorithm-II (NSGAII) and its revised version (ε-NSGAII), respectively, to optimize representative Low Impact Development (LID) scenarios in a small urbanized catchment. Herrera et al., (2006) also used NSGA-II with SWMM to analyze the trade-offs between low, medium, and high flows. Barco et al., (2008) utilized a weighted multi-objective function and alternating starting points or constraints to optimize coupled GIS/SWMM4 model for the large urban catchment. Zaghloul et al., (2001) used Generalized Regression Neural Network to improve PCSWMM98 model simulation with inverse calibration technique, which was applied in an impervious test area.

In the application of event-based hydrological models for peak flow prediction, the question of which calibrated model parameter sets should be used, can create a practical dilemma, unlike with continuous models. Despite the above efforts in improving the simulation and prediction capabilities of event-based models, novel methods are still required to address the uncertainties associated with model parameterization and temporal variations of input storm events. Robust calibration and validation approaches are requisite to identify optimum model parameters and improve runoff predictions (Krauße et al., 2012). Calibration procedures of hydrological models vary by their intended purpose, characteristics of the watershed, and the type and complexity of the models. The traditional approach is to calibrate the entire catchment (lumped or distributed) parameters according to model predictive performance at the basin outlet assessed via single or multiple objectives. Some authors have proposed advancing the single site calibration with a sequential/hierarchical approach (Hay et al., 2006; Ozdemir et al., 2017; Singh & Bárdossy, 2015). While the first authors sequentially calibrate a model's performance of potential evapotranspiration, water balance, and daily runoff , the second authors divided sub-basins into two hydrologic response units (HRU) and two further child HRUs based on influential parameters such as curve number and hydraulic conductivity. However, the limitation of single site approach in improving runoff simulation at interior sites of a distributed catchment has motivated multi-site calibration methods.

One straightforward and efficient way of calibrating models to a set of distinct events would be using all calibration events in a series, yielding a unique parameter set per event, and then select the final parameter set as the one that performs best in terms of average performance across all the events (in this paper, multi-event multi-site calibration approach). However, this could lead to under- or over-estimation of flows for any arbitrary event and marks a high compromise in searching parameters sets that satisfies all events at once.

A fairly reasonable and default multi-site calibration approach to consider internal gauges is by using a weighted average of performance metrics across the gauging sites (Asadzadeh et al., 2014; Engeland et al., 2006; Haghnegahdar et al., 2014; Khu et al., 2006; Khu et al., 2008; Madsen et al., 2002; Shinma & Reis, 2014; Xia et al, 2002; Zhang et al., 2009). These studies applied continuous calibration with different types of models. Haghnegahdar et al., (2014), for example, used this approach to calibrate the Canada's Modélisation Environmentale-Surface et Hydrologie (MESH) model (Pietroniro et al., 2007) by aggregating the objective function of multiple sites into a single objective and highlighted that the method has lower computational cost than other methods involving multi-objective optimization techniques.

As an alternative to the above approach, some authors proposed multi-site simultaneous calibration approach to exclusively implement multi-objective optimization technique and generate a set of non-dominated calibration solutions (Leta et al., 2017; Zhang, et al., 2010). With this approach, objective functions at the interior sites are optimized at the same time and the optimization result shows the tradeoffs between objective functions. Leta et al., (2017) applied a multi-site simultaneous calibration in developing SWAT Model for a heterogeneous catchment. Zhang et al., (2010) compared three optimization algorithms for multi-site simultaneous calibration of the SWAT model. The study highlighted that a multi-algorithm, genetically adaptive multi-objective method (AMALGAM) outperforms commonly used evolutionary multi-objective optimization such as Strength Pareto Evolutionary Algorithm 2 (SPEA2) and Non-dominated Sorted Genetic Algorithm II (NSGA-II). The above two studies were applied in continuous calibration approach for SWAT model. Other authors also considered multi-site step-wise/cascade (Brocca et al., 2011; Cao et al., 2006; Wang et al., 2012; Wi et al., 2015; Xue et al., 2016). Brocca et al., (2011), for example, used a distributed model with a sequential (step by step) calibration procedure to investigate its importance in flood forecasting and argued that the model improved peak flow estimation at internal sites.

To overcome the challenge of high computational cost in iterating through each sub-basin of a distributed catchment in multi-objective global search, the adaptation of tools with parsimonious characteristics is non-trivial. Asadzadeh & Tolson, (2009) developed a promising optimization tool, Pareto Archived Dynamically Dimensioned Search (PA-DDS), which is the multi-objective version of Dynamically Dimensioned Search (DDS) (Tolson & Shoemaker, 2007). PA-DDS has been compared with benchmark algorithms of NSGA-II and AMALGAM (Asadzadeh & Tolson, 2009), ε-NSGAII and AMALGAM (Asadzadeh & Tolson, 2013), and NSGAII and SPEA2 (Asadzadeh & Tolson, 2012) and the authors concluded that PA-DDS showed improved performances with limited computational cost compared to alternative algorithms.

Behavioral parameter sets of distributed models should be identified with an efficient optimization algorithm to help overcome problems of uncertainty and over-parameterization. For example, parameters derived from the calibration process do not always give improved performances in a validation period (Beven, 1989; Beven & Freer, 2001; Brocca et al., 2011; Madsen, 2003). Mediero et al.,

(2011) claim that the presence of multiple acceptable parameter sets not only avoid "equifinality", but also leads to an ensemble of flood event simulations, which provide probabilities. During the calibration process, they identified the Pareto solutions and fitted a distribution function to estimate bias and confidence intervals of ensembles in the validation period.

One way of solving the problem associated with distributed catchment parameters is through the use of spatial regularization as demonstrated by Pokhrel & Gupta, (2010). The authors used a non-linear transformation to reduce the number of parameters from Ng * Np (number of grid cells * number of parameters) to 3*Np by applying an adjustable multiplier, power term and additive constant to each prior estimated parameter value.

The above literature reviews indicate that the majority of multi-objective optimizations were conducted either for continuous distributed and lumped models or for application other than flood prediction in semi-urban watersheds. The objective of this study is to develop and test different event-based calibration approaches for enhanced flood prediction in semi-urban distributed catchments. A second objective is to analyze the spatio-temporal parameter variability of calibrated parameter sets to address the uncertainty in event-based parametrizations. The recent version of Storm Water Management Model (SMWM5) with DDS and PA-DDS optimization algorithms are used as calibration tools in this study. Section 2 describes the study area and data. Section 3 outlines the methodology including details of the model and optimization formulations, whereas the results and discussion are provided in Section 4. Finally, conclusion is presented in Section 5.

2. Study area and data

The research is conducted in the Humber River Watershed (Figure 1), which is located in Southern Ontario, Canada. The catchment area covers 911 km^2, and the main Humber River drains to Lake Ontario. The distributed catchment is configured by dividing the basin into 714 sub-catchments with areas spanning between 4.3 ha (0.043 km^2) and 860 ha (8.6 km^2). Humber River watershed is characterized as a semi-urban area with 54% rural, 33% urban and 13% urbanizing land covers and is administered by Toronto and Region Conservation Authority (TRCA, 2013). The hydrology and drainage patterns of the watershed are affected by its distinct topographic regions, which contain four hydrologic soil types (A, AB, B, BC, C, and D) (TRCA, 2008). The dual hydrologic soil groups AB and BC denote Sandy loam and Silt Loam soil types respectively (NVCA, 2006).

Gauge rainfall and discharge measurements were collected from Environment Canada and Toronto and Region Conservation Authority. The temporal resolution of received data ranges from 5 to 30 minutes for rainfall data and 15 minutes to 1 hour for discharge records depending on the availability. Ground-based rainfall data were used instead of gridded satellite or radar data because of unavailability of sub-hourly high-resolution temporal precipitation data in the study area. Niemi et al., (2017) also claimed that on-site gauge rainfall data showed better runoff simulation performance than radar-based data in urbanizing catchments. In the Humber River Watershed (Figure 1), eleven rain gauges spatially distributed across the basin and five river flow gauging stations along the main tributaries including one near the outlet have been used for this study. To separate the base flow from direct runoff, a simple straight line hydrograph separation method is used (Ajmal et al., 2016; Deshmukh et al., 2013).

Significant rainfall events in spring periods are screened and selected based on criteria of (1) total rainfall amount larger than 20 mm (TRCA & AMEC, 2012), (2) spatial coverage and distribution in the watershed (rainfall amounts measured at most of the rain gauges in the watersheds), and (3) their consistency with the associated discharge measurement. As such, ten calibration events and four validation events were captured in the period spanning between 2007 and 2014 (Table 2).

Manuscript

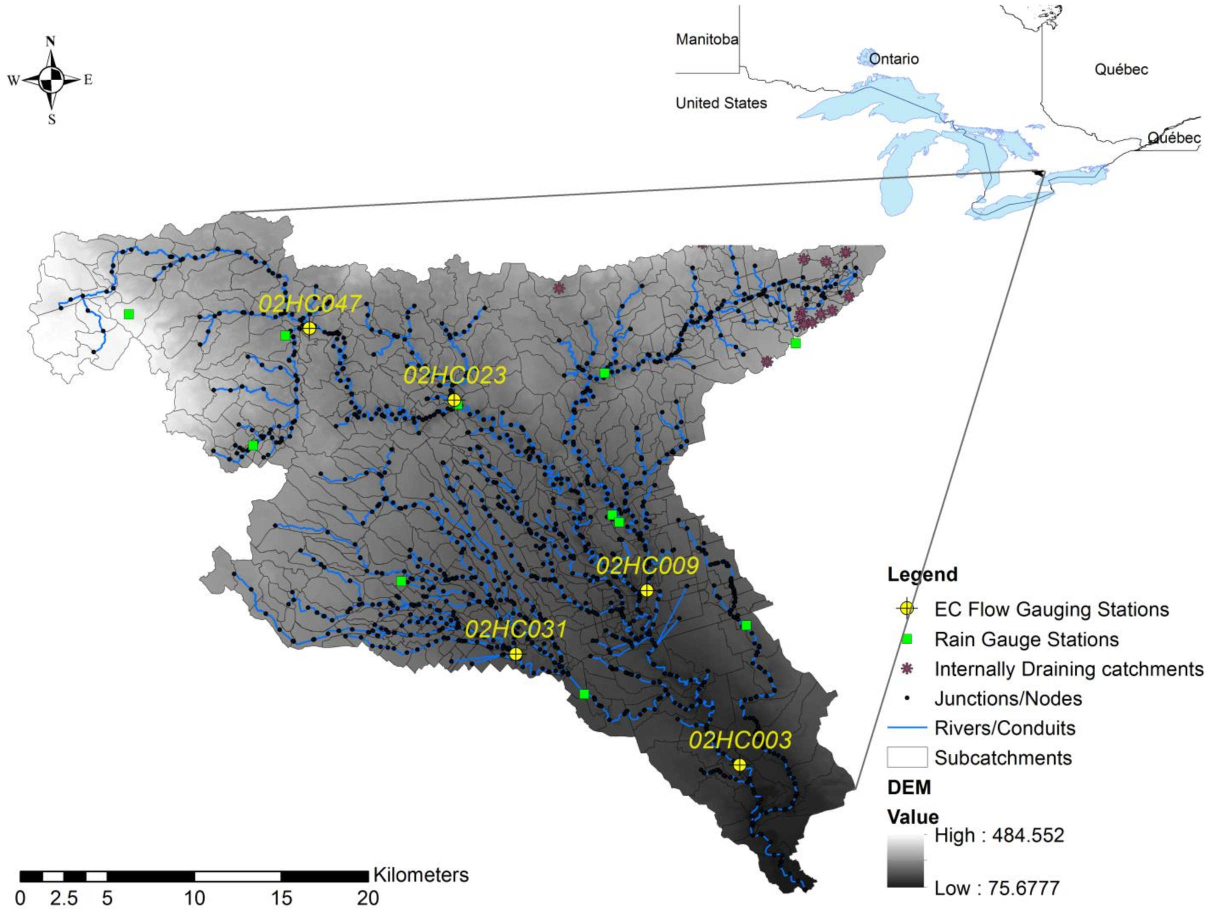


Figure 1: Location of the study area in Humber River Watershed, Southern Ontario.

Table 1: Description of SMWM5 model parameters

Parameter Codes	Description	Initial range of parameters**
IM*	Imperviousness [%]	0-99
W*	Characteristics Width of Overland flow [m]	163-124000
SP*	Depression storage in Pervious areas [mm]	1-600
CN*	Curve Number [-]	1-99
SL*	Catchment slope [%]	0.3-4.5
NI	Manning's n for overland flow in Impervious areas [-]	0.008-0.025
DT*	Drying time [days]	4-12
SM*	Depression storage in Impervious areas [mm]	0.2-5
NP	Manning's n for overland flow in Pervious areas [-]	0.08-0.4

*parameters used in calibration process

** The initial values of SWMM parameters were collected from the Toronto Region Conservation Authority (TRCA & AMEC, 2012)

Manuscript

3. Methods

3.1. Model setup

The Storm Water Management Model (SWMM) is a well-established event-based and continuous semi-distributed model used to simulate extreme events and peak flows in urban and semi-urban watersheds (Huber & Dickinson, 1988; Rossman, 2010). Due to the semi-urban characteristics of the study area and SWMM's wide application in operational flood forecasting (Randall et al., 2014; Robert et al., 2008), the recent version of SWMM (SWMM5) engine within PCSWMM platform is used in this study. Curve number method and dynamic wave routing method have been used as an infiltration model and routing method respectively.

A sub-catchment in SWMM5 is represented by a non-linear reservoir model, where the conservation of mass is applied to generate overland flow (Rossman & Huber., 2015). By combining Conservation of Mass and Manning's equation, SWMM5 solves first the depth of a pond in sub-catchment (d) and then runoff at each time step using the following equations. More detailed information can be obtained from Rossman & Huber., (2015).

$$\frac{\partial d}{\partial t} = i - e - f - \alpha (d - d_s)^{5/3} \qquad (1)$$

Where, $\alpha = \frac{W S^{1/2}}{A\, n}$, in which each sub-catchments area (A) can be partitioned into pervious and impervious areas using the 'Percent Imperviousness' parameter. And the roughness (n) will be defined for each partition using the 'pervious manning's n' and 'impervious manning's n' parameters.

i = rate of rainfall + snowmelt (m/s)

e = surface evaporation rate (m/s)

i = infiltration rate (m/s)

d = ponded depth (m)

d_s= depression storage depth (m)

W = sub-catchment width (m)

Manuscript

S = sub-catchment slope (-)

Once d (ponded depth) is solved using equation 1 at each time step, the volumetric flow rate (Q in m^3/s) can be estimated by:

$$Q = \frac{WbcshS^{1/2}}{n}(d - d_s)^{5/3} \quad (2)$$

Using the Curve Number method (in the current research) as an infiltration method and assuming the cumulative precipitation and infiltration at the start of the time step as P_1 and F_1 respectively, the infiltration rate (in m/s) is solved as follows (Rossman & Huber., 2015).

$$f = (F_2 - F_1)/\Delta t \quad (3)$$

Where, $F_2 = P_2 - \frac{{P_2}^2}{P_2 + S_{max}}$

And, $\mathrm{S_{max}} = \frac{25400}{\mathrm{CN}} - 254$, where CN is the curve number and, S_{max} is the maximum soil moisture storage capacity (in mm).

Finally, the drying time (DT in days) is used to calculate a recovery constant (hr^{-1}), that is used to model the depletion and replenishment of the soil moisture storage capacity in wet periods and dry period, respectively (Rossman & Huber., 2015).

SWMM5 consists of several physical and hydrological parameters to generate flow hydrograph, out of which nine catchment parameters (Table 1) are investigated to check their sensitivity against peak flow. 714 sub-catchments of Humber River watershed are assigned with unique parameter values. In Table 1, column three indicates the range of initial parameter values for 714 sub-catchments that are collected from previous studies and guidelines (CIVICA & TRCA, 2015; James, 2005). Event-by-event calibration and model testing are performed with simulation time steps of 15 or 30 minutes depending on input data time resolution. For defining the initial wetness of the watershed, the model was run for 1 to 2 weeks before each storm events as a ‘warm up’ period.

Manuscript

The methodology proposed in this study is summarized by a flowchart shown in Figure 2, which breaks down the calibration procedure into a series of phases. Phase 1 is the model setup and calibration/validation data selection phase, which is described above. Phase 2 is the sensitivity analysis phase, the purpose of which is to find most sensitive model parameters in semi-urban watersheds such as Humber River Basin. Phase 3 is the spatial and temporal parameter variability assessment that aims to analyze the uncertainty associated with event-based calibration and variability of candidate parameter sets. In Phase 4, two calibrations steps are introduced. The first one compares four different types of calibration approaches and proposes ten individual candidate parameter sets obtained from the best optimization approach. The second step tests the candidate parameter sets to all calibration events and selects a certain number of parameters sets that have higher scores over the entire events and gauging sites. Phase 5 evaluates the candidate parameter set(s) in different events to refine the calibration output and select the best representative parameter set. The details and methodology associated with each of these phases are described sequentially in the following Sections (Section 3.2 to 3.5).

Table 2: Events selected for calibration and model testing

No.	Calibration Events	Amount of rainfall (mm)	Avg. Discharge* (mm)	Avg. Discharge* (m3/s)	No.	Validation Events	Amount of rainfall (mm	Avg. Discharge* (mm)	Avg. Discharge* (m3/s)
1	19-Aug-05	53.3	30.4	282.4	1	15-May-07	47.1	8.7	81.0
2	10-Jul-06	66.7	8.7	81.0	2	20-Oct-11	75.6	9.8	90.6
3	28-May-13	64.5	10.8	100.0	3	5-Sep-14	84.1	8.3	76.8
4	8-Jul-13	81.9	29.0	269.0	4	29-Nov-11	75.2	15.9	147.2
5	31-Jul-13	74.5	5.1	47.0					
6	27-Jul-14	29.8	7.3	67.3					
7	20-Aug-09	19.9	6.8	62.7					
8	28-Sep-10	41.4	5.4	50.3					
9	13-May-11	64.2	9.7	90.1					
10	7-May-10	37.6	9.0	83.1					

* Average discharge measured at the outlet (HC003).

Manuscript

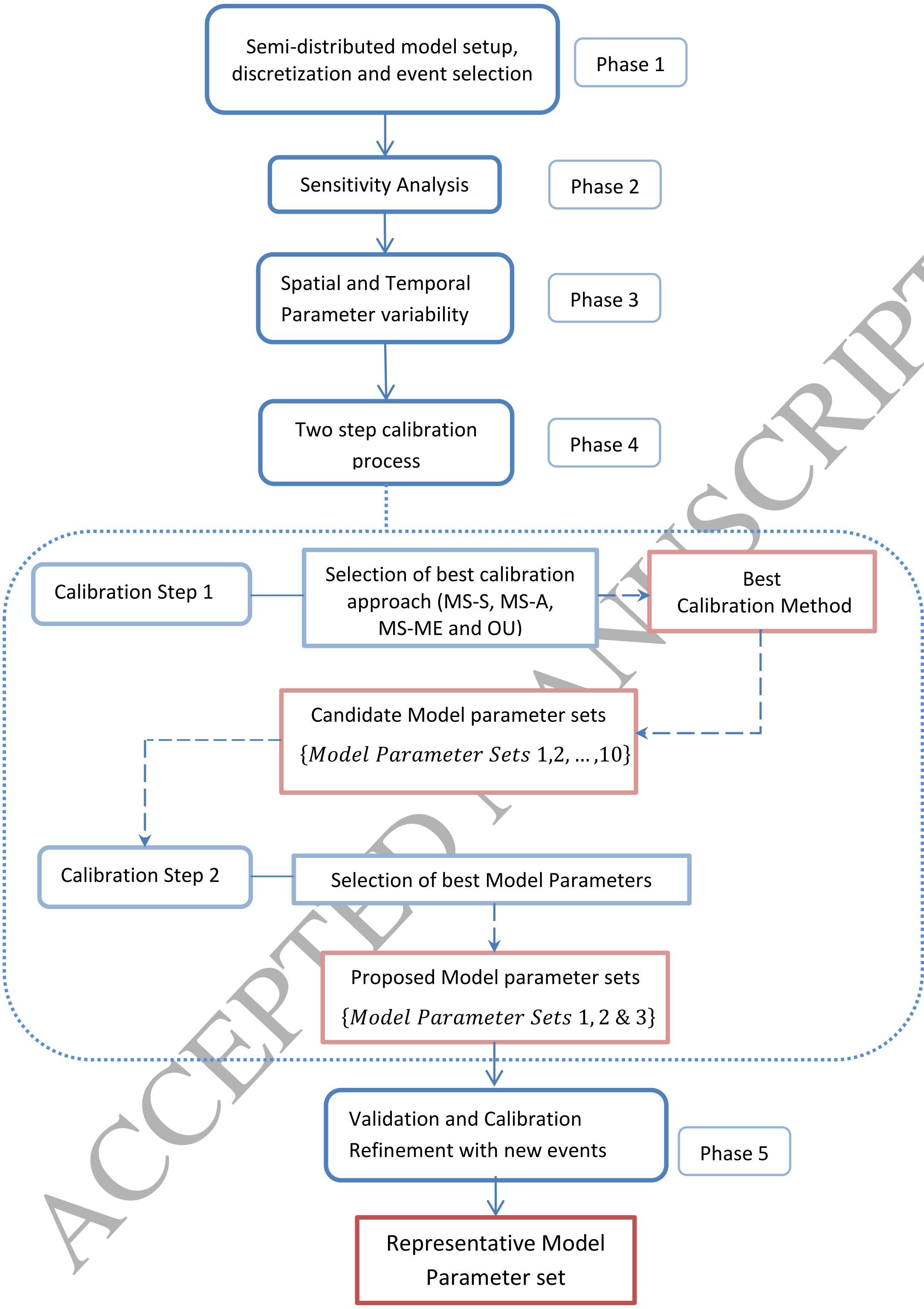


Figure 2: Flowchart of proposed approach for selecting representative parameter set in event-based models

3.2. Sensitivity analysis

The sensitivity of different versions of SWMM model parameters has been tested in different rural and urban watersheds (Barco et al., 2008; Irvine, et al., 1993). In this research, the purpose of sensitivity analysis of SWMM5 model is to identify the most sensitive parameters for the study basin. It was conducted by using two methods: Regionalized Sensitivity Analysis (RSA) (Spear & Hornberger, 1980) and Cumulative Sum of the Normalized Reordered Output (CUSUNORO) (Plischke, 2012).

Regionalized Sensitivity Analysis (RSA): Also called Generalized Sensitivity analysis or Hornberger-Spear-Young-method (Spear & Hornberger, 1980), RSA is used to identify the most sensitive parameters by distinguishing behavioral and non-behavioral parameter sets for Nash-Sutcliffe Efficiency (NSE), Peak flow Error (PE) and Volume Error (VE) model performances. 3500 parameter sets were generated by using Pareto Archived Dynamically Dimensioned Search (PA-DDS) (Asadzadeh & Tolson, 2013) optimization algorithm. The sensitivity was measured by Kolmogorov–Smirnov test statistics, which evaluates the maximum vertical distance between the curves of the cumulative distribution function of behavioral $F_n(x)$ and non-behavioral $F_{n'}(x)$ parameter sets as defined by:

$$d_{n,n'} = \sup_x |F_n(x) - F_{n'}(x)| \tag{4}$$

Where, $d_{n,n'}$ is the maximum vertical distance and *sup* is the supremum function. $d_{n,n'}$ (hereafter called RSA index) value ranges between 0 and 1 representing the limit between the most insensitive and sensitive parameters, respectively. Most sensitive parameters would have higher maximum vertical distance between the curves of $F_n(x)$ and $F_{n'}(x)$.

Cumulative Sum of the Normalized Reordered Output (CUSUNORO): Initially proposed by Plischke, 2012, CUSUNORO is a graphical post-processing method to represent the first-order sensitivity index. Its principle is withdrawn from the ideas of Contribution to the Sample Mean (CSM) plot (Bolado-Lavin et al., 2009). CSM and CUSUNORO are found to be suitable for estimating the main effect, the first-order variance based sensitivity index for cases where there is no direct access to the sampling procedure and the simulation model to map input-output relationship (Plischke, 2012).

Manuscript

Let π denote an arrangement of ordered values of input parameters sorted in ascending order, i.e, $x_{\pi(i)} = \{x_{i,1}, x_{i,2}, \dots, x_{i,n}\}$; hence its corresponding sorted series of outputs $y_{\pi(i)}$ can be created for all x. A scaling factor, which resembles the output variance is then created using the square root of the sum of squares $s_{yy} = \sum_{i=1}^{n}(y_{(i)} - \bar{y})$ (Plischke, 2012). Finally, the cumulative sum of normalized reordered output is defined as:

$$z(i) = \frac{1}{\sqrt{n.s_{yy}}}\sum_{j=1}^{i}\left(y_{\pi(i)} - \bar{y}\right) \quad (5)$$

The CUSUNORO values, $z(i)$, can then be plotted against the empirical cumulative distribution of input parameters x_i to visualize the sensitivity of individual parameters on the output statistics.

SWMM5 model parameters (Table 1) are considered as inputs and different performances metrics were used as outputs. Input-output mapping is performed externally by using Pareto Archived Dynamically Dimensioned Search (PA-DDS) (Asadzadeh & Tolson, 2013).

3.3. Spatial and temporal parameter variability

The primary objective of this section is to address the variability in event-based parametrizations in a semi-urban watershed and how it can be quantified by different calibration approaches. Before starting applying alternative and new methods of calibration formulations and optimization algorithms, we perform this exercise using a benchmark calibration approach at the catchment outlet involving limited manual and multi-objective calibration. The calibration process is described in detail together with the other proposed approaches in Section 3.4.1. The outcome assists to formulate and compare alternative event-based calibration approaches in reducing the uncertainties. Different parameterizations of the SWMM5 model represent several realizations of the physical process in the event of extreme spring rainfalls. Ten individual event-based calibrations result with ten SWMM5 model parameter sets. The variability of these sets regarding the model output as well as differences of calibrated sensitive parameters among the events was assessed.

First, the spread of two sensitive model parameters (Imperviousness and Drying Time) in each model parameter sets were assessed by developing box plots for different percentile values. Parameter values, collected from 714 sub-catchments, were ranked in ascending order and their percentiles were extracted accordingly. The variability of calibrated parameters in space can be observed by the degree of the spread.

Second, the uncertainty of event-based parametrization in a distributed catchment was evaluated by analyzing the peak flow response. We re-run the ten model sets for ten calibration events by regarding each model sets as an individual model and the peak flow simulation results were extracted. The specific objective of this method is to check how variable simulated peak flows are within each model sets as well as with the observation at multiple interior sites. Various boxplots were used to display standardized peak flow variability. The Standardized peak flow is calculated by normalizing the deviation of the simulated peak flows from observed peak flow by their standard deviation.

3.4. Model Calibration

3.4.1. Event-based Calibration approaches

Three multi-site event-based calibration approaches are compared with a benchmark 'At-catchment outlet' method to select potential parameters sets in Humber River basin. The calibration parameters in each of these four approaches are the same and are determined from the sensitivity analysis described above.

i) At Catchment Outlet (OU)

The conventional calibration approach of many hydrological models is to calibrate the entire catchment using a gauging station located at the basin outlet. In this approach, calibration to each of the ten events is completed independently. This calibration method is used as a benchmark to compare its results with other considered calibration approaches. Limited manual calibration is performed before using the following optimization formulation in order to get initialized solutions.

Single and multi-objective optimization techniques could be used to calibrate distributed models at basin outlets. Here, in order to find the best achievable parameter sets, multi-objective optimization with three different performance metrics (Nash-Sutcliffe Efficiency (NSE) (Nash & Sutcliffe, 1970), Peak flow Error (PE) (Liong et al., 1995), and Volume Error (VE) (Niemi et al., 2017) are used to calibrate Humber River Watershed at HC003 gauging station. This formulation is similar to the one used by Barco et al., 2008, where they minimized a weighted objective function summing the total flow volume, peak flow rate and instantaneous flow rate errors (each as percentage). The basic difference is that Barco et al., 2008 minimize/maximize a single weighted objective function by changing the weights depending on target flow type (e.g. peak flow or volume) whereas the approach here gives equal weight to individual objective functions and used a multi-objective PA-DDS algorithm to identify non-dominated solutions. The exercise is repeated ten times for ten calibration events with maximum iteration of 500 set for each optimization.

The multi-objective target is to maximize NSE and minimize PE and VE at station this station (a). i.e

$$O_{humber} = \{o_1 = NSE_a,\ \ o_2 = VE_a,\ o_3 = PE_a\} \tag{6}$$

In which:

$$NSE = 1 - \frac{\sum\left(Q_{o,i} - Q_{s,i}\right)^2}{\sum\left(Q_{o,i} - \overline{Q_o}\right)^2}$$

$$VE = \frac{|V_o - V_s|}{V_o} \tag{7}$$

$$PE = \frac{\left|Q_{p,o} - Q_{p,s}\right|}{Q_{p,o}}$$

Where, $Q_{o,i}$ & $Q_{s,i}$ are observed and simulate discharge at each time step, in cubic meter per second and $\overline{Q_o}$ is the average observed discharge; $Q_{p,o}$ & $Q_{p,s}$ are observed and simulated peak flows respectively; and V_o & V_s are the volume of water under observed and simulated flow hydrographs respectively, in million cubic meter. NSE value ranges between –∞ and 1 with 1 indicating best performance. PE, and VE have values spanning between 0 and ∞ and better performing model sets would have values close to 0. The result of the OU calibration approach is ten parameter sets (for ten calibration events), with each set being made up of the average of non-dominated solutions corresponding to a specific flow event.

ii) Multi-Site Simultaneous multi-objective (MS-S)

Multi-objective optimization techniques have been frequently used to calibrate distributed models. A multi-objective optimization algorithm is used to find a feasible set of Pareto-optimal parameter solutions by minimizing or maximizing the objective function vector. i.e. $Min/Max\ \ \boldsymbol{O}(\boldsymbol{p}) = [o_1(p), o_2(p), o_3(p), \dots, o_m(p)]$ where the objective function vector $\boldsymbol{O}(\boldsymbol{p})$ is comprised of m objective functions or performance metrics (Zhang et al., 2010).

Multi-site simultaneous multi-objective optimization was previously considered for continuous calibration (Leta et al., 2017; Zhang et al., 2010). In the current study, it is applied for an event-based calibration process. In this calibration approach, optimization is performed independently for ten individual calibration events. For each event, the model's performance is assessed simultaneously across multiple gauging stations using Nash-Sutcliffe Efficiency (Nash & Sutcliffe, 1970) performance metrics. In other words, the performance at each site in the study area is assessed by a different objective function so that performances at multiple locations are accounted for *simultaneously*. That is, for the five gauging stations in Humber River Watershed (represented by a, b, c, d, and e):

$$O_{humber} = \{o_1 = NSE_a,\ \ o_2 = NSE_b,\ o_3 = NSE_c\ \ o_4 = NSE_d\ \ o_5 = NSE_e\} \tag{8}$$

For optimization, Pareto Archived Dynamically Dimensioned Search (PA-DDS) (Asadzadeh & Tolson, 2013) algorithm is applied to find the Pareto-optimal parameters sets. PA-DDS was used within OSTRICH (Matott, 2005) framework toolkit. The selection operation in PA-DDS of non-dominated solutions (Pareto-optimal solution) is performed using estimated Hypervolume Contribution (HVC) (Asadzadeh & Tolson, 2013). The maximum number of iteration is set as 500 and the perturbation parameter is left as the default value of 0.2. Since there are 10 calibration events, 10 PADDS optimization is performed to evaluate the objective function values of each solution.

The result of the MS-S calibration approach is multiple parameter sets or non-dominated solutions corresponding to a specific flow event. Then, equal weight is given to each objective functions (o_1, o_2, o_3, o_4& o_5 *in equation* 3) to find the average of the non-dominated parameter sets and solution for each calibration event.

iii) Multi-Site Average objective function (MS-A)

This calibration method is frequently used by several researchers to account for the interior sites of a semi- or fully distributed catchment in the calibration process by taking the weighted average of multiple objective functions. The objective functions at multiple gauging stations are aggregated into a single objective function. Then optimization is performed to maximize the aggregated single objective function.

The five sites of Humber River Watershed are evaluated by their respective Nash-Sutcliffe Efficiency index:

$$NSE = (NSEa + NSEb + NSEc + NSEd + NSEe)/5$$
$$O_{humber} = \{NSE\} \tag{9}$$

The single-objective function (O_{humber}) is optimized by using Dynamically Dimensioned Search (DDS) (Tolson & Shoemaker, 2007) optimization algorithm within OSTRICH framework (Matott, 2005). Similar to the MS-S approach, the MS-A DDS optimization is performed independently for 10 individual calibration events and the result is 10 candidate parameter sets. In addition, the maximum number of iteration of 500 and perturbation value of 0.2 was set.

With perfect algorithms that converge to true optimal solution/true set of non-dominated solutions, MS-A would yield one of the non-dominated solutions generated by solution of MS-S formulation. In all practical calibration situations, convergence to true optimal/Pareto-optimal set of solutions is not guaranteed and thus all results are approximate. The quality of the approximations to the true, but unknown solutions is dependent on the algorithm quality (DDS and PADDS) and is also dependent on the algorithm computational budget. PADDS and DDS computational budgets in terms of number of solutions evaluated in MS-A and MS-S are equivalent and set to 500 and replicated 10 times for 10 calibration events.

The main difference between the MS-A and the MS-S approach is on the optimization method. While MS-A is based on a single objective optimization scheme (see Equation 7, p.18), the MS-S approach employs multi-objective optimization function (see Equation 6, p.17). In the former MS-A approach, although it involves aggregating several objectives, it is based on a single objective calibration process with the help of Dynamically Dimensioned Search (DDS) method: i.e. the objective is to maximize a single NSE value which is the average NSE of all sites (including at the outlet). Conversely, MS-S approach aims to find a feasible Pareto front by maximizing the objective function vector (rather than a single value): in which the vector comprises of NSEs at multiple sites including the outlet. In MS-S approach, the non-dominated (Pareto-optimal) solutions are generated by finding a tradeoff between individual objective functions using Pareto Archived Dynamically Dimensioned Search (PA-DDS) algorithm. At each iteration, MS-S searches for a tradeoff of optimum parameters that simultaneously satisfies individual objective functions or simultaneously maximizes the performances of each NSEs (interior as well as outlet), whereas MS-A searches the best parameters of the whole 714 sub catchments that maximize a single NSE value (average of NSEs).

iv) Multi-event multi-site calibration (ME-MS)

This approach involves concatenating the simulated and observed discharge of separate events and treating it as a single time series. For the combined multi-event series, the performance metrics (NSE) are then computed at each gauging stations. The multi-site objective function is basically defined in a similar manner as the previous calibration approach (MS-A) (equation 7) and thus is also formulated as a single-objective optimization problem. One of the differences between ME-MS and the above two (MS-S and MS-A) approaches is that ME-MS is applied over all ten events, whereas the others performed event by event. The optimization was performed by DDS algorithm with maximum iteration of 500. The calibration result is one set of candidate parameter sets that are somehow appropriate for all ten flow events.

3.4.2. Calibration steps

In order to identify the best parameters sets across the calibration events, the results of the above four calibration approaches described in section 3.4.1 are processed and compared in the following two calibration steps.

Step 1:- Select best set of candidate solutions, (e.g. select best calibration approach):

Each calibration approach generates a set of candidate parameter sets. The calibration approach with better performance and score at each calibration event and gauging station is selected for the next step. This step comprises of a couple of processes. Initially, we calibrate the model to ten individual events (Table 2) using MS-S, MS-A and OU approaches. At the end of each optimization or calibration approaches ten candidate parameters sets are foreseen for ten flow events. The performance of the final calibrated sets of parameters would be different for different optimization formulation. Therefore in the next process we compared the result of these calibration approaches at each individual event. Here, since ME-MS approach is formulated by aggregating over ten calibration events, it results with one set of calibrated parameters for all events as opposed to the output of MS-S, MS-A and OU approaches, which have ten sets of calibrated parameters. For comparison purpose, we re-apply the final calibrated parameter sets of ME-MS to ten events so that the results of four calibration approaches could be compared at individual events. In addition, comparison is also made at individual gauging stations (five sites). Finally, the best calibration approach that performed well at ten calibration events and five sites is proposed to the next calibration step. The final outcome of this step is ten calibrated parameter sets from one of the calibration approaches.

Comparison of calibration approaches is performed using model improvement scale or Prediction Error Decrease (PED) in percentage (Coulibaly, 2003) and Taylor Diagram (Taylor, 2001). The PED shows the model performance improvement of Multi-site simultaneous (MS-S) and Multi-site Average objective function (MS-A) and Multi-event multi-site (ME-MS) calibration approaches when compared to the benchmark At-Catchment Outlet (OU) approach at five gauging stations. Taylor diagram is used to precisely quantify and display the pattern similarity and statistics of different calibrated model parameter sets and the observation at multiple gauging sites. A revised normalized Taylor Diagram is constructed based on Kärnä & Baptista, 2016 by relating normalized centered root-mean-squared error with ratio of standard deviation of observed and simulated discharge and correlation coefficient through a Law of Cosines. The attributes of Taylor Diagram will be able to show the statistical proximity of individual model sets derived from two calibration approaches with the observation at five gauging stations. Details regarding Taylor Diagram can be found in Taylor, 2001.

<u>Step 2</u>:- From best approach candidate parameter sets, filter out poor candidates (e.g. select top three):

From the first step, ten candidate parameter sets are produced by the best calibration approach. But the performance of each candidate parameter sets in a different calibration event is not yet evaluated. In this step, we re-apply each candidate parameter set to all events and aggregate performance across the events and sites to score parameter sets. Then the most representative parameter sets are chosen based on the highest score. Normalized NSE is used to score the performance across the events and sites. Here the performance criterion (NSE) is normalized by using the maximum and minimum values of the candidate model parameters sets at each site and event. Then the sum of the normalized NSE over the entire calibration events is estimated for each candidate model parameters sets. The top three potential model parameters sets with the highest total normalized NSE is registered and proposed for model testing and calibration refinement.

3.5. Validation

Validation was performed to test and refine top three model parameter sets selected during calibration process using a data set independent of calibration period. We have selected four validation events (Table 2) that qualify the event selection criteria described Section 2. This phase is dedicated to select the most representative model parameter sets. The model testing and refinement is performed in four new events (Table 2). The three model sets are evaluated by using Taylor Skill Score (Taylor, 2001) to further corroborate the outcome of the previous two step calibration processes. This score summarizes a Taylor diagram and defines a single skill score that measures the correlation coefficient and centered root-mean-squared error along with standard deviation (Taylor, 2001). It is defined as:

$$S = \frac{4(1+R)}{\left(\frac{\sigma_s}{\sigma_o} + \frac{1}{\sigma_s/\sigma_o}\right)^2 (1+R_o)} \tag{10}$$

Where: S indicates the Taylor Skill Score; σ_s is model variance; σ_o is observed variance; R is model correlation coefficient, and R_o is maximum correlation attainable, here taken as the maximum of model's correlation coefficient. The skill increases (approaches one) as σ_s and R get closer to σ_o and R_o respectively.

4. Results and discussion

4.1. Sensitivity analysis

The sensitivity analysis (Figure 3) indicates that Imperviousness (IM) is the most sensitive SWMM5 parameter to NSE, PE, and VE model performances in Humber River watershed. The RSA indexes show that after Imperviousness and Drying time (DT), Depression storage in Impervious areas (SM) and Pervious areas (SP) appear to be slightly sensitive to the model performances, particularly to Peak flow Error. This result is analogous to the plots of Cumulative Sum of the Normalized Reordered Output (CUSUNORO) (Figure 4). The CUSUNORO plots indicate that Imperviousness (IM) followed by Drying time (DT) have the largest first order contribution to NSE, VE, and PE as the departure of their cumulative sum of the normalized output from the horizontal line (y=0) is considerable. The different direction of CUSUNORO plots for NSE, VE and PE indicates that the contribution of each parameter to the mean and variance and the output is positive if above the horizontal and negative if below the horizontal.

The results of both sensitivity analyses are reasonable for semi-urban areas like Humber River watershed, which covers about 50% pervious and 50% impervious areas. The rainfall-runoff response is governed by the percentage of imperviousness in the sub-catchments upstream of the gauging station and recovery time (drying time) of the saturated soil in pervious areas of the sub-catchments. In general, Imperviousness and Depression storage are found to be the most sensitive parameters of SWMM model to peak flow and volume in urbanizing watersheds, which is also supported by Barco et al., (2008). For calibration, the SWMM parameters except Manning's n are considered as it has relatively less impact to NSE and Peak flow in both Impervious and Pervious areas.

Manuscript

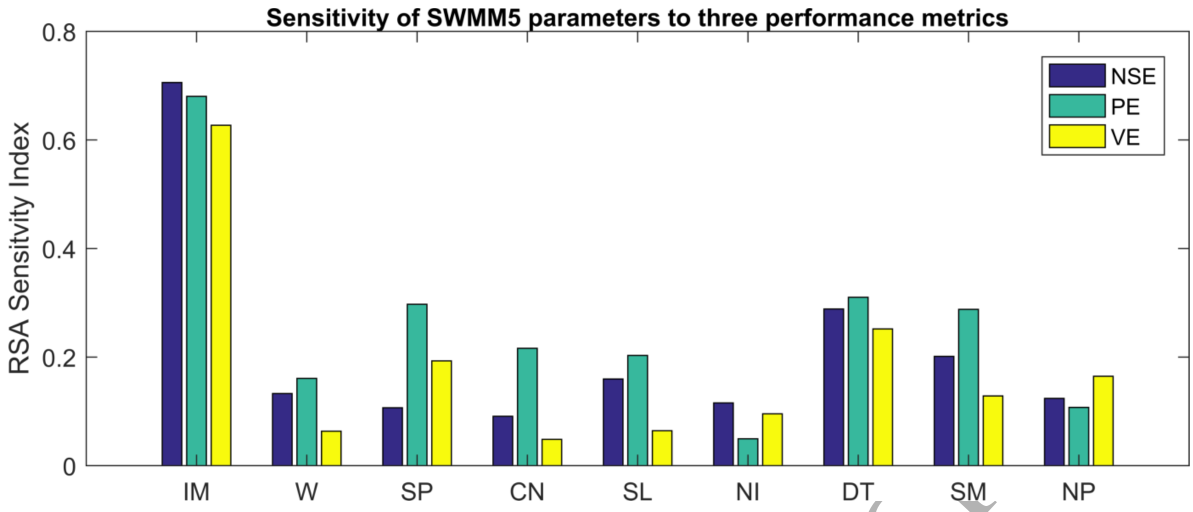


Figure 3: Output of Regionalized Sensitivity Analysis. Figure displays the sensitivity index value of nine SWMM5 parameters for Nash-Sutcliffe Efficiency (NSE), Peak Flow Error (PE) and Volume Error (VE). Higher RSA index corresponds to higher sensitivity of parameters to the output performance. Description of parameter letter codes (x-axis) is presented in Table 1.

Manuscript

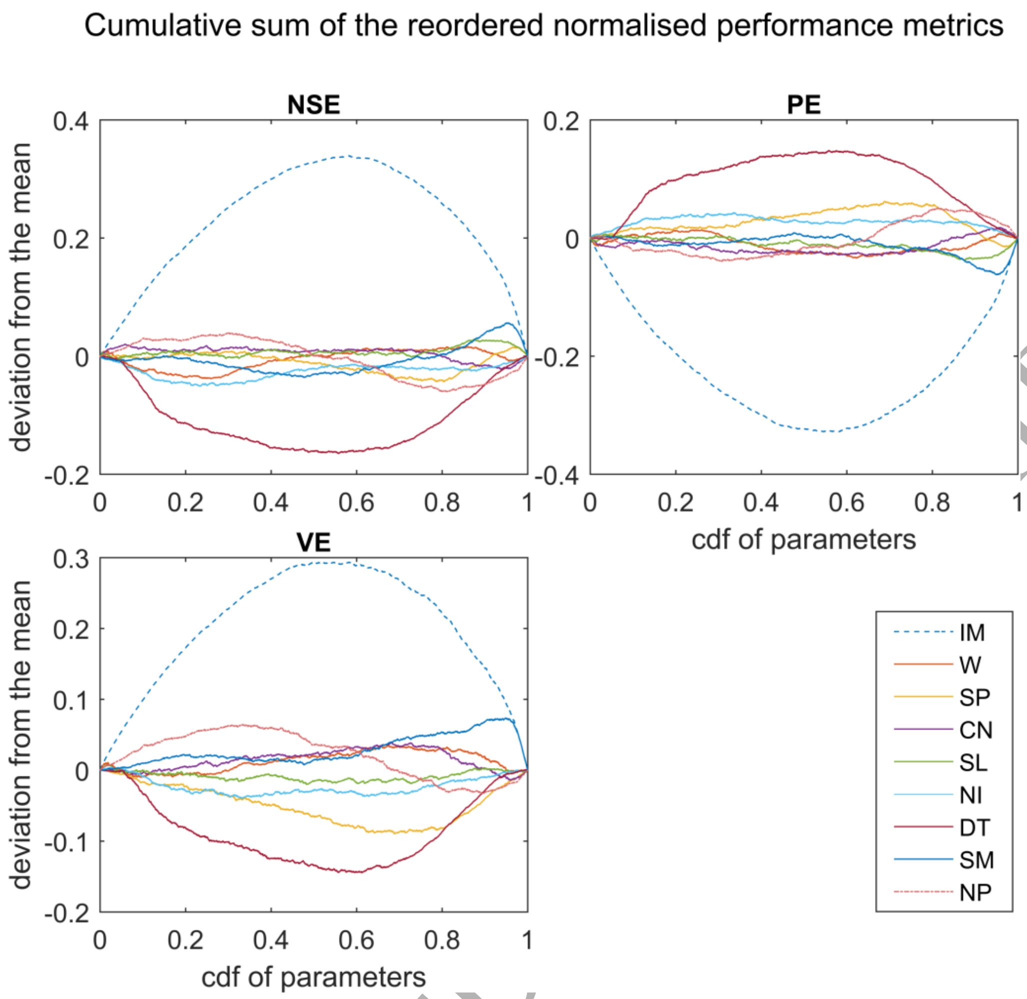


Figure 4: Cumulative Sum of the Normalized Reordered Output (CUSUNORO) used as first order sensitivity of SWMM5 parameters to three performance metrics (NSE, PE and VE). The deviation from the mean (CUSUNORO values or z(i) in Eqn. 5) is plotted against the empirical cumulative distribution of input parameters (x-axis). Higher deviation from the mean indicates higher sensitivity of parameters to corresponding performance metrics. Descriptions of parameter letter codes (for each colored lines of the plots) are presented in Table 1.

4.2. Spatial and temporal parameter variability

The study assessed the degree of uncertainty in event-based calibration of SWMM5 distributed model parameters sets that were obtained by an event-based calibration processes performed for ten calibration events. The parameter variability (uncertainty) was demonstrated by temporal scale (among calibration events) and spatial scale (within 714 sub-catchments). In Figure 5, the spatial variability of the two most sensitive parameters (Imperviousness and Drying Time) that are generated by ten calibrated parameter sets is shown. The medians and the interquartile ranges (IQR) of the box plots in higher percentile imperviousness values show variability between individual calibration events. Lower and medium percentiles values of imperviousness have relatively similar medians and IQRs among the parameter sets. In general, higher uncertainty is observed among the sub-catchments with higher imperviousness (>80% Imperviousness). This result can be reasonably expected from a semi-urban watershed where high impervious areas highly influence the rainfall-runoff response in the time of extreme events. Figure 5 also shows that pervious areas that have relatively faster recovery time to be in a drying state when saturated (<20% Drying Time or less than 5.5 days) shows higher variability or uncertainty. Rapid recovery time is often recognized in hydrologic soil group D such as medium and coarse sandy soils, which pertains to high rate of water transmission or infiltration (Rossman, 2010; NRCS, 2007).

Figure 6 shows the peak flow variability of the ten potential representative SWMM5 model parameter sets in different calibration events. The uncertainty is expressed by standardized peak flow deviation from the observation recorded at multiple gauging stations. The degree of the deviation is quite significant in almost all events and measuring stations. The medians and associated IQRs are either above or below the green horizontal line (observation), which depicts underestimation and overestimation of peak flows by the potential model parameters sets. Outliers were also observed on many occasions. This investigation indicates the existence of high uncertainty in reproducing peak flows by the majority of model parameters sets. Within each boxplot, it can be seen that only one point (one model parameter set) matches (or close to matching) with observed peak flow, which is, in fact, the calibrated model parameter set for each event that the boxplot is constructed. The results of this variability analysis give an overview of the difficulty in selecting representative parameter sets in distributed semi-urban watersheds and the need for a robust method of calibration when dealing with event-based model parametrization.

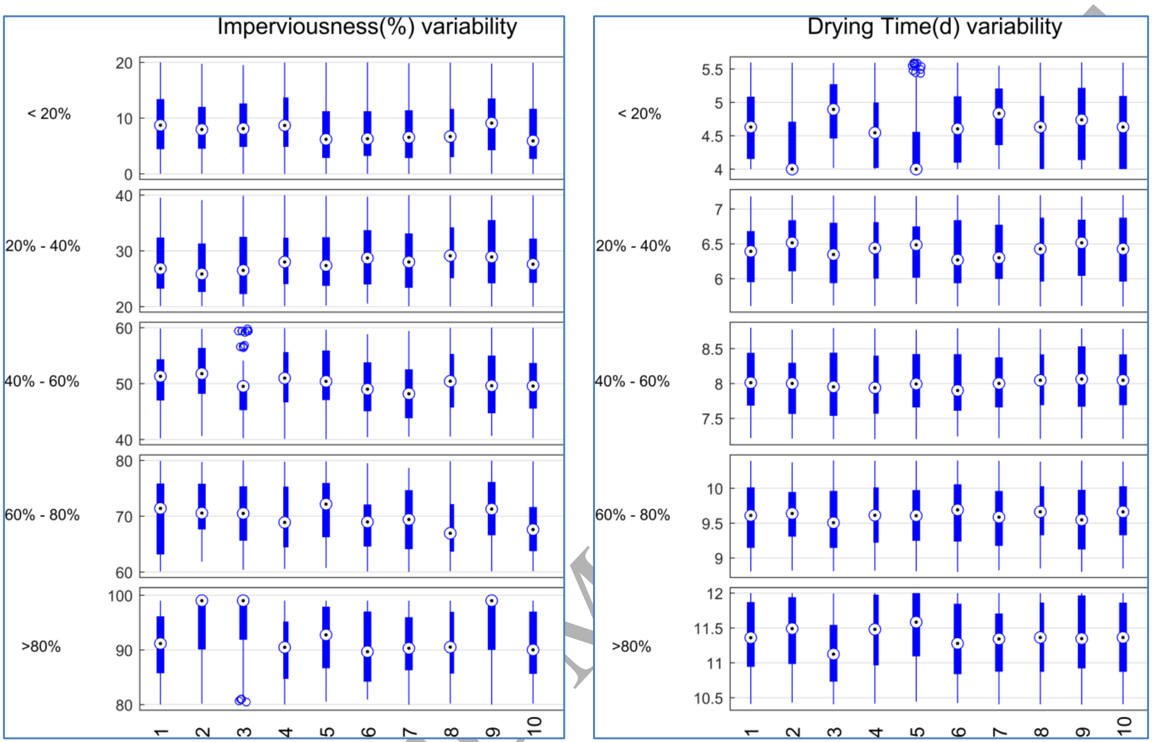


Figure 5: Box plots showing the spread of the lower, middle three and upper percentile values of most sensitive calibrated parameters (Imperviousness-Left and Drying Time-Right) to illustrate their variability in ten Model Sets (x-axis). Parameter values, collected from 714 sub-catchments, were ranked in ascending order. Model parameter sets represent different realization of the PCSWMM model in ten calibration events.

Manuscript

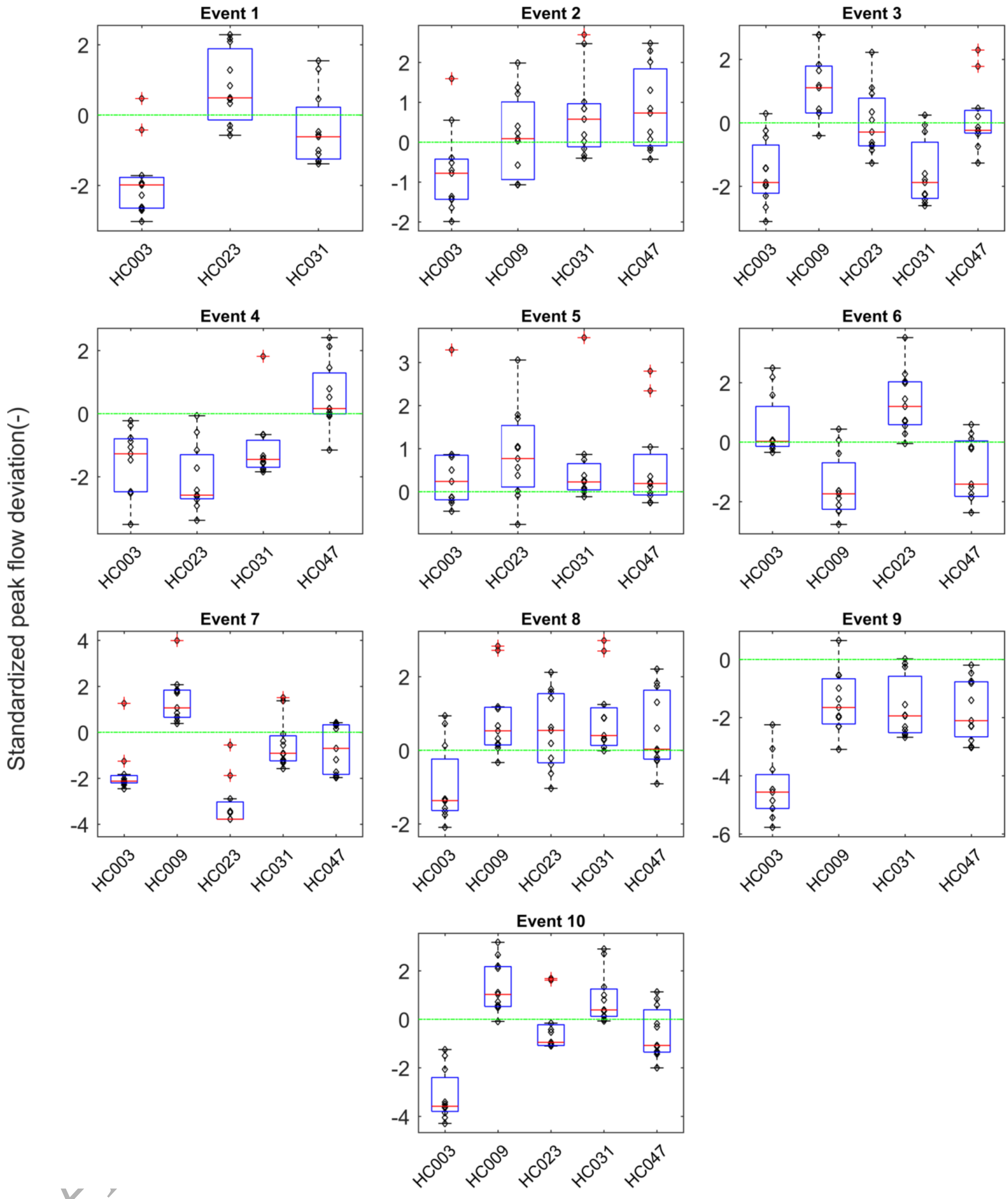


Figure 6: Figure showing Peak Flow variability of model parameter sets. 10 plots are constructed for 10 calibration events and each boxplots within a plot corresponds to different gauging stations. Individual boxplots are developed from 10 standardized peak flows, which are generated by ten different Model Parameter Sets in order to demonstrate the variability of different realizations of SWMM5 model. Standardized peak flows are calculated by normalizing the deviation of the simulated peak flow from observed peak flow by the standard deviation of the simulated peak flow. Green horizontal line along the zero y-axis is computed based on observed peak flow.

4.3. Calibration Approaches

The outputs from the four multi-objective calibration approaches presented in section 3.4.1 (MS-S, MS-A, ME-MS and OU) are evaluated in ten individual calibration events at five gauging stations. Their performances are compared at each calibration steps mentioned in section 3.4.2.

Figure 7 and Figure 8 present the comparison of calibration approaches for the first calibration step. The relative improvement of Multi-site average objective function (MS-A) and Multi-site simultaneous (MS-S) over the benchmark At-catchment outlet (OU) is quantified by the prediction (simulation) error decrease (PED) percentage. The PED (in Figure 7) shows the improvement of NSE of both MS-A and MS-S approaches when compared to OU at five gauging stations. Using either of the multi-site calibration approaches improves the model performance by about 28% in the interior sites when compared to the conventional at catchment outlet calibration method. Comparing the two multi-site optimization methods, aggregating the objective functions over the gauging stations (MS-A) gives a fairly better performance than calibrating the multiple sites simultaneously (MS-S). With a reference to the benchmark OU calibration, the NSE performance metric of MS-A is improved by an average of 43% as compared to MS-S where it was improved by only 29%. In fact, only 4 out of 42 calibration events and stations show slightly higher NSE performance for MS-S; out of which 3 are at the outlet. At the outlet, there are some occasions where the benchmark OU calibration shows improved performance over both MS-S and MS-A. This is a reasonable because it is generally easier to improve the performance at one location during optimizing. The calibrated parameter sets from Multi-event multi-site (ME-MS) calibration approach is re-applied for each calibration event to evaluate and compare its result with the other methods. It is found that the performance of ME-MS is significantly lower than both multi-site optimizations as well the benchmark calibration approach. Although not shown in Figure 7 due to its high percentage difference to present in PED metrics with other calibration approaches, the comparison is shown in Figure 8.

The performance of the four calibration approaches was tested at six calibration events, and statistical comparison is shown by the Taylor Diagram in Figure 8. Confirming the model comparison using PED metrics in Figure 7, the MS-S and MS-A calibration approaches have better statistical proximity and pattern with the observation than ME-MS and OU methods. The Taylor diagrams indicate that MS-A approach has relatively more confined points towards the observation ('OBS' black dot and line) and consistently proves to be a better calibration approach than MS-S and other methods. The multi-event multi-site (ME-MS) optimization has more sparse points away from the 'OBS' proximity and produces an inconsistent performance over the calibration events.

In general, the calibrated model parameters sets generated by multi-site average objective function (MS-A) approach achieved improved model performance (NSE) and statistical measures (standard deviation, root mean squared error and correlation coefficient) during calibration step-1 and hence selected for calibration step-2.

Ten calibrated parameter sets generated by MS-A optimization approach were applied again to each of the ten calibration events and the results were extracted. Figure 9 demonstrates the normalized NSE performance metrics evaluated at five gauging stations. The summation of the normalized NSE over each gauging sites and calibration events indicates that Model parameter Set 5 has the highest performance followed by Model Set 2 and 3. The result indicates that it is fairly reasonable to represent distributed semi-urban watersheds by qualifying model parameter sets generated from multiple even-based calibration process.

With the above results in mind, the DDS algorithm used by MS-A appears to converge to a better approximate true solution than the PADDS algorithm employed by MS-S approach. One of the key reasons is that MS-S result quality is summarized by precisely the objective function being optimized by MS-A. Another reason is likely that when solving the MS-S formulation, PADDS is spending substantial effort to approximate a Pareto-set in five dimensions and as such, PADDS is generating candidate solutions from much diversified parts of parameter space. In contrast, DDS is generating candidate solutions concentrated in the area of parameter space that leads to a good average objective function value.

Manuscript

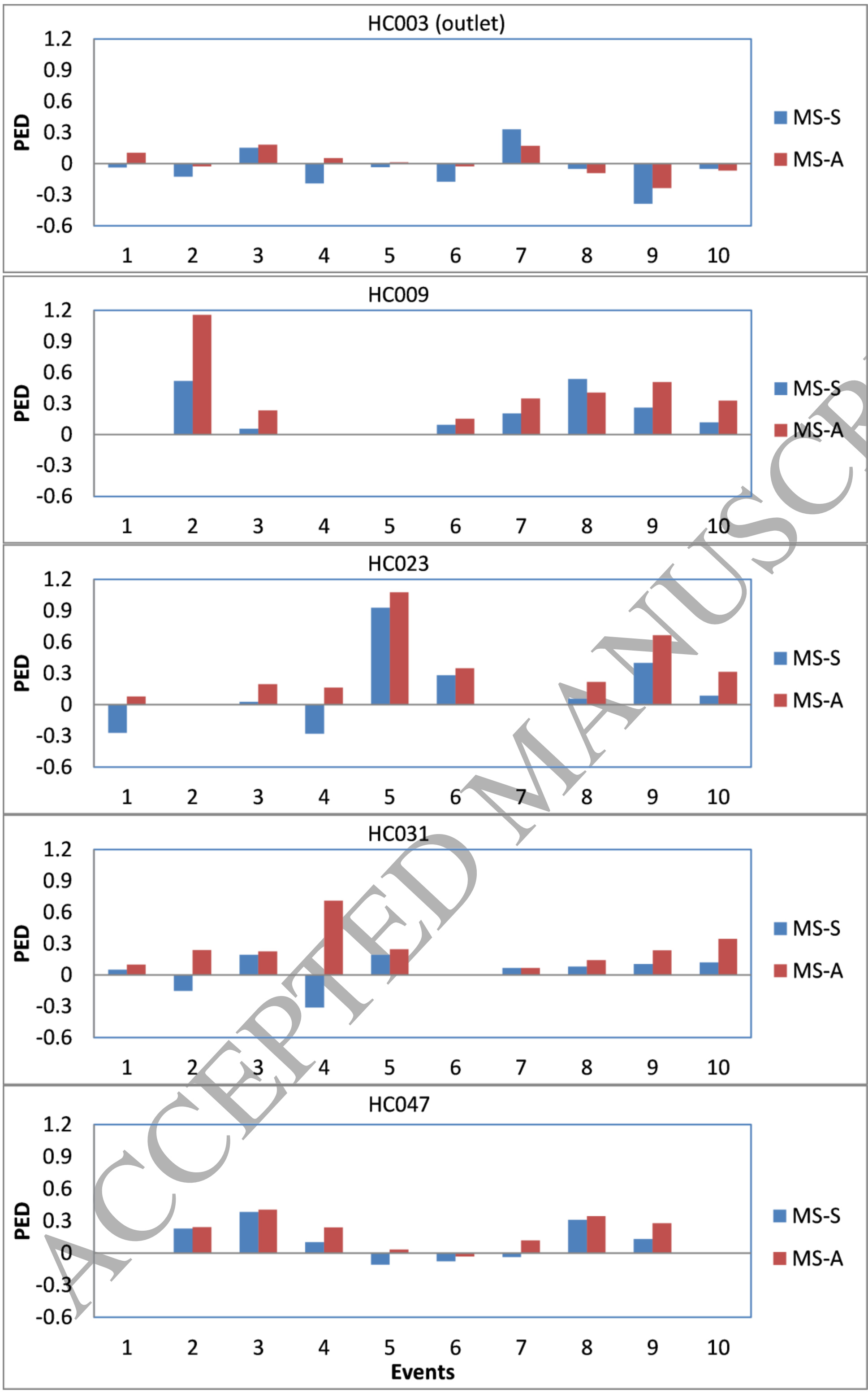


Figure 7: Model Improvement (defined by Prediction Error Decrease in percentage (PED *100)) of Multi-site Simultaneous (MS-S) and Multi-site Average objective function (MS-A) calibration approaches when compared with Catchment Outlet (OU) approach at five gauging stations and ten calibration events.

Manuscript

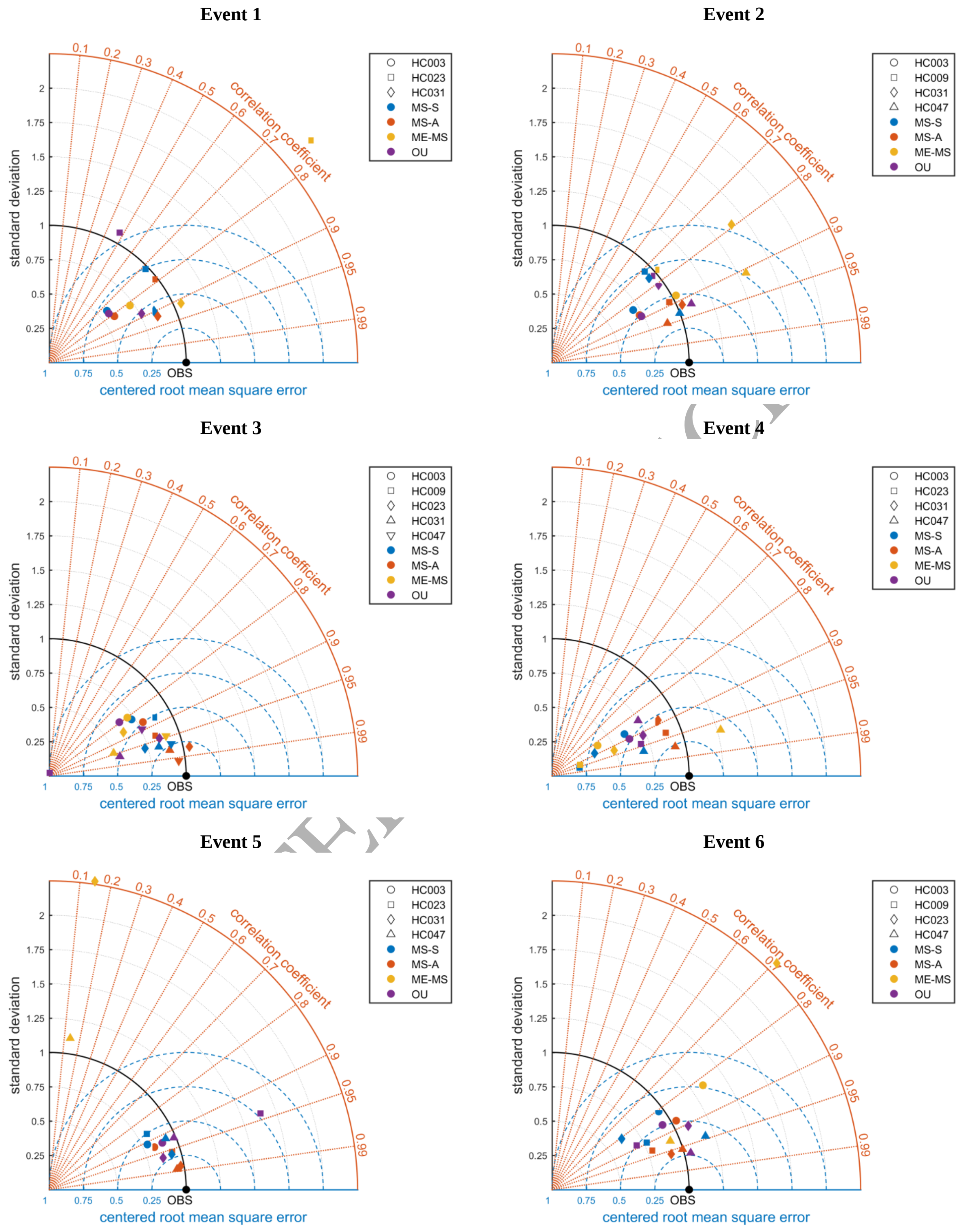


Figure 8: Comparison of Taylor diagrams showing an event-by-event statistical evaluation of simulated flows from four calibration approaches (MS-S, MS-A, ME-MS, & OU) evaluated at six calibration events. The Taylor Diagrams summarized three statistical performances at five gauging stations for each event. Different colors denote respective calibration approaches while different shapes correspond to different stations (gauging sites). Perfect models sets would align themselves closer to the black arc as well as point 'OBS', which depict agreement with observations.

Manuscript

Calibration Events — Metric -> Normalized NSE

Model Set	1			2				3					4				5			
Stations	HC003	HC023	HC031	HC003	HC009	HC031	HC047	HC003	HC009	HC023	HC031	HC047	HC003	HC023	HC031	HC047	HC003	HC023	HC031	HC047
1	1.0	1.0	1.0	1.0	0.9	0.9	0.0	1.0	0.7	0.8	0.4	0.5	0.7	0.1	0.6	0.0	1.0	1.0	0.9	0.0
2	0.7	0.8	0.9	1.0	1.0	1.0	1.0	0.9	0.7	1.0	0.2	0.4	0.4	0.8	0.3	0.9	0.9	0.9	1.0	0.9
3	0.7	0.9	0.3	0.7	0.9	0.2	0.7	0.9	1.0	1.0	1.0	1.0	0.9	0.6	1.0	0.9	0.5	1.0	0.1	1.0
4	0.4	0.2	0.1	0.4	0.2	0.0	0.9	0.6	0.1	0.5	1.0	0.7	1.0	1.0	1.0	1.0	0.5	0.3	0.0	0.8
5	0.5	1.0	0.8	0.9	0.9	1.0	0.9	0.7	0.9	0.9	0.3	0.8	0.8	0.3	0.7	1.0	0.8	1.0	1.0	0.9
6	0.4	1.0	0.8	0.9	0.5	1.0	0.4	0.7	0.4	0.7	0.0	0.6	0.0	0.0	0.1	0.5	0.6	1.0	0.9	0.1
7	0.5	0.0	0.8	0.9	0.9	0.9	0.9	0.7	0.4	0.0	0.4	0.4	0.5	0.8	0.4	0.9	0.9	0.0	1.0	0.8
8	0.3	1.0	0.7	0.8	0.9	1.0	1.0	0.7	0.0	0.9	0.0	0.0	0.1	0.2	0.0	0.9	0.2	1.0	0.9	0.8
9	0.0	0.7	0.0	0.0	0.0	0.1	0.1	0.0	0.1	0.9	0.8	0.7	1.0	0.8	0.8	0.6	0.5	0.8	0.5	0.6
10	0.3	1.0	0.7	0.7	0.9	1.0	0.8	0.7	0.5	0.9	0.1	0.5	0.1	0.3	0.0	0.9	0.0	1.0	0.9	0.4

Model Set	6				7				8					9					10				SUM	Rank
Stations	HC003	HC009	HC023	HC047	HC003	HC009	HC031	HC047	HC003	HC009	HC023	HC031	HC047	HC003	HC009	HC023	HC031	HC047	HC003	HC009	HC023	HC031		
1	0.9	0.9	1.0	0.0	0.9	0.6	0.9	0.0	1.0	0.8	0.9	0.9	0.0	0.7	0.9	0.2	0.4	0.5	1.0	1.0	0.9	0.0	28	5
2	1.0	0.5	0.7	0.8	1.0	0.6	1.0	0.3	1.0	0.9	0.7	1.0	0.8	0.4	0.9	0.8	0.2	0.3	1.0	0.7	1.0	0.5	32	2
3	0.2	0.4	0.9	0.9	0.9	0.7	0.1	0.3	0.6	0.8	1.0	0.2	1.0	0.9	0.9	0.7	0.7	0.8	0.7	0.7	0.2	1.0	30	3
4	0.0	0.1	0.3	0.9	0.8	0.0	0.0	0.4	0.0	0.0	0.3	0.0	0.8	1.0	0.9	0.8	0.6	0.4	0.6	0.2	0.0	0.6	19	10
5	0.9	0.6	1.0	1.0	0.9	0.5	1.0	0.3	1.0	0.9	1.0	1.0	0.9	0.3	0.9	0.5	0.2	0.5	0.8	0.7	1.0	0.8	**33**	**1**
6	1.0	0.8	1.0	0.3	0.9	0.6	0.9	0.2	0.9	0.7	0.9	1.0	0.4	0.2	1.0	0.0	0.0	0.6	0.8	0.9	1.0	0.1	25	7
7	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	0.6	1.0	0.7	0.9	0.9	0.5	0.9	0.3	0.5	0.0	1.0	1.0	0.9	0.4	29	4
8	0.7	0.0	1.0	0.5	0.9	0.8	1.0	0.4	1.0	1.0	1.0	1.0	1.0	0.1	0.0	0.4	0.0	0.3	0.1	0.0	1.0	0.3	24	8
9	0.3	0.4	0.8	0.2	0.0	0.4	0.9	0.1	0.0	0.8	0.0	1.0	0.4	0.8	0.0	1.0	1.0	1.0	0.3	0.1	0.2	0.9	20	9
10	0.7	0.1	1.0	0.6	0.8	1.0	1.0	0.9	1.0	1.0	1.0	1.0	0.9	0.0	0.0	0.5	0.0	0.5	1.0	1.0	1.0	1.0	28	6

Normalized NSE

1 Good — 0 Bad

Figure 9: Performance ranking of 10 model parameter sets in ten calibration events. Normalized Nash-Sutcliffe Efficiency index (NSE) is used to score the performances at each gauging stations with sum over all sites and over all events displayed on the right side. Highest score corresponds to best performing model parameter set and vice versa. The heatmap shows the Normalized NSE values according to color palette displayed at the bottom side.

Manuscript

		Validation Events																				
		1					2					3					4					
Metric ->		Taylor Skill Score																			SUM	Rank
Model Sets	2	0.90	0.73	0.37	0.90	0.91	0.95	0.88	0.64	0.90	0.42	0.46	0.74	0.96	0.33	0.79	0.97	0.03	0.79	1.00	0.42	14	2
	3	1.00	0.87	0.35	0.56	0.99	0.83	0.58	0.64	0.48	0.37	0.44	0.71	0.67	0.36	0.87	0.88	0.00	0.96	0.09	0.22	12	3
	5	0.76	0.91	0.30	0.87	0.98	0.95	0.68	0.79	0.91	0.56	0.88	1.00	0.88	0.95	0.96	0.98	0.81	0.91	0.95	0.25	**16**	**1**
Stations		HC003	HC009	HC023	HC031	HC047	HC003	HC009	HC023	HC031	HC047	HC003	HC009	HC023	HC031	HC047	HC003	HC009	HC023	HC031	HC047		

Taylor Skill Score

1 Good — 0 Bad

Figure 10: Model validation of top three model sets of MS-A approach in different events. The Taylor Skill Scores are evaluated at each of the five gauging sites for four different events. Most skillful models would have a score of 1 and the least ones have a score of 0.

4.4. Validation

To verify the outcome of the above calibration processes, the top three model parameter sets (Model Set 5, 2 and 3) were evaluated at validation events because their performance from calibration step 2 are not significantly different (Summation of Normalized NSE: 30, 32 and 33 in Figure 9). The Taylor skill score was used to evaluate these SWMM5 model parameter sets at multiple sites and results are presented in

		Validation Events																				
		1					2					3					4					
Metric ->		Taylor Skill Score																			SUM	Rank
Model Sets	2	0.90	0.73	0.37	0.90	0.91	0.95	0.88	0.64	0.90	0.42	0.46	0.74	0.96	0.33	0.79	0.97	0.03	0.79	1.00	0.42	14	2
	3	1.00	0.87	0.35	0.56	0.99	0.83	0.58	0.64	0.48	0.37	0.44	0.71	0.67	0.36	0.87	0.88	0.00	0.96	0.09	0.22	12	3
	5	0.76	0.91	0.30	0.87	0.98	0.95	0.68	0.79	0.91	0.56	0.88	1.00	0.88	0.95	0.96	0.98	0.81	0.91	0.95	0.25	**16**	**1**
Stations		HC003	HC009	HC023	HC031	HC047	HC003	HC009	HC023	HC031	HC047	HC003	HC009	HC023	HC031	HC047	HC003	HC009	HC023	HC031	HC047		

Taylor Skill Score

1 Good — 0 Bad

Figure 10. Based on the scores, Model Set 5 appears to be more skillful than Model Set 2 and 3 as its score is close to 1 for majority of gauging stations and events. The summation of the Taylor Score over the gauges and evens (Sum=16) is the highest. Conversely Model Set 2 and 3 have lower scores because Taylor Skill Score penalizes models with little statistical pattern similarity and weak correlation with observations. In general, Taylor Skill Score is found to be a precise evaluation tool to select skillful SWMM5 model parameter sets that could represent the distributed watershed in space and time.

5. Conclusion

A proposed event-based calibration approach integrating multi-site and multi-objective optimizations was used to select representative SWMM5 model parameter sets in a distributed semi-urban watershed. We compared the performance of four calibration approaches in reproducing the desired spring flow responses at interior sites of Humber River Watershed. These are Multi-site simultaneous (MS-S), Multi-site average objective function (MS-A), Multi-event multi-site (ME-MS) and a benchmark At-catchment outlet (OU) calibration approaches. MS-S and OU approaches utilized PA-DDS optimization algorithm, whereas the others applied DDS algorithm.

A spatio-temporal variability of calibrated model parameter sets among different calibration events was initially assessed in anticipation of capturing the uncertainty of event-based parametrization. The results indicated that there is considerable uncertainty in calibrating highly impervious sub-catchments (>80% Imperviousness) and pervious areas with rapid recovery time (< 5.5 days of Drying Time). Another remark from the variability analysis is the presence of uncertainty in peak flow response by the model parameter sets. The uncertainty in reproducing peak flows by the majority of model parameters sets at multiple interior sites is a clear indication of a need for a robust calibration approaches in event-based distributed models.

The output from the proposed calibration approaches and steps demonstrated that multi-site average objective function (MS-A) and multi-site simultaneous (MS-S) calibration approaches showed superior performances against the Multi-event multi-site and benchmark calibration approaches. The desired flows at interior upstream sites were better reproduced using MS-A and MS-S methods as compared to calibrating using the outlet (OU); a finding similar to Leta et al., (2017).

Most importantly, aggregating the objective functions across the multiple sites into a single objective function (MS-A) outperformed the multi-site simultaneous (MS-S) approach. Individually calibrated model parameter sets from MS-A calibration approach shows significant improvement of NSE performance metrics when compared to MS-S at the majority of stations. This is also supported by Taylor diagrams, which demonstrated that the MS-A approach attained better statistical pattern and amplitude of observed hydrographs. Using MS-A method, ten parameter sets extracted from ten individual calibration events were cross-tested again at all events in the second calibration step. This step was able to identify the top three parameter sets out of ten potential model sets using their aggregated normalized NSE estimated at multiple sites. Model parameter sets 5 followed by 2 and 3 appear to outperform the rest of the model parameter sets. Validation was made at four different events to test the statistical performances using Taylor Skill Scores. And the result indicates that Model Parameter Set 5, which is calibrated using MS-A approach, is the most skillful and representative SWMM5 model parameter set in the study area.

In General, using the single objective DDS algorithm in MS-A approach to find the best average NSE of five gauging stations in the catchment area is found to be more efficient than using the multi-objective PA-DDS algorithm to find non-dominated Pareto-front of five NSE performances.

The study discovered that combination of efficient optimization tools with a series of calibration approaches is important in finding candidate parameters sets and representing distributed catchments by event-based hydrological models. The study takes advantage of the DDS and PA-DDS algorithms to select non-dominated solutions and representative model parameter sets. Finally, the authors strongly believe that the methods and calibration approaches employed in this research can also be applied in other watersheds. An interesting result from the study is that averaging/aggregating objective functions during calibration provide better simulation output, which can be applied for any cases.

Manuscript

Acknowledgments: This work was supported by the Natural Science and Engineering Research Council (NSERC) Canadian FloodNet (Grant number: NETGP 451456).

Manuscript

6. Reference

Ackerman, D., Schiff, K. C., & Weisberg, S. B. (2005). EVALUATING HSPF IN AN ARID, URBANIZED WATERSHED. *Journal of the American Water Resources Association*, *41*(2), 477–486. http://doi.org/10.1111/j.1752-1688.2005.tb03750.x

Ajmal, M., Waseem, M., Ahn, J.-H., & Kim, T.-W. (2016). Runoff Estimation Using the NRCS Slope-Adjusted Curve Number in Mountainous Watersheds. *Journal of Irrigation and Drainage Engineering*, *142*(4), 4016002. http://doi.org/10.1061/(ASCE)IR.1943-4774.0000998

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assesment Part I: Model development. *JAWRA Journal of the American Water Resources Association*, *34*(1), 73–89. http://doi.org/10.1111/j.1752-1688.1998.tb05961.x

Asadzadeh, M., Razavi, S., Tolson, B. A., & Fay, D. (2014). Pre-emption strategies for efficient multi-objective optimization: Application to the development of Lake Superior regulation plan. *Environmental Modelling and Software*, *54*, 128–141. http://doi.org/10.1016/j.envsoft.2014.01.005

Asadzadeh, M., & Tolson, B. (2012). Hybrid Pareto archived dynamically dimensioned search for multi-objective combinatorial optimization: application to water distribution network design. *Journal of Hydroinformatics*, *14*(1), 192–205. http://doi.org/10.2166/hydro.2011.098

Asadzadeh, M., & Tolson, B. (2013). Pareto archived dynamically dimensioned search with hypervolume-based selection for multi- objective optimization. *Engineering Optimization ISSN:*, *45*(12), 1489–1509. http://doi.org/10.1080/0305215X.2012.748046

Asadzadeh, M., & Tolson, B. A. (2009). A New Multi-Objective Algorithm , Pareto Archived DDS. In *11th Annual Conference Companion on Genetic and Evolutionary Computation Conference* (pp. 1963–1966). Montreal, Quebec, Canada: Late Breaking Papers.

Barco, J., Wong, K. M., Stenstrom, M. K., & Asce, F. (2008). Automatic Calibration of the U.S. EPA SWMM Model for a Large Urban Catchment. *J. Hydraul. Eng.*, *134*(4), 466–474. http://doi.org/10.1061/ASCE0733-94292008134:4466

Beven, K. (1989). CHANGING IDEAS IN HYDROLOGY-THE CASE OF PHYSICALLY-BASED MODELS. *Journal of Hydrology Elsevier Science Publishers B.V*, *105*, 157–172. Retrieved from http://ac.els-cdn.com/0022169489901017/1-s2.0-0022169489901017-main.pdf?_tid=2565f7b2-4a06-11e7-a4f7-00000aab0f01&acdnat=1496677771_7917c0ab5092577de570d5147ff94df8

Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, *249*(1–4), 11–29. http://doi.org/10.1016/S0022-1694(01)00421-8

Bolado-Lavin, R., Castaings, W., & Tarantola, S. (2009). Contribution to the sample mean plot for graphical and numerical sensitivity analysis. *Reliability Engineering & System Safety*, *94*(6), 1041–1049. http://doi.org/10.1016/j.ress.2008.11.012

Brocca, L., Melone, F., & Moramarco, T. (2011). Distributed rainfall-runoff modelling for flood frequency estimation and flood forecasting. *Hydrological Processes*, *25*(18), 2801–2813.

http://doi.org/10.1002/hyp.8042

Cao, W., Bowden, W. B., Davie, T., & Fenemor, A. (2006). Multi-variable and multi-site calibration and validation of SWAT in a large mountainous catchment with high spatial variability. *Hydrological Processes*, *20*(5), 1057–1073. http://doi.org/10.1002/hyp.5933

CIVICA & TRCA. (2015). *Humber River Hydrology Update*. Retrieved from https://trca.ca/wp-content/uploads/2016/07/Humber-Hydrology-Update-Final-Report-v19.1.pdf

Coulibaly, P. (2003). Impact of meteorological predictions on real-time spring flow forecasting. *Hydrological Processes*, *17*(18), 3791–3801. http://doi.org/10.1002/hyp.5168

Cuo, L., Lettenmaier, D. P., Mattheussen, B. V, Storck, P., & Wiley, M. (2008). Hydrologic prediction for urban watersheds with the Distributed Hydrology–Soil–Vegetation Model. *Hydrological Processes*, *22*, 4205–4213. http://doi.org/10.1002/hyp.7023

Deshmukh, D. S., Chaube, U. C., Ekube Hailu, A., Aberra Gudeta, D., & Tegene Kassa, M. (2013). Estimation and comparision of curve numbers based on dynamic land use land cover change, observed rainfall-runoff data and land slope. *Journal of Hydrology*, *492*, 89–101. http://doi.org/10.1016/j.jhydrol.2013.04.001

El Hassan, A. A., Sharif, H. O., Jackson, T., & Chintalapudi, S. (2013). Performance of a conceptual and physically based model in simulating the response of a semi-urbanized watershed in San Antonio, Texas. *Hydrological Processes*, *27*(24), 3394–3408. http://doi.org/10.1002/hyp.9443

Engeland, K., Braud, I., Gottschalk, L., & Leblois, E. (2006). Multi-objective regional modelling. *Journal of Hydrology*, *327*(3–4), 339–351. http://doi.org/10.1016/j.jhydrol.2005.11.022

Fletcher, T. D., Andrieu, H., & Hamel, P. (2013). Understanding, management and modelling of urban hydrology and its consequences for receiving waters: A state of the art. *Advances in Water Resources*, *51*, 261–279. http://doi.org/10.1016/j.advwatres.2012.09.001

Furl, C., Sharif, H. O., El Hassan, A., Mazari, N., Burtch, D., Mullendore, G. L., … Mullendore, G. L. (2015). Hydrometeorological Analysis of Tropical Storm Hermine and Central Texas Flash Flooding, September 2010. *Journal of Hydrometeorology*, *16*(6), 2311–2327. http://doi.org/10.1175/JHM-D-14-0146.1

Gironás, J., Asce, A. M., Niemann, J. D., Asce, M., Roesner, L. A., Asce, F., … Andrieu, H. (2010). Evaluation of Methods for Representing Urban Terrain in Storm-Water Modeling. *J. Hydrol. Eng.*, *15*(1), 1–14. http://doi.org/10.1061/ASCEHE.1943-5584.0000142

Haghnegahdar, A., Tolson, B. A., Davison, B., Seglenieks, F. R., Klyszejko, E., Soulis, E. D., … Matott, L. S. (2014). Calibrating Environment Canada's MESH Modelling System over the Great Lakes Basin. *Atmosphere-Ocean*, *52*(4), 281–293. http://doi.org/10.1080/07055900.2014.939131

Hay, L. E., Leavesley, G. H., Clark, M. P., Markstrom, S. L., Viger, R. J., & Umemoto, M. (2006). Step Wise Multiple Objective Calibration of a Hydrologic Model for a Snowmelt Dominated Basin. *Journal of the American Water Resources Association (JAWRA)*, *42*(4), 877–890. Retrieved from ftp://brrcrftp.cr.usgs.gov/pub/mows/pubs/leavesley_pubs/leavesley_pdf/j.1752-1688.2006.tb04501.x.pdf

Manuscript

Herrera, M., I. Heathcote, W. J. and A. B. (2006). Multi-Objective Calibration of SWMM for Improved Simulation of the Hydrologic Regime. *Journal of Water Management Modeling*, *R225-15*. http://doi.org/10.14796/JWMM.R225-15

Huber, W. C., & Dickinson, R. E. (1988). *STORM WATER MANAGEMENT MODEL, VERSION 4: USER'S MANUAL*. Athens, GA. http://doi.org/EPA/600/3-88/001a

Irvine, K. N., Loganathan, B. G., Pratt, E. J., & Sikka, H. C. (1993). Calibration of PCSWMM to Estimate Metals, PCBs and HCB in CSOs from an Industrial Sewershed. *Journal of Water Management Modeling*, *R175*(10). http://doi.org/I 0.14796/JWMM.RI75-10.

Jacobson, C. R. (2011). Identification and quantification of the hydrological impacts of imperviousness in urban catchments: A review. *Journal of Environmental Management*, *92*(6), 1438–1448. http://doi.org/10.1016/j.jenvman.2011.01.018

James, W. (2005). *Rules for Responsible Modeling-4th Edition* (4th ed.). Guelph, Ontario: CHI (Computational Hydraulics International). Retrieved from http://chiwater.com/Files/R184_CHI_Rules.pdf

Kärnä, T., & Baptista, A. M. (2016). Evaluation of a long-term hindcast simulation for the Columbia River estuary. *Ocean Modelling*, *99*, 1–14. http://doi.org/10.1016/j.ocemod.2015.12.007

Kennedy, J. R., Goodrich, D. C., Asce, M., & Unkrich, C. L. (2013). Using the KINEROS2 Modeling Framework to Evaluate the Increase in Storm Runoff from Residential Development in a Semiarid Environment. *JOURNAL OF HYDROLOGIC ENGINEERING ©*, *18*, 698–706. http://doi.org/10.1061/(ASCE)HE.1943-5584.0000655

Khu, S.-T., Di Pierro, F., Savi, D., Djordjevi, S., & Walters, G. A. (2006). Incorporating spatial and temporal information for urban drainage model calibration: An approach using preference ordering genetic algorithm. *Advances in Water Resources 29*, *29*(8), 1168–1181. http://doi.org/10.1016/j.advwatres.2005.09.009

Khu, S.-T., Madsen, H., & di Pierro, F. (2008). Incorporating multiple observations for distributed hydrologic model calibration: An approach using a multi-objective evolutionary algorithm and clustering. *Advances in Water Resources*, *31*(10), 1387–1398. http://doi.org/10.1016/j.advwatres.2008.07.011

Krauße, Cullmann, J., Saile, P., & Schmitz, G. H. (2012). Robust multi-objective calibration strategies – possibilities for improving flood forecasting. *Hydrol. Earth Syst. Sci*, *16*, 3579–3606. http://doi.org/10.5194/hess-16-3579-2012

Krebs, G., Kokkonen, T., Valtanen, M., Koivusalo, H., & Setaïa, H. (2013). A high resolution application of a stormwater management model (SWMM) using genetic parameter optimization. *Urban Water Journal*, *10*(6), 394–410. http://doi.org/10.1080/1573062X.2012.739631

Leta, O. T., van Griensven, A., & Bauwens, W. (2017). Effect of Single and Multisite Calibration Techniques on the Parameter Estimation, Performance, and Output of a SWAT Model of a Spatially Heterogeneous Catchment. *Journal of Hydrologic Engineering*, *22*(3), 5016036. http://doi.org/10.1061/(ASCE)HE.1943-5584.0001471

Liong, S.-Y., Chan, W. T., & ShreeRam, J. (1995). Peak-Flow Forecasting with Genetic Algorithm and

SWMM. *Journal of Hydraulic Engineering*, *121*(8), 613–617. http://doi.org/10.1061/(ASCE)0733-9429(1995)121:8(613)

Madsen, H. (2003). Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Advances in Water Resources*, *26*(2), 205–216. http://doi.org/10.1016/S0309-1708(02)00092-1

Madsen, H., Wilson, G., & Ammentorp, H. C. (2002). Comparison of different automated strategies for calibration of rainfall-runoff models. *Journal of Hydrology*, *261*(1), 48–59. http://doi.org/10.1016/S0022-1694(01)00619-9

Matott, L. S. (2005). *OSTRICH : An Optimization Software Tool ; Documentation and User ' s Guide, Version 1.6*. Buffalo.

Mediero, L., Garrote, L., & Martín-Carrasco, F. J. (2011). Probabilistic calibration of a distributed hydrological model for flood forecasting. *Hydrological Sciences Journal*, *56*(7), 1129–1149. http://doi.org/10.1080/02626667.2011.610322

Miller, J. D., Kim, H., Kjeldsen, T. R., Packman, J., Grebby, S., & Dearden, R. (2014). Assessing the impact of urbanization on storm runoff in a peri-urban catchment using historical change in impervious cover. *Journal of Hydrology*, *515*, 59–70. http://doi.org/10.1016/j.jhydrol.2014.04.011

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. http://doi.org/10.1016/0022-1694(70)90255-6

Niemi, T. J., Warsta, L., Taka, M., Hickman, B., Pulkkinen, S., Krebs, G., … Kokkonen, T. (2017). Applicability of open rainfall data to event-scale urban rainfall-runoff modelling. *Journal of Hydrology*, *547*, 143–155. http://doi.org/10.1016/j.jhydrol.2017.01.056

NRCS. (2007). Chapter 7 Hydrologic Soil Groups. In *National Engineering Handbook, Part 630 Hydrology*. United States Department of Agriculture. Retrieved from https://directives.sc.egov.usda.gov/OpenNonWebContent.aspx?content=17757.wba

NVCA. (2006). *Innisfil Creek Subwatershed Plan: APPENDIX D - Hydrologic Modeling*. Utopia, ON. Retrieved from https://www.nvca.on.ca/Shared Documents/Innisfil Creek Appendices D-F.pdf

Ogden, F. L., Sharif, H. O., Senarath, S. U. S., Smith, J. A., Baeck, M. L., & Richardson, J. R. (2000). Hydrologic analysis of the Fort Collins, Colorado, flash flood of 1997. *Journal of Hydrology*, *228*(1–2), 82–100. http://doi.org/10.1016/S0022-1694(00)00146-3

Ozdemir, A., Leloglu, U. M., & Abbaspour, K. C. (2017). Hierarchical approach to hydrological model calibration. *Environmental Earth Sciences*, *76*(8), 318. http://doi.org/10.1007/s12665-017-6560-6

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*, *11*, 1279–1294. http://doi.org/10.5194/hessd-3-2473-2006

Plischke, E. (2012). An adaptive correlation ratio method using the cumulative sum of the reordered output. *Reliability Engineering and System Safety*, *107*, 149–156.

http://doi.org/10.1016/j.ress.2011.12.007

Pokhrel, P., & Gupta, H. V. (2010). On the use of spatial regularization strategies to improve calibration of distributed watershed models. *Water Resources Research*, *46*(1), 1–17. http://doi.org/10.1029/2009WR008066

Randall, M., James, R., James, W., Finney, K., & Heralall, M. (2014). PCSWMM Real Time Flood Forecasting – Toronto, Canada. *PCSWMM Real Time Flood Forecasting*. Retrieved from http://academicworks.cuny.edu/cc_conf_hic

Robert, W. C., Robert James, W. C., Heralall, M., & James, W. (2008). Integrated Web-Based Automated Radar Acquisition, Processing and Real-time Modeling for Flow/Flood Forecasting. Retrieved from https://web.sbe.hw.ac.uk/staffprofiles/bdgsa/11th_International_Conference_on_Urban_Drainage_CD/ICUD08/pdfs/748.pdf

Rossman, L. A. (2010). *Storm Water Management Model User's Manual Version 5.0*. Cincinnati, OH. http://doi.org/EPA/600/R-05/040

Rossman, L., & Huber., W. (2015). *Storm Water Management Model Reference Manual: Volume I-Hydrology (Revised)*. Washington, DC. http://doi.org/EPA/600/R-15/162A

Sharif, H. O., Chintalapudi, S., Hassan, A. A., Xie, H., & Zeitler, J. (2013). Physically Based Hydrological Modeling of the 2002 Floods in San Antonio, Texas. *Journal of Hydrologic Engineering*, *18*(2), 228–236. http://doi.org/10.1061/(ASCE)HE.1943-5584.0000475

Sharif, H. O., Sparks, L., Hassan, A. A., Zeitler, J., & Xie, H. (2010). Application of a Distributed Hydrologic Model to the November 17, 2004, Flood of Bull Creek Watershed, Austin, Texas. *Journal of Hydrologic Engineering*, *15*(8), 651–657. http://doi.org/10.1061/(ASCE)HE.1943-5584.0000228

Shinma, T. A., & Reis, L. F. R. (2014). Incorporating multi-event and multi-site data in the calibration of SWMM. *Procedia Engineering*, *70*, 75–84. http://doi.org/10.1016/j.proeng.2014.02.010

Singh, S., & Bárdossy, A. (2015). Hydrological Model Calibration by Sequential Replacement of Weak Parameter Sets Using Depth Function. *Hydrology*, *2*(2), 69–92. http://doi.org/10.3390/hydrology2020069

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. http://doi.org/10.1016/0043-1354(80)90040-8

Sun, N., Hall, M., Hong, B., & Zhang, L. (2014). Impact of SWMM Catchment Discretization: Case Study in Syracuse, New York. *Journal of Hydrologic Engineering*, *19*(1), 223–234. http://doi.org/10.1061/(ASCE)HE.1943-5584.0000777

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, *106*(D7), 7183–7192. http://doi.org/10.1029/2000JD900719

Tolson, B. A., & Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, *43*(1), 1–16. http://doi.org/10.1029/2005WR004723

Tramblay, Y., Bouaicha, R., Brocca, L., Dorigo, W., Bouvier, C., Camici, S., & Servat, E. (2012). Estimation of antecedent wetness conditions for flood modelling in northern Morocco. *Hydrology and Earth System Sciences*, *16*(11), 4375–4386. http://doi.org/10.5194/hess-16-4375-2012

TRCA. (2008). *Humber River Watershed Scenario Modelling and Analysis Report*. Downsview, ON. Retrieved from http://www.trca.on.ca/dotAsset/196551.pdf

TRCA. (2013). *Humber River Watershed Report Card 2013*. Retrieved from https://trca.ca/wp-content/uploads/2016/04/2173_WatershedReportCards_Humber_rev15_forWeb.pdf

TRCA, & AMEC. (2012). HYDROLOGIC IMPACTS OF FUTURE DEVELOPMENT ON FLOOD FLOWS AND MITIGATION REQUIREMENTS IN THE HUMBER RIVER WATERSHED-Draft Report Submited by AMEC Environment & Infrastructure. Toronto, ON, Canada.

Wang, S., Zhang, Z., Sun, G., Strauss, P., Guo, J., Tang, Y., & Yao, A. (2012). Multi-site calibration, validation, and sensitivity analysis of the MIKE SHE Model for a large watershed in northern China. *Hydrol. Earth Syst. Sci*, *16*, 4621–4632. http://doi.org/10.5194/hess-16-4621-2012

Wi, S., Yang, Y. C. E., Steinschneider, S., Khalil, A., & Brown, C. M. (2015). Calibration approaches for distributed hydrologic models in poorly gaged basins: implication for streamflow projections under climate change. *Hydrol. Earth Syst. Sci*, *19*, 857–876. http://doi.org/10.5194/hess-19-857-2015

WMO. (2011). *Manual on Flood Forecasting and Warning: WMO-No. 1072*. Geneva. Retrieved from http://www.wmo.int/pages/prog/hwrp/publications/flood_forecasting_warning/WMO 1072_en.pdf

Xia, Y., Pitman, A. J., Gupta, H. V., Leplastrier, M., & Henderson-Sellers, A. (2002). Calibrating a Land Surface Model of Varying Complexity Using Multicriteria Methods and the Cabauw Dataset. *Journal of Hydrometeorology*, *3*(2), 181–194. http://doi.org/10.1175/1525-7541(2002)003<0181:CALSMO>2.0.CO;2

Xue, X., Zhang, K., Hong, Y., Gourley, J. J., Kellogg, W., Mcpherson, R. A., … Austin, B. N. (2016). New Multisite Cascading Calibration Approach for Hydrological Models: Case Study in the Red River Basin Using the VIC Model. *Journal of Hydrologic Engineering*, *21*(2). http://doi.org/10.1061/(ASCE)HE.1943-5584.0001282

Zaghloul, N. A., & Abu Kiefa, M. A. (2001). Neural network solution of inverse parameters used in the sensitivity-calibration analyses of the SWMM model simulations. *Advances in Engineering Software*, *32*(7), 587–595. http://doi.org/10.1016/S0965-9978(00)00072-7

Zhang, G., Hamlett, J. M., Reed, P., & Tang, Y. (2013). Multi-Objective Optimization of Low Impact Development Designs in an Urbanizing Watershed. *Open Journal of Optimization*, *2*, 95–108. http://doi.org/10.4236/ojop.2013.24013

Zhang, X., Srinivasan, R., & Van Liew, M. (2010). On the use of multi-algorithm, genetically adaptive multi-objective method for multi-site calibration of the SWAT model. *Hydrological Processes*, *24*(8), 955–969. http://doi.org/10.1002/hyp.7528

Zhang, X., Zhang, X., Hu, S., Liu, T., & Li, G. (2013). Runoff and sediment modeling in a peri-urban artificial landscape: Case study of Olympic Forest Park in Beijing. *Journal of Hydrology*, *485*, 126–138. http://doi.org/10.1016/j.jhydrol.2012.01.038

Manuscript

Zhou, F., Xu, Y., Chen, Y., Xu, C.-Y., Gao, Y., & Du, J. (2013). Hydrological response to urbanization at different spatio-temporal scales simulated by coupling of CLUE-S and the SWAT model in the Yangtze River Delta region. *Journal of Hydrology*, *485*, 113–125. http://doi.org/10.1016/j.jhydrol.2012.12.040