

ON THE USE OF SURROGATE MODEL ENSEMBLES FOR
BIORETENTION CELL DESIGN OPTIMISATION UNDER CURRENT AND
FUTURE CLIMATE CONDITIONS

RAHMAH KHALID

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF MASTER OF APPLIED
SCIENCE

GRADUATE PROGRAM IN CIVIL ENGINEERING

YORK UNIVERSITY

TORONTO, ONTARIO

AUGUST 2021

© Rahmah Khalid, 2021

Abstract

Urbanisation and climate change are contributing to increased stormwater runoff, which exacerbates the risk of urban flooding and adversely affects our receiving waters. Bioretention cells (BRCs) can be used to mitigate these impacts, however finding their optimal design can be a time-consuming process, further exacerbated by using physical-based models. Surrogate models, an alternative to physical-based models, are investigated, to determine the impact architecture has on the performance of these stochastic models. Ultimately, the quality of the input set is the biggest determinant of model performance, and selecting the appropriate input set significantly improves model prediction accuracy. This surrogate is then used in a simulation-optimisation framework to streamline the optimisation of BRC design under current and future climate. BRC performance under future climate decreases in comparison with current climate, and existing design guidelines should be updated to reflect those findings.

Acknowledgments

Thank you Dr. Khan for all your guidance and support.

Thank you Ramy, Bitu, Donia. You guys were the real support system during this journey.

Thank you Everett and Michael. I was lucky to benefit from your super brains.

Thank you to York University and NSERC for funding this research.

Table of contents

<i>Abstract</i>	<i>ii</i>
<i>Acknowledgments</i>	<i>iii</i>
<i>Table of contents</i>	<i>iv</i>
<i>List of Figures</i>	<i>vii</i>
<i>List of Tables</i>	<i>x</i>
<i>List of Abbreviations</i>	<i>xi</i>
<i>Chapter 1 - Introduction</i>	<i>1</i>
1.1. Urbanisation	1
1.2. Climate change	2
1.2.1. General Circulation Models (GCMs)	3
1.2.2. Regional Climate Models (RCMs)	4
1.3. Climate Change Impacts on Stormwater	4
1.4. Low Impact Development (LID)	5
1.5. Bioretention cells (BRCs)	6
1.6. Surrogate Models	8
1.7. Optimisation	11
1.8. Knowledge gaps	12
1.9. Research Objectives	13
1.9.1. Surrogate model development objectives	13
1.9.2. BRC design objectives	13
1.10. Thesis outline	14
<i>Chapter 2 – Bioretention cell modelling</i>	<i>15</i>
2.1. Modelling of LIDs	15
2.1.1. SWMM	15
2.1.2. L-THIA-LID	16
2.1.3. HEC-HMS	17
2.1.4. SUSTAIN	17

2.2.	Study site.....	19
2.2.1.	BRC design configuration	21
2.3.	Artificial Neural Networks – Surrogate Model.....	23
2.4.	Rainfall scenarios.....	24
<i>Chapter 3 – An investigation of surrogate model development for bioretention cell design in place of SWMM.....</i>		<i>27</i>
3.1.	Introduction.....	27
3.1.1.	Physics – based models	27
3.1.2.	Surrogate Models.....	28
3.1.3.	Objectives	30
3.2.	Methods.....	31
3.2.1.	Artificial neural networks (ANNs)	31
3.2.2.	Surrogate model synthesis	31
3.2.3.	ANN model configurations	32
3.2.4.	Performance Metrics.....	36
3.3.	Results and Discussion.....	38
3.3.1.	Number of training samples per rainfall event	38
3.3.2.	Random vs stratified sampling for datasplitting	41
3.3.3.	Input selection.....	43
3.4.	Conclusion.....	49
<i>Chapter 4 – Surrogate model assisted optimisation of bioretention cell design under current and future climate conditions</i>		<i>52</i>
4.1.	Introduction.....	52
4.1.1.	Climate change	53
4.1.2.	Bioretention cells.....	55
4.1.3.	Simulation-optimisation framework for BRC design	55
4.1.4.	Objectives	57
4.2.	Methods.....	58
4.2.1.	BRC design configuration	58
4.2.2.	SM development	58
4.2.3.	Rainfall Scenarios	59
4.2.4.	Optimisation of BRC design.....	60

4.3.	Results and discussion	67
4.3.1.	Surrogate model reliability	67
4.3.2.	Generation number selection.....	69
4.3.3.	Optimisation under current design guidelines	70
4.3.4.	BRC performance under adapted design guidelines.....	74
4.4.	Conclusion.....	79
<i>Chapter 5 - Conclusion</i>		<i>81</i>
5.1.	Limitations	83
5.2.	Future recommendations	83
<i>References.....</i>		<i>84</i>
<i>Appendices</i>		<i>95</i>
Appendix A: Surrogate Model performance for all storms		95
Appendix B: BRC optimised design performance under current and future climate conditions for all storms.....		104
Appendix C: Pseudo code		112

List of Figures

Figure 1-1: Schematic of a BRCs showing the structural components.	7
Figure 1-2: Schematic of multi-objective optimisation showing the Pareto optimal front.....	12
Figure 2-1: Proposed study site where the BRC (shown by the purple rectangle) receives runoff from the impervious areas delineated by the dashed lines.....	20
Figure 2-2: Schematic showing the recommended design values for the main parameters in a BRC (growing media depth, gravel storage depth and underdrain)	22
Figure 2-3: A schematic of a feedforward ANN showing the neurons, weights, biases and layers	24
Figure 2-4: AES-SO distribution for the 100-year 12-hour storm	25
Figure 3-1: Schematic showing one possible variation of LHS.....	33
Figure 3-2: Schematic explaining stratified sampling – from top to bottom the schematic shows the sorting of input data into respective strata resulting in an equal representation of input data	34
Figure 3-3: Schematic showing the probability distribution functions of two ensembles. Ensemble 1 is denoted in black, ensemble 2 is denoted in the dashed purple line and the observed value is marked.....	37
Figure 3-4: Comparison of model predictive capability for the LHS 50, 100 and 200 scenarios	39
Figure 3-5: comparison of the RHs for all three models. Left - LHS 50. Center - LHS 100. Right - LHS 200.	41
Figure 3-6: Comparison of model predictive performance under different stratified sampling goals. Right: Event based stratified sampling. Left: Flow based stratified sampling.....	42
Figure 3-7: Performance plots for the new improved surrogate. Left: Observed vs predicted. Right: RH.....	43
Figure 3-8: Model performance of the hydrological and water quality SM	45
Figure 3-9: Water quality surrogate results for all three design parameters of the BRC	46
Figure 3-10: Event-based model performance of the hydrological and water quality SM.....	48
Figure 4-1: SWMM/SM/NSGA-II framework for the BRC optimisation.....	66
Figure 4-2: Hydrologic SM performance for the 2-year 6-hour, 25-year 12-hour and 100-year 24- hour storms design storms.	68
Figure 4-3: Environmental SM performance for the 2-year 6-hour, 25-year 12-hour and 100-year 24-hour storms design storms. The top row contains the observed and predicted plots. The bottom row contains the corresponding RHs.....	69

Figure 4-4: Comparison of Pareto fronts under different generations	70
Figure 4-5: BRC Pareto fronts for the 2-year 6-hour storm for future and current climate under current design guidelines.....	71
Figure 4-6: BRC Pareto fronts for the 25-year 12-hour storm for future and current climate under current design guidelines.....	72
Figure 4-7: BRC Pareto fronts for the 100-year 24-hour storm for future and current climate under current design guidelines	73
Figure 4-8: BRC performance under design recommendations adapted for climate change for the 2-year 6-hour	75
Figure 4-9: BRC performance under design recommendations adapted for climate change for the 25-year 12-hour	76
Figure 4-10: BRC performance under design recommendations adapted for climate change for the 100-year 24-hour	77
 Figure A - 1: SM results for 2-year 6-hour storm.....	95
Figure A - 2: SM results for 2-year 12-hour storm	95
Figure A - 3: SM results for 2-year 24-hour storm	96
Figure A - 4: SM results for 5-year 6-hour storm.....	96
Figure A - 5: SM results for 5-year 12-hour storm	97
Figure A - 6: SM results for 5-year 24-hour storm	97
Figure A - 7: SM results for 10-year 6-hour storm	98
Figure A - 8: SM results for 10-year 12-hour storm	98
Figure A - 9: SM results for 10-year 24-hour storm	99
Figure A - 10: SM results for 25-year 6-hour storm	99
Figure A - 11: SM results for 25-year 12-hour storm	100
Figure A - 12: SM results for 25-year 24-hour storm	100
Figure A - 13: SM results for 50-year 6-hour storm	101
Figure A - 14: SM results for 50-year 12-hour storm	101
Figure A - 15: SM results for 50-year 24-hour storm	102
Figure A - 16: SM results for 100-year 6-hour storm	102
Figure A - 17: SM results for 100-year 24-hour storm	103
 Figure B - 1: BRC optimisation results - Existing versus Updated designs – 2-year 12-hour...	104
Figure B - 2: BRC optimisation results - Existing versus Updated designs – 2-year 24-hour...	104

Figure B - 3: BRC optimisation results - Existing versus Updated designs – 5-year 6-hour.....	105
Figure B - 4: BRC optimisation results - Existing versus Updated designs – 5-year 12-hour...	105
Figure B - 5: BRC optimisation results - Existing versus Updated designs – 5-year 24-hour...	106
Figure B - 6: BRC optimisation results - Existing versus Updated designs – 10-year 6-hour...	106
Figure B - 7: BRC optimisation results - Existing versus Updated designs –10-year 12-hour..	107
Figure B - 8: BRC optimisation results - Existing versus Updated designs – 10-year 24-hour.	107
Figure B - 9: BRC optimisation results - Existing versus Updated designs – 25-year 6-hour...	108
Figure B - 10: BRC optimisation results - Existing versus Updated designs – 25-year 24-hour	108
Figure B - 11: BRC optimisation results - Existing versus Updated designs – 50-year 6-hour.	109
Figure B - 12: BRC optimisation results - Existing versus Updated designs –50-year 12-hour	109
Figure B - 13: BRC optimisation results - Existing versus Updated designs –50-year 24-hour	110
Figure B - 14: BRC optimisation results - Existing versus Updated designs – 100-year 6-hour	110
Figure B - 15: BRC optimisation results - Existing versus Updated designs – 100-year 12-hour	111

List of Tables

Table 1-1 - Description of AR5 RCP scenarios (van Vuuren et al., 2011)	3
Table 1-2: A comparison of the performance metrics for various SMs used in hydrological studies.....	10
Table 2-1: A comparison of the 4 physical-models for LID modelling	19
Table 2-2: Subcatchment names and characteristics of impervious areas draining into the BRC	21
Table 2-3: Minimum and maximum BRC design parameters	23
Table 3-1: Comparison of model performance (in terms of NSE, CRPS and PC) of the three models (LHS 50, 100 and 200).....	39
Table 3-2: Comparison of model performance for flow and event based stratified sampling	42
Table 3-3: Performance results of the improved surrogate	43
Table 3-4: Model performance metrics for the hydrological and water quality SM	45
Table 3-5: Event-based model performance metrics for the hydrological and water quality SM	48
Table 4-1: Rainfall depths (mm) for all climate scenarios.....	60
Table 4-2: Maximum runoff volume (m ³) values under all storms	63
Table 4-3: Maximum TSS mass (kg) under all climate scenarios	64
Table 4-4: Design constraints of the decision variables.....	65
Table 4-5: Hydrologic SM performance results for the 2-year 6-hour, 25-year 12-hour and 100-year 24-hour storms.....	67
Table 4-6: Environmental SM performance results for the 2-year 6-hour, 25-year 12-hour and 100-year 24-hour storms	69
Table 4-7: BRC hydrologic and environmental performance under historic and all future climate scenarios for the largest BRC design possible	74

List of Abbreviations

AMALGAM	A Multi-Algorithm Genetically Adaptive Multiobjective
ANNs	Artificial Neural Networks
AES-SO	Atmospheric Environmental Services – Southern Ontario
BMPs	Best Management Practices
BRC	Bioretention Cell
CFM	Change Factor Methodology
CRPS	Continuous Ranked Probability Score
CMIP5	Fifth Coupled Model Intercomparison Project
CVC	Credit Valley Conservation
CDF	Cumulative Distribution Function
CN	Curve Number
DoE	Design of Experiments
DTM	Digital Terrain Model
ESMs	Earth System Models
EMC	Event Mean Concentration
AR5	Fifth Annual Report
GCMs	General Circulation Models
GAs	Genetic Algorithms
GI	Green Infrastructure
HEC-HMS	Hydrologic Engineering Centre – Hydrologic Modelling System
HPSF	Hydrological Simulation Program Fortran
IDF	Intensity Duration Frequency
LHS	Latin Hypercube Sampling
LID	Low Impact Development
MSE	Mean Square Error
mGLUE	Modified Generalized Likelihood Uncertainty Estimation
MC	Monte-Carlo
NSE	Nash-Sutcliffe Efficiency
NRCS	Natural Resources Conservation Service
NSGA- II	Non-Dominated Sorting Genetic Algorithm
NPS	Non-Point Source
PCIC	Pacific Climate Impacts Consortium
PC	Percent Capture
PDF	Probability Distribution Function
RH	Rank Histogram
RCMs	Regional Circulation Models
RCP	Representative Concentration Pathway
SBX	Simulated Binary Crossover
SCS	Soil Conservation Services
SWMM	Stormwater Management Model
SMs	Surrogate Models
STEP	Sustainable Technologies Evaluation Program
SUDS	Sustainable Urban Drainage Systems
SUSTAIN	System for Urban Stormwater Treatment and Analysis Integration
L-THIA-LID 2.1	Long-Term Hydrologic Impact Assessment-Low Impact Development 2.1
TSS	Total Suspended Solids

Chapter 1 - Introduction

Increasing urbanisation means the conversion of natural landscapes to impervious surfaces. Adding to that, the potential increases in frequency of extreme climate events, such as heavy rainfall, results in increased stormwater runoff, increased peak flowrate, and increased pollutant concentration. This chapter provides a brief overview of urbanisation, climate change and their impacts on stormwater runoff, as well as potential methods to mitigate these impacts. Knowledge gaps in research are identified, followed by research objectives and an outline of the thesis.

1.1. Urbanisation

With the increasing spread of commercial landscapes, rural and undeveloped lands are continuously converted to urban areas. These developments replace vegetated surfaces with impervious surfaces such as parking lots, roofs and streets. Consequently, the natural water balance of the watershed is disrupted leading to increased stormwater surface runoff volume and diminished infiltration capacity due to reduced vegetation (Davis, 2005). This excess runoff is usually transported to nearby receiving waters to prevent localized flooding.

This stormwater runoff transports with it increased nutrients, bacteria and toxic metals collected from impervious surfaces, and deposits them into our receiving waters leading to reduced aquatic diversity (Ahiablame *et al.*, 2012; Bedan & Clausen, 2009; Dietz & Clausen, 2008). The greater volumes of surface runoff coupled with the strong momentum of flowing water create favourable conditions for the mobilization of dissolved contaminants and solids (Davis, 2005). These contaminants can originate from a single point of pollution such as discharge from an industrial area, or from multiple sources without a direct pollution discharge such as rainfall and snowmelt depositing pollutants onto impervious surfaces (Wu & Chen, 2013). Stormwater runoff deposits these various contaminants into our receiving waters, and therefore, urbanisation is considered to be the biggest contributor to non-point source (NPS) pollution.

Given these adverse impacts of urbanisation on stormwater runoff, it is necessary to implement mitigative measures to ensure future resiliency. Such measures include the use of Low Impact Development (LID) units, such as permeable pavements, bioretention cells and green roofs,

which may be used to mimic pre-development hydrology, as well as mitigate against the impacts of climate change. An overview of climate change, and its impacts on runoff are presented, followed by an introduction into LIDs and their potential use to mitigate the combined impacts of urbanisation and climate change.

1.2. Climate change

In addition to urbanisation, climate change also contributes to increased runoff. However, to precisely define what constitutes climate change, an explanation of the difference between weather and climate is necessary. Weather describes the conditions of a place's meteorological elements (such as temperature, pressure, humidity and others), precipitation, and special phenomena (such as thunderstorms, tornados and others) at a specific instance in time (Cubasch *et al.*, 2013). Climate is defined as the average change in weather over a sustained period of time. Climate change is thus identified by the changes in climate that persist over a prolonged period of time, commonly decades (Cubasch *et al.*, 2013).

Indicators of climate change can be observed from changing historical trends such as increasing global temperatures or melting of glaciers. Combined global mean temperatures of land and ocean surface have increased by 0.85°C over the period of 1880 to 2012, with 16 of the 17 warmest years since 1880 having occurred in 2001 to 2016 (Fahey *et al.*, 2017). Changes in ocean characteristics have also been observed according to Fahey *et al.* (2017) with a warming rate of 0.11°C per decade from 1971 to 2010 in the top 75 m. Further, global mean sea levels have risen by 0.19 m with a rising rate greater than that of the previous two millennia since the mid-19th century. Additionally, their report indicates a decrease in the annual average Arctic sea-ice extent, with a rate in the range of 3.5 to 4.1% per decade over the period of 1979 to 2012. Although these findings point to a global increasing trend, these changes may be spatially variable, such as different warming rates at the poles than the equator.

Based on historical and recent trends, further changes in climate can be predicted using a variety of methods. Prediction of future climate change and its impacts is a necessity to ensure communities can both mitigate, by proactively addressing the impacts of climate change, and adapt, by ensuring the implemented measures protect against the possible future climate conditions. An increasingly popular trend worldwide, is the use of LIDs as an adaptive method to climate change impacts (Barah *et al.* 2021). LIDs can also act as a supplement to existing urban

drainage systems alleviating the stress on these systems, and consequently reducing the potential for localised flooding. To determine the severity of future climate and the consequent need for LIDs, a number of prediction methods exist, including: General Circulation Models (GCMs) and Regional Climate Models (RCMs), amongst others.

1.2.1. General Circulation Models (GCMs)

General Circulation Models (GCMs) are three dimensional mathematical models that represent the interactions between the Earth's atmosphere, ocean, cryosphere and land surface, and are used to simulate global climate responses to external forcings. Current GCMs are known as Earth System Models (ESMs) which are the most comprehensive models to date. Expanding on previous models that incorporated representation of only physical processes of the climate system, ESMs include representations of the various biogeochemical cycles (such as the carbon cycle and ozone) in climate modelling (Flato *et al.*, 2013). GCMs predict future climate change trends by incorporating inputs related to climate change scenarios developed by the Intergovernmental Panel on Climate Change (IPCC). The scenarios for driving climate change in the fifth Annual Report (AR5) are known as the Representative Concentration Pathways (RCP), named according to the target radiative forcing level for 2100. Table 1-1 outlines the four RCP scenarios used in IPCC's AR5. IPCC's sixth annual report, including new climate change scenarios known as the Shared Socioeconomic Pathways (SSP) has been released, but is not discussed here as it was released at the time this thesis work was completed.

*Table 1-1 - Description of AR5 RCP scenarios (van Vuuren *et al.*, 2011)*

Scenario	Description
RCP2.6	Radiative forcing peak at $\sim 3 \text{ W/m}^2$ before 2100 followed by a decline to 2.6 W/m^2 by 2100
RCP4.5	Stabilization to 4.5 W/m^2 without a rise and fall by 2100.
RCP6	Stabilization to 6 W/m^2 without a rise and fall by 2100.
RCP8.5	Rising trend to 8.5 W/m^2 by 2100.

GCMs have very coarse spatial resolutions ranging from 10,000 km² to 90,000 km² per grid cell (Anandhi *et al.*, 2011). This can create biases in the simulation results due to the mismatched spatial scales of the study sites in comparison to the GCM resolution. Anandhi *et al.* (2011) reported a number of existing methods to overcome this problem including: Change Factor Methodology (CFM), statistical downscaling and dynamical downscaling. CFM involves the manipulation of observed climate data to produce relevant spatial data that follows the trend of the GCM. A change factor is applied to observed rainfall data to project future rainfall data in line with trends suggested by the GCM scenario. Statistical downscaling methods form relationships between the local and future climate variables. Dynamical downscaling methods involve using regional finer scale models such as Regional Climate Models.

1.2.2. Regional Climate Models (RCMs)

The spatial scale of GCMs is often too coarse, making their use impractical for small area climate predictions. The GCM outputs are thus transformed to finer resolutions using dynamic downscaling methods like RCMs. The RCMs often utilize the same initial and boundary conditions used in their coarser GCM counterparts (Willems *et al.*, 2012). Using RCMs can amplify the climate uncertainty at the regional scale due to numerical computation limitations, and therefore they need to be further combined with other downscaling methods to correct resulting systematic bias (Tirpak *et al.*, 2021; Willems *et al.*, 2012).

These different methods of simulating climate change are valuable tools to facilitate predictions of future climate and its impacts on stormwater.

1.3. Climate Change Impacts on Stormwater

In addition to increasing surface global temperatures, climate change is expected to alter the global water cycle. An increase in mean annual precipitation is expected to occur in high latitude countries such as Canada (IPCC, 2014). Further, given the potential increase in extreme rainfall events, cities are at a heightened risk of urban flooding. The risks associated with these impacts include; social risks such as illness outbreaks due to contact with contaminated runoff during flooding events, loss of human life, environmental risks due to the transport of these contaminated waters to nearby rivers or lakes, reducing the ecological diversity, and finally systemic risks such as substantial economic losses due to the breakdown of infrastructure.

Conventional urban drainage systems, consisting of pipes and underground basins, are designed to quickly transport surface runoff from urban areas via a sewer system network to nearby receiving water bodies (Goulden *et al.*, 2018; Zhou, 2014; Zhou *et al.*, 2019). Traditionally their design has been based on historical climate trends and the assumption of stationarity in climate trends (Pour *et al.*, 2020). However, given the aforementioned studies identifying trends of increased precipitation and stormwater runoff under future climate change, existing urban drainage networks will likely be overburdened as they are undersized and in poor condition, further intensifying the risk of localized flooding in urban areas (Barah *et al.*, 2021; Ngamalieu-Nengoue *et al.*, 2019). Replacing existing stormwater networks to pipes of larger sizes can be extremely expensive and may not be feasible due to surrounding infrastructure (Barah *et al.*, 2021). Additionally, upsizing pipe diameters does not address the issue of contaminant transport in runoff and the deposition of these dissolved pollutants into our receiving waters. An alternative to this expensive and disruptive replacement process would be through the installation and use of LID units (Barah *et al.*, 2021; Tirpak *et al.*, 2021; Wang *et al.*, 2019; Zhang *et al.*, 2019).

1.4. Low Impact Development (LID)

The term LID emerged in Maryland in the 1990s and has since then been the widely adopted term in North America and New Zealand. Other countries have adopted different terms such as Water Sensitive Urban Design (WSUD) in Australia, Sustainable Urban Drainage Systems (SUDS) in the United Kingdom. Though the names differ, very similar concepts and principles are shared with LIDs (Dietz, 2007; Eckart *et al.*, 2017; Sohn *et al.*, 2019). LIDs are commonly employed adaption strategies which aim to ease the increased stress on conventional stormwater management systems due to urbanisation and climate change (Eckart *et al.*, 2017). Ideally, LIDs aim to replicate the pre-development hydrological conditions of the developed watershed, through increased on-site infiltration and evapotranspiration. (Dietz, 2007; Eckart *et al.*, 2017; Sohn *et al.*, 2019). This will help maintain pre-development stormwater runoff volume, reduce peak flow rates, increase the lag time, increase groundwater recharge and reduce pollutant mobilization (Sohn *et al.*, 2019).

In addition to their runoff reduction benefits, LIDs that fall under Green Infrastructure (GI) can provide a multitude of additional environmental and socioeconomic benefits (Kaykhosravi *et al.*, 2019). GI practices focus on landscape and urban planning and the integration of as much green spaces as possible (Eckart *et al.*, 2017; Sarah Kaykhosravi *et al.*, 2019). Therefore, environmental benefits include improved water quality through filtration of pollutants, reduction of air pollutant

concentrations, persevering existing habitats and improving biodiversity (Kaykhosravi *et al.*, 2019). The socioeconomic benefits include increasing access to green spaces, enhancing student cognitive performance in schools, improving the mental health status of hospital patients and finally, enhancing immune function and lowering fasting blood-glucose levels in both adults and children (Kaykhosravi *et al.*, 2019).

Bioretention cells (BRCs), a type of LID, are commonly incorporated in watersheds with a proven performance in reducing runoff volume and peak flow rate (Ahiablame *et al.*, 2012; Khan *et al.*, 2012a, 2012b; Tirpak *et al.*, 2021; Wang *et al.*, 2019). Additionally, they are the most commonly researched LID according to the findings of Sohn *et al.* (2019). Therefore, given their popularity, they are worthy of further examination to determine their ability to mitigate the combined effects of urbanisation and climate change.

1.5. Bioretention cells (BRCs)

BRCs, also known as rain gardens, are one of the most popular LID practices in urban watersheds (Johnson & Hunt, 2019; Kratky *et al.*, 2017). BRCs are generally small, depressed units consisting of plant and soil media, designed to infiltrate, filtrate and temporarily store runoff they receive (Davis, 2005; Khan *et al.*, 2012a). Underlying the vegetation layer is a porous growing media layer, mainly responsible for the BRCs pollutant removal through a series of chemical (e.g. adsorption), physical (e.g. sedimentation) and biological processes (e.g. denitrification) (Kratky *et al.*, 2017). Once the runoff percolates through the growing media it then exits the cell either through an underdrain or through infiltration into the surrounding subsoils.

Though their specific design may differ in each application, their general profile (shown in Figure 1-1) follows a similar layout; a top vegetation layer, a growing media layer, a gravel storage layer and possibly an underdrain (City of Toronto, 2017). The need for an underdrain is mainly a function of the surrounding soil permeability, as well as the exiting water volume and regional groundwater usage (Kratky *et al.* 2017). As the ponded surface water percolates into the highly porous growing media, it undergoes a series of physical, biological and chemical processes that aim to remove contaminants, thereby improving the quality of runoff (Khan *et al.* 2012a; 2012b). Recommended BRC designs for Southern Ontario are offered in the Credit Valley Conservation and Toronto and Region Conservation Authority LID Stormwater Management Planning and Design Guide (2010), and are outlined in Chapter 2. It should be noted that BRC designs are subject to change depending on the particular jurisdiction within which they are placed.

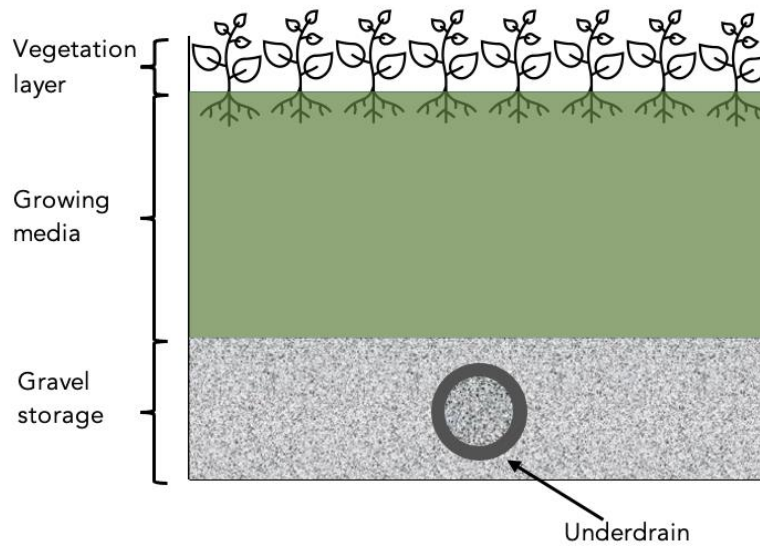


Figure 1-1: Schematic of a BRCs showing the structural components.

When built on impervious surfaces, BRCs can mitigate the impacts of urbanisation and climate change as they act as pervious surfaces aiding in runoff infiltration and filtration. This can help prevent localized flooding, alleviate existing stresses on aging infrastructure and enhance the runoff water quality exfiltrating into the surrounding soils.

BRC performance in runoff retention, peak flow reduction and water quality improvement is highly variable across research studies (Batalini de Macedo *et al.*, 2019; Jia *et al.*, 2015; Khan *et al.*, 2012b, 2012a; LeFevre *et al.*, 2015; Singh *et al.*, 2019), since BRC performance is highly dependent on its design configuration. Reported results have varying efficiencies ranging from 66% to 96% in peak flow reductions, 62% to 93.5% in runoff reduction (Jia *et al.*, 2015; U.T. Khan *et al.*, 2012a) and 48% to 99% in Total Nitrogen (TN) change of mass (Jia *et al.*, 2015; U.T. Khan *et al.*, 2012b). To arrive at an optimal site specific design, several evaluations of design configuration iterations are necessary, as existing design guidelines offer a broad range of design values (Bacys *et al.*, 2019). This can result in a design process that is iterative and time consuming. Computer modelling is often used to evaluate proposed BRC designs, with the most common model being the Stormwater Management Model (SWMM) (Eckart *et al.*, 2017).

SWMM, a physics-based model, uses principles of mass conservation to model a system. Using SWMM for optimisation purposes is highly time-consuming due to the significant allocation of time

for models requiring a large number of simulations and iterative design evaluations, such as the case with BRC design optimisation (Ghodsi *et al.*, 2020). Additionally, these models may be difficult to calibrate and validate for studies relating to BRCs if the BRC has not yet been built. Surrogate Models (SMs) are one alternative to physical-based models, favoured for their benefits, discussed below.

1.6. Surrogate Models

Surrogate models are numerical approximators of more complex models, such as SWMM, trained on a set of known inputs generated from the original model (Brunetti *et al.*, 2017). SMs are capable of evaluating simulations in a fraction of the time original physical-based models require, all while maintaining high levels of accuracy (Reichstein *et al.*, 2019). Artificial Neural Networks (ANNs) are commonly used SMs in literature (Ghodsi *et al.*, 2020; Khu & Werner, 2003; Mohammadiun *et al.*, 2018; Yazdi & Salehi Neyshabouri, 2014; Zhang *et al.*, 2019). ANNs are widely used in rainfall-runoff simulations and are considered universal approximators, capable of simulating complex non-linear relationships (Dawson & Wilby, 2001), making them an ideal candidate for SMs.

The use of SMs in place of their physics-based model counterpart is widely prevalent in the field of water resources. Brunetti *et al.* (2017) used a kriging model as the SM in place of the Hydrus-2D mechanistic model to evaluate and demonstrate the efficacy of using SMs in place of mechanistic models for LID analysis. Since the kriging model is computationally inexpensive, 40,000 optimisation simulations were run, a task that would've taken almost a month of continuous simulations using HYDRUS-2D (Brunetti *et al.*, 2017). Comparisons of the SM with the measured hydrographs showed a Nash-Sutcliffe Efficiency (NSE) value close to 0.8, proving their efficacy in producing highly accurate and computationally inexpensive results. Khu and Werner (2003) introduce a Monte-Carlo (MC) simulation technique known as the Modified Generalized Likelihood Uncertainty Estimation (mGLUE) method. The sampling technique of the mGLUE method remains unchanged, however, a Genetic Algorithm (GA) and an ANN are used to map the response surface in place of the Monte Carlo simulation. Mohammadiun *et al.* (2018) aimed to integrate external loading (extreme rainfall) and internal loading (structural failure) to develop a rehabilitation framework for Urban Stormwater Drainage Systems (USDSSs). An ANN was used as the SM to SWMM to run the optimisation of the USDSS designs. The 50th generation GA optimisation Pareto front is evaluated using both the SWMM and ANN model to compare results, where it is revealed that good agreement exists between both models with only a relative

error of 25%. Zhang *et al.* (2019) used an ensemble of ANN's as the SM to the MIKE-FLOOD physics-based model in an optimisation framework to identify the optimal design of detention tanks under different rainfall conditions. An ensemble of ANN's was used as opposed to one ANN to compare the optimal results produced by the single ANN and the ensemble. A summary of the performance of SMs and their respective error indices is provided in Table 1-2.

Table 1-2: A comparison of the performance metrics for various SMs used in hydrological studies

Source	Type of SM	Type of physical model	R ² of SM to observed outputs	NSE of SM to observed outputs	MAE of SM to observed outputs	MSE of SM to observed outputs	Relative error between physical and SM
Brunetti et al. (2017)	Kriging Model	Hydrus – 2D	0.98	0.8	N/A	N/A	N/A
Khu and Werner (2003)	ANN	SWMM	0.999 (training) 0.997 (testing)	N/A	N/A	N/A	N/A
Mohammadiun et al. (2018)	ANN	SWMM	N/A	N/A	N/A	N/A	25%
Yazdi and Salehi Neyshabouri (2014)	ANN	HEC-HMS	N/A	N/A	0.003	N/A	N/A
Raei et al. (2019)	ANN	SWMM	0.99-validation	N/A	N/A	6550 (Model 1 – Runoff) 12138 (Model 2 - TSS) 451 (Model 3 - BOD)	N/A

Using SMs ensures a high level of accuracy is maintained whilst also expediting the optimisation process of BRC design significantly. Optimisation algorithms can provide many optimal site

specific design configurations, of which the design that best caters and adheres to the requirement of decision makers and designers can be selected.

1.7. Optimisation

BRC performance in reduction of runoff volume and peak flows is highly dependent on where it is situated in the watershed. Watershed hydrological properties such as slope, soil permeability, depression storage, land use and upstream drainage area all greatly affect various aspects of runoff and consequently BRC performance. Additionally, the sizing, shape, and depth of each BRC unit can significantly impact its performance. BRCs should also be implemented with the least cost, as this is a common requirement for municipal planners. Therefore, given the many factors affecting BRC performance, the primary aim of optimisation is determining which combination of these factors maximises BRC performance.

Optimisation algorithms generally fall under two categories: single and multi-objective optimisation. Single objective optimisation usually consists of the minimisation or maximisation of a single objective function to arrive at the optimal solution (or solutions) (Deb, 2014). Objective functions are a measure of the quality of the solution. In other words, if the objective function is the cost of LID design and it is to be minimised then the most optimal solution will be one that provides the lowest LID design cost.

Multi-objective optimisation refers to the minimisation or maximisation of multiple objectives simultaneously. Often this results in a number of trade-off optimal solutions which provide an easily identifiable front plotted on the objective space with the objective values, as shown in Figure 1-2. This front is known as the Pareto-optimal front and all solutions found along this front are called Pareto-optimal solutions (Deb, 2014). In a minimisation of objective functions, a vector solution x_i belongs in the Pareto-optimal set over a solution x_j if x_i is no worse than x_j in all objectives and x_i is less than x_j in at least one objective. Solution x_i is thus considered Pareto dominant over x_j , as the objectives are being minimised. It is possible for a solution to be non-dominated (belong in the Pareto-optimal set) without dominating all other solutions.

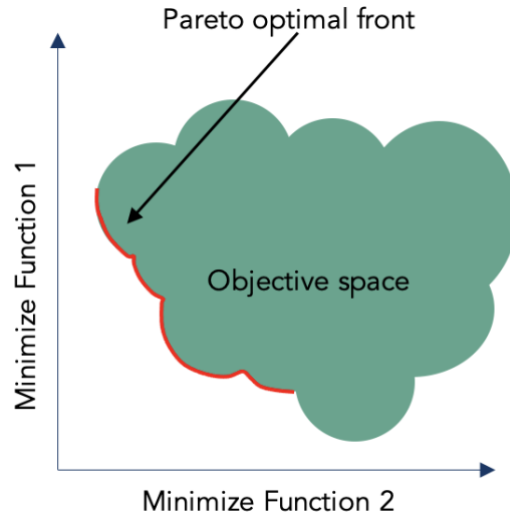


Figure 1-2: Schematic of multi-objective optimisation showing the Pareto optimal front

1.8. Knowledge gaps

Though ANNs are widely applied in the field of SM, there are a number of limitations associated with their use. By default, ANNs are stochastic models, meaning their results cannot be replicated (Zhang *et al.*, 2019). This can create some bias and uncertainty associated with their results. However, the uncertainty associated with replacing original models with a single SM is often overlooked in literature. Therefore, there is a clear motivation to identify the uncertainty stemming from the use of a single SM and incorporate methods that capture and reduce this inherent uncertainty.

Additionally, input selection and ANN architecture are often overlooked when ANNs are used as surrogates for BRC optimisation design. Zhang *et al.* (2019) presented their ANN ensemble surrogates without a formal investigation/description of the processes undertaken to justify the input set selected. Similarly, Raei *et al.* (2019) selected an 80% - 20% split for training and testing for the ANN SM without including a discussion on the uncertainty stemming from arbitrarily selecting the data points for partitioning in this manner. Neither study conducted a formal investigation into increasing/decreasing the number of data points per input for the SM and its effect on model prediction performance. ANN architecture, selection of inputs and the number of training samples per input may have a significant effect on model performance. Selecting too few hidden neurons in the hidden layer could result in a model that is unable to fully form connections that accurately describe a system. Conversely, too many hidden neurons could result in an ANN

that requires significant computational effort (Dawson & Wilby, 2001). Selecting inputs that are not relevant to the output (i.e., do not aid the ANN in predicting the output) could introduce noise to the system and therefore hinder the ANN's ability to accurately predict the output, as well as increase computational demand (Snieder *et al.*, 2020). Finally, if the model is simulating a physical system, having a large number of inputs means additional data monitoring and collection which is expensive and may be unnecessary.

Developing a reliable and efficient SM is essential in an optimisation-simulation framework, as it reduces the computational time of the original model to a fraction of the time (Brunetti *et al.*, 2017; Mohammadiun *et al.*, 2018). Current studies investigating optimising BRC design under climate change impacts, are limited (Tirpak *et al.*, 2021; M. Wang *et al.*, 2019; Yang & Chui, 2018), and usually employ a single SM (Ghodsi *et al.*, 2020). This can amplify the uncertainty of BRC performance under future climate, as there is already a lot of uncertainty stemming from the use of climate change models. Therefore, there is a need to address this additional uncertainty.

1.9. Research Objectives

The following section relates to specific research objectives pertaining to Chapter 3 and Chapter 4

1.9.1. Surrogate model development objectives

Based on the identified knowledge gaps, the following objectives have been identified:

- I. Determine the ability of an ensemble of ANN SMs in replicating SWMM predictions for BRC hydrological and water quality performance.
- II. Determine the impact ANN architecture and input selection has on model performance.

Achieving these objectives can provide a development framework that is easily applicable in other research works, providing a way to unify the process used to design SM's. This can promote more consistency across the field and thus increase the confidence in results obtained from SM's.

1.9.2. BRC design objectives

- I. Quantify the performance difference for BRCs designed using existing design guidelines under current and future climate.

- II. Determine the necessary changes to BRC design recommendations to achieve past performance under future climate.
- III. Optimise BRC design for future climate by focusing on minimizing runoff volume at the outlet, minimizing cost of design and maximizing removal efficiency of Total Suspended Solids (TSS)

Achieving these objectives provides insight into the suitability of current BRC design guideline recommendations under changing climate trends. Additionally, it provides an estimate of the necessary sizing upgrades and their associated costs, which can help policymakers and municipal planners in the sizing of BRC for future resilience as well as the planning of future fiscal budgets.

1.10. Thesis outline

Chapter 2 describes the study area, the rainfall scenarios, and SM inputs and outputs. Chapter 3 describes the research methods, findings and discussion for the SM development research. Chapter 4 describes the research methods, findings and discussion for the BRC design research. Finally, major research findings are presented in the conclusion and future research opportunities are outlined.

Chapter 2 – Bioretention cell modelling

This chapter includes an overview of common BRC modelling software, as well as description of the study site where the BRC is to be located, an overview of the SWMM model, a description of the ANN SM and finally the various rainfall scenarios to be used in this research.

2.1. Modelling of LIDs

A literature review of common physical models used to model LIDs was conducted to select the most appropriate model to evaluate BRC performance in this research. A simulation approach rather than a field approach was selected for this study as computer modelling is a highly effective alternative to field experiments when it comes to studying the performance of LIDs. Computer models allow the user to examine LID behaviour under several storms differing in severity, intensity, duration and profile. Additionally, computer modelling offers the chance to simulate historical long term precipitation to examine LID behaviour under different antecedent soil moisture conditions. Various LIDs can be modelled, which is hugely advantageous since researchers no longer need to construct physical models to perform testing.

The choice of hydrologic/hydraulic model is usually dependent on the problem at hand and project objectives (Ahiablame *et al.*, 2012). An extensive list of possible simulation models exists in literature. For a review of these models the reader is invited to refer to Bosley (2008), Elliott and Trowsdale (2007) and Kaykhosravi *et al.* (2018). Four popular computer models have been chosen to be discussed in the following section due to their prevalence in literature. These models are: SWMM (Eckart *et al.*, 2018), L-THIA-LID (Ahiablame *et al.*, 2013), HEC-HMS (Yazdi & Salehi Neyshabouri, 2014) and SUSTAIN (Lee *et al.*, 2012).

2.1.1. SWMM

The Storm Water Management Model (SWMM) is prominently used in the field of LID modelling as it is a powerful and advanced simulation tool. SWMM is a spatially distributed model conceptualizing catchments with a link-node drainage system (Bosley, 2008; Elliott & Trowsdale,

2007). It can be used to model both single event and continuous simulations, with the ability to accept a wide range of temporal distributions ranging from an annual to a second's resolution.

Subcatchments in SWMM can be divided into pervious and impervious components for runoff generation, with the option of directing runoff from the impervious section directly to the pervious section. A number of subcatchment characteristics that govern the hydrological behaviour of the subcatchment can be defined such as depression storage, overland flow width, impervious area, and slope. Overland flow width is defined as the longest perpendicular distance to the direction of flow. The Manning's equation is used for overland flow routing. When runoff is generated, either Horton's method, Green-Ampt or Curve Number (CN) method may be used for infiltration into the catchment. Runoff is then transported through the subcatchment through a series of links known as conduits representing pipes. SWMM is capable of steady state, kinematic and dynamic flow routing.

Pre-defined LID units are integrated into the SWMM LID module. SWMM is capable of modelling the following LIDs: BRCs, rain gardens, porous pavements, swales, green roofs, rain barrels, and infiltration trenches. Users can specify the type of LID they wish to work with and can enter the necessary design parameters to simulate the LID they require. Ponding in SWMM is possible by specifying a maximum berm height in the LID design editor. Layering of soil types in SWMM is not permitted, with only one layer specified for the growing media. LIDs can be modelled as a distinct subcatchment in the study area or within a parent subcatchment. Flow from the impervious surfaces can be directly routed to the LID surface (Kaykhosravi *et al.*, 2018). SWMM is a moderately complex software that is better suited for advanced modelling phases, preliminary and detailed design.

2.1.2. L-THIA-LID

Long Term Hydrologic Impact Assessment Model – Low Impact Development (L-THIA-LID) is a simplified conceptual model suitable for conceptual design. Runoff calculations are performed using the Natural Resources Conservation Service (NRCS) CN method provided user defined land use, soils data and long term rainfall time series. L-THIA-LID is a refined version of the original L-THIA model using CN number and Event Mean Concentration (EMC) to model annual runoff volume and NPS pollutants. L-THIA-LID supports the following predefined LIDs: BRCs/rain gardens, porous pavements, green roof and rain barrel/cistern. Runoff and NPS pollution can be simulated with LID for a single lot to a watershed scale. Unlike SWMM, LIDs cannot be modelled

as a separate subcatchment and are instead modelled explicitly within the watershed (Kaykhosravi *et al.*, 2018).

2.1.3. HEC-HMS

Hydrologic Engineering Centre – Hydrologic Modelling System (HEC-HMS) was originally developed to model the hydrological behaviour of a dendritic watershed, and has not been extensively used for LID modelling purposes. HEC-HMS classified hydrologic and hydraulic elements into three categories: basin, meteorological, and time series component. HEC-HMS is sufficiently suited for advanced modelling, providing detailed stormwater results such as peak flow, runoff amount, flow depth and flow volume within conduits. It does not have pre-defined LIDs like the previous two models, but LIDs can be implicitly defined. This is done through modelling the LID as a subcatchment and assigning it the required LID properties. It does not allow for explicit surface ponding in LIDs, however, it can be implicitly modelled by assigning ponding as a loss method. The soil is considered to have two layers which allows for ponding implicitly. A limitation of this method is its underlying assumption of a linear infiltration versus time relationship, which is a simplification of the nonlinear process. This assumption is too simplified for LID applications and can thus be a major limitation. Modelling of rainfall data is possible for fine (i.e. in minutes) resolution. Infiltration modelling is vital for accurate LID modelling, and several infiltration methods are possible in HEC-HMS such as the Green-Ampt method and CN method. Capable of steady state and kinematic modelling, it cannot be used for dynamic modelling, (i.e. there can be no backwater effect modelling) (Kaykhosravi *et al.*, 2018).

2.1.4. SUSTAIN

Created in 2003 by the EPA, the System for Urban Stormwater Treatment and Analysis Integration (SUSTAIN) was developed to provide a decision support system to aid in selecting and placing LIDs/Best Management Practices (BMPs) at strategic locations in urbanized watersheds based on cost and effectiveness. SUSTAIN is different from the above mentioned models in that it integrates the physical model, the optimisation model and the BMP module. Given that SWMM was created by the US EPA it comes as no surprise that SUSTAIN's underlying physical processes for its land and conveyance module are the same as those of SWMM 5 (Lee *et al.*, 2012).

Two methods exist to model runoff and pollutant transport in the land simulation module. The first is through the internal simulation option which generates a hydrograph and a pollutograph using

the same algorithms as those of SWMM. Alternatively, the user can import a hydrograph and pollutograph from an external pre-calibrated watershed model. Hydrological processes such as snowmelt, evapotranspiration, infiltration and pollutant build up and wash off are all identical to those in SWMM5. Watershed surfaces are treated as planar non-linear reservoirs with user-defined depression storage. Contaminant calculation includes sediment loading, which is not simulated in SWMM. Sediment is calculated by using algorithms adapted from Hydrological Simulation Program Fortran (HPSF) and non-sediment pollutant is calculated as normal from SWMM's build-up and wash off coefficients for each land use. Flow and water quality routing is modelled through conduits. Conduits represent pipes or 1-D open channels that transport flow from one junction to another. Flow routing is simulated using the kinematic wave routing option (Lee *et al.*, 2012).

A pre-defined library of BMPs is available and the user can adjust the physical parameters of the design to suit their needs. The BMP module can simulate both hydrological and pollutant behaviour of BMPs. The following hydrological processes can be modelled: Infiltration, storage, evapotranspiration, underdrain infiltration and pollutant removal (Lee *et al.*, 2012).

A quick comparison of the four mentioned models is provided in Table 2-1. From Table 2-1 it is evident that SWMM is the most comprehensive and advanced model available for use to model LIDs. This is demonstrated in its ability to accept both coarse and fine temporal resolutions, as well as its variety of infiltration methods, its explicit existing LID directory and its hydraulic routing capabilities which are not all possible in the other compared modelling software's. This, coupled with its popularity in the field of LID modelling makes it ideal to be used in this thesis, as a SM developed for SWMM will have the most widespread benefits.

Table 2-1: A comparison of the 4 physical-models for LID modelling

	SWMM	L-THIA-LID	HEC-HMS	SUSTAIN
Temporal resolution				
Coarse (ex. Annual)		Lumped – Annual runoff volume		
Fine (ex. Seconds)			✓	
Both	✓			✓
Infiltration Method				
CN	✓	✓	✓	✓
Green-Ampt	✓	✗	✓	✓
Horton's	✓	✗	✗	✓
LID modelling				
Implicit or Explicit	Explicit – predefined library	Explicit – predefined library	Implicit – Defined as new subcatchment	Explicit – predefined library
Surface ponding	✓	✗	✗	✓
Underdrain	✓	✗	✗	✓
Multiple soil layers	✗	✗	✓	✗
Hydraulic routing				
Kinematic	✓	✗	✗	✓
Dynamic	✓	✗	✓	✗
Steady State	✓	✓	✓	✗

2.2. Study site

The proposed BRC to be built by the City of Toronto is to be located close to York University, at the intersection of Evelyn Wiggins and Murray Ross Parkway, shown in Figure 2-1.

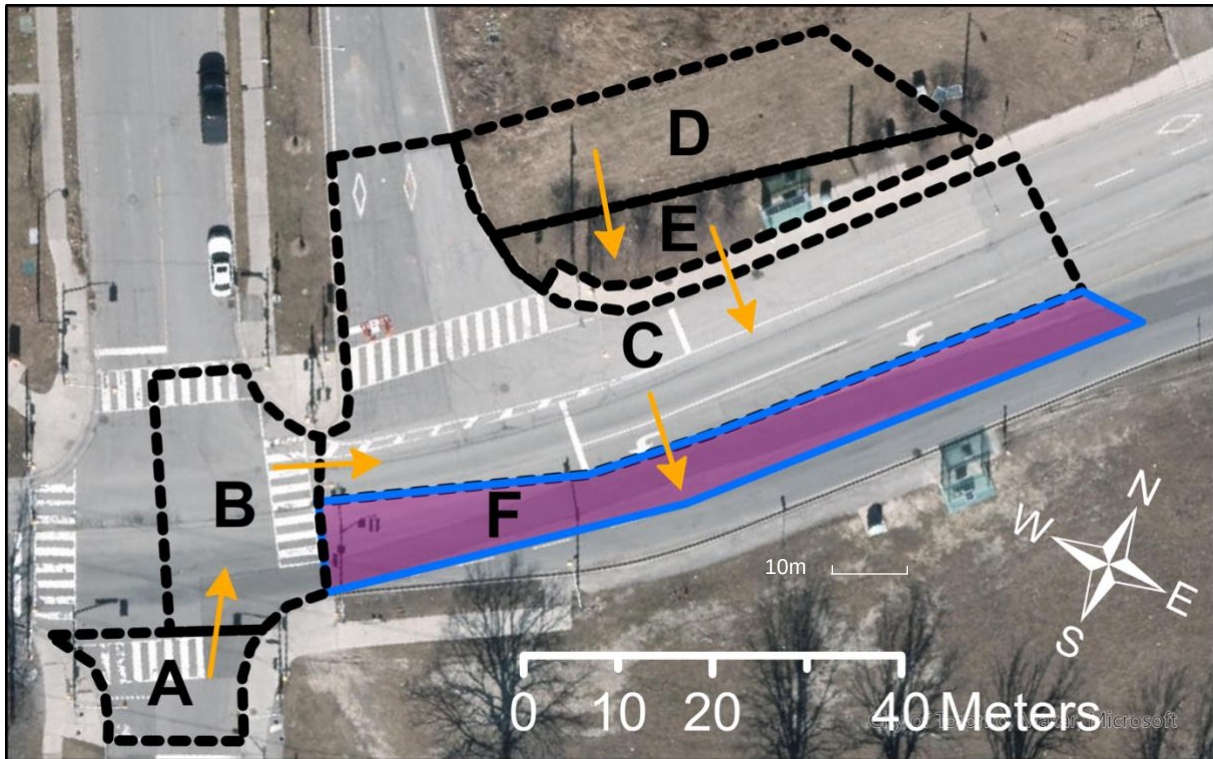


Figure 2-1: Proposed study site where the BRC (shown by the purple rectangle) receives runoff from the impervious areas delineated by the dashed lines

The intersection is located between York University's Village, a residential area, as well being in close proximity with the Keele and Murray Ross Parkway intersection which receives a lot of public transport traffic as it is nearby the Finch West subway station. The area can therefore be classified to be mixed landuse. Figure 2-1 shows the impervious subcatchments draining into the BRC, with the black arrows denoting flow direction. The BRC is delineated by the purple rectangle and the letter F. To develop the hydrological model in SWMM, the BRC subcatchment characteristics were established using ESRI's ArcGIS Pro. A 1 m resolution Digital Terrain Model (DTM) was obtained from the Government of Canada (Government of Canada, 2019) and used to determine the slope, flow direction and flow accumulation. The BRC watershed was then delineated to accurately calculate and map out the area draining into the BRC. The characteristics of the areas draining into the BRC are given in Table 2-2. This was done based on the soil characteristics of the surrounding area. The native soil permeability in that area is less than 15 mm/hr, found from the reference maps in Appendix E provided by the City of Toronto (City of Toronto, 2017). The soil type was assumed to be a mixture of silty loam and loam, and the field capacity and wilting point of the site were set to 0.25 and 0.13 respectively.

Table 2-2: Subcatchment names and characteristics of impervious areas draining into the BRC

Subcatchment ID	Subcatchment	Width (m)	Area (m ²)	Slope (%)
A	Collection Point 1	13	229	0.5
B	Collection Point 2	15	183	1.5
C	Roadway	83	1000	0.5
D	Grasshill	47	561	0.5
E	Grass Slope	47	121	10
F	BRC	3	170	2

To predict water quality improvement, TSS was selected as the target pollutant. TSS is frequently selected as the target pollutant, as many LIDs, BRCs included, are commonly designed for suspended solid control (Li & Davis, 2008). To simulate TSS reduction by a BRC in SWMM, a mixed landuse was defined. Three different TSS EMC values were defined as the washoff coefficients, further discussed in Chapter 3. TSS was the only parameter selected in this study as it is one of the most consistent indicators of water quality improvement used in BRC design studies.

Calibration and validation of the SWMM model was not possible as the BRC has not been built yet. However, model parameters were selected to represent the characteristics of the surrounding area as closely as possible. It is important to note however, calibration and validation of the model does not affect the objectives of this research since the interest lies in both investigating the SM's ability to replicate any physical process as well as a comparison between BRC future and baseline performance for runoff retention as well as TSS removal.

2.2.1. BRC design configuration

The current recommended BRC design guidelines are discussed below and summarized in Figure 2-2. To maintain BRC performance and avoid early failure or frequent overflows the recommended maximum upstream area draining into a BRC should not exceed 0.8 ha (Credit Valley Conservation (CVC), 2010). The upstream impervious area to BRC area ratio is also given as a range from 15:1 to 5:1. Underdrains should generally be installed if the native soil has a low

permeability (less than 15 mm/hr) (City of Toronto, 2017). Ponding depth should be 150 – 200 mm in height, the growing media depth should be 1 m deep although if space is limited a minimum of 500 mm depth is sufficient for pollutant removal, the gravel storage layer should be a minimum of 300 mm deep and if an underdrain is placed it should be placed at least 100 mm above the bottom of the gravel storage layer (CVC, 2010).

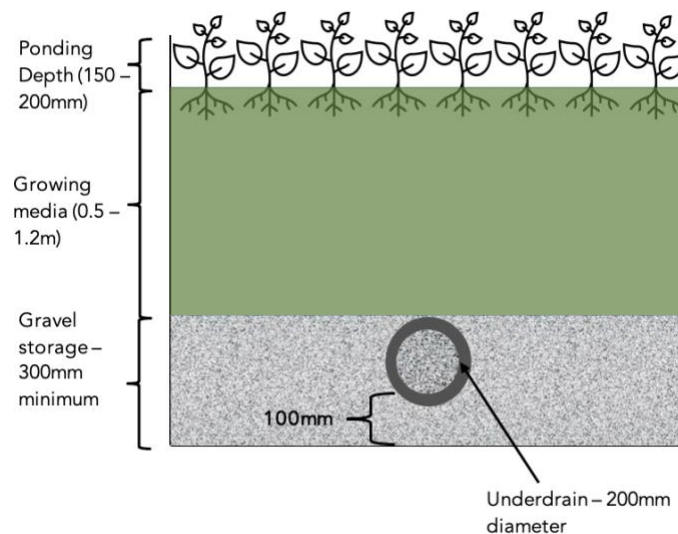


Figure 2-2: Schematic showing the recommended design values for the main parameters in a BRC (growing media depth, gravel storage depth and underdrain)

For the study site, the total upstream area draining into the BRC is approximately 2100 m². Therefore, according to the design guidelines the minimum and maximum BRC area is 150 m² and 400 m² respectively. An underdrain of 200 mm was also added to the design, as per design guidelines. The minimum and maximum design limits used in Chapters 3 and 4 are outlined below in Table 2-3.

Table 2-3: Minimum and maximum BRC design parameters

Design parameter	Minimum	Maximum
Area (m ²)	150	400
Growing media depth (m)	0.5	1
Gravel storage depth (m)	0.3	0.4

2.3. Artificial Neural Networks – Surrogate Model

The SM to be developed in this research will be used to model BRC performance both in runoff retention and TSS removal. An ANN was selected as the basis for the SMs as it has been shown that ANNs are the most commonly used type of SM.

A feedforward multi-layer perceptron ANN (shown in Figure 2-3) is used in this research to create an SM to the SWMM BRC model. ANNs commonly consist of multiple layers, starting with the input layer, followed by the hidden layer and ending with the output layer. In a feedforward ANN information is carried forward, where it is moved from the input to the hidden nodes and finally into the output layer. Each connection from the input and output nodes respectively to the hidden node has an associated multiplicative and additive weight, w_k , and bias, b_k . These weights and biases are iteratively adjusted at each training step, using the Levenberg-Marquardt backpropagation method, to minimize the difference between the predicted and observed model output. This is the most common training algorithm across ANNs, known as error backpropagation (Snieder *et al.*, 2020). A generalized method for selecting the total number of hidden nodes in the hidden layer does not exist, instead it is usually determined based on a trial-and-error approach specific to the complexity of the problem, this is further discussed in Chapter 3.

The Levenberg-Marquardt backpropagation method is used as the training algorithm to minimise the Mean Square Error (MSE). Typically, the input dataset is subdivided into three subsets: training, validation and testing. The training subset is used to adjust the network weights and biases, the validation subset is used as an early stopping technique to avoid overfitting, and the testing subset is used to quantify model performance. The training, validation and testing datasplit was 60% - 20% - 20% respectively to ensure a higher fraction of data is available for testing, since this is preferred in literature (Khan *et al.*, 2018).

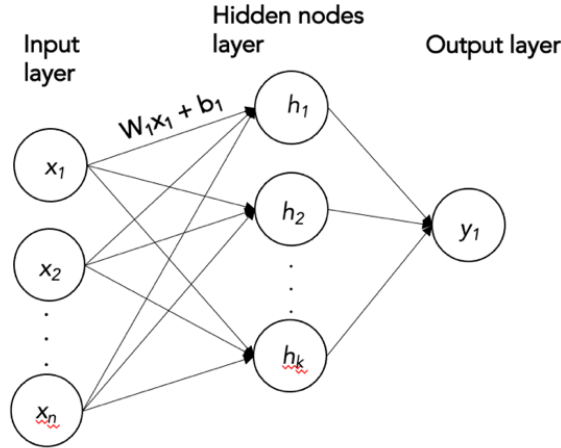


Figure 2-3: A schematic of a feedforward ANN showing the neurons, weights, biases and layers

To address the uncertainty associated with using a single ANN as a SM, an ensemble of 100 ANNs was created. The inputs used in these ensembles is further discussed in Chapters 3 and 4. The bootstrapping method was used to create the ensembles, generating different realizations by randomly resampling the original input dataset with replacement, where each new training set is used to train and generate a new ANN. Each model is trained, validated and tested on a different subset of the original dataset, therefore each model is also distinctly different in terms of weights and biases and thus its predictive capabilities. Obtaining predictions from a larger set of distinct models aids in quantifying model uncertainty (Parasuraman & Elshorbagy, 2008; Zhang *et al.*, 2019).

2.4. Rainfall scenarios

To evaluate BRC performance under different rainfall conditions, a series of design storms are entered into SWMM. Different design storm distributions, varying in the shape of the hydrograph, exist, such as the Chicago, the Soil Conservation Services (SCS) and the Atmospheric Environmental Services – Southern Ontario (AES-SO) distribution. Using the AES-SO storm distribution is advantageous over other distributions as it was created based on Canadian data (Marsalek & Watt, 1984). A total of 18 design rainfall events following the AES-SO distributions were used in this research. This facilitated the evaluation of BRC performance under different rainfall events. Six storms of return periods 2, 5, 10, 25, 50 and 100 years were used, for 3 different rainfall durations of 6, 12 and 24 hours each. This produces a total of 18 storms. Total rainfall depths and intensities were obtained for the Toronto North York station (ID: 615S001) using the

Intensity-Duration-Frequency developed by Schardong *et al.* (2020). The 100-year 12-hour AES-SO distribution is shown in Figure 2-4 as an example.

Chapter 3 explores the reliability of using SMS in place of its physical-based model counterpart, SWMM, to evaluate BRC hydrological and environmental performance. Therefore, the SMS performance in simulating BRCs under a number of different rainfall scenarios is compared to that of SWMM's.

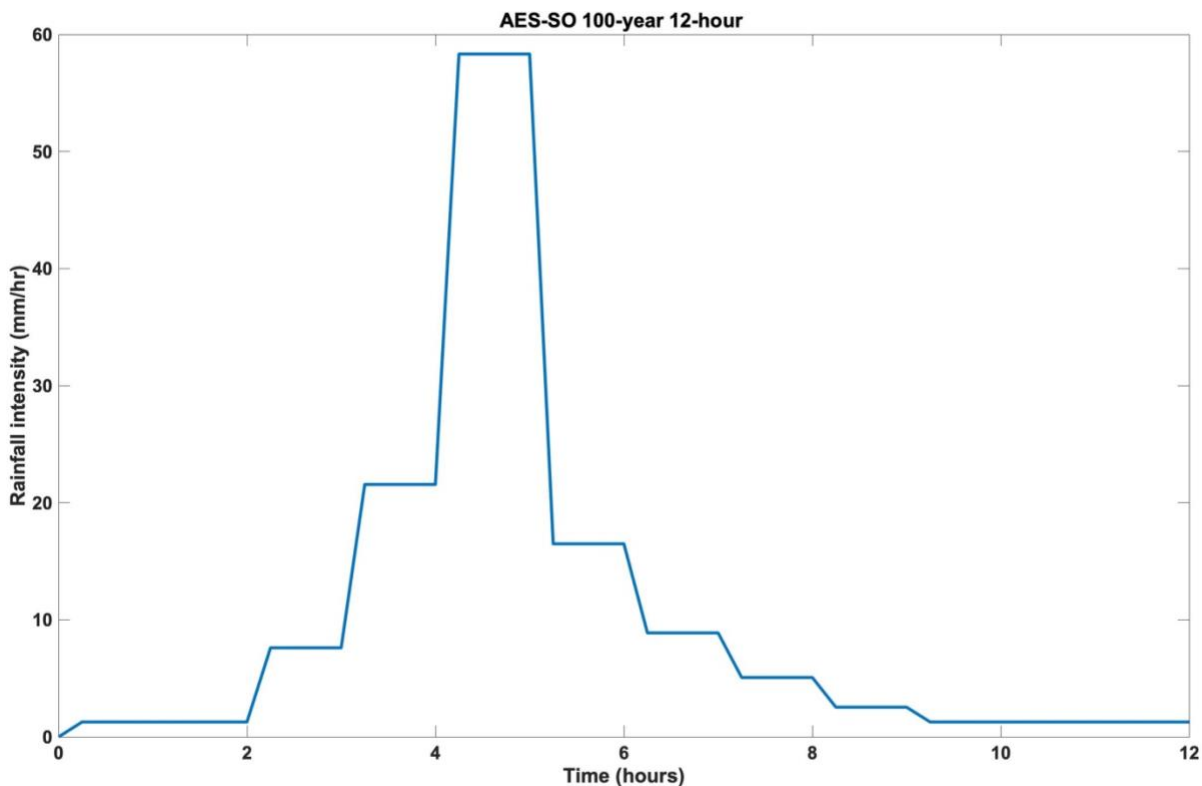


Figure 2-4: AES-SO distribution for the 100-year 12-hour storm

Chapter 4 examines BRCs hydrological and environmental performance under current and future climate conditions for the 18 different rainfall storms. This is important as it highlights the difference in performance under different rainfall events. Additionally, this approach can help identify at which future climate storm does BRC performance begin to deteriorate relative to the historical performance.

Of the most common rainfall-runoff models, SWMM was chosen to be used in this thesis work due to its many capabilities and its prevalence in the rainfall-runoff modelling field. Developing a

SM for SWMM facilitates its widespread use as it can be applied across these studies. The SWMM model for this research was then developed by first establishing the physical characteristics of the BRC subcatchment using GIS and applying those findings to develop the hydrological model. A feed-forward ensemble of ANN's were chosen for this thesis work. Two sets of ANN's were developed, one to predict runoff rate at the outlet and the other to predict TSS at the outlet. Finally, the AES-SO distribution was used and applied to obtain 18 different rainfall scenarios. These rainfall scenarios were used to design the BRCs under different rainfall conditions.

Chapter 3 – An investigation of surrogate model development for bioretention cell design in place of SWMM

3.1. Introduction

Rapid urbanisation alters the natural hydrological cycle as stormwater runoff no longer infiltrates the soil layer. This, coupled with the increased extreme precipitation events due to climate change, contribute to the heightened risk of localized flooding in urban cities (Brunetti *et al.*, 2017; Zhang *et al.*, 2019). Low Impact Development (LID) units have become an increasingly popular solution to restore pre-development hydrological conditions by promoting infiltration and evapotranspiration (Ahiablame *et al.*, 2012; Eckart *et al.*, 2017). Common examples of LIDs include; Bioretention cells (BRCs), infiltration trenches, grass swales, green roofs and permeable pavements (Ahiablame *et al.*, 2012).

BRCs are one of the most common types of LIDs installed in urban watersheds (Johnson & Hunt, 2019; Kratky *et al.*, 2017), and the most commonly researched LID practice (Sohn *et al.*, 2019). BRCs have proven capabilities in reducing runoff (Jia *et al.*, 2015), reducing peak flows (Winston *et al.*, 2016), reducing pollutant loading in effluent (Khan *et al.* 2012b; LeFevre *et al.*, 2015; Singh *et al.*, 2019) as well also improving air quality, providing shade, reducing noise pollution levels and improving landscape aesthetics (Muerdter *et al.*, 2018; Roy-Poirier *et al.*, 2010; Wang *et al.*, 2013).

3.1.1. Physics – based models

Predicting BRC performance through simulation modeling provides valuable information, that may not be necessarily captured through time limited monitoring periods, as modelling can be applied across a range of time frames, location and conditions (Ahiablame *et al.*, 2012; Brunetti *et al.*, 2017). A number of modelling software's exist to assess the performance of BRC design in stormwater management. Boancă *et al.* (2018) used the RECARGA model to evaluate the effectiveness of BRCs in 4 different types of urban sites; commercial, industrial, high-density residential and low-density residential for an urban catchment in Romania. He and Davis (2011) developed a 2D mechanistic model using COMSOL to evaluate the hydrologic performance (i.e.,

reduction of outflow volume) of a BRC through its underdrain outflow. Additionally, they examined bottom and side exfiltration rates into the surrounding soils, as well as the effect of varying design parameters such as the ratio of drainage area to BRC surface area on hydrologic performance (He & Davis, 2011). To evaluate models catering to long term simulation, Brown *et al.* (2013) used DRAINMOD and compared its performance against observed field data obtained from monitoring two BRC for a period of 24 months. Physical-based models are also often used in optimisation purposes. Coupling a hydrological physical-based simulation model with an optimisation algorithm is a widely adopted practice (Zhang *et al.* 2019), with the Stormwater Management Model (SWMM) being the most commonly used hydrological model (Eckart *et al.*, 2018; Ghodsi *et al.*, 2020; Mohammadiun *et al.*, 2018; Raei *et al.*, 2019; Tirpak *et al.*, 2021; Zhang *et al.*, 2019). However, physically-based hydrological simulations often entail solving complex, non-linear differential equations requiring a high computational demand. This, coupled with repeatedly running the hydrological model to evaluate different designs in a simulation-optimisation framework significantly increases the computational burden, presenting a major limitation to the feasibility of a simulation-optimisation framework. A possible remedy is the use of surrogate models.

3.1.2. Surrogate Models

Surrogate Models (SMs) are numerical approximators of an original model, e.g. SWMM, favoured for their computational efficiency and high accuracy. SMs have been widely applied in the field of hydrological modelling and optimisation. Lu *et al.* (2019) used quadratic regression models as SMs to SWMM in their optimisation of storage ponds for urban drainage systems. Raei *et al.* (2019) used an Artificial Neural Network (ANN) in place of SWMM in their multi-objective optimisation framework for LID design. Three SMs were developed in their work to simulate runoff volume, TSS and BOD contamination in the effluent. An R^2 value of 0.99 for all three models for the validation dataset was achieved, indicating the high predictive accuracy and reliability of the SM. Mohammadiun *et al.* (2018) used an ANN as an SM to SWMM in their optimisation of Urban Stormwater Drainage Systems (USDSSs). Results of the 50th Pareto front was evaluated using both the SWMM and ANN model to compare performance. Results indicated good agreement between both models with only a relative error of 25%. Additionally, significant reductions in computational time were achieved, where 50 generations resulted in a run time of 85 days using SWMM, and only a fraction of the time using the ANN. Therefore, the need and effectiveness of an SM in place of SWMM has been demonstrated in research.

Though ANN - SMs are frequently used in water resources research, a number of limitations exist. ANNs are inherently stochastic, meaning their results cannot be replicated twice (Zhang *et al.*, 2019). Therefore, using a singular ANN in these surrogate modeling studies introduces bias, uncertainty and may affect the reliability of the solutions. This uncertainty is rarely discussed, and to the best of the author's knowledge has only been discussed in Zhang *et al.* (2019). They used an ensemble of SMs coupled with the optimisation model to quantify this uncertainty. Additionally, the development of the SM is seldom discussed, rather the singular SM is presented without discussion of the development process or alternative methods undertaken to ensure the best SM is produced.

Model performance is dictated by the choice of input variables. If the selected set of inputs is not relevant (i.e., comprised of good predictors) or too few variables are selected then the input set is not sufficiently informative and the outcome is a poorly performing model (May *et al.*, 2011). Similarly, ANN accuracy is highly dependent on the quantity of training samples per input provided at the onset of model evaluation. Too few training samples may hinder the ANN's ability to infer the system's underlying relationships, thereby causing the ANN to under or overpredict. Conversely, providing too many training samples may result in overfitting, negatively impacting the ANN's ability to generalize and accurately predict results for input variables not yet encountered (Dawson & Wilby, 2001). Finally, the sampling methodology used for data splitting in ANN's is of great importance and is yet often ignored or understated. If the training, validation and testing dataset of the ANN are not equally representative of the problem domain, the uncertainty in SM performance due to sampling variance may be greater than other sources such as model architecture (May *et al.*, 2010). Using a sampling method such as stratified sampling for datasplitting can ensure equal representation of the input data in the training, validation and testing dataset. This can eliminate some of the uncertainty stemming from a poor choice of datasplitting. Uncertainties associated with the impact the number of training samples, the type of datasplit sampling method and input selection have on model performance remain unanswered.

This chapter explores the development process of an ANN as a surrogate to SWMM with an emphasis on the impacts of stratified sampling, the number of training samples and the impacts input selection have on model performance. The physical model SWMM was first used to simulate a BRC located at York University, in Toronto, Canada. Subsequently, a set of distinct ANN SMs ensembles were developed to evaluate BRC design and predict runoff flowrate and mass of TSS

at the outlet. The predictive accuracy of the ANNs constructed in this way is compared to that of SWMM. If the SMs are deemed a reliable substitute to SWMM, the ANN can then be used in place of SWMM in optimisation-simulation applications. The investigation of the impacts ANN architecture and the development process have on SM performance is the first of its kind.

3.1.3. Objectives

The main objective of this research is to develop a robust, efficient and accurate SM that may be used in place of SWMM to simulate BRC performance in runoff retention and contaminant removal. The difference in model performance between the SM and SWMM is quantified, to determine the suitability of SMs to be used in place of SWMM. An ensemble of ANN's are proposed in this research as the basis of the SM, since they are widely used in rainfall-runoff simulations and are considered universal approximators capable of simulating complex non-linear relationships (Dawson & Wilby, 2001), making them an ideal candidate for SMs.

Secondly, the minimum number of training samples necessary to maintain high model performance is determined. To investigate this, the number of training samples is varied from 50 to 200 per input. Determining the minimum number of samples required, beyond which including additional training samples would not significantly increase model performance, can help reduce computational training effort.

Next, the difference in model performance stemming from using random and stratified sampling for data splitting is quantified. This is done to investigate the effects an unequal representation of the problem domain has on model performance. Additionally, the impact of input selection on model performance is quantified. Originally, only BRC design parameters and rainfall depth will be supplied as inputs. This will serve as the baseline model performance, to quantify the improvement in model performance as other inputs are added in. Finally, the performance difference between an event-based and universal SM is quantified. An event-based SM is trained on only one rainfall event, whereas the universal SM is trained on all rainfall events. This is done to determine if a superior model exists, thus providing a comprehensive overview of SM architecture. A thorough investigation of the various components making up ANN development is useful as it serves as a baseline framework for future ANN SM development. Having a common framework promotes unity in ANN SM development, increasing the reliability of SM results.

The research is structured in the following manner: firstly, the number of training samples per input is varied to determine the minimum number of training samples necessary to achieve a high level of accuracy. Next, the impact of random versus stratified datasplit sampling on model performance is quantified. Finally, the impact input selection has on the model performance is quantified.

3.2. Methods

All the calculations conducted in this study were performed using MATLAB 2020a. The EPA SWMM model used was version 5.1.015.

3.2.1. Artificial neural networks (ANNs)

For this research two ANNs were developed, one to predict runoff flowrate at the outlet (Hydrological ANN) and one to predict the mass of TSS remaining in the effluent (Environmental ANN). Two ANN models with a single output were developed, since models with singular outputs have better prediction performance than a model with multiple outputs (Snieder *et al.*, 2020). Both ANNs are made up of 3 layers with 4 inputs, 10 hidden nodes and one output. BRC area, growing media depth, gravel storage depth and rainfall depth are used as the input variables.

3.2.2. Surrogate model synthesis

The primary use of this SM is to facilitate BRC design evaluations without having to use SWMM under different rainfall scenarios. Therefore, the inputs to the SM consisted of the BRC design parameters (area, growing media depth and gravel storage depth), as well as the rainfall depth pertaining to the different rainfall scenarios. These inputs were selected as they are assumed to be the only parameters changed during BRC design iterations, since the surrounding characteristics remain unchanged.

SMs are trained on a known set of inputs, which are obtained from first running the original model, in this case, SWMM. Therefore, the original datasets used to train the ensemble of ANNs were first obtained from SWMM. Latin Hypercube Sampling (LHS) was used to ensure a well-distributed set of initial BRC designs consisting of different areas, growing media depths and gravel storage depths is selected. A poorly distributed set of initial BRC designs may result in a poorly performing ANN, as the ANN will not be trained on a representative distribution of the problem domain, and may rather be trained on a small subset. This will impede its generalisability

resulting in a poorly performing ANN. The runoff retention and pollutant removal performance of the BRC was evaluated in SWMM under the 18 different rainfall scenarios (2,5,10,25,50 and 100-year storms with durations of 6,12 and 24 hours each). These inputs (BRC area, growing media depth, gravel storage depth and rainfall depth) and associated outputs (runoff rate and mass of TSS remaining in effluent) were then used to train the ANN ensemble.

3.2.2.1. *Event based and universal models*

An SM for each of the 18 rainfall events can be created, known as an event-based surrogate, whereas a single SM for all rainfall events can be used, known as universal surrogate. The distinction between the two lies in their predictive capabilities for different rainfall scenarios. The event-based surrogate can be used to evaluate BRC performance under a single specific rainfall scenario. Alternatively, the universal surrogate can be used to evaluate BRC performance under multiple different rainfall scenarios, with the choice of rainfall scenario easily selected. Both have their benefits. The event-based scenario may be advantageous when only one design storm is of interest, thereby saving computational time and effort in the initial training of the SM as only designs under that specific rainfall event need to be evaluated. Conversely, the universal based surrogate is advantageous as it allows for easy alternating between different rainfall scenarios, suitable for BRC design under many different storms. The predictive performance of both is tested to determine if there is a superior model, to better recommend which should be used in the future if users have no preference.

3.2.3. ANN model configurations

The following sections include descriptions of the different experiments conducted on ANN configurations.

3.2.3.1. *Training Samples*

One of the key components in the development of SMs is the initial selection of sampling points. A well-distributed initial design set ensures the entire design space is explored (Brunetti *et al.*, 2017; Garud *et al.*, 2017). Various sampling methods exist such as Random Sampling, Stratified Sampling and Latin Hypercube Sampling (McKay *et al.*, 1979). LHS was used as the sampling method for initial data points as it is widely applied in existing literature (Brunetti *et al.*, 2017; Davis *et al.*, 2018; Lu *et al.*, 2019; Yazdi & Salehi Neyshabouri, 2014; Zhang *et al.*, 2019). LHS is an extension of stratified sampling first proposed by McKay, Beckman, and Conover (1979). The range of each input X_i , $i = 1, \dots, N$, is divided into N strata of equal marginal probability $1/N$, with

only one sample taken from each stratum (McKay *et al.*, 1979), this is demonstrated in Figure 3-1. The axes in Figure 3-1 represent a 2-dimensional input set, where each cross represents a selection of values from each input. The crosses are selected such that only one set of inputs is selected from each stratum per input. This example is only one possible variation of selecting a set of inputs.

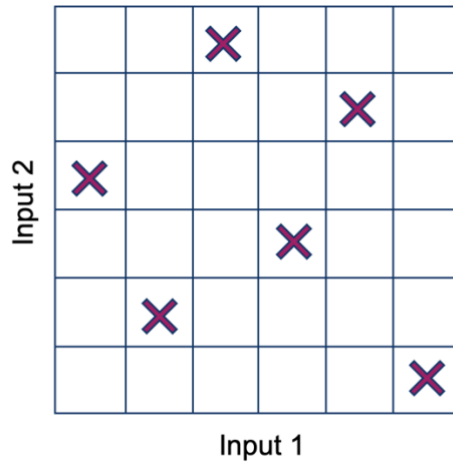


Figure 3-1: Schematic showing one possible variation of LHS

In this study, the number of training samples refers to the number of different BRC configurations under each of the 18 rainfall design storms, used to train, validate and test the SMs. Three values of training samples were used in this research, 50, 100 and 200 samples per rainfall event. The 50 training samples refer to 50 different configurations of BRC design parameters (area, growing media depth and gravel storage depth), ranging from the minimum to maximum value expressed in design guidelines, found in Chapter 2. The rainfall depth under the 50 training sample scenario is also repeated 50 times. Therefore, given 18 design storms, there are 900 total training samples for this scenario. The inputs supplied to the ANN consisted of 900 rows and 4 columns, with each row acting as a distinct design configuration of the BRC under a specific rainfall depth. Each of the distinct design configurations was obtained using LHS. A matrix of 900 rows and 3 columns was created using MATLAB's *lhsdesign* function, with the columns representing area, growing media depth and gravel storage depth respectively. The initial range of values in this matrix were from 0 – 1, therefore, to obtain the desired area values, the first column was multiplied by 250 and 150 was added to each row value. This was done to ensure that the minimum area was 150 m² and the maximum area was 400 m² in accordance with the minimum and maximum values

expressed in design guidelines. A similar approach was carried out for the growing media and gravel storage depth columns. For the 100 training samples scenario, a total of 1800 training samples were used to train, validate and test the SMs. Similarly, for the 200 training samples scenarios, a total of 3600 training samples were used.

3.2.3.2. *Random vs Stratified sampling for datasplitting*

Data splitting is the process of subdividing the initial dataset into three groups to be used for training, validation and testing. There are several ways to do this, with the most common being random sampling. As the name implies, random sampling entails selecting data with uniform probability. This approach however may suffer from bias, especially if the data is non-uniformly sampled.

Stratified sampling partitions the original dataset into X strata with similar characteristics, and data is sampled from within each stratum, this process is shown in Figure 3-2. The top row of Figure 3-2 represents a poorly distributed training set, with an unequal representation of input data points (represented by the coloured circles). Using stratified sampling, the inputs are sorted into the respective strata (this being colours in the figure), and finally the input data is resampled from each stratum using stratified sampling to ensure an equal representation of input data in the training set. The same process can be carried out for the validation and calibration. This partitioning of the data ensures a well-distributed and equally representative dataset is used for training, validation and testing (May *et al.*, 2010).

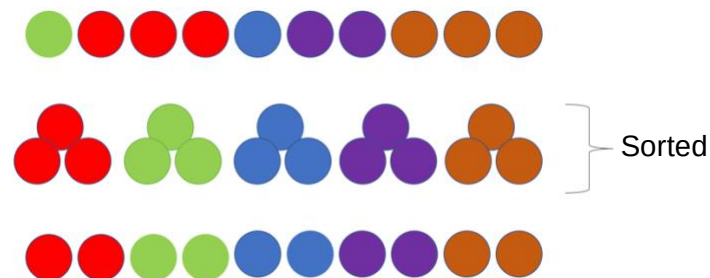


Figure 3-2: Schematic explaining stratified sampling – from top to bottom the schematic shows the sorting of input data into respective strata resulting in an equal representation of input data

The training sample was stratified by events or by low and high flow storms. Of the 18 design storms used in this research, the 50 and 100-year design storms were deemed to be high flow

storms, i.e., generating high flows of runoff at the outlet, and the remaining storms deemed low flows. Based on this, the flows were subdivided into high and low flow categories. Of these categories, 60% of the data were selected. Out of the 600 low flows, 360 flows were selected and of the 300 high flows 180 flows were selected. These selected flows were used for the training. Similarly, 20% of the high flows and low flows were selected for the validation and testing datasets. This was done to ensure equal representations of both high and low flows were present in the training.

To further this investigation, stratified sampling was also used to ensure 60% of each rainfall event was included in the training set, with the remaining 40% split equally in the validation and testing. For the LHS 50 model this meant 30 design configurations from each rainfall event were used in training, for a total of 540 samples. The validation and testing datasets used 180 samples each. This was done separately from the flow-based stratification, to investigate whether one stratification strategy is superior to another.

3.2.3.3. *Input Selection*

The four baseline inputs are: BRC area, growing media depth, gravel storage depth and rainfall depth of the design storm. Given the importance of input selection on model performance, it is necessary to investigate whether an improvement in model performance occurs from the addition of other input variables, or whether these base inputs are sufficient. These 4 inputs were chosen as they are the only ones altered during BRC design evaluations. Of course, other BRC design parameters can be varied such as diameter of underdrain, and ponding depth to name a few, however the three chosen design parameters were deemed sufficient enough to be representative of overall BRC design.

The additional inputs include average rainfall intensity and peak rainfall intensity. These were selected as it was assumed that the only variable affecting BRC performance other than its physical design is the amount of rainfall acting on the BRC. All other external factors (such as slope, soil hydraulic conductivity etc) were assumed to remain the same.

The impact of input selection was first tested on the Hydrological ANN predicting runoff rate at the outlet. Once the input set providing the best performance was identified, the Event Mean Concentration (EMC) was added as an input and the second Environmental ANN predicting mass of TSS remaining in effluent was developed. The addition of EMC as an input variable to the

network was also quantified to ensure it didn't adversely impact model performance. EMC is a necessary input to the water quality ANN as it dictates the concentration of pollutants per rainfall event, directly relating to the BRCs ability to remove pollutants from the system. Three EMC values of 50 mg/L, 130 mg/L and 300 mg/L were selected as inputs representing the 25th, 50th and 75th percentile concentration of TSS per rainfall input respectively (Behera *et al.*, 2000; City of Toronto, 2006).

3.2.4. Performance Metrics

To measure model performance, a number of performance metrics are used in this study: the Nash-Sutcliffe Efficiency (NSE), Continuous Ranked Probability Score (CRPS), Rank Histogram (RH), and Percent Capture (PC). The NSE and CRPS are defined as follows:

$$NSE = 1 - \left[\frac{\sum_{i=1}^n (obs_i - pred_i)^2}{\sum_{i=1}^n (obs_i - \overline{obs})^2} \right] \quad (3-1)$$

$$CRPS = \int_{-\infty}^{\infty} [F(pred) - H\{pred \geq obs\}]^2 dx \quad (3-2)$$

where obs_i is the runoff value observed for design configuration i , $pred_i$ is the predicted runoff value for design configuration i , $\overline{obs}$ is the mean observed value, n is the number of design configurations and $F(pred)$ is the predictive cumulative distribution function. The function $H\{pred \geq obs\}$ is the Heaviside function which equals 1 for predicted values equal to and greater than the observed function, and equals 0 for predicted values lower than the observed value.

NSE was used in this case to compare the predictions of the SM to those obtained from SWMM. It is a unitless value ranging from $-\infty$ to 1.0, where a value of 1.0 represents perfect agreement. The NSE for the ensemble was calculated as follows: firstly, the NSE of each individual model within the ensemble, for the testing dataset, was calculated. Next, the average of the 100 NSE values was calculated, and this was used as the performance metric.

The CRPS score represents the squared difference of the area under the curve bounded by the predictive Cumulative Distribution Function (cdf) and Heaviside function (Boucher *et al.*, 2009; Hersbach, 2000). Therefore, if the Probability Distribution Function (pdf) of the ensemble is concentrated around the observed function with minimal spread, the CRPS score will be

minimised. This is demonstrated in Figure 3-3. The pdf of ensemble 1 is both sharper in prediction and centered around the observed value, whereas the pdf of ensemble 2 is more diffuse in prediction and not centered around the observed value. Therefore, the CRPS score of ensemble 1 will be much lower than ensemble 2. The CRPS has a negative orientation, favouring smaller values, and should be minimised. A value of 0 in the CRPS score indicates a perfect score.

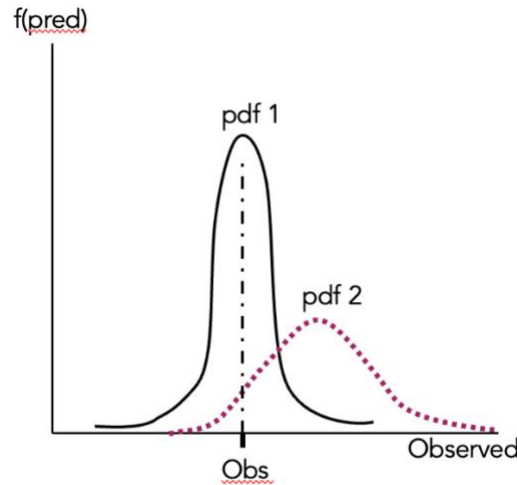


Figure 3-3: Schematic showing the probability distribution functions of two ensembles. Ensemble 1 is denoted in black, ensemble 2 is denoted in the dashed purple line and the observed value is marked

The RH, also known as the Talagrand diagram, is a visual assessment tool used to verify the calibration of predicted distribution. Meaning it can be used to assess whether the ensemble is over or underpredicting or whether the observed value has equal probability of occurring anywhere within the set of predictions. Its construction begins with adding with the observed value to the set of ensemble predicted values. If the predicted set has n members, the new set with the added observed value will consist of $n + 1$ values. The rank of the observed value is then determined and tabulated. This process is repeated for all N ensemble models (the 100 ensembles in this study) and a histogram of the resulting N ranks is produced.

This process is slightly more complicated if the observed value is equal to some of the members in the ensemble. In this case, the tied members (M) are counted. A set of $M + 1$ uniform random numbers are generated and the rank of the observed random number in the $M + 1$ set is noted. An insignificantly small number is subtracted from all ensemble members with a rank lower, and an insignificantly small number is added to the ensemble members with a higher rank. This

ensures that ranks are randomly assigned without affecting subsequent calculations (Thomas M & Colucci, 1997). A relatively flat histogram indicates the ensemble is well calibrated, while asymmetry indicates bias.

PC is a quantitative measure of how frequently the observed value is found within the bounds of the ensemble predictions, which can be indicative of whether the model is under or overpredicting. This method has been used for ensemble verification (Santi *et al.*, 2021), as well as evaluating other possibilistic models (Khan & Valeo, 2017; Khan & Valeo, 2016).

PC, CRPS and RH were evaluated on the entire dataset to obtain a more holistic evaluation of model performance. The use of multiple performance metrics is considered good practice, as different metrics are indicative of different model characteristics (Snieder *et al.*, 2020).

3.3. Results and Discussion

The following section includes a discussion into the performance of the SM model in place of SWMM under the different experiments. All shown results will be for the 100-year 12-hour storm event for consistency. The remaining results for all other storms are shown in Appendix A.

3.3.1. Number of training samples per rainfall event

First, the effect of using 50, 100 or 200 design configurations per rainfall event is demonstrated. Increasing the number of design configurations increases the overall number of data points used to train the ANN. However, it also increases the computational time associated with running the original SWMM model.

Predictive model performance is summarized in Figure 3-4 and Table 3-1. The light blue shaded region delineates the ensemble spread, the blue crosses are the averaged ensemble predicted value and the purple stars are the observed runoff values obtained from SWMM. Thus, these plots show how closely the ensemble average aligns with the observed SWMM values. Ideally, the blue crosses would be centered around the purple stars indicating perfect alignment between the SWMM and SM values. However, as can be seen from the plots that is not the case. The model frequently underpredicts, and often the ensemble range does not include the observed data points, with the observed values being greater than the maximum ensemble prediction.

A more quantitative analysis reveals that although the models appeared to be performing poorly, their NSE values indicate very high accuracy between the SM and SWMM. This is likely due to the small difference in error values between the predicted and observed values. Though it seems the model is underpredicting, the differences between the observed and averaged predicted values are minimal.

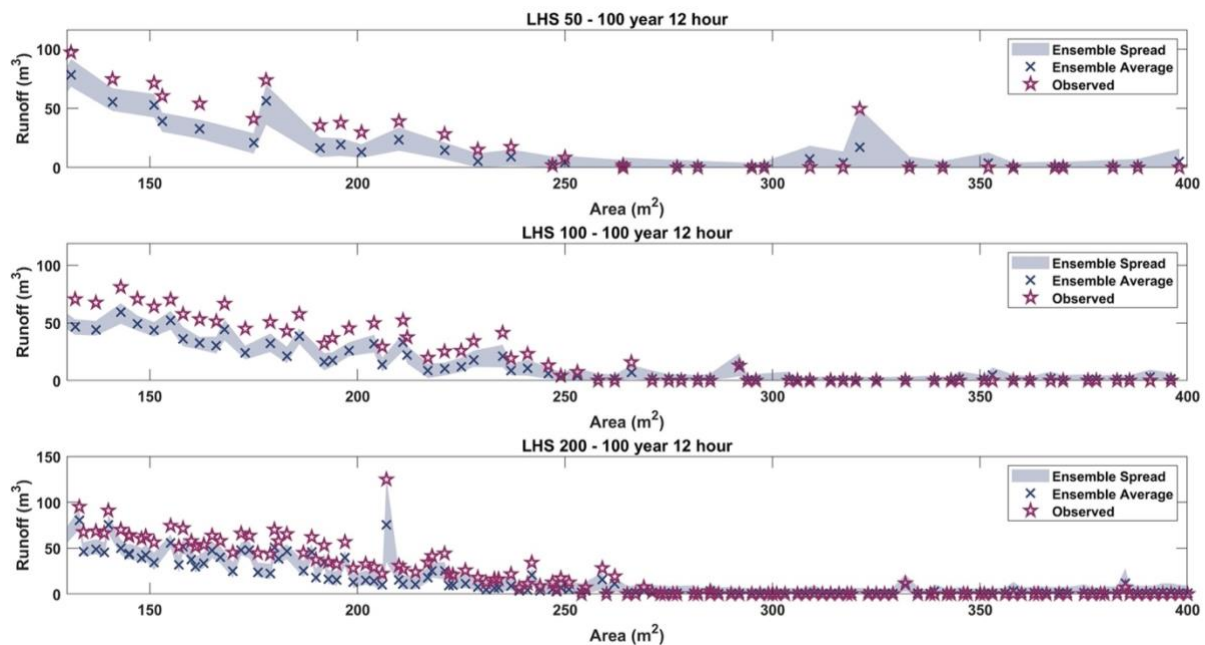


Figure 3-4: Comparison of model predictive capability for the LHS 50, 100 and 200 scenarios

The CRPS results are indicative of a similar result. Though the model underpredicts, the difference in error is small thus causing the area bounded by the ensemble CDF and observed values to be small, resulting in a small CRPS. PC is a more accurate depiction of model bias, as it shows that the observed value does not lie within the range of the ensemble half the time.

Table 3-1: Comparison of model performance (in terms of NSE, CRPS and PC) of the three models (LHS 50, 100 and 200)

Model	NSE (-)	CRPS (m ³)	PC (%)
LHS 50	0.91	0.030	52
LHS 100	0.92	0.028	43
LHS 200	0.94	0.029	48

To further analyze the model performance, the RHs for all three models are shown in Figure 3-5. The horizontal axis in Figure 3-5 represents the rank of the observed value within the ensemble predictions. Since there were 100 ensemble models, the ranks extend to 101 to include the observed value. The vertical axis represents the frequency of the occurrence of the observed value in a specific rank. Not surprisingly, the RHs are asymmetrical, indicating the ensembles are biased. The observed value is frequently ranked at the 101st location, demonstrated by the high frequency of 101st bin, suggesting inherent underprediction of the models. These results demonstrate the important of using multiple performance metrics to capture different model characteristics.

The difference in performance between the three models is minimal. This indicates that using more than 50 design configurations per rainfall event does not contribute to improved model performance, and given the substantial increase in computational demand associated with evaluating 100 or 200 design configuration in SWMM, it is not recommended. Therefore, for this current study, it is noted that using 50 design configurations are sufficient to adequately train a SM. It should be noted however that this number of training samples is highly dependent on the complexity of the underlying system, and the number of training samples should be individually determined on a problem basis.

Given the exhibited underprediction behaviour of the ensembles, the current ensembles cannot be reliably used in place of SWMM. Further model refinement, such as using stratified sampling for data splitting, or introducing new variables, is needed.

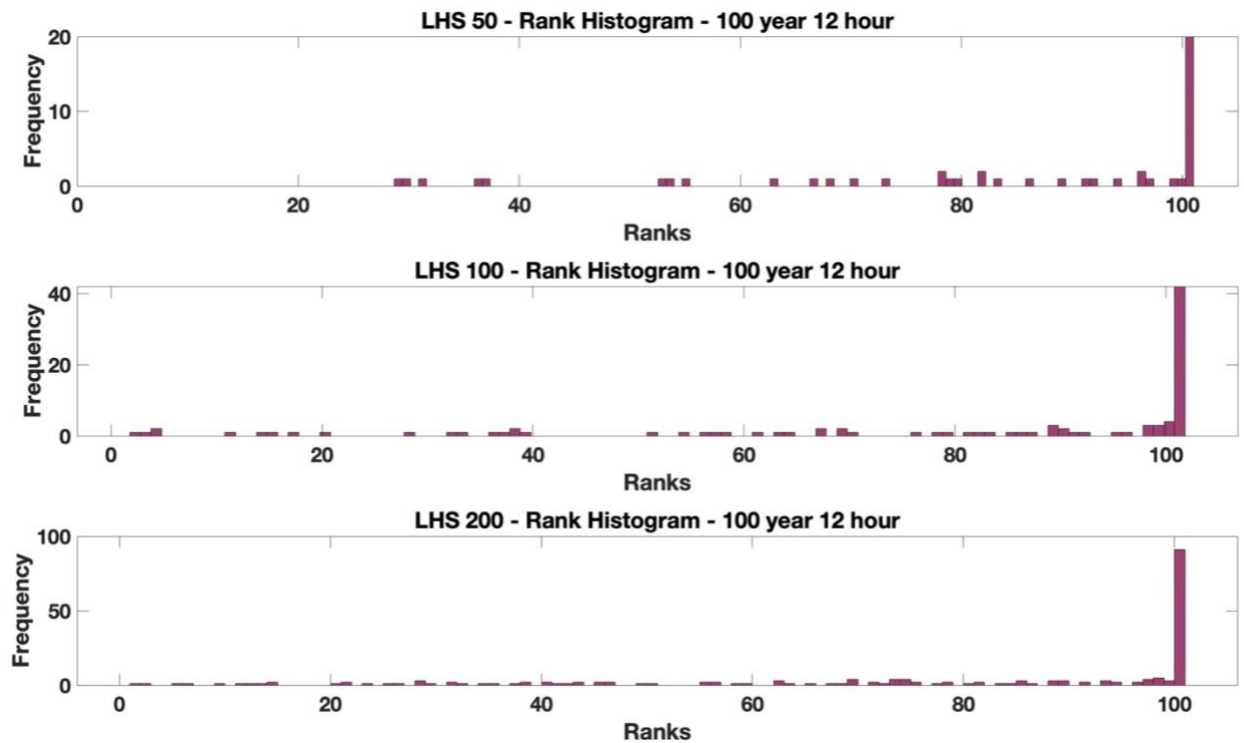


Figure 3-5: comparison of the RHs for all three models. Left - LHS 50. Center - LHS 100. Right - LHS 200.

3.3.2. Random vs stratified sampling for datasplitting

To determine whether stratified sampling improves model performance, event-based and flow-based stratification was carried out. Experiment results are provided in Figure 3-6 and Table 3-2. Figure 3-6 contains the observed versus predicted model plots for both experiments in the top row, with the accompanying RH in the panel below.

Comparing results for the random sampling LHS 50 model in Figure 3-4 and Table 3-1 to those in the stratified sampling scenario, no significant improvement in model performance is observed. For both event and flow-based stratification, the NSE remained at 0.91. Similarly, CRPS values were 0.029 and 0.032 for event and flow-based stratification, which was very close to the 0.03 CRPS value of the original model. PC remained stagnant at approximately 50% for all models. These results therefore indicate the model still suffers from underprediction, further demonstrated through the RH results as well as the location of predicted data points relative to the observed. These results indicate that stratified sampling alone is not sufficient to improve model performance.

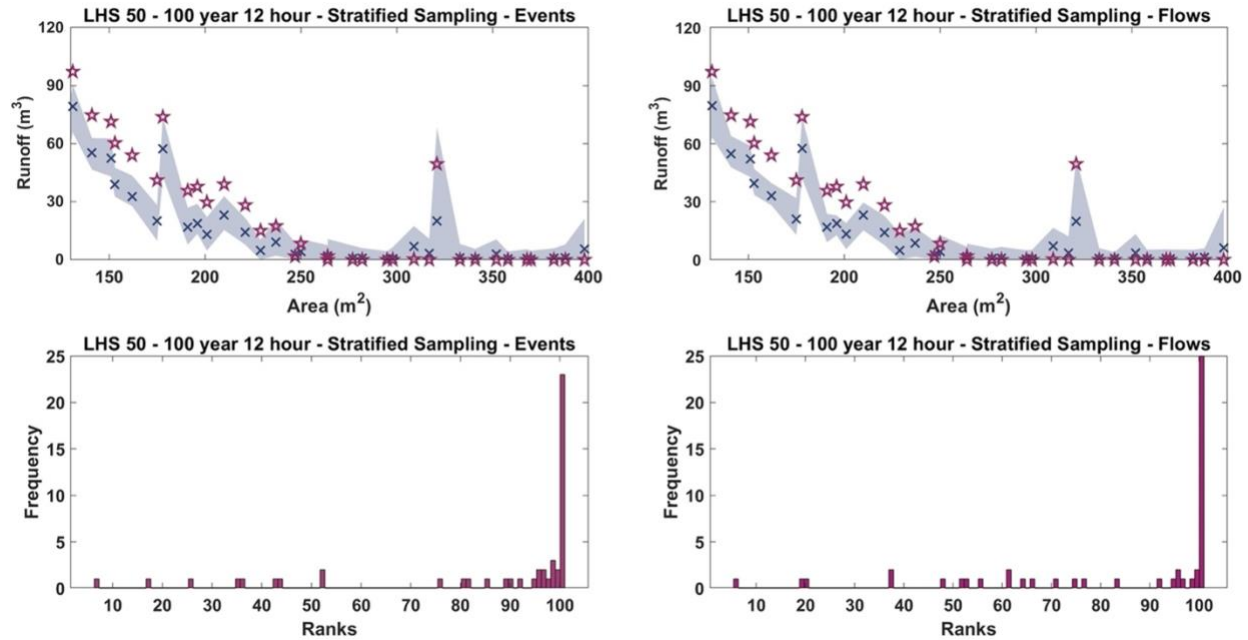


Figure 3-6: Comparison of model predictive performance under different stratified sampling goals. Right: Event based stratified sampling. Left: Flow based stratified sampling.

Table 3-2: Comparison of model performance for flow and event based stratified sampling

Model	NSE (-)	CRPS (m ³)	PC (%)
Flow based	0.91	0.029	54
Event based	0.91	0.032	50

Additionally, MATLAB's built-in random sampling proved to be just as effective as the more sophisticated stratified sampling method. The datasplit for training, validation and testing was 60% - 20% - 20% respectively, and it was found that random sampling inherently selects approximately 60% of the data from each rainfall event for training automatically.

Results of the event-based and flow-based stratification were almost identical. Both models had the same NSE values, with very minimal variation in CRPS and PC values. Neither method addressed model underprediction, suggesting that the underprediction of the model is not stemming from training bias.

3.3.3. Input selection

The impact of input selection on model performance improvement is investigated in this section. The original model used four inputs namely BRC area, growing media depth, gravel storage depth and rainfall depth. Given that rainfall depth was repeated for each design configuration, it is possible that these four inputs alone were sufficiently informative, thus the ANN was not able to accurately infer the underlying system relationships. To this end, two additional inputs were included: average and peak rainfall intensity.

Results of model performance are provided in Figure 3-7 and Table 3-3. Evidently, the addition of the two inputs contributed to significant model improvement, noted in Figure 3-7. The model no longer exhibits consistent underprediction, shown through the increased proximity of the observed and ensemble average markers in the leftmost plot in Figure 3-7 as well as through the relatively uniform RH.

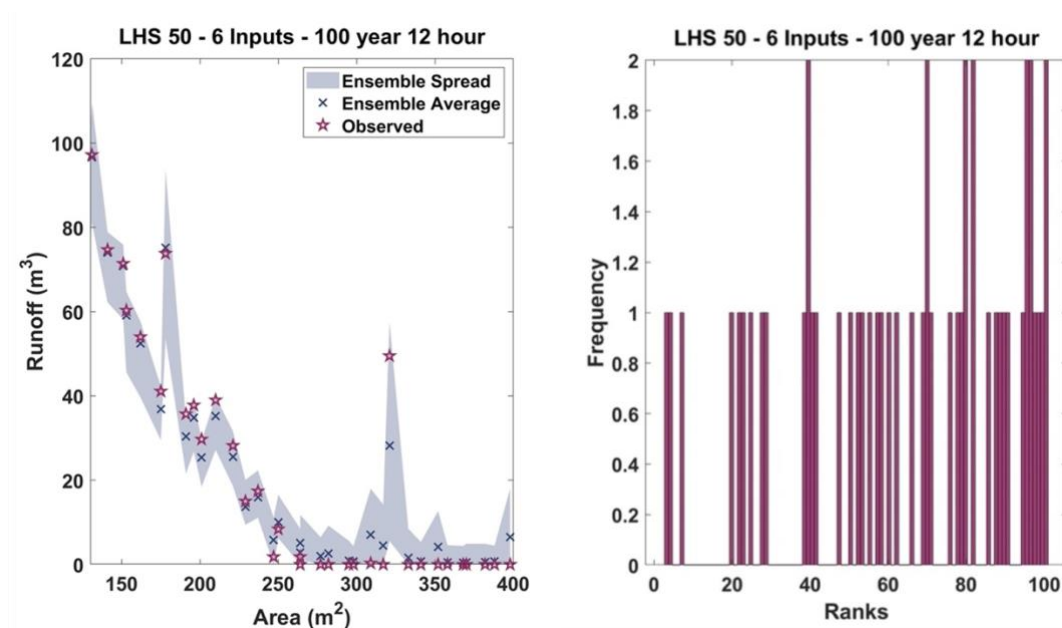


Figure 3-7: Performance plots for the new improved surrogate. Left: Observed vs predicted. Right: RH.

Table 3-3: Performance results of the improved surrogate

Model	NSE (-)	CRPS (m³)	PC (%)
LHS 50 - Improved	0.98	0.005	96

The observed value's frequency of occurrence in a certain bin is no longer concentrated in the 101st bin, rather it is much more varied. Previously, the observed value occurred in the 101st bin

approximately 23 instances (Figure 3-5), whereas in the improved model it is 2 instances. Additionally, the improved surrogate performs much better with NSE increasing from 0.91 in the original LHS 50 model (Table 3-1) to 0.98, CRPS decreasing from 0.03 to 0.005 and most significantly PC increasing from 52% to 96%.

These experiments were carried out on the hydrological SM to establish a baseline SM with reliable performance before developing the second water quality SM. The water quality SM required the addition of EMC as an input variable, therefore, to ensure consistency across both models, EMC was also added as an input to the hydrological model. EMC values for both models were 50 mg/L, 130 mg/L and 300 mg/L. These values were also repeated 50 times per rainfall event, giving a total of 2700 instances of the 7 inputs variables. The results of both the hydrological and water quality SM for EMC 50 under the 100-year 12-hour storm are given in Figure 3-8. It should be noted that the EMC value has no bearing on the runoff results of the hydrological SM however it was included to ensure both models had the same number of inputs and data points for consistency. The range of inputs was also expanded, to now include area values from 1 m² to 400 m² as well as growing media depths from 1 mm to 1000 mm to cater to policy-makers and designers who may want to design smaller BRC's.

From the results of Table 3-4, it is clear that the inclusion of EMC as an input variable did not have a negative impact on model performance. The performance of both the hydrological and water quality model maintain very high performance, as indicated by the three performance metrics. NSE remains very high at 0.99 and 0.89 for the Hydrological and Environmental SM respectively. Similarly, CRPS remains at approximately the 0.004 range and PC is also approximately within the 100% range for both SMs. The RH of both models is indicative of a well-calibrated ensemble of models.

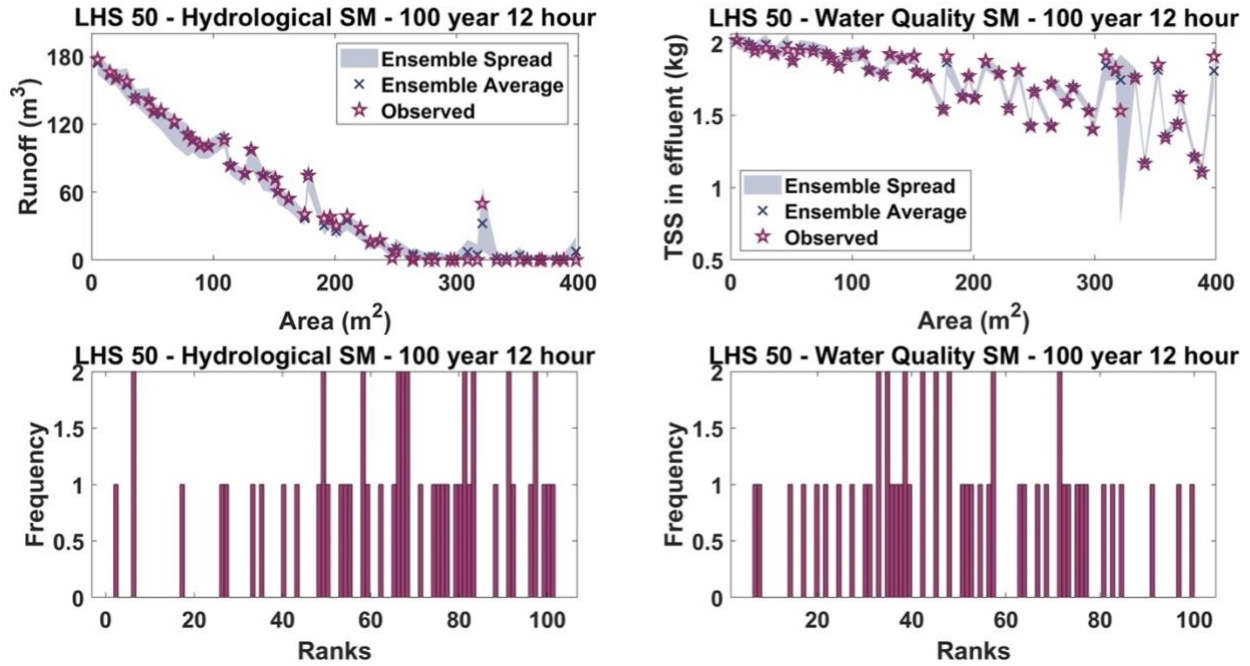


Figure 3-8: Model performance of the hydrological and water quality SM

Table 3-4: Model performance metrics for the hydrological and water quality SM

Model	NSE (-)	CRPS (m ³)	PC (%)
Hydrological	0.99	0.004	96
Water Quality	0.89	0.005	100

The fluctuations in mass of TSS remaining in effluent stem from the fact that different design configurations have different growing media depths. For example, the last two points in the water quality plot correspond to areas of 388 m² and 398 m² with growing media depths of 921 mm and 57 mm respectively. Even though the BRC area increases, the growing media depth is much smaller, explaining the sudden increase in TSS at the outlet.

The SMs can be used in the design of BRCs as they're capable of instantaneously evaluating BRC performance for numerous design configurations. To generate multiple preliminary designs, the results can be plotted against each design parameter and used as a design tool. An example of this plot is show in Figure 3-9 for the water quality SM.

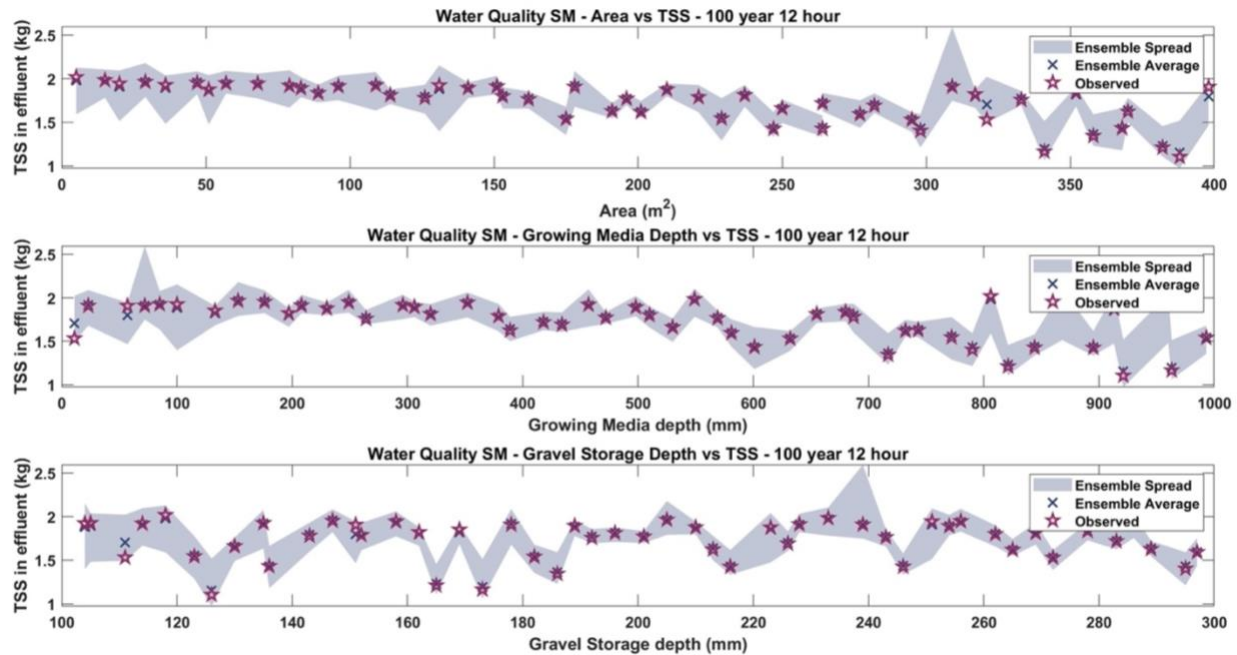


Figure 3-9: Water quality surrogate results for all three design parameters of the BRC

TSS remaining in effluent is plotted against area, growing media depth and gravel storage depth. The required BRC design for a specific performance target can be read off the graphs. For example, for approximately 1.5 kg of TSS remaining in effluent for the 100-year 12-hour storm, an area of 175 m² with a growing media depth of 600mm and a gravel storage depth of 180 mm. Further, given the demonstrated high levels of prediction accuracy, these SMs can be reliably used in simulation-optimisation frameworks in place of physical models. The significant computational demand associated with using a physical model to evaluate a large number of design configurations can be overcome by using these highly efficient and accurate SMs.

The improvement in model performance following the addition of two inputs highlights the importance of input selection on model performance. Prior to the inclusion of the additional inputs, the model displayed inherent bias, frequently underpredicting and failing to include the observed values within the ensemble range prediction, even though its NSE values were extremely high. Consequently, this also highlights the need to use different performance metrics as relying on one value (such as NSE) could be misleading. Finally, the uncertainty associated with using a single SM versus an ensemble is highlighted in the ensemble spread, shown in the observed vs predicted plots. Using a single SM could result in a predicted value that falls anywhere within that range, and given the spread of the ensemble range, the prediction values could either be over or

underpredicted. For example, for a BRC with an area of 36 m², a growing media depth of 960 mm and a gravel storage layer depth of 205 mm, for the 100-year 12-hour storm, the runoff rate would be 142.2 m³. The minimum prediction of the ensemble is 135 m³ and the maximum prediction is 150 m³, with an average of 142.5 m³. Therefore, it is clear to see that if a single model is selected, and it has a tendency to over or underpredict, there would be uncertainty associated with its use. However, averaging out the ensemble predictions results in a much more accurate and precise reading, which increases the reliability of using SMs.

It should be noted that running this SM with 2700 inputs, 50 training samples per EMC value, resulted in a run time of 1.59 seconds. Conversely, running the SWMM model for the same number of inputs resulted in a run time of 19940 seconds (5.4 hours). These results demonstrate the benefit of using SMs in place of physical-based models, especially in an optimisation application. The large number of iterations in an optimisation framework and their associated computational demand in a physical-based model could be a significant deterrent to policy-makers. Using a SM can streamline this process significantly, resulting in a more expedited and highly accurate process.

3.3.3.1. Universal and event-based surrogate

The results provided so far have been obtained from the universal surrogate. The distinction between the universal and event-based surrogate lies in the number of data points and consequently the number of rainfall events. For the LHS 50 model, the universal surrogate had 900 instance of the seven inputs (BRC area, growing media depth, gravel storage depth, rainfall depth, peak and average rainfall intensity and EMC), whereas the event-based surrogate will have only 50 samples per rainfall event it is modelling. In that case, there would be 18 event-based surrogates instead of the one universal model. The event-based SM performance is presented in Figure 3-10 and Table 3-5 and can be compared against the results provided in Table 3-4 and Figure 3-8. Both inputs used the same multi-start ANN method, meaning both models used an ensemble of 100 ANNs. The same 50 design configurations that were used for the 100-year 12-hour event in the universal surrogate are used in the event-based surrogate to ensure consistency when comparing the two models.

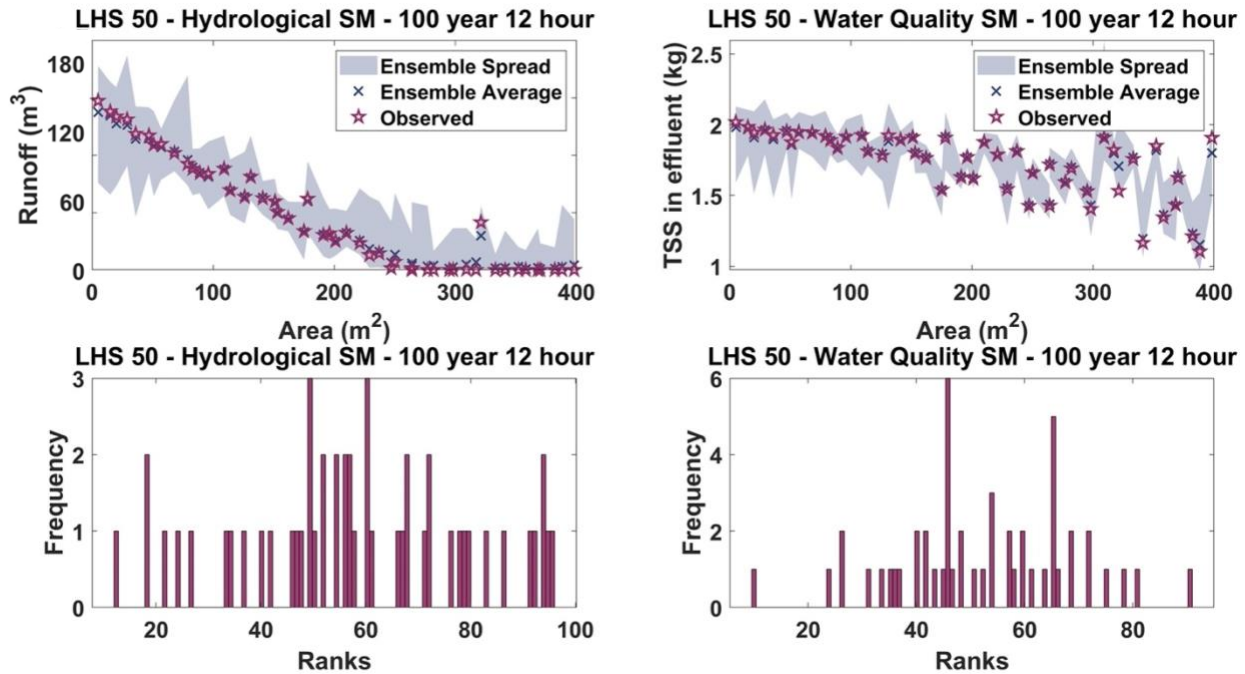


Figure 3-10: Event-based model performance of the hydrological and water quality SM

Table 3-5: Event-based model performance metrics for the hydrological and water quality SM

Model	NSE (-)	CRPS (m ³)	PC (%)
Hydrological	0.91	0.005	100
Water Quality	0.72	0.009	100

The results of Table 3-5 demonstrate a decrease in the NSE index from 0.99 and 0.89 for the universal Hydrological and Environmental SM respectively to 0.91 and 0.72 for the event-based SM. Additionally, the RH of the universal model displays more uniformity, suggesting better calibration. The ensemble spread of the event-based SM is larger than that of the universal surrogate, which is indicative of greater intra-model variability. This is also potentially the reason why the NSE index is slightly lower, as the larger spread of the ensemble could result in less accurate averaged predictions. Though the decrease in performance is not substantially different from the original, it does point to a decrease in model performance.

Appendix A contains the remaining plots showing SM model performance for the remaining design storms. As shown in the figures, model performance maintained high accuracy throughout all the other design storms, demonstrated by the accuracy of the predicted points (SM predictions)

in relation to the observed point (SWMM predictions). These results demonstrate the reliability of using a universal SM, as all the design storms exhibit a high level of predictive accuracy.

Model performance of the SMs discussed thus far has shown to be relatively unaffected by the number of training samples, unaffected by the sampling technique used for data splitting, and only affected by the quality of input variables. It was shown that adding more relevant inputs to the output substantially improved model performance, as these extra inputs provided additional explanatory information of output behaviour, resulting in a better predicted system. It is therefore evident that SM development should be more thoroughly reported and studied, to provide better understanding of the reliability and the workings of SMs in place of physical-based models.

3.4. Conclusion

With the alteration of the natural water cycle and the amplified risks of localised flooding in urban cities due to the impacts of urbanisation and climate change, more innovative mitigatory solutions are sought. One of the most commonly implemented adaptive measures are LIDs, and in particular BRCs. To ensure optimal BRC performance wherever they are placed, their designs are often iteratively revised and evaluated using physical-based models. To streamline this design process, optimisation algorithms are frequently used. However, coupling optimisation algorithms to physical-based models can result in a time-consuming design process. SMs are therefore used in place of physical-based models due to their efficiency and accuracy. ANNs are commonly used as SMs to physical-based models in rainfall-runoff research. The development process of ANNs as SMs is often unreported in research and many uncertainties associated with ANN architecture remain.

This research evaluated the impact of varying the number of training samples, method of datasplit sampling and input selection on SM model performance. Model performance was evaluated through a comparison between SM predictions of BRC performance and SWMM predictions, to assess the reliability of using SM's in place of SWMM. Four performance metrics were used, namely NSE, CRPS, PC and RH. A number of important findings were established. Firstly, the importance of varying SM parameters and investigating their effect on model performance. It was found that using 50 data points per rainfall event instead of 100 or 200 data points resulted in equally accurate predictions. This shows the model can maintain high levels of prediction accuracy without the need for a large number of training points, which is beneficial in the cases of SM synthesis since they are trained on a set of known inputs obtained from the original physical

model. Needing less data points to train the surrogate significantly cuts down the computational demand associated with running the original physical model. Running the SWMM model for 18 rainfall events, with 150 design configurations, 50 designs based on each EMC value resulted in a computational time of 19440 seconds (5.4 hours) whereas running either SM for the same amount of data instances took 1.59 seconds.

Next, a comparison of the impact of random and stratified sampling for datasplitting revealed no discernable difference between the two methods on model performance. Event-based and flow based stratified sampling also proved to have no positive impact on model performance. The reasons for the lack of model improvement using stratified sampling were two-fold. Firstly, MATLAB's built-in random sampling was already selecting approximately 60% of inputs from each rainfall event, rendering the event-based stratified sampling unnecessary. Secondly, since the flow based stratification did not improve model performance, it was clear that the model's bias is not a result of training bias. Therefore, the effect of input selection was evaluated. Including two additional inputs, average and peak rainfall intensity per rainfall event, greatly improved model performance. Results indicated the model no longer suffered from underprediction bias. This finding highlights the importance of input selection and its impact on the model's ability to fully infer the underlying system relationship's. Further, a comparison between model performance of an event-based and a universal SM revealed that both models had similar performances, with the universal SM performing slightly better. This finding further enforces the SM's ability to be a universal approximator. Additionally, universal SM's are more efficient, facilitating BRC design under different rainfall events in one SM. This research also demonstrated the benefit of using an ensemble of SM's rather than a singular ANN. The uncertainty associated with using a singular ANN, shown in the ensemble range plots, was minimized by using the averaged ensemble predictions.

This thorough investigation of the dependence of ANN configurations on model performance provides a basis for a universal framework that facilitates ANN development for SM application. Additionally, it highlights the importance of investigating all aspects of ANN-SM development, to promote the development of better performing models. Having a common framework such as this presents a unified approach to ANN-SM development, thereby increasing the reliability of ANN-SM use in rainfall-runoff simulation applications. This in turn can promote more widespread use of ANN-SM for BRC design and optimisation, which can streamline the previously time-consuming process, encouraging more widespread adoption of BRCs in urban cities. Ultimately, the more

implementation of mitigative approaches, the more resilient urban cities will be in the face of increasing urbanisation and climate change.

Chapter 4 – Surrogate model assisted optimisation of bioretention cell design under current and future climate conditions

This chapter introduces the concepts of bioretention cell design under current and future climate conditions with a focus on surrogate model integrated optimisation.

4.1. Introduction

With increasing urbanisation, rural and undeveloped lands are increasingly converted to urban areas. The resulting impervious areas, such as parking lots, buildings and streets replace the native vegetation, consequently disrupting the natural water balance of a watershed (Davis, 2005). Additionally, the rate of urbanisation has been steadily growing worldwide. In Indiana, through a 10-year period dating from 1992 to 2002, more than 4000 km² of farmland were converted to urban areas (Gao *et al.*, 2016). In fact, in the past decade the impervious area coverage has exceeded approximately 100,000 km² in the contiguous United States (Gao *et al.*, 2016). In another study conducted in Watford, Connecticut, the rate of urbanisation was tabulated against the increase in yearly runoff depth. It was found that for a 0.02 km² subdivision, as impervious area increased from 1% to 32%, the runoff depth increased from 0.1 cm to over 50 cm, an increase of 49,000% (Dietz & Clausen, 2008). On the same study area, Bedan and Clausen (2009), found that average flow depth increased 16 times for an increase of 32% in impervious cover. There are several other factors that could influence surface runoff increase such as the severity, intensity or duration of a rainfall event, the upstream drainage area and antecedent moisture conditions in the soil. However, there exists a consensus in the literature that points to urbanisation adversely altering surface runoff volume.

Further, urbanisation contributes to deteriorated stormwater quality due to increased nutrients, bacteria and toxic metals (Ahiablame *et al.*, 2012; Bedan & Clausen, 2009; Dietz & Clausen, 2008; Qin *et al.*, 2013). Pollutants such as toxic metals resulting from the wear of vehicle parts such as tires, nutrients from fertilizers and detergents and organic compounds such as hydrocarbons and surfactants are dissolved in stormwater and pose a significant threat to ecological and human health (LeFevre *et al.*, 2015). These pollutants are a threat since current

conventional stormwater management practices such as retention ponds, catch basins and gutters aim to quickly transport stormwater to the nearest receiving water body without treatment, therefore depositing all these dissolved pollutant in our receiving waters. Dietz and Clausen (2008) reported that an increase in surface runoff was the dominant driver behind pollutant loading increase, with Nitrate-Nitrogen ($\text{NO}_3\text{-N}$) and Total Nitrogen (TN) exhibiting exponential increases with impervious area increase. Exceeding a threshold of approximately 12% of impervious area is strongly linked to degradation of fish species and number of fish in streams (Wang *et al.*, 2001).

4.1.1. Climate change

Historical trends are indicative of a changing climate, with increasing surface temperatures globally, intensification of heavy precipitation events in certain areas of the world, changes to precipitation patterns, changes in ocean salinity, melting of sea ice and increase in ocean heat content (IPCC, 2014). Prediction of future climate change and its impacts is a necessity to ensure communities can both mitigate and adapt to possible future conditions. Prediction methods include General Circulation Models (GCMs) and Regional Circulation Models (RCMs), amongst others.

4.1.1.1. Climate change modelling

General Circulation Models, which are coarse resolution mathematical models representing the interactions between the Earth and surrounding environment, can be used to predict future climate change trends. The use of GCMs is very common in studies investigating the impacts of future climate change. Zahmatkesh *et al.* (2015) used 134 GCMs from the fifth phase of the Coupled Model Intercomparison Project (CMIP5), a tool developed to ensure standardized boundary conditions across all developed GCMs, to project future rainfall. Projections for future rainfall were obtained for the period from 2030-2059 with a complementary baseline period from 1950-1979. The future time period matched the baseline time period to facilitate the process of obtaining change factors through CFM. Similarly, Ghodsi *et al.* (2020) investigated 14 GCM models to obtain the minimum, maximum and mean scenarios for future precipitation. In their study, Zhang *et al.* (2019) used an ensemble of 8 GCMs to investigate the impacts climate change has on pollutant removal, flow rate reduction and harvesting reliability on biofilter systems, wetlands and rainwater tanks. GCM's are commonly used in an ensemble, as each model tends to predict future climate differently due to their unique implementation of the governing atmospheric physical equations; therefore, using an ensemble is recommended in research (Barah *et al.*, 2021; Rosenberg *et al.*, 2010; Tirpak *et al.*, 2021; Zhang *et al.*, 2019).

RCMs are also commonly used in research. To determine the modifications to BRC design needed to ensure past performance under future climate, Tirpak *et al.* (2021) utilized 10 RCMs downscaled from GCMs for their work. Rosenberg *et al.* (2010) used a combination of two RCMs to assess the difference in past and future trends in extreme precipitation across Washington State. Their recommendations include using an ensemble size greater than two models to ensure a more robust set of conclusions. Additionally, it is considered good practice to combine RCMs with secondary downscaling methods to correct resulting systematic bias (Tirpak *et al.*, 2021; Willems *et al.*, 2012). In their review of commonly used downscaling and bias-correction methods for future climate modelling, Willems *et al.* (2012) identify a number of methods to correct RCM bias such as; empirical transfer functions making use of observed empirical data, resampling which seeks historical events that are similar to the future climate output to be used as the downscaled future event and finally stochastic rainfall methods with probability distributions based on the coarser scale climate predictors.

4.1.1.1. *Climate change impact on stormwater*

GCMs and RCMs can also be used to predict the effect climate change can have on stormwater characteristics. Gaur *et al.* (2019) demonstrate the increasing flooding frequency in Canadian cities. According to their study 40%-60% of the cities are projected with future flooding increases. Under the RCP 8.5 scenario, storms with a severity similar to that of a storm expected to occur once every 100 years are now expected to occur once every 15 years for the City of Toronto. In other works, Semadeni-Davies *et al.* (2008) investigated the impact of climate change on the Lussebacken catchment (2474 ha) in Sweden, using an RCM for the 2071-2100 period. Under no climate change and continued urbanisation, minor flooding of 50cm above bank full occurred in the Lussebacken creek. However, under the worst climate change scenario, flooding of 2.55 m above bank full was recorded (Semadeni-Davies *et al.*, 2008). M. Wang *et al.* (2019) utilized future Intensity Duration Frequency (IDF) curves for a hypothetical catchment in Guangzhou, China to analyze the impacts of climate change on stormwater runoff. Under the RCP 8.5 scenario for the period of 2040-2059 peak flow increased 32% for the 1-year 1-hour storm, and 14% for the 100-year 1-hour storm. Zahmatkesh *et al.* (2015) considered the impacts of climate change on the Bronx River watershed in New York, for the period of 2030-2059 under all possible RCP scenarios. Their results showed increases of up to 48% in monthly surface runoff values under RCP 8.5 scenario's when compared to the historical levels. Therefore, with the demonstrated probable increases in intense precipitation events, runoff and subsequent risk of flooding, it is

imperative to implement adaptive infrastructure that simulates the natural water cycle to mitigate the impacts of climate change on stormwater runoff. A common solution is through the use of Low Impact Development (LID) units.

LIDs aim to restore pre-development hydrological conditions by promoting infiltration and evapotranspiration (Ahiablame *et al.*, 2012; Eckart *et al.*, 2017). Common examples of LIDs include; Bioretention cells (BRC), infiltration trenches, permeable pavements and green roofs (Ahiablame *et al.*, 2012).

4.1.2. Bioretention cells

A number of studies have reported BRCs ability to successfully mitigate the changes in stormwater runoff (Batalini de Macedo *et al.*, 2019; Jia *et al.*, 2015; Johnson & Hunt, 2019; Khan *et al.*, 2012b, 2012a; Kratky *et al.*, 2017; Winston *et al.*, 2016). Their reported hydrological (runoff retention) and water quality (pollutant removal) performances vary in literature, due to the influence of their design parameters (e.g., surface area, growing media depth and gravel storage depth) as well as the characteristics of the rainfall event and the upstream impervious area draining into the BRC.

Given the probable increases in rainfall event frequency due to climate change, and consequent increase in stormwater runoff quantity and pollutant concentration, it is necessary to evaluate the performance of BRCs under future climate. BRCs can be designed to mitigate the impacts of climate change, however their design process is quite iterative and time-consuming as current design guidelines offer a range of recommended parameter design values.

4.1.3. Simulation-optimisation framework for BRC design

Optimisation algorithms are often integrated into a design simulation framework, known as simulation-optimisation, which facilitates the design process of BRCs. Eckart *et al.* (2018) combined the BORG multi-objective optimisation algorithm with the Stormwater Management Model (SWMM) to optimise the design of multiple LID units, including BRCs, to reduce runoff and peak flow while maintaining low cost. Li *et al.*, (2020) used the physical model DRAINMOD in combination with Response Surface Methodology to optimise the design of BRCs for runoff retention and nitrogen removal from effluent. Simulation-optimisation frameworks are also used to optimise BRC design under climate change. For example, Liu *et al.* (2016) used the Long-Term Hydrologic Impact Assessment-Low Impact Development 2.1 (L-THIA-LID 2.1) physical model

coupled with A Multi-Algorithm Genetically Adaptive Multiobjective (AMALGAM) method to optimise LID design, including BRCs. Their study assessed optimal placement and selection of each type of LID, and the necessary improvements to stormwater runoff (quantity and quality) required to maintain the same characteristics of runoff in 2050 as in 2001 (Liu *et al.*, 2016). The limitation of the simulation-optimisation approach is the substantial computational demand associated with running the physical-based model repeatedly. An alternative approach is the use of Surrogate Models (SMs), which are efficient and accurate numerical approximators of the original model (Brunetti *et al.*, 2017). A number of studies employ the use of these models, with Artificial Neural Networks (ANNs) being the most common type of surrogate. Raei *et al.* (2019) used an ANN as a surrogate to SWMM in their optimisation of LID design. Their work combined the use of the Non-Dominated Sorting Genetic Algorithm (NSGA- II) method to optimise runoff retention, pollutant removal and cost under competing objectives from different stakeholders (Raei *et al.*, 2019). Ghodsi *et al.* (2020) also used an ANN as an SM to SWMM in their single objective optimisation of LIDs, including BRCs, under climate change. BRC design under event-based rainfall scenarios and under a combinatory multiple rainfall scenario was evaluated to compare the difference in performance and determine whether an individual design or a universal design was more efficient (Ghodsi *et al.*, 2020). The use of a single ANN as such assumes the predictions are deterministic in nature. Whereas it is well known that ANNs are stochastic in nature, meaning their results cannot be identically replicated in consecutive occasions (Zhang *et al.*, 2019). Therefore, to avoid the potential additional bias and uncertainty stemming from the use of a single SM, an ensemble of surrogates is used in this study. To the author's best knowledge Zhang *et al.* (2019) are the only study to utilise an ensemble of SMs, however, their research does not consider BRCs and instead focuses on detention tanks, and they do not take in account optimisation under future climate. Additionally, MIKE FLOOD is used as the physical-based model, as opposed to SWMM which is used in this research.

The primary focus of this chapter is to determine the necessary modifications to BRC design to ensure the same level of performance under current climate conditions is maintained under future climate. This will be done by comparing the optimal BRC design under current and future climate conditions. Optimal BRC design will ensure maximised runoff retention and TSS removal performance whilst minimising design costs. Firstly, the BRC will be evaluated using SWMM. An ensemble of ANN SMs will then be developed to be coupled with the NSGA-II optimisation algorithm. This ensemble surrogate-based ensemble simulation-optimisation is the first of its kind

to evaluate BRC design under future climate change and identify the necessary modifications to ensure resiliency against future climate.

The research is structured in the following manner, a comparison of optimised BRC design under current and future climate change, for two climate periods (2030 – 2059 and 2060 – 2089), is performed. Consequently, the modifications to the following BRC design parameters, BRC area, growing media depth and gravel storage depth, are determined, with the aim of maintaining past performance under climate change.

4.1.4. Objectives

The main objective of this research is optimisation BRC design parameters under future climate change for two climate period, namely 2030 – 2059 and 2060 – 2089, which will be referred to as near future and extended future respectively moving forward. A comparison of optimised BRC design performance under current and future climate will be evaluated to determine the change in performance caused by the impacts of climate change. This gap in performance will be addressed by modifying the BRC design parameters (area, growing media depth and gravel storage depth) to ensure that past performance is maintained under future climate conditions. The established additional modifications needed to ensure future climate resiliency can be implemented in future design guidelines. In turn, this will facilitate implementing proactive approaches to impending climate changes approaches, that can easily be applied throughout urban watersheds.

Further, an ensemble of ANN SMs will be trained to be used in place of SWMM for the simulation-optimisation framework. This is done to eliminate the computational burden associated with running physical models for optimisation purposes. SMs are much more efficient running at a fraction of the time their physical model counterparts do, whilst also maintaining extremely high levels of accuracy. Establishing the reliability of using SMs in place of physical models can promote widespread use of this approach, addressing the computational demand limitation of their counterparts.

4.2. Methods

All the calculations conducted in this study were performed using MATLAB 2020a and the deep learning toolbox for the development of the SMs. The EPA SWMM model used was version 5.1.015.

4.2.1. BRC design configuration

For this study, the design parameters to be varied are BRC area, growing media depth and gravel storage depth. Optimisation of these design parameters under current climate conditions will be performed using the design recommendations outlined in Chapter 2. Under future climate, these design parameter values will be adjusted accordingly to ensure that past performance is maintained.

4.2.2. SM development

The primary objective of the SMs is to facilitate BRC design evaluations in a simulation-optimisation framework without having to rely on physical-based models. Therefore, the inputs to the SMs consisted of the BRC design parameters (area, growing media depth and gravel storage depth), as well as the rainfall characteristics (depth, average and peak intensity) pertaining to the different rainfall scenarios, and EMC values to model pollutant concentration. These inputs were selected as they are assumed to be the only parameters changed during BRC design iterations, since the surrounding characteristics remain unchanged. Both BRCs had 10 hidden neurons and 1 output each respectively.

Based on the results of Chapter 3, universal surrogates with 50 design samples per rainfall event using random sampling for datasplitting were developed. Latin Hypercube Sampling (LHS) was used to obtain a well-distributed set of 900 initial BRC designs consisting of different areas, growing media depths and gravel storage depths. Next, these designs were evaluated in SWMM to obtain both the hydrological and environmental performance, under the 18 different rainfall scenarios. The ANN surrogates were then trained on these inputs and outputs. To minimise the bias associated with using a singular SM, a set of 100 ANNs were used as ensemble for each SM. SM ensemble model performance was evaluated using the Nash-Sutcliffe Efficiency (NSE) index, the Continuous Ranked Probability Score (CRPS), Percent Capture (PC) and the Rank Histogram (RH) was used to visually assess the performance. Given the deterministic nature of

NSE, the NSE was calculated for each ensemble model and finally averaged to obtain the average NSE performance of the 100 ensembles.

4.2.3. Rainfall Scenarios

BRC designs were evaluated for the 18 different rainfall scenarios, to determine their efficacy under different rainfall conditions for current and future climate. To evaluate BRC performance in removing pollutants from stormwater under different loadings, three EMC values were used in this study. These values were 50 mg/L, 130 mg/L and 300 mg/L, chosen to represent the 25%, 50% and 75% percentile respectively, of TSS concentration in Toronto. Exact EMC values for the study site were unavailable, therefore the EMC information was obtained from design guidelines (City of Toronto 2006) and from the work of Behera *et al.*, (2000) into runoff quality in Toronto.

4.2.3.1. *Climate change modelling*

Two periods of future climate were selected for this research: near future (2030 – 2059) and extended future (2060 – 2089). This was done to compare the difference in modifications necessary under the two future climate periods, i.e. to see how much larger or how much deeper BRCs need to be designed to perform the same in extended future period as they did in the near future period. This will help determine whether these modifications are very different and much more expensive, or whether they will be relatively minor.

For both periods, three RCP scenarios were modelled. RCP 2.6, 4.5 and 8.5. This was done to compare the difference in modifications and cost associated with optimising BRC design for the most severe climate scenario (RCP 8.5), the intermediate scenario (4.5) and the least severe (2.6). If the differences in design costs are minimal, then it would be recommended to incorporate the modifications for the most severe climate scenario in design guidelines. This will ensure future resiliency against both the least severe and intermediate scenario.

The climate change data was obtained from the Intensity Duration Frequency (IDF) tool developed by Schardong *et al.* (2020) for the Toronto North York Station (ID: 615S001). The station was operational for approximately 30 years, therefore the future climate periods selected were also 30 year intervals. The tool includes 24 downscaled bias-corrected global climate models, obtained from the Pacific Climate Impacts Consortium (PCIC), with a gridded resolution of 300 arc-seconds (approximately 10 km). The tool offers the choice of selecting a single downscaled GCM or an ensemble of the GCMs in climate change projections. To minimise the

uncertainty associated with using only one GCM, the ensemble option was selected, which combines all 24 GCM and outputs the median value. Therefore, the obtained rainfall depth under each RCP scenario was the median value obtained from the ensemble of GCM's. The outputted rainfall depth for each of the 18 modelled storms was used to produce the future AES design storm distributions. Total rainfall depths (mm) for all scenarios are given in Table 4-1.

Table 4-1: Rainfall depths (mm) for all climate scenarios

Storm	Historical	2030 – 2059 RCP 2.6	2030 – 2059 RCP 4.5	2030 – 2059 RCP 8.5	2060 – 2089 RCP 2.6	2060 – 2089 RCP 4.5	2060 – 2089 RCP 8.5
2-year 6-hour	38.89	37.83	38.11	38.43	37.59	38.67	42.50
25-year 12-hour	99.01	110.43	106.38	1100.76	107.98	107.19	114.27
100-year 24-hour	128.93	172.71	182.28	181.94	176.00	177.29	184.69

4.2.4. Optimisation of BRC design

BRC hydrological and environmental performance is strongly dependent on its design configuration. Using an optimisation algorithm can ensure maximised performance under current and future climate scenarios. One of the most common type of optimisation algorithms are Genetic Algorithms (GAs). GAs have their foundations rooted in principles of natural selection and Darwin's theory of survival of the fittest (Sivanandam & Deepa, 2008a). A GA's basic framework consists of parent population initialisation, encoding the initial population set, selection of the fittest parents, recombination to produce an offspring solution from the reproduction of the parents, mutation to introduce random genetic information and replacement of the parent solutions with better offspring solutions (Sastry *et al.*, 2014; Sivanandam & Deepa, 2008b).

The NSGA-II optimisation algorithm (Deb *et al.*, 2000) is widely used in simulation-optimisation frameworks. It's performance has been tested against a number of other optimisation algorithms (Kollat & Reed, 2006; Reed *et al.*, 2013; Yazdi *et al.*, 2018), and it has proven to be highly efficient, generating diverse solutions. Therefore, it will be used as the optimisation algorithm in this research. The NSGA-II process begins by evaluating the original parent population and assigning

them a fitness rank. In this study the original parent population is composed of three decision variables, namely: BRC area, growing media depth and gravel storage depth. The initial population set is evaluated according to three objectives functions, those being runoff retention, TSS removal and design cost. The fitness rank is a quantitative measure of the strength of solutions. All solutions within the first rank are non-dominated, meaning no other solutions outrank them. All solutions in the following rank are dominated by the first rank, meaning they are outperformed in objective function values. For example, the first rank (referred to as the first front) will contain solutions that have the highest runoff retention, TSS removal and lowest cost, and they cannot be made better without compromising performance in one or another objective. However, all solutions in the second rank are outperformed by those in the first rank, meaning they do not provide solutions that yield the highest runoff retention, TSS removal or lowest design cost. Solutions are based on ranks, however in this case that two solutions belong to the same rank, a crowding distance calculation is used to rank the solutions. The process for crowding distance is as follows:

1. all solutions in front i are ranked in ascending order for the m^{th} objective, and given a crowding distance of 0.
2. The maximum and minimum objective values are identified and saved as f_{max} and f_{min} respectively. These values are also given a crowding distance of infinity.
3. Starting from j onto the penultimate value, the crowding distance for each solution is calculated as follows:

$$I(d_j) = I(d_j) + \frac{I(j+1).m - I(j-1).m}{f_{max} - f_{min}} \quad (4-1)$$

where $I(d_j)$ is the crowding distance for the j^{th} ranked value in rank i and objective m , and $I(j+1).m$ corresponds to the value of the m^{th} objective function for the $j+1$ individual in rank i .

4. This process is repeated for all the remaining m objectives.
5. Finally, the m crowding distances are summed together and assigned to all the individuals within rank i . This process is repeated for the remaining ranks.

Once this is complete, recombination begins. This is the process of combining the two parent individuals to produce an offspring population. In this study the Simulated Binary Crossover (SBX) operator was used. This is done as follows:

1. First, a uniform random number u from 0 to 1 is sampled.

$$\beta = (2u)^{\left(\frac{1}{n_c+1}\right)} \text{ if } u \leq 0.5 \quad (4-2)$$

$$\beta = \frac{1}{(2(1-u))^{\left(\frac{1}{n_c+1}\right)}} \text{ if } u > 0.5 \quad (4-3)$$

where n_c is the crossover index, and β is a crossover index.

2. Once the constant β is calculated, the two offspring children are created as follows:

$$C_1 = 0.5[(1 - \beta)(p_1) + (1 + \beta)(p_2)] \quad (4-4)$$

$$C_2 = 0.5[(1 + \beta)(p_1) + (1 - \beta)(p_2)] \quad (4-5)$$

Where C_1 and C_2 are child 1 and child 2 respectively, and p_1 and p_2 are parent 1 and parent 2 respectively.

Recombination is preformed upon the selection of a random number k , where if k is less than a certain value, recombination occurs, otherwise, mutation occurs. Mutation is the process of slightly altering the original parent to produce a different offspring. It is calculated as follows:

1. A random number r is drawn from 0 to 1. The mutation variation is calculated as follows

$$\delta = (2r)^{\left(\frac{1}{n_m+1}\right)} - 1 \text{ if } r < 0.5 \quad (4-6)$$

$$\delta = 1 - [2(1 - r)^{\left(\frac{1}{n_m+1}\right)}] \text{ if } r \geq 0.5 \quad (4-7)$$

where n_m is the mutation index. This mutation variation is then added onto the child as follows:

$$C_1 = P_1 + \delta \quad (4-8)$$

Once recombination/or mutation are completed the offspring population is generated. The offspring population is combined with the original parent population (size N) to produce a population size of $2N$ and the population is sorted based on the non-domination ranks. The next generation is filled by each front sequentially until the population size equals N . This process is repeated for the specified number of generations. The pseudo code for this process is provided

in Appendix C. This code was originally obtained from Seshadri (2009), and adapted to incorporate using an ensemble of SMs to evaluate objective functions.

Runoff retention and TSS removal were predicted by the two SMs, Hydrological and Environmental, respectively. The 100 ensemble members of each SM were individually used to predict the runoff and TSS at the outlet, and the final predicted output was the averaged value of the 100 predictions. Using an ensemble of SMs in this way ensured no single SM was depended on, therefore minimising the potential for bias and uncertainty.

4.2.4.1. Objective 1: Runoff retention

Minimisation of total runoff volume leaving the BRC is preferred. This leads to less stormwater runoff and thus less risk of localized flooding. The objective function is thus expressed as:

$$F_1 = \frac{V_i}{V_{imax}} \quad (4-9)$$

where V_i represents the total runoff volume from the BRC for rainfall scenario i and V_{imax} represents the total maximum runoff volume to the outlet assuming there was no BRC. For brevity, the results for this study will relate only to the 2-year 6-hour, 25-year 12-hour and 100-year 24-hour storms. These three storms were selected so that BRC performance under current and future climate is evaluated under storms of mild, medium and severe rainfall conditions. Therefore, V_{imax} for all design storms under all RCP scenarios is given in Table 4-2.

Table 4-2: Maximum runoff volume (m^3) values under all storms

Storm	Historical	2030 – 2059 RCP 2.6	2030 – 2059 RCP 4.5	2030 – 2059 RCP 8.5	2060 – 2089 RCP 2.6	2060 – 2089 RCP 4.5	2060 – 2089 RCP 8.5
2-year 6-hour	75	69	69	69	66	69	77.4
25-year 12-hour	198	228	216	231	222	219	237
100-year 24-hour	234	366	381	381	372	375	387

4.2.4.2. Objective 2: TSS removal

As climate change affects the characteristics of stormwater, it is likely that pollutant concentration will increase. Therefore, to mitigate this, the BRC's second design objective is to minimise the mass of TSS remaining in the effluent at the outlet. This is expressed as:

$$F_2 = \frac{TSS_i}{TSS_{imax}} \quad (4-10)$$

where TSS_i represents the mass of TSS remaining in the effluent for rainfall scenario i , and TSS_{imax} represents the maximum mass of TSS remaining in the effluent assuming no BRC is implemented in the watershed. For the 100-year 12-hour storm results, an EMC value of 130 mg/L is used as the washoff coefficient. The maximum TSS mass is given in Table 4-3 for all climate scenarios.

Table 4-3: Maximum TSS mass (kg) under all climate scenarios

Storm	Historical	2030 – 2059 RCP 2.6	2030 – 2059 RCP 4.5	2030 – 2059 RCP 8.5	2060 – 2089 RCP 2.6	2060 – 2089 RCP 4.5	2060 – 2089 RCP 8.5
2-year 6-hour	22.45	21.79	21.95	22.16	21.69	22.30	24.73
25-year 12-hour	63.22	71.42	68.49	71.65	69.64	69.07	74.19
100-year 24-hour	84.98	117.67	124.93	124.67	120.16	121.17	126.76

4.2.4.3. Objective 3: Cost of BRC

The BRC design cost is minimised in this study. To calculate the cost, the Sustainable Technologies Evaluation Program (STEP) cost tool which can be found at <https://sustainabletechnologies.ca/lid-lcct/> was used (Toronto and Region Conservation Area, 2019). It should be noted that only the major costs associated with constructing and placing a BRC were considered for simplicity. The costs considered are as follows:

$$\text{Underdrain} = \text{Length}_{\text{BRC}} * 73.05 \quad (4-11)$$

$$\text{Stone: } (\text{Length}_{\text{BRC}} * \text{Width}_{\text{BRC}} * \text{Gravel storage depth}) * 49.88 \quad (4-12)$$

$$\text{Curb and Gutter inlets: } ((\text{L}_{\text{BRC}} * 2) + (\text{W}_{\text{BRC}} * 2)) * 107.78 \quad (4-13)$$

$$\text{Filter media: } \text{Area}_{\text{BRC}} * \text{Growing media depth} * 1.25 * 50.64 \quad (4-14)$$

$$\text{Vegetation: } \frac{2}{3} * \text{Length}_{\text{BRC}} * \text{Width}_{\text{BRC}} * 94.50 \quad (4-15)$$

4.2.4.4. Optimisation framework

$$\text{Minimize } F_1 = \frac{V_i}{V_{imax}} \quad (4-16)$$

$$\text{Minimize } F_2 = \frac{TSS_i}{TSS_{imax}} \quad (4-17)$$

$$\text{Minimize } F_3 = \text{Underdrain} + \text{Stone} + \text{Curb and Gutter} + \text{Filter media} + \text{Vegetation Cost} \quad (4-18)$$

Subject to the following constraints which are related to the maximum and minimum values of the design parameters, provided are given in Table 4-4.

$$A_{min} \leq A_i \leq A_{max} \quad (4-19)$$

$$GM_{min} \leq GM_i \leq GM_{max} \quad (4-20)$$

$$GS_{min} \leq GS_i \leq GS_{max} \quad (4-21)$$

Where A_i , GM_i and GS_i are the decision variables used in the optimisation algorithm representing Area, Growing Media depth and Gravel Storage depth respectively.

Table 4-4: Design constraints of the decision variables

Design parameter	Minimum	Maximum
Area (m ²)	150	400
Growing media depth (m)	0.5	1
Gravel storage depth (m)	0.3	0.4

The SWMM – SM – NSGA-II optimiser connection is shown in Figure 4-1. The SWMM model is used to train the set of SMs to predict both runoff volume and TSS at the outlet. These SMs are then called at every optimisation iteration to evaluate the runoff and TSS for each set of BRC design configuration. The third cost objective function is calculated by minimising the cost

objective function equation output, whereas the runoff and TSS objectives are minimised by minimising the value of the SM output.

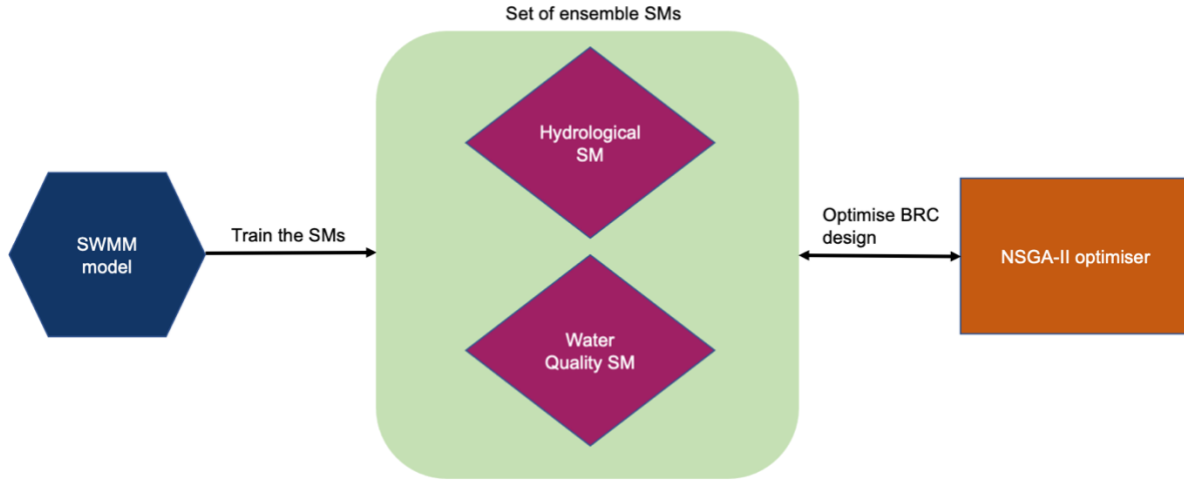


Figure 4-1: SWMM/SM/NSGA-II framework for the BRC optimisation

4.2.4.5. Optimisation parameters

To determine the number of generations, 5 trials with 20, 40, 60, 80, and 100 generations were evaluated for each rainfall event. The Pareto optimal front for each generation iteration was plotted and compared against the others. The best Pareto front, i.e. the one that resulted in the best hydrological and environmental performance whilst maintaining the lowest cost was selected as the best one, and the associated number of generations was used for all future optimisation iterations. For all optimisation iterations, the crossover rate was set to 90% and the mutation rate was set to $1/n$, n being the population number. A random number is first drawn before crossover and mutation occur, and if the number is less than the crossover rate (0.9), crossover occurs. Otherwise, mutation will occur. For this study a population of 100 data points was selected.

4.2.4.6. Design of Experiments (DoE)

The optimisation algorithm performance is dependent on the initial sample population from which it begins its evolution process. The computational efficiency of the optimisation process can be increased through providing an input set that is representative of the entire decision domain (Poles *et al.*, 2009). LHS has been shown to have uniform sampling (Garud *et al.* 2017), assist in faster convergence of the NSGA-II algorithm (Poles *et al.*, 2009) and improves the quality of

solutions for optimisation problems with less than 500 decision variables (Bi *et al.* 2015). Therefore, LHS was used to create the initial population sample of BRC decision variables.

4.3. Results and discussion

4.3.1. Surrogate model reliability

SM performance was evaluated using the averaged NSE index for the testing dataset, which is totally independent from the training and validation set, as well as the CRPS, PC and RH, which were evaluated on the entire dataset for completeness. Results for the 2-year 6-hour, 25-year 12-hour and 100-year 24-hour storm are provided in Figure 4-2 and Table 4-5. The top row of Figure 4-2 shows the SM predicted runoff plotted against an area value, as well as the observed runoff (the SWMM prediction) plotted against the same area value. The light blue filled area in each graph represents the range of the ensemble predictions for that particular area value.

From the results of Figure 4-2 it clear to see that the SM predictions very closely followed the observed values obtained from SWMM. This is indicative of the high accuracy of the SM. Additionally, the RH of each storm event is displayed in the second row. The RH's are generally uniform with a few outliers.

Table 4-5: Hydrologic SM performance results for the 2-year 6-hour, 25-year 12-hour and 100-year 24-hour storms

Design Storm	NSE (-)	CRPS (m³)	PC (%)
2-year 6-hour	0.96	0.008	96
25-year 12-hour	0.99	0.007	92
100-year 24-hour	0.99	0.01	85

The results in Table 4-5 are indicative of accurate model performance as well. NSE for all three design storms is very high, with an average value of 0.98 across all three storms. Similarly, CRPS has very low values indicating accurate model ensemble performance. PC is the lowest for the 100-year 24-hour storm with a value of 85%, however 85% is still considered very good performance, capturing 85% of all points within the ensemble range. These results indicate that the SM can be reliably used in place of SWMM in the optimisation-simulation framework.

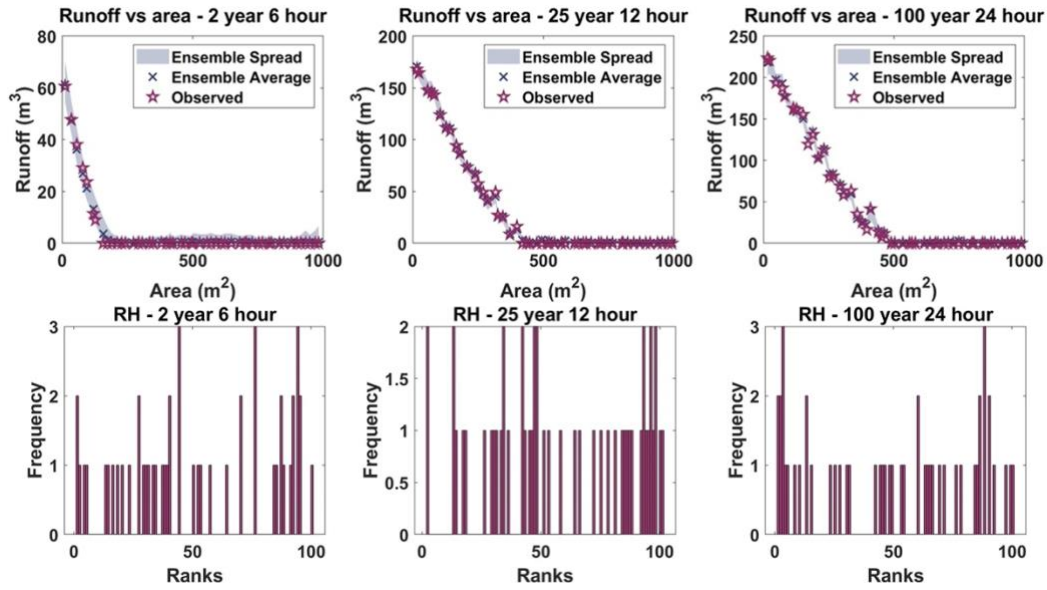


Figure 4-2: Hydrologic SM performance for the 2-year 6-hour, 25-year 12-hour and 100-year 24-hour storms design storms.

Results for the Environmental SM (predicting mass of TSS remaining in effluent) for an EMC of 130 mg/L are also given in Figure 4-3 and Table 4-6. These results are very similar to those of the Hydrologic SM, suggesting that the environmental SM can also be reliably used in the simulation-optimisation framework to predict TSS. From these results, the 2-year 6-hour Environmental SM has the lowest NSE value equalling 0.88, which is also lower than it's Hydrological counterpart, having an NSE of 0.96. Given the large ensemble range shown in Figure 4-3, it can be expected that the SM underperforms in terms of NSE and CRPS. This performance can be improved by retraining the SM until better performance is achieved. However, it should also be noted that due to the small range between the starting value of TSS and the maximum value of TSS for the 2-year 6-hour, the ensemble range appears greater than the other storms with much wider ranges. Therefore, the ensemble ranges appear to be smaller in the other storms due to the large range in the vertical axis, when in fact, the 2-year 6-hour storm differences in values are very minimal, as indicated by its high performance indicated by the performance metrics. It should be noted that the TSS values fluctuate and do not gradually decline with increasing area as would be predicted due to the values of the other two design parameters (growing media depth and gravel storage depth). For example, for the 100-year 24-hour storm the penultimate data point, the BRC area is 997 m² with a growing media depth of 103 mm and a gravel storage depth of 507 mm. Conversely, the previous data point corresponds to a BRC with

an area of 979 m² with a growing media depth and gravel storage depth of 1629 mm and 468 mm respectively. Therefore, although the areas are very similar, the other design parameters not shown contribute greatly to the value of TSS at the outlet.

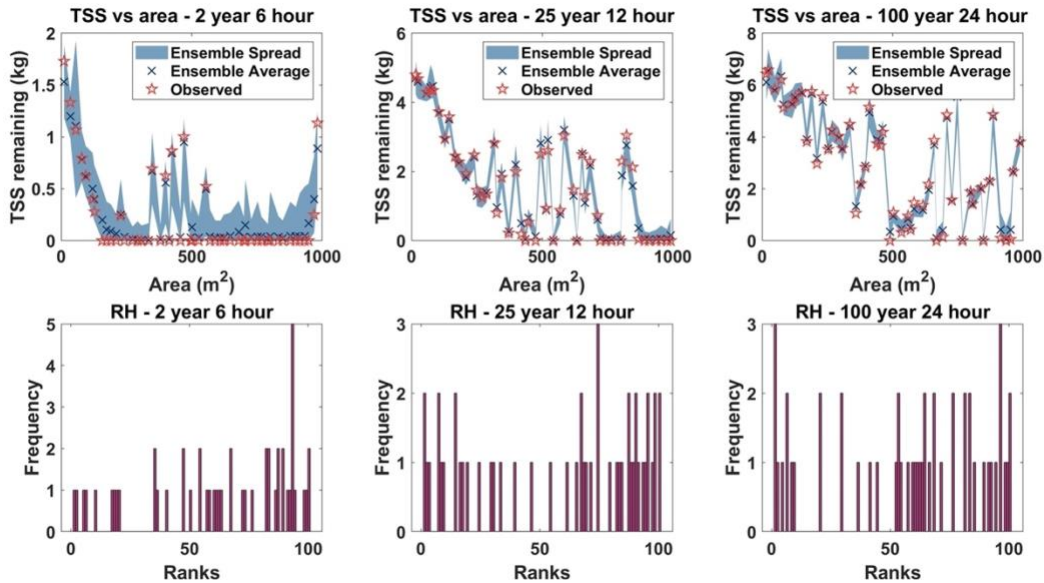


Figure 4-3: Environmental SM performance for the 2-year 6-hour, 25-year 12-hour and 100-year 24-hour storms design storms. The top row contains the observed and predicted plots. The bottom row contains the corresponding RHs

Table 4-6: Environmental SM performance results for the 2-year 6-hour, 25-year 12-hour and 100-year 24-hour storms

Design Storm	NSE (-)	CRPS (kg)	PC (%)
2-year 6-hour	0.88	0.05	98
25-year 12-hour	0.94	0.12	93
100-year 24-hour	0.99	0.07	92

4.3.2. Generation number selection

Since there is no guidance in literature on selecting the optimal number of generations for the optimisation algorithm, a sensitivity analysis was performed. The number of generations was varied from 20 to 100 in increments of 20. The results of this analysis are presented in Figure 4-4.

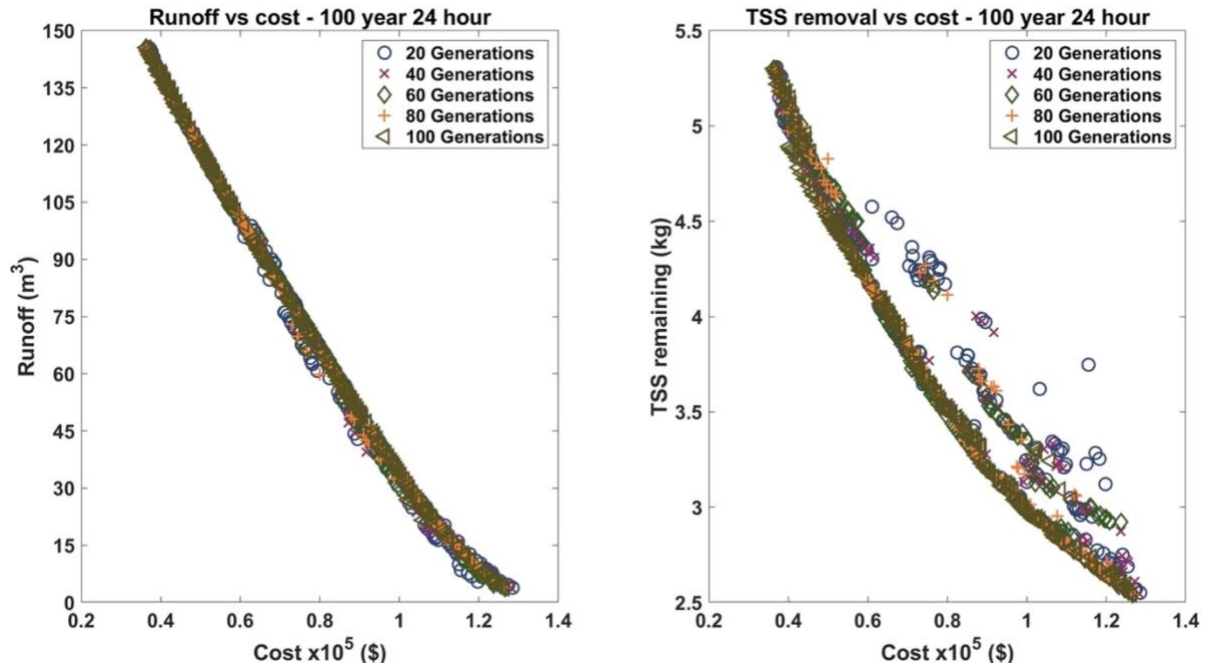


Figure 4-4: Comparison of Pareto fronts under different generations

As shown in Figure 4-4 there is very minimal difference in Pareto optimal fronts between any of the generations, and therefore, 100 generations will be used for the optimisation algorithm.

4.3.3. Optimisation under current design guidelines

To determine whether BRC's designed under the current design recommendations will have sufficient hydrologic and environmental performance under future climate change, the optimised results under current and future climate were compared. Results for the 2-year 6-hour, 25-year 12-hour and 100-year 24-hour storm are shown in Figure 4-5 - Figure 4-7 respectively. Results for the remaining storms are presented in Appendix B.

The optimised results are presented for the near and extended future periods, as well as the historical climate record. All three IPCC climate scenarios are presented in the results as well. BRC performance is given in the vertical axis for runoff retention and pollutant removal. Under the extended and near future period respectively.

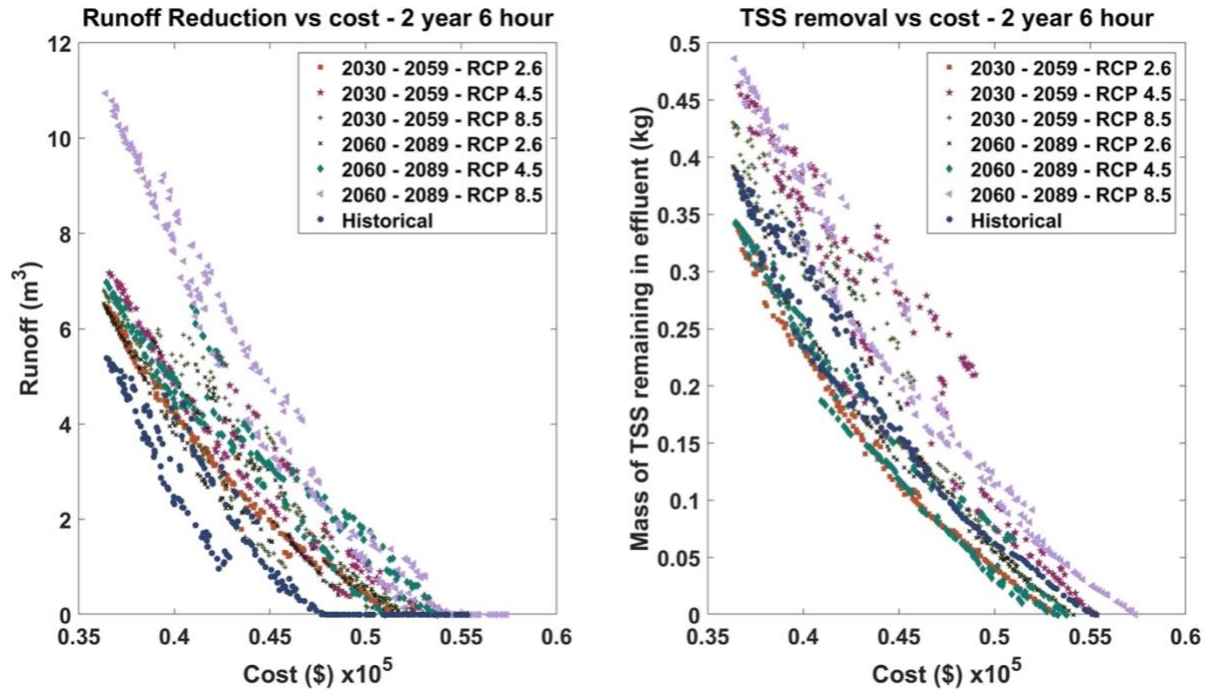


Figure 4-5: BRC Pareto fronts for the 2-year 6-hour storm for future and current climate under current design guidelines

For the 2-year 6-hour storm, BRC performance is similar for both the historical and two future periods. The smallest BRC built according to the design guidelines under historical conditions had an area of 150 m² with a growing media and gravel storage depth of 500 mm and 300 mm respectively, resulted in 6 m³ of runoff and 0.39 kg of TSS at the outlet for an EMC of 130 mg/L. Conversely the BRC resulting in 0 m³ of runoff and 0 kg of TSS had an area of 206 m² with a growing media and gravel storage depth of 870 mm and 302 mm respectively. The BRC designed under all the other future periods, had similar designs of areas approximately 200 m², 850 mm growing media depth and 320 mm of gravel storage depth. Evidently, BRCs designed under the current design guideline recommendations have sufficient performance under climate change for smaller storms.

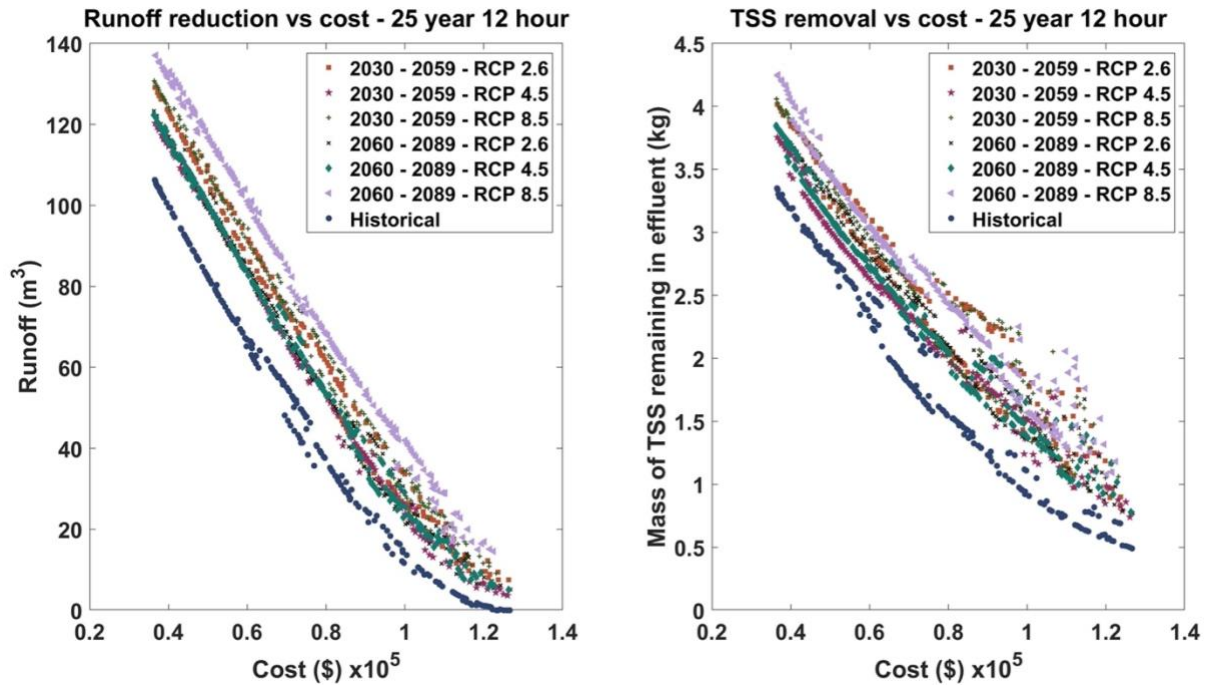


Figure 4-06: BRC Pareto fronts for the 25-year 12-hour storm for future and current climate under current design guidelines

For the 25-year 12-hour storm, the BRC performance starts decreasing under future climate conditions, and does not achieve the same level of performance as it does for the historical climate conditions. Performance for similar BRC designs with an approximate area of 400 m² and growing media and gravel storage depths of approximately 1 m and 300 mm respectively was compared. Under the extended future period RCP 8.5, the runoff and TSS leaving the BRC was 27 m³ and 1.31 kg respectively. Similarly, under the near future period, the runoff and TSS leaving the BRC was 21 m³ and 2.87 kg respectively. Finally, under the historical period, the runoff and TSS values were 5.4 m³ and 0.71 kg respectively.

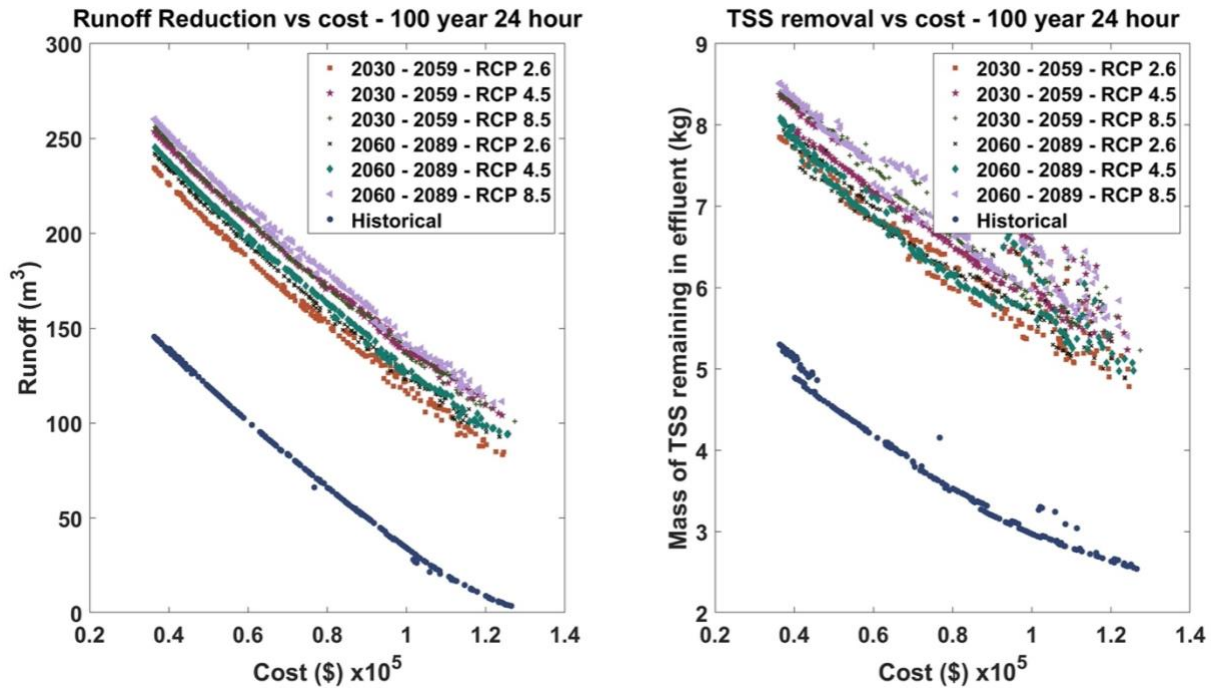


Figure 4-7: BRC Pareto fronts for the 100-year 24-hour storm for future and current climate under current design guidelines

These differences in performance become more pronounced under the 100-year 24-hour storm period. For similar BRC design configurations as previously mentioned, i.e. the largest possible recommended design, the RCP 8.5 extended future period results in a runoff volume of 129 m³ and 5.67 kg of TSS. Similarly, under the near future period RCP 8.5 scenario, the runoff volume was 96 m³ and 5.58 kg of TSS. Evidently, past the 25-year return period the BRC performance begins to decrease. This is indicative of the need to increase the recommended design guideline limits as they do not provide adequate performance in the future. BRC performance under the largest possible optimised design of an area of approximately 400 m² with a growing media depth of 1 m and a gravel storage depth of 300 mm is given in Table 4-7, along with the historic performance for comparison. It is clear to see that if the design guidelines are not updated to reflect the changing climate, BRC performance is significantly diminished.

Table 4-7: BRC hydrologic and environmental performance under historic and all future climate scenarios for the largest BRC design possible

		Historical	2060 – 2089 RCP 8.5	2060 – 2089 RCP 4.5	2060 – 2089 RCP 2.6	2030 – 2059 RCP 8.5	2030 – 2059 RCP 4.5	2030 – 2059 RCP 2.6
100- year 24- hour	Runoff (m ³)	21	132	114	111	126	126	105
	TSS (kg)	2.79	5.88	5.33	5.16	5.58	5.55	5.23
25- year 12- hour	Runoff (m ³)	6	27	15	15	21	15	21
	TSS (kg)	0.71	1.31	1.06	1.02	1.14	1.12	1.08
2-year 6-hour	Runoff (m ³)	0	0	0	0	0	0	0
	TSS (kg)	0	0	0	0	0	0	0

To assess the increases needed to achieve past performance, the design limits of the decision variables were increased by 50%, meaning that the maximum allowable area was set to 600 m², the growing media depth was set to 1.5 m and the gravel storage was set to 600 mm. These changes are chosen to update the current design lines to ensure resiliency against future climate change.

4.3.4. BRC performance under adapted design guidelines

Results of BRC performance following the increase of recommended design limits for the decision variables are presented in Figure 4-8 and Figure 4-10 for the three design storms as previous.

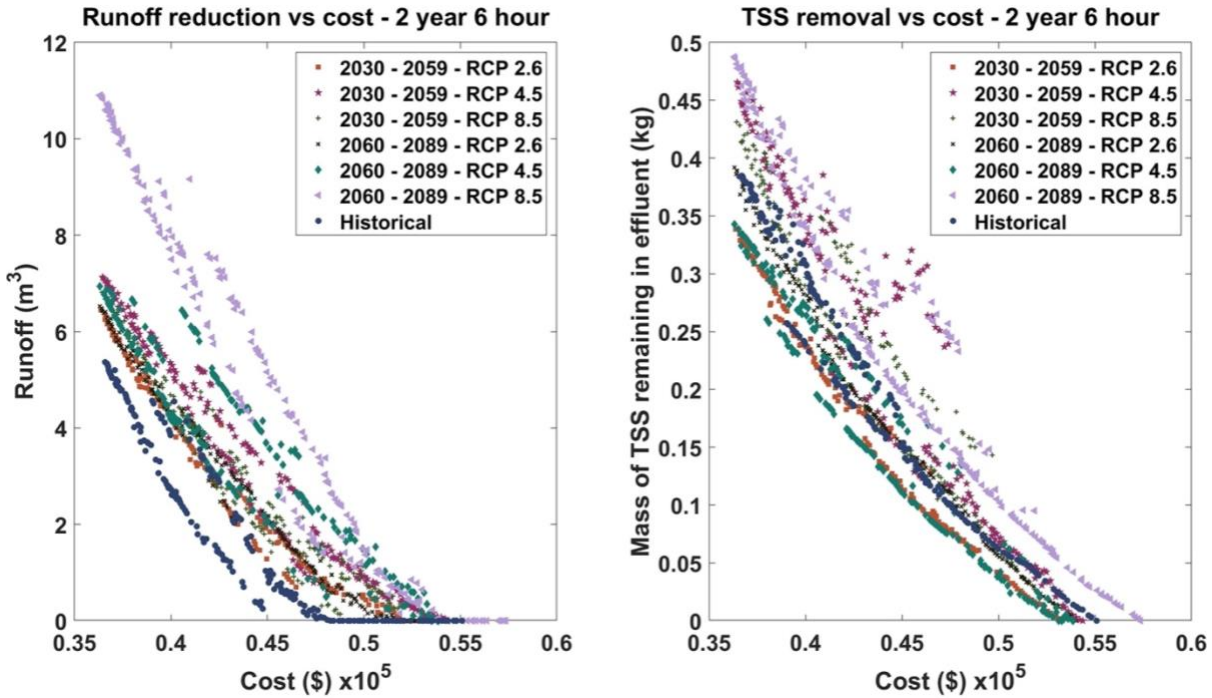


Figure 4-8: BRC performance under design recommendations adapted for climate change for the 2-year 6-hour

For the 2-year 6-hour storm, following the original design guideline recommendations resulted in sufficient BRC performance. Adapting the design guidelines to double the upper limits of the decision variables results in similar performance for runoff retention, however, TSS removal is improved. This is made possible due to the increases in growing media depth. For example, under the original design guidelines for the RCP 8.5 extended future scenario, to achieve a runoff volume of 28.8 m³ and 1.28 kg of TSS at the outlet, a BRC with an area of 400 m², a growing depth of 1 m and a gravel storage layer depth of 300 mm is necessary. Under the adapted guidelines, for the same scenario, a BRC with an area of 393 m², a growing media depth of 1.034 m and a storage depth layer of 326 mm results in 28 m³ runoff and 0.88 kg of TSS at the outlet. Therefore, under the adapted guidelines, this results in a 68% reduction of TSS at the outlet for a 7% increase in cost. Additionally, comparing Figure 4-8 with Figure 4-5, it is clear to see that the cost of the BRC built for the near future and extended future under the adapted design guidelines is similar to the original guidelines. For example, under original guidelines near future RCP 8.5 scenario, to achieve 0 m runoff and 0 kg TSS, the BRC had an area, growing media depth and gravel storage of 194 m², 0.98 m and 308 mm respectively for a total cost of \$53,462. Under the adapted guidelines, for the same performance, the BRC had an area, growing media depth and gravel storage depth of 193 m², 1.04 m and 318 mm respectively for a total cost of \$53,991.

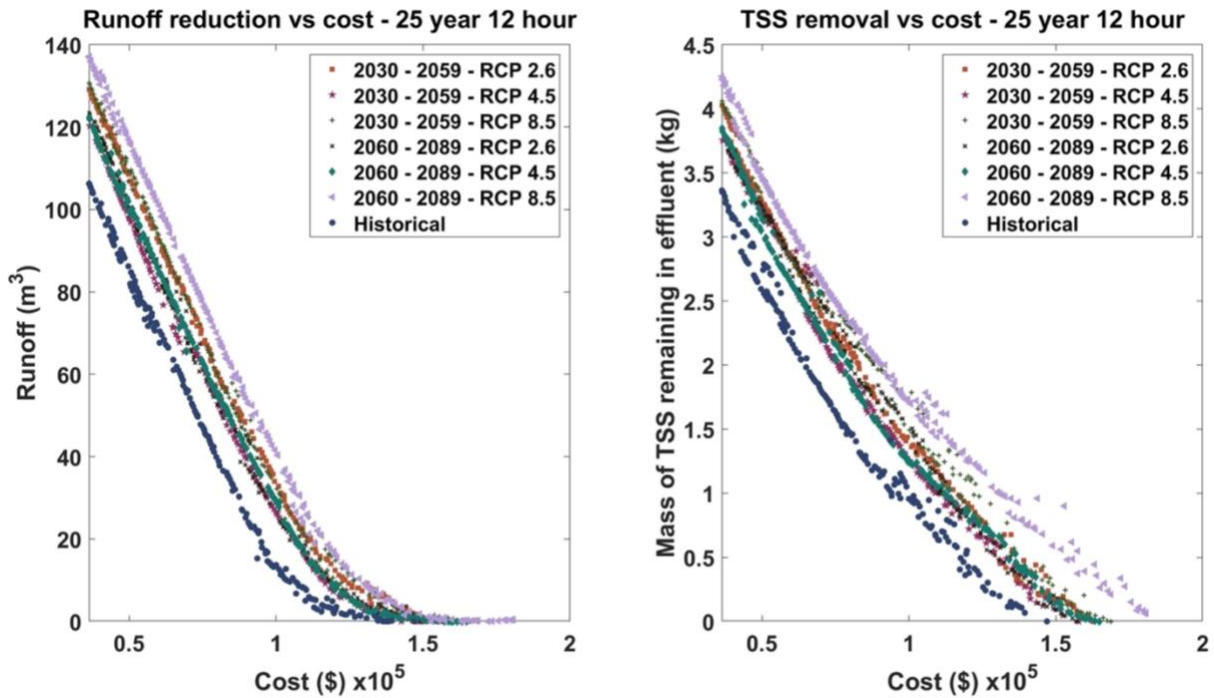


Figure 4-9: BRC performance under design recommendations adapted for climate change for the 25-year 12-hour

For the 25-year 12-hour storm, the increase to the decision variables' upper limits results in similar BRC performance under future and historic scenarios, both designs achieving 100% runoff attenuation and TSS removal. This was not the case under the original design guideline recommendations, as shown in Figure 4-06. To achieve the 100% performance under the historic climate scenario for the adapted guidelines, a BRC area of 481 m² with a growing media depth of 1.39 m and a gravel storage depth of 333 mm is required, with a cost of \$146,803. To achieve similar results under the extended future RCP 8.5 scenario, an increase of 23% in area, 0.72% in growing media depth and no increase in gravel storage is required, with a cost increase of 23%. Similarly, under the near future scenario an increase of 13% in area, 3% in growing media depth and 27% in gravel storage is required, for an increase in cost of 14%. These results demonstrate that the difference in cost between BRC designed for the extended and near future scenario is negligible. Additionally, under the original design guideline recommendations for the RCP 8.5 extended future scenario, to achieve a runoff volume of 27 m³ and a TSS mass of 1.31 kg at the outlet, the cost of the BRC would be \$109,540, for an area of 400 m², a growing media depth of 1 m and a gravel storage of 300 mm. Under the adapted design guidelines, to achieve the same performance in runoff flowrate, and a TSS mass of 1.19 kg, the BRC was built to have an area of 401 m² with a growing media depth of 1.03 m and a gravel storage layer depth of 413 mm. This

amounted to a cost of \$113,863, which is a 3% increase in cost from the original design for the same runoff retention performance, and improved TSS removal.

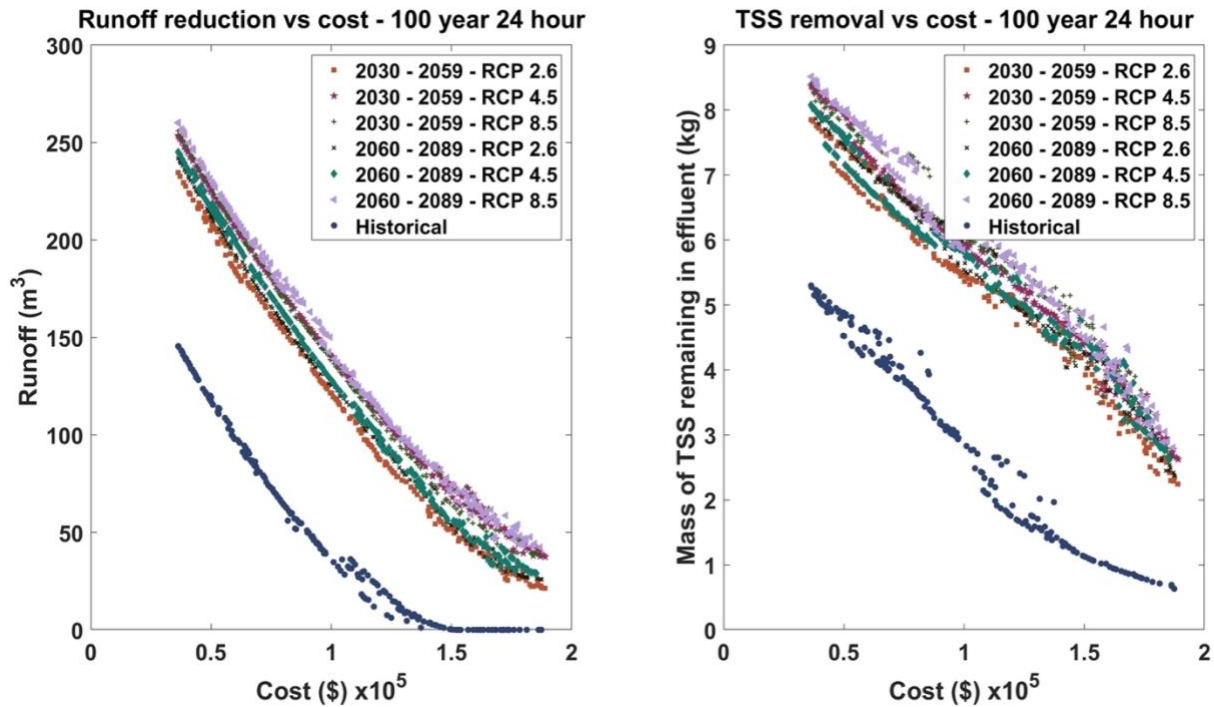


Figure 4-10: BRC performance under design recommendations adapted for climate change for the 100-year 24-hour

Results for the 100-year 24-hour storm, shown in Figure 4-10, reveal that increasing the upper recommended design limits of the decision variables achieves better performance than under original design guidelines. A comparison between Figure 4-10 and Figure 4-7 reveal that BRC performance under both climate change periods, for all three scenarios, improves when compared to the original design guideline performances in Figure 4-7. The biggest possible BRC constructed with an area of 400 m², growing media depth of 1 m and gravel storage of 300 mm achieved a runoff volume and TSS value of 20.1 m³ and 2.79 kg at the outlet respectively, for the historical original design guidelines scenario for a cost of \$109,911. While this past performance is not achieved under the RCP 8.5 extended future scenario, building the biggest BRC with an area of 600 m², growing media depth of 1.5 m and gravel storage of 353 mm achieved a runoff volume and TSS value of 42 m³ and 2.8 kg at the outlet respectively with a cost of approximately \$187,769. A BRC achieving 42 m³ and 3.2 kg under historical conditions, with an area of 349 m², growing media depth of 0.89 m and gravel storage of 347 mm was necessary, for a cost of \$92,555. Therefore, an increase of 41% in area and growing media depth, and a 2% increase in

gravel storage depth is necessary to achieve past performance under future climate, with a cost increase of 102%.

Further, under the proposed increases, there is a 97% reduction of TSS mass, whereas previously it was approximately 95%. Similarly, for runoff retention, there is an 89% reduction of runoff flowrate, whereas previously it was approximately 71%. However, the relative performance difference between the existing and proposed recommendations is approximately 61%. Meaning the maximum achieved runoff flowrate was 108 m³ previously under the extended RCP 8.5 scenario whereas currently it is close to 42 m³. This also results in an increase of approximately 70% in costs when the original designs are compared to the updated designs.

These results highlight that there is a cost increase to achieve past performance under future climate for the large return storms such as the 100-year 24-hour storm, however, studies have shown that proactive mitigation against future climate may be cheaper than reactive solutions at the time (Visitation *et al.*, 2009). Additionally, as seen from Figure 4-8 to Figure 4-10, the difference in cost between BRC designs for the different RCP scenarios is negligible. Therefore, it is recommended that BRCs be designed for the most severe RCP 8.5 scenario to ensure resilience against the other two scenarios.

Results for the 2-year 6-hour storm indicate that an increase to current design guideline recommendation is not necessary as BRCs are still able to achieve 100% runoff retention and TSS removal. However, doubling the upper allowable limits resulted in cost savings not previously seen, which is beneficial. For the 25-year increases of approximately 50% in the maximum allowable design limits result in past performance under climate change. Meaning that BRC performance achieved under the historic scenario is also achieved under the RCP 8.5 climate change scenario. The 50% increase in allowable limits however, resulted in a 100% runoff retention and TSS removal under the 25-year storm, which was previously not possible. This was not the case for the 100-year storm, however performances of 89% runoff retention and 97% TSS removal were obtained. Therefore, it is strongly recommended that increases in design guidelines be applied to ensure resiliency against impending climate impacts.

Results for all other design storms are given in Appendix B. Apart from the 50-year and 100-year storms, all other design storms achieved 100% runoff retention and TSS removal under future

climate change. The 50-year and 100-year storms achieved greater than 50% improvement in performance in comparison with the original design guidelines under future climate change.

4.4. Conclusion

Urbanisation and climate change are increasing the risks of localised flooding in urban cities. These risks can be mitigated through the implementation of adaptive measures such as BRCs. To ensure the constructed BRCs are capable to mitigating future climate conditions, their current designs must be evaluated under future climate conditions. Additionally, to reach this optimal BRC design under future climate conditions, several iterations of BRC design modifications are necessary. These can be time-consuming as they are quite often done using physical-based models. Therefore, SMs such as ANNs are used to expedite this process. Given the stochastic nature of ANNs, their performance must be tested to ensure accurate predictions are obtained.

This chapter evaluated BRC performance under current design guidelines for current and future climate change scenarios. This was done through using an ensemble of ANNs as an SM to SWMM in a simulation-optimisation framework. The NSGA-II optimisation framework was used to maximise BRC runoff retention, TSS removal and minimise design cost. Three decision variables, namely BRC area, growing media depth and gravel storage depth were used as the decision variables and varied in the optimisation algorithm to achieve optimised BRC designs. The reliability of the ANN to be used in SWMM was evaluated to ensure the ANN maintained high predictive accuracy. The ANN performance metrics exhibited high levels of accuracy, and the ensemble was well-calibrated indicating its reliability for use in place of SWMM.

Results of the optimised BRC designs under current design guideline recommendations indicated sufficient performance for the 2-year 6-hour storm under historic and both extended and near future scenarios, achieving 100% runoff retention and TSS pollutant removal. BRC performance began to diminish under the 25-year and 100-year storms for future climate conditions. Past performance was no longer achieved, and there was a large relative difference in performance under the climate change and historic scenarios. To remedy this, the maximum allowable design limits for the decision variables expressed in design guidelines were increased. The maximum area was increased from 400 m² to 600 m², the maximum growing media depth of 1 m was increased to 1.5 m and finally the recommended 300 mm of gravel storage depth was increased to 600 mm.

To achieve past performance under the extended climate change scenario RCP 8.5, an increase of approximately 23% in area, 13% and 3% increase in growing media depth for the 25-year and 100-year storms respectively and a 20% increase in gravel storage is necessary. Further, to achieve the same performance under the new measures relative to existing guidelines, BRCs with smaller areas and a deeper growing media depth were constructed. Additionally, 100% runoff retention and TSS removal was achieved for the 25-year storm under all future climate scenarios, which was not possible under the existing design guidelines. For the 100-year storm, a total reduction of 89% in runoff flow rate and 97% in TSS removal was observed under the new design guidelines. Further analysis into the performance differences between the existing and new design guidelines revealed an increase of approximately 60% in both runoff retention and TSS removal under the new design guidelines.

Therefore, this research demonstrated the reliability of using SMs in place of SWMM in a simulation-optimisation framework. Additionally, it demonstrated that existing BRC design guidelines do not provide sufficient performance under impending climate change impacts. It is therefore strongly recommended to update design guidelines to ensure BRC design is adapted and resilient in the face of climate change.

This research revealed the necessary additions to existing design guidelines to ensure BRCs are designed to be resilient, achieving 100% runoff retention and TSS removal under future climate. The increases suggested were not sufficient to achieve 100% performance under the 50-year and 100-year storms. Existing design guidelines can be redrafted to reflect the necessary modifications identified in this research. Doing so will ensure that BRCs designed for Toronto city are able to mitigate the alleviated risk of urbanised flooding due to urbanisation and climate change.

Chapter 5 - Conclusion

Urbanisation, coupled with the impacts of increased extreme rainfall events due to climate change, is exacerbating the risk of urban flooding in cities. Existing urban drainage networks are aging, undersized and expensive to replace, further contributing to this risk. A potential solution is through the use of Low Impact Development (LID) units, such as Bioretention cells (BRCs), which aim to mimic the predevelopment hydrology of the catchment. BRC design is a highly iterative and time-consuming process, especially when done using a physical-based model. An alternative to using physical-based models is the use of Surrogate Models (SMs), a topic which is explored in chapter 3. BRC performance is also highly dependent on their design configuration, which can be improved using an optimisation algorithm. Optimisation algorithms can also be used to design BRCs under future climate scenarios to determine the necessary increases to existing designs to ensure past performance under climate change, a topic explored in chapter 4.

Chapter 3 provided a thorough overview of the development of a Surrogate Model (SM) and the impacts varying model characteristics has on its performance. The impact number of training samples, method of datasplit sampling and input selection has on model performance was quantified. An ensemble of Artificial Neural Networks (ANN) was trained as the SM to the Stormwater Management Model (SWMM). Model performance was assessed through comparing the difference between SM runoff and TSS predictions and SWMM predictions, to evaluate the reliability of using SM's in place of SWMM. 18 rainfall design storms following the Atmospheric Environmental Services – Southern Ontario (AES-SO) distribution were used.

The number of training samples was varied from 50, 100 to 200 data points per rainfall event. Results indicated that using only 50 data points per rainfall event as opposed to 100 or 200 was equally sufficient, with no clear indication of superior model performance for an increasing number of data points. Ultimately, this indicates that the SM can maintain high levels of prediction accuracy without needing a large number of training points. This is beneficial in a SM approach, as SMs are trained on a set of known inputs obtained from the original physical model, therefore, needing less training points results in significant computational time savings.

Flow-based and event-based stratified sampling was used and compared against the efficacy of using random sampling. A 60% stratification was used to ensure the training dataset had 60% of the high flows (50 and 100-year storm) and 60% of the low flows (flow-based) or 60% of the design configurations from each rainfall event (event-based). Results indicated no model improvement using stratified sampling in comparison to MATLAB's built-in random sampling. Finally, the impact of input selection on model performance was evaluated. Two additional inputs, average and peak rainfall intensity, were included into the input set. Model performance greatly improved, and the model no longer suffered from underprediction bias. These findings highlighted the importance of thoroughly investigating SM model characteristics and their impacts on its performance. Additionally, a comparison between universal and event-based SM performance revealed similar performance among both models, with the universal SM having slightly superior performance.

Chapter 4 evaluated BRC performance following existing design guidelines for current and future climate scenarios. Similar to Chapter 3, an ensemble of ANNs was used as an SM to SWMM in a simulation-optimisation framework. The reliability of the ANN in place of SWMM was quantified through different performance metrics, with results indicated high levels of predictive accuracy and a well-calibrated ensemble. Next, using the NSGA-II optimisation algorithm, three decision variables, namely BRC area, growing media depth and gravel storage depth, were varied to obtain the optimal BRC design that maximises runoff retention, TSS removal and minimises cost.

Results of BRC performance under existing design guidelines indicated sufficient performance for the 2-year 6-hour storm under all climate scenarios, achieving 100% runoff retention and TSS removal. However, performance started to diminish under the 25-year 12-hour storm and 100-year 24-hour storm. Performance achieved under the historic scenario was no longer possible for both design storms under future climate scenarios. To determine the necessary increases to existing design guideline recommendations to achieve past performance, the upper maximum limits of the decision variables was doubled.

The proposed increases resulted in 100% runoff retention and pollutant removal for the 25-year storm, which was previously not possible. While achieving 100% BRC performance was not possible under the 100-year storm, results indicated a performance increase of approximately 50% in both runoff retention and TSS removal under the new increases relative to the existing guidelines.

5.1. Limitations

The lack of SWMM model validation presents a limitation in this research as the results obtained for the BRC performance are entirely dependent on the setup of the physics-based model. It is important to note however that this study underlines the relative changes needed to the current design guidelines in the face of future changing climate. Therefore, if the BRC absolute performance were to decrease or increase due to model validation, the relative percent changes in BRC design parameters should remain the same. Another limitation of this research is the assumption that external factors such as pollutant loading and rate of urbanisation remain the same under future climate. An increase in these external factors should be implemented to reflect this, or uncertainty bounds relating to the uncertainty of these parameters in the future should be implemented. Finally, the use of synthetic design storms, though a widespread practice for BRC design, presents a limitation. This can be addressed by implementing historical storms that can be scaled up to reflect future climate as they present a more realistic storm distribution.

5.2. Future recommendations

Future research opportunities include investigating the additional necessary increases in design guideline recommendations to achieve 100% performance under the 100-year and 50-year storms. Additionally, these increases should also be investigated for continuous modelling, to investigate the impact of antecedent moisture on BRC performance.

Finally, recommendations for the SM research work includes focusing on improving ensemble dispersion, perhaps by altering the neural network architecture. Prior to the addition of the additional inputs, investigations into increasing the number of hidden neurons, adding an additional hidden layer, varying the initial weights and biases, and using a different sampling technique for datasplitting should all be investigated. This coupled with including the additional inputs can help eliminate some of the uncertainty surrounding model architecture.

References

- Ahiablame, L. M., Engel, B. A., & Chaubey, I. (2012). Effectiveness of low impact development practices: Literature review and suggestions for future research. *Water, Air, and Soil Pollution*, 223(7), 4253–4273. <https://doi.org/10.1007/s11270-012-1189-2>
- Ahiablame, L. M., Engel, B. A., & Chaubey, I. (2013). Effectiveness of low impact development practices in two urbanized watersheds: Retrofitting with rain barrel/cistern and porous pavement. *Journal of Environmental Management*, 119, 151–161. <https://doi.org/10.1016/j.jenvman.2013.01.019>
- Anandhi, A., Frei, A., Pierson, D. C., Schneiderman, E. M., Zion, M. S., Lounsbury, D., & Matonse, A. H. (2011). Examination of change factor methodologies for climate change impact assessment. *Water Resources Research*, 47(3), 1–10. <https://doi.org/10.1029/2010WR009104>
- Bacys, M., Khan, U. T., Sharma, J., & Bentzen, T. R. (2019). Hydrological Efficacy of Ontario's Bioretention Cell Design Recommendations: A Case Study from North York, Ontario. *Journal of Water Management Modeling*, 47. <https://doi.org/10.14796/jwmm.c468>
- Barah, M., Khojandi, A., Li, X., Hathaway, J., & Omitaomu, O. F. (2021). Optimizing green infrastructure placement under precipitation uncertainty. *Omega*, 100. <https://doi.org/10.1016/j.omega.2020.102196>
- Batalini de Macedo, M., Ambrogi Ferreira do Lago, C., Mendiondo, E. M., & H. Giacomoni, M. (2019). Bioretention performance under different rainfall regimes in subtropical conditions: A case study in São Carlos, Brazil. *Journal of Environmental Management*, 248. <https://doi.org/10.1016/j.jenvman.2019.109266>
- Bedan, E. S., & Clausen, J. C. (2009). Stormwater Runoff Quality and Quantity From Traditional and Low Impact Development Watersheds. *JAWRA Journal of the American Water Resources Association*, 45(4), 998–1008. <https://doi.org/10.1111/j.1752-1688.2009.00342.x>
- Behera, P., Li, J., & Adams, B. J. (2000). Characterization of Urban Runoff Quality: A Toronto Case Study. *Journal of Water Management Modeling*, 8. <https://doi.org/10.14796/jwmm.r206-14>
- Bi, W., Dandy, G. C., & Maier, H. R. (2015). Improved genetic algorithm optimization of water distribution system design by incorporating domain knowledge. *Environmental Modelling and Software*, 69, 370–381. <https://doi.org/10.1016/j.envsoft.2014.09.010>
- Boancă, P., Dumitraș, A., Luca, L., Bors-Opriș, S., & Laczi, E. (2018). Analysing bioretention

- hydraulics and runoff retention through numerical modelling using RECARGA: A case study in a Romanian urban area. *Polish Journal of Environmental Studies*, 27(5), 1965–1973. <https://doi.org/10.15244/pjoes/79271>
- Bosley, E. K. (2008). *Hydrologic Evaluation of Low Impact Development Using a Continuous, Spatially-Distributed Model*. Virginia Tech. <https://vtechworks.lib.vt.edu/handle/10919/34263>
- Boucher, M. A., Perreault, L., & Anctil, F. (2009). Tools for the assessment of hydro logical ensemble forecasts obtained by neural networks. *Journal of Hydroinformatics*, 11(3–4), 297–307. <https://doi.org/10.2166/hydro.2009.037>
- Brown, R. A., Skaggs, R. W., & Hunt, W. F. (2013). Calibration and validation of DRAINMOD to model bioretention hydrology. *Journal of Hydrology*, 486, 430–442. <https://doi.org/10.1016/j.jhydrol.2013.02.017>
- Brunetti, G., Šimůnek, J., Turco, M., & Piro, P. (2017). On the use of surrogate-based modeling for the numerical analysis of Low Impact Development techniques. *Journal of Hydrology*, 548, 263–277. <https://doi.org/10.1016/j.jhydrol.2017.03.013>
- City of Toronto. (2006). *Wet Weather Flow Management*. [https://doi.org/10.1061/40792\(173\)167](https://doi.org/10.1061/40792(173)167)
- City of Toronto. (2017). *Toronto Green Streets Technical Guidelines*.
- Credit Valley Conservation (CVC), & Toronto and Region Conservation Authority (TRCA). (2010). *Low Impact Development Stormwater Management Planning and Design Guide*. https://cvc.ca/wp-content/uploads/2014/04/LID-SWM-Guide-v1.0_2010_1_no-appendices.pdf
- Cubasch, U., Wuebbles, D., Chen, D., Facchini, M. C., Frame, D., Mahowald, N., & Winther, J.-G. (2013). Introduction. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, & P. M. Midgley (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 119–159). Cambridge University Press.
- Davis, A. P. (2005). Green engineering principles promote low-impact development. *Environmental Science and Technology*, 39(16). <https://doi.org/10.1021/es053327e>
- Davis, S. E., Cremaschi, S., & Eden, M. R. (2018). Efficient Surrogate Model Development: Impact of Sample Size and Underlying Model Dimensions. *Computer Aided Chemical Engineering*, 44, 979–984. <https://doi.org/10.1016/B978-0-444-64241-7.50158-0>
- Dawson, C. W., & Wilby, R. L. (2001). Hydrological modelling using artificial neural networks. *Progress in Physical Geography: Earth and Environment*, 25(1), 80–108.

<https://doi.org/10.1177/030913330102500104>

- Deb, K. (2014). Multi-objective Optimization. In E. K. Burke & G. Kendall (Eds.), *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques* (pp. 403–449). Springer US. https://doi.org/10.1007/978-1-4614-6940-7_15
- Deb, K., Agrawal, S., Pratap, A., & Meyarivan, T. (2000). A Fast Elitist Non-dominated Sorting Genetic Algorithm for Multi-objective Optimization: NSGA-II. In M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo, & H.-P. Schwefel (Eds.), *Parallel Problem Solving from Nature PPSN VI* (pp. 849–858). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45356-3_83
- Dietz, M. E. (2007). Low impact development practices: A review of current research and recommendations for future directions. *Water, Air, and Soil Pollution*, 186(1–4), 351–363. <https://doi.org/10.1007/s11270-007-9484-z>
- Dietz, M. E., & Clausen, J. C. (2008). Stormwater runoff and export changes with development in a traditional and low impact subdivision. *Journal of Environmental Management*, 87(4), 560–566. <https://doi.org/10.1016/j.jenvman.2007.03.026>
- Eckart, K., McPhee, Z., & Bolisetti, T. (2017). Performance and implementation of low impact development – A review. *Science of the Total Environment*, 607–608, 413–432. <https://doi.org/10.1016/j.scitotenv.2017.06.254>
- Eckart, K., McPhee, Z., & Bolisetti, T. (2018). Multiobjective optimization of low impact development stormwater controls. *Journal of Hydrology*, 562, 564–576. <https://doi.org/10.1016/j.jhydrol.2018.04.068>
- Elliott, A. H., & Trowsdale, S. A. (2007). A review of models for low impact urban stormwater drainage. *Environmental Modelling and Software*, 22(3), 394–405. <https://doi.org/10.1016/j.envsoft.2005.12.005>
- Fahey, D. W., Doherty, S. J., Hibbard, K. A., Romanou, A., & Taylor, P. C. (2017). Physical drivers of climate change. In D. J. Wuebbles, D. W. Fahey, K. A. Hibbard, D. J. Dokken, B. C. Stewart, & T. K. Maycock (Eds.), *Climate Science Special Report: Fourth National Climate Assessment, Volume I* (pp. 73–113). U.S. Global Change Research Program. <https://doi.org/10.7930/J0513WCR>
- Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., & Rummukainen, M. (2013). Evaluation of Climate Models. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, & P. M. Midgley (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of*

- Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 741–867). Cambridge University Press.
- Gao, Y., Babin, N., Turner, A. J., Hoffa, C. R., Peel, S., & Prokopy, L. S. (2016). Understanding urban-suburban adoption and maintenance of rain barrels. *Landscape and Urban Planning*, 153, 99–110. <https://doi.org/10.1016/j.landurbplan.2016.04.005>
- Garud, S. S., Karimi, I. A., & Kraft, M. (2017). Design of computer experiments: A review. *Computers and Chemical Engineering*, 106, 71–95. <https://doi.org/10.1016/j.compchemeng.2017.05.010>
- Gaur, A., Gaur, A., Yamazaki, D., & Simonovic, S. P. (2019). Flooding Related Consequences of Climate Change on Canadian Cities and Flow Regulation Infrastructure. *Water*, 11(1), 63. <https://doi.org/10.3390/w11010063>
- Ghodsi, S. H., Zahmatkesh, Z., Goharian, E., Kerachian, R., & Zhu, Z. (2020). Optimal design of low impact development practices in response to climate change. *Journal of Hydrology*, 580, 124266. <https://doi.org/10.1016/j.jhydrol.2019.124266>
- Goulden, S., Portman, M. E., Carmon, N., & Alon-Mozes, T. (2018). From conventional drainage to sustainable stormwater management: Beyond the technical challenges. *Journal of Environmental Management*, 219, 37–45. <https://doi.org/10.1016/j.jenvman.2018.04.066>
- Government of Canada. (2019). *High Resolution Digital Elevation Model (HRDEM) - CanElevation Series - Open Government Portal*. <https://open.canada.ca/data/en/dataset/957782bf-847c-4644-a757-e383c0057995>
- He, Z., & Davis, A. P. (2011). Process Modeling of Storm-Water Flow in a Bioretention Cell. *Journal of Irrigation and Drainage Engineering*, 137(3), 121–131. [https://doi.org/10.1061/\(asce\)ir.1943-4774.0000166](https://doi.org/10.1061/(asce)ir.1943-4774.0000166)
- Hersbach, H. (2000). Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather and Forecasting*, 15(5), 559–570. [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2)
- IPCC. (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Core writing team, R. . Pachaurim, & L. . Meyer (eds.)).
- Jia, H., Wang, X., Ti, C., Zhai, Y., Field, R., Tafuri, A. N., Cai, H., & Yu, S. L. (2015). Field monitoring of a LID-BMP treatment train system in China. *Environmental Monitoring and Assessment*, 187, 373–391. <https://doi.org/10.1007/s10661-015-4595-2>
- Johnson, J. P., & Hunt, W. F. (2019). A retrospective comparison of water quality treatment in a bioretention cell 16 years following initial analysis. *Sustainability (Switzerland)*, 11(7),

- 1945–1957. <https://doi.org/10.3390/su11071945>
- Kaykhosravi, Sahereh, Khan, U., & Jadidi, A. (2018). A Comprehensive Review of Low Impact Development Models for Research, Conceptual, Preliminary and Detailed Design Applications. *Water*, 10(11), 1541–1569. <https://doi.org/10.3390/w10111541>
- Kaykhosravi, Sarah, Abogadil, K., Khan, U. T., & Jadidi, M. A. (2019). The Low-Impact Development Demand Index: A New Approach to Identifying Locations for LID. *Water*, 11(11), 2341–2369. <https://doi.org/10.3390/w11112341>
- Khan, U. T., & Valeo, C. (2017). Comparing a Bayesian and fuzzy number approach to uncertainty quantification in short-term dissolved oxygen prediction. *Journal of Environmental Informatics*, 30(1), 1–16. <https://doi.org/10.3808/JEI.201700371>
- Khan, U.T., Valeo, C., Chu, A., & van Duin, B. (2012a). Bioretention cell efficacy in cold climates: Part 1 — hydrologic performance. *Canadian Journal of Civil Engineering*, 39(11), 1210–1221. <https://doi.org/10.1139/l2012-110>
- Khan, U.T., Valeo, C., Chu, A., & van Duin, B. (2012b). Bioretention cell efficacy in cold climates: Part 2 — water quality performance. *Canadian Journal of Civil Engineering*, 39(11), 1222–1233. <https://doi.org/10.1139/l2012-111>
- Khan, Usman T., He, J., & Valeo, C. (2018). River flood prediction using fuzzy neural Networks: An investigation on automated network architecture. *Water Science and Technology*, 2017(1), 238–247. <https://doi.org/10.2166/wst.2018.107>
- Khan, Usman T., & Valeo, C. (2016). Dissolved oxygen prediction using a possibility theory based fuzzy neural network. *Hydrology and Earth System Sciences*, 20(6), 2267–2293. <https://doi.org/10.5194/HESS-20-2267-2016>
- Khu, S.-T., & Werner, M. G. F. (2003). Reduction of Monte-Carlo simulation runs for uncertainty estimation in hydrological modelling. *Hydrology and Earth System Sciences*, 7(5), 680–692. <https://doi.org/10.5194/hess-7-680-2003>
- Kollat, J. B., & Reed, P. M. (2006). Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design. *Advances in Water Resources*, 29(6), 792–807. <https://doi.org/10.1016/j.advwatres.2005.07.010>
- Kratky, H., Li, Z., Chen, Y., Wang, C., Li, X., & Yu, T. (2017). A critical literature review of bioretention research for stormwater management in cold climate and future research recommendations. *Frontiers of Environmental Science and Engineering*, 11(4), 1–15. <https://doi.org/10.1007/s11783-017-0982-y>
- Lee, J. G., Selvakumar, A., Alvi, K., Riverson, J., Zhen, J. X., Shoemaker, L., & Lai, F. hsiung. (2012). A watershed-scale design optimization model for stormwater best management

- practices. *Environmental Modelling and Software*, 37, 6–18.
<https://doi.org/10.1016/j.envsoft.2012.04.011>
- LeFevre, G. H., Paus, K. H., Natarajan, P., Gulliver, J. S., Novak, P. J., & Hozalski, R. M. (2015). Review of Dissolved Pollutants in Urban Storm Water and Their Removal and Fate in Bioretention Cells. *Journal of Environmental Engineering*, 141(1).
[https://doi.org/10.1061/\(ASCE\)EE.1943-7870.0000876](https://doi.org/10.1061/(ASCE)EE.1943-7870.0000876)
- Li, H., & Davis, A. P. (2008). Urban Particle Capture in Bioretention Media. I: Laboratory and Field Studies. *Journal of Environmental Engineering*, 134(6), 409–418.
[https://doi.org/10.1061/\(ASCE\)0733-9372\(2008\)134:6\(409\)](https://doi.org/10.1061/(ASCE)0733-9372(2008)134:6(409))
- Li, J., Liu, F., & Li, Y. (2020). Simulation and design optimization of rain gardens via DRAINMOD and response surface methodology. *Journal of Hydrology*, 585.
<https://doi.org/10.1016/j.jhydrol.2020.124788>
- Liu, Y., Theller, L. O., Pijanowski, B. C., & Engel, B. A. (2016). Optimal selection and placement of green infrastructure to reduce impacts of land use change and climate change on hydrology and water quality: An application to the Trail Creek Watershed, Indiana. *Science of the Total Environment*, 553, 149–163. <https://doi.org/10.1016/j.scitotenv.2016.02.116>
- Lu, W., Qin, X., & Yu, J. (2019). On comparison of two-level and global optimization schemes for layout design of storage ponds. *Journal of Hydrology*, 570, 544–554.
<https://doi.org/10.1016/j.jhydrol.2018.10.078>
- Marsalek, J., & Watt, W. E. (1984). Design storms for urban drainage design. *Canadian Journal of Civil Engineering*, 11(3), 574–584. <https://doi.org/10.1139/l84-075>
- May, R., Dandy, G., & Maier, H. (2011). Review of Input Variable Selection Methods for Artificial Neural Networks. In *Artificial Neural Networks - Methodological Advances and Biomedical Applications*. InTech. <https://doi.org/10.5772/16004>
- May, R. J., Maier, H. R., & Dandy, G. C. (2010). Data splitting for artificial neural networks using SOM-based stratified sampling. *Neural Networks*, 23(2), 283–294.
<https://doi.org/10.1016/j.neunet.2009.11.009>
- McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, 21(2), 239–241. <https://doi.org/10.2307/1268522>
- Mohammadiun, S., Yazdi, J., Salehi Neyshabouri, S. A. A., & Sadiq, R. (2018). Development of a stochastic framework to design/rehabilitate urban stormwater drainage systems based on a resilient approach. *Urban Water Journal*, 15(2), 167–176.
<https://doi.org/10.1080/1573062X.2018.1424218>

- Muerdter, C. P., Wong, C. K., & Lefevre, G. H. (2018). Emerging investigator series: The role of vegetation in bioretention for stormwater treatment in the built environment: Pollutant removal, hydrologic function, and ancillary benefits. *Environmental Science: Water Research and Technology*, 4(5), 592–612. <https://doi.org/10.1039/c7ew00511c>
- Ngamalieu-Nengoue, U. A., Martínez-Solano, F. J., Iglesias-Rey, P. L., & Mora-Meliá, D. (2019). Multi-objective optimization for urban drainage or sewer networks rehabilitation through pipes substitution and storage tanks installation. *Water (Switzerland)*, 11(5), 935–949. <https://doi.org/10.3390/w11050935>
- Parasuraman, K., & Elshorbagy, A. (2008). Toward improving the reliability of hydrologic prediction: Model structure uncertainty and its quantification using ensemble-based genetic programming framework. *Water Resources Research*, 44(12), W12406. <https://doi.org/10.1029/2007WR006451>
- Poles, S., Fu, Y., & Rigoni, E. (2009). The effect of initial population sampling on the convergence of multi-objective genetic algorithms. *Lecture Notes in Economics and Mathematical Systems*, 618, 123–133. https://doi.org/10.1007/978-3-540-85646-7_12
- Pour, S. H., Wahab, A. K. A., Shahid, S., Asaduzzaman, M., & Dewan, A. (2020). Low impact development techniques to mitigate the impacts of climate-change-induced urban floods: Current trends, issues and challenges. *Sustainable Cities and Society*, 62, 102373. <https://doi.org/10.1016/j.scs.2020.102373>
- Qin, H. peng, Li, Z. xi, & Fu, G. (2013). The effects of low impact development on urban flooding under different rainfall characteristics. *Journal of Environmental Management*, 129, 577–585. <https://doi.org/10.1016/j.jenvman.2013.08.026>
- Raei, E., Reza Alizadeh, M., Reza Nikoo, M., & Adamowski, J. (2019). Multi-objective decision-making for green infrastructure planning (LID-BMPs) in urban storm water management under uncertainty. *Journal of Hydrology*, 579, 124091. <https://doi.org/10.1016/j.jhydrol.2019.124091>
- Reed, P. M., Hadka, D., Herman, J. D., Kasprzyk, J. R., & Kollat, J. B. (2013). Evolutionary multiobjective optimization in water resources: The past, present, and future. *Advances in Water Resources*, 51, 438–456. <https://doi.org/10.1016/j.advwatres.2012.01.005>
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195–204. <https://doi.org/10.1038/s41586-019-0912-1>
- Rosenberg, E. A., Keys, P. W., Booth, D. B., Hartley, D., Burkey, J., Steinemann, A. C., & Lettenmaier, D. P. (2010). Precipitation extremes and the impacts of climate change on

- stormwater infrastructure in Washington State. *Climatic Change*, 102(1–2), 319–349.
<https://doi.org/10.1007/s10584-010-9847-0>
- Roy-Poirier, A., Champagne, P., & Fillion, Y. (2010). Review of Bioretention System Research and Design: Past, Present, and Future. *Journal of Environmental Engineering*, 136(9), 878–889. [https://doi.org/10.1061/\(asce\)ee.1943-7870.0000227](https://doi.org/10.1061/(asce)ee.1943-7870.0000227)
- Santi, M. De, Khan, U. T., Arnold, M., Fesselet, J.-F., & Ali, S. I. (2021). Forecasting point-of-consumption chlorine residual in refugee settlements using ensembles of artificial neural networks. *Npj Clean Water* 2021 4:1, 4(1), 1–16. <https://doi.org/10.1038/s41545-021-00125-2>
- Sastry, K., Goldberg, D. E., & Kendall, G. (2014). Genetic Algorithms. In E. K. Burke & G. Kendall (Eds.), *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques* (pp. 93–117). Springer US. https://doi.org/10.1007/978-1-4614-6940-7_4
- Schardong, A., Simonovic, S. P., Gaur, A., & Sandink, D. (2020). Web-based tool for the development of intensity duration frequency curves under changing climate at gauged and ungauged locations. *Water (Switzerland)*, 12(5), 1243. <https://doi.org/10.3390/W12051243>
- Semadeni-Davies, A., Hernebring, C., Svensson, G., & Gustafsson, L. G. (2008). The impacts of climate change and urbanisation on drainage in Helsingborg, Sweden: Suburban stormwater. *Journal of Hydrology*, 350(1–2), 114–125.
<https://doi.org/10.1016/j.jhydrol.2007.11.006>
- Seshadri, A. (2009). *NSGA - II: A multi-objective optimization algorithm*.
<https://www.mathworks.com/matlabcentral/fileexchange/10429-nsga-ii-a-multi-objective-optimization-algorithm>
- Singh, R. P., Zhao, F., Ji, Q., Saravanan, J., & Fu, D. (2019). Design and performance characterization of roadside bioretention systems. *Sustainability (Switzerland)*, 11(7), 2040. <https://doi.org/10.3390/su11072040>
- Sivanandam, S. N., & Deepa, S. N. (2008a). Genetic Algorithms. In *Introduction to Genetic Algorithms* (pp. 15–37). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-73190-0_2
- Sivanandam, S. N., & Deepa, S. N. (2008b). Technologies and Operators of GA. In *Introduction to Genetic Algorithms* (pp. 39–82). Springer Berlin Heidelberg.
- Snieder, E., Shakir, R., & Khan, U. T. (2020). A comprehensive comparison of four input variable selection methods for artificial neural network flow forecasting models. *Journal of Hydrology*, 583, 124299. <https://doi.org/10.1016/j.jhydrol.2019.124299>

- Sohn, W., Kim, J. H., Li, M. H., & Brown, R. (2019). The influence of climate on the effectiveness of low impact development: A systematic review. *Journal of Environmental Management*, 236, 365–379. <https://doi.org/10.1016/j.jenvman.2018.11.041>
- Solomatine, D. P., & Ostfeld, A. (2008). Data-driven modelling: Some past experiences and new approaches. *Journal of Hydroinformatics*, 10(1), 3–22. <https://doi.org/10.2166/hydro.2008.015>
- Thomas M, H., & Colucci, S. J. (1997). Verification of Eta–RSM Short-Range Ensemble Forecasts. *Monthly Weather Review*, 125(6), 1312–1327.
- Tirpak, R. A., Hathaway, J. M., Khojandi, A., Weathers, M., & Epps, T. H. (2021). Building resiliency to climate change uncertainty through bioretention design modifications. *Journal of Environmental Management*, 287, 112300. <https://doi.org/10.1016/j.jenvman.2021.112300>
- Toronto and Region Conservation Area. (2019). *STEP Low Impact Development Life Cycle Costing Tool*. <https://sustainabletechnologies.ca/lid-lcct/>
- van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J. F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., & Rose, S. K. (2011). The representative concentration pathways: An overview. *Climatic Change*, 109(1), 5–31. <https://doi.org/10.1007/s10584-011-0148-z>
- Visitacion, B. J., Booth, D. B., & Steinemann, A. C. (2009). Costs and Benefits of Storm-Water Management: Case Study of the Puget Sound Region. *Journal of Urban Planning and Development*, 135(4), 150–158. [https://doi.org/10.1061/\(ASCE\)0733-9488\(2009\)135:4\(150\)](https://doi.org/10.1061/(ASCE)0733-9488(2009)135:4(150))
- Wang, L., Lyons, J., Kanehl, P., & Bannerman, R. (2001). Impacts of urbanization on stream habitat and fish across multiple spatial scales. *Environmental Management*, 28(2), 255–266. <https://doi.org/10.1007/s0026702409>
- Wang, M., Zhang, D., Cheng, Y., & Tan, S. K. (2019). Assessing performance of porous pavements and bioretention cells for stormwater management in response to probable climatic changes. *Journal of Environmental Management*, 243, 157–167. <https://doi.org/10.1016/j.jenvman.2019.05.012>
- Wang, R., Eckelman, M. J., & Zimmerman, J. B. (2013). Consequential environmental and economic life cycle assessment of green and gray stormwater infrastructures for combined sewer systems. *Environmental Science and Technology*, 47(19), 11189–11198. <https://doi.org/10.1021/es4026547>
- Willems, P., Arnbjerg-Nielsen, K., Olsson, J., & Nguyen, V. T. V. (2012). Climate change impact

- assessment on urban rainfall extremes and urban drainage: Methods and shortcomings. *Atmospheric Research*, 103, 106–118. <https://doi.org/10.1016/j.atmosres.2011.04.003>
- Winston, R. J., Dorsey, J. D., & Hunt, W. F. (2016). Quantifying volume reduction and peak flow mitigation for three bioretention cells in clay soils in northeast Ohio. *Science of the Total Environment*, 553, 83–95. <https://doi.org/10.1016/j.scitotenv.2016.02.081>
- Wu, Y., & Chen, J. (2013). Investigating the effects of point source and nonpoint source pollution on the water quality of the East River (Dongjiang) in South China. *Ecological Indicators*, 32, 294–304. <https://doi.org/10.1016/J.ECOLIND.2013.04.002>
- Yang, Y., & Chui, T. F. M. (2018). Optimizing surface and contributing areas of bioretention cells for stormwater runoff quality and quantity management. *Journal of Environmental Management*, 206, 1090–1103. <https://doi.org/10.1016/j.jenvman.2017.11.064>
- Yazdi, J., Mohammadiun, S., Sadiq, R., Salehi Neyshabouri, S. A. A., & Alavi Gharahbagh, A. (2018). Assessment of different MOEAs for rehabilitation evaluation of Urban Stormwater Drainage Systems – Case study: Eastern catchment of Tehran. *Journal of Hydro-Environment Research*, 21, 76–85. <https://doi.org/10.1016/j.jher.2018.08.002>
- Yazdi, J., & Salehi Neyshabouri, S. A. A. (2014). Identifying low impact development strategies for flood mitigation using a fuzzy-probabilistic approach. *Environmental Modelling and Software*, 60, 31–44. <https://doi.org/10.1016/j.envsoft.2014.06.004>
- Zahmatkesh, Z., Burian, S. J., Karamouz, M., Tavakol-Davani, H., & Goharian, E. (2015). Low-Impact Development Practices to Mitigate Climate Change Effects on Urban Stormwater Runoff: Case Study of New York City. *Journal of Irrigation and Drainage Engineering*, 141(1), 04014043. [https://doi.org/10.1061/\(ASCE\)IR.1943-4774.0000770](https://doi.org/10.1061/(ASCE)IR.1943-4774.0000770)
- Zhang, K., Manuelpillai, D., Raut, B., Deletic, A., & Bach, P. M. (2019). Evaluating the reliability of stormwater treatment systems under various future climate conditions. *Journal of Hydrology*, 568, 57–66. <https://doi.org/10.1016/j.jhydrol.2018.10.056>
- Zhang, W., Li, J., Chen, Y., & Li, Y. (2019). A Surrogate-Based Optimization Design and Uncertainty Analysis for Urban Flood Mitigation. *Water Resources Management*, 33(12), 4201–4214. <https://doi.org/10.1007/s11269-019-02355-z>
- Zhou, Q. (2014). A review of sustainable urban drainage systems considering the climate change and urbanization impacts. *Water (Switzerland)*, 6(4), 976–992. <https://doi.org/10.3390/w6040976>
- Zhou, Q., Leng, G., Su, J., & Ren, Y. (2019). Comparison of urbanization and climate change impacts on urban flood volumes: Importance of urban planning and drainage adaptation. *Science of the Total Environment*, 658, 24–33.

<https://doi.org/10.1016/j.scitotenv.2018.12.184>

Appendices

Appendix A: Surrogate Model performance for all storms

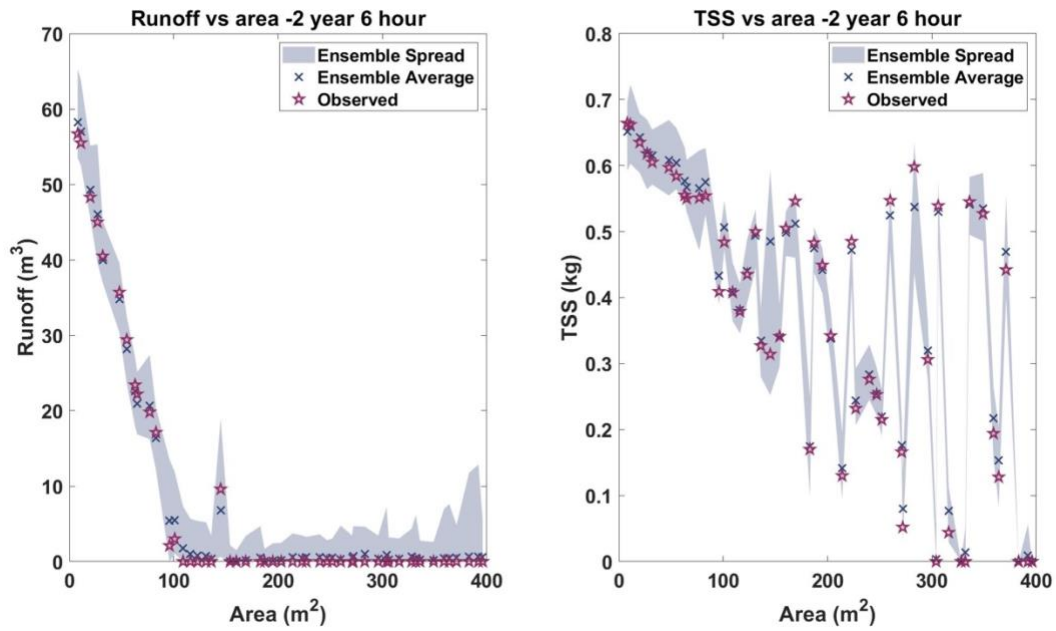


Figure A - 1: SM results for 2-year 6-hour storm

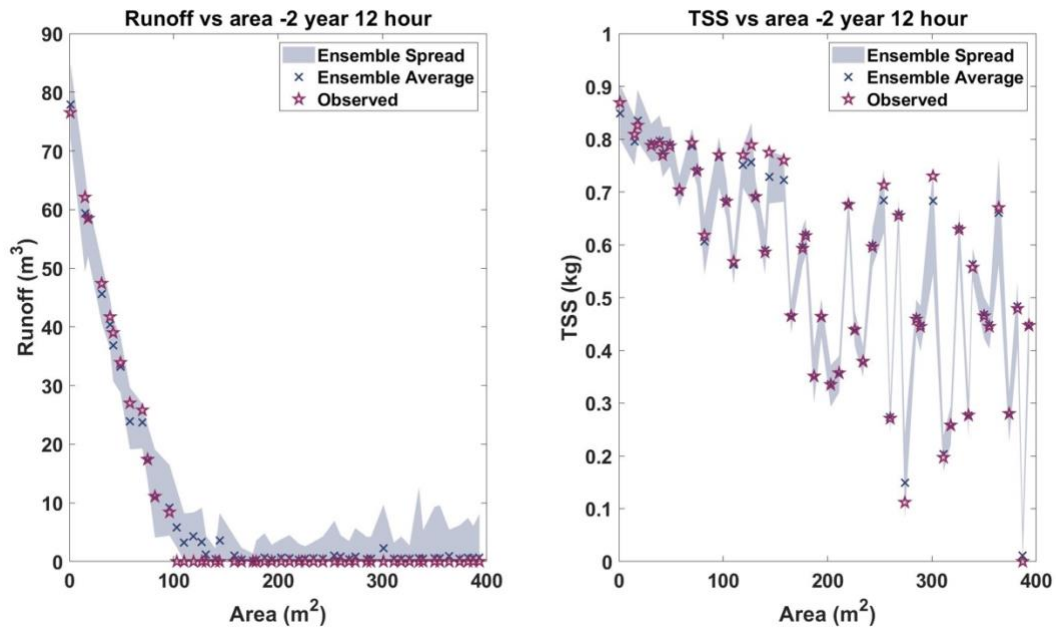


Figure A - 2: SM results for 2-year 12-hour storm

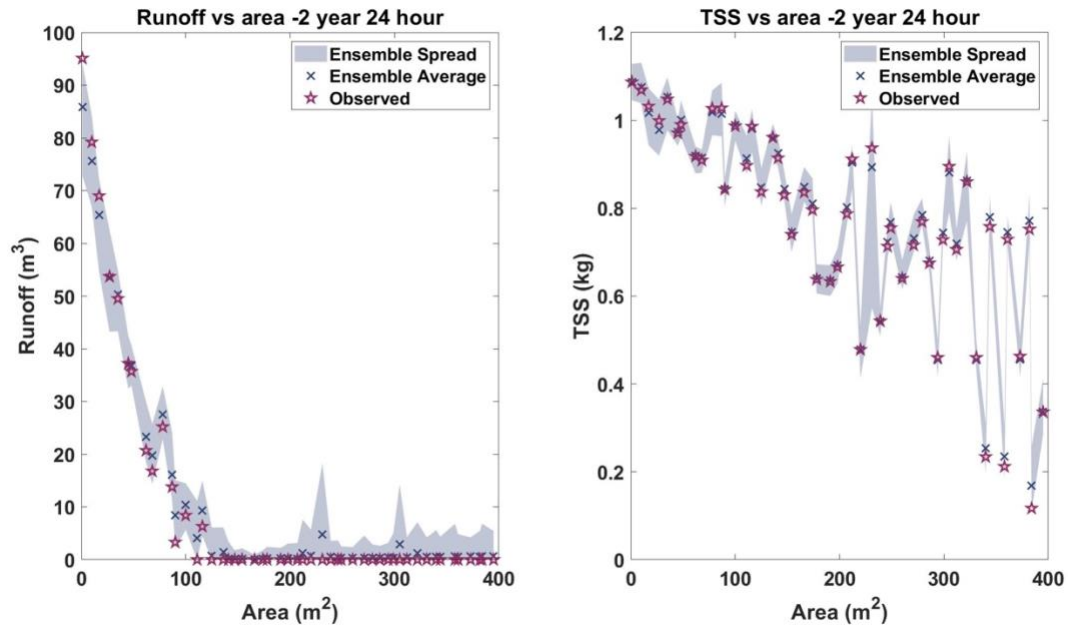


Figure A - 3: SM results for 2-year 24-hour storm

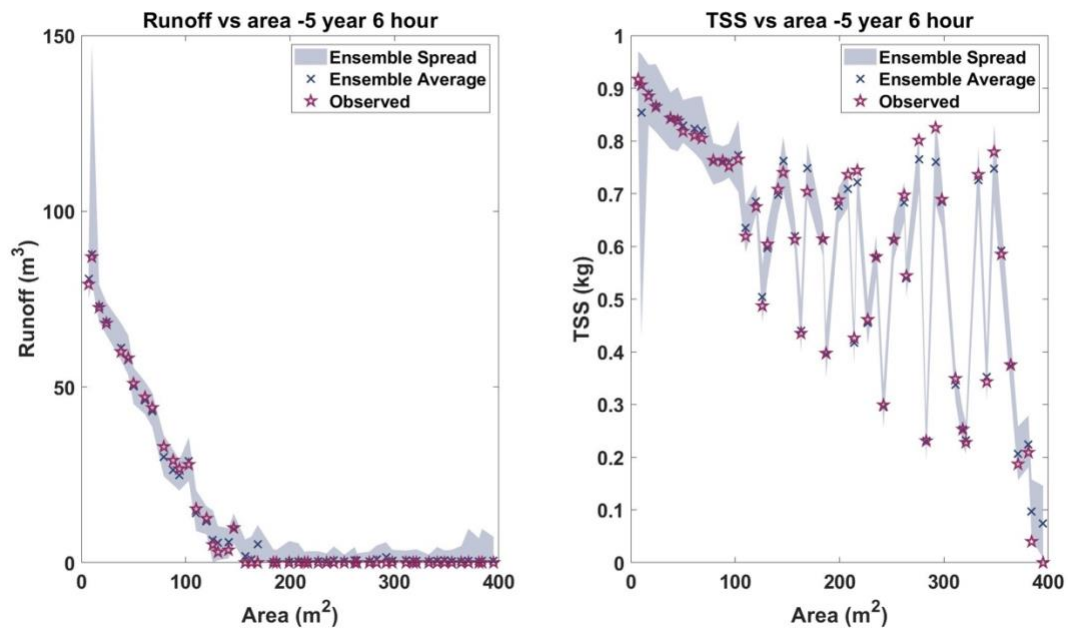


Figure A - 4: SM results for 5-year 6-hour storm

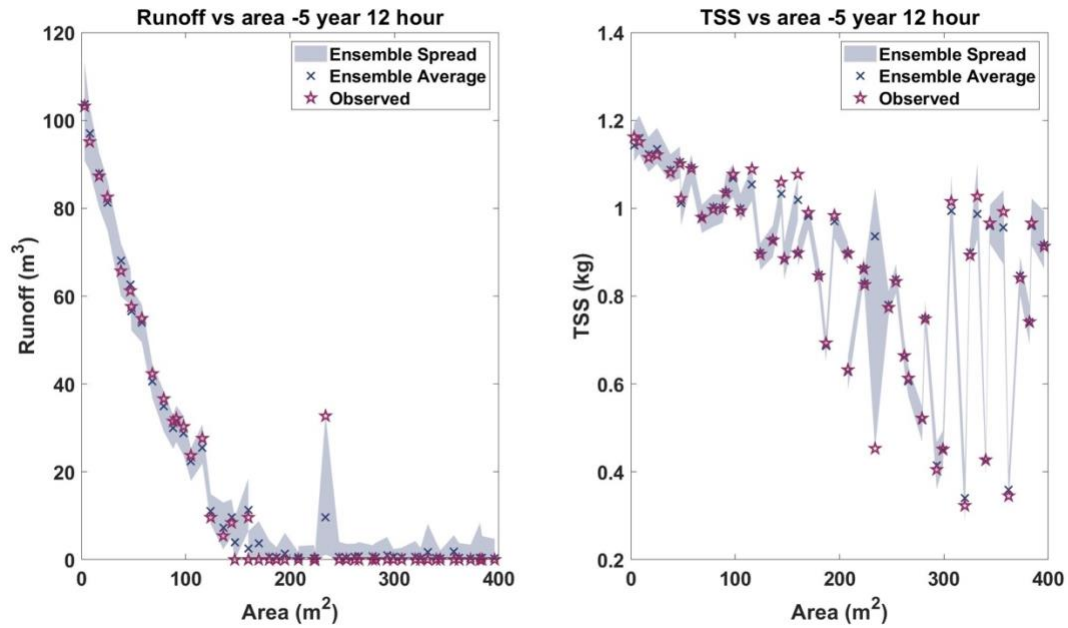


Figure A - 5: SM results for 5-year 12-hour storm

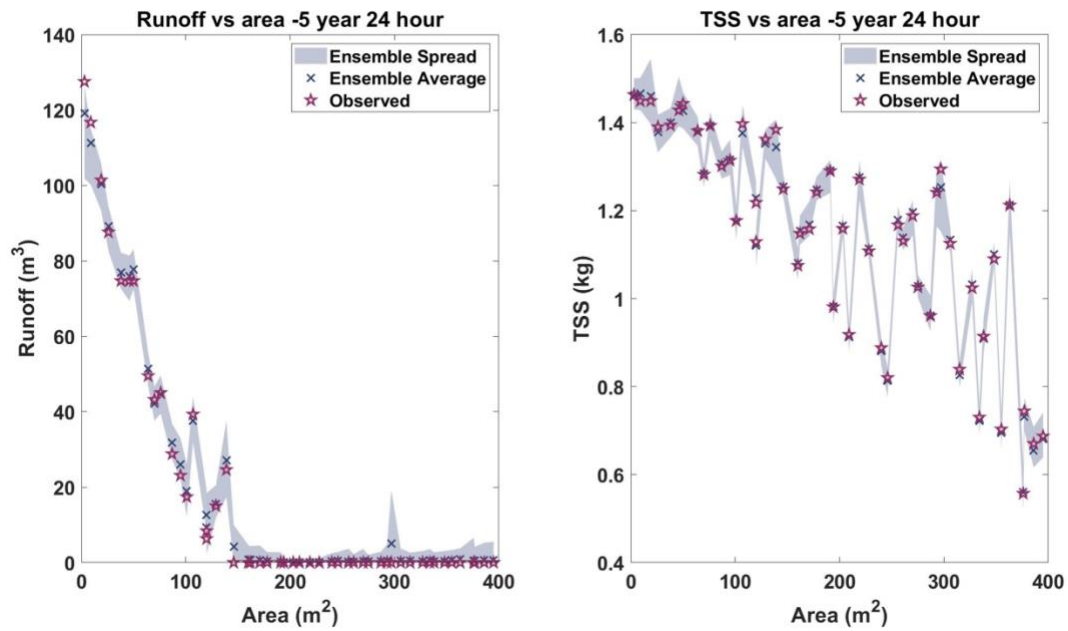


Figure A - 6: SM results for 5-year 24-hour storm

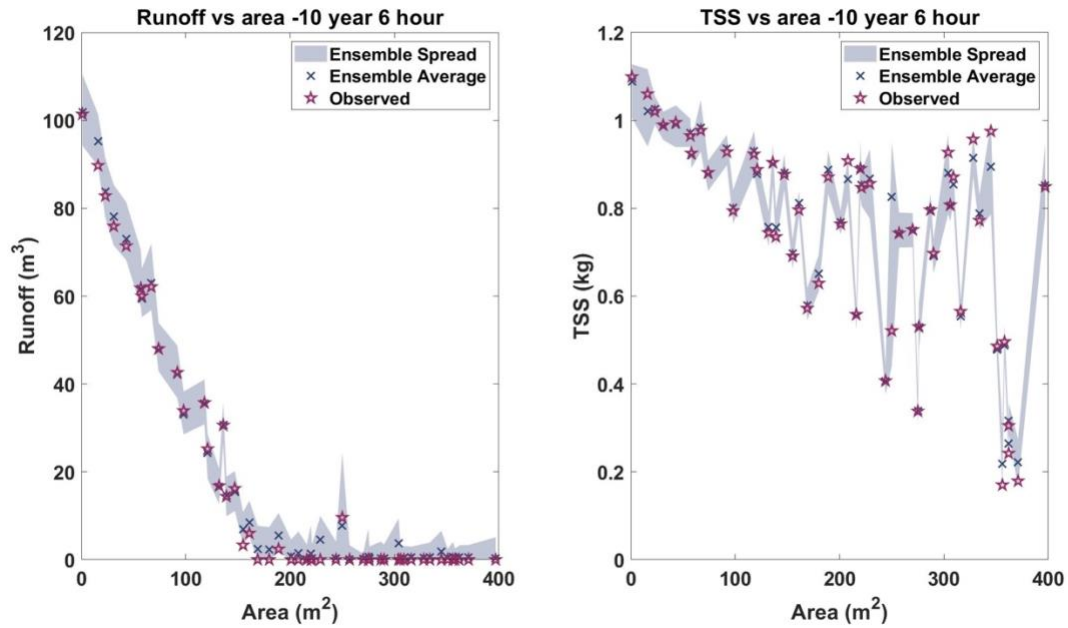


Figure A - 7: SM results for 10-year 6-hour storm

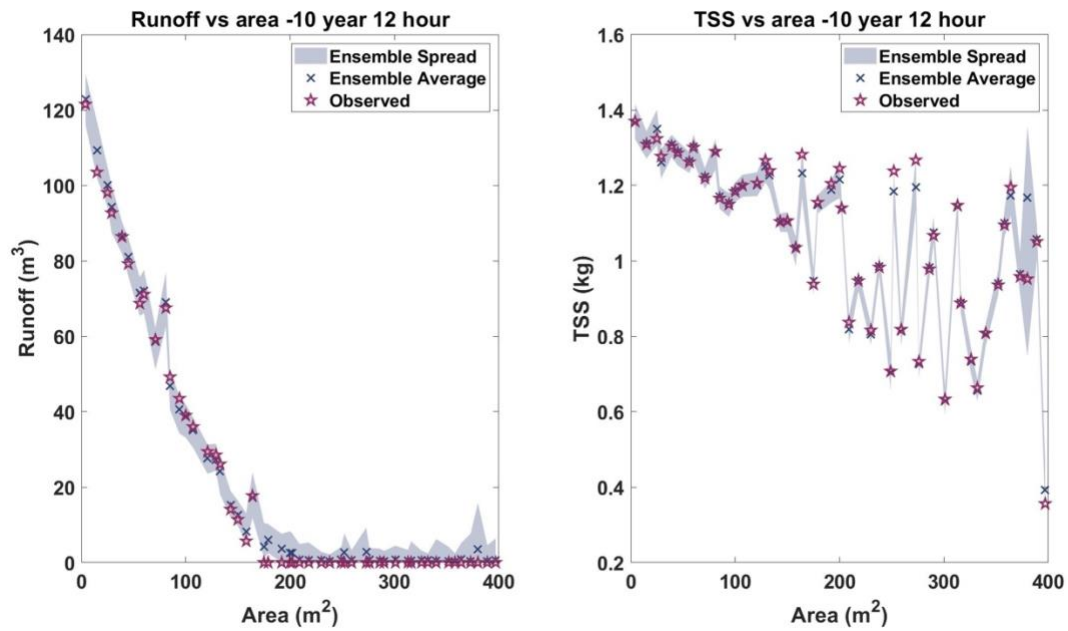


Figure A - 8: SM results for 10-year 12-hour storm

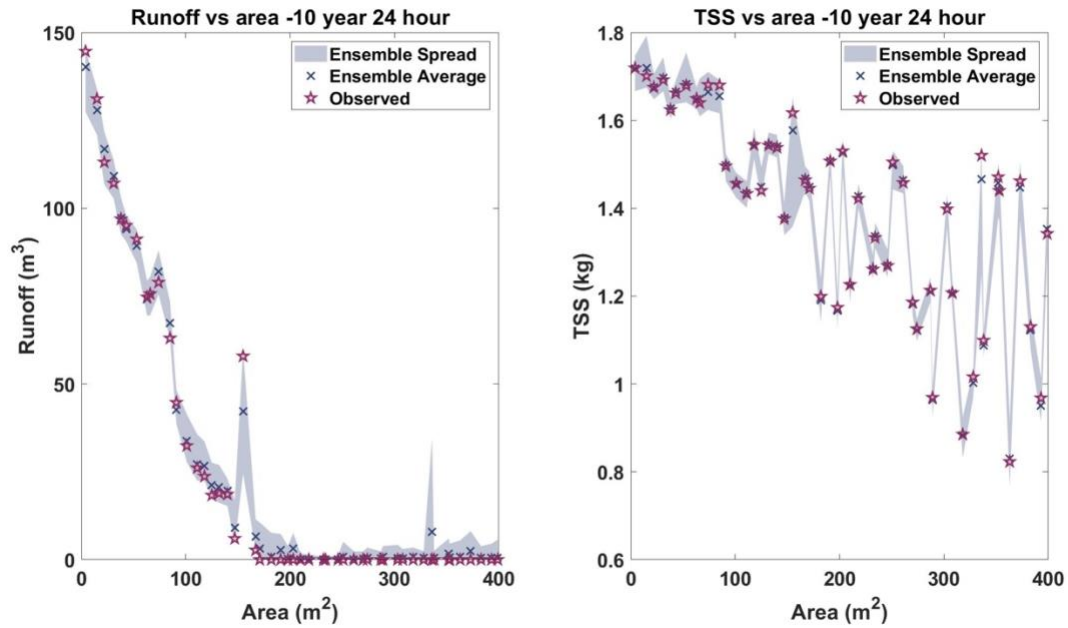


Figure A - 9: SM results for 10-year 24-hour storm

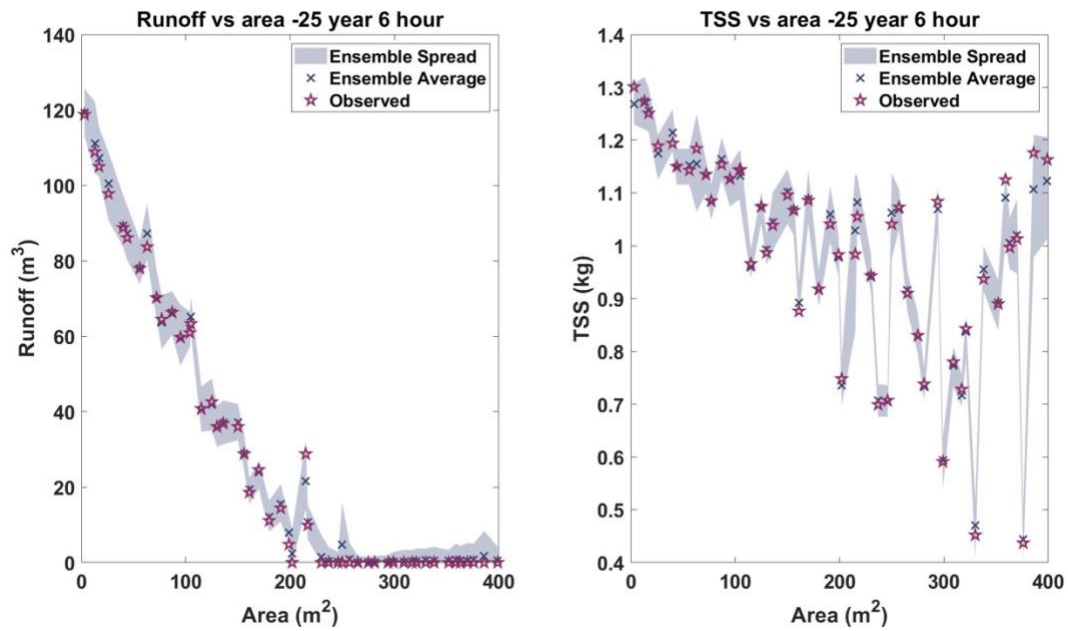


Figure A - 10: SM results for 25-year 6-hour storm

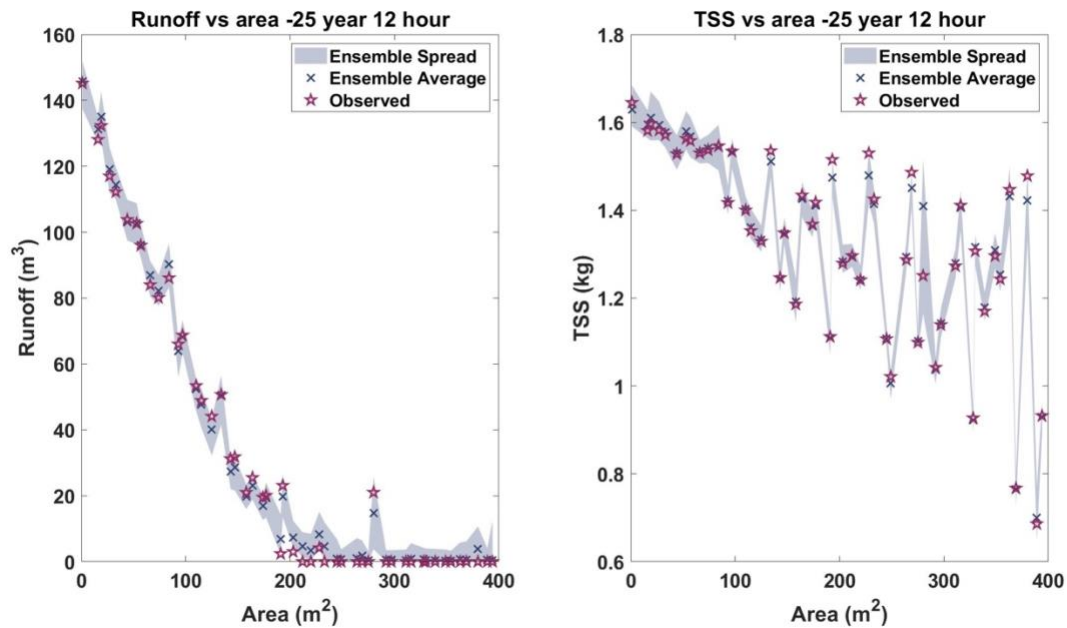


Figure A - 11: SM results for 25-year 12-hour storm

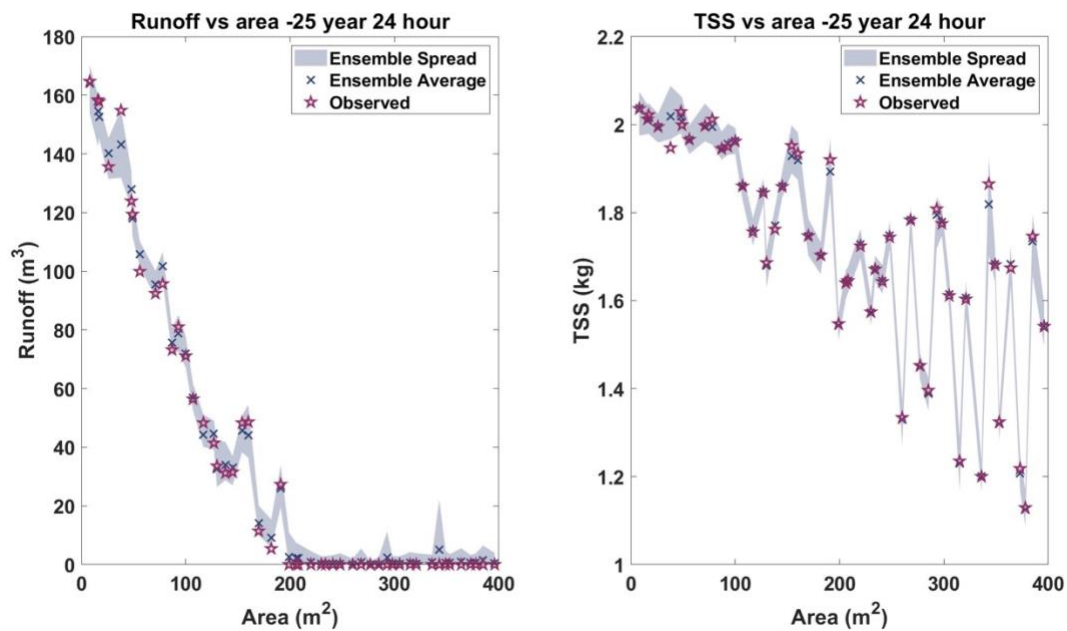


Figure A - 12: SM results for 25-year 24-hour storm

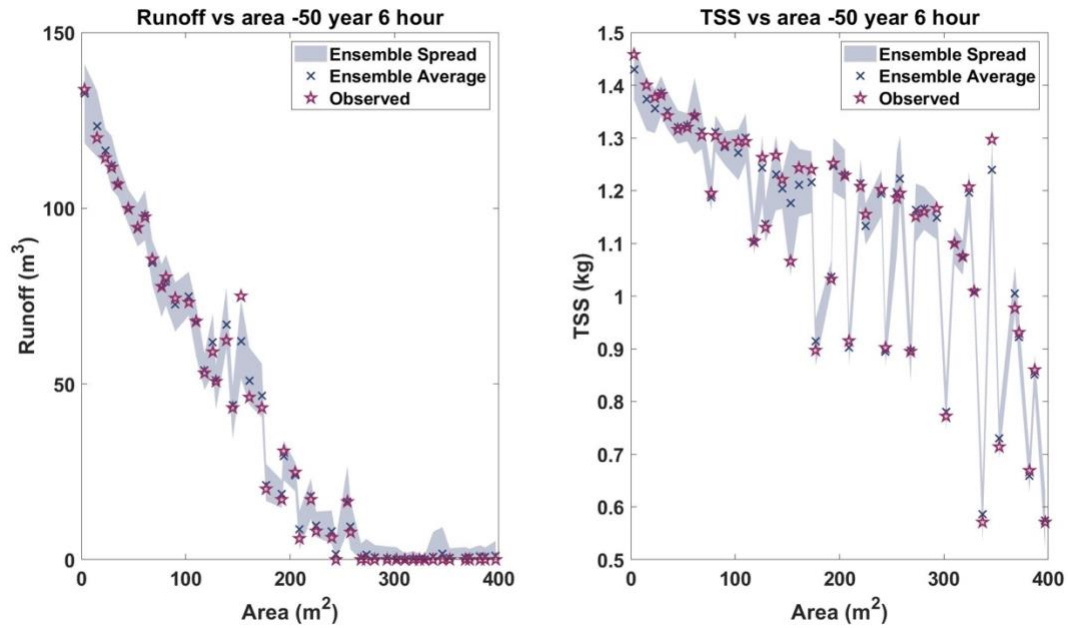


Figure A - 13: SM results for 50-year 6-hour storm

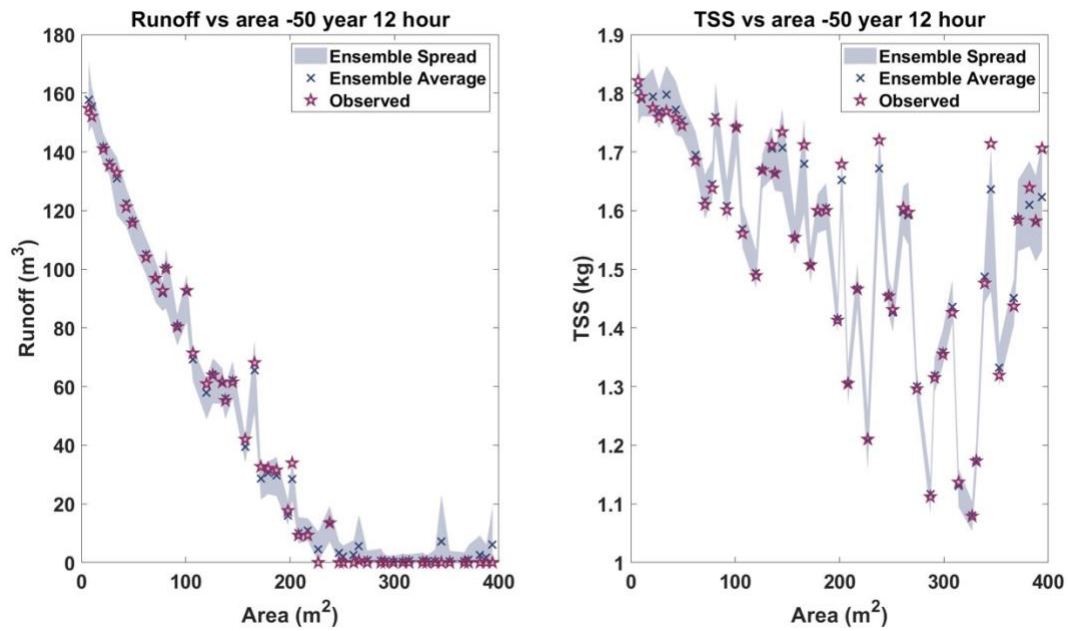


Figure A - 14: SM results for 50-year 12-hour storm

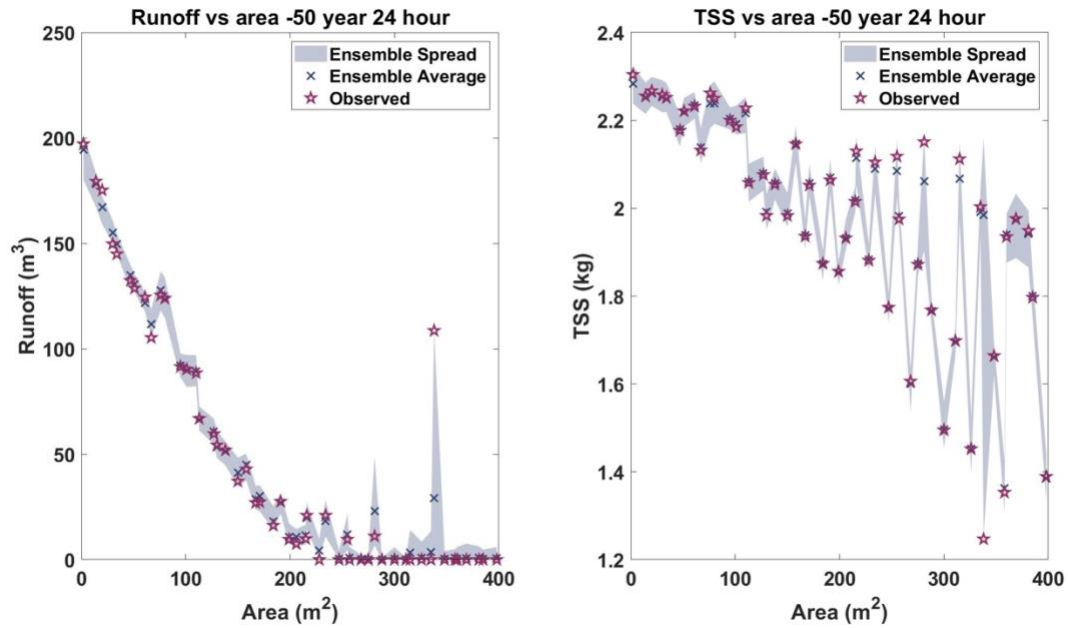


Figure A - 15: SM results for 50-year 24-hour storm

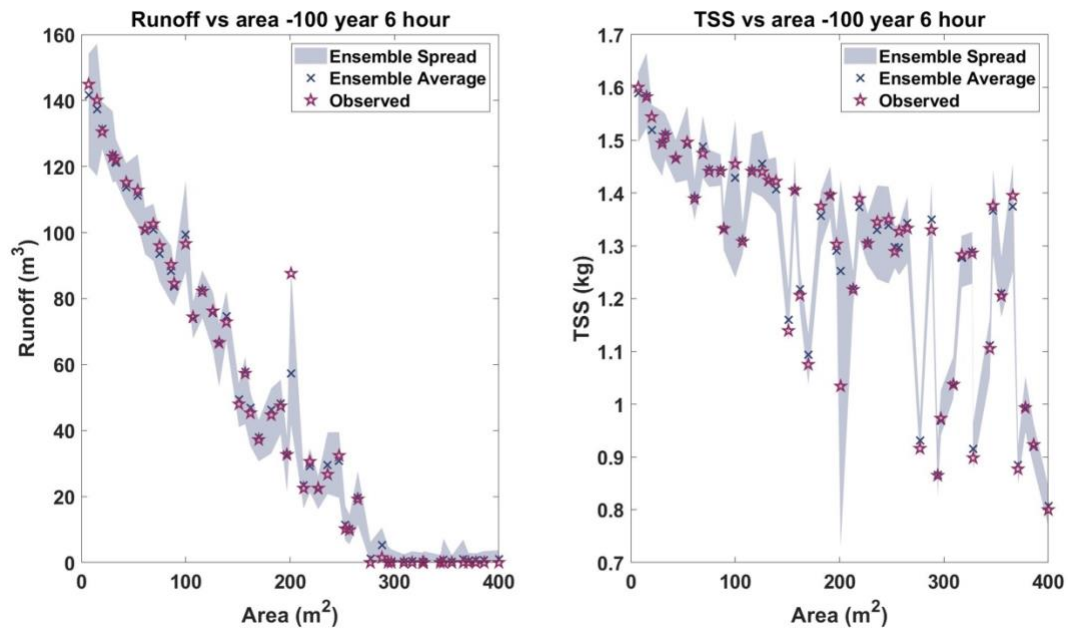


Figure A - 16: SM results for 100-year 6-hour storm

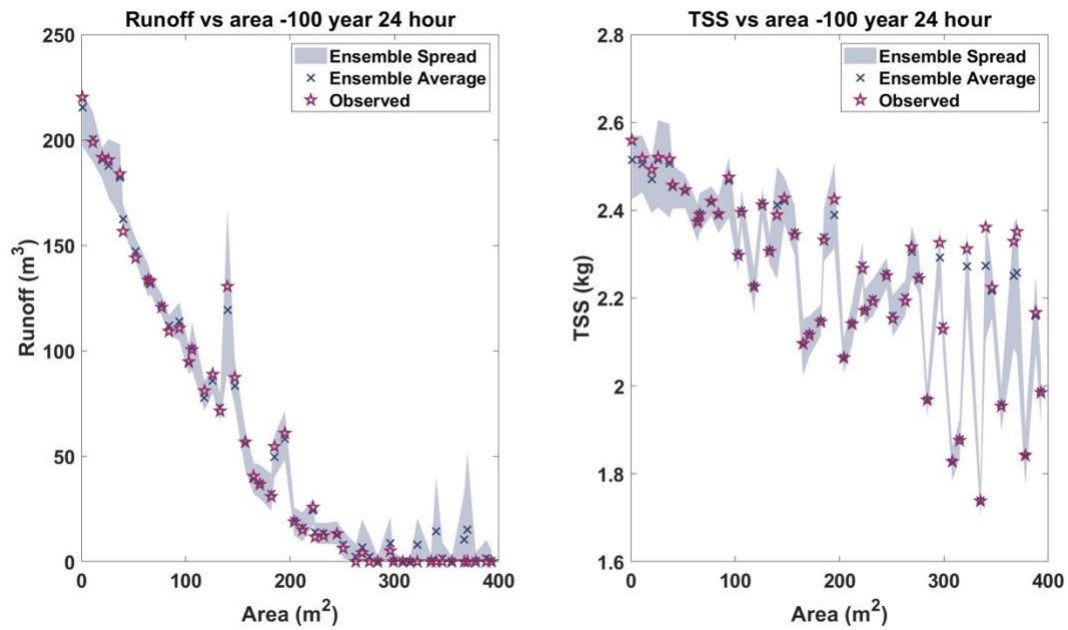


Figure A - 17: SM results for 100-year 24-hour storm

Appendix B: BRC optimised design performance under current and future climate conditions for all storms

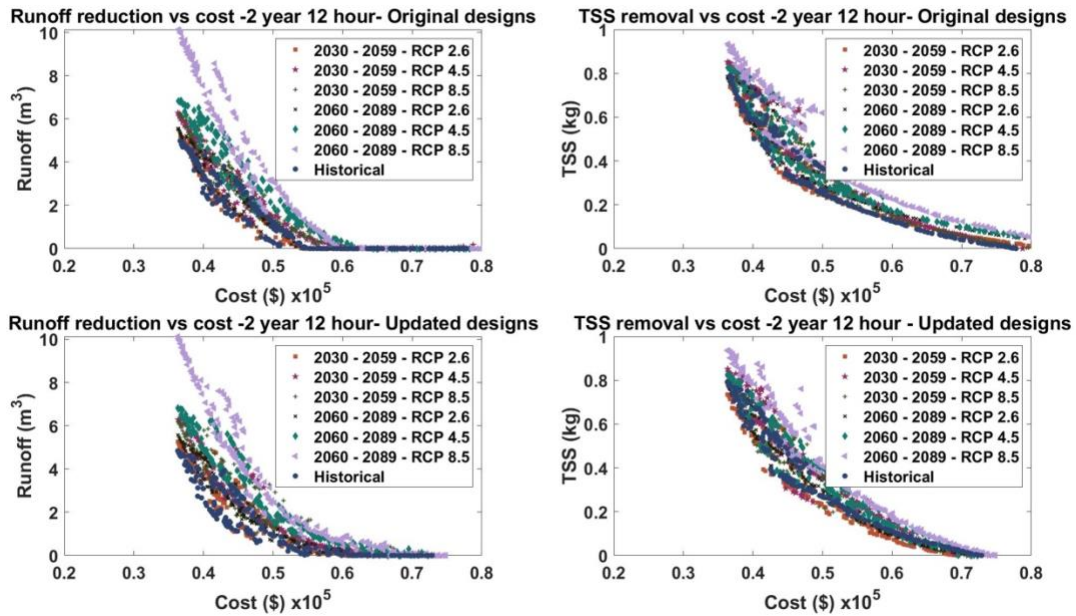


Figure B - 1: BRC optimisation results - Existing versus Updated designs – 2-year 12-hour

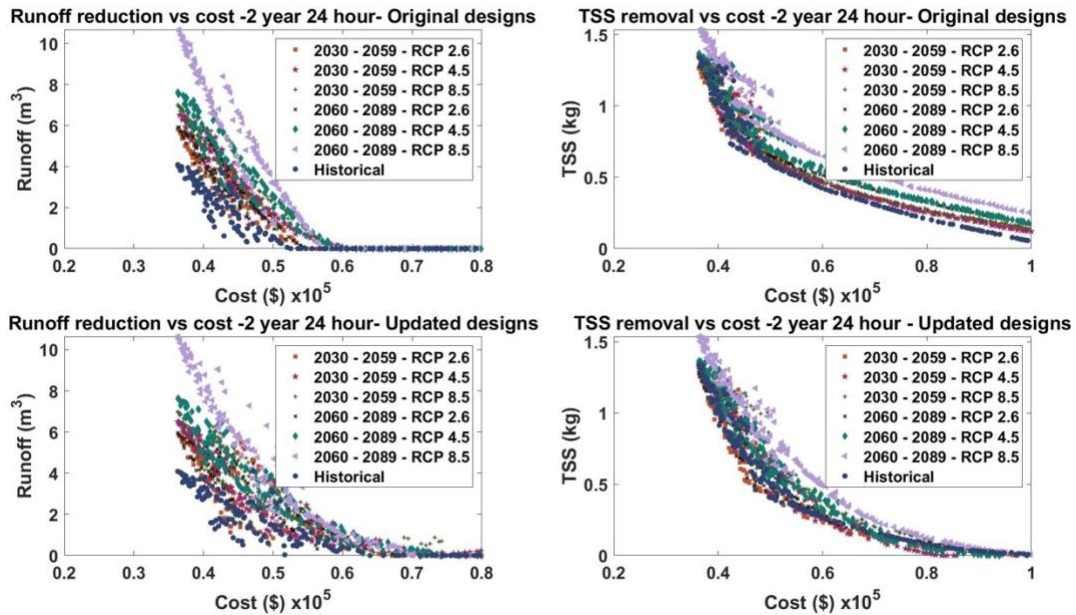


Figure B - 2: BRC optimisation results - Existing versus Updated designs – 2-year 24-hour

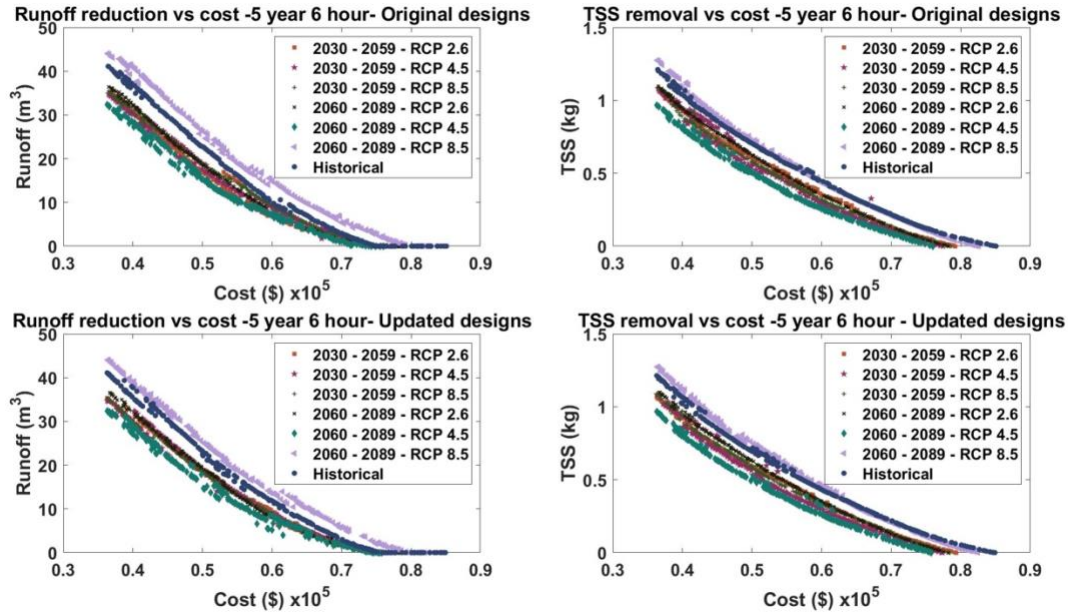


Figure B - 3: BRC optimisation results - Existing versus Updated designs – 5-year 6-hour

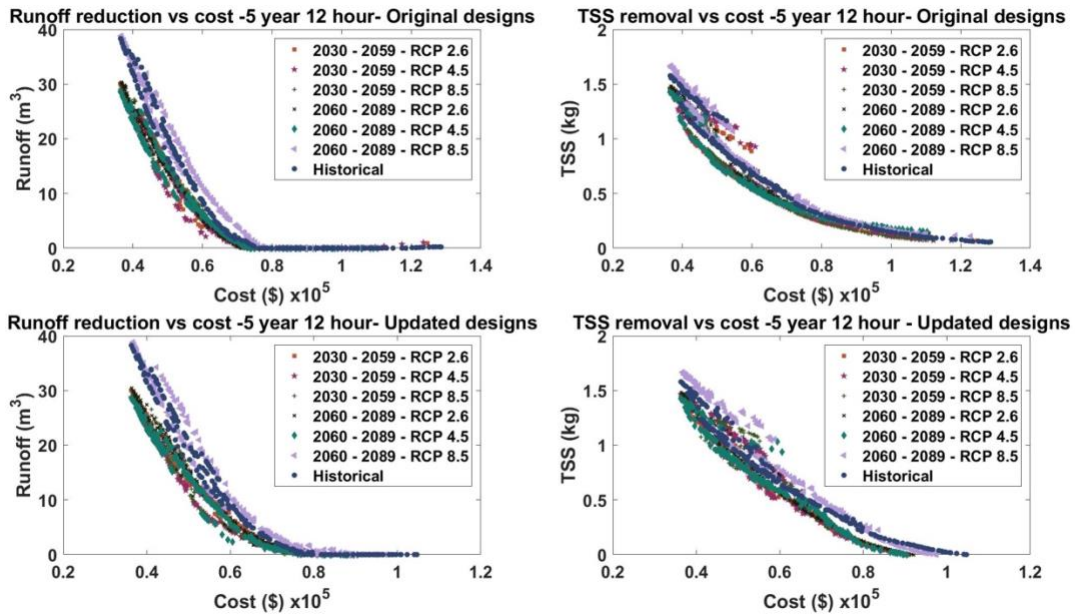


Figure B - 4: BRC optimisation results - Existing versus Updated designs – 5-year 12-hour

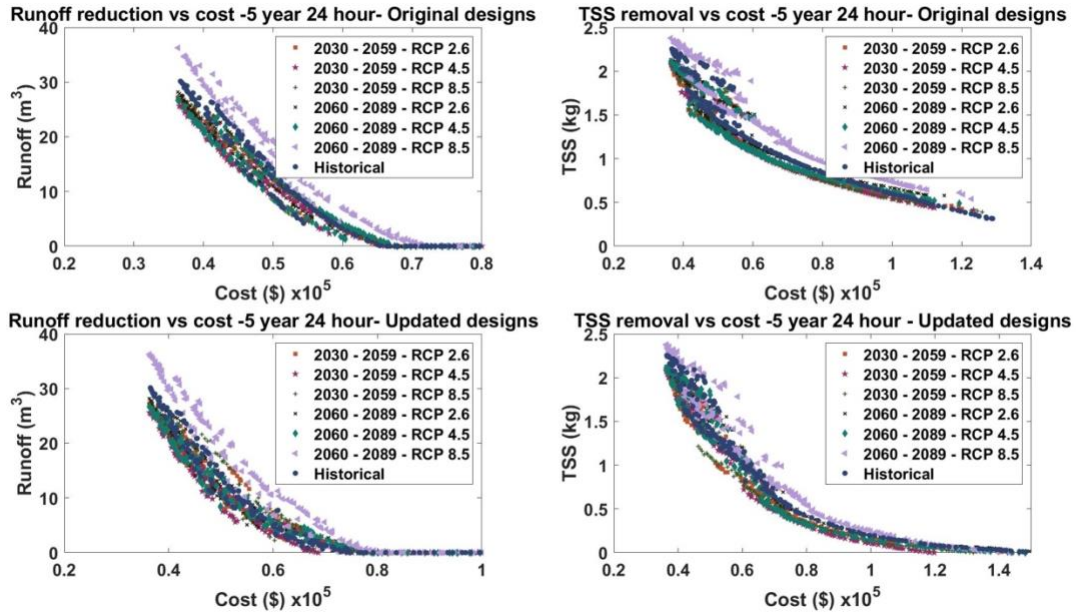


Figure B - 5: BRC optimisation results - Existing versus Updated designs – 5-year 24-hour

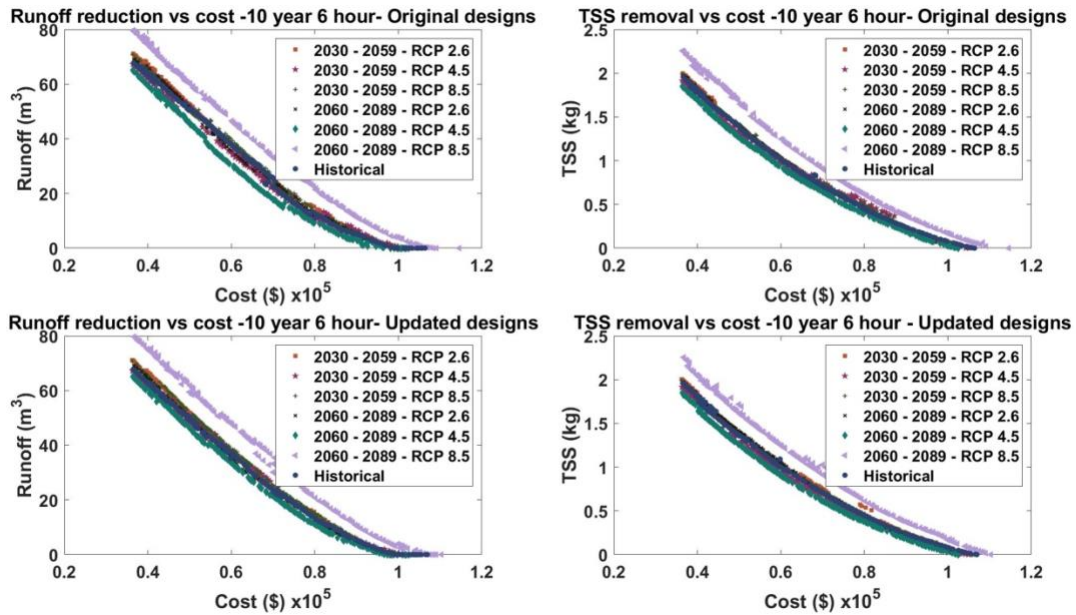


Figure B - 6: BRC optimisation results - Existing versus Updated designs – 10-year 6-hour

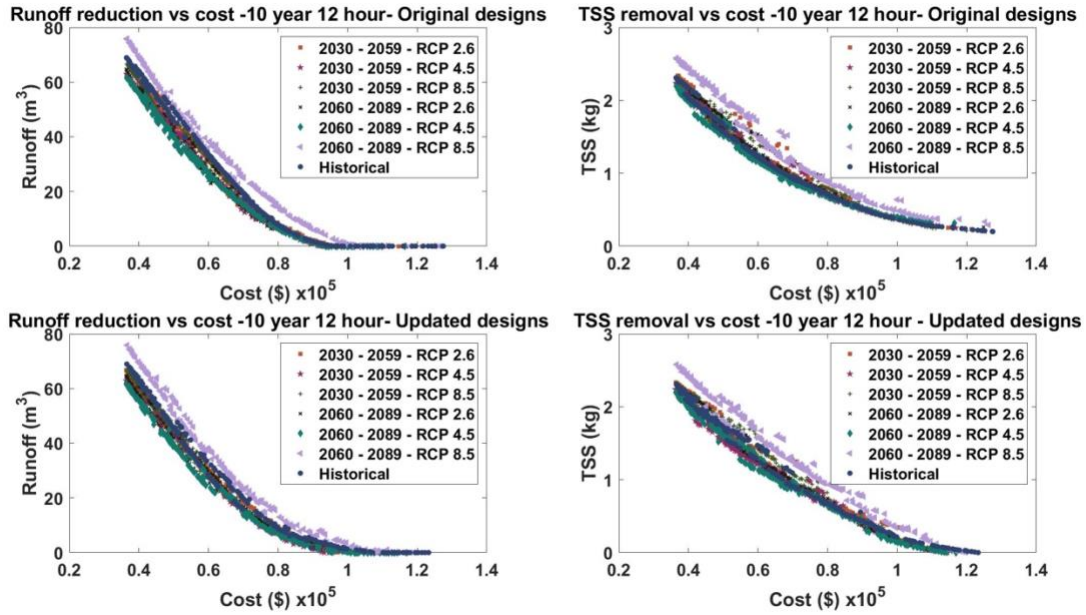


Figure B - 7: BRC optimisation results - Existing versus Updated designs –10-year 12-hour

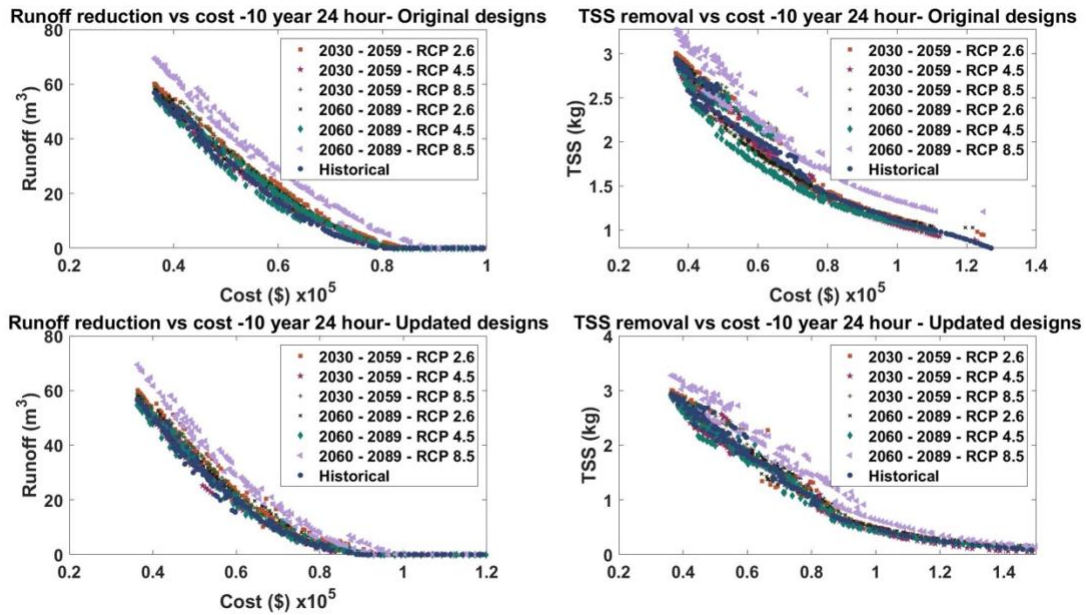


Figure B - 8: BRC optimisation results - Existing versus Updated designs – 10-year 24-hour

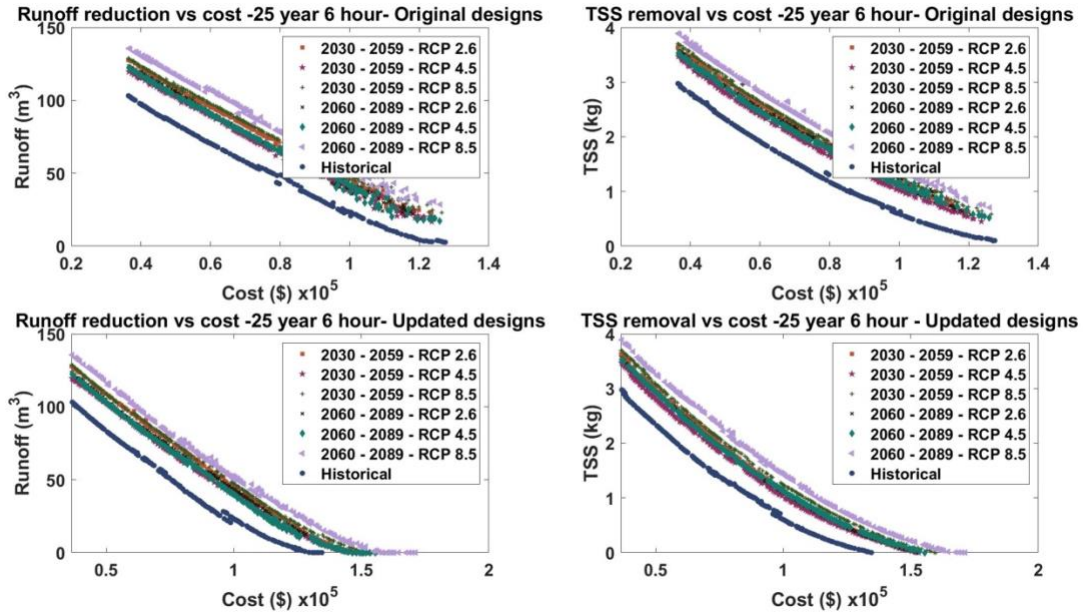


Figure B - 9: BRC optimisation results - Existing versus Updated designs – 25-year 6-hour

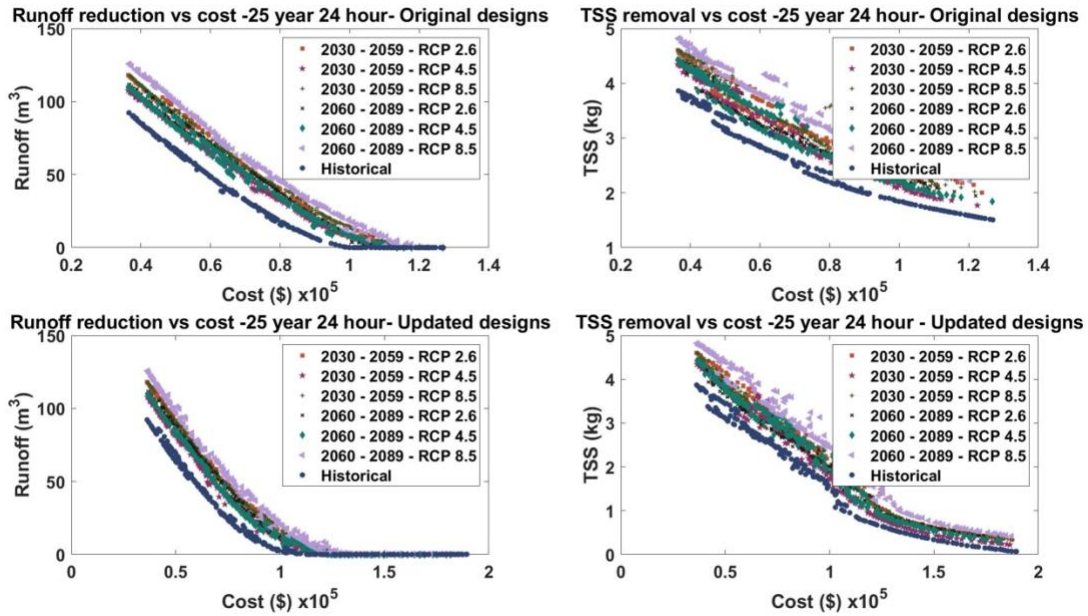


Figure B - 10: BRC optimisation results - Existing versus Updated designs – 25-year 24-hour

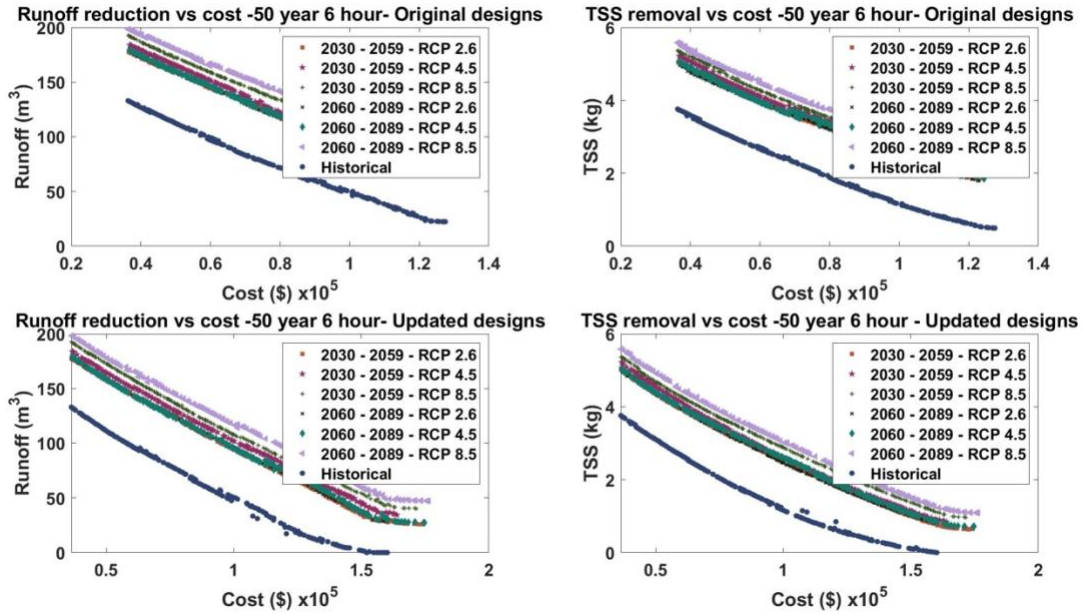


Figure B - 11: BRC optimisation results - Existing versus Updated designs – 50-year 6-hour

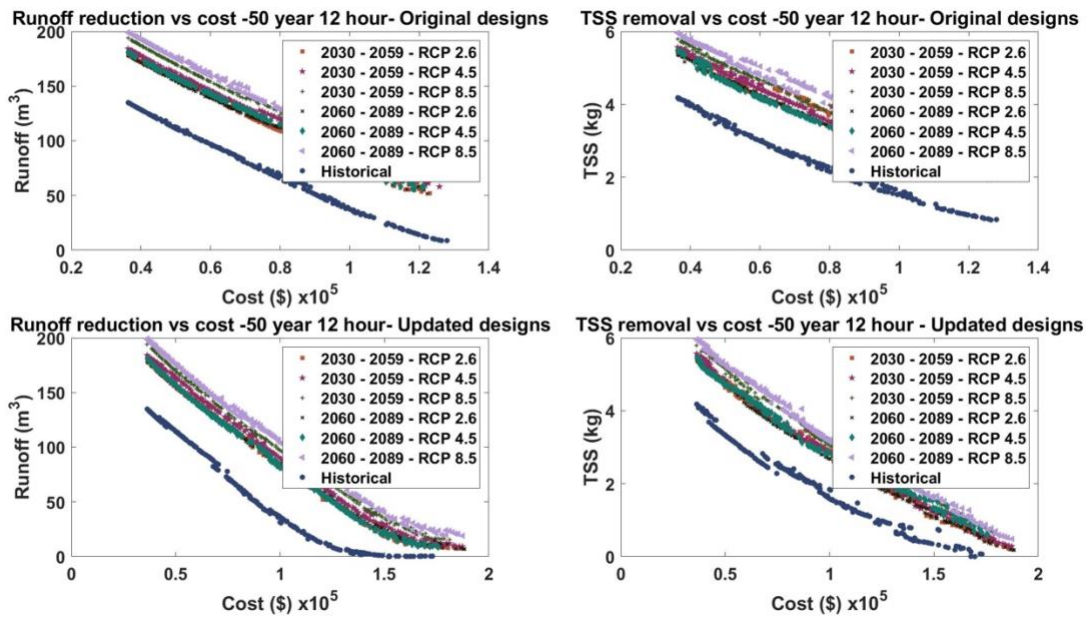


Figure B - 12: BRC optimisation results - Existing versus Updated designs – 50-year 12-hour

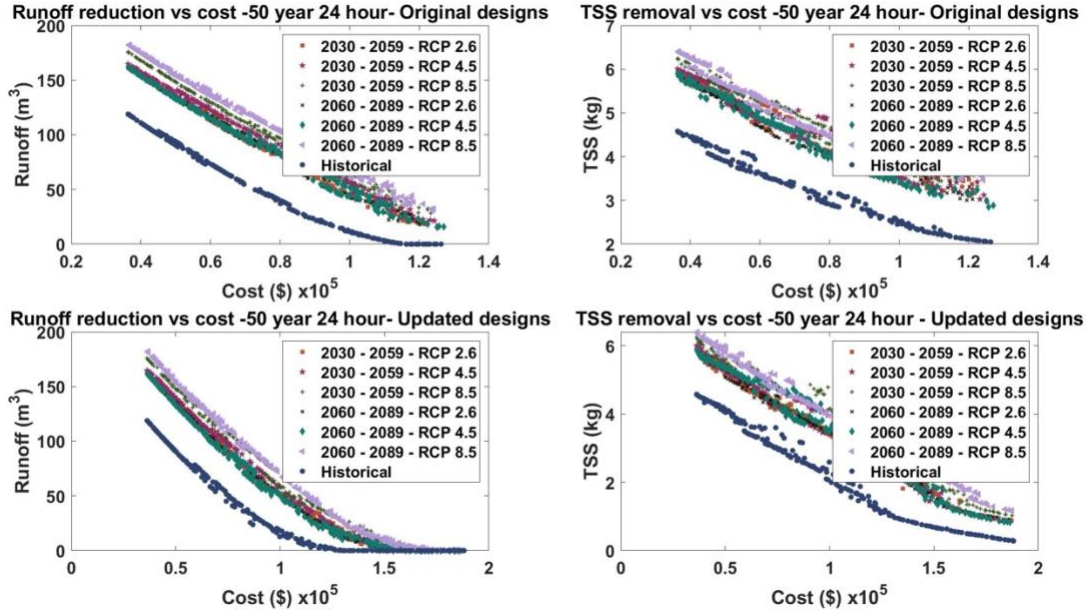


Figure B - 13: BRC optimisation results - Existing versus Updated designs –50-year 24-hour

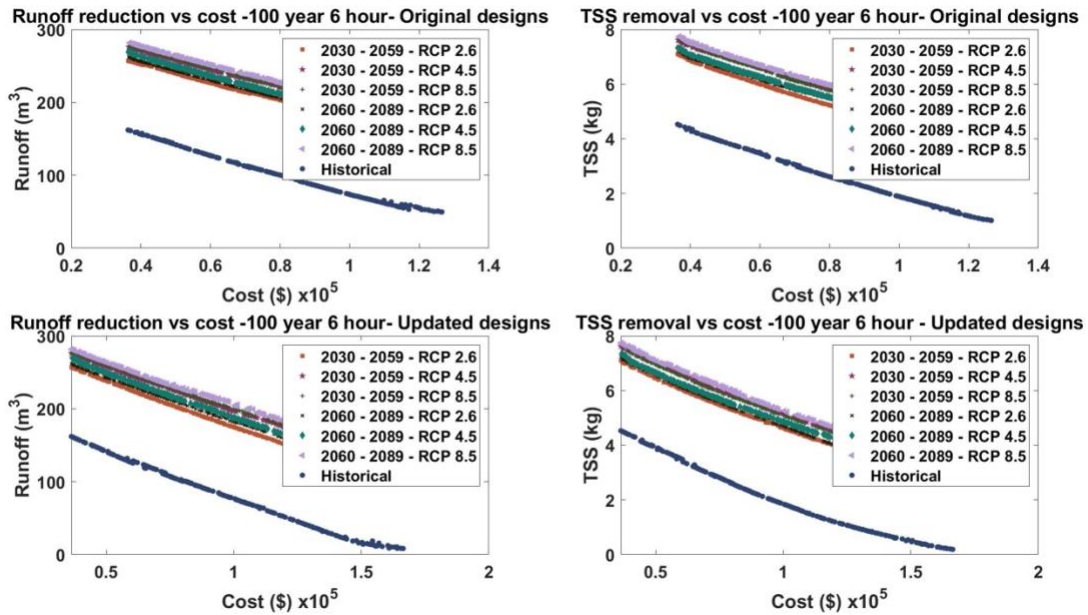


Figure B - 14: BRC optimisation results - Existing versus Updated designs – 100-year 6-hour

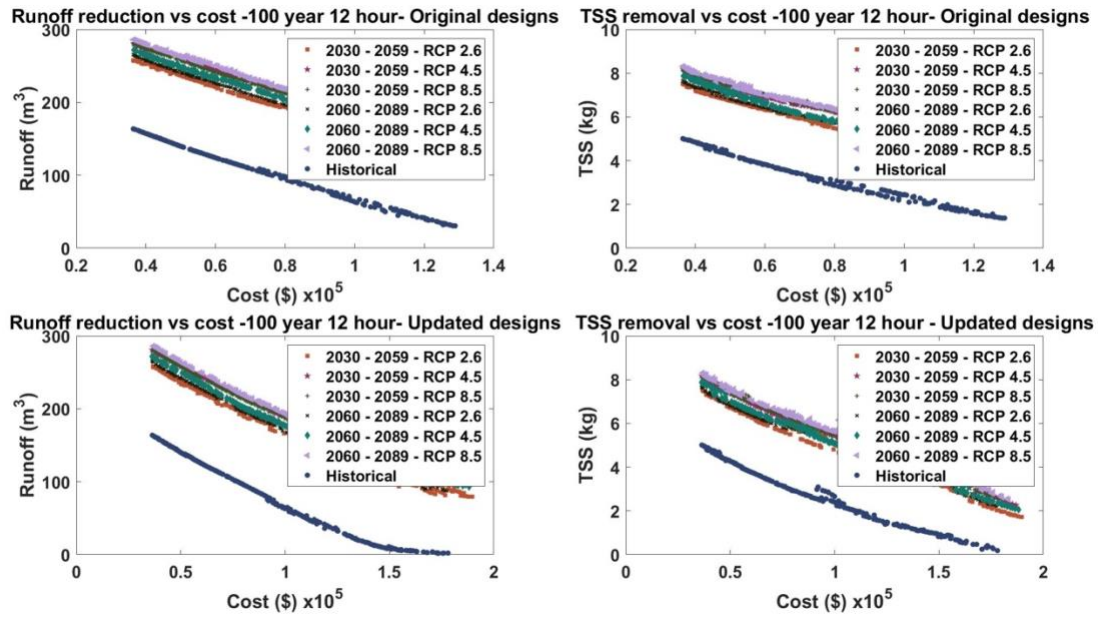


Figure B - 15: BRC optimisation results - Existing versus Updated designs – 100-year 12-hour

Appendix C: Pseudo code

Optimisation pseudo Code

V = LHS(I) Ve = E(V,SM)	Initialize parent population using LHS Evaluate the initial population set using the SMs in terms of the objective functions
for j = 1:num_population	Rank initial population in terms of rank
for u = 1:num_population	
if individual(j,m(1)) > individual(u,m(1))	
Increase domination count of individual u	
else	
Increase domination count of individual j	
end	
end	
for j = 1:num_population	Calculate crowding distance of each individual for each m objective
for u = 1:num_objective	
l(individual(j)) = O(j+1,m(u)) - O(j-1,m(u)) / fmax - fmin	
end	
end	
C = Nd(Ve)	Sort the initial population by rank and crowding distance
for i = 1:gen	
Op(i) = GO(C)	Initialize the offspring population using the genetic operator and perform mutation and crossover
D(i) = [Op(i);C]	Combine the initial population and parent population
D_s(i) = Nd(D(i))	Sort the combined population using nondomination
C(i) = R(D_s(i))	Replace the existing solution with all those better than it
end	