

On Comparison of Two-Level and Global Optimization Schemes for Layout Design of Storage Ponds

Wei Lu¹, Xiaosheng Qin^{1,2*} and Jianjun Yu³

¹ School of Civil and Environmental Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798

² Environmental Process Modelling Centre (EPMC), Nanyang Environment and Water Research Institute (NEWRI), Nanyang Technological University, 1 Cleantech Loop, Singapore 637141

³ Environmental Change Institute, School of Geography and the Environment, University of Oxford, Dyson Perrins Building, South Parks Road, Oxford OX1 3QY, UK.

***Corresponding author:** Dr. Qin Xiaosheng, School of Civil and Environmental Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798; Tel: +65-67905288; Fax: +65-67921650; E-mail: xsqin@ntu.edu.sg

Abstract

Optimization techniques have emerged as robust tools to aid the planning and design of urban drainage facilities in cost-effective ways. Such an effort was traditionally aided by heuristic methods (like genetic algorithm), which was **generally** time-consuming and also challenging in reaching convergence for large-scale problems with wide decision spaces. This study proposed a novel optimization method, denoted as two-level optimization (TO) scheme, for supporting rainwater storage pond design in an urban drainage system. Polynomial regression models were established as surrogate models to facilitate the solution of the optimization framework using traditional iteration algorithm. The TO scheme firstly sought the optimal layout of storage ponds on tributary sub-watersheds, and then proceeded to that of the mainstream one to yield the final solution. Through a case study, the TO scheme was compared with the traditional global optimization (GO) scheme where the physical simulation model was dynamically linked with genetic algorithm (GA) to seek the global optimal solution. The performance of two schemes under different constraint settings was analyzed. Effects of related issues such as start-point selection and mainstream design on tributary sub-watersheds were also discussed. The results showed that the proposed TO scheme is a prominent alternative to the traditional GO scheme to support urban water managers for a more science-based decision making towards storage pond implementation in large-scale practical problems.

Key words: Urban drainage system; Storage pond; Two-level optimization; Regression; Surrogate; Genetic Algorithm

1. INTRODUCTION

Urbanization has been taking place worldwide since the 1950s, resulting in rapid change of city geography and increasing runoff discharges (Wang et al., 2017; Cimorelli et al., 2015). Climate change is also believed to have resulted in escalated frequency and severity of extreme rainfalls around the world (Chen et al., 2016; Kirshen et al., 2014). Both effects are threatening the existing urban drainage systems by posing higher flooding risks and bringing severe damages to property assets, transportation systems, industrial manufacturers, and human lives (Lu et al., 2017a). As one of the efficient solutions to reduce peak flows and cope with flood impact, hydraulic structures such as storage tanks, floodplain storages and stormwater capture tanks are widely adopted (Chapman and Horner, 2010; De Martino et al., 2012; De Paola and Ranucci, 2012; De Paola et al., 2013; Tao et al., 2014; Huang et al., 2015; Li et al., 2015; Topa et al., 2015; Bellu et al., 2016). Since implementation of these facilities generally involves significant investment, their designs would need careful consideration on tradeoffs among environmental, social and economic factors (Cunha et al., 2016). Over the past years, optimization techniques have emerged as robust tools to aid the planning and design of such facilities.

From the previous studies, one of the popular methods was to dynamically link a simulation model with an optimization framework that targeted at a specific design of drainage network or layout (Cibin and Chaubey, 2015; Kalcic et al., 2014; Tao et al., 2014). Hydrological simulation models such as storm water management model (SWMM) (Rossman, 2010) and PCSWMM (Lu et al., 2017b) could be utilized for describing the relationship between control measures and hydrological responses. It is also found from previous studies that choosing efficient optimization algorithm is critical for achieving effective convergence of solutions (Cibin and

Chaubey, 2015). Evolutionary optimization algorithms such as Simulated Annealing (SA), Genetic Algorithms (GA) (Holland, 1992) and NSGA-II (Deb et al. 2002) are popular choices among many optimal designs of storage facilities (Tao et al.,2014; Oxley and Mays, 2014; Li et al., 2015; Cimorelli et al., 2015; Cunha et al., 2016; Yu et al., 2017).

Nevertheless, GA and its peer algorithms still have problems in global convergence and high computational cost if the optimization problem is of a large scale and involves significant number of decision variables (Andre et al., 2001; Lin and Lee, 2004; Kadaba and Nigard, 2014). To tackle such an issue, several multi-level optimization approaches were proposed for splitting massive and complex problems into smaller ones. For example, James and Dovit (1985) proposed a method for decomposing an optimization problem into several sub-problems, and introduced a coordination mechanism that preserved coupling between sub-problems. The method was tested for an assembled structural optimization problem where a shorter design time was achieved. Cijin and Chaubey (2015) developed a multi-level spatial optimization approach (MLSOPT) for dealing with watershed-scale spatial optimization problems by disaggregating the watershed into sub-watersheds and identifying optimum layout for each one. The study demonstrated that MLSOPT achieved robust convergence and high-efficiency in computation compared with the single-level option. Wang et al. (2017) presented a two-stage multi-objective optimization scheme for designing storage tanks. The scheme included an analytical module (stage I, to obtain a preliminary scheme) and an iteration module (stage II, to search for optimal scheme based on the preliminary scheme). Generally, few of these studies took into account the hydrological characteristics of an urban watershed in dividing sub-watersheds, and carried out optimization in consideration of tributary and mainstream interactions.

47
48 Moreover, building and calibrating a large-scale hydrological model is normally time-consuming
49 (Krebs et al., 2013), and the high computational burden of tautologically running the simulation-
50 aided optimization models may restrict the applicability of the method (Lu et al., 2017a). One
51 possible solution is to use surrogate models like artificial neural network (Sreekanth and Datta,
52 2011), moving least squares (Yu et al., 2015b) and response surface models (Seo et al., 2012), to
53 replace the physical models, instead of linking them directly into the optimization modules.
54 These surrogate models could be built up with multiple scenarios simulated from hydrological
55 models, and applied to approximate the input-output hydrological relations for achieving
56 computationally more efficient predictions (Yu et al., 2015a). **Previously, rare studies were**
57 **reported in adopting surrogate models in multi-level optimization frameworks in the field of**
58 **drainage infrastructure design.**
59
60 Therefore, this study aims to propose a new optimization scheme for supporting optimal design
61 of storage ponds by embedding surrogate models into a two-level optimization framework. The
62 level-I optimization will focus on upstream area, mainly involving tributary sub-watersheds.
63 Upon receiving the outputs from the level-I, the level-II optimization will embark on
64 downstream area comprising the mainstream sub-watershed. Quadratic regression models (i.e.
65 the surrogate models) are introduced to describe the hydraulic relationships between the layout
66 (i.e., location and sizes) of storage ponds and the overloaded flood volume in the drainage
67 system. The proposed scheme will be compared with traditional single-level optimization
68 scheme in terms of applicability, reliability and computational efficiency. A hypothetical urban
69 watershed is introduced as the study case and discussed on comparing the two schemes.

70

71 2. METHODOLOGY

72 Figure 1 shows a flowchart of the overall methodology. Two optimization schemes for layout
73 design of storage ponds are introduced, including (i) a two-level optimization (TO) scheme with
74 quadratic regression and gradient-based solution algorithm; (ii) a global optimization (GO)
75 scheme with genetic algorithm aided solution. Details of the two schemes are given below.

76 -----

77 Place Figure 1 here

78 -----

79

80 2.1 TO scheme

81 *Establishment of the physical model*

82 A physical model is normally utilized to demonstrate the hydraulic or hydrologic relations in an
83 urban drainage system. SWMM is a dynamic model for linking rainfall to runoff in the drainage
84 networks and tracking the water quality of runoff generated within each sub-watershed (Aad et
85 al., 2009), and it has been used worldwide for planning, design and analysis related to urban
86 drainage systems (US EPA, 2017). Generally, mainstream and tributaries can be identified
87 according to the hydraulic structure of a drainage system. **The criteria of division could be based**
88 **on considerations of (i) computational efficiency (i.e., no excessive number of sub-watersheds)**
89 **and (ii) local hydrological and topographical condition of the study case (i.e., runoff from each**
90 **sub-watershed independently flows into the mainstream sub-watershed). Similar criteria of sub-**
91 **watershed definition could be applied to many other urban areas; however, from practical point**
92 **of view, the decision-makers or stake holders may also get involved in the discussion of**

division of urban watershed for a better consideration of local conditions. In this study, sub-watersheds contributing to tributaries will be grouped into tributary sub-watersheds, whereas those contributing to the mainstream will be categorized as mainstream sub-watersheds. In TO scheme, the tributary sub-watersheds are the target areas for level-I optimization, and the mainstream sub-watershed will be tackled at level-II.

Establishment and validation of surrogate models

Polynomial surrogate models are adopted to explore the relationships between several explanatory variables (e.g., locations and size of interested hydraulic units) and one or more response variables (e.g., performance and construction cost of these hydraulic units). Polynomial models are essential intermediaries for replacing the physical models during the optimization process in TO scheme. The main reason lies on the avoidance of using simulation model under an optimization framework, which may bring along technical difficulty and high computational burden. Also, comparing with other surrogate models like artificial neural network (ANN), polynomial models have explicit mathematical expressions which are more favorable for optimization model expression and solution. Taking quadratic models as an example, the formulation can be written as (Lucas, 1976):

$$y = b_0 + \sum_{q=1}^s b_q x_q + \sum_{q=1}^s b_{qq} x_q^2 + \sum_{1 \leq q < r \leq s} b_{qr} x_q x_r + e \quad (1)$$

where y is the response or dependent variable; x_q ($q = 1, 2, \dots, s$) are the independent experimental variables; b_0 , b_q , b_{qq} and b_{qr} ($q = 1, 2, \dots, s$; $r = 1, 2, \dots, s$) are parameters to be estimated using the data from experiment, and e is an error term; q and r are indexes for x and b , and s denotes the total number of variables x .

For model establishment, all the junctions over the study area are assumed potential locations to install storage ponds. Random samples of storage pond layout (i.e. locations and surface area of ponds) can be generated by using Latin Hypercube Sampling (LHS) (McKay et al., 1979). For each sample of layout, SWMM model will run under a specific design rainfall to compute the total surcharged runoff (i.e., flood volume) from all nodes (i.e., junctions and ponds) within sub-watersheds. Then a polynomial surrogate model is introduced to mimic the relationship between the layout of storage ponds and the corresponding flood volume. For model validation, the samples can be split into training and testing groups. The testing group does not contribute to the model construction, but only be used to test the predicted values from the training group.

Formulation of optimization models

In this study, the total number of implemented storage ponds within each sub-watershed is firstly determined. The criteria could be set by the decision makers considering budget availability, land susceptibility, or flood control balance. Here we use the evenness of flood distribution over sub-watersheds under the design rainfall. After the number of storage ponds are determined, the optimal layout (including location and size) of ponds within each sub-watershed can be identified through the following two-level optimization process. In level-I optimization, the model will generate the optimal layout of storage ponds in each tributary sub-watershed. By controlling the total flood volume in the tributary area, it lays a foundation for the level-II optimization, which focuses on the mainstream area. Let i denote the i^{th} sub-watershed (S_i), the corresponding optimization model can be formulated as (Tao et al. 2014):

$$MinTC_i = \sum_{k=1}^{j_i} [a \cdot (A_k \cdot H_k)^e + b \cdot A_k + g(A_k)] \quad (2a)$$

subject to

$$j_i \leq p_i \quad (2b)$$

$$F_i = f_i(A_k) \leq c_i, \forall k$$

(2c)

$$a, b, g(\cdot) = \begin{cases} \text{const}, A_k \geq A_{\min}, \forall k \\ 0, A_k < A_{\min} \end{cases}$$

(2d)

$$A_{\min} \leq A_k \leq A_{\max}$$

(2e)

where TC_i stands for the total cost of building the storage ponds in S_i ; j_i denotes the number of storage ponds in S_i ; p_i is the maximum number of storage ponds in S_i ; k is the index of j_i ($k = 1, 2, \dots, j_i$); A_k denotes the surface area of the k^{th} storage pond, varying within the range of (A_{\min}, A_{\max}); H_k denotes the depth of the k^{th} storage pond; a and e are coefficients pertaining to the dimension of storage pond and b is a coefficient reflecting the unit land cost in the area; $g(\cdot)$ is the launching cost of each storage pond (Li et al., 2015; Huang et al., 2015), which can also be considered as an additional cost for initializing the storage pond construction. Eq.2d denotes that a , b and $g(\cdot)$ will automatically turn to 0 for leaving out the construction cost of a specific storage pond when the iterated input of A_k is less than A_{\min} ; otherwise, they will be set constant, regardless of the exact size of the storage pond. F_i denotes the total flood volume from all junctions (or ponds) in S_i , which is equal to the excessive runoff that beyond the capacity of junctions and storage ponds in S_i . F_i is calculated based on the corresponding variables A_k of the regression model (here denoted as function f_i) for S_i ; c_i is defined as the total flood volume threshold from all junctions (or ponds) in S_i .

The mainstream sub-watershed is distinguished from the rest sub-watersheds located in tributaries, as the hydraulic-hydrological performance of the mainstream drainage is significantly influenced by the runoff from tributaries. Thus, level-II optimization will adopt the optimization results from level-I (i.e., layout of storage ponds in all the tributary sub-watersheds) as its inputs. The corresponding regression model will be established for the mainstream sub-watershed, keeping all tributary watersheds fixed at A_k ($k = 1, 2, \dots, j_i; i = 1, 2, \dots$) from level-I. The objective of level-II optimization is the optimal layout of storage pond in mainstream sub-watershed. The optimization formulation at level-II is formulated as:

$$MinTC_M = \sum_{m=1}^{j_M} [a \cdot (A_m \cdot H_m)^e + b \cdot A_m + g(A_m)] \quad (3a)$$

subject to

$$j_M \leq p_M$$

(3b)

$$F_M = f_M(A_m) \leq c_M, \forall m$$

(3c)

$$a, b, g(\cdot) = \begin{cases} const, A_m \geq A_{\min} \\ 0, A_m < A_{\min} \end{cases}, \forall m$$

(3d)

$$A_{\min} \leq A_m \leq A_{\max}$$

(3e)

$$A_k = const, \forall k, \forall i$$

(3f)

where TC_M stands for the total cost of building the storage ponds in mainstream; j_M denotes the

number of storage ponds in mainstream; p_M is the maximum number of storage ponds in mainstream; m is the index of j_M ($m = 1, 2, \dots, j_M$); A_m denotes the surface area of the m^{th} storage pond; H_m denotes the depth of the m^{th} storage pond; a , e , b and $g(\cdot)$ are similar coefficients from level-I. F_M denotes the total flood volume from all junctions (or ponds) in mainstream, which is calculated based on the corresponding regression model (here denoted as function f_M); c_M is defined as the total flood volume threshold from all junctions (or ponds) in mainstream. In addition, Eq. 3f emphasizes that the areas and locations of storage ponds outside the mainstream sub-watershed are kept unchanged in the level-II optimization process.

2.2 GO scheme

The GO scheme searches optimal solutions at a global scale. Compared with TO scheme, the GO one does not need division of sub-watersheds, and the number of storage ponds is unconstrained in the study area. GA is adopted for searching solutions that could minimize the construction cost and satisfy local constraints (e.g., total flood volume, area limit of storage ponds, etc.). As the problem involves much more decision variables when compared with TO scheme, the sampling space in GA has been largely expanded. The optimization formulation is shown as:

$$MinTC_N = \sum_{n=1}^N \left[a \cdot (A_n \cdot H_n)^e + b \cdot A_n + g(A_n) \right] \quad (4a)$$

subject to

$$F_N = s(A_n) \leq c_N, \forall n$$

(4b)

$$a, b, g(\cdot) = \begin{cases} const, A_n \geq A_{\min}, \forall n \\ 0, A_n < A_{\min} \end{cases}$$

(4c)

$$A_{\min} \leq A_n \leq A_{\max}$$

(4d)

where TC_N and N stands for the total cost of building the storage ponds and total number of storage ponds in the entire study area, respectively; n is the index of N ; F_N denotes the total flood volume from all the overloaded junctions and ponds, which is calculated based on the output of SWMM model (s) under different scenarios of A_n ; c_N is defined as the total flood volume threshold in the study area. Definitions and principles of the rest variables and constants remain the same as in the TO scheme.

3. CASE STUDY

3.1 Case Background and Model Setup

Description of the study site

A hypothetical urban catchment adapted from a real tropical urban area is selected for demonstrating the proposed design framework. Figure 2 shows the general map and layout of the existing drainage network of the study site. The site is characterized by urban commercial and residential areas under a tropical climate, where the average rainy days is about 167 days of the year and the maximum hourly rainfalls could reach up to 147 mm (Meteorological Service Singapore, 2017).

Place Figure 2 here

Design rainfall

Rainfall are designed based on intensity-duration-frequency (IDF) curves and Huff distribution. IDF curves reveal the characteristics of the rainfall extremes, which are established based on annual maximum rainfalls with various durations (Arnbjerg-Nielsen, 2012). Huff storm distribution includes four types of rainfall patterns, with peak intensity occurring in different quarters of the duration in each pattern (Chow et al., 1988). This study assumes a design rainfall (with total depth at 107.5 mm) with 1-hour duration, Type-II Huff distribution and 25-years return period (PUB, 2011). Then the design rainfall is established to set up the model and generate flood scenarios. Figure 3 shows the corresponding hyetograph with an interval of 5 minutes.

Place Figure 3 here

SWMM model setup and division of sub-watersheds

An urban drainage model based on SWMM is established for the study case with an area of 25.14 km². In this study, the dynamic routing model is adopted in SWMM; it solves the complete one-dimensional Saint-Venant flow equations and therefore produces the most theoretically accurate results. Dynamic wave routing can address issues related to channel storage, backwater, flow reversal and pressurized flow, and it is suitable for systems subjected to significant backwater effects due to downstream flow restrictions (e.g. the effect of downstream pond design on flooding of upstream areas) (Rossman, 2010). Originally, the model consists of 44

subcatchments, 81 junctions, 81 conduits and 1 outfall. These subcatchments are basic hydrological units in SWMM, and they are grouped into several sub-watersheds. Depending on the criteria of division, the drainage system will be aggregated into 1 mainstream sub-watershed and 3-6 tributary sub-watersheds. Based on this structure, 34 subcatchments are considered as in tributary areas, and the remaining 10 subcatchments contribute their runoff directly to the mainstream. The 34 subcatchments in tributary areas are disaggregated into three sub-watersheds ($S1$, $S2$ and $S3$), which have 11, 12 and 11 subcatchments (20, 21 and 20 junctions), respectively. The rest subcatchments form the mainstream sub-watershed $S4$, which has 10 subcatchment (20 junctions). **Figure 4 shows the result of sub-watershed division based on the criteria mentioned in the methodology section. For convenience in identifying pond locations, the potential storage ponds are indexed (denoted as Pond Index) based on the junction indexes as shown in Figure 2.**

Place Figure 4 here

Setup of optimization framework

In TO scheme, the Constrained Nonlinear Minimization Solver (CNMS) (Coleman et al., 1999; Matlab, 2015) is adopted as the iterative algorithm for searching the optimal layout in TO scheme, and the CNMS is embedded with the regression model as one of model constraints (see Eq.2c and Eq.3c). In GO scheme, the SWMM model is directly linked with MATLAB platform. Therefore, the execution results of SWMM model can be automatically extracted into MATLAB for the iteration process of GA. Also, a penalty-based algorithm is adopted in GA process for model solution.

275

276 To test the TO scheme, the following parameter settings are adopted (as Scenario I).

277 The coefficients of a , b and e are set as: $a = \$500$, $b = \$100$, and $e = 0.69$. As the launching cost

278 for each storage pond, the value of $g(\cdot)$ is assumed to be \$5000 if the pond is actually

279 implemented. The minimum and maximum areas for each storage pond are set as $A_{\min} = 500$ and

280 $A_{\max} = 10000 \text{ m}^2$, respectively. The depth of each storage pond is set to be equal to its original

281 junction for ease of construction. If the iterated input of A_k and A_m in TO is less than 500 m^2 , it

282 will not be implemented and there is no construction cost calculated for that pond. For the TO

283 scheme, the total flood volume threshold for all nodes in each sub-watershed (c_i and c_M) is set to

284 $20,000 \text{ m}^3$. For consistency, the c_N is set at $80,000 \text{ m}^3$ for the four sub-watersheds in GO scheme.

285

286 For the TO scheme under Scenario I, the maximum number of storage ponds is set to 42. In

287 practical applications, this number could be determined based on financial capability, available

288 land area for pond installation, and policy requirement. Then, the number of implemented

289 storage ponds p_i within sub-watershed S_i can be determined either by decision-makers'

290 preference or local condition. In this study, we use design rainfall to test the flooding condition

291 for all sub-watersheds and then apply the percentage of flood volume at each sub-watershed to

292 estimate the maximum number of ponds that should be deployed into each sub-watershed. The

293 detailed results are shown in Table 1 (p_2 is set as 10, because S_2 originally has more junctions

294 than other sub-watersheds).

295

296 -----

297 Place Table 1 here

298 -----

Setup of surrogate models

The quadratic regression models are adopted and established in TO scheme. Eq.1 gives the formulation of the models. The sampling number for each sub-watershed is customized to 1000. To ensure the accuracy of quadratic models, 90% of the samples (i.e., 900 sample sets) and their outputs are used for model establishment, while the rest 10% are used for verification. The testing results for three tributary sub-watersheds *S1*, *S2* and *S3* are shown in Figure 5 (sub-figures 5a, 5b and 5c correspond to *S1*, *S2* and *S3*, respectively). In Figure 5, each scatter represents a sample, plotted by using the predicted value of total flood volume from quadratic model on the x axis and simulated results from testing group on the y axis. The straight line in each sub-figure represents the line ' $y=x$ '. From Figure 5, it is clear that most of the regressions achieved a decent fitness with R^2 above 0.95, except for sub-watershed *S2* which has a R^2 of 0.89. This is mainly because of the over-estimation of the flood condition in a small portion of samples by the regression model. After verifying the accuracy of quadratic models, the optimal layout of sub-watersheds *S1* to *S3* is identified by running level-I of TO optimization (details are provided in the Result Analysis and Discussion). Afterwards, the quadratic regression model is built for the mainstream sub-watershed *S4* (see Figure 5d) and the R^2 level of regression performance is 0.977. The detailed coefficients of quadratic models can be referred to the Supplementary Materials.

Place Figure 5 here

3.2 Result Analysis and Discussion

Results of TO and GO scheme

To investigate the two optimization schemes under different settings of constraints, two more scenarios, denoted as Scenarios II and III, are added in addition to Scenario I. Scenario II assumes a lower maximum number of storage ponds (i.e. 38), and a higher threshold of flood volume for each sub-watershed (i.e. 30,000 m³). The maximum storage pond area is limited to 8,000 m². Scenario III allows more storage ponds to be constructed (maximum number set to 50) but their maximum storage area is reduced to 6,000 m². The threshold of flood volume for this scenario is set to 15,000 m³. There are no strict rules to set these parameters, as the main purpose is to test the consistency of TO performance under different conditions. Since the maximum number of ponds has changed, the allocation patterns are also adjusted. Under Scenario III, the number of ponds is still allocated based on the same percentage of flooding used under Scenario I. The allocation pattern for Scenario II has some adjustment assuming decision-makers prefer to add more storage ponds in the mainstream sub-watershed. Scenario IA is added to test the effect of changing the maximum number of ponds for the TO scheme under Scenario I. For a fair comparison between TO and GO schemes, Scenario IA adopts the same settings as Scenario I, except that the maximum number of storage ponds is fixed at 71 which is the same as the optimal number of ponds obtained from GO scheme under Scenario I. Scenario IB is another scenario with different sub-watershed division scheme and it will be elaborated later.

Table 2 shows the optimal layouts of storage ponds for all sub-watersheds and the overall performance of TO scheme under Scenario I. Generally, for the TO scheme, the solution with a layout of 42 storage ponds could be obtained, with an overall cost around \$42.41 million and

overall flood volume at 78,418 m³. For the GO scheme, after a 50-generation iteration (with a population size of 200 for each generation) in the GA-SWMM interaction process, the solution with a layout of 71 storage ponds among 81 potential locations is found (details can be referred to Supplementary Material); the corresponding overall cost is \$48.09 million and the overall flood volume is 79,934 m³. The cost of GO scheme is about 13.4% higher than that of TO scheme, while the simulated flood volume under both schemes are both close to the sum of thresholds for all sub-watersheds (i.e. 80,000 m³).

Place Tables 2 here

Table 3 summarizes the results of the two optimization schemes under Scenarios I, IA, II and III. Generally, both schemes would lead to similar flood levels, and the TO scheme would achieve a lower cost under Scenarios I, II and III (reduction of cost varied from 8.7% to 11.8%) due to reduced number of storage ponds while satisfying local criteria of flood reduction. Whereas, the Scenario IA for TO scheme achieved a final layout of 49 ponds and the overall cost would be somewhat higher than that with TO scheme in Scenario I due to the heavy investment in mainstream sub-watershed (which has a layout of 17 storage ponds at a cost of \$23.21 million based on its level-II optimization). This means that, in Scenario IA, the ponds in the mainstream area are suggested to accommodate higher runoff and demand a higher cost of construction. This also explains the increase of total cost for this scenario compared with Scenario I. The cost from TO scheme under Scenario IA is slightly higher than that from GO scheme, possibly due to the fact that GO's searching is more global, and could obtain solutions with lower average area of

ponds (i.e. 3282 m²) than TO scheme under Scenario IA (i.e. 5905 m²). The above results demonstrate that, in terms of optimization performance, TO scheme could lead to solutions with superior or at least comparable results than GO scheme. For detailed optimization results, readers can refer to Supplementary Material.

In terms of efficiency of model runs, the total run times for the TO and GO schemes are approximately 3.3 and 6.5 hours, based on a desktop computer (Xeon™ W3550 CPU @3.07 GHz, 12G RAM). The run time for the TO scheme is estimated based on the following considerations: (1) each SWMM run is about 3s and the time required to run all 1000 samples for 4 sub-watersheds is about $3s \times 4 \times 1000 / 3600 = 3.3$ hours; (2) the optimization process generally takes less than 1 minutes to complete. For GO scheme, the simulation model is called within the optimization framework multiple times depending on the population size for each generation and the total number of generations. It is worth mentioning that the workload of the sampling process to set up the regression surrogate model could be easily alleviated by using parallel computing (e.g. multiple computers, multiple CPU cores, or cloud computing). It is technically more challenging to reduce computational time for the GO scheme.

Place Table 3 here

Impact of selecting different start points on optimization results

Based on model test, the start points (SP) (i.e., initial value of variables, which refer to the areas of ponds) was found to have a noticeable effect on the optimization results. Figure 6 shows the

testing results for 5 runs with different SPs for Scenario III. Figures 6a shows the solutions of storage pond layout and their corresponding objective function values (i.e. construction costs) of 5 runs for sub-watersheds *S1* to *S4*. Each sub-watershed adopts 5 random sets of SPs generated by LHS. It is found that the majority of storage ponds have stable areas among five runs. From the upper right of Figure 6a, the objective function values (i.e. construction cost) of 5 runs for 4 sub-watersheds show somewhat fluctuations, especially for sub-watersheds *S3* under Run 4 (~12.1% lower than average) and *S2* under Run 5 (~14.6% higher than average), respectively.

The performance of TO scheme is generally stable towards different SPs. For TO scheme, the solutions of optimization models show a strong dependence on SP; some sets of SPs cannot even support a feasible solution (i.e., they cannot satisfy the constraints before reaching the stopping criteria) and reselection of different SPs may be necessary. Such an infeasibility issue is caused by the inherent error associated with the quadratic regression model and the limitation of solution algorithm (i.e. CNMS) that is affected by the selection of SPs and local optima.

For GO scheme, Figure 6b shows that the overall cost (i.e. GA fitness values) would range from 49.8 to 55.6 million, which is comparable to the 48.1 million achieved by TO scheme under the same scenario. Moreover, Figure 6b demonstrates that the optimized layout from GO is highly diversified among various runs; while the final solutions in the TO scheme are much more stable (as shown in Figure 6a). Apparently, when the scale of the problem is large (like the studied case), GA may not perform satisfactorily in seeking global optimal solutions.

Place Figure 6 here

Effect of mainstream pond design on sub-watershed flooding

During the sampling stage for building up the quadratic regression model, the physical model (SWMM) would have to run based on each sample input. The physical model takes the entire watershed into consideration, including the interactions between the mainstream and tributary sub-watersheds. It is found that the layout of storage ponds in the mainstream sub-watershed would also cause flood conditions in the tributary sub-watersheds, mainly through the connecting junctions. To control or alleviate such an effect, if desired, we could consider modify the level-II optimization model structure. For example, we could add additional constraints of limiting the flood volume at the junctions of concern (e.g. sensitive areas like residential buildings, main roads, and commercial centers) within a certain range. For demonstration purpose in this study, if the reflux of excessive runoff from the mainstream junctions to the adjacent tributary junctions (defined as the “backflow effect”) is of the major concern to decision makers, three extra constraints (denoted as Extra-con) could be added into the level-II optimization framework to control the gap between the flood volume of the adjacent junctions in the mainstream model and that in the tributary model within an allowable level (i.e. 500 m^3). Examples of adjacent tributary junctions are J09, J21, and J63, as shown in Figure 2. Specifically, from ordinary condition of node flooding under Scenario I, the flood volume for J09, J21 and J63 is 6348, 0 and 2400 m^3 , respectively. Thus, the extra constraints limit the maximum flood volume of J09, J21 and J63 below 6848, 500 and 2900 m^3 , respectively. The way to limit the flood volume of these junctions is to add additional constraints to the optimization model (i.e. Equation 3). These constraints are established using regression models which reflect the relationship between flood volumes of adjacent tributary junctions and area of mainstream ponds.

Figure 7 shows the 1000 sample scenarios of total flood volumes at three tributary sub-watersheds (i.e. *S1*, *S2* and *S3*) and the mainstream sub-watershed *S4*, where the *x* axis denotes the flood volume in the mainstream sub-watershed in an ascending order. It is indicated that the flood volumes from the sampled layouts of storage ponds would vary in both the mainstream sub-watershed and three tributary sub-watersheds. The correlation coefficient of flood volume between the mainstream and the tributary sub-watersheds is found to be 0.765, indicating a generally positive correlation and possible backflow effect from mainstream to the tributary sub-watersheds. Figure 8 shows the flooded volume at various junctions (with a threshold of 500 m³) from the two types of models under the TO scheme, which consist of the original model (Scenario I) and the model with Extra-con on controlling backflow effect. It can be found that the junctions of J38 and J63 are heavily flooded in the original model; with the Extra-con, the flooding in J38 and J63 would disappear and other junctions would have somewhat minor fluctuations of flood volume. The details of pond layout can be referred to the Supplementary Materials. The results demonstrated that, for TO scheme, more specific constrains at sensitive junctions could be adopted to mitigate the effect of mainstream pond design on flooding conditions of the sub-watersheds. In terms of the overall flood magnitude and cost, the model with Extra-con would decrease the total flood volume from 78,418 to 68,977 m³ and total cost from \$42.41 to 41.97 million, in comparison to those from the original model under Scenario I.

Place Figures 7 and 8 here

Effect of different division of sub-watersheds on optimized results

In this study, the urban watershed is divided into sub-watersheds according to computational efficiency and local hydrological and topographical condition. In order to examine the effect of choosing different division schemes, we have tested another way of dividing the sub-watersheds. Figure 9 shows the new division over the study domain, where there are 11, 8, 16 and 9 subcatchments for *S1*, *S2*, *S3* and *S4*, respectively. Extra scenario simulations, denoted as Scenario IB, are carried out based on the new division scheme and the relevant results are also summarized in Table 3. The Scenario IB adopts similar settings as Scenario I (e.g. maximum number of total storage ponds and maximum area for each ponds), and also shows similar outputs in terms of overall cost and flooding. This implies that the effect of using different division of sub-watersheds on optimized results is insignificant. The detailed layout of storage pond is summarized in the Supplementary Materials.

Place Figures 9 here

Effect of different GA options on optimized results of GO scheme

In GO scheme, besides start points, options of GA parameters (e.g. elite count, crossover fraction and mutation function) may also affect the algorithm's efficiency for searching global optimal solutions and satisfying local constraints. In this study, the GA adopts default values for most of the parameters defined in MATLAB (MATLAB, 2015). For example, the rates for elite count and crossover fraction are set to 10 (twentieth of the Population Size) and 0.8, respectively and

the “adaptive feasible” is chosen as the mutation function, which is the recommended function for constraint-dependent optimization (MATLAB, 2015). To further investigate the sensitivity of GO scheme towards options of GA, three extra GO scheme runs (with different settings on crossover fraction and elite count rate) under Scenario I are carried out, and the results are summarized in Table 4. It is shown that the overall cost would range from \$46.6 to 49.8 million. Runs No. 2 and No. 4 achieve lower overall costs than that from original GA setting (i.e. No. 3), and their total flood volume is relatively closer to the threshold (i.e. 80000 m³). In general, the reliability and convergence of GO scheme under various GA options are acceptable. More detailed information of iterations and final solutions can be referred to Supplementary Materials.

Place Table 4 here

Further discussion

This study attempts to integrate various techniques into a TO scheme for storage pond planning and design. These techniques involve urban drainage modeling, sub-watersheds division, surrogate model construction and CNMS iterative algorithm. The performance of the proposed TO scheme is compared with traditional GO scheme. It is revealed that the TO scheme has the following advantages:

(1) Applicability

Applicability indicates the scope of the proposed scheme to be used in applications (Yu et al, 2015a). The TO scheme could be applied to optimize a large-scale catchment which involves a

significant number of decision variables. With the aid of the two-level (or even multi-level) optimization structure, the number of variables in each level would significantly decrease; furthermore, within each level, the regression models are combined with an iterative algorithm for more efficient searching of the optimized results. As for GO scheme, it is proven to be difficult or time-consuming in reaching a global optimal solution if the problem involves a large number of decision variables. Also, under various modelling scenarios in this study, it is found that the TO scheme managed to converge to better or comparable results in terms of overall cost compared with GO scheme (see Table 3).

(2) Reliability

Reliability is an indicator showing logical coherence of the scheme and the feasibility of the optimized results under uncertainty. In TO scheme, the regression models are verified (see Figure 5) for carrying on the optimization framework. Uncertainty issues, including setup of start points and mainstream effect, and their impact on optimized results were also investigated. Figure 6 proves a more stable performance of TO scheme over GO scheme.

(3) Computational efficiency

In this study, TO scheme showed a relatively more efficient computational effort than GO scheme. However, TO is deemed potentially more advantages in undertaking a parallel computing using multiple cores or computers, as its main time-consuming part is the running of samples to build up the surrogate models and it needs almost negligible time to run the optimization model. GO scheme requires both physical model and optimization run at the same time, which is more challenging to be parallelized.

535

536 As the first attempt in proposing TO scheme, we also encountered some limitations. Firstly, the

537 scale control of each tributary sub-watershed is an important issue. For TO scheme in this study,

538 the decision variables (i.e., number of junctions) for each sub-watershed is controlled around 20,

539 ensuring sufficient accuracy of regression models (e.g., in Figure 5, all the regression models

540 achieved a decent fitness with R^2 around 0.9 or above). However, when the number of decision

541 variables increases significantly (e.g. each sub-watershed has over 100 junctions), it is likely that

542 the regression models could not achieve an acceptable level of fitness. By simply increasing the

543 number of sub-watersheds may bring extra run time and may not be applicable for optimizing

544 large-scale watersheds. Multi-level optimization may be a good solution, but it needs further

545 investigations. Secondly, as indicated in **Scenarios I and IA**, the performance (i.e. overall cost) of

546 TO scheme is subjected to the user-defined settings (e.g. patterns and maximum number of

547 storage ponds), which may bring along uncertainty issues for the optimization results. Users are

548 recommended to set a relatively lower limit of the number of ponds for saving cost, if the

549 constraints of flood volume can be satisfied. **Moreover, TO scheme could offer a comparable**

550 **(may not be necessarily more optimal) solution to GO scheme, with respect to minimization of**

551 **total system cost and satisfaction of system constraints. But we should bear in mind that the main**

552 **advantage of TO scheme lies in the saving of much computational efforts when the physical**

553 **model is large and/or the dynamic linkage of GA and physical model becomes impractically**

554 **complex. Thirdly, in practical applications, if the number of storage ponds exceeds the capacity**

555 **of construction at one time, sequence of construction may need to be considered. The sensitivity**

556 **of flood from partial deployment of storage ponds at different stages could be analyzed in order**

557 **to prioritize the construction efforts for a better flood control effect. For example, if there is no**

storage ponds for the case in this study, the flood volume for the entire system could reach 693.2 thousand m^3 . Assume we construct ponds for one sub-watershed at a time based on the model solution for Scenario I, the flood volume would drop to 537.4, 572.2, 537.7 and 509.8 thousand m^3 with individual deployment of ponds in *S1*, *S2*, *S3*, and *S4*, respectively. The implementation plan could also be guided by the model with reduced total number of ponds.

4. CONCLUSIONS

A two-level optimization (TO) scheme was proposed for supporting optimal design of storage ponds in urban drainage systems. The scheme focused on optimization of storage pond layout in tributary sub-watersheds (level-I optimization) first, and then proceeded to optimization for the mainstream sub-watershed (level-II optimization). The performances of the proposed TO scheme and the traditional global optimization (GO) scheme were tested and compared in terms of applicability, reliability and computational efficiency. A hypothetical case adapted from a tropical urban community was selected for demonstrating the applicability of methodology. The results showed that both schemes could achieve convergences under different settings of optimization constraints. Compared with the GO scheme, the TO one could achieve a comparable or lower system cost, save almost half of the run time and obtain a more stable solution with different setups of start points. It is a good alternative of GO scheme for tackling large-scale problems, although there still much room for improvement in the future. Also, for both schemes, the effect of backflow from mainstream to tributary areas and their propagation to optimization results need to be further investigated.

ACKNOWLEDGEMENT

The research work was supported by Singapore's Ministry of Education (MOE) AcRF Tier 1 Project (Ref No. RG170/16; WBS No.: 4011766.030).

REFERENCES

- Aad, M. P. A., Suidan, M. T., Shuster, W. D., 2009. Modeling techniques of best management practices: Rain barrels and rain gardens using EPA SWMM-5. *J. Hydrol. Eng.*, 10.1061/(ASCE)HE.1943-5584.0000136.
- Andre, J., Siarry, P., Dognon, T., 2001. An improvement of the standard genetic algorithm fighting premature convergence in continuous optimization. *Adv. Eng. Software*, 32(1), 49-60.
- Arnbjerg-Nielsen, K., 2012. Quantification of climate change effects on extreme precipitation used for high resolution hydrologic design. *Urban Water J.*, 9(2), 57-65.
- Bellu, A., Fernandes, L. F. S., Cortes, R. M., Pacheco, F. A., 2016. A framework model for the dimensioning and allocation of a detention basin system: The case of a flood-prone mountainous watershed. *J. Hydrol.*, 533, 567-580.
- Chapman, C., Horner, R. R., 2010. Performance assessment of a street-drainage bioretention system. *Water Environ. Res.*, 82(2), 109-119.
- Chen, Y., Samuelson, H. W., Tong, Z., 2016. Integrated design workflow and a new tool for urban rainwater management. *J. Environ. Manag.*, 180, 45-51.
- Chow, V.T., Maidment, D. R., Mays, L.W., 1988. *Applied hydrology*. McGraw-Hill Book Company, Maidenherd.

602 Cibin, R., & Chaubey, I., 2015. A computationally efficient approach for watershed scale spatial
603 optimization. *Environ. Modell. Software*, 66, 1-11.

604 Cimorelli, L., Morlando, F., Cozzolino, L., Covelli, C., Della Morte, R., Pianese, D., 2015.
605 Optimal positioning and sizing of detention tanks within urban drainage networks. *J. Irrig.*
606 *Drain. Eng.*, 142(1), 04015028.

607 Coleman, T., Branch, M. A., Grace, A., 1999. Optimization toolbox. For Use with MATLAB.
608 User's Guide for MATLAB 5, Version 2, Release II.

609 Cunha, M. C., Zeferino, J. A., Simões, N. E., Saldarriaga, J. G., 2016. Optimal location and
610 sizing of storage units in a drainage system. *Environ. Modell. Software*, 83, 155-166.

611 Deb, K., Pratap, A., Agarwal, S., Meyarivan, T. A. M. T., 2002. A fast and elitist multiobjective
612 genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.*, 6(2), 182-197.

613 De Martino, G., De Paola, F., Fontana, N., Marini, G., Ranucci, A., 2012. Experimental
614 assessment of level pool routing in preliminary design of floodplain storage. *Sci. Total*
615 *Environ.*, 416, 142-147.

616 De Paola, F., Ranucci, A., 2012. Analysis of spatial variability for stormwater capture tank
617 assessment. *Irrig. Drain.*, 61(5), 682-690.

618 De Paola, F., Ranucci, A., Feo, A., 2013. Antecedent moisture condition (SCS) frequency
619 assessment: a case study in southern Italy. *Irrig. Drain.*, 62(S2), 61-71.

620 Holland, J. H., 1992. Adaptation in natural and artificial systems: an introductory analysis with
621 applications to biology, control, and artificial intelligence. MIT press.

622 Huang, C. L., Hsu, N. S., Wei, C. C., Luo, W. J., 2015. Optimal spatial design of capacity and
623 quantity of rainwater harvesting systems for urban flood mitigation. *Water*, Vol. 7, No. 9,
624 pp. 5173-5202.

James, B., Dovit, A., 1985. Structural optimization by multilevel decomposition.

Kadaba, N., Nygard, K. E., 2014. Improving the performance of genetic algorithms in automated discovery of parameters. In *Machine Learning: Proceedings of the Seventh International Conference* (pp. 140-148).

Kalcic, M., Prokopy, L., Frankenberger, J., Chaubey, I., 2014. An in-depth examination of farmers' perceptions of targeting conservation practices. *Environ. Manage.*, 54(4), 795-813.

Kirshen, P., Caputo, L., Vogel, R. M., Mathisen, P., Rosner, A., Renaud, T., 2014. Adapting urban infrastructure to climate change: A drainage case study. *J. Water Resour. Plann. Manage.*, 141(4), 04014064.

Krebs, G., Kokkonen, T., Valtanen, M., Koivusalo, H., Setälä, H., 2013. A high resolution application of a stormwater management model (SWMM) using genetic parameter optimization. *Urban Water J.*, 10(6), 394-410.

Li, F., Duan, H.F., Yan, H.X., Tao T., 2015. Multi-Objective Optimal Design of Detention Tanks in the Urban Stormwater Drainage System: Framework Development and Case Study. *Water Resour. Manage.*, 29(7), 2125 – 2137. <https://doi.org/10.1007/s11269-015-0931-0>

Lin, C. C., Lee, Y. J., 2004. Stacking sequence optimization of laminated composite structures using genetic algorithm with local improvement. *Compos. Struct.*, 63(3), 339-345.

Lu, W., Qin, X.S., Yu, J.J., 2017a. Emulator-aided optimization of detention tanks for flood reduction, In *Proceedings of the 37th IAHR World Congress, Kuala Lumpur, Malaysia*, 13-18 August, 2017, pp.4165-4170.

Lu, W., Qin, X., Jun, C., 2017b. (In press) A Parsimonious Framework of Evaluating WSUD Features in Urban Flood Mitigation. *J. Environ. Inf.*, 0. Retrieved from <http://www.jeionline.org/index.php?journal=mys&page=article&op=view&path%5B%5D=>

201700373

Lucas, J. M., 1976. Which response surface design is best: a performance comparison of several types of quadratic response surface designs in symmetric regions. *Technometrics*, 18(4), 411-417.

MATLAB, 2015. Global Optimization Toolbox. Use's Guide (R2015b).

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.

Meteorological Service Singapore., 2017. Climate of Singapore.

<http://www.weather.gov.sg/climate-climate-of-singapore/> (accessed Nov. 19, 2017)

Oxley, R. L., Mays, L. W., 2014. Optimization–simulation model for detention basin system design. *Water Resour. Manage.*, 28(4), 1157-1171.

Public Utility Board (PUB). 2011. Code of Practice on Surface Water Drainage. (Sixth Edition) https://www.pub.gov.sg/Documents/COP_Final.pdf (accessed Nov. 19, 2017)

Rossman, L. A., 2010. Storm Water Management Model–User manual Version 5.0. Revised July 2010. US Environmental Protection Agency (US EPA), Cincinnati, USA.

Seo, J., Dueñas-Osorio, L., Craig, J. I., Goodno, B. J., 2012. Metamodel-based regional vulnerability estimate of irregular steel moment-frame structures subjected to earthquake events. *Eng. Struct.*, 45, 585-597.

Sreekanth, J., Datta, B., 2011. Comparative evaluation of genetic programming and neural network as potential surrogate models for coastal aquifer management. *Water Resour. Manage.*, 25(13), 3201-3218.

Tao, T., Wang, J., Xin, K., Li, S., 2014. Multi-objective optimal layout of distributed storm-

671 water detention. *Int. J. Environ. Sci. Technol.*, 11(5), 1473-1480.

672 Topa, M. E., Giugni, M., De Paola, F., 2014. Off-stream floodplain storage: numerical modeling
673 and experimental analysis. *J. Irrig. Drain. Eng.*, 141(1), 04014040.

674 US Environmental Protection Agency (US EPA), 2017. Storm Water Management Model
675 (SWMM) - Version 5.1.010 with Low Impact Development (LID) Controls.

676 <https://www.epa.gov/water-research/storm-water-management-model-swmm> (accessed Nov.
677 19, 2017)

678 Wang, M., Sun, Y., Sweetapple, C., 2017. Optimization of storage tank locations in an urban
679 stormwater drainage system using a two-stage approach. *J. Environ. Manage.*, 204, 31-38.

680 Yu, J. J., Qin, X. S., Larsen, O., 2015a. Applying ANN emulators in uncertainty assessment of
681 flood inundation modelling: a comparison of two surrogate schemes. *Hydrol. Sci. J.*, 60(12),
682 2117-2131.

683 Yu, J. J., Qin, X. S., Larsen, O., 2015b. Uncertainty analysis of flood inundation modelling using
684 GLUE with surrogate models in stochastic sampling. *Hydro. Processes*, 29(6), 1267-1279.

685 Yu, J., Qin, X., Chiew, Y. M., Min, R., Shen, X., 2017. Stochastic Optimization Model for
686 Supporting Urban Drainage Design under Complexity. *J. Water Resour. Plann. Manage.*,
687 143(9), 05017008.

688

689 **Table Captions:**

690 Table 1. Flooding information and allocated storage ponds for sub-watersheds in Scenario I.

691 Table 2. Optimal layout of storage ponds in each sub-watersheds and overall performance of TO
692 scheme under Scenario I.

693 Table 3. Results of two optimization schemes under Scenarios I, IA, IB, II and III.

694 Table 4. Results of GO scheme under different GA options.

695

Table 1. Flooding information and allocated storage ponds for sub-watersheds in Scenario I.

Sub-watershed No. (S_i)	Flood volume (10^3 m^3)	Flood percentage (%)	Number of allocated ponds (p_i)	Number of original junctions
S_1	192.5	27.6	11	20
S_2	155.3	22.2	10	21
S_3	181.1	26.0	11	20
S_4	169.0	24.2	10	20
Sum	697.9	100	42	81

Table 2. Optimal layout of storage ponds in each sub-watersheds and overall performance of TO scheme under Scenario I.

Sub-watershed	Optimal layout of storage pond												Cost (\$)
<i>S1</i>	Pond	J09	J36	J61	J51	J60	J50	J48	J25	J17	J37	J10	10,293,580
	Area (m ²)	9800	6298	5554	9604	6954	3435	5713	7464	2115	500	2468	
<i>S2</i>	Pond	J47	J23	J21	J32	J44	J42	J57	J41	J40	J31	9,316,161	
	Area (m ²)	3655	7221	3457	9193	6241	3274	6068	6491	5549	2298		
<i>S3</i>	Pond	J68	J78	J82	J72	J65	J80	J69	J63	J74	J77	J67	10,855,816
	Area (m ²)	4088	7058	4768	7471	8509	4297	4758	5820	5391	5521	4446	
<i>S4</i>	Pond	J04	J05	J06	J11	J12	J13	J19	J28	J38	J54	11,945,893	
	Area (m ²)	6969	7604	6680	6478	6942	6219	7567	5867	5911	6244		
Overall flood volume (m ³)					78,418				Overall Cost				42,411,450

Table 3. Results of two optimization schemes under Scenarios I, IA, IB, II and III.

Scenarios	Optimized number of ponds: TO					Optimized number of ponds: GO	Overall cost (\$)		Total flood volume (m ³)		Total run time (hour)	
	<i>S1</i>	<i>S2</i>	<i>S3</i>	<i>S4</i>	Sum		TO	GO	TO	GO	TO	GO
I	11	10	11	10	42		42,411,450		78,418			
IA	12	10	10	17	49	71	50,335,665	48,090,220	79,483	79,934	3.3	6.5
IB	11	8	15	9	43		43,456,687		79,366			
II	9	9	9	11	38	66	38,866,884	43,252,806	117,866	111,278		
III	13	11	13	12	49	74	48,108,047	52,684,022	59,283	58,422		

708

Table 4. Results of GO scheme under different GA options.

709

No. of run	Options of GA		Results of GO	
	Elite count	Crossover fraction	Overall cost (\$)	Total flood volume (m ³)
1	3	0.6	49,763,148	79,392
2	6	0.7	46,628,648	79,991
3	10	0.8	48,090,220	79,934
4	15	0.85	46,795,460	79,984

710

711

712 **Figure Captions:**

713 Figure 1. Flow diagram of the proposed design framework.

714 Figure 2. Layout of the studied drainage network and model setting.

715 Figure 3. Hyetograph of the design rainfall.

716 Figure 4. Division of sub-watersheds over the study domain.

717 Figure 5. Regression performance of quadratic surrogate model for tributary sub-watersheds *S1*
718 to *S4*.

719 Figure 6. Results based on Scenario III with different SPs for (a) optimized pond areas and costs
720 of each sub-watershed from TO scheme and (b) optimized pond areas and final costs from GO
721 scheme.

722 Figure 7. Sample scenarios of flood volumes at tributary sub-watersheds (*S1*, *S2* and *S3*) and
723 mainstream sub-watershed *S4*.

724 Figure 8. Flood volumes at various junctions (with a threshold of 500 m³) from two optimal
725 solutions (i.e. original Scenario I model and the model with Extra-con on controlling backflow
726 effect) under TO scheme.

727 Figure 9. The new division of sub-watersheds over the study domain for Scenario IB.

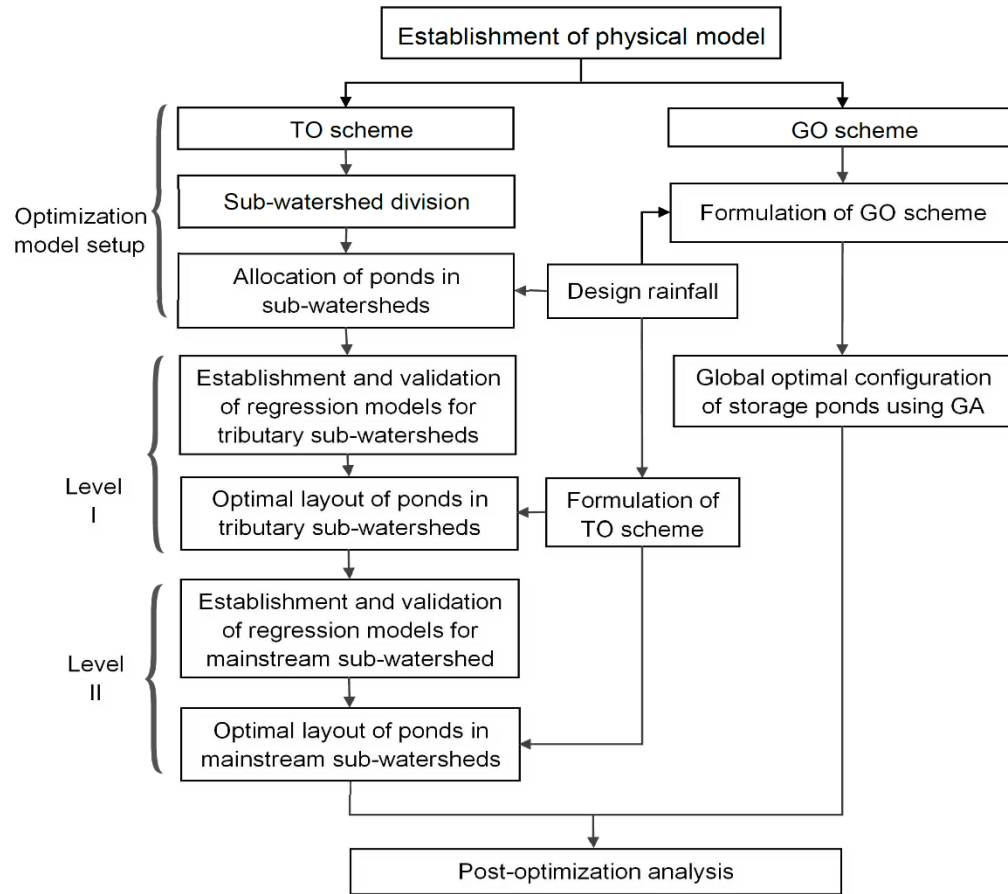
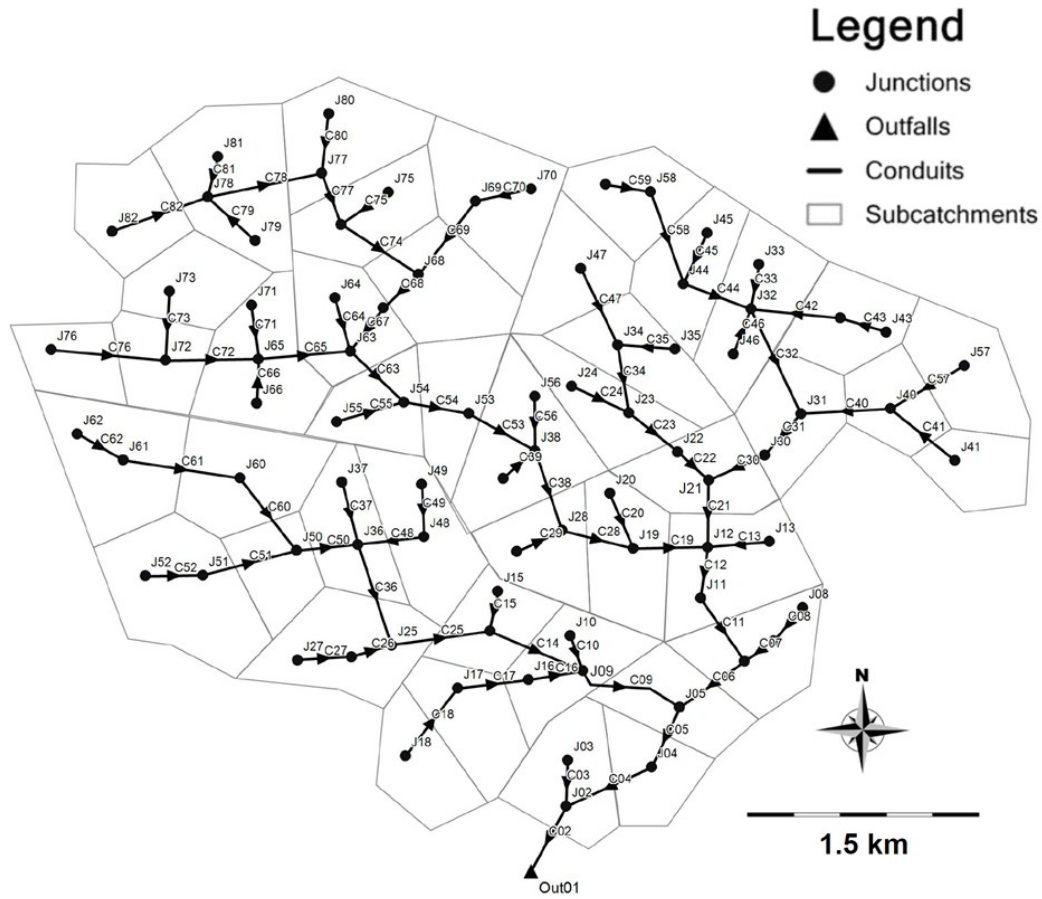


Figure 1. Flow diagram of the proposed design framework.

731



732

733

734

Figure 2. Layout of the studied drainage network and model setting.

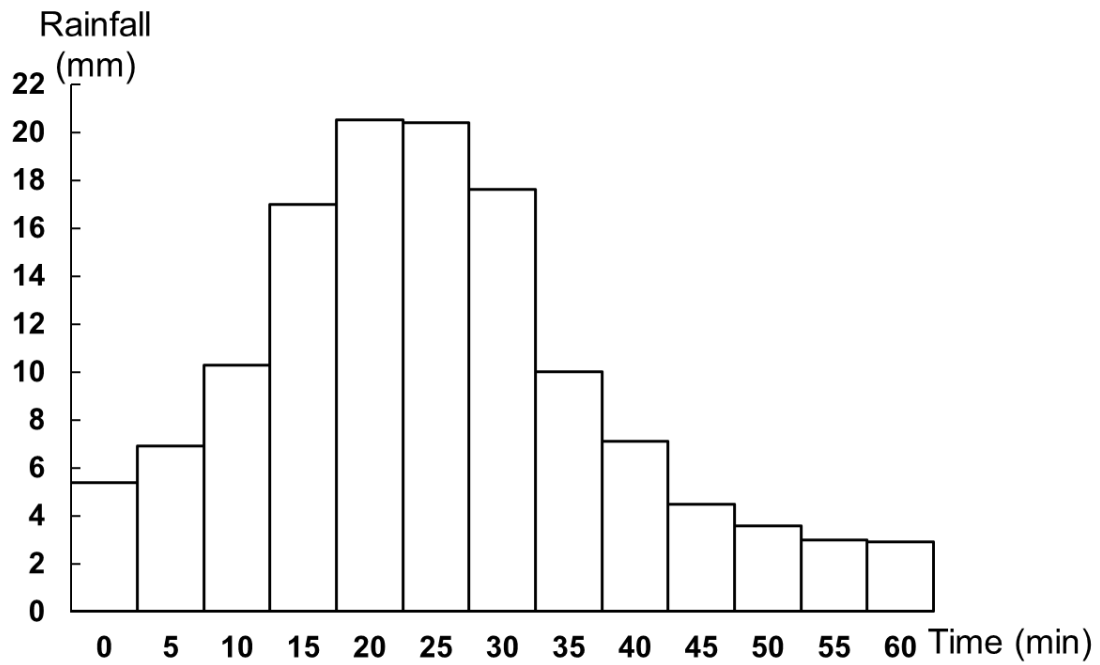
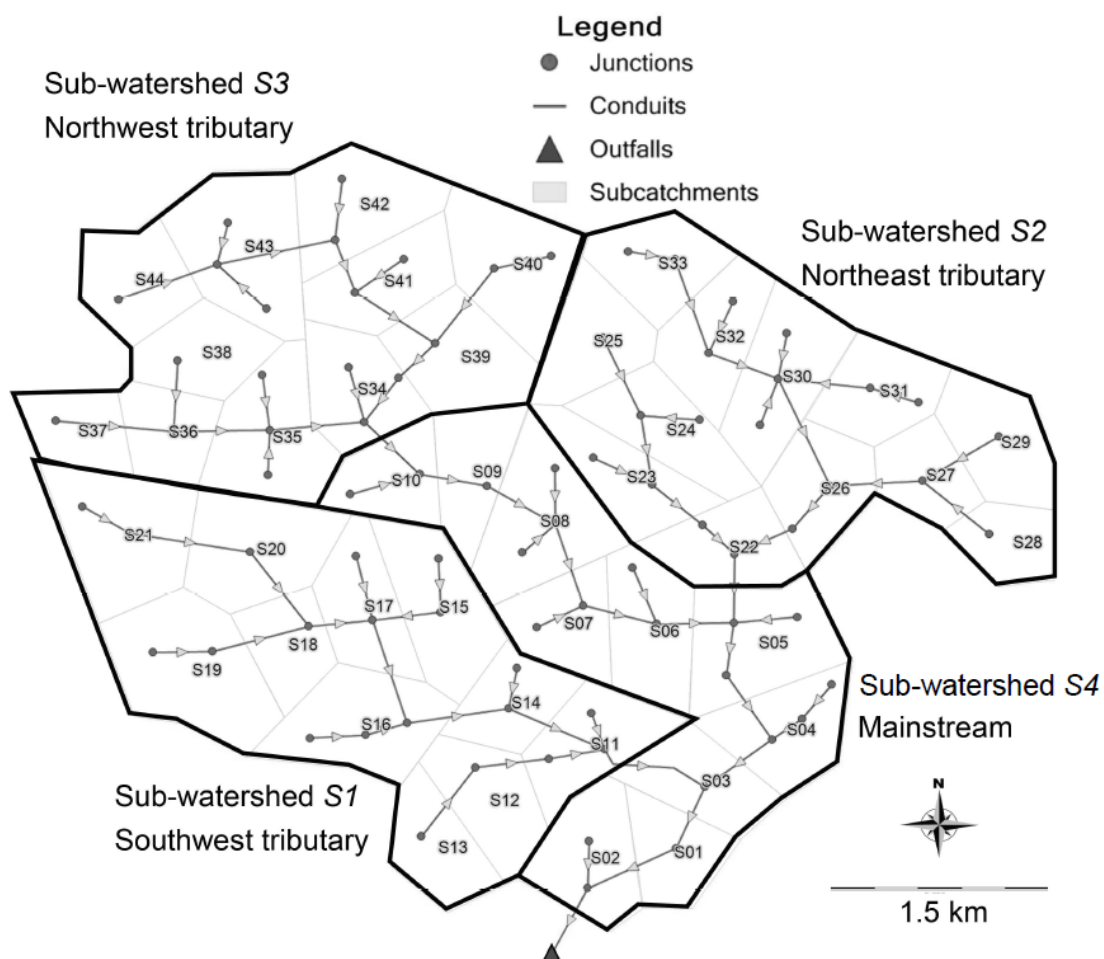


Figure 3. Hyetograph of the design rainfall

738



739

740

741

Figure 4. Division of sub-watersheds over the study domain.

742

743

744

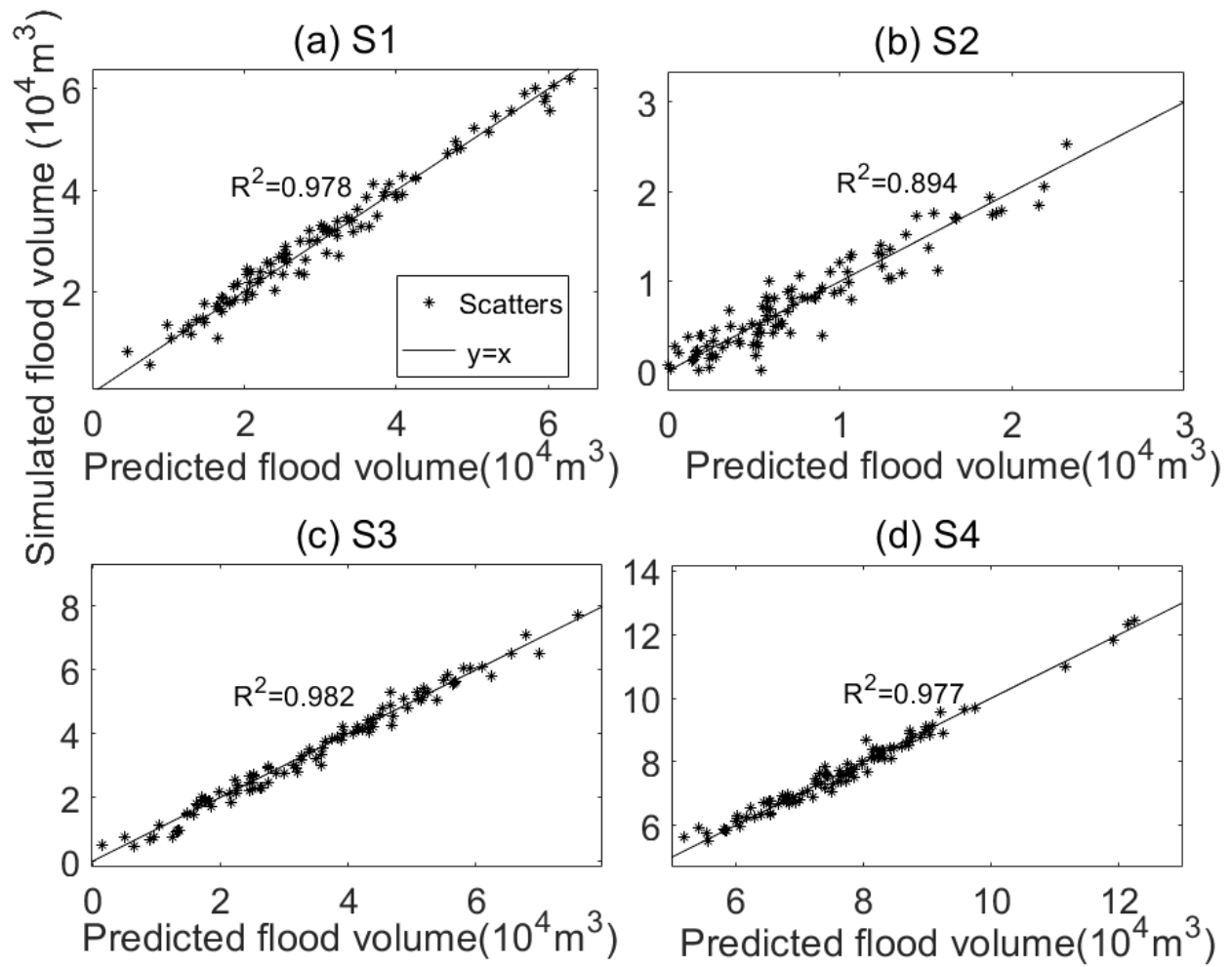


Figure 5. Regression performance of quadratic surrogate model for tributary sub-watersheds *S1* to *S4*.

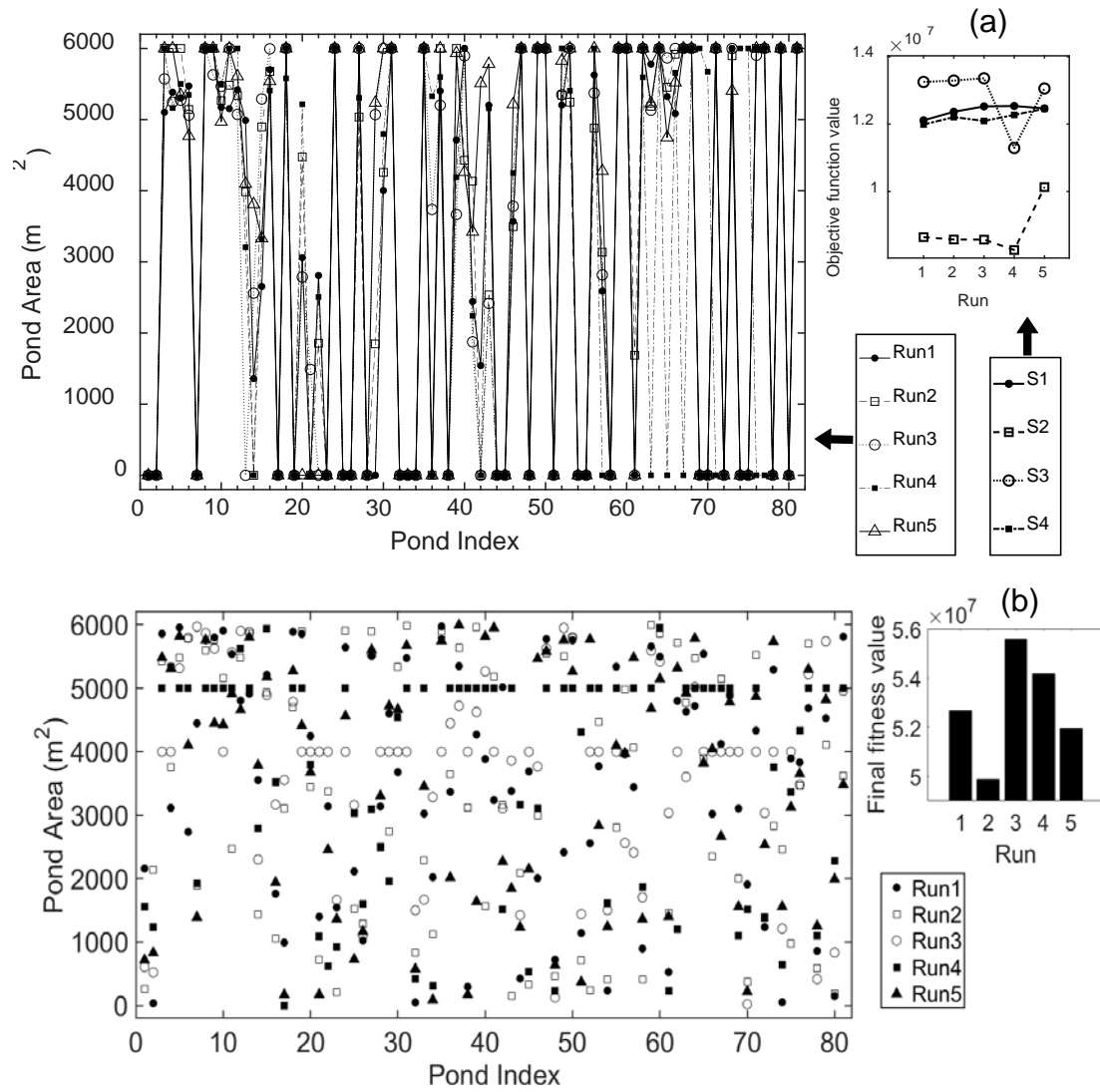


Figure 6. Results based on Scenario III with different SPs for (a) optimized pond areas and costs of each sub-watershed from TO scheme and (b) optimized pond areas and final costs from GO scheme.

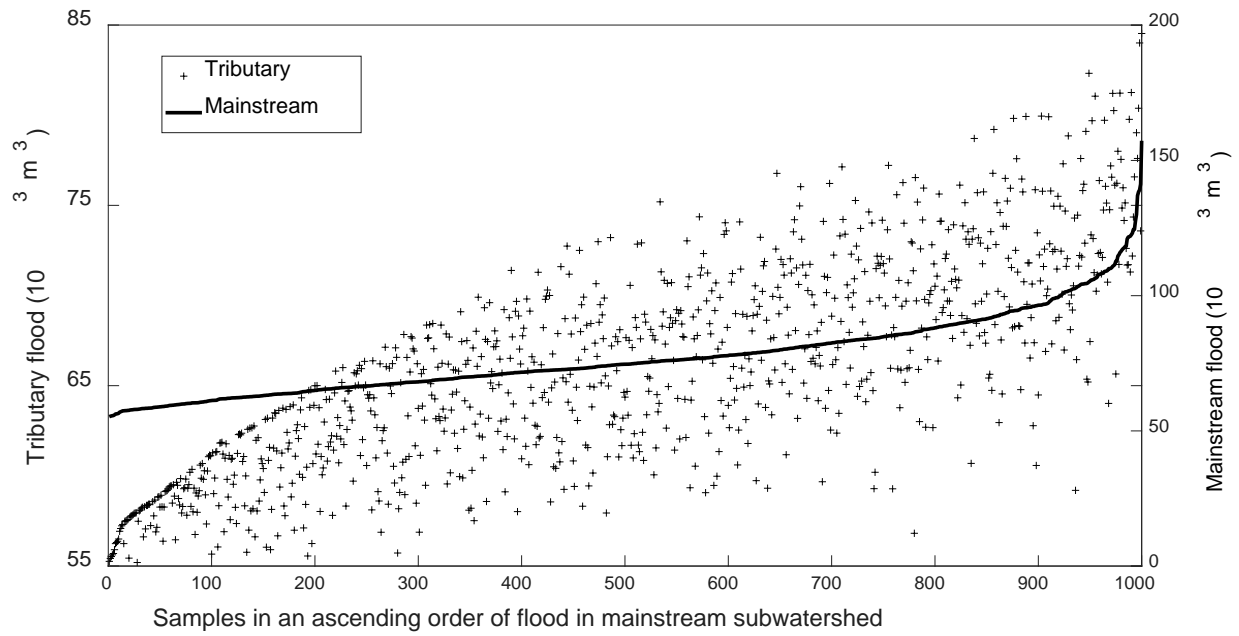
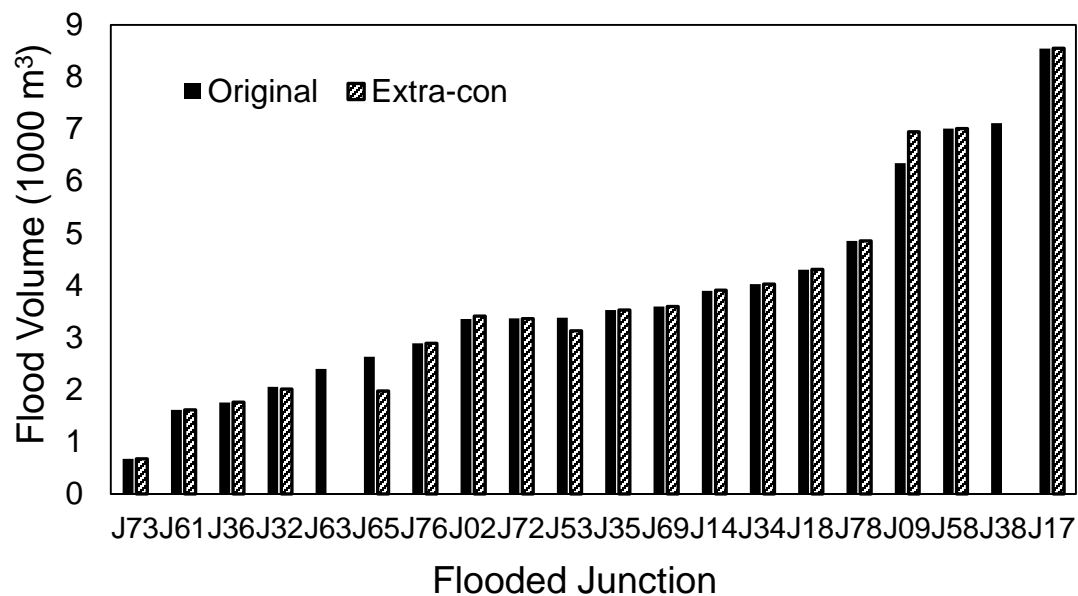


Figure 7. Sample scenarios of total flood volumes at tributary sub-watersheds (*S1*, *S2* and *S3*) and mainstream sub-watershed *S4*.

761



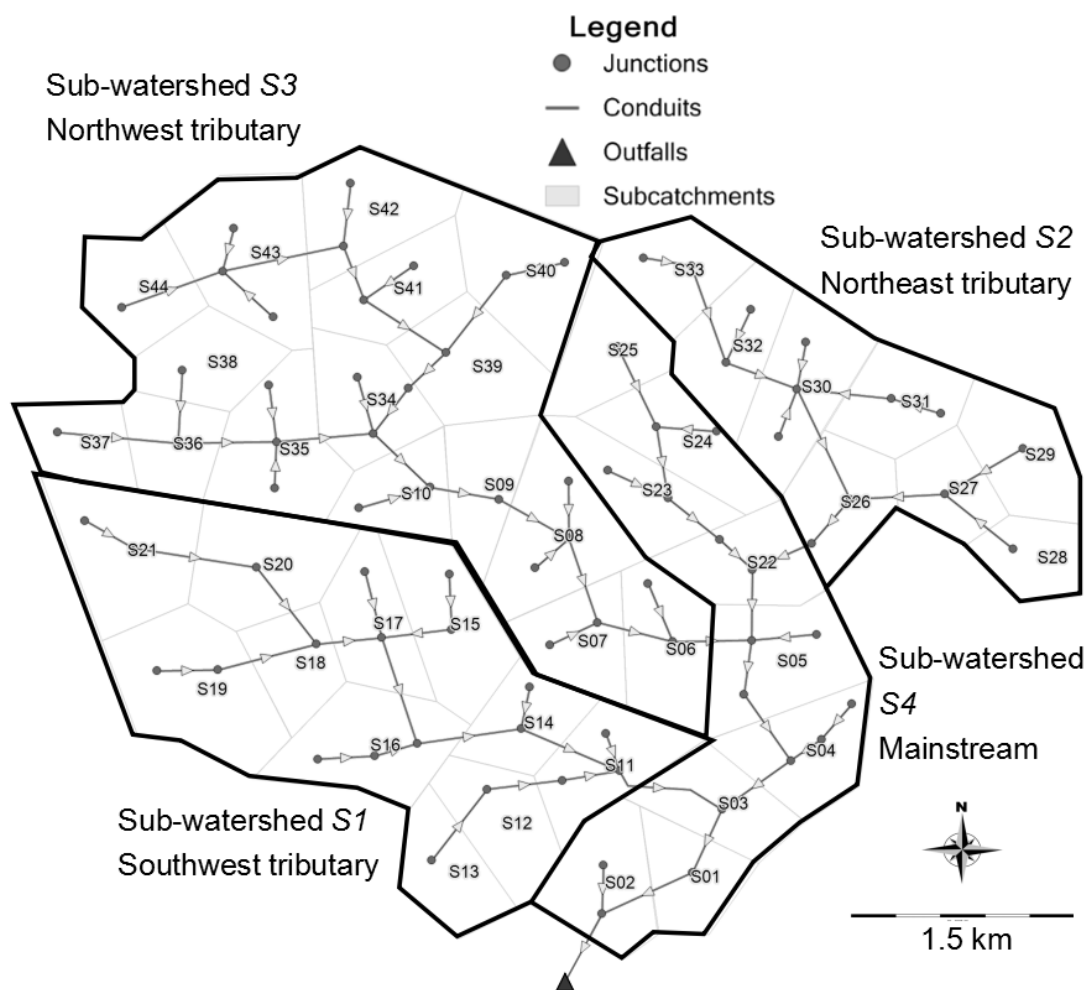
762

763

764 Figure 8. Flood volume at various junctions (with a threshold of 500 m³) from two optimal
 765 solutions (i.e. original Scenario I model and the model with Extra-con on controlling backflow
 766 effect) under the TO scheme.

767

768



769

770

771 **Figure 9. The new division of sub-watersheds over the study domain for Scenario IB.**