Response of Urban Stormwater Quality to Rainfall Characteristics and Land-Use/Land-Cover

by

Haibin Yan

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Water Resources Engineering

Department of Civil and Environmental Engineering
University of Alberta

Abstract

Urban stormwater quality is a critical concern in urban planning and stormwater management due to its significant influence on receiving water bodies and ecosystems. The dynamics of stormwater quality is influenced by complex factors, with rainfall characteristics and land-use/land cover (LULC) playing primary roles. The joint effects of rainfall characteristics and LULC remain inadequately understood. This research aims to enhance the understanding of how urban stormwater quality responds to these variables and utilize the created knowledge to improve the performance of urban stormwater quality models.

First, a data mining framework was designed to study the impacts of rainfall characteristics on urban stormwater quality. Specifically, the relationship between rainfall characteristics and stormwater quality was studied. Rainfall events were classified using a K-means clustering method. A rainfall event type-based (RTB) calibration approach was used to improve water quality model performance. Antecedent dry days, average rainfall intensity, and rainfall duration were the most critical rainfall characteristics affecting the event mean concentrations (EMCs) of total suspended solids (TSS), total nitrogen (TN), and total phosphorus (TP). The RTB calibration approach can improve water quality model accuracy. The calibrated stormwater quality parameters could be transferred to adjacent catchments with similar characteristics.

Secondly, to study the impacts of land cover on the simulation of stormwater runoff and pollutant loading, a land-cover based (LCB) PCSWMM model was built. The LCB approach performed better than the watershed delineation tool (WDT) approach in hydrological simulation. The two models showed comparable performances in simulation of TSS, TN, and TP. The LCB approach parameters could be regionalized based on land cover types. The hydrologic-hydraulic parameters

could potentially be transferred to similar catchments. The transferring of water quality parameters did not perform as satisfactory. The LCB approach could quantitively evaluate the contribution to runoff and pollutant loads of different land covers. Roads and roofs were found to be the major contributors to urban runoff and pollutants in the two urban catchments.

Thirdly, to study the effects of mixed land use on urban stormwater quality under different rainfall event types, the dataset from the two-year field monitoring program in four urban catchments in Calgary was utilized. EMC and event pollutant load (EPL) were employed to evaluate TSS and nutrients. EMC of TSS was positively correlated with most nutrient components. EPLs exhibited higher correlation compared to EMCs. Mixed land use could influence the generation of stormwater pollutants. Intense rainfall and long antecedent dry days could yield higher EMC and EPL and long rainfall duration could generate elevated EPL. Seasonal variations were found in EMC and EPL, with higher values in the spring and summer than the fall.

Fourthly, to investigate the effects of land use and rainfall conditions on the particle size distribution (PSD), the dataset from a six-year water sampling program across 15 study sites in Calgary was used to characterization of PSDs. The median particle size decreased in the order: paved residential, commercial, gravel lane residential, mixed land use, industrial and roads. Fine particles are the dominant particles of suspended sediments in runoff in Calgary. The impact of rainfall event types could vary depending on land use types. Long antecedent dry period tends to result in the accumulation of fine particles. High rainfall intensity and long duration could wash off more coarse particles. The season is a factor in the PSD of suspended sediment in runoff. The PSD in spring exhibits the finest particles. Particles in snowmelt are finer for the same land use than that during rainfall events.

Finally, to predict stormwater quality in data-deficient areas using data-driven models, a semi-supervised machine learning framework was proposed. This approach was applied in four catchments in Calgary to predict EMCs of TSS, TN and TP. This study demonstrated that machine learning is an effective tool for predicting stormwater quality. Relying on limited data and features may lead to overfitting. By integrating data from different catchments, the model performance could be enhanced. Consideration of catchment characteristics could improve the model capacity. The pseudo-labeling learning could enrich the training dataset, and this semi-supervised machine learning approach could elevate the model predictive ability. Rainfall characteristics are important variables for predicting all three pollutants, with varying focus on duration, amount, intensity and antecedent dry days.

Preface

This thesis is an original work by Haibin Yan. This study is a part of a collaborative research project by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the City of Calgary, Water Resources. It is presented in a paper format and consists of seven chapters.

Chapter 1 is a general introduction on the background and the objectives of this research.

Chapter 2 to 7 are the main contents of this thesis.

Chapter 2 was published as: Yan, H., Zhu, D.Z., Loewen, M.R., Zhang, W., Liang, S., Ahmed, S., van Duin, B., Mahmood, K., Zhao, S., 2023. Impact of rainfall characteristics on urban stormwater quality using data mining framework. Sci. Total Environ. 862, 160689.

Chapter 3 was published as: Yan, H., Fernandez, A., Zhu, D.Z., Zhang, W., Loewen, M.R., van Duin, B., Chen, L., Mahmood, K., Zhao, S., Jia, H., 2022. Land cover based simulation of urban stormwater runoff and pollutant loading. J. Environ. Manage. 303, 114147.

Chapter 4 has been submitted as a journal manuscript: Yan, H., Zhu, D.Z., Loewen, M.R., Zhang, W., Yang Y., Zhao, S., van Duin, B., Chen, L., Mahmood, K. 2023. Particle Size Distribution of Suspended Sediments in Urban Stormwater Runoff: Effect of Land Uses and Precipitation Conditions. Water Res. Manuscript ID WR77877, under review.

Chapter 5 is currently being prepared as a journal manuscript.

Chapter 6 is currently being prepared as a journal manuscript.

Chapter 7 discuss the practical application of this research in urban stormwater quality management and design.

Chapter 8 is a conclusion chapter. The implications of the results and the suggestions for future research on this topic are discussed.

I was responsible for the concept formation, data analysis, computational modelling, and manuscript composition. Dr. David Z. Zhu and Mark R. Loewen were the supervisory authors

involved with concept formation, data analysis and manuscript composition. Dr. Wenming Zhang contributed to data analysis and the manuscript edits of Chapters 2 through 6. Dr. Sherif Ahmed and Shuntian Liang contributed to the two-year field monitoring work in four catchments and data collection used in Chapter 2, 3, 4 and 6. They also contributed to the manuscript edits of Chapter 3. Arlette Fernandez contributed to the setup of PCSWMM model using WDT approach in Chapter 2 and 3. Mr. Bert van Duin, Chen Lei, Stacey Zhao and Khizar Mahmood facilitated data collection and contributed to manuscript edits of Chapters 2 through 6. Dr. Haifeng Jia and Dr. Yang Yang contributed to manuscript edits of Chapters 3 and 4, respectively. The City of Calgary provided the field monitoring data in Chapter 5.

Acknowledgements

First and foremost, I am deeply grateful to my supervisor, Dr. David Z. Zhu. His guidance, encouragement, and friendship have been the lighthouse in my academic journey. His wisdom and insights have been invaluable throughout this process. Equally, my gratitude extends to my co-supervisor, Dr. Mark Loewen, who has shown remarkable patience and provided guidance and suggestions in shaping my research career.

I would like to extend my appreciation to the committee members for contributions to my PhD candidacy and final exams: Dr. Caterina Valeo (University of Victoria), Dr. Kevin Devito, Dr. Tong Yu, Dr. Wenming Zhang. Their insights and feedback have been critical in refining both my candidacy and final exams. A special thanks to Dr. Zhang, who continuously contributed to the writing and publication of my papers.

My sincere thanks also go out to the staff in the City of Calgary: Bert van Duin, Stacey Zhao, Lei Chen, Khizar Mahmood and others. Their willingness to provide resources and their attention during the editing of my papers have been greatly appreciated.

Furthermore, to my colleagues, Dr. Sherif Ahmed, Shuntian Liang and Arlette Fernandez, I am very grateful for their dedication and collaborative spirit during the monitoring program.

This research is being funded by Natural Sciences and Engineering Research Council (NSERC) of Canada and the City of Calgary through a Collaborative Research and Development (CRD) grant. I also acknowledge the scholarships from the China Scholarship Council (CSC) and the University of Alberta.

Last, but certainly not least, I am eternally thankful to my family. Their unconditional support and belief in my studies have been the foundation upon which I have built my ambitions. Their love and encouragement have been my constant source of strength and inspiration.

List of Tables

List of Figures

Abbreviations

AB	Auburn Bay
ADD	Antecedent dry days
ANOVA	Analysis of variance
B	Build-up
C_1	Maximum buildup possible
C_2	Buildup rate constant
C_3	Wash-off coefficient
C_4	Wash-off exponent
CA	Catchment area
CO	Commercial
CR	Cranston
CS	Continuous simulation
CV	Coefficient of variation
D_{50}	Median particle size
DEM	Digital Elevation Model
EB	Event-based
EMC	Event mean concentrations
EPL	Event pollutant load

GA	Genetic Algorithm
GIS	Geographic Information System
GLR	Gravel lane residential
HSD	Honest significant difference
Imp	Imperviousness
IN	Industrial
K-S	Kolmogorov–Smirnov test
LCB	Land-cover based
LID	Low Impact Development
LS	Landscaped
LULC	Land-use/land-cover
MaxNumSplits	Maximum number of splits
MFFn	Mass first flush ratio
MFR	Multi-family residential
MinLeafSize	Minimum leaf size
ML	Machine learning
MLU	Mixed land use
NRP	Non-reactive Phosphorus
NSE	Nash-Sutcliffe efficiency

NSQD	National Stormwater Quality Database
NumTrees	Number of trees
OS	Open space
PCA	Principal Component Analysis
PR	Paved residential
PSD	Particle size distribution
q	Runoff rate per unit area
R^2	Coefficient of determination
RD	Rainfall duration
RDS	Road deposited sediment
RF	Random forest
RI-avg	Average rainfall intensity
RI-max	Maximum rainfall intensity
RMSE	Root Mean Square Error
RO	Royal Oak
RR	Rocky Ridge
RT	Total rainfall amount
RT-30	Initial 30-min total rainfall amount
RTB	Rainfall event type-based

SCM	Stormwater control measure
SFR	Single family residential
SRTC	Sensitivity-based Radio Tuning Calibration
SSE	Sum of squared errors
SVR	Support Vector Regression
S-W	Shapiro–Wilk test
SWMM	Storm Water Management Model
TDP	Total Dissolved Phosphorus
TKN	Total Kjeldahl Nitrogen
TN	Total nitrogen
TON	Total Organic Nitrogen
TP	Total phosphorus
T-peak	Time to peak rainfall intensity
TPP	Total Particulate Phosphorus
TR	Transportation
TRP	Total Reactive Phosphorus
TSS	Total suspended solids
W	Wash-off rate
WDT	Watershed delineation tool

Table of Contents

1. General introduction

1.1 Research background

Urbanization considerably changes the characteristics of urban stormwater runoff, including hydrology and water quality (Goonetilleke et al. 2005). The overdevelopment of urban areas has considerably changed the land use pattern. The surface permeability has decreased due to increased impervious areas such as roads, roofs, and parking lots. A variety of anthropogenic activities in urban areas generate a range of pollutants. With the increase of surface runoff, a number of pollutants accumulated on the surface enter the urban drainage system and directly enter the receiving water body, which poses severe threats to public health and the environment (Hernandez-Crespo et al. 2019).

In order to control stormwater runoff and improve the quality of runoff, many stormwater control measures (SCMs) have been developed, such as wet ponds and constructed stormwater wetlands (Ahilan et al. 2019). The efficient design of SCM is closely dependent on the knowledge of the response of urban stormwater quantity and quality to rainfall and catchment characteristics (Liu et al. 2013). In this regard, stormwater modeling outcomes provide designers with important guidance and dataset such as the prediction of storm-driven inflow hydrographs and the corresponding sediment/nutrients loads. Therefore, the accuracy of modeling approaches and the reliability modeling outcomes are of particular concern.

The accuracy of a stormwater model closely relies on the accuracy of the model replicating the natural system, especially for the pollutant build-up and wash-off process (Egodawatta et al. 2007). Pollutant build-up refers to pollutant accumulation on urban surfaces such as roads and roofs during dry weather periods (Wang et al. 2020). Pollutant wash-off refers to the phenomenon that the accumulated pollutants are mobilized by the kinetic energy of raindrops and the turbulence created by stormwater runoff (Muthusamy et al. 2018). Rainfall and catchment characteristics, especially land-use/land-cover (LULC), are among the most important influential factors concerning build-up/wash-off processes (Li et al. 2015). Therefore, stormwater quality modelings are primarily impacted by these factors.

Pollutant build-up and wash-off processes are considerably affected by rainfall characteristics. The impact of rainfall characteristics on water quality is very complex. For example, Egodawatta et al. (2007) reported that the relationship between pollutant wash-off and rainfall intensity is a stepwise function. The influence of rainfall intensity on pollutant wash-off could differ based on intensity thresholds. However, the models such as PCSWMM usually use simple continuous functions of rainfall characteristics to describe the processes, ignoring the different responses of water quality to different rainfall. Data mining techniques have potential to explore the water quality responses to their influential factors and apply the knowledge to stormwater models.

Land cover refers to the physical material on the land surface. Urbanization increases impervious land cover, resulting in an increase in runoff which creates additional avenues for the transportation of pollutants from the landscape into waterbodies (Zhou et al. 2021). Accurate delineation of the catchment is required to achieve accurate and successful applications of stormwater models. Subcatchments covered by different land cover types are nonhomogeneous and a single set of parameters cannot describe their spatial heterogeneity. Many studies have shown that spatial heterogeneity can affect the accuracy of runoff simulation (Elliott et al. 2009, Krebs et al. 2014, Leandro et al. 2016). Therefore, it is important to study the effects of land cover heterogeneity on simulating urban stormwater runoff and pollutant loading from the modeling perspective.

Different land use types have various anthropogenic activities, such as traffic volume, vehicle types, vegetation cover and maintenance activities and population density, playing a vital role in generating pollution (Wijesiri et al. 2018). Many studies have investigated the impact of single land use type on urban stormwater quality, but the effects of mixed land uses are seldom investigated (Miguntanna et al. 2010, Li et al. 2015, Yazdi et al. 2021). Typically, the water quality from different land uses was directly compared in the existing studies, ignoring the influences of rainfall characteristics. However, the water quality responses on land use may vary with rainfall event types. In addition, land uses have impacts on the form of nutrients (such as organic, inorganic, dissolved and particle forms), and the nutrient composition is also essential for the effective treatment of urban stormwater (Taylor et al. 2005, Miguntanna et al. 2013). Though nutrient composition from different urban land uses has been reported in the literature (Yang and Toor 2017, Lucke et al. 2018), no specific attention has been paid to the temporal variation of nutrient

composition under different rainfall conditions. Therefore, the nutrient composition from mixed land use during different rainfall events need to be investigated.

Road deposited sediment (RDS), also known as surface particles or road dust, is the sediment that accumulates on paved surfaces such as roads and parking lots (Wang et al. 2020). RDS on the urban surface is a significant stormwater pollution source and the primary carrier of other particulate pollutants such as particulate nutrients and heavy metals constituents (Zhao and Li 2013, Shen et al. 2016). Land use types can considerably impact PSD (Bian and Zhu 2009). Some studies have investigated PSD from different land use types (Bian and Zhu 2009, Shen et al. 2016). PSD of suspended sediments in runoff can be also impacted by rainfall characteristics. Rainfall characteristics has been shown to influence PSD during sediment build-up and wash-off processes (Wijesiri et al. 2016). However, there are knowledge gaps in the study of the impacts of rainfall event types on PSD in runoff. Therefore, there is need to examine the effects of land use and precipitation conditions on PSD in stormwater and snowmelt runoff.

Data-driven models, especially those utilizing machine learning (ML) algorithms, are based on statistical relationships derived from observed data to apply in environmental research (Granata et al., 2017, Ahmed et al., 2019, Behrouz et al., 2022). Process-driven models (or physically based) (e.g., PCSWMM) are designed to represent the physical processes involved in stormwater generation and pollutant transport. Data-driven models have emerged as a simpler alternative of process-driven models due to their fewer extensive data requirements, relying predominantly on weather, land use and water quality data. For example, using six factors as input features, effective machine learning models were developed to predict total suspended solids (TSS) (Moeini et al., 2021). Their predictive power is contingent on the quantity and quality of training data (Zhi et al., 2021). Collecting stormwater quality data, is often time-consuming and expensive, leading to a lack of available stormwater quality data (McKenzie et al., 2013). The majority of urban catchments suffer from a dearth of comprehensive water quality datasets, which can lead to poor model performance and generalization in ML applications (Zhu et al., 2022). Therefore, there is need to develop a machine learning framework capable of predicting urban stormwater quality in data-deficient regions in Calgary.

1.2 Thesis outline

This thesis is written in paper format and composed of five contributions. This research aims to understand the responses of urban stormwater quality to rainfall characteristics and LULC and improve the performance of urban stormwater quality models. Each study is presented in a separate chapter. Following is the brief introduction of each chapter.

Chapter 2 proposed a data mining framework to study the impacts of rainfall characteristics. A rainfall event type-based calibration approach was developed to improve water quality model performance.

Chapter 3 built a new land-cover based (LCB) PCSWMM model, and a systematic comparison with the traditional watershed delineation tool (WDT) model was made to study the impacts of land cover on the simulation of stormwater runoff and pollutant loading.

Chapter 4 investigated the effects of mixed land use on urban stormwater quality (TSS, TN and TP) under different rainfall event types using the dataset from a comprehensive field monitoring program in Calgary, Alberta, Canada from 2018 to 2019.

Chapter 5 studied the effects of land use and rainfall conditions on the particle size distribution (PSD) of suspended sediments in stormwater and snowmelt runoff utilizing the dataset from a six-year monitoring program across 15 study sites in Calgary, Alberta, Canada from 2016 to 2021.

Chapter 6 proposed a semi-supervised machine learning (random forest) framework to improve the prediction performance of models in data-deficient areas. This approach was applied in four catchments in Calgary, Alberta, Canada to predict EMC of TSS, TN and TP.

Chapter 7 discussed the practical application of this research in urban stormwater quality management and design.

2. Impact of rainfall characteristics on urban stormwater quality using data mining framework

2.1 Introduction

Urban stormwater quality models are essential for effective stormwater management (Liu et al., 2013, Yan et al., 2022). Stormwater quality models may generally be classified into process-driven (or physically based) models such as PCSWMM and data-driven (or data-based) models. Data-driven models are computer tools that can use algorithms to model stormwater quality parameters (Orouji et al., 2013). Process-driven stormwater quality models can better capture physical processes such as pollutant build-up/wash-of process in a catchment compared to data-driven models (Damle and Yalcin 2007). Stormwater quality is greatly affected by rainfall characteristics (Liu et al., 2013), so it is vital to fully understand the response of the stormwater quality to rainfall characteristics for process-driven models.

Data mining involves extracting meaningful information and discovering patterns in data sets and is widely used for outlier detection and clustering analysis (Deng and Wang 2017). Data-driven models usually use data mining techniques to explore the nonlinear relationship between various input conditions and outputs (Kusiak et al., 2013, Deng and Wang 2017). Data mining techniques could fully use the existing data and process the complex nonlinear relationships between input and output data (Saber et al., 2019). The use of data mining techniques to explore the response of stormwater quality to its governing factors and applying the knowledge gained to advance process-driven models is a promising method to improve the performance of stormwater quality models. Very few studies focus on using data mining techniques to improve the performance of stormwater quality models.

Pollutant build-up and wash-off processes, which are the most common representation of the behaviors of pollutants within a catchment in process-driven models, are considerably affected by rainfall characteristics (Goonetilleke et al., 2005, Li et al., 2015, Wijesiri et al., 2015, Gong et al., 2016). The impacts of rainfall characteristics on stormwater quality are complex. For example, the relationship between pollutant wash-off and rainfall intensity may differ based on intensity thresholds (Egodawatta et al., 2007, Liu et al., 2013). However, models such as PCSWMM usually

only employ wash-off functions with the same parameter set to describe the processes for all kinds of rainfall events, ignoring the different responses of the stormwater quality to different rainfall characteristics (Rossman et al., 2010).

Due to the complex effects of rainfall characteristics on stormwater quality, the response of stormwater quality shows a highly nonlinear behavior on an event scale, which results in significant challenges with model calibration (Bonhomme and Petrucci 2017, Tscheikner-Gratl et al., 2019). A number of studies have focused on using data mining techniques to improve the performance of hydrological models. For example, researchers developed a new multi-period model using a Fuzzy C-mean Clustering method to improve the calibration of hydrological models based on hydroclimatic similarities (Zhang et al., 2011). Another study integrated data mining techniques to develop a clustering framework for preprocessing climatic-land surface indices such as precipitation, evaporation, antecedent streamflow, and the runoff coefficient (Lan et al., 2018). Hydrological data were clustered into sub-annual periods for calibration to simulate seasonal dynamic behaviors. Clustering methods were used to develop a discretization methodology that helped portion the simulation periods and subsequent calibration (Sudheer 2021). These studies employed data mining technologies to divide long time series data into sub-periods for independent calibration to improve model performance. Data mining technology has been widely used in studying the impact of rainfall characteristics on stormwater quality (Liu et al., 2013, Li et al., 2015, Deng and Wang 2017, Liu et al., 2018, Jin et al., 2020). However, it has been rarely used in stormwater quality simulation, especially in process-driven models.

Stormwater quality modeling relies heavily on data availability and reliability, as no models can be successfully developed without local data (Vaze and Chiew 2003). Collecting data, especially stormwater quality data, is often time-consuming and expensive, leading to a lack of available stormwater quality data (Yan et al., 2022). Due to the lack of data, opinions regarding the impacts of rainfall characteristics on the pollutant build-up/wash-off mechanisms may diverge (Wijesiri et al., 2015, Bonhomme and Petrucci 2017). The quality of the collected data is not guaranteed for various reasons (McKenzie and Young 2013). Model calibration requires a large amount of data, and the reliability of data affects the accuracy of the model. Therefore, the collected data should be preprocessed to evaluate the reliability while retaining the maximum amount of information. In addition, stormwater quality models are challenging to apply in ungauged catchments due to the

lack of stormwater quality data. As such, transferring the parameters of a calibrated stormwater quality model from gauged catchments to ungauged catchments is a common approach to create stormwater quality models for ungauged catchments.

Stormwater quality models are often divided into continuous simulation (CS) models and event-based (EB) models based on their temporal domain, each of which has its own limitation (Hossain et al., 2019). CS approach apply an "optimal" set of parameters to cover the long-time scale. However, a single parameter set has major challenges in representing anything beyond the average stormwater quality response, as the current generation of models does not account for the time-varying nature of the stormwater quality responses (Guse et al., 2014). For EB models, parameters can be calibrated based on a single event or several rainfall events (Wang and Altunkaynak 2012, Tang et al., 2021). Compared to CS approach, EB models need to consider antecedent conditions for each event. Many studies have considered antecedent dry days in predicting pollutant load on urban impervious surface (Soonthornnonda and Christensen., 2008, Liu et al., 2015, Charters et al., 2020). However, when considering urban pervious areas like grass land, some other antecedent conditions such as initial soil moisture content could affect the generation of runoff, which in turn affect model estimates of pollutant loading. These parameters need to be estimated for each rainfall event in the EB models, which poses a great challenge to the reliability of the model. Therefore, a rainfall event type-based continuous simulation approach could overcome the above limitations. Rainfall event type-based models yield an optimal parameter set for each rainfall event type and thus could account for the temporal variation of stormwater quality response.

Data mining techniques have potential in the application of process-driven stormwater quality models. This study aims to propose a data mining framework for calibrating rainfall event type-based stormwater quality models considering the response of urban stormwater quality to rainfall characteristics. The specific objectives are as follows: (1) apply a statistical analysis method to detect outliers; (2) use Principal Component Analysis (PCA) to study the impacts of rainfall characteristics on stormwater quality and determine the representative rainfall characteristics; (3) use a K-means clustering method to classify rainfall based on the selected rainfall characteristics and study the stormwater quality for different rainfall event types; (4) develop a new approach to improve model calibration based on potential rainfall event types; (5) test the transferability of the parameters of the resulting stormwater quality model obtained from the above new approach. The

PCSWMM model was applied to four catchments in Calgary, Alberta, Canada, to demonstrate the effectiveness of the proposed framework.

2.2 Study sites and data collection

2.2.1 Description of the study areas

The study sites include four urban residential catchments located in Calgary, Alberta, Canada, which are the Royal Oak (RO), Rocky Ridge (RR), Cranston (CR) and Auburn Bay (AB) catchments, as shown in Fig. 2-1. Calgary is located in a semi-arid, cold temperate climate region, with cold winters and mild to warm summers. The four catchments are mature without a significant new development in the past five years. Table 2-1 shows the catchment characteristics of the study areas. The land use types in all four catchments are predominantly residential (i.e., more than 50%), the AB and CR catchments are more complex. The AB and CR catchment contains partial institutional and commercial land use in addition to the residential, transportation and open space land uses. The imperviousness of these catchments is around 50%, among which RR has the lowest imperviousness at 48% and AB has the highest imperviousness at 58%. RR and RO have similar catchment areas of 14.0 and 14.5 ha, respectively, and they are located only 300 m apart. The land cover types are primarily composed of asphalt/concrete, gravel lanes, landscaped area and roofs. Both RO and RR have a wetland to receive the generated runoff. CR and AB have larger areas (117.4 and 222.6 ha, respectively) and are approximately 1.5 km away from each other. The generated runoff in the CR catchment is directed to a wet pond. An upstream pond in the AB catchment treats approximately 69 ha drainage area before the runoff drains to the piping system and is discharged to a wet pond.

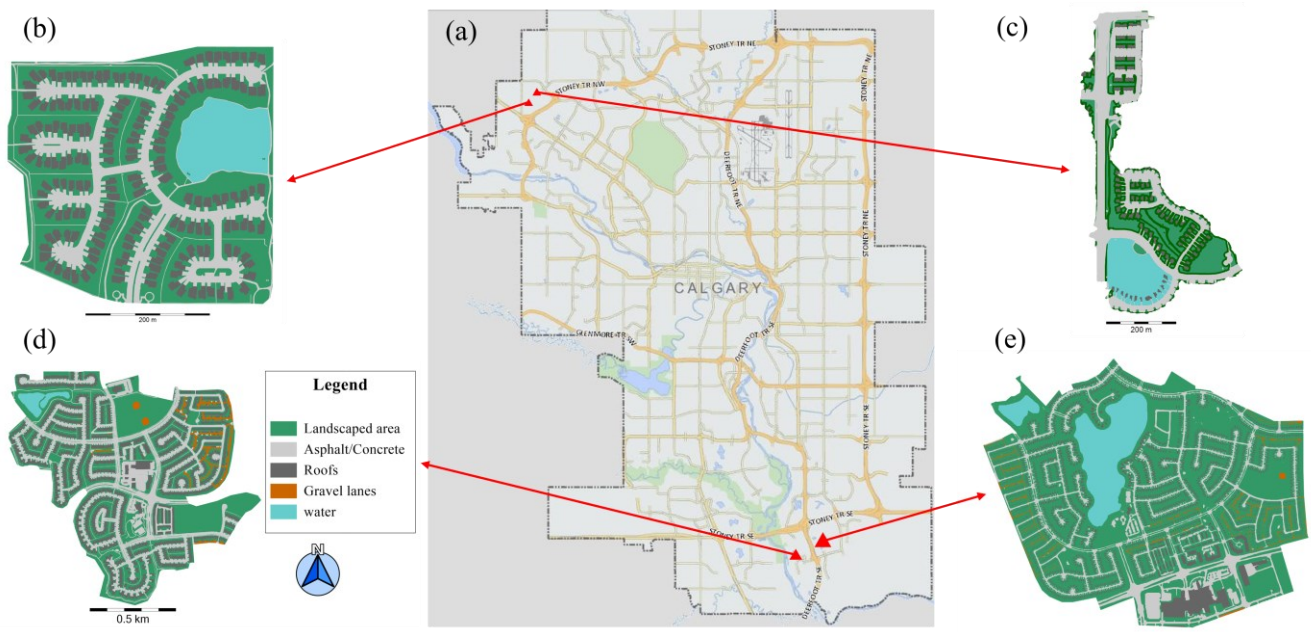


Fig. 2-1. Location photos of the study sites: (a) locations of four catchments in Calgary (Google, 2022); (b) RR catchment; (c) RO catchment; (d) CR catchment; (e) AB catchment.

Table 2-1. Catchment characteristics of the study areas.

Catchment		RR	RO	CR	AB
Area (ha)		14	14.5	118.5	222.6
Imperviousness (%)		48	51	52	58
Land Use (%)	Residential	72	58	61	50
	Transportation	18	28	19	20
	Open space	10	14	13	18
	Institutional	0	0	4	6
	Commercial	0	0	3	6
Land Cover (%)	Asphalt/Concrete	25	37	26	28
	Gravel lanes	0	0	2	3
	Landscaped area	52	49	51	46
	Roofs	23	14	21	23

2.2.2 Field measurements

Field monitoring was conducted during the ice-free seasons (May - October) of 2018 and 2019. A weather station, flow meters and autosamplers were installed in each catchment to collect data.

The weather station consisted of a HOBO RX300 Remote Monitoring Station equipped with a HOBO RG3-M Data Logger (Onset Computer Corp., U.S.). Rainfall was measured with a tipping bucket rain gauge, triggered at a minimum of 0.2 mm cumulative rainfall with data recorded at 5-minute intervals.

Flow meters and autosamplers were installed in the inlets of the stormwater wetlands and wet ponds. Table 2-2 provides information about the instruments that were used at the outlets of the four catchments during the two years field program, 2018 and 2019. Six ISCO 6712 autosamplers (Teledyne ISCO, USA) equipped with ISCO 750 area/velocity flow modules and four ISCO 2150 area/velocity flow modules were installed. The flow modules sensors were installed in the inlet pipes to automatically calculate the inflow rates from the recorded water depths/velocities and the given pipes diameters.

Table 2-2. Equipment installation at four catchments (RR, RO, AB and CR).

Catchment	2018	2019
RR	1×HOBO RX300 1×6712 Autosampler 1×750 Flow sensor	1×HOBO RX300 1×6712 Autosampler 1×750 Flow sensor
RO	1×HOBO RX300 1×6712 Autosampler 1×750 Flow sensor 1×2150 Flow sensor	1×HOBO RX300 1×6712 Autosampler 1×750 Flow sensor 1×2150 Flow sensor
CR	1×HOBO RX300 2×6712 Autosampler 2×750 Flow sensor	1×HOBO RX300 2×6712 Autosampler 2×750 Flow sensor 1×2150 Flow sensor
AB	1×HOBO RX300 2×6712 Autosampler 2×750 Flow sensor 2×2150 Flow sensor	1×HOBO RX300 2×6712 Autosampler 2×750 Flow sensor 2×2150 Flow sensor

The collected water samples were analyzed for total suspended solids (TSS), total nitrogen (TN), and total phosphorus (TP). The flow meters monitored flow rates continuously at 5-min intervals. The ISCO 6712 autosamplers collected water samples with a maximum collection capability of 24×1 L bottles. A flow-paced program was used to collect samples that would be representative of the flow for the day of event in question. The ISCO 6712 autosampler was matched with the ISCO 750 AV flow module or the ISCO 2150 flow module, which recorded the water level and flow velocity continuously to trigger and pace the ISCO 6712 sampler program. Given the limitations of the maximum collection capability of the ISCO 6712 unit and the desire to maximize the volume of water quality samples during a runoff event, three flow pacing configurations were used: low flow pacing, high flow pacing and extreme event pacing. Low flow pacing targets events in the ~5 mm to 10 mm range, high flow pacing targets events up to 20 mm, and extreme event pacing targets isolated large events more than 20 mm. To select the appropriate flow pacing configuration, weather forecasts and expected precipitation were monitored. The water level to trigger or activate the ISCO 6723 unit was then adjusted at the AV sensor location.

In this study, the aliquots samples during each storm event were combined to generate composite samples, which represented approximately 75% of the storm runoff volume, resulting in event mean concentrations (EMC). To preserve sample quality, composite sample bottles are placed in a cooler until samples are delivered to the laboratory within 24-hours of sample collection. Further, database analysts ensured that all data were consolidated in a consistent manner, with over 200 quality checks performed on the data prior to import into the database. The number of captured events for the individual catchments was: Rocky Ridge (17), Royal Oak (15), Cranston (18) and Auburn Bay (11), which amounted to 61 events in total. The rainfall data and EMCs of TSS, TN and TP are summarized in Table A1 in Appendix A.

2.3 Methodology of data mining framework

The proposed data mining framework that used MATLAB as a platform is illustrated in Fig. 2-2, which involves five main steps:

(1) Pre-select appropriate rainfall and stormwater quality indices for the subsequent analysis.
(2) Detect outliers within the stormwater quality data using a boxplot approach and analyze the outliers before deleting/keeping them to obtain the most representative data set.

(3) Analyze the correlation between the rainfall indices and stormwater quality indices by PCA to screen out the key rainfall parameters that can represent the rainfall characteristics and reduce the dimension of the data set.

(4) Classify the rainfall event types applying a K-means clustering method based on the reduced rainfall indices.

(5) Calibrate the RTB PCSWMM model based on the different rainfall event types and compare its model performance with conventional CS approach.

The data mining components (boxplot approach, principal component analysis and K-means clustering algorithm) were conducted using MATLAB as a platform. The outcome of data mining (i.e., rainfall events clustering) will be used as foundation of rainfall event type-based calibration of PCSWMM. Each of these steps is explained in the sections below.

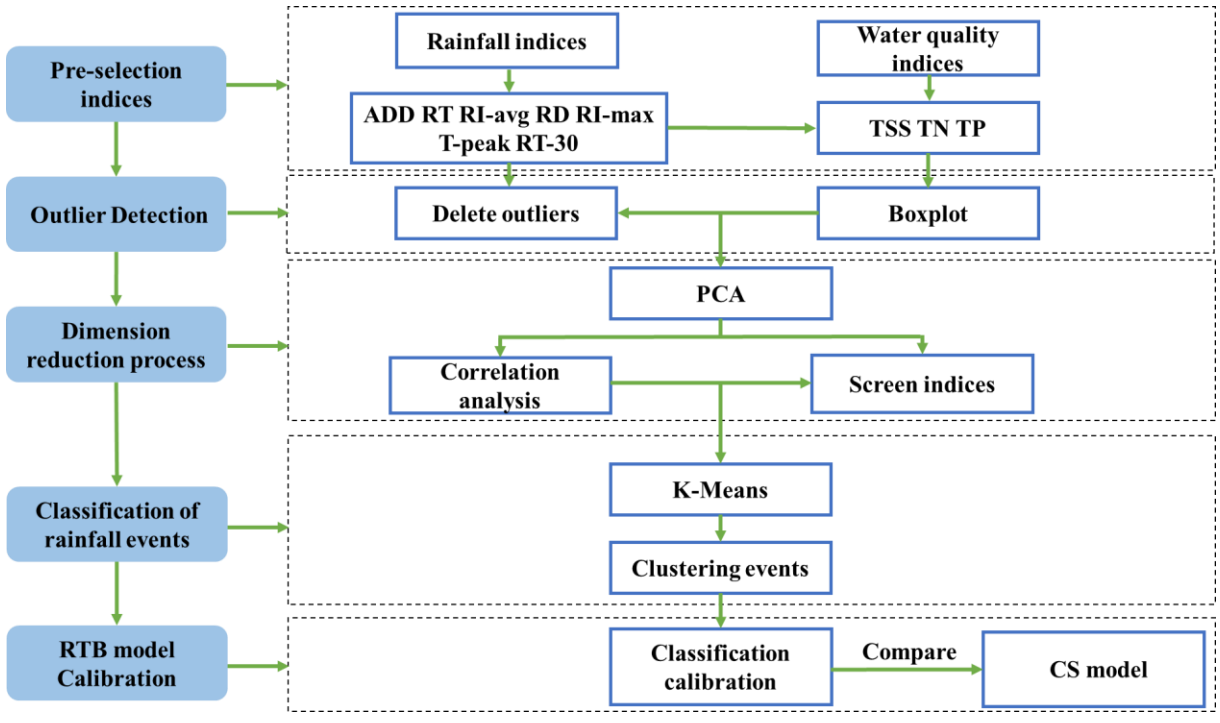


Fig. 2-2. The proposed data mining framework for model calibration.

2.3.1 Pre-selection of indices

In order to establish the inherent relationship between rainfall characteristics and stormwater quality, it is necessary to select appropriate rainfall indices and corresponding stormwater quality indices. Potentially suitable rainfall indices were initially selected based on information from

studies in the literature. The number of antecedent dry days had been found to exert a significant influence on pollutant concentration for residential land use in stormwater runoff (Maestre and Pitt, 2006, Murphy et al., 2015). The total rainfall amount, average rainfall intensity and duration were significant factors for stormwater runoff quality (Salim et al., 2019). The maximum intensity was the critical factor defining EMCs for TSS, TN, TP, and chemical oxygen demand (Li et al., 2015). The rainfall process (hyetograph) can considerably affect the shape and peak value of TSS pollutograph (Gong et al., 2016). Time to peak rainfall intensity, which can reflect the rainfall process, is a potential parameter affecting stormwater quality. The initial runoff in urban catchments during rainfall events contained the highest pollutant levels due to the first flush effect, so initial rainfall was also a potential parameter affecting stormwater quality (Li et al., 2015). Sampling during the first 30 minutes of the event is an effective approach to evaluate the first flush effect (Maestre and Pitt, 2004). Seven rainfall indices were selected as candidate rainfall indices for this study, including the number of antecedent dry days (ADD), total rainfall amount (RT), average rainfall intensity (RI-avg), rainfall duration (RD), maximum rainfall intensity (RI-max), time to peak rainfall intensity (T-peak) and initial 30-min total rainfall amount (RT-30).

Three common stormwater pollutants of interest to regulators, and thus designers and governmental agencies, are sediment, nitrogen and phosphorus. They are typically quantified in stormwater runoff using total suspended solids (TSS), total nitrogen (TN) and total phosphorus (TP). EMC is a critical analytical parameter for assessing stormwater quality, and EMCs of different catchments are comparable. Therefore, EMCs of TSS, TN and TP were selected as the stormwater quality indices of interest for this study.

2.3.2 Outliers detection

Outliers are defined as data points that have considerably different characteristics from other observations (Hawkins, 1980). Outliers often exhibit unreasonable characteristics in the dataset, leading to biased conclusions when using data mining techniques such as the K-means clustering algorithm. Therefore, it is necessary to detect these outliers in stormwater quality data and deal with them appropriately during data exploration. A boxplot approach is a straightforward but effective way to visualize outliers for univariate data (Li et al., 2016). A boxplot consists of five values, the lower extreme, the first quartile (Q_1), the second quartile (Q_2), the third quartile (Q_3)

and the upper extreme (Smiti 2020). The interquartile (IQR) equals $Q_3 - Q_1$. The lower extreme is assigned as $Q_1 - 1.5IQR$ and the upper extreme is assigned as $Q_3 + 1.5IQR$. Any data point above the upper or below the lower extreme can be identified as an outlier (Smiti 2020). Therefore, the boxplot approach was chosen to detect stormwater quality outliers.

Stormwater quality is affected by rainfall characteristics, so the stormwater quality data cannot be analyzed in isolation. They must be analyzed with the rainfall data to make the data set more representative while ensuring data integrity to the greatest extent. Boxplots of the EMCs for TSS, TN and TP are shown in Fig. B1 in Appendix B: in four study sites, there were 6, 2 and 2 outliers in total for TSS, TN and TP, respectively. For example, two outliers (1328 mg/L and 562 mg/L) for TSS represent events that occurred on May 2nd in the Royal Oak and Rocky Ridge catchments, respectively. These two events were small rainfall events with short antecedent dry days, which should not have produced such large EMCs for TSS. This was likely due to the pollutants that accumulated over the winter being flushed down the drainage systems. This allows large concentrations of pollutants to occur during small rainfall events. Therefore, these events should be eliminated from this research. In the end, four events were deleted, resulting in a 57×10 data set (i.e., 57 events with 10 rainfall and stormwater quality indices).

2.3.3 Dimension reduction processing

Principal component analysis (PCA) is a dimensionality reduction method that can transform multiple indices into a few comprehensive indices (Granato et al., 2018). PCA is a useful for analyzing relationships between objects and variables and has been applied in stormwater quality analysis (Liu et al., 2018, Yazdi et al., 2021). As part of a visual representation, an axis or principal component (PC) is usually used as a factor to construct a two-dimensional or three-dimensional projection of a sample, i.e., a biplot. The variances of PCs were calculated in MATLAB and when the accumulated variances of the first few PCs exceed 60 %, these PCs are considered sufficient to present the majority of the original information (Yazdi et al., 2021). In this study, subsequent to the pre-selection of indices and outlier detection, the ten indices (i.e., rainfall and stormwater quality indices) were selected as variables, and the 57 monitored rainfall events were selected as objects. Various biplots were used to analyze the influence of rainfall characteristics on stormwater quality, and the most important rainfall parameters were screened out. If the angles between the

vectors in a PCA biplot are less than 90 degrees, the variables are considered correlated (Liu, 2015). An obtuse angle indicates a low correlation, a 90-degree angle indicates that the variables are uncorrelated, while a 180-degree angle indicates a negative correlation. The vector length represents the importance of the variable in the individual principal components (Borris et al., 2014). The selected key rainfall indices were included for further analysis (Section 3.4).

Multicollinearity describes the correlation among several independent variables (Alin, 2010). Multicollinearity may have negative effects and therefore, it is preferable to detect multicollinearity in advance and use uncorrelated variables in data analysis (Mansfield and Billy, 2012). Use of PCA prevented the multicollinearity and eliminated the redundant information among the rainfall indices. A correlation matrix analysis was used to verify the results of the PCA. The correlation was divided into five types based on Pearson correlation: 1) .00 to .30 (.00 to -.30): negligible correlation, 2) .30 to .50 (-.30 to -.50): low positive (negative) correlation, 3) .50 to .70 (-.50 to -.70): moderate positive (negative) correlation, 4) .70 to .90 (-.70 to –.90): high positive (negative) correlation, and 5) .90 to 1.00 (-.90 to -1.00): very high positive (negative) correlation (Mukaka, 2012).

2.3.4 Classification of rainfall events

Rainfall indices provide the basis for dividing rainfall events into different clusters based on rainfall similarity. The K-means clustering algorithm is an unsupervised learning method widely used to identify groups in a dataset (Choi and Beven, 2007). As a point clustering method, the K-means clustering algorithm cannot be directly applied to the clustering analysis of high-dimensional data. As such, PCA (Section 3.3) was conducted to reduce the rainfall dimension. K-means clustering divides the data into K clusters by treating each observation in the data as an object with a location in space (Steinley and Michael, 2007). Using this approach, a partition is found in which objects in each cluster are as close to each other as possible and as far away as possible from objects in other clusters. The best cluster number is determined through an elbow method, which is an iterative algorithm using the sum of squared errors (SSE) as a performance indicator (Yuan and Yang, 2019). SSE will decrease with an increase in the cluster number. In the process of SSE reduction, there is an inflection point or elbow point where the rate of decline suddenly slows down, which is considered the optimal cluster number (Yuan and Yang, 2019).

The first flush in urban runoff has been an essential but controversial phenomenon among many researchers (Bach et al., 2010). First flush effects showed a relation with EMCs, but the relationship of the first flush and rainfall characteristics is unclear, and no rule could be employed universally (Li et al., 2015, Mamun et al., 2020, Chaudhary et al., 2022). Therefore, this research collected some discrete water quality samples to assess the first flush effects. The first flush effects can be quantified using a mass first flush ratio (MFFn) (Li et al., 2015, Mamun et al., 2020). MFFn is a dimensionless indicator that can quantify the relationship between cumulative runoff and pollutant load. MFFn is defined as follows:

$$MFFn = \frac{\int_0^{t_i} C_t Q_t dt / M}{\int_0^{t_i} Q_t / V} \quad (2\text{-}1)$$

where n represents the percentage of the runoff, ranging from 0% to 100%; t_i is the time up to point n in the event; C_t is the pollutant concentration at time t; Q_t is the runoff flow rate at time t; M is the total pollutant mass during the event; V is the total runoff volume.

2.3.5 Rainfall event type-based (RTB) PCSWMM model calibration and validation

In this study, an RTB approach was established based on the clustering results of the rainfall indices and using the PCSWMM model as a foundation. The conventional CS approach runs PCSWMM continuously for a long-time series, and the water quality parameters are calibrated for all observed events. The RTB approach is also a long-term continuous simulation, but the water quality parameters are only calibrated for events of a single rainfall event type. For example, if the number of rainfall event types is four, the RTB approach will run four times for the entire period and four water quality parameter sets will be optimized independently. The pollutant build-up process in PCSWMM is a function of the number of antecedent dry days which are included in the rainfall input data. In this long-term continuous simulation, antecedent conditions will also be accurately reflected. The RTB approach simulations will provide the optimal water quality parameter set for different rainfall event types and avoid having to set initial conditions (i.e., avoid the hot start problem in PCSWMM). Therefore, the RTB approach is expected to take advantage of both CS (accounting for the antecedent conditions) and EB models (calibrating parameters for specific events). To evaluate the performance of the RTB approach, a conventional CS approach was created as a comparison. The CS approach was run for the entire period to obtain an optimal

set of water quality parameter. An accurate hydrological simulation in PCSWMM is required to ensure the accuracy of the stormwater quality simulation, so the hydrological simulation was calibrated first. The hydrological simulation for a specific period will be considerably affected by the antecedent conditions, leading to unreliable simulation results (James 2010). This kind of approach will result in a discontinuous hydrological simulation, which cannot be used for continuous water quality simulation. Therefore, the hydrological parameters are calibrated for the entire period rather than for different rainfall event types.

Two main components of the water quality routines in PCSWMM are the build-up and wash-off of pollutants from a given land cover type: asphalt/concrete, gravel lanes, landscaped area and roofs. Users can select three build-up functions (power function, exponential function and saturation function) and three wash-off functions (EMCs, rating curves and exponential function) in PCSWMM. EMCs and rating curves cannot account for the build-up process. In contrast to EMCs and rating curves, exponential curves have a wash-off load that is proportional to the runoff rate and the amount of pollutant remaining on the catchment (Rossman et al., 2016). Build-up functions are all based on a certain cumulative velocity approaching maximum cumulant. In the literature, exponential build-up equations are widely used to simulate pollutants (Tu et al., 2018). Therefore, this study only modeled the stormwater quality using the exponential build-up and wash-off functions.

The exponential build-up function is given by

$$B = C_1(1 - e^{-C_2 t}) \tag{2-2}$$

where B = buildup (mass per unit area), C_1 = maximum buildup possible (mass per unit area or curb length), C_2 = buildup rate constant (1/day) and t = buildup time interval (day). This function is applicable when the buildup follows an exponential growth that asymptotically approaches a maximum limit (James et al., 2010).

The exponential wash-off function is given by

$$W = C_3 q^{C_4} B \tag{2-3}$$

where W = wash-off rate (mass/hr), C_3 = wash-off coefficient, C_4 = wash-off exponent and q = runoff rate per unit area (mm/hr) (James et al., 2010).

The Sensitivity-based Radio Tuning Calibration (SRTC) tool, a built-in tool in PCSWWM, was used to calibrate the parameters. The user inputs a percentage of estimated uncertainty (e.g., +/- 30%) for each parameter that is considered for calibration. This gives a low-end value and a high-end value for each parameter. The SRTC tool executes a model run using the high and low values of the designated uncertainty range, generating sensitivity gradients for each parameter defined for calibration. The Cranston catchment has the most events (i.e., 18 events), so the events in 2019 (i.e., 12 events) in the Cranston catchment were used to calibrate the water quality models, while the events in 2018 (i.e., 6 events) in the Cranston catchment were used to validate the water quality models subsequently.

The Nash-Sutcliffe efficiency (NSE) and coefficient of determination (R^2) were chosen to assess the agreement between the simulated and observed flow rate. Model performance was evaluated on the following criteria: very good (NSE > 0.65, R^2> 0.8), good (0.54 < NSE < 0.65, 0.7 < R^2< 0.8) and satisfactory (0.5 < NSE < 0.54, 0.6 < R^2< 0.7) (Donigian et al., 2002, Rai et al., 2017). The absolute value of relative error between simulated and observed pollutant concentrations was used to evaluate the model performance for stormwater quality. A criterion of 50% for the simulated pollutant concentrations was assumed to be an acceptable deviation from the measured value with a targeted relative error lower than 25% (Tuomela et al., 2019).

To test the transferability of the stormwater quality model's parameters obtained from the calibration of the CR catchment to the other catchments, and to compare the performance of the CS and RTB approach on parameter transferability, the AB, RO and RR catchments were considered as ungauged catchments. The stormwater quality model's parameters resulting from calibration of the RTB and CS approach in the CR catchment were directly transferred to the AB, RO and RR catchments. To avoid the influence of the hydrological simulation on the stormwater quality simulation, the parameters of the hydrological model for the AB, RO and RR catchments were calibrated independently rather than transferring from the CR catchment.

2.4 Results and discussion

2.4.1 Impacts of rainfall characteristics on stormwater quality

Fig. 2-3 reveals the underlying and complex associations among the selected rainfall and stormwater quality indices. The first two PCs explained 42.9% and 18.0% of the data variance, respectively, leading to a total of approximately 60.9%. Therefore, the first two PCs were considered to be adequate to construct biplot to visualize PCA results. In Fig. 2-3, ADD is projected to the positive PC1 axis and shows an acute angle with the stormwater quality indices, which indicates ADD is positively correlated with pollutant EMC values. This implies that ADD is a important factor affecting the buildup process of pollutants. RI-max, RI-avg and RT-30 strongly correlate because the angles between these vectors are very small. They are also positively correlated with pollutant EMCs as they have small angles with pollutant EMCs. The angles of RT-30 and pollutant EMCs are less than the angles of RI-max, RI-avg with pollutant EMCs, which indicates the effect of RT-30 on pollutant EMCs is greater than that of RI-max and RI-avg. This could be explained by the first flush effect, where most pollutants are usually removed by runoff in the initial phase (Perera et al., 2021, Chaudhary et al., 2022). Kim et al (2007) reported that pollutant concentrations decreased considerably in the initial 30–40-minute storm period with the greatest decrease rate occurring in the initial 30-minute period. The angles of RD, T-peak and EMCs are close to 180º, indicating that RD and T-peak are inversely correlated with EMCs. A similar conclusion was drawn by Pitt et al. (2005) and Kim et al. (2007). They found that pollutant concentration decreases with rainfall duration. The angles between RT and other indices are almost 90º, indicating that RT has a negligible relationship with other parameters since total rainfall is determined by RI-avg and RD. RT cannot be explained by a single parameter. Pollutant EMCs depend on the available initial pollutants and the kinetic energy provided by the rainfall (Wijesiri et al., 2016). ADD affects the available pollutants, and the rain intensity determines the kinetic energy, so the effect of the total rainfall can be replaced by other parameters.

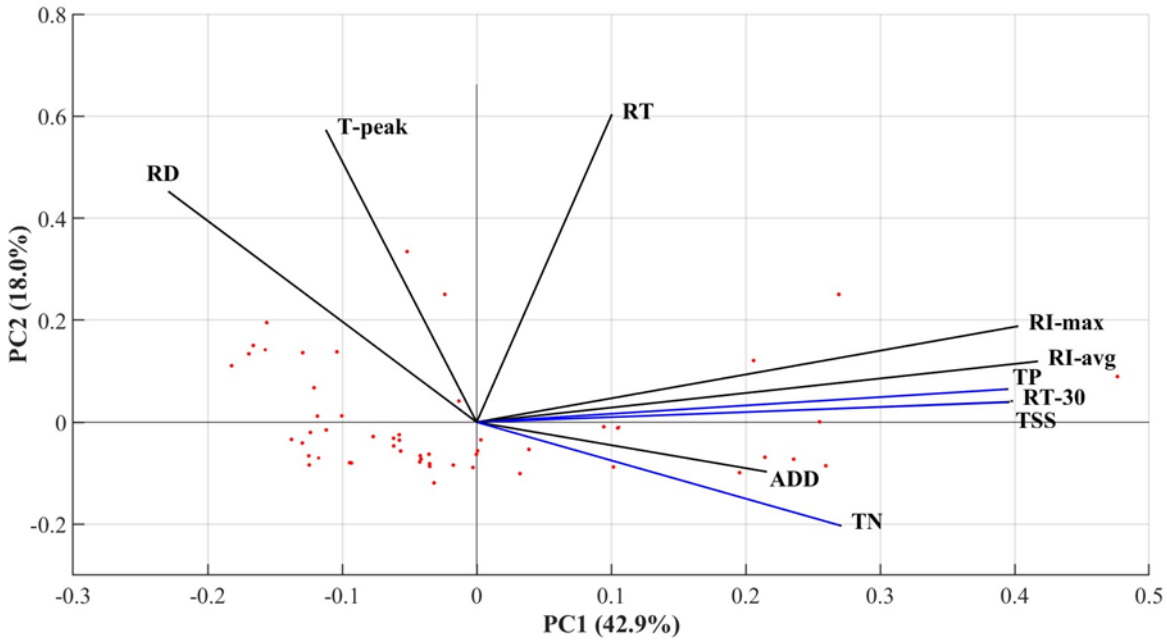


Fig. 2-3. PCA biplots for rainfall indices and stormwater quality indices: the dots represent rainfall-runoff samples.

To verify the results of the PCA analysis, a correlation matrix analysis was used, as shown in Fig. 2-4. The correlation coefficient between RD and T-peak reached 0.51, showing a moderate correlation, indicating that the time to the peak rain intensity would be longer in a longer-duration event. RD was negatively correlated with TSS, TN and TP, consistent with the PCA analysis. The correlation coefficients of the three indices related to rain intensity (RI-max, RI-avg and RT-30) are all over 0.8, showing a strong correlation. Meanwhile, the correlation coefficients between the rain intensity indices and TSS and TP range from 0.48 to 0.61, showing a moderate correlation but range from 0.16 to 0.28 or TN, indicating a weak correlation with TN. There is no apparent correlation between RT and the other rainfall parameters because the total rainfall is determined by the rainfall time and the average rainfall intensity (Liu et al., 2013). There is negligible correlation between RT and the stormwater quality indices and between ADD and the other rainfall indices because ADD is an independent parameter. The correlation coefficients of ADD with TSS, TN and TP ranged from 0.36 to 0.44 indicating low correlations but it is the only rainfall characteristic that shows a significant (i.e., non-negligible) positive correlation with all stormwater quality indices, indicating that ADD is an important parameter affecting stormwater quality (Murphy et al., 2015, Li et al., 2015, Perera et al., 2021). TSS was moderately and highly correlated

with TN and TP indicating that sediment is an important carrier of nitrogen and phosphorus (Zhao et al., 2022).

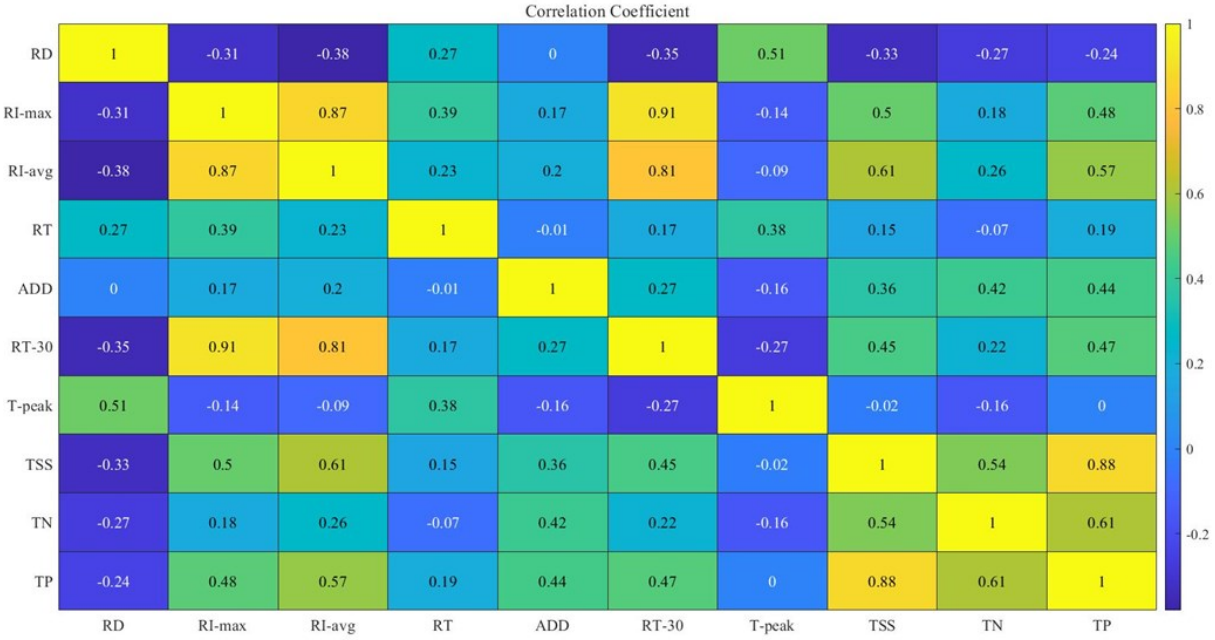


Fig. 2-4. Correlation coefficient matrix heatmap of rainfall and stormwater quality indices.

Combining the results of the PCA and correlation analyses, RT was eliminated, given its poor correlation with the stormwater quality indices, while RI-avg was retained from the three rain intensity parameters because of its stronger correlations. In addition, both ADD, positively correlated with all stormwater quality indices, and RD, negatively correlated with all stormwater quality indices were retained. RI-avg, RD and ADD were finally selected as the indices for the rainfall event type clustering, thus reducing the dimension of the data set.

2.4.2 Clustering results of rainfall indices

Adopting the screening results of the PCA, the smaller data matrix (57x3) consisting of three rainfall indices (i.e., ADD, RI-avg and RD) was used to classify the rainfall events. Fig. C1 shows the variation of SSE with the number of clusters. The SSE decreases with an increase in cluster number, with the magnitude of the decline slowing down after 4 clusters, indicating that the classification effect is not noteworthyly improved by increasing the cluster number beyond 4 clusters. Therefore, the optimal number of clusters is determined to be 4. Fig. C2 in Appendix C

displays how close each point in one cluster is to the points in the neighboring clusters. The silhouette value of each object is a measure of how similar that object is to other objects in the same cluster. Large silhouette values indicate that the clusters are well separated. The average silhouette value is 0.66, which is larger than that of Sönmez's research (2011). They reclassified rainfall regions using the K-means methodology with an average silhouette value of 0.261. Eventually, our data set was divided into four categories, as shown in Fig. 2-5.

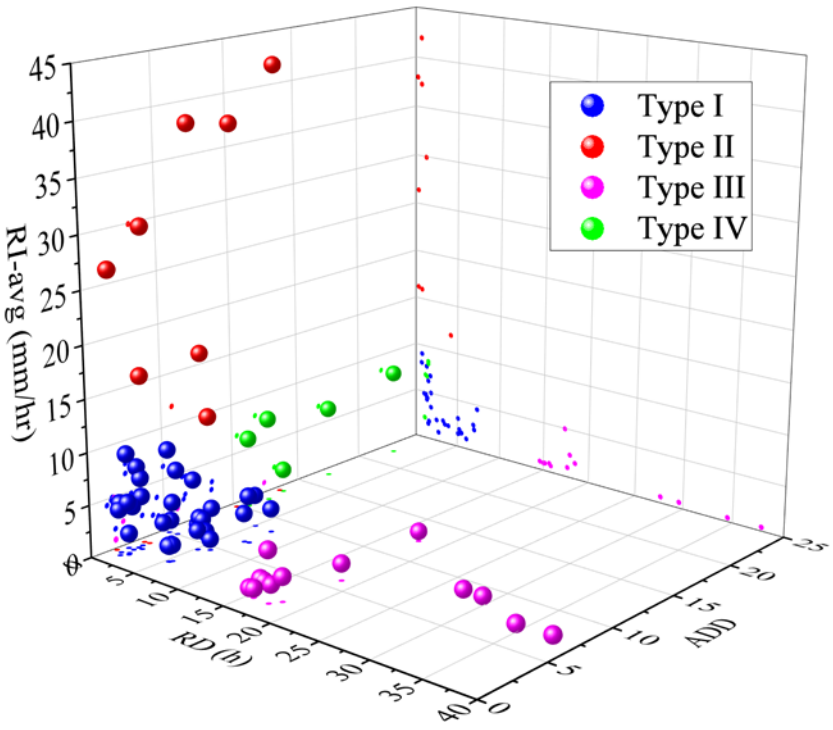


Fig. 2-5. Distributions of four rainfall event types: RI-avg is average rainfall intensity, RD is rainfall duration, ADD is antecedent dry days.

The mean value and range of rainfall and stormwater quality parameters of the four rainfall event types are summarized in Table 2-3. Some typical histograms of four rainfall event types are shown in Fig. D1 in Appendix D. The characteristics of the four types are as follows:

(1) Type I rainfall events have a shorter rainfall duration (3.2 h), a lower average rainfall intensity (3.5 mm/h) and a smaller number of antecedent dry days (4.3 days) with small EMC values for

TSS, TN and TP. Type I rainfall events were small and frequent rainfall events in the four catchments, which amounted to 30 out of 57 events.

(2) Type II rainfall events have a shorter rainfall duration and a small number of antecedent dry days, while they have the highest average rainfall intensity (25.6 mm/h). As shown in Fig. D1b in Appendix D, Type II rainfall events have high rainfall intensity of a rather short duration, consistent with the characteristics of typical convective rainfall events (Thyregod et al., 1999). This kind of rainfall event represents more kinetic energy that is able to mobilize more of the accumulated pollutants, so Type II rainfall events display higher average EMC values of TSS, TN and TP (i.e., 302 mg/L, 1.96 mg/L and 0.39 mg/L, respectively). Type II rainfall events amounted to 9 out of 57 events.

(3) The average rainfall intensity and the number of antecedent dry days of the Type III rainfall events were similar to those of the Type I rainfall events, but the rainfall duration was much longer. Fig. D1c in Appendix D shows that Type III rainfall events are relatively continuous and uniform in intensity, consistent with the typical stratiform rainfall events (Löwe et al., 2016). There are limits to the wash-off capacity of rainfall events, and typically only a fraction of pollutants can be mobilized (Egodawatta et al., 2007, Muthusamy et al., 2018). Type III rainfall events with small rainfall intensity could only wash away pollutants in the initial period and fail to mobilize pollutants in the later stage of long duration, resulting in the smallest average EMC values of TSS, TN and TP (i.e., 73 mg/L, 1.10 mg/L and 0.14 mg/L, respectively). Type III rainfall events amounted to 13 out of 57 events.

(4) Type IV rainfall events have the largest number of antecedent dry days (14.8 days) among the four rainfall event types. More pollutants were able to build up during the longer antecedent dry periods, so Type IV rainfall events display the highest average EMC values with 344 mg/L, 2.57 mg/L and 0.41mg/L for TSS, TN and TP, respectively. Type IV rainfall events amounted to 5 out of 57 events.

Compared with the Type I rainfall events, the latter three rainfall event types reflect the impacts of average rainfall intensity, rainfall duration and antecedent dry days on stormwater quality. More antecedent dry days could provide more available pollutants, while a higher rainfall intensity could provide more kinetic energy, which results in larger pollutant concentrations. It was pointed out

that the higher kinetic energy and longer rainfall durations may produce higher EMCs (Liu et al., 2013). However, there is no corresponding rainfall event type with long duration and high average rainfall intensity as the study areas are located in a semi-arid area with less rainfall. Therefore, the long rainfall duration mainly decreases the pollutant concentration because there is insufficient energy to mobilize the pollutants (Pitt et al., 2005). These four rainfall event types appear to cover most of the rainfall events in the study areas, at least for the 2018-2019 period. It should be noted that the number and type of classified rainfall will vary with the study area when the method is applied to other catchments.

Table 2-3. Summary of four rainfall event types.

Rainfall event types		RD (h)	RI-avg (mm/h)	ADD (days)	TR (mm)	Event number	TSS (mg/L)	TN (mg/L)	TP (mg/L)
Type I	Average	3.2	3.5	4.3	8.6	30	95	1.45	0.15
	Range	0.7-7.3	0.9-9.1	1.2-8.5	2.4-33.2		13-313	0.61-3.74	0.03-0.38
Type II	Average	1.1	25.6	6.4	21.4	9	302	1.96	0.39
	Range	0.3-4.2	11.8-42	1.5-13.1	5.3-49.4		148-593	0.85-3.74	0.05-0.66
Type III	Average	21.3	1.2	5.0	22.8	13	73	1.10	0.14
	Range	14.6-37.8	0.4-5	1.6-12.4	10.6-85.8		11-175	0.41-1.95	0.04-0.29
Type IV	Average	1.3	6.5	14.8	8.8	5	344	2.57	0.41
	Range	1.1-1.5	2.2-8.4	10.4-22	8.0-12.6		190-588	2.09-3.17	0.25-0.58

This study uses MFF curves (Fig. 2-6) to assess the first flush effect. The MFF curves are dimensionless curves of MFFn against cumulative runoff volume. Based on Geiger's definition of the first flush, when the values of MFFn exceed the balanced line (MFFn = 1), the first washout is observed (Geiger 1987, Li et al., 2015). The difference of the MFFn curve from the balanced line is used to assess the magnitude of the first flush.

Fig. 2-6a and b show that the MFFn of the event of 5/2/2019 for RO and RR exceeded the balance line, indicating the occurrence of the first flush phenomenon. Under similar rainfall conditions, the strength of the first flush effect of RR was slightly higher than that of RO under similar rainfall conditions, which could be attributed to the catchment shapes. The hydrograph, which influences the generation of pollution and the first flush effect, could be affected by the shape of the catchment. Square-shaped catchments are more likely to have a stronger first flush than elongated catchments (Mamun et al., 2020). As shown in Fig. 2-1, RO and RR catchment have almost the same catchment area around 14 ha, while RR is square-shaped and RO is elongated. Therefore, under similar rainfall conditions, the RR catchment exhibited a stronger first flush effect than the RO.

The values of MFFn of the event of 6/27/2019 for RO and RR are below the balance line, indicating that the first flush phenomenon did not occur. This event is a Type III event with small rain intensity in the early period and the rain intensity peak occurring later in the long-duration event (Fig. D1c in Appendix D). The later high-intensity rainfall produced higher pollutant concentrations than that of the initial period. A similar trend was observed by Pitt et al. (2004), who found that there likely will not be a noticeable first flush when the highest rainfall intensity occurs later in the event. Type III event (5/17/2019), which occurs in the AB basin, shows first flush effect (Fig. 2-6c). This event has a long duration with small rain intensity and uniform distribution (Fig. D1c in Appendix D), so the runoff from the earlier stage carried away a large load of pollutants. The values of MFFn of the event of 7/16/2019 (Type I) are almost consistent with the balance line, indicating that pollutants are evenly distributed throughout the rainfall event. However, Type I rainfall events (8/22/2019, 8/31/2019 for CR and 6/27/2019 for AB) show first flush effects, except for TN and TP on 8/31/2019. Although these rainfall events belong to Type I, the rainfall processes are different (Fig. D1a in Appendix D). The rain intensity peak of the event of 6/27/2019 (in RO and RR) occurred at the beginning of the event, while the rain intensity peak of the event of 6/27/2019 (in AB) and 8/22/2019, 8/31/2019 (in CR) occurred later in the events. High rainfall intensity at the beginning of the event could contribute to the high wash-off of some pollutants, such as TSS. When the rain intensity peak occurs at the beginning of the event, the first flush effect is exaggerated (Pitt et al., 2004).

Overall, the first flush phenomenon was observed at times in all four catchments. Although the common consensus is that the first flush generally occurs in small catchments and such a

phenomenon is uncertain in large catchments (Mamun et al., 2020), first flush effects also occurred in larger catchments (AB and CR) in this study. This may be related to the fact that discrete water quality samples were collected for only a few rainfall events. This study emphasizes the importance of rainfall processes on the first flush phenomenon. Since discrete water quality samples were only collected for a few Type I and Type III rainfall events, no clear relationship between first flush and rainfall event type was found.

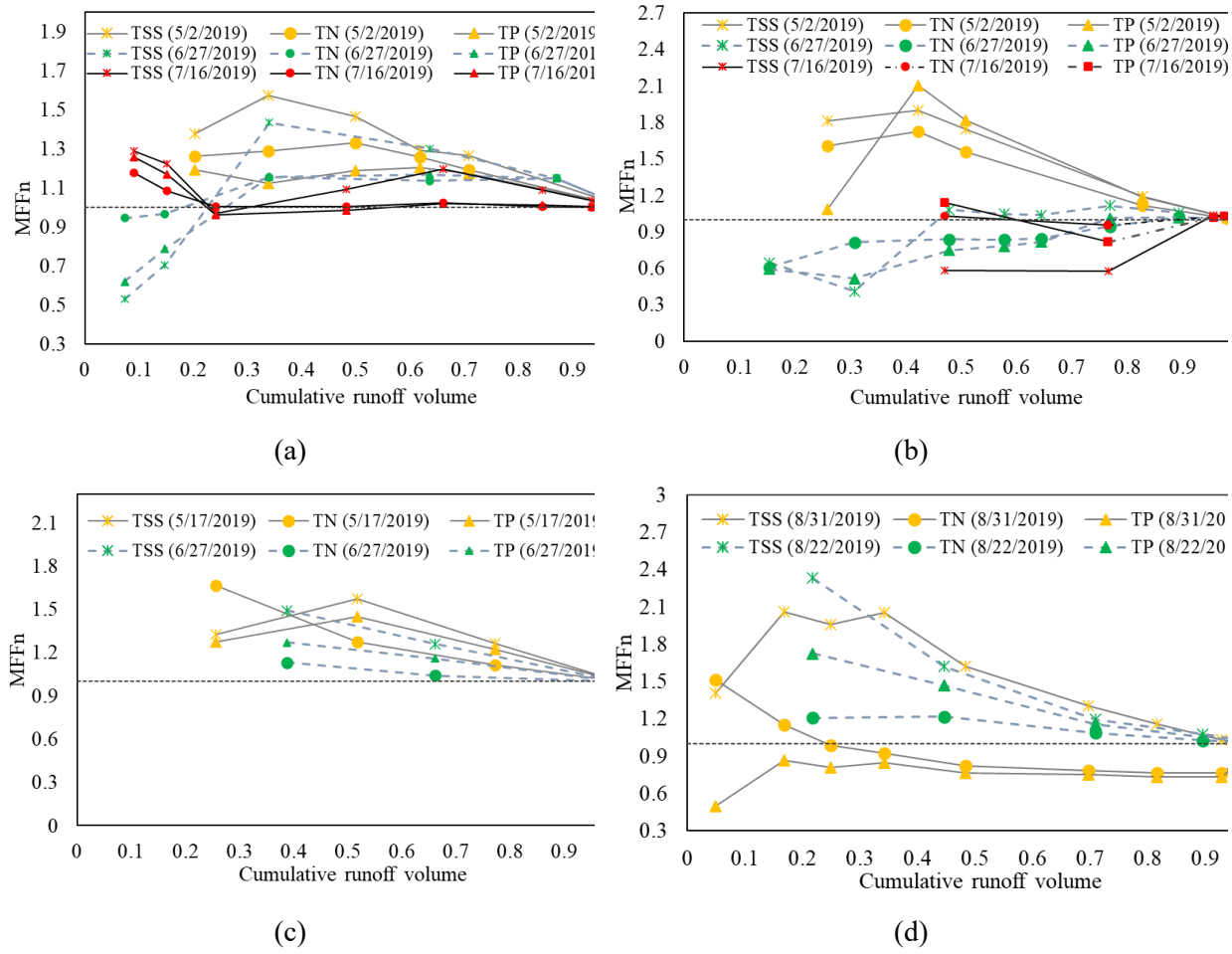


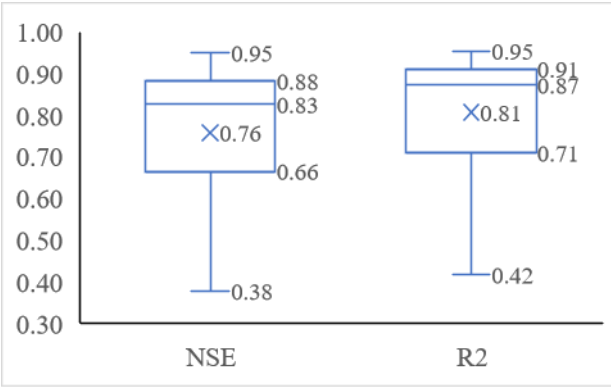
Fig. 2-6. Mass first flush ratio (MFF) curves of (a) RO; (b) RR; (c) AB; (d) CR.

2.4.3 Evaluation of rainfall event type-based model performance

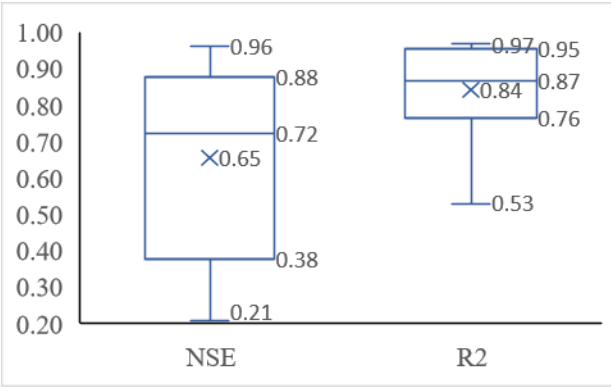
In this study, water quality data was not collected for all recorded rainfall events. In some small events, autosamplers were not triggered or activated and “No Liquid Detected” errors appeared, indicating missed aliquots. Battery power failures also resulted in missed aliquots or termination of the sampler program. Therefore, the number of recorded rainfall events is larger than the number

of composite water quality samples. In order to obtain a reliable hydrological model, recorded rainfall events in 2019 (23 events) were used for the model calibration and recorded rainfall events in 2018 (12 events) were used for the model validation. Fig. 2-7 provides boxplots of different evaluation indices of the hydrological simulation. The NSE and R^2 values were calculated based on the modeled vs. recorded flow rate for the entire hydrograph of individual events. For the calibration period of 2019, the values of NSE ranged from 0.38 to 0.95 (with an average of 0.76). The values of R^2 ranged from 0.42 to 0.95, with an average of 0.81. Therefore, the calibrated hydrological model performed well. The hydrological model also achieved very good performance during the validation period of 2018. The NSE ranged from 0.21 to 0.96 (with an average of 0.65) and the R^2 values ranged from 0.53 to 0.97, with an average of 0.84. Detailed hydrological modeling results in Cranston catchment can be found in Appendix E.

Table 2-4 summarizes the hydrological performance of the different rainfall event types. During the calibration period, the performance of the model on criteria of NSE (> 0.65) and R^2 (> 0.8) for Type I, Type III and Type IV rainfall events was in the very good range. Similarly, the hydrological performance for Type I, Type III and Type IV rainfall events was very good for the validation period. Model performance for Type II rainfall events during the calibration and validation period was not good based on the NSE criterion (< 0.5) but was satisfactory based on the R^2 criterion (> 0.6). To sum up, the calibrated hydrological model could be used as the base for the subsequent simulation of stormwater quality.



(a)



(b)

Fig. 2-7. Boxplots of different evaluation indices of hydrological simulation: (a) NSE and R^2 for the calibration period of 2019; (b) NSE and R^2 for the validation period of 2018 (central bar = median; hinges = 25% and 75%; empty circle = outliers; cross = average).

Table 2-4. Average NSE and R^2 for the different rainfall event types.

Rainfall event type	Calibration period		Validation period	
	NSE	R^2	NSE	R^2
Type I	0.76	0.85	0.62	0.84
Type II	0.38	0.61	0.48	0.77
Type III	0.79	0.82	0.86	0.87
Type IV	0.80	0.83	0.88	0.95

The events in 2019 (12 events) in the Cranston catchment were used to calibrate the CS and RTB water quality model. The events in 2018 (6 events) in the Cranston catchment were used to validate the water quality models. Optimal parameter sets ($C_1 - C_4$) for each land cover type were obtained. A sensitivity analysis was conducted using the sensitivity-based radio tuning calibration (SRTC) tool. The values of C_1 and C_3 for CS and RTB approach, identified as insensitive parameters, are similar, as shown in Table 2-5. C_1 and C_3 represent the maximum possible build-up on the surface and the wash-off coefficient, respectively. These two parameters are mainly affected by land use/land cover types, so they did not show noticeable changes in the two models. While C_2 and C_4 reflect the build-up and wash-off rate, which were affected by rainfall characteristics. Therefore, these two parameters showed non-negligible variability in different models, which were primarily associated with the temporal aspects of the build-up and wash-off equations.

Table 2-5. Calibrated stormwater quality parameters for RTB approach and CS approach.

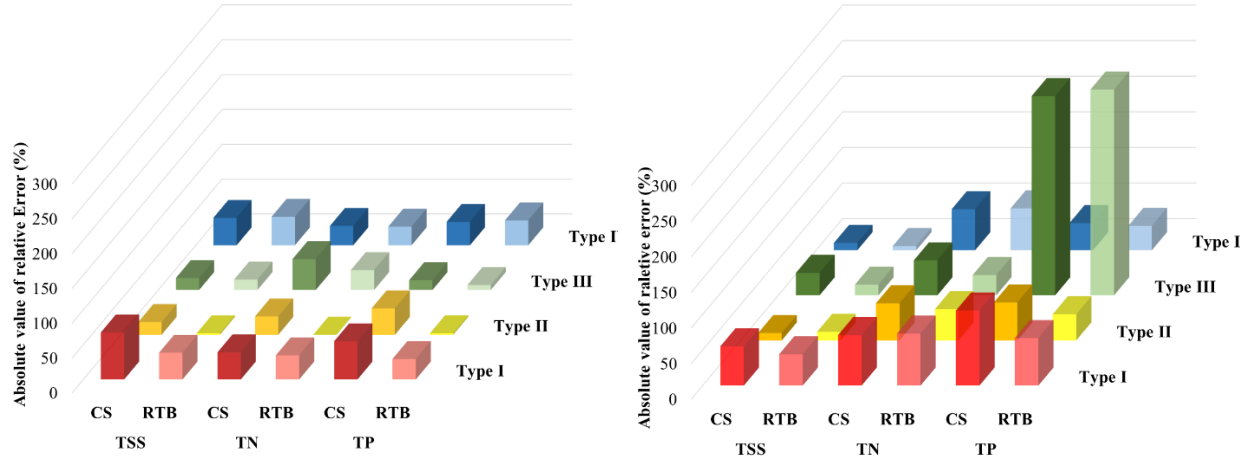
Model		TSS				TN				TP			
		C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
CS	Asphalt/Concrete	369	1.21	0.0007	2.00	2.00	0.70	0.0040	1.29	0.29	1.21	0.0081	1.37
	Gravel	141	1.01	0.0013	1.22	1.00	0.58	0.0020	1.14	0.15	1.01	0.0200	1.55
	Landscaped	100	0.73	0.0013	1.68	3.12	0.42	0.0010	1.00	0.49	0.73	0.0001	1.75
	Roofs	299	0.97	0.0013	1.47	1.50	0.56	0.0050	1.43	0.19	0.97	0.0003	1.67

RTB	Type I	Asphalt/Concrete	373	1.43	0.0007	1.83	2.00	0.78	0.0041	1.22	0.30	1.44	0.0080	1.19
		Gravel	142	1.19	0.0013	1.12	1.00	0.65	0.0020	1.09	0.15	1.20	0.0198	1.33
		Landscaped	101	0.86	0.0013	1.54	3.12	0.47	0.0010	0.95	0.50	0.87	0.0001	1.51
		Roofs	302	1.14	0.0013	1.35	1.50	0.62	0.0051	1.36	0.20	1.16	0.0003	1.44
	Type II	Asphalt/Concrete	369	0.96	0.0007	2.08	2.00	0.96	0.0040	1.38	0.30	0.96	0.0081	1.54
		Gravel	141	0.80	0.0013	1.27	1.00	0.80	0.0020	1.23	0.15	0.80	0.0200	1.73
		Landscaped	100	0.58	0.0013	1.75	3.12	0.58	0.0010	1.08	0.50	0.58	0.0001	1.96
		Roofs	299	0.77	0.0013	1.53	1.50	0.77	0.0050	1.54	0.20	0.77	0.0003	1.87
	Type III	Asphalt/Concrete	369	0.96	0.0007	2.07	2.16	0.64	0.0042	1.20	0.29	1.20	0.0081	1.42
		Gravel	141	0.80	0.0013	1.26	1.08	0.53	0.0021	1.07	0.15	1.00	0.0200	1.60
		Landscaped	100	0.58	0.0013	1.74	3.37	0.39	0.0010	0.93	0.49	0.73	0.0001	1.81
		Roofs	299	0.77	0.0013	1.52	1.62	0.51	0.0052	1.33	0.20	0.97	0.0003	1.72
	Type IV	Asphalt/Concrete	380	0.96	0.0007	2.00	2.00	0.96	0.0040	1.28	0.30	0.96	0.0083	1.38
		Gravel	145	0.80	0.0014	1.22	1.00	0.80	0.0020	1.13	0.15	0.80	0.0206	1.55
		Landscaped	103	0.58	0.0014	1.68	3.12	0.58	0.0010	0.99	0.50	0.58	0.0001	1.76
		Roofs	308	0.77	0.0014	1.47	1.50	0.77	0.0050	1.42	0.20	0.77	0.0003	1.67

Note: C_1 and C_2 are build-up parameters, and C_3 and C_4 are wash-off parameters, seen equations (2) and (3).

Fig. 2-8a compares the two models (RTB vs. CS) concerning the different stormwater quality parameters and rainfall event types during the calibration period. The CS approach performed poorly for the Type I rainfall, with an absolute value of the relative error of TSS and TP over 50%, which was deemed unacceptable. Application of the RTB approach reduced the absolute value of the relative error to about 30%, indicating that the RTB approach was acceptable for the simulation of Type I rainfall events. The RTB approach showed improved performance for the Type II rainfall events, reducing the absolute value of the relative error to less than 5%. The comparison demonstrated that the RTB approach is well suited for simulating stormwater quality for severe rainfall events. Both models had acceptable accuracy of simulating TSS and TP for Type III rainfall events. The RTB approach performed better for TN compared to the CS approach with relative errors of 44.2% and 28.5%, respectively. The RTB approach did not performed better for type IV rainfall events, but both models were acceptably accurate. Overall, compared to the CS approach, the RTB approach reduced the relative error for the three pollutants by 11.4% to 16.4% for calibration period. The RTB approach was better than the CS approach, especially for rainfall

event types I and II while the CS approach sacrificed simulation accuracy for the Type 1 events to achieve good average performance.



(a) (b)

Fig. 2-8. Comparison of the performance of the two models for the different rainfall event types: (a) during the calibration period; (b) during the validation period.

Fig. 2-8b compares the performance of the two models (RTB vs. CS) with respect to the different stormwater quality parameters and rainfall event types during the validation period. For the Type I rainfall events, although the RTB approach improves the model performance to some extent, the absolute value of the relative error of the two models is almost above 50%, which is deemed unacceptable. For the Type II rainfall events, the simulation of TN and TP by the RTB approach was improved compared to the CS approach. Both models failed to represent TP for the Type III rainfall events, but the RTB approach improved the simulation for TSS and TN. Similar to the calibration period, the two models performed similarly for the Type IV rainfall events. Overall, the RTB approach performed better than the CS approach during the validation period, although the simulation of TP was generally unsuccessful. As expected, the performance of the two models decreased during the validation period in comparison to their performance during the calibration period. Detailed water quality modeling results in Cranston catchment can be found in Appendix E.

2.4.4 Transferability of RTB approach's parameters

The transferability of the stormwater quality model parameters resulting from the calibration of the RTB and CS approach in the CR catchment to the other catchments was tested by applying the parameters to the AB, RO and RR catchments. The absolute value of the relative errors between the observed and simulated EMCs was used to assess the parameter transferability. As shown in Table 2-6, the model performance of the AB catchment was the best. Compared to the CR catchment, the performance of the stormwater quality model for the AB catchment decreased, but the average relative error was still less than 50%, which is deemed acceptable. Except for the TSS simulation, the performance of the RTB approach for the AB catchment was better than the CS approach, with the relative error reduced to less than 35%. For the RR catchment, the transferability of the model parameters was not ideal for both the CS and the RTB approach. Even though the relative error of the RTB approach for the TN simulation was less than 50%, the other simulation results are unacceptable (i.e., with relative errors more than 50%). However, the RTB approach still outperformed the CS approach. The RO catchment was the worst with respect to the transferability of the stormwater quality model parameters. The absolute value of the relative error exceeded 100% except for the simulation of TN by the RTB approach, which means that the stormwater quality parameters of the CR catchment cannot be transferred to the RO catchment.

The performance of the models declined when the stormwater quality parameters derived for the CR catchment were transferred to the other catchments. However, the RTB approach remained better than the CS approach. The RTB approach performed well when the parameters were transferred to the adjacent catchment (AB), which has a similar catchment area. However, the performances of the two models were poor when the parameters were transferred to the two catchments (RR and RO) with smaller areas. RR and RO are situated close to each other, whereas they and Cranston (CR) are located at opposite ends of the city. The influence of the catchment characteristics should be carefully evaluated when calibrated stormwater quality model parameters are transferred to other catchments with similar precipitation conditions. For example, conduct a comprehensive assessment of catchment characteristics of the source catchment and the target catchments. This could include land use, soil types, topography, impervious surface percentage, drainage systems, vegetation cover, and current stormwater management practices. Then develop

criteria to assess the similarity of catchments characteristics. This could involve statistical analyses, similarity indices, or machine learning techniques to quantify the degree of similarity.

Table 2-6. Average absolute value of relative error (%) between the observed and simulated EMCs of the two models for the four catchments.

Catchments	TSS		TN		TP	
	CS	RTB	CS	RTB	CS	RTB
CR	35.5	24.1	34.3	22.5	35.0	18.6
AB	44.3	51.6	78.1	35.3	54.6	29.0
RR	61.3	51.9	98.8	41.1	152.5	98.4
RO	98.6	164.5	214.1	49.7	283.9	103.7

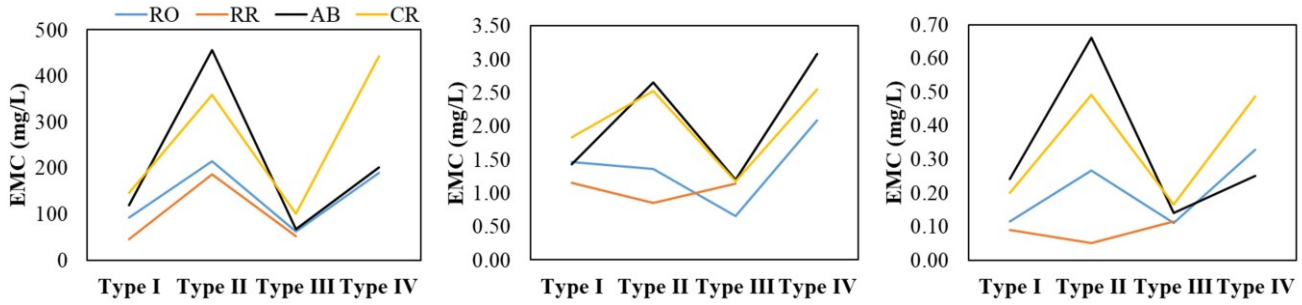
To investigate the possible reasons for the differences in parameter transferability to the different catchments, we compared the average EMC in the four catchments with respect to the different rainfall event types in Fig.9. It should be noted that the EMC of the type IV rainfall events in the RR basin is empty in Fig. 2-9 because no rainfall event type IV rainfall events were recorded in this catchment in 2018 and 2019. As can be seen from Fig. 2-9, for the types I and type III rainfall events, the EMCs of the four catchments were not overly different, with the EMCs of the CR and AB catchments slightly higher than those of the RO and RR catchments. This could be attributed to the fact that these two rainfall event types have neither longer antecedent dry days to build up pollutants nor a stronger rainfall intensity to wash off pollutants, so the EMCs of the four catchments were relatively small.

The four catchments showed differences for the type II and type IV rainfall event type events, which may be related to the catchment characteristics. For the type II rainfall events with higher average rainfall intensity, the EMCs showed a good relationship with the catchment area size and imperviousness. The AB catchment, which has the largest area size and the largest imperviousness, yielded the highest EMCs (457 mg/L of TSS, 2.66 mg/L of TN, and 0.66 mg/L of TP, respectively) and the RR catchment which has the smallest area size and the smallest imperviousness corresponded to the smallest EMCs. This indicates that events with stronger rainfall intensity could produce higher EMCs in catchments with a larger area and imperviousness. However, an impervious surface increases runoff volume and rate, resulting in pollutants being more easily

washed off, therefore larger imperviousness could increase pollutant concentrations in stormwater (Pitt et al., 2001). In addition, pollutants could be temporarily deposited in the larger pipes of a larger drainage system during small rainfall events (e.g., Type I and III rainfall events) (Chow et al., 2013). These deposited pollutants could be transported by large storm events with sufficient energy (e.g., Type II), resulting in a higher pollutant concentration.

For type IV rainfall events with longer antecedent dry days, CR catchment yield the highest EMCs of TSS and TP (442mg/L and 0.49mg/L, respectively) and higher EMC of TN (2.55mg/L). This may be related to the land uses and land covers in the catchments. As shown in Table 2-1, although the land use types in all four catchments are predominantly residential (i.e., more than 50%), the AB and CR catchments are more complex. The AB and CR catchment contains some institutional and commercial land use in addition to the residential, transportation and open space land uses. The commercial site could produce more pollutants due to the high traffic volume in parking area (Pitt et al., 2006, Chow et al., 2013). The commercial and institutional sites have higher impervious areas where runoff could directly enter the drainage system. However, residential sites have smaller impervious surface area. The portion of runoff from the roofs of residential areas may enter the landscaped area and will not drain into the drainage system (Pitt et al., 2006). It was estimated that half of the roofs in residential area of the study sites drain directly to storm pipes and the rest of the runoff from roofs may flow to the pervious area, resulting in smaller pollutant concentrations. In addition, as shown in Table 2-1, there are partial gravel lanes in the CR and AB catchments but not in the RO and RR catchments. Gravel lanes could be an important source of high concentrations of TSS. This could be attributed to the increased wear and tear of vehicles on gravel surfaces (Helmreich et al., 2010) and gravel erosion effects (Nelson et al., 2002). The road surfaces (asphalt/concrete) in commercial and institutional areas in the CR and AB catchments have higher traffic volume than road surfaces in the residential areas in RO and RR catchment. Higher pollutant concentrations were seen from higher-trafficked roads in comparison to the lower trafficked roads (Liu et al., 2016, Charters et al., 2021). Road surfaces contribute to stormwater pollution load mainly through mechanical wear of road surface by vehicle tires and vehicle exhaust (Müller et al. 2020). Therefore, this may help to explain why CR and AB catchments yield higher EMCs compared to RR and RO catchments.

In general, the catchments with similar catchment characteristics (i.e., including the catchment area, imperviousness, and land use) were able to yield similar EMCs for the same rainfall event types. This explains why the model performance is better when the parameters derived for the CR catchment are transferred to the adjacent AB catchment, which has similar catchment characteristics, while the model errors are larger when the parameters are transferred to both the RO and RR catchments. Therefore, the influence of the catchment characteristics should be carefully considered when calibrated stormwater quality model parameters are transferred to other catchments with similar precipitation conditions.



(a) TSS (b) TN (c) TP

Fig. 2-9. Comparison of the average EMCs in the four catchments under the different rainfall event types.

2.4.5 Limitations and future work

The RTB approach shows better performance to the CS model during both calibration and validation periods, with an obvious reduction in relative error for simulating pollutants. It improves simulation performance for all rainfall types, especially for Types I and II. However, the RTB approach's performance diminishes when calibrated parameters are transferred to different catchments, especially those with differences in characteristics.

The proposed framework applied various data mining techniques using MATLAB, but some manual operations were still needed. These operations require the experience and knowledge of the analysts, which may result in a degree of subjectivity. For example, Boxplots were generated in Section 3.2 to screen possible outliers, but the setting of the threshold for rejecting outliers is a subjective exercise. These rejected outliers may have some influence on the subsequent modeling analysis.

The hydrological simulation results strongly influence the performance of the stormwater quality model. Therefore, the accuracy of the hydrological simulations must be improved in applying the proposed method. Some studies have accomplished this by dividing long time series into different sub-periods for independent calibration to improve the performance of the hydrological models (Zhang et al., 2011, Lan et al., 2018). Independent calibration of the hydrological models for the different rainfall event types is also a potential approach. The hydrological simulation for a specific period will be considerably affected by the initial hydrological conditions, leading to unreliable computed responses (James 2010). Although this is a big challenge for the application of the RTB approach for hydrological simulation, the benefits of this approach with respect to hydrological simulations are worth exploring.

The RTB approach is a long-term continuous simulation, but the water quality parameters are only calibrated for events of a single rainfall event type. This approach generates a separate optimal model for each rainfall event type. At present, to obtain a continuous long simulation series, it would be necessary to combine the simulation results of each rainfall event type events into a long series. One possible way is to develop a tool that can automatically modify the model parameters in the modeling process so that the stormwater quality model parameters can be automatically updated as the model moves from one rainfall event type to another one.

2.5 Conclusions

This paper designed a data mining framework to improve the performance of urban stormwater quality models. The influence of rainfall characteristics on stormwater quality was studied by using data mining technology which allowed for the screening of major rainfall characteristics. Rainfall events were clustered using a K-means method and a rainfall event type-based model (RTB) was developed. The main findings and conclusions are as follows:

(1) PCA and correlation analysis showed that the number of antecedent dry days (ADD), average rainfall intensity (RI-avg) and rainfall duration (RD) were the most critical rainfall characteristics affecting the event mean concentrations (EMCs) of TSS, TN and TP. ADD was the only rainfall indices that showed an obvious positive correlation with all stormwater quality indices because ADD affected the build-up process of the pollutants. RI-avg was positively correlated with the EMCs because a high average rainfall intensity could provide more kinetic energy to wash off

pollutants. RD was negatively correlated with the pollutant concentrations, that is, the EMC values were lower for the longer duration rainfall events. The total rainfall depth had a negligible relationship with the stormwater quality indices and other rainfall characteristics, as it was determined jointly by RI-avg and RD.

(2) The K-means method facilitated effective clustering of the rainfall events in the study areas into four types with their own characteristics. Type I was the most common rainfall event type with a shorter rainfall duration, a lower average rainfall intensity and a shorter number of antecedent dry days while Types II, III and IV had a higher average rainfall intensity, a longer rainfall duration and a higher number of antecedent dry days, respectively. The pollutants showed different responses to the four rainfall event types. The rainfall event types with a stronger average rainfall intensity and a higher number of antecedent dry days produced higher pollutant concentrations, while longer duration events produced lower pollutant concentrations. The first flush phenomenon was observed in all four catchments. Rainfall process plays a non-negligible role in the first flush phenomenon.

(3) The RTB approach was calibrated for each rainfall event type independently and optimal sets of stormwater quality parameter were obtained for the different rainfall event types. Overall, the RTB approach was better than the CS approach during both the calibration and the validation period. Compared with the CS approach, the relative error of the RTB approach with respect to the simulation of the three pollutants was reduced by 11.4% to 16.4% in the calibration period and 6.5% to 12.5% in the validation period, respectively. The RTB approach improved the simulation performance for all rainfall event types, especially for Types I and II, while the CS approach sacrificed simulation accuracy for some events in order to achieve good average performance. In the validation period, the RTB approach was still better than the CS approach, indicating that application of a RTB approach could yield superior simulation performance.

(4) The performance of the models diminished when the calibrated stormwater quality parameters were transferred to other catchments, but the RTB approach remained better than the CS approach. The performance of the RTB approach was acceptable (i.e., < 50 % relative error) when the parameters were transferred to an adjacent catchment with similar characteristics. However, the performance of both models was unacceptable when the parameters were transferred to catchments

located at a long distance from the originally calibrated catchment and that had non-negligible differences in catchment characteristics. The influence of the catchment characteristics should be carefully evaluated when calibrated stormwater quality model parameters are transferred to other catchments with similar precipitation conditions.

3. Land cover based simulation of urban stormwater runoff and pollutant loading

3.1 Introduction

In urban areas, the presence of impervious surfaces greatly increases volumes and peaks of surface runoff (Ebrahimian et al., 2016). Various pollutants contained in urban runoff pose serious threats to public health and environment (Li et al., 2016). Urbanized areas are typically comprised of a variety of land covers such as roads, buildings, parking and green spaces with the urban hydrological cycle being considerably impacted by different land covers (Niemi et al., 2019).

Urban stormwater models such as the Storm Water Management Model (SWMM) are important tools for managing stormwater quantity and quality (Gironás et al., 2010). The complexity of land covers poses a number of challenges in urban stormwater modeling. Semi-distributed stormwater models require a large number of parameter sets to describe a catchment (Petrucci et al., 2014). Accurate description of the catchment is required to achieve accurate and successful applications of stormwater models (Wu et al., 2017). The standard methods of catchment delineation are based on Digital Elevation Models (DEMs) and Geographic Information System (GIS), such as the built-in Watershed Delineation Tool (WDT) in PCSWMM. Typically, subcatchments are considered to have the same hydrological characteristics (i.e., they share the same set of model parameters) and are called Hydrological Response Units (Sun et al., 2014). This traditional method is acceptable for undeveloped, natural areas (Gorgoglione et al., 2016). Subcatchments covered by different land types are nonhomogeneous and a single set of parameters cannot describe their spatial heterogeneity. A number of studies have shown that spatial heterogeneity can affect the accuracy of runoff simulation (e.g., Elliott et al., 2009; Krebs et al., 2014; Leandro et al., 2016).

Urban stormwater models have a number of calibration parameters that are difficult to estimate a priori, so they are usually calibrated using measured data (Bárdossy et al., 2007). Collection of data is often not only time-consuming and expensive, but also the quality of the data is not guaranteed for a variety of reasons (McKenzie et al., 2013). For example, the main sources of error in rainfall measurement are related to both rainfall catching and counting errors. Battery, logger and computer clock failures are also sources of uncertainties in rainfall measurements. The lack

of reliable measurement data makes it difficult to apply the model directly to ungauged catchments (Patil et al., 2014).

A number of studies have researched spatial transferability of model parameters in neighboring catchments. The term regionalization referred in rainfall-runoff modeling context to the transfer of parameters from neighboring gauged catchments to an ungauged catchment (Oudin et al., 2008). Model parameters can be transferred from gauged catchments to similar ungauged catchments. Similarity between gauged catchments and ungauged catchments is typically identified using an approach that is either based on spatial proximity or physical similarity (Patil et al., 2015). These approaches usually use physical descriptors such as catchment area, mean slope, median altitude, river network density and fraction of forest cover to evaluate watershed similarity without considering the composition of land cover types (Oudin et al., 2008). These approaches regionalize the parameters of the entire catchment without considering the complex composition of the land cover types, so it is not suitable for application in the urban catchments. If we determine parameter sets based on land cover types, it would make sense to apply these parameters to ungauged catchment because the same land cover type in different catchments are comparable. However, there are few studies on parameterization of urban stormwater models based on land cover types and on parameter transferability between urban catchments. Krebs et al. (2014) reported that that the parameter values calibrated for the high-resolution model are suitable for regionalization to larger, ungauged urban areas. Sun et al. (2014) stated that the posterior parameters from the high-resolution model could provide higher confidence in parameter transferability for other ungauged urban areas. Rio et al. (2020) found that discretized catchment into uniform surface types enhances the hydrological and water quality outputs compared to the representation as pervious and impervious surfaces.

To control runoff and contaminants near these source areas, management concepts such as Low Impact Development (LID) have been developed (Elliott et al., 2007). Urban stormwater management increasingly requires water quality models that account for urban land cover types to assess the impact of land use planning scenarios and effectiveness of LID (Rio et al., 2020). In many applications of PCSWMM, different water quality functions and parameters have been applied to different land cover types (Bonhomme et al., 2017). However, PCSWMM cannot report runoff and pollutant loads on each land cover type because they did not discretize catchment by

land cover types. Therefore, traditional catchment delineation methods in PCSWMM cannot be used to assess the different runoff and pollutant sources and pathways.

Recent research focused on improving PCSWMM model's performance by discretizing a catchment based on land cover types (Tu et al., 2018; Rio et al., 2020; Palli et al., 2020). For example, an 11.4 ha catchment was discretized into 600 subcatchments representing vegetation, roofs, asphalt, sand/gravel, stone/tile pavers or open rock (Tuomela et al., 2019). They showed that the performance of the model can be improved by increasing spatial variability and dividing the catchment into several areas with homogeneous land cover and using separate parameter sets. In another study, an automated tool was developed to delineate adaptive subcatchments with homogeneous land cover (Niemi et al., 2019). The results showed that a model with adaptive subcatchments could reproduced the observed discharge with high model-performance. However, Niazi et al. (2017) concluded that the effects of spatial resolution can vary considerably on model predictions. Ghosh et al. (2012) reported that the total outflow volume was relatively insensitive to spatial resolution in their study catchment.

A comprehensive comparison between traditional delineation methods and land-cover based methods has not been reported. Based on the above considerations, this paper developed a land-cover based PCSWMM model and compared it with a traditional catchment delineation method to study the impacts of land cover heterogeneity on simulating urban stormwater runoff and pollutant loading. We selected two mature, urban residential catchments as the study areas and applied the two models. The specific objectives were to (1) evaluate the performance of the two methods in simulating hydrographs and TSS, TN and TP loading; (2) parameterize different land cover types for application to an ungauged area and test the transferability of model parameters; (3) use land cover representation to assess the individual contributions of different land cover types to runoff and water quality.

3.2 Materials and methods

3.2.1 Description of the study areas

The study areas include two residential catchments located in the City of Calgary, Canada: Rocky Ridge (RR) and Royal Oak (RO). The climate of Calgary is continental, with cold winters and

mild warm summers. The two neighboring catchments are mature with no considerable new development for the past 5 years.

The RR catchment has a 14.0 ha drainage area that is directed to the RR wetland. The altitude of RR catchment ranges from 1263m to 1270m and the average slope is about 9.5%. The land cover types in RR had the following areal distribution: 23.6% roofs, 5.7% parking, 16.4% roads, 51.4% greenspace and 2.9% trails. The RO catchment has a 14.5 ha drainage area that is directed to the RO wetland. The altitude of RR catchment ranges from 1265m to 1283m and the average slope is about 9.9%. The land cover types in RR had the following areal distribution: 20.2% roofs, 5.5% parking, 25.2% roads, 46.6% greenspace and 2.5% trails. The RR and RO catchments are located only 300 m apart. Therefore, they are similar in terms of their physical characteristics and climate.

3.2.2 Field measurements

In each catchment, one weather station, flow meters and autosamplers were installed to collect data. Field monitoring was conducted during the ice-free seasons (May - October) of 2018 and 2019. The weather station consisted of a HOBO RX300 Remote Monitoring Station Data Logger equipped with a HOBO RG3-M Data Logger (Onset Computer Corp., USA) for collecting meteorological data. Rainfall was measured with a tipping bucket, which was triggered at a minimum of 0.2 mm cumulative rainfall and data was recorded at a 5-min interval.

Flow meters and autosamplers were installed in the inlets of the wetlands. An ISCO 750 Area Velocity Flow Module (Teledyne Isco, USA) and an ISCO 6712 autosampler (Teledyne Isco, USA) were installed at the RR inlet. An ISCO 750 Area Velocity Flow Module, an ISCO 2150 Area Velocity Flow Module (Teledyne Isco, USA) and an ISCO 6712 autosampler were installed at the RO inlet. The collected water samples were analyzed for TSS, TN, and TP. The flow meters monitored flow rates continuously at 5-min intervals. ISCO 6712 autosamplers collected water samples with a maximum collection capability of 24×1 L bottles. Flow paced program was used to collect samples that could be representative of the flow for the day. ISCO 6712 with 750 AV sensor and ISCO 2150 flow module recorded water level and flow velocity on a continuous basis for the purpose of triggering and pacing ISCO 6712 sampler programs. Given the limitations of ISCO 6712 unit's maximum collection capability, and the desires to maximize water quality samples during a runoff event, three pacing configurations were used: low flow pacing, high flow pacing and extreme event pacing. Therefore, weather conditions were monitored to forecast

precipitation, then the water level at AV sensor location was adjusted to trigger or activate ISCO 6712 unit.

The volume of collected samples per rainfall event ranged from 1 L to 24 L, with an average of 11 L. A total of 25 composite water samples was successfully collected for around 60-70% of the rainfall events over the two years. The study area is in a semi-arid cold temperate climate region. In some small events, autosamplers were not triggered or activated and "No Liquid Detected" errors appeared indicating missed aliquots. Battery power failures will also result in missed aliquots or termination of the sampler program. These led to the incomplete samples for some rain events. Nevertheless, larger rain events accounted for around 90% of the total rainfall of all events. Water samples were successfully collected for most larger rainfall events in this study. In this study, the aliquots sampled during each storm event were combined to generate composite samples, resulting in event mean concentrations (EMC).

The monitoring campaign recorded rainfall, flow, TSS, TN and TP in 2018 and 2019. Rainfall/runoff data related to all events, including maximum rainfall intensity (mm/hr), total rainfall (mm), peak runoff (m^3/s), and total flow (m^3), are summarized in Table 3-1. The rainfall and flow data have been checked to ensure the data quality. EMCs of TSS, TN and TP are also listed in Table 3-1.

Table 3-1. Summary of rainfall-runoff events at the RR and RO catchment in 2018 and 2019 including EMCs for TSS, TN and TP.

Catchment	Event	Maximum Rainfall (mm/hr)	Total Rainfall (mm)	Peak Flow (m^3/s)	Total flow (m^3)	TSS (mg/L)	TN (mg/L)	TP (mg/L)
RR	6/15/2018	19.2	4	0.054	86			
	6/16/2018	7.2	5	0.048	141			
	6/23/2018	21.6	33.2	0.118	1188	144	1.21	0.18
	6/28/2018	4.8	7	0.026	169	13	0.96	0.03
	7/04/2018	7.2	2.1	0.013	26			
	7/10/2018	4.8	2.2	0.025	49	40	1.57	0.12
	7/18/2018	19.2	2.6	0.069	54			
	7/24/2018	24	12	0.08	249	38	0.82	0.11

Continued Table 3-2. Summary of rainfall-runoff events at the RR and RO catchment in 2018 and 2019 including EMCs for TSS, TN and TP.

Catchment	Event	Maximum Rainfall (mm/hr)	Total Rainfall (mm)	Peak Flow (m^3/s)	Total flow (m^3)	TSS (mg/L)	TN (mg/L)	TP (mg/L)
RR	7/25/2018	12	1.2	0.017	16			
	8/02/2018	7.2	2.2	0.024	47			
	8/03/2018	9.6	6.4	0.044	138			
	8/12/2018	50.4	7	0.148	128			
	8/24/2018	14.4	9.8	0.053	154	50	3.65	0.11
	8/26/2018	7.2	8.2	0.026	145	28	1.07	0.06
	9/03/2018	7.2	3	0.023	46			
	9/10/2018	4.8	2	0.028	47	22	0.95	0.06
	9/20/2018	4.8	6.4	0.021	119	15	1.43	0.6
	9/26/2018	7.2	10.8	0.016	188	17	0.66	0.04
	5/07/2019	12	4.6	0.071	229			
	6/19/2019	52.8	11.8	0.37	461	184	1.52	0.25
	6/27/2019	55.2	42	0.651	2853	142	1.83	0.25
	7/09/2019	14.4	4	0.091	151	41	1.27	0.06
	7/16/2019	26.4	10.4	0.096	308			
	7/18/2019	33.6	5.6	0.192	199			
	7/20/2019	12	5.8	0.071	257			
	8/6/2019	40.8	10.2	0.215	213	37	0.64	0.07
	8/22/2019	12	5.6	0.066	100			
	9/01/2019	72	9.6	0.576	328	232	1.44	0.13
RO	6/23/2018	19.2	34.2	0.105	888			
	6/28/2018	4.8	8.2	0.026	141	71.7	1.55	0.42
	7/18/2018	16.8	3.2	0.044	48			
	7/23/2018	28.8	15.2	0.243	412	116	2.08	0.25
	8/03/2018	12	7	0.045	186			
	8/24/2018	14.4	9.8	0.05	81	96	3.13	0.1
	8/26/2018	7.2	8.2	0.022	44	22	0.76	0.07
	9/03/2018	7.2	3	0.019	17			

Continued Table 3-3. Summary of rainfall-runoff events at the RR and RO catchment in 2018 and 2019 including EMCs for TSS, TN and TP.

Catchment	Event	Maximum Rainfall (mm/hr)	Total Rainfall (mm)	Peak Flow (m^3/s)	Total flow (m^3)	TSS (mg/L)	TN (mg/L)	TP (mg/L)
RO	9/26/2018	7.2	11.6	0.009	48	25	0.5	0.05
	4/28/2019	2.4	1.6	0.013	24	193	1.19	0.23
	6/19/2019	72	12.8	0.215	293	148	1.1	0.2
	7/06/2019	12	4.4	0.101	253			
	7/09/2019	14.4	4.2	0.07	110			
	7/16/2019	43.2	12	0.15	301	77	1	0.06
	7/18/2019	24	5.6	0.077	127			
	8/06/2019	4.8	6	0.022	121	190	2.09	0.33
	8/16/2019	45.6	11	0.23	241	137	0.6	0.13
	8/22/2019	12	7	0.058	176	101	1.03	0.13

Note: Blanks mean that no water samples were collected.

3.2.3 Modeling methods

The built-in Watershed Delineation Tool (WDT) in PCSWMM delineates watersheds based on a digital elevation model (DEM). The WDT automatically discretizes the watershed and generates subcatchments and flow directions with assigned slopes and flow lengths. Usually, after this step, modelers check the results and adjust it manually taking into consideration the topography and drainage infrastructure. The subcatchments generated by this method include pervious areas and impervious areas. The flow routing elements consist of an outlet and an internal subarea routing. Users can specify the outlet and internal flow direction: pervious subarea to impervious subarea or vice versa. RR and RO were divided by the WDT into 427 and 449 subcatchments, respectively (Fig. 3-1a and 3-1c).

In the LCB approach, the catchment is delineated based on land cover such that each subcatchment has only one land cover type. First, the built-in tile map service in PCSWMM is used to access high-resolution geo-referenced satellite images that are used to identify the land types contained in the catchment. The high-resolution satellite imagery can provide spatial resolution down to

5m×5m which is needed to accurately delineate these small urban catchments. Typically, the land covers of a catchment are classified into three categories: road, roof, and greenfield (Li et al., 2016). In this paper, the catchment was divided into five land cover types manually based on the built-in map in PCSWMM: roofs, roads, parking, trails, and greenspace. In this study, manually delineation of urban surface is an effective method to identify land covers because the study areas are small urban catchments. For large urban area, advanced automatic catchment-discretization approach with homogeneous land cover is proposed (Niemi et al., 2019). Second, the catchment was subdivided into subcatchments with a single land cover type. This method divided RR into 477 subcatchments: 23.6% roofs, 5.7% parking, 16.4% roads, 51.4% greenspace and 2.9% trails; and RO into 482 subcatchment: 20.2% roofs, 5.5% parking, 25.2% roads, 46.6% greenspace and 2.5% trails (Fig. 3-1b and 3-1d). As each subcatchment is homogeneous, it is assigned as either pervious or impervious: roofs, roads and parking are 100% impervious, and greenspace and trails have 0% imperviousness.

In the LCB approach, impervious cover is divided into direct impervious cover and indirect impervious cover. The runoff from direct impervious covers such as roads is collected by the drainage system directly (Direct impervious surfaces are defined as areas directly connected to urban drainage system.). If an impervious cover i.e., roofs is connected to pervious cover and runoff from the impervious cover flows to the pervious cover, the impervious cover is considered indirect impervious cover because runoff from such impervious cover may infiltrate on the pervious cover. However, how to accurately reflect building/rooftop runoff connectivity is a common challenge in the modelling work in general. A portion of runoff from a certain percentage of roofs may flow to pervious area and the rest may drain to the drainage system. Determination of the exact portion is often difficult due to the scarcity of survey data. Therefore, it is assumed that runoff from rooftops drains to pervious space in this work. Note this can be different in other catchments. RR catchment was used to calibrate and validate PCSWMM model and RO catchment was used to evaluate the transferability of model parameters.

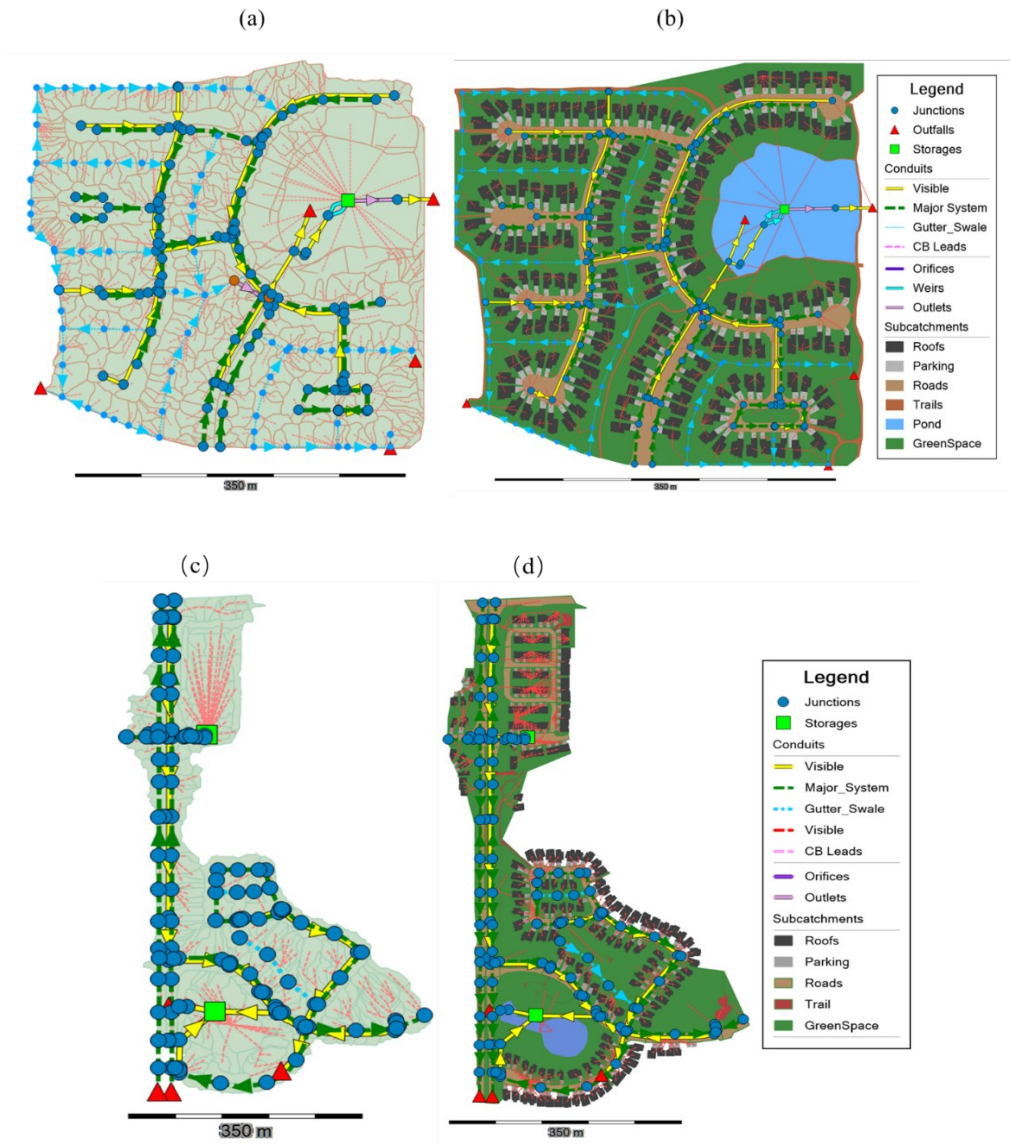


Fig. 3-1. WDT and LCB approach configurations for RR and RO catchment: (a-b) WDT approach and LCB approach for RR; (c-d) WDT approach and LCB approach for RO.

PCSWMM is a GIS-based dynamic hydrologic-hydraulic simulation model used either single-event or long-term continuous simulation of runoff quantity and quality primarily in urban areas. In this study, the model was setup with subcatchments, conduits, manhole, storage junctions, and a dual-drainage system (consisting of sewers as minor system and roads as major system). The Green-Ampt method was applied to represent infiltration of rainfall in pervious subcatchments. Flow routing was modeled using dynamic wave routing. The model was used to continuously simulate the rainfall and runoff from May to October 2018 and 2019. The buildup and washoff of

pollutants from the different land cover types were simulated using the exponential buildup and washoff functions. The land cover types for the LCB approach include roofs, roads, parking, greenspace and trails. The land cover types for the WDT approach include roofs, paved areas and greenspace.

The exponential buildup function is given by,

$$B = C_1(1 - e^{-C_2 t}) \tag{3-1}$$

where B = buildup (mass per unit area), C_1 = maximum buildup possible (mass per unit area or curb length), C_2 = buildup rate constant (1/day) and t = buildup time interval (day). This function is applicable when the buildup follows an exponential growth that asymptotically approaches a maximum limit (James et al., 2003).

The exponential washoff function is given by,

$$W = C_3 q^{C_4} B \tag{3-2}$$

where W = washoff rate (mass/hr), C_3 = washoff coefficient, C_4 = washoff exponent and q = runoff rate per unit area (mm/hr) (James et al., 2003).

The Sensitivity-based Radio Tuning Calibration (SRTC) tool, a built-in tool in PCSWMM, was used to quantify the sensitivity of the calibration parameters. The user inputs a percentage of estimated uncertainty (e.g., +/- 30%) for each parameter that is considered for calibration based on the accuracy and source of the initial input. This gives a low-end value and a high-end value for each parameter. The SRTC tool executes a SWMM5 run by using both the high and low values of the designated uncertainty range generating sensitivity gradients for each parameter defined for calibration. The sensitivity analysis can be performed with the SRTC tool before calibration. Events in 2018 of RR were used to calibrate models, and events in 2019 of RR were used to validate models. The Nash-Sutcliffe efficiency (NSE) and coefficient of determination (R^2) were chosen to assess the agreement between simulated and observed flows. An NSE greater than 0.5 and $R^2 \geqslant 0.7$ indicate acceptable model performance for PCSWMM simulations (Zhao et al., 2009; Donigian et al., 2002). Relative error between simulated and observed pollutant concentrations was used to evaluate model performance for water quality. A criterion of ±50% for the simulated pollutant concentrations was assumed to be an acceptable deviation from the measured value with a targeted relative error lower than ±25% (Tuomela et al., 2019).

PCSWMM can automatically produce subcatchment summaries including contributions from upstream subcatchments. Each subcatchment in the LCB approach is homogeneous with a single land cover type. The runoff is either directly routed into the drainage system or into a downstream catchment. The runoff routed into a downstream catchment area is called runon. Therefore, the total runoff from a subcatchment includes the runoff generated by rainfall over the subcatchment and any runon routed from upstream subcatchments. First, in order to accurately assess the contribution of different land covers, it is necessary to classify the subcatchments into five categories according to the land cover types. Second, the local runoff is then calculated based on the runoff coefficients reported by PCSWMM for each subcatchment. The runon from upstream subcatchments will be reduced by losses due to evaporation, infiltration and depression storage. Therefore, in addition to local losses and these upstream losses must also be calculated. All constituents can be calculated by the following equations:

$$R_L = P \cdot \alpha \tag{3-3}$$

$$I_L = P - E \cdot \frac{R_L}{R_L + R_{on}} - R_L \tag{3-4}$$

$$I_{on} = R_{on} + R_L - R_{total} = I - I_L \tag{3-5}$$

where R_L is the local runoff, P is the total precipitation, α is the runoff coefficient, I_L is infiltration of local precipitation, E is evaporation, R_{on} is runon, I_{on} is infiltration of runon, R_{total} is the net runoff from the subcatchment, I is total infiltration. Note that α , E , I, R_{on}, R_{total} are reported by PCSWMM. All variables except the runoff coefficient have been converted into m^3.

3.3 Results and discussion

3.3.1 Model parameter sensitivity and calibration

A parameter sensitivity analysis was conducted for the two models and the ranked lists of the most sensitive parameters are presented in Table 3-2 and Table 3-3. It was shown that five of the eight most sensitive parameters were common to both models. It is noteworthy that Percent Routed, defined as the percent of runoff routed between subareas, was the most sensitive parameter in the WDT approach but was the least sensitive in the LCB approach. Knighton (2016) applied PCSWMM in a wetland without consideration of land cover types. Their sensitivity analysis results also showed that percent routed was the most sensitive parameter particularly for runoff

volume and peak flow. In the LCB approach, each subcatchment contained just one type of land cover so that there was no internal runoff exchange and as a result the LCB approach was completely insensitive to Percent Routed. Imperviousness was sensitive in the WDT approach, but for the LCB approach there was no need to calibrate it because it had been specified for each type of land cover. Therefore, Percent Routed and imperviousness did not need to be calibrated in the LCB approach. DStore Imperv, defined as the depth of depression storage on the impervious area of the subcatchment, was the first and second most sensitive parameter for the LCB and the WDT approach, respectively. This parameter should vary with land cover type, for example it should be relatively small for inclined roofs and larger for roads with numerous cracks and seams on the surface. Therefore, Dstore Imperv was sensitive in the LCB approach.

Table 3-4. Ranked list and calibrated value of hydrologic-hydraulic parameters based on sensitivity (from most sensitive to least) for the WDT approach.

Order	Parameter Name	Calibrated value
1	Percent Routed (%)	25
2	Dstore-Imperv (mm)	1.1
3	Conductivity (mm/hr)	2
4	% Imperv	45
5	Width (m)	8.4
6	% Zero-Imperv	43.8
7	Dstore-Perv (mm)	1.8
8	N-Imperv	0.015

Table 3-5. Ranked list and calibrated value of hydrologic-hydraulic parameters based on sensitivity (from most sensitive to least) for the LCB approach.

Order	Parameter Name	Calibrated value				
		Roofs	Roads	Parking	Greenspace	Trails
1	Dstore-Imperv (mm)	0.4	2.2	1.5	/	/
2	Width (m)	13.3	29.9	6.6	32	17.2
3	N-Imperv	0.012	0.02	0.062	/	/
4	% Zero-Imperv	66	36.6	58.6	/	/
5	% Slope	8.2	3.4	6	9.4	10.7
6	Suction head (mm)	/	/	/	405	363
7	Initial deficit (mm/hr)	/	/	/	2	2
8	Dstore-Perv (mm)	/	/	/	0.45	0.45

Note: Percent Routed is the percent of runoff routed between subareas; Dstore-Imperv is the depth of depression storage on the impervious portion of the subcatchment; Width is Characteristic width of the overland flow path for sheet flow runoff; N-Imperv is Manning's n for overland flow over the impervious portion of the subcatchment; % Imperv is the percent of land area which is impervious; % Zero-Imperv is the percent of the impervious area with no depression storage; % Slope is the average percent slope of the subcatchment; Dstore-Perv is the depth of depression storage on the pervious portion of the subcatchment; Conductivity is the soil's saturated hydraulic conductivity; Suction head is the average value of soil capillary suction along the wetting front; Initial deficit is the fraction of soil volume that is initially dry.

As shown in Table 3-2 and Table 3-3 that Dstore-Imperv for the WDT approach was 1.1 mm which falls between the values for roofs and roads in the LCB approach. Dstore-Imperv for the WDT approach demonstrated the average depth of depression storage on the entire impervious portion of the catchment. The parameter values of soil properties for the WDT approach and the LCB approach were close. For example, the conductivity for both models was 2 mm/hr.

Urban runoff pollution is heavily influenced by local site conditions. Therefore, the coefficients C_1, C_2, C_3 and C_4 in Eq. 1 and 2, used to estimate buildup and washoff, are expected to vary with land cover type. The LCB approach included five types of land use: roofs, roads, parking, greenspace and trails. The land cover types in the WDT approach were grouped into three general categories: paved areas, roofs and greenspace. SRTC Tool was used to quantify the sensitivity of

the buildup-washoff parameters. Parameter sensitivity analysis showed that maximum buildup possible (C_1) and washoff exponent (C_4) are the most sensitive coefficients. The value of the four coefficients were calibrated for the various land cover types in each model. The most sensitive parameters were calibrated first, followed by the parameters with lesser sensitivity. Parameter calibration results are shown in Table 3-4.

Table 3-6. Calibrated water quality parameters for WDT approach and LCB approach.

Model	Land Cover	TSS				TN				TP			
		C_1	C_2	C_3	C_4	C_1	C_2	C_3	C_4	C_1	C_2	C_3	C_4
WDT	Roofs	299	0.77	0.0013	1.45	0.69	1.49	0.0001	1.32	0.10	0.8	0.0003	1.94
	Greenspace	100	0.58	0.0013	1.66	3.12	1.49	0.0006	1.22	0.22	1.1	0.0001	2.04
	Paved areas	369	0.96	0.0007	1.97	1.39	2.09	0.0078	1.73	1.72	4.2	0.0081	1.60
LCB	Roofs	154	0.54	0.0012	1.36	6.30	0.29	0.0016	0.86	0.57	0.6	0.0011	1.02
	Roads	193	0.72	0.0016	1.82	8.40	0.39	0.0022	1.14	0.71	0.8	0.0015	1.36
	Parking	150	0.10	0.0100	1.00	2.00	0.05	0.0010	1.00	0.30	0.1	0.0010	1.00
	Greenspace	39	0.18	0.0004	1.36	2.10	0.05	0.0006	0.86	0.28	0.2	0.0007	1.02
	Trails	100	0.10	0.0010	1.60	0.50	0.05	0.0010	1.60	0.10	0.1	0.0010	1.60

3.3.2 Comparison of two models' performance

Fig. 3-2 showed comparisons of NSE and R^2 values of all storm events simulated using both models during the calibration and validation periods, 2018 and 2019, respectively, in RR. The NSE and R^2 values were calculated for individual events. Each data point in Fig. 3-2 represented a NSE or R^2 value of a single event calculated based on flow rate. As shown in Fig. 3-2a and 3-2b, in the calibration period, the NSE and R^2 values for the LCB simulations were greater than or equal to those for the WDT approach in 83% (15 out 18 events) and 72% (13 out of 18 events) of the simulations, respectively. Based on NSE, the LCB and the WDT approach simulated 16 and 15 events out of 18 with acceptable accuracy (i.e., NSE $\geq$ 0.5), respectively. Using the value of $R^2 \geq$ 0.7 as the standard, the LCB approach simulated all events with acceptable accuracy while for the WDT approach 15 of 18 events were acceptably accurate.

The comparisons in Fig. 3-2c and 3-2d showed that in validation period of 2019, the LCB approach was better than the WDT approach simulating all ten events. The LCB simulations were sufficiently accurate 100% of the time using either the NSE or R^2 standard while the WDT

approach was only accurate for 30% to 40% of the events. Therefore, the LCB approach was more accurate simulating flows during the validation period. Detailed hydrological modeling results of LCB and WDT approaches can be found in Appendix F.

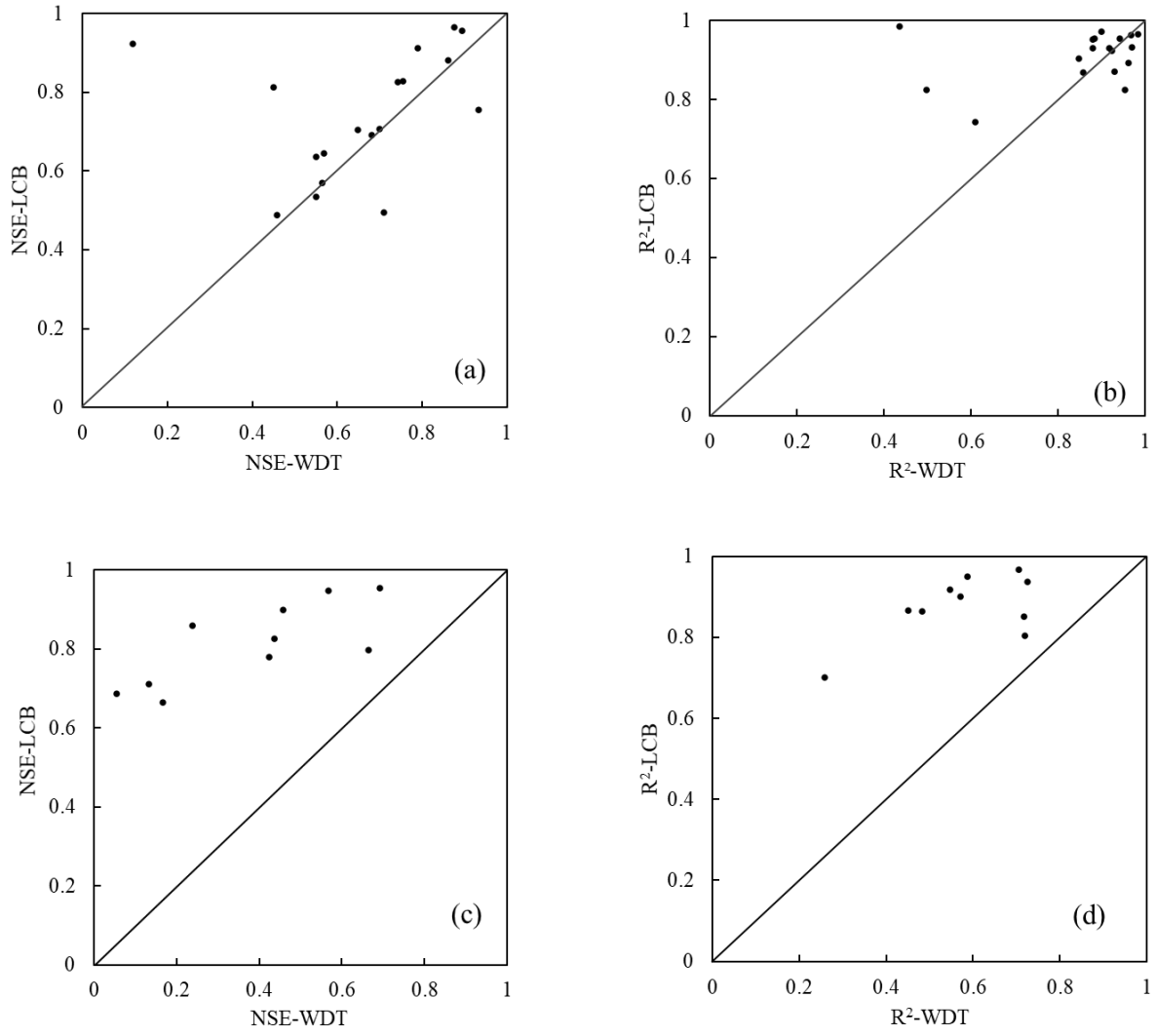


Fig. 3-2. Comparison of the LCB and WDT approach performance simulating stormwater runoff in RR: (a-b) NSE and R^2 for the calibration period of 2018; (c-d) NSE and R^2 for the validation period of 2019.

Fig. 3-3 showed the comparison of two models' performance for water quality in the RR catchment in both 2018 and 2019. The dotted gray lines and dotted black lines showed the 25% and 50% relative error envelope, respectively. Relative error lower than 25% indicates a good agreement between simulated and observed EMC's and relative error lower than 50% are deemed to be

acceptable (Tuomela et al., 2019). For TSS simulation, 9 out of 14 events of the LCB approach were acceptable and 6 of them had good agreement with observed TSS concentrations. Similarly, 9 out of 14 events of the WDT approach were acceptable but 7 of them had good agreement with observed TSS concentrations. For TN simulation, 11 out of 14 events of the LCB approach were acceptable and 8 of them have good agreement with observed TN concentrations. While only 9 out of 14 events of the WDT approach were acceptable, 7 of them had good agreement with observed TN concentrations, and 3 out of 14 events are obviously overestimated. For TP simulation, the LCB approach and the WDT approach had 8 and 6 events that were acceptable, respectively, and 6 out of 14 events of the WDT approach are obviously underestimated. Both models showed good performance in four events (i.e., relative was lower than 25%).

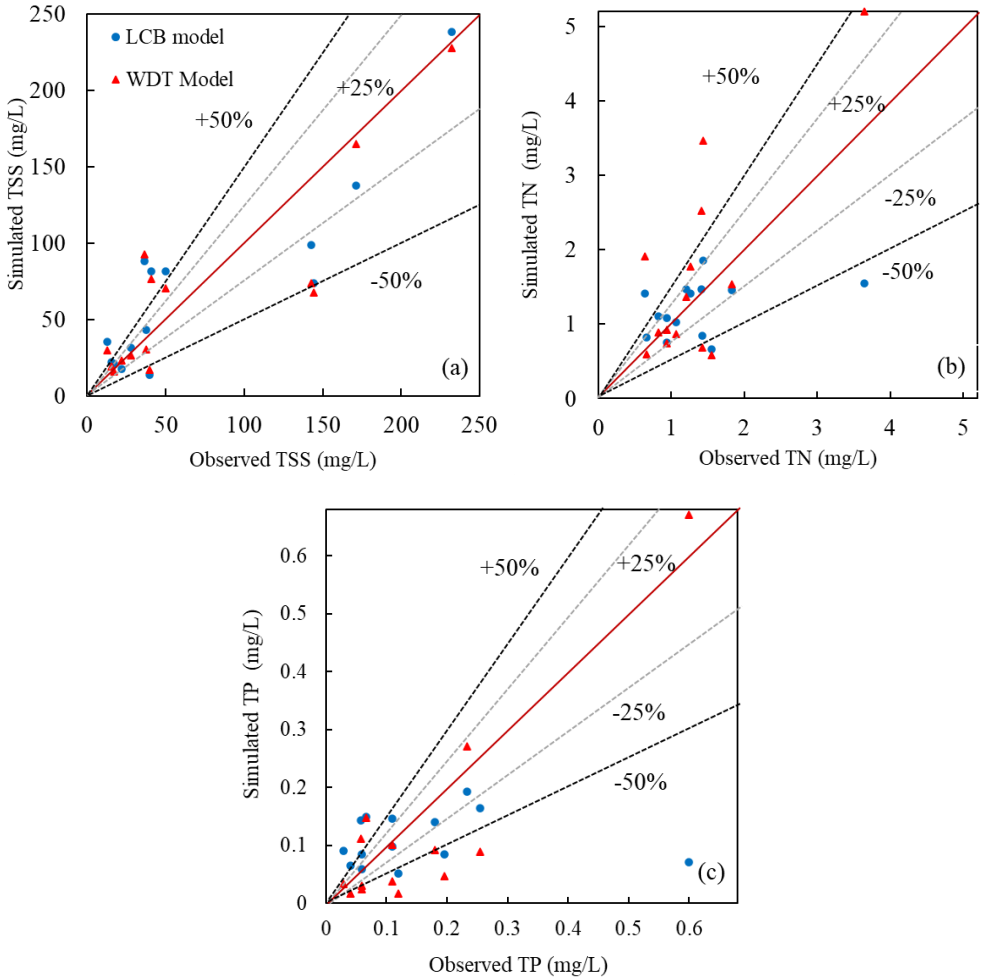


Fig. 3-3. Measured versus predicted concentrations of (a) TSS, (b) TN and (c) TP for storm events in 2018 and 2019 in the RR catchment.

Table 3-5 summarized two models' performance for water quantity and quality simulations in both 2018 and 2019. Table 3-5 provided the range and mean value of NSE and R^2 to quantify overall performance for water quantity simulation and presents the percentages of acceptable and accurate events to all storm events to assess model performance for water quality simulation. It was shown that both minimum and maximum value of NSE and R^2 for the LCB approach were larger than that of the WDT approach. The mean value of NSE and R^2 for the LCB and the WDT approach were 0.56, 0.77, 0.75 and 0.90, respectively, which indicates the LCB approach was more superior than the WDT approach for water quantity simulation. This finding is in line with the studies of Petrucci et al. (2014) and Rio et al. (2020). They reported that model performance for water quantity simulation was clearly improved by discretizing catchment into homogeneous unit based on land cover information. As can be seen from Table 3-5, 67% of storm events can be simulated with acceptable accuracy by the LCB approach in term of water quality simulation, which was slightly larger than that of the WDT approach (57%). Both models can simulate water quality with good accuracy for 43% of storm events. Petrucci et al. (2014) reported that model with sufficient land cover information had slight better performance than model with less land cover information, which showed that that an increase in variability associated to land uses had no considerable effects.

Table 3-7. Summary of model performance for water quantity and quality simulations.

Water quantity				Water quality			
Criterion	Model	Range	Mean	Criterion	Model	% Acceptable events*	% Accurate events*
NSE	WDT	0.12~0.93	0.56	Relative Error	WDT	57%	43%
	LCB	0.49~0.97	0.77				
R^2	WDT	0.26~0.98	0.75		LCB	67%	43%
	LCB	0.7~0.98	0.90				

Note: * % Acceptable events mean the percentages of acceptable events to all storm events calculated by relative error, and % Accurate events mean the percentages of accurate events to all storm events calculated by relative error.

3.3.3 Parameter transferability

To test the parameter transferability, the RO catchment was considered as an ungauged area. The parameters resulting from calibration of the two models in the RR catchment were directly transferred to the RO catchment. Fig. 3-4 showed the comparison of NSE and R^2 values of all

events simulated for 2018 and 2019. By comparing the application of the two models, it was shown in Fig. 3-4a and 3-4b that in 56% (five out of nine) of the events, the NSE of the LCB approach lied at or close to the 1:1 line in 2018 and the NSE of the LCB approach was lower than WDT approach in other four events. In 78% (seven out of nine) of events, R^2 of the LCB approach was equal or higher than that of the WDT approach. Based on the Nash efficiency, the LCB and the WDT approach both had five events that are acceptable (i.e., NSE $\geqslant$0.5). Based on coefficient of determination, the LCB approach and the WDT approach had seven and eight events that were simulated accurately, respectively (i.e., $R^2 \geqslant 0.7$).

Fig. 3-4c and 3-4d showed that in 2019, the LCB approach was better than the WDT approach for all ten events. Based on the Nash efficiency, the LCB approach simulated 80% (eight) of events with acceptable accuracy (i.e., NSE was higher than 0.5), while the WDT approach only simulated 10% (one) of event acceptably. Based on the coefficient of determination, the LCB and WDT approach performed acceptably (i.e., R^2 was higher than 0.7) for 70% and 40% of the events.

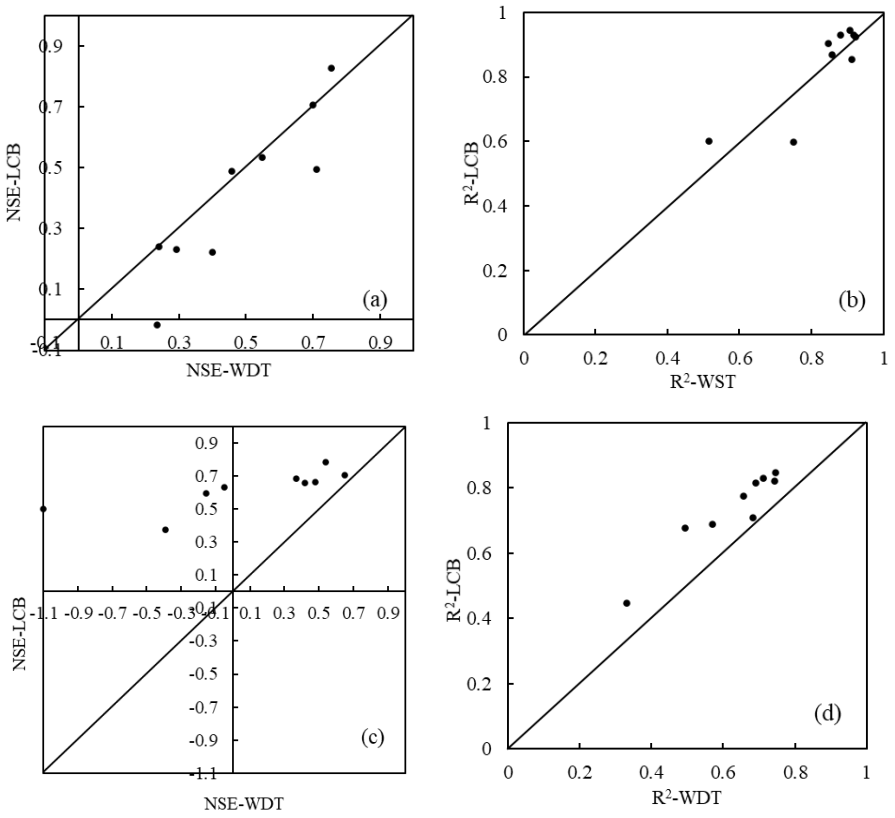
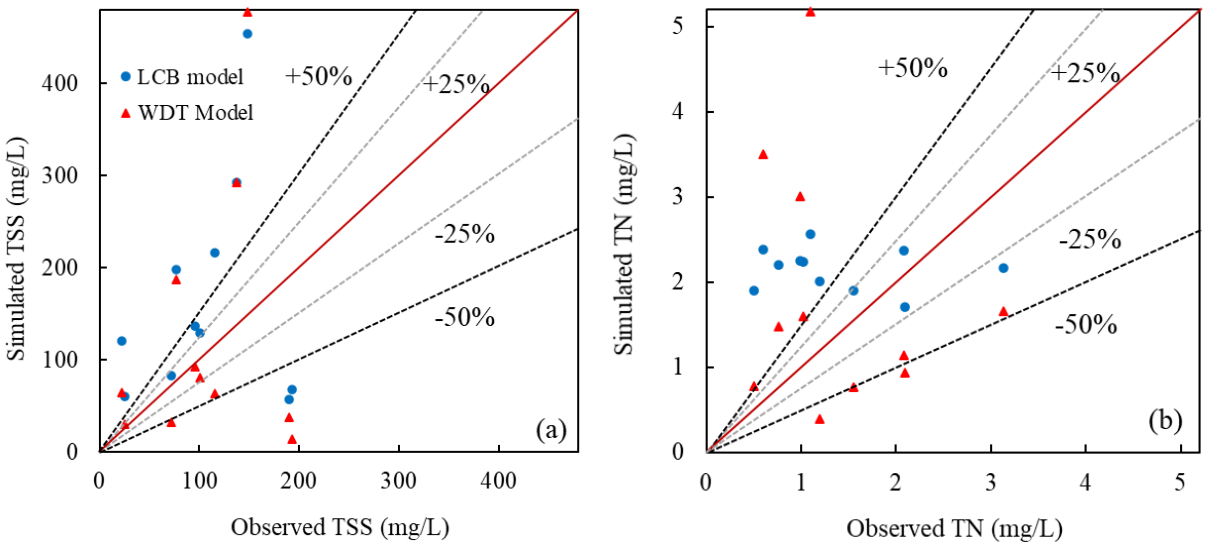


Fig. 3-4. Comparison of the LCB and WDT approach performance simulating stormwater runoff in RO: (a-b) NSE and R^2 in 2018; (c-d) NSE and R^2 in 2019.

In Fig. 3-5, a comparison of the two models' performance predicting water quality in the RO catchment in 2018 and 2019 was presented. Measurements of TSS, TN and TP are plotted versus predicted values along with lines showing the 25% and 50% error envelopes. Relative error lower than 25% and 50% indicates satisfactory and acceptable agreement between simulation and observation pollutant concentrations, respectively (Tuomela et al., 2019). For TSS simulation, 27% (3 out of 11) of simulated events of the LCB approach were acceptable and only 9% (1 out of 11) had good agreement with observed TSS concentrations. Similarly, 4 out of 11 events of the WDT approach were acceptable and two of them had good agreement with observed TSS concentrations. For TN simulation, 4 out of 11 events of the LCB approach were acceptable and 3 of them had good agreement with observed TN concentrations. While none events of the WDT approach had good agreement with observed TN concentrations. TN concentrations of 82% of events were overestimated by the LCB approach. The simulated TN concentrations for 55% of events were larger than observed TN concentrations, while 45% of events are underestimated by the WDT approach. For TP simulation, the LCB approach and the WDT approach had 3 and 2 events that were acceptable, respectively. The range of TP concentrations simulated by the LCB approach was 0.11~0.33mg/L, which was smaller than the measured concentrations' range (0.05~0.42mg/L). The simulated TP EMCs by the WDT approach were scattered and did not fit well with the measured TP concentrations. When the parameters were transferred to the elongated catchment (RO catchment), both LCB and WDT approach could result in overestimation or underestimation of TSS, TN and TP.



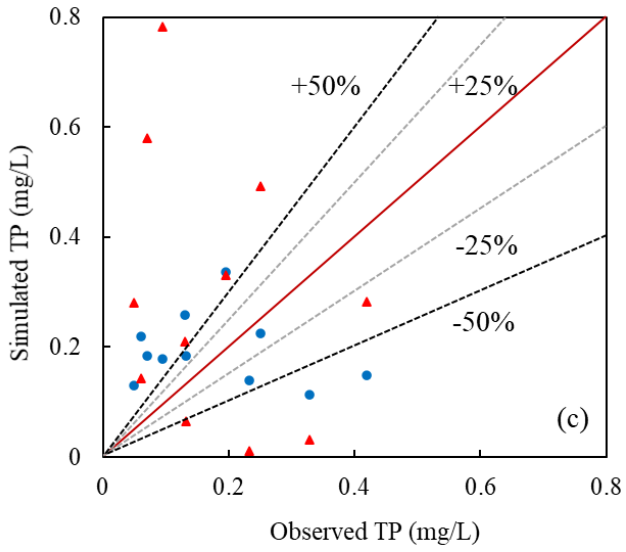


Fig. 3-5. Measured versus predicted concentrations of (a) TSS, (b) TN and (c) TP for storm events in 2018 and 2019 in the RO catchment.

Table 3-6 summarized parameter transferability of the two models for water quantity and quality in both 2018 and 2019. Results for water quantity showed that the performance of the WDT approach markedly decreased when model parameters were transferred from RR to RO catchments, and the WDT approach was generally unacceptable (mean NSE < 0.5). Despite the performance of the LCB approach also declined, it still showed better performance (NSE > 0.5, R^2 > 0.7). Both minimum and maximum value of NSE and R^2 for the LCB approach were larger than that of WDT approach. It showed that it was feasible to transfer hydrologic-hydraulic parameters to an ungauged catchment with similar land cover types. As presented in Table 3-6, less than 30% of storm events could be simulated with acceptable accuracy by the LCB approach and the WDT approach, which means the two models had similar performances for water quality modelling. Although the LCB approach could simulate water quality with good accuracy for more storm events than the WDT approach, its performance was not satisfactory. Therefore, even in adjacent catchments with similar catchment characteristics, the transfer of water quality parameters should be done cautiously.

Table 3-8. Summary of parameter transferability for water quantity and quality.

Water quantity				Water quality			
Criterion	Model	Range	Mean	Criterion	Model	% Acceptable events*	% Accurate events*
NSE	WDT	-1.1~0.76	0.32	Relative Error	WDT	27%	6%
	LCB	-0.02~0.82	0.53				
R^2	WDT	0.33~0.92	0.73		LCB	30%	15%
	LCB	0.45~0.94	0.79				

Note: * % Acceptable events mean the percentages of acceptable events to all storm events calculated by relative error, and % Accurate events mean the percentages of accurate events to all storm events calculated by relative error.

3.3.4 Assessment of runoff and pollutant loading from different land covers

Fig. 3-6 showed the contribution of land covers to runoff and the losses of rainfall runoff in RR based on the water quantity simulation results. As shown in Fig. 3-6a, in the dryer year of 2018, almost all the rainfall on the roofs generated runoff, with a runoff coefficient as high as 0.98. This was because the roofs in this catchment were inclined roofs, which have negligible detention storage. Parking was similar, with a runoff coefficient as high as 0.97. Even though they were impervious, the runoff coefficient for roads was 0.83. This was likely because there were many cracks in the roads, where rainfall was temporarily stored and eventually lost as evaporation. As shown in Fig. 3-6a, in the dryer year of 2018, the volume of rainfall and runon for greenspace was 14,521 m^3 and 6102 m^3, respectively. The local infiltration and infiltration of runon were 14,442 m^3 and 6,065 m^3, respectively. The infiltration losses of rainfall and runon reached up to 99%. Not only was the rainfall of greenspace mostly infiltrated, but also the runoff from any upstream subcatchment was mainly lost through infiltration in greenspace. The runoff flowing into greenspace was primarily from impervious areas, which shows the importance of greenspace in reducing urban runoff. Although trails were also considered pervious, they had less impact on runoff due to their small area. This analysis shows that Roof, Roads and Parking were main sources of runoff for RR in 2018.

Fig. 3-6b showed the results for the wetter year of 2019. The results for performance of impervious areas were similar to that of dryer year. Almost all rainfall on roofs and parking produced runoff, and the runoff coefficient was higher than 0.98. Despite a lot of cracks and seams existed in the

roads, the runoff coefficient was 0.89 which was higher than that of dryer year of 2018 due to the heavier rainfall. The runoff coefficient reached 0.15, compared with 0.003 in the drye year of 2018. Greenspace was not able to treat all rainfall and runon from upstream subcatchments. Infiltration loss of runon was 84%, compared to 99% in 2018.

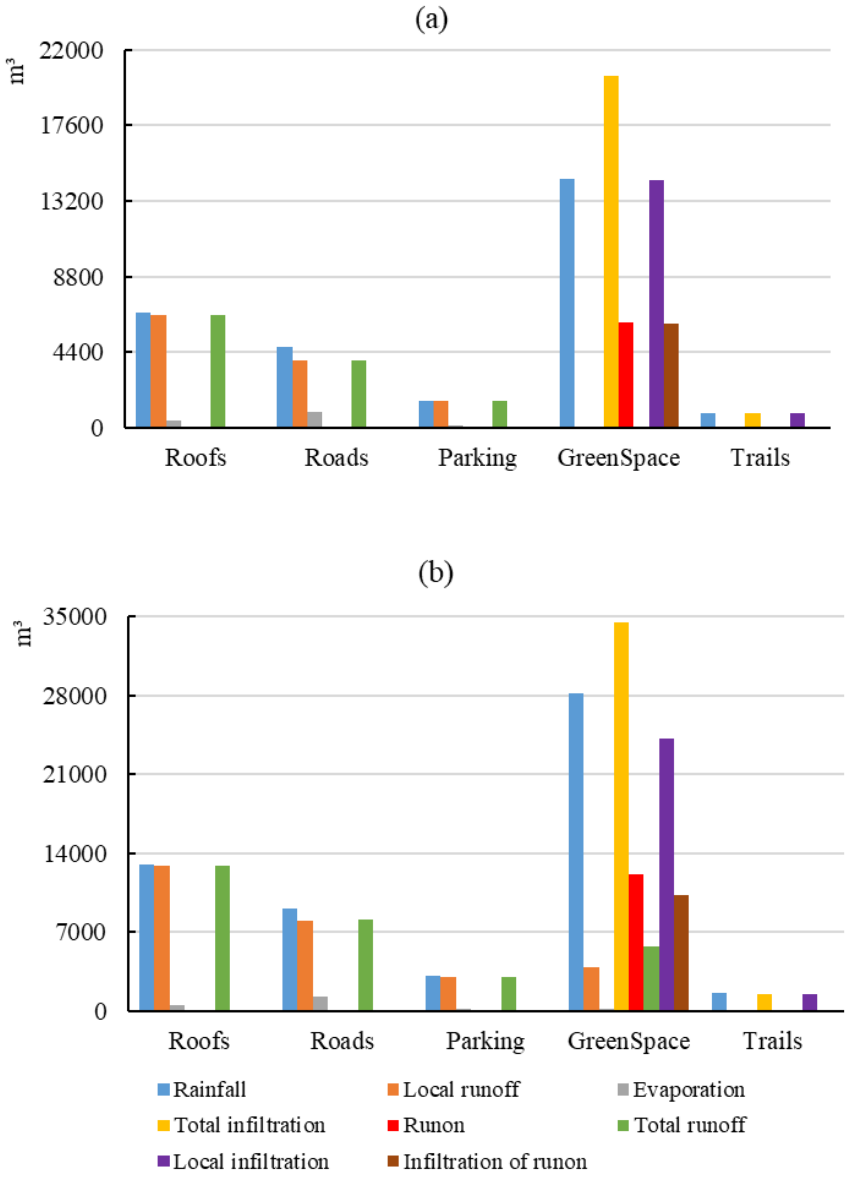


Fig. 3-6. The contribution of land covers to runoff and the losses of rainfall runoff in RR based on the water quantity simulation results. (a) 2018; (b) 2019.

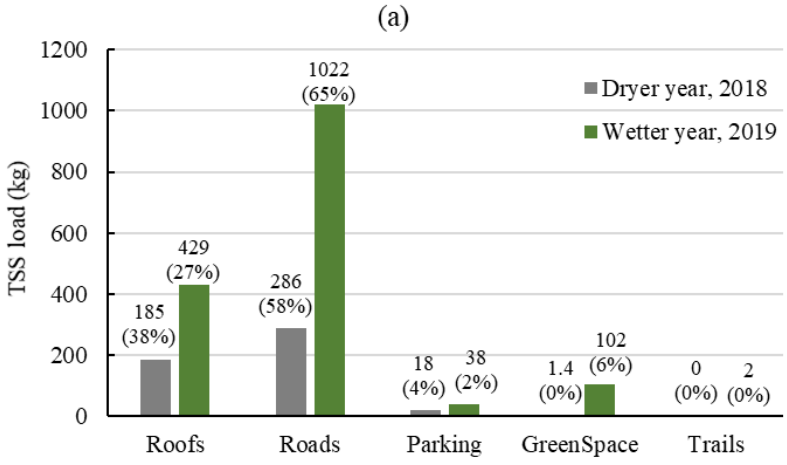
The calibrated water quality module of the LCB approach in RR catchment was used to assess the contribution of different land cover types to the output of pollutants in different precipitation

conditions. It should be noted that the uncertainty of water quality modelling results could have impacts on the assessment of runoff and pollutant loading. Model structure, input parameters and measured data uncertainty are the main sources of uncertainty (Liu et al., 2015). Uncertainty resulting from the model structure is primarily related to the lack of knowledge on the influential factors on stormwater quality. For example, uncertainty inherent to pollutant build-up and wash-off processes is not adequately represented in stormwater models (Wijesiri et al., 2016). Insufficient input parameters are among the most important sources of model uncertainty. For example, lumped parameters usually cannot accurately represent the variation of water quality within a specific land use area (Liu et al., 2012). Measured datasets, such as rainfall data, have inherent uncertainty, which increases the data requirements for model calibration (Dotto et al., 2014). Even though the LCB approach improved the performance of water quality modelling, the uncertainty is still inevitable. Therefore, the analysis was based on the assumption that the LCB approach functioned well in water quality modelling.

Fig. 3-7 showed the contribution of different land cover types to pollutant loadings. Precipitation from May to October 2018 was 201mm, and in 2019 it was almost double at 390 mm. As shown in Fig. 3-7, roads accounted for only 14.8% of the catchment total area, but produced more than 55% of pollutants, with up to 58% of TSS loads. Roofs were the largest impervious area, accounting for 21.3% of the catchment area and account for ~40% of the pollutant loading, which was less than the contribution from the roads. In the dryer year of 2018, roads and roofs were the main sources of pollutants. This is consistent with Ashley et al. (2004) who concluded that roads and roofs are a principal source of pollutants of runoff water. They reported that pollutants washed off from roofs during storm events have a number of different origins: pollutants from atmosphere deposited during dryer weather periods; pollutants from chemical and physical degradation of the roof and gutter material itself; pollutants from bird droppings, branches and leaves. Although impervious areas accounted for only 41% of the catchment area, they were the source of 96-100% of all pollutants. It is noteworthy that the contribution of impervious areas to water quality would change seasonally. Carroll et al. (2013) reported that the influence exerted on the water environment by seasonal factors is secondary to that of land use. The seasonal variation of application of fertilizer and degradation of leaf litter could result in the change of inputs of nutrients into the surface water. The most considerable difference between the wetter year of 2019 and the dryer year 2018 was the greenspace pollutant loading which increased from approximately zero to

~7%. However, compared with the contribution of pollutants from impervious areas, the pollutants produced by greenspace were still small. Tuomela et al. (2019) found that vegetation in wet year was the source of for ~10% of pollutants, which is similar to this study. During the wetter year, the proportion of TP and TSS produced from roads increased 2% and 6%. In addition, the percentage of pollution produced from roofs decreased 7% to 11% in the wetter year.

The pollutant loading in the two years was also compared in Fig. 3-7. The TSS loading from roofs and roads was 2.3 and 3.6 times higher, respectively in the wetter year. TSS loading from greenspace was 1.4 kg in 2018 and increased to 102 kg in 2019. Although parking was shown to be an important source of pollution in some studies (Bannerman et al., 1993; Tuomela et al., 2019), its contribution was small in this study as parking areas account for only 5.2% of the catchment. Trails were only 2.6% of the area, and as a result its contribution was negligible. The total precipitation in 2019 is ~2 times that of 2018, but the TSS load generated in 2019 is 3.3 times that of 2018. TN and TP loads showed similar conclusions. Pollutant loading was larger for all land cover types in the wetter year compared to the dryer year. Therefore, precipitation clearly had an important influence on the pollutant loading during ice-free seasons.



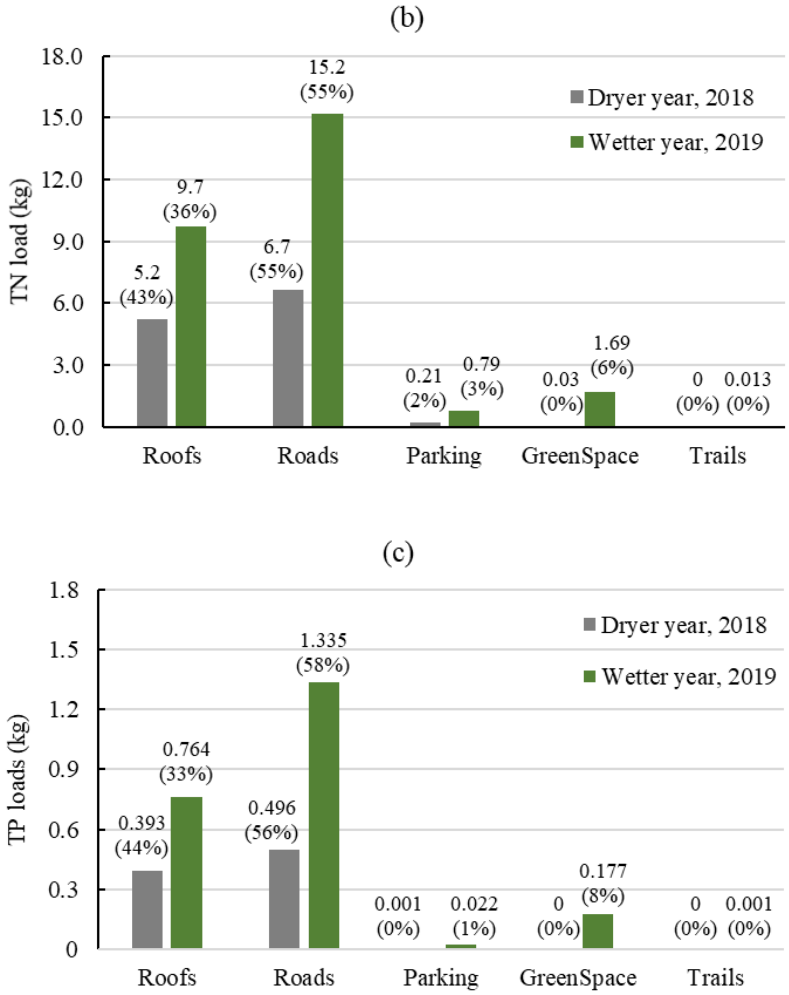


Fig. 3-7. Contribution of different land cover types to the output of pollutants in different precipitation conditions in RR based on water quality simulation results: (a)TSS; (b) TN; (c)TP. The values in bracket present the percentage of pollutant loading from different land cover types.

3.3.5 Limitations and future work

The LCB approach outperforms the WDT approach in hydrological simulation, with better accuracy due to enhanced land cover consideration. It can quantitatively evaluate the contribution of different land covers to runoff and pollutant loads. However, the transfer of water quality parameters in the LCB approach is not as satisfactory as its hydrologic-hydraulic parameters.

The result of this study may have been influenced by the limitation of the data used. Firstly, water quantity and quality data were only collected during ice-free seasons in two years for the two

catchments. Water samples were not adequately collected in some small rainfall events due to some reasons (e.g., battery power failures, "No Liquid Detected" errors). Therefore, the water quality conditions may not be sufficient for the analysis of model performance in small rainfall events. In addition, land survey data might not be sufficient to accurately reflect the building/rooftop connectivity, which may influence the model output of water quantity and quality. Future work should also focus on more comprehensive sampling programs with a longer duration and more detailed survey of building/rooftop drainages in order to produce more accurate model predictions.

LCB approach usually has a high resolution which could describe some spatial variability of the catchment characteristics. Manual delineation approach could be time consuming and laborious when it is applied for large urban areas due to the diversity of urban land cover. Developing advanced automatic delineation tool with detailed land cover will be necessary for the application of LCB approach in large urban areas. Five major land cover types were considered in this study. The relatively simple classification approach could conceal some important information. For example, common roofs and green roofs behave differently in terms of runoff export. Pollutant outputs of artificial garden and natural grassland are also different due to variable fertilization patterns. Clustering of land cover may have an impact on the study of model performance and parameter transferability. Therefore, whether increasing the number of land cover types can improve simulation accuracy needs to be studied.

3.4 Conclusions

This paper built a new land-cover based (LCB) PCSWMM model to study the impacts of land cover on the simulation of stormwater runoff and pollutant loading. The parameter sensitivity analysis showed that the most sensitive parameters in the WDT approach were the Percent Routed and Imperviousness, while these two parameters were completely insensitive in the LCB approach. The hydrological simulations showed that the NSE range of the WDT approach was 0.12 ~ 0.93, and that of the LCB approach was 0.49~0.97. The mean NSE values of the LCB and WDT approaches were 0.56 and 0.77, respectively. It indicates that the LCB approach is better than the WDT approach for hydrological simulation. In general, the LCB approach simulated all three water quality variables with acceptable accuracy more often than the WDT approach, while the model performance was comparable in simulating with good accuracy for both models.

Testing the transferability of the model parameters showed that the mean NSE value of the LCB approach (0.53) was larger than that of the WDT approach (0.32), which indicates that the hydrologic-hydraulic parameters could be satisfactorily transferred from neighboring gauged catchments to similar ungauged catchments. However, the transfer of water quality parameters did not perform as satisfactorily. Therefore, even in adjacent catchments with similar catchment characteristics, the transfer of water quality parameters should be done cautiously.

The LCB approach provides an opportunity to quantitively evaluate the contribution of different land covers to runoff and pollutant loads. Roofs and roads were found to be major contributors to urban runoff. In drier years, green spaces produced small proportion of pollutants, while in wetter years they become a important contributor of pollutants. Roofs and roads were the main sources of pollutants. Although impervious areas accounted for only 41% of the RR catchment area, over 92% of pollutants loading came from these areas. Precipitation conditions had an important influence on the contribution of pollutant output from different land covers.

4. Effects of mixed land use on urban stormwater quality under different rainfall event types

4.1 Introduction

Urban stormwater quality is a concern in urban planning and environmental management due to its considerable impact on water bodies and ecosystems (Goonetilleke et al., 2005; Roy et al., 2008). The quality of stormwater runoff is influenced by complex factors, with land use patterns and rainfall characteristics playing prominent roles (Reichwaldt et al. 2012; Beck et al. 2016; Simpson et al. 2022; Yan et al. 2022). Nutrients are primary constituents of concern in stormwater due to their potential to cause eutrophication and harmful algal blooms when present at excessive levels. Suspended solids can impair water quality by reducing light penetration and oxygen levels essential for aquatic life (Aryal et al. 2010; Behrouz et al. 2022). Land uses such as residential, commercial, industrial, and recreational areas, coupled with varying rainfall characteristics, contribute to the variability of urban stormwater quality (Liu et al. 2015; Guzman et al. 2022).

Numerous studies have explored the relationship between land uses and urban stormwater quality. Research has shown that different land uses generate distinct types and levels of pollutants in stormwater runoff. Industrial land use tends to produce more solids, particularly fine particles, than other land uses from Gold Coast, Queensland State, Australia (Miguntanna et al., 2010). A three-year study by Li et al. (2015) revealed that residential and commercial catchments in Dongguan City, Southeast China exhibited the highest median event mean concentrations (EMCs) for various pollutants, surpassing industrial and parking areas. Yazdi et al. (2021) found that EMCs of total suspended sediment (TSS) in runoff from open space and industrial areas in City of Virginia Beach, USA were notably higher than other urban land uses after analyzing water samples from catchments with six different urban land uses during 30 storm events. Although studies have primarily examined the impact of specific land uses on stormwater quality, the complexity introduced by mixed land use, wherein diverse land uses coexist in close proximity, can further complicate these influences (Lee et al., 2009, Zivkovich et al., 2018). Generally, diverse urban catchments tend to produce a wider variety of pollutants. Previous research has conducted monitoring and modelling works of urban stormwater quality from mixed land use catchments (Kellner et al., 2017; Paule-Mercado et al., 2018; Kshirsagar et al., 2023). While there have been

extensive studies on the impacts of individual land uses on urban stormwater quality, the specific effects of mixed land use patterns, particularly their spatial configurations, have not been thoroughly investigated.

Rainfall characteristics, such as intensity, duration, and antecedent dry days, also substantially impact urban stormwater quality (Gong et al. 2016, Goonetilleke et al. 2005, Li et al. 2015, Wijesiri et al. 2015). Rainfall intensity, driven by its kinetic energy, plays a crucial role in pollutant wash-off. Research found that both the volume and intensity of rainfall typically elevate TSS concentrations in runoff, regardless of land use (Yazdi et al. 2021). Antecedent dry days considerably impacted TSS and metal loads in stormwater runoff, with pollutants accumulating quickly at the beginning of the dry period (Murphy et al. 2015). Rainfall characteristics exert a notable influence on water quality, not merely through their individual impacts but through their combined effects, particularly when considering rainfall event types (Yan et al., 2023). Although research has extensively examined the relationship between specific rainfall characteristics and stormwater quality, the joint effects of rainfall event type and land use on urban stormwater quality need to be further studied.

Land use can also affect nutrient forms, with nutrient components being crucial for effective urban stormwater treatment (Taylor et al. 2005). For example, stormwater wetlands prioritize removing coarse materials, then finer particulates, and lastly dissolved components (Tanner et al., 2011). A study in Melbourne found stormwater nitrogen to be mainly dissolved, with ammonia being the least common (Taylor et al. 2005). Research in Tampa Bay, Florida, United States emphasized how urban heterogeneity, like residential development patterns, can influence nitrogen and phosphorus transport even in similar geologic and climatic conditions (Yang and Toor 2017). A study in South-east Queensland, Australia found that runoff from urban residential catchments contained on average nitrogen oxides at 16% and ammonia at 9% of total nitrogen, with Total Kjeldahl Nitrogen (TKN) averaging 84% of event-based total nitrogen concentrations (Lucke et al. 2018). Despite existing research on nutrient composition across urban land uses, the temporal variation under different rainfall conditions remains understudied.

This research aims to address these gaps by conducting a comprehensive two-year water field program across four mixed land use areas in Calgary, Alberta, Canada. The primary objective is

to investigate the effects of mixed land use on urban stormwater quality under different rainfall event types. By assessing the joint impacts of land use and rainfall characteristics, this study seeks to provide insights into the complex interactions shaping stormwater quality dynamics. Additionally, the research aims to contribute to a deeper understanding of nutrient composition variations across different urban land uses under varying rainfall conditions. The findings from this study have the potential to inform more effective and targeted urban stormwater management strategies.

4.2 Materials and methodology

4.2.1 Description of the study areas

The study sites include four urban catchments located in Calgary, Alberta, Canada, which are the Royal Oak (RO), Rocky Ridge (RR), Cranston (CR) and Auburn Bay (AB) catchments, as shown in Fig. 4-1. Calgary is in a semi-arid, cold temperate climate region, with cold winters and mild to warm summers. Table 4-1 lists the catchment characteristics of the study areas. The four urban catchment areas exhibit slightly land use and land cover characteristics. RR and RO, located only 300 m apart, are smaller suburban catchments with areas of 14 ha and 15 ha, respectively. They primarily consist of single-family residential areas, with RR at 72% and RO at 51%. RO is slightly more diverse, incorporating 7.3% multi-family residential area. Both catchments have transportation and open space components, with imperviousness levels of 48% for RR and 51% for RO. The land cover types are primarily composed of asphalt/concrete, landscaped area and roofs. Both RO and RR have a wetland to receive the generated runoff.

In contrast, CR and AB are larger and more urbanized, located approximately 1.5 km apart. CR has an area of 119 ha with 52% imperviousness, while AB, is 223 ha with 58% imperviousness. Both catchments mix multiple land uses, incorporating commercial, institutional, and multi-family residential areas. A unique feature of CR and AB is the presence of gravel lanes, covering 2% and 3% of their areas, respectively. The runoff generated in CR flows into a single wet pond, while AB has an upstream pond that collects runoff from 69-ha that conveys this to a larger wet pond that also receives runoff from the remaining 154 ha.

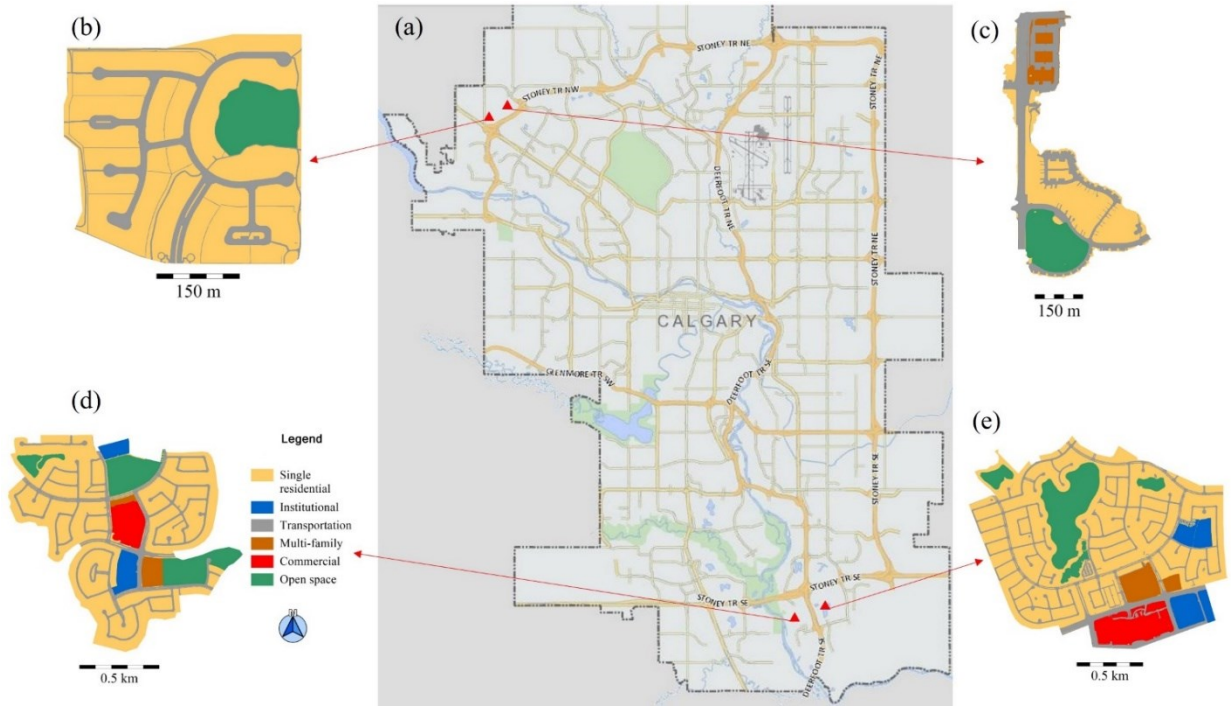


Fig. 4-1. Location photos of the study sites: (a) locations of four catchments in Calgary (Google, 2023); (b) RR catchment; (c) RO catchment; (d) CR catchment; (e) AB catchment.

Table 4-1. Catchment characteristics of the study areas.

Catchment		RR	RO	CR	AB
Area (ha)		14	15	119	223
Imperviousness (%)		48	51	52	58
Land Use (%)	Single family residential	72	51	58	47
	Multi-family residential	0.0	7.3	3.4	3.1
	Transportation	18	28	19	20
	Open space	10	14	13	18
	Institutional	0.0	0.0	3.7	6.1
	Commercial	0.0	0.0	3.1	5.9
Land Cover (%)	Roads	25	37	26	28
	Roofs	23	16	21	23
	Landscaped area	52	49	51	46
	Gravel lanes	0.0	0.0	2.2	3.3

4.2.2 Field measurements

Field monitoring was carried out during the ice-free periods (May to October) in 2018 and 2019. Each catchment was equipped with a weather station, flow meters, and autosamplers to gather required data. The weather station included a HOBO RX300 Remote Monitoring Station paired with a HOBO RG3-M Data Logger (Onset Computer Corp., USA). A tipping bucket rain gauge was used to measure rainfall, activated at a minimum of 0.2 mm cumulative rainfall, and data were logged every five minutes. Flow meters and autosamplers were installed in the inlet pipes at the stormwater wetlands and wet ponds to measure flowrates and collect water samples at catchment outlets. Note that the catchment outlets are the inlets of stormwater wetlands and wet ponds. Table 4-2 provides details about the equipment used at the catchments' outlets during the two-year field program in 2018 and 2019. The setup involved ISCO 6712 autosamplers (Teledyne ISCO, USA) equipped with ISCO 750 area velocity flow modules and ISCO 2150 area velocity flow modules. The utilization of multiple flow modules was to ensure the accuracy of measured flowrates. These flow module sensors, installed in the inlet pipes, automatically calculated inflow rates using the pipe diameter and the measured water depths and velocities.

Table 4-2. Equipment installed at the four catchments outlets for flow measurement and water sampling.

Catchment	Equipment	2018 Count	2019 Count
RR	HOBO RX300	1	1
	6712 Autosampler	1	1
	750 Flow sensor	1	1
RO	HOBO RX300	1	1
	6712 Autosampler	1	1
	750 Flow sensor	1	1
	2150 Flow sensor	2	2
CR	HOBO RX300	1	1
	6712 Autosampler	2	2
	750 Flow sensor	2	2
	2150 Flow sensor	0	2

AB	HOBO RX300	1	1
	6712 Autosampler	2	2
	750 Flow sensor	2	2
	2150 Flow sensor	2	2

Note: The "/" in the table represents no equipment of that type installed during that year.

Flow meters recorded flow rates at five-minute intervals. ISCO 6712 autosamplers, equipped with 24×1 L sample bottles, were used for water sample collection. Water samples were used for analyzing suspended solids and nutrient (nitrogen and phosphorus) concentrations. Each autosampler was paired with either an ISCO 750 AV flow module or an ISCO 2150 flow module. These recorded water levels and flow velocity control the ISCO 6712 sampler program. Considering the maximum collection capacity of ISCO 6712, and to optimize water quality sample volume during runoff events, three flow-paced configurations were utilized: low (targeting ~5 mm to 10 mm rainfall events), high (up to 20 mm rainfall events), and extreme (above 20 mm rainfall events). The selection of the suitable flow-paced setup was made by monitoring weather and anticipated rainfall. The triggering water levels of ISCO 6723 unit were accordingly adjusted.

For this research, aliquot samples from each storm event were mixed to create composite samples, representing approximately 75% of the storm runoff volume. To maintain sample quality, composite sample bottles were sent to the laboratory within a day of collection or were frozen. At the RR, RO, CR and AB catchments 15, 13, 18, and 11 events were recorded, respectively, totaling 57 events.

4.2.3 Analytical parameter and nutrient composition

To explore the effects of land use on stormwater quality, this study selected event mean concentration (EMC) and event pollutant load (EPL) as the most relevant parameters. EMC is a crucial parameter used in water quality studies to represent the average concentration of a pollutant in runoff during a specific rainfall event (Li et al., 2015). EPL is another important parameter that represents the total amount of pollutants discharged from a specific area during a rainfall event (Li et al., 2015). These parameters effectively reflect the levels and loading of pollutants, offering

insight into the environmental impacts under various land use conditions. EMC and EPL are defined as follows:

$$EMC(mg/L) = \frac{M}{V} = \frac{\sum_{i=1}^{n} C_i Q_i \Delta t}{\sum_{i=1}^{n} Q_i \Delta t} = \frac{\sum_{i=1}^{n} C_i V_i}{\sum_{i=1}^{n} V_i} \quad (4\text{-}1)$$

$$EPL(mg/m^2) = \frac{M}{A} = \frac{\sum_{i=1}^{n} C_i V_i}{A} \quad (4\text{-}2)$$

where M is the pollutant mass during the rainfall event; V is the runoff volume during the rainfall event; C is the pollutant concentration at time i; Q is the flow rate at time i ; Δt is the interval time of sampling and A is the catchment area.

To ensure a comprehensive assessment, five forms of both nitrogen and phosphorus were measured. The forms of nitrogen included Ammonia (NH_4^+), Nitrate and Nitrite(NO_2^-/ NO_3^-), Total Organic Nitrogen (TON), Total Kjeldahl Nitrogen (TKN) , and Total Nitrogen (TN). Concurrently, the phosphorus evaluation included Total Dissolved Phosphorus (TDP), Total Particulate Phosphorus (TPP), Total Reactive Phosphorus (TRP), Non-reactive Phosphorus (NRP), and Total Phosphorus (TP). This extensive nutrient assessment contributes to a more comprehensive understanding of the varied water quality dynamics across different land use and rainfall conditions.

To explore the effects of mixed land use on urban stormwater quality under different rainfall event types, it is important to classify rainfall events into different types that can reflect the rainfall characteristics in the studied areas. The rainfall events classification follows the same procedure as section 2.3.4. The approach and results will not be repeated here.

4.2.4 Statistical analysis

Statistical tests were performed on both EMCs and EPLs to assess whether there were significant differences in pollutant levels between the study areas. The normality of the dataset was assessed using the Shapiro–Wilk test method using a significance level of α = 0.05 before further statistical analysis (Das and Imon 2016). The Kruskal–Wallis test, an analysis of variance (ANOVA) using a significance level of α=0.05, was performed to test whether there are significant differences between the pollutant concentrations from the different study areas. Finally, the Tukey honest

significant difference (HSD) test, a multiple comparison test, was conducted to determine which land use differs from others.

Correlation matrix analysis was performed to determine the linear correlation between all pollutant components. The degree of correlation was classified into five types based on the magnitude of the Pearson correlation: 1) .00 to .30 (.00 to -.30): negligible correlation, 2) .30 to .50 (-.30 to -.50): low positive (negative) correlation, 3) .50 to .70 (-.50 to -.70): moderate positive (negative) correlation, 4) .70 to .90 (-.70 to –.90): high positive (negative) correlation, and 5) .90 to 1.00 (-.90 to -1.00): very high positive (negative) correlation (Mukaka, 2012).

4.3 Results and discussion

4.3.1 Correlation analysis

In Fig. 4-2 the correlation matrix between EMCs of TSS and the nitrogen and phosphorus components is presented. TSS was highly positively correlated with NRP (0.87) and TP (0.88), indicating that these phosphorus forms may predominantly exist in particulate form in urban runoff (Li et al., 2018). This suggested that strategies focusing on reducing particulate matter in stormwater could also substantially lower phosphorus levels (River et al., 2018). The moderate positive correlation between TSS and TON (0.53), TKN (0.56), and TN (0.54) implies that these nitrogen forms were partially present in particulate form (Janke et al., 2014), contributing to the overall nitrogen levels in the stormwater. NH4+ (0.30) and NO_2^- / NO_3^- (0.06) showed low positive and negligible correlation with TSS, indicating they were predominantly dissolved in runoff (Jani et al., 2020).

In the nitrogen components, a very high positive correlation (0.98) was observed between TKN and TON, which was expected as TKN includes TON. The very high positive correlation (0.94) between TON and TN emphasized the contribution of organic nitrogen to the total nitrogen in stormwater (Jani et al., 2020), stressing the need to consider organic nitrogen sources such as plant debris and leaf litter in nitrogen management strategies. The correlation between other nitrogen components such as NH4+ and NO_2^- / NO_3^- (0.30) was relatively low, indicating diverse sources and different forms of nitrogen in the stormwater, which may not always be interconnected (Jani et al., 2020).

Regarding phosphorus components, a very high positive correlation between NRP and TP (0.98) suggested that NRP is the dominant form of TP in stormwater. The low to negligible correlations between TDP, TPP, and TRP indicated that these phosphorus forms may have varied sources and pathways in the urban environment, contributing independently to the overall phosphorus levels in stormwater (Yang et al., 2018).

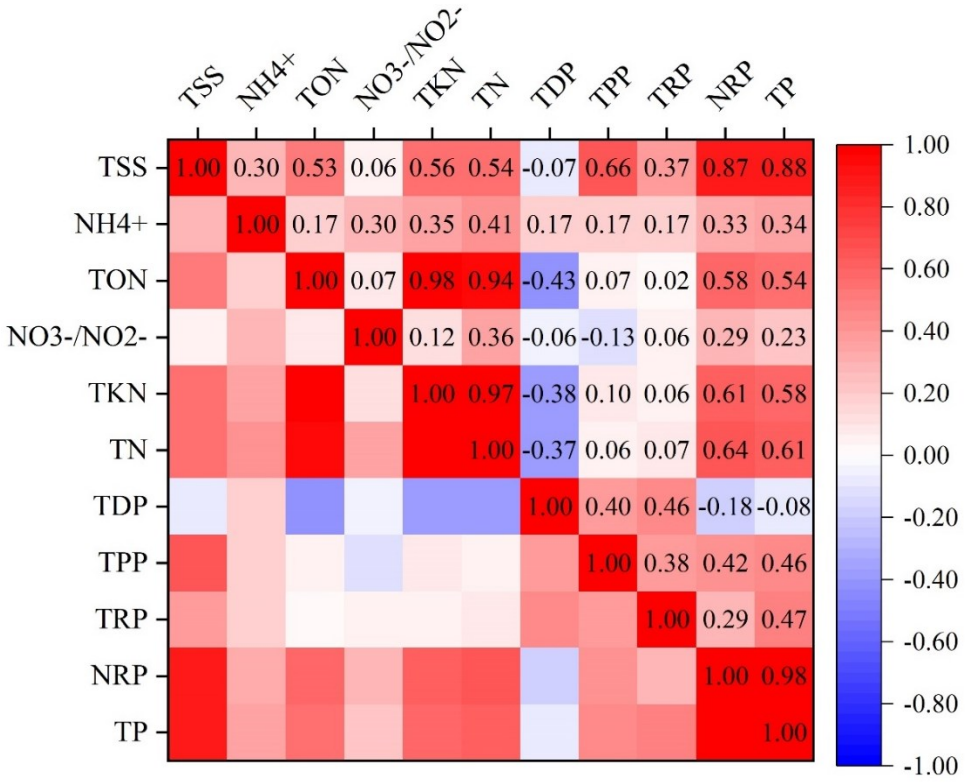


Fig. 4-2. Correlation coefficient matrix heatmap for EMCs of different pollutant components for all sites and years.

Fig. 4-3 presents the correlation matrix between EPLs of TSS and various nitrogen and phosphorus components. EPL of TSS showed a stronger correlation with nitrogen and phosphorus components compared to the EMC. For instance, the correlation between TSS and NH4+ for EPL was 0.78, while it was only 0.30 for EMC. This could be attributed to the relationship between EMC and EPL. The relationship between EMC and EPL can be derived from equation 4-1 and 4-2 as follows:

$$EMC = \frac{M}{V} = \frac{M}{A \cdot P \cdot \alpha} = \frac{EPL}{P \cdot \alpha} \quad (4\text{-}3)$$

where P is total rainfall of the rainfall event, α is rainfall-runoff coefficient of the rainfall event.

Equation 4-3 indicates that EPL is a function of EMC, total rainfall (P), and the rainfall-runoff coefficient (α), which means EPL is influenced by both concentration (EMC) and the volume of runoff ($P \cdot \alpha$). The volume of runoff is influenced by rainfall characteristics. For example, the rainfall-runoff coefficient typically increases with rainfall intensity and duration in pervious areas. Therefore, the relationship between EMC and EPL is determined by the characteristics of rainfall. Higher runoff volumes in events may carry larger quantities of both TSS and other pollutants, leading to a stronger correlation in EPLs compared to EMCs, which only consider concentrations.

Similarly, the correlations between different nitrogen components were generally higher for EPLs. The correlation between NH4+ and TN was 0.83 for EPL compared to 0.41 for EMC. This phenomenon could be explained by the cumulative effect of washed pollutant loads over the event. Even if the mean concentrations do not correlate highly, the total loads might, as they were influenced by the rainfall duration and intensity (Fletcher et al., 2013). The phosphorus components also exhibited higher correlations for EPLs. For example, the correlation between EPL of TRP and TP was 0.86, whereas it was 0.47 for EMC. This indicated that the total phosphorus load was more closely related to the total reactive phosphorus load than the average concentrations were.

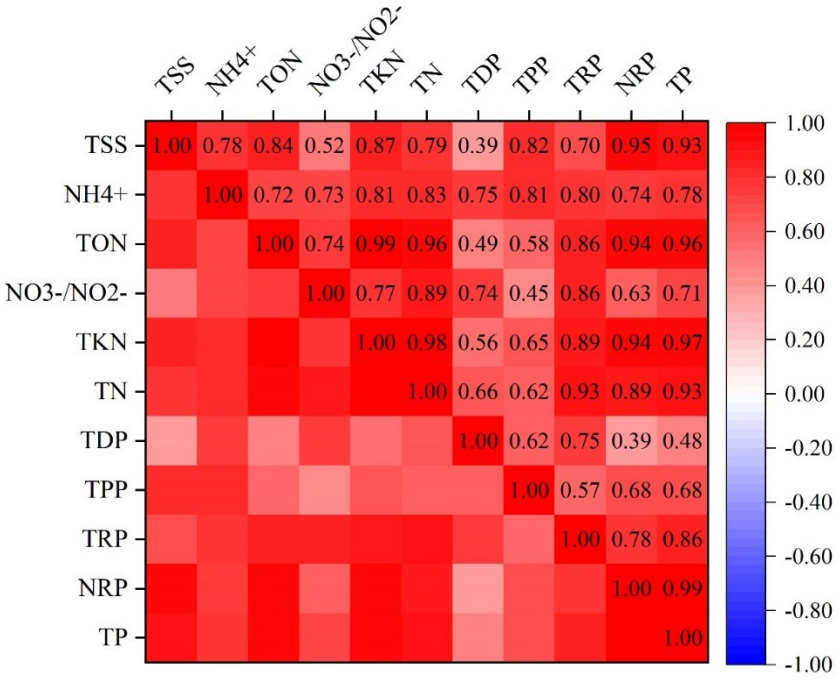


Fig. 4-3. Correlation coefficient matrix heatmap for EPLs of different pollutant components for all sites and years.

4.3.2 EMCs and EPLs from different mixed land uses

The Shapiro–Wilk tests indicated that the raw data were not normally distributed for all study areas and pollutant components. Therefore, a log-transformation was applied to meet the requirement of normality. Table 4-3 and Table 4-4 present the results of the Tukey test for EMCs and EPLs. The catchments were listed both horizontally and vertically. If a pollutant appeared in the intersection of two catchments, it indicated that the pollutant was significantly higher in the catchment listed horizontally compared to the one listed vertically. Take Table 4-3 as an example, NO_2^-/NO_3^- was listed in the cell intersecting RR (horizontal) and RO (vertical), indicating that the EMC for Nitrate and Nitrite was significantly higher in RR compared to RO.

Table 4-3 showed that RR had a higher EMC of Nitrate and Nitrite compared to RO, which was likely related to the largest percentage of landscaped areas in RR (Simpson et al., 2023). A large proportion of nitrogen fertilizer used in landscaped areas (i.e., gardens and lawns) would be nitrite and nitrate, which is known to be a significant source of nitrogen in urban stormwater runoff (Yang

et al., 2018). RO did not exhibit significant higher EMCs for any pollutants compared to the other catchments.

For AB catchment, the EMCs for NRP and TP were significantly higher compared to RR and higher NO_2^-/NO_3^- compared to RO. Given the complex mix of all land uses in AB, including multi-family residential, institutional, and commercial alongside significant transportation and open space, the more diverse land cover characteristics might be contributing to these higher phosphorus levels and higher NO_2^-/NO_3^- concentrations. CR, with the varied land use and the largest presence of gravel lanes, showed higher EMCs for a range of pollutants compared to RR and higher NO_2^-/NO_3^- compared to RO. The mixed land use and land cover characteristics in CR, including the presence of gravel lanes (Thenoux et al., 2002), might be contributing to the elevated levels of various pollutants, reflecting the impact of diverse urban development on stormwater quality (Paule-Mercado et al. 2018).

Table 4-3. Tukey test for EMCs.

	RR	RO	AB	CR
RR		NO_2^-/NO_3^-		
RO				
AB	NRP TP	NO_2^-/NO_3^-		
CR	TSS TON TKN TN TRP NRP TP	NO_2^-/NO_3^-		

In Table 4-4, which examined the significant differences between EPLs from different catchments, CR catchment again stood out with higher EPLs for TSS, NRP, and TP compared to RR. This was consistent with Table 4-3 where CR also showed higher values for these pollutants, reinforcing the observation that the unique land use and land cover characteristics of CR catchment contribute to elevated pollutant levels.

Comparing Table 4-3 and Table 4-4, it was notable that while the EMCs in CR were higher for a broader range of pollutants (TSS, TON, TKN, TN, TRP, NRP, TP) compared to RR, the EPLs were higher specifically for TSS, NRP, and TP. This could be attributed to the fact that EPLs, representing the total amount of pollutants discharged from a specific area during a rainfall event, were influenced not just by the concentration of pollutants (as in EMCs) but also by other factors such as the volume and rate of runoff (Brezonik et al., 2002). The higher EPLs for TSS, NRP, and

TP in CR could indicate that these pollutants were not only present in higher concentrations but were also being discharged in larger volumes during rainfall events.

Table 4-4. Tukey test for EPLs.

	RR	RO	AB	CR
RR				
RO				
AB				
CR	TSS NRP TP			

To further evaluate the difference of EMCs in four catchments, the violin plots presented in Fig. 4-4 illustrate the distribution of EMC for various pollutants for the different catchments. Violin plot can show a box plot and a kernel density plot. The thick black bar in the center represents the interquartile range, the thin black line extending from it represents the 95% confidence interval, and the white dot is the median. On each side of the black line is the kernel density plot that shows the shape of the distribution of the data. Wider portions of the violin plot indicate a higher probability of a given value and narrower portions indicate a lower probability. The violin plots show a trend in median EMC values across different catchments: CR>AB>RO>RR, except for NO_2^-/ NO_3^-, TDP, TPP and TRP. The plots demonstrate a wider distribution and higher median EMC values for CR and AB, indicating a larger range and concentration of pollutants. For example, the median EMC of TSS for CR and AB catchments (195 mg/L) were on average 122% larger than median EMC of TSS for RR and RO catchments (88 mg/L).

The commercial and institutional areas in CR and AB, absent in RR and RO (Fig. 4-1), could contributes to their elevated EMC values compared to RR and RO (McLeod et al., 2006, Bakri et al., 2008). Despite the similarity in composition and proportions of land uses between CR and AB, they exhibited notable differences in spatial distribution (Fig. 4-1). In CR (Fig. 4-1d), commercial, institutional, and multi-family residential areas were more centrally located within the catchment. Conversely, in AB (Fig.4-1e), these areas were situated upstream, farther from the catchment outlet. Furthermore, AB, is double the size of CR, resulting in longer flow pathways. This means pollutants from these land uses in AB took a longer route to the outlet. In the large sewer system of AB, particulate pollutants would tend to settle in pipes during smaller rainfall events, reducing concentrations measured at the pond inlet (Tang et al. 2020). This indicates that the spatial

distribution of land uses can affect generation and transportation of pollutants. Additionally, the upstream pond in the AB catchment treats approximately 69 ha drainage area, which may also reduce the pollutant concentration (Yang et al., 2023).

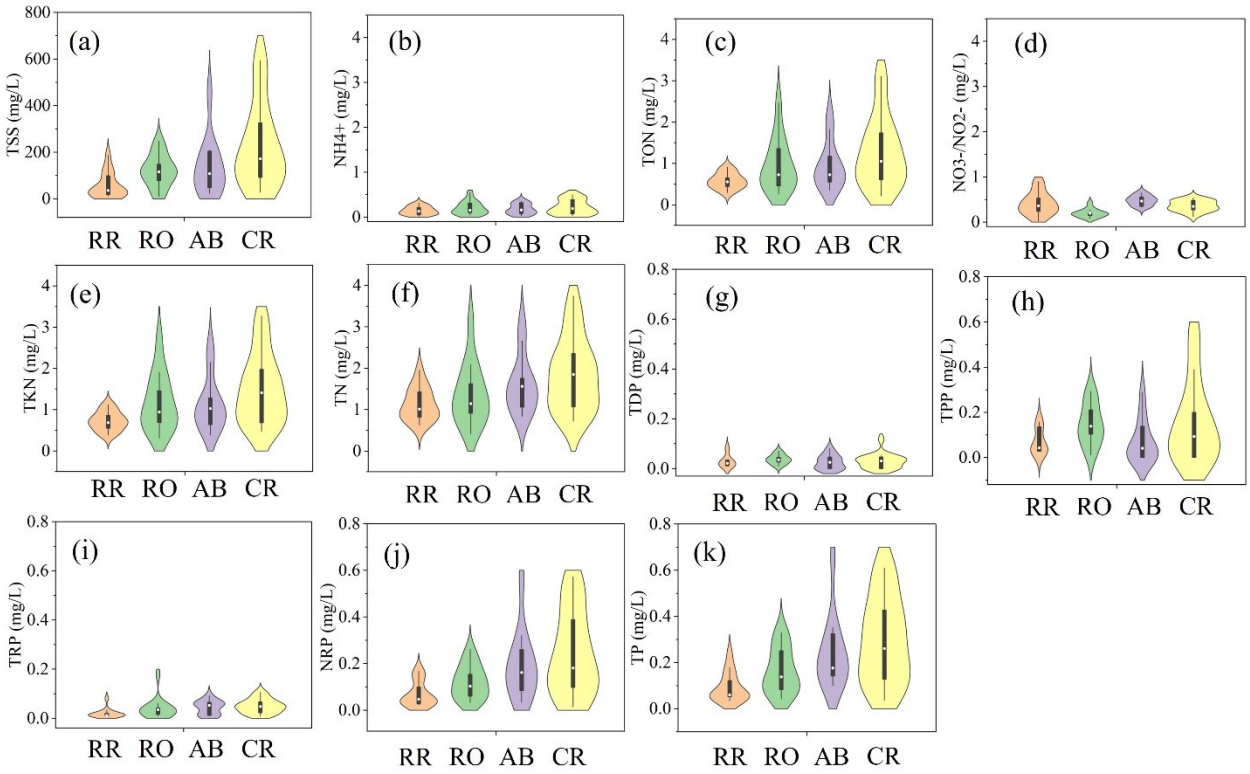


Fig. 4-4. Comparative violin plots of EMCs for (a) total suspended sediment (TSS); (b) ammonia (NH4+); (c) total organic nitrogen (TON); (d) nitrate and nitrite (NO_2^-/ NO_3^-); (e) total kjeldahl nitrogen (KN); (f) total nitrogen (TN); (g) total dissolved phosphorus (TDP); (h) total particulate phosphorus (TPP); (i) total reactive phosphorus (TRP); (j) non-reactive phosphorus (NRP); (k) total phosphorus (TP).

Examining the violin plots for EPLs (Fig. 4-5), a different pattern emerged that is different from the EMC distributions. Except for TSS, NRP and TP, no significant differences at 5% level in median values of EPLs across the study areas were observed. The fact that most of the EPLs across different catchments were statistically similar suggests that variations in land use, area, and rainfall characteristics do not necessarily lead to different pollutant loading levels per unit area. This could imply that varying factors between catchments (such as land use and rainfall conditions) counterbalance each other in terms of their impact on pollutant loads.

Fig. 4-4 shows that there were obvious patterns and non-negligible differences in EMC values between catchments, emphasizing that pollutant concentrations in runoff could be influenced by site-specific factors. This difference between EMC and EPL trends emphasizes the complexity of urban stormwater quality dynamics, where concentrations and loads are not always consistent (Simpson et al., 2023). It emphasized the importance of understanding and addressing both concentration and load in stormwater quality across varied urban landscapes (Cahn et al., 2014).

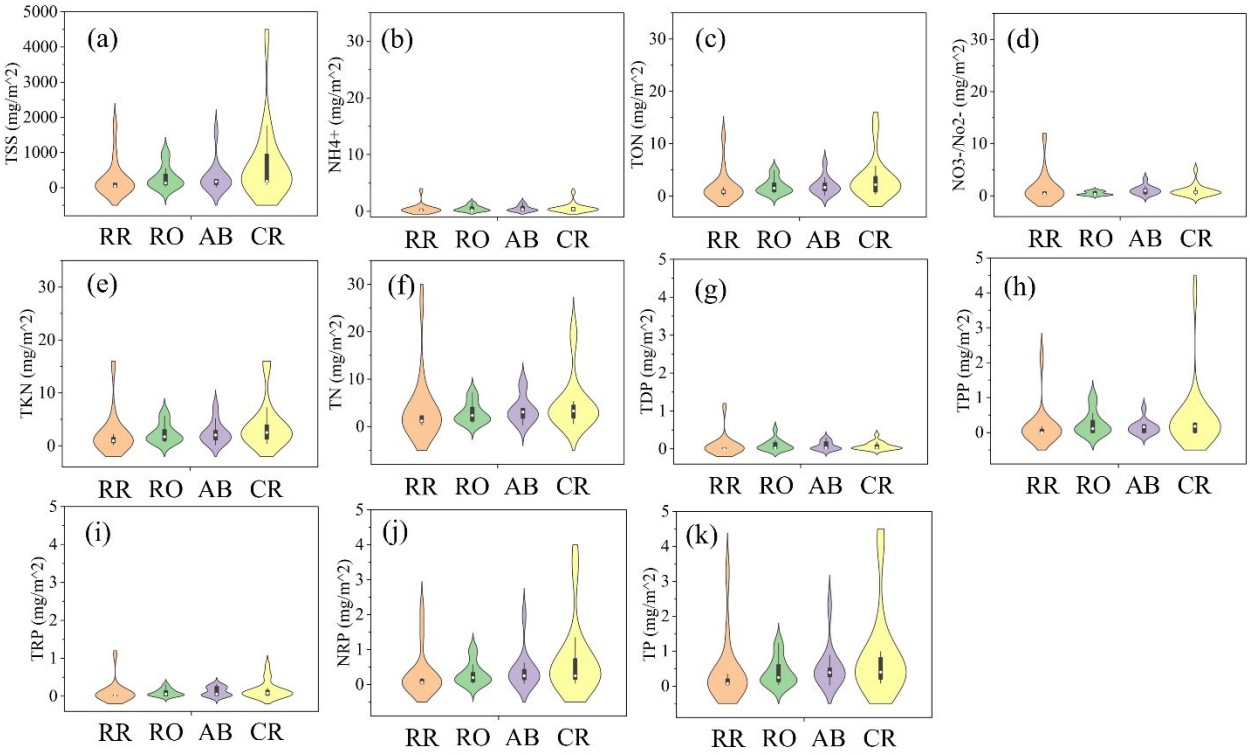


Fig. 4-5. Comparative violin plots of EPLs for (a) total suspended sediment (TSS); (b) ammonia (NH4+); (c) total organic nitrogen (TON); (d) nitrate and nitrite (NO_2^-/ NO_3^-); (e) total kjeldahl nitrogen (KN); (f) total nitrogen (TN); (g) total dissolved phosphorus (TDP); (h) total particulate phosphorus (TPP); (i) total reactive phosphorus (TRP); (j) non-reactive phosphorus (NRP); (k) total phosphorus (TP).

4.3.3 Rainfall event type effects

In this section the extent EMCs and EPLS from the different catchments are a function of rainfall event types is investigated. In Fig. 4-6 (a-f) median EMCs and EPLs for the different rainfall event types are compared. The EMC of TSS (Fig. 4-6a) revealed a pattern across different rainfall event types and catchments. During type I rainfall events, the EMC of TSS increased from RR to CR,

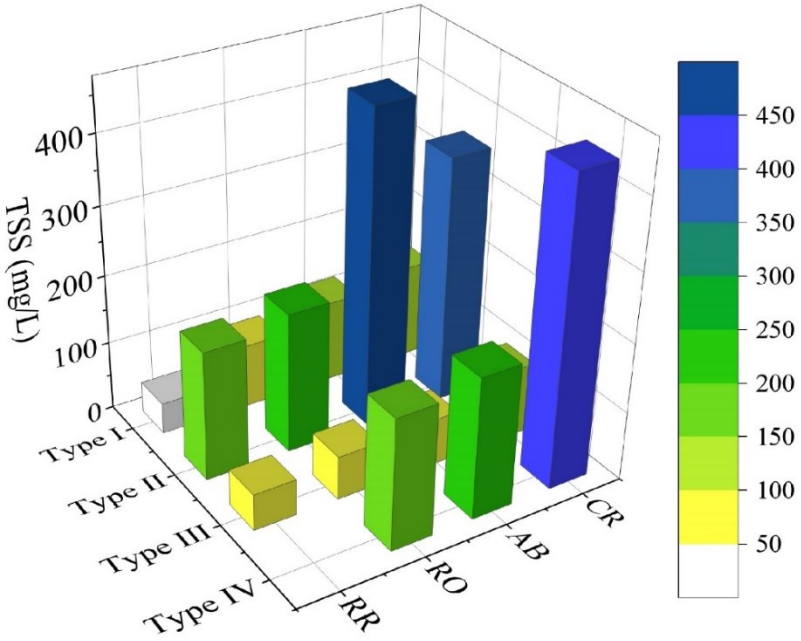
with CR having the highest value (137 mg/L). Type II rainfall events, characterized by high rainfall intensity and short duration, led to elevated TSS concentrations, particularly in the AB (457 mg/L) and CR (358 mg/L) catchments. These elevated values are likely due to the ability of intense rainfall to mobilize and wash off accumulated pollutants rapidly (Liu et al., 2013, Yazdi et al., 2021). The TSS values were generally lower for Type III and then increase for Type IV, especially in CR. The long antecedent dry days in Type IV might led to more pollutant build-up, resulting in higher EMCs of TSS during rainfall events (Murphy et al. 2015, Yan et al., 2023), while long rainfall durations in Type III may cause lower pollutant concentrations (Pitt et al., 2005, Kim et al., 2007, Schiff et al., 2016).

Analyzing the EMC of nitrogen (Fig. 4-6b), most of the nitrogen components generally followed a similar trend to TSS, with CR and AB showing higher values compared to RR and RO. Type II and Type IV rainfall events again had higher EMCs (except for NO_2^-/ NO_3^-), particularly in the AB and CR catchments. These catchments have more complex mixed land uses and this may have contributed to the higher EMCs during these event types and indicating that catchment characteristics have a non-negligible impact on nitrogen concentrations in stormwater. The EMCs of phosphorus components (Fig. 4-6c) followed a similar trend, with Types II and IV rainfall events showing elevated concentrations (except for TDP). This could be attributed to the accumulation of phosphorus over long antecedent dry periods during Type IV rainfall events and wash-off during brief-intense rainfall events like Type II.

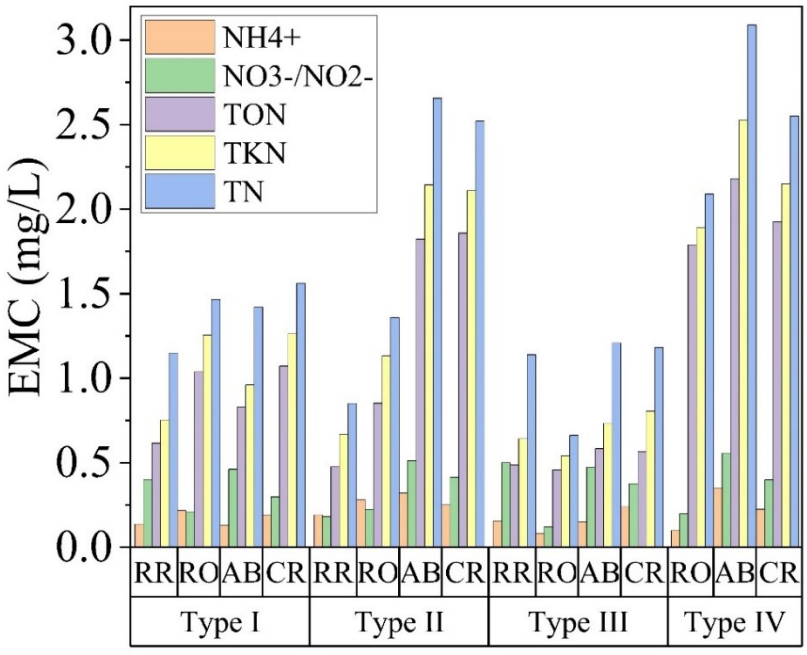
The plot of EPLs for TSS in Fig. 4-6d shows that Type II rainfall events contributed to higher TSS loads, particularly in the CR catchment. These brief-intense rainfall events would tend to cause substantial erosion and particle transport, resulting in elevated TSS loads (Li et al., 2015). Type III rainfall events, despite their low intensity, also contributed to slightly higher TSS loads compared to Type I rainfall events, reflecting the prolonged wash-off period due to their long rainfall duration (Egodawatta et al., 2007, Muthusamy et al., 2018).

For nitrogen (Fig. 4-6e), Type II rainfall events led to highest EPLs in AB and CR. Type III rainfall events even produced higher EPLs in RR compared to Type IV rainfall events for other catchments. This pattern emphasized the role of rainfall duration in transporting more pollutant loads (Jani et al., 2020). The EPLs of phosphorus (Fig. 4-6f) also showed similar pattern. Fig. 4-6 shows differential influence of rainfall event types on EMCs and EPLs. While Type IV rainfall events

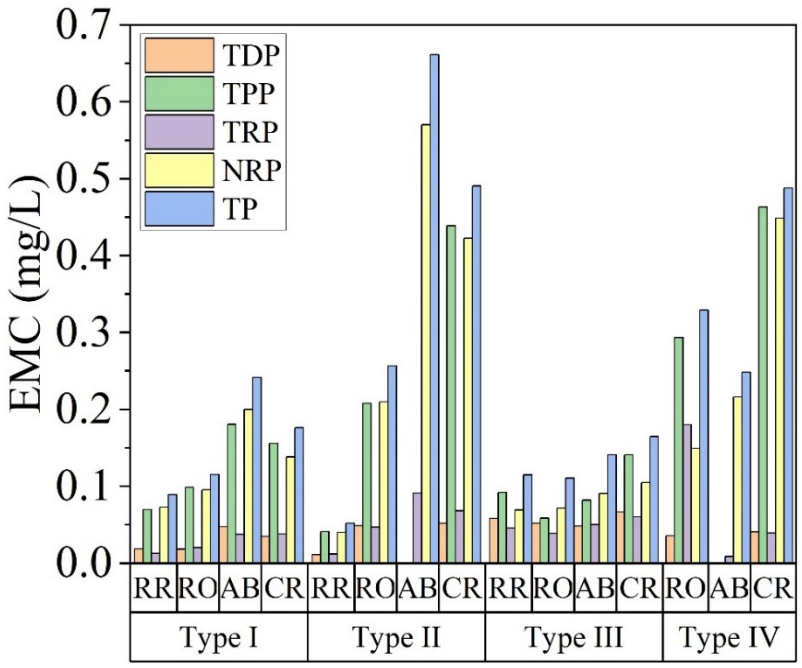
led to higher EMCs due to accumulated pollutants, the longer duration of Type III rainfall events compensates for the lower concentrations by washing off pollutants over a long period, leading to similar or even higher EPLs.



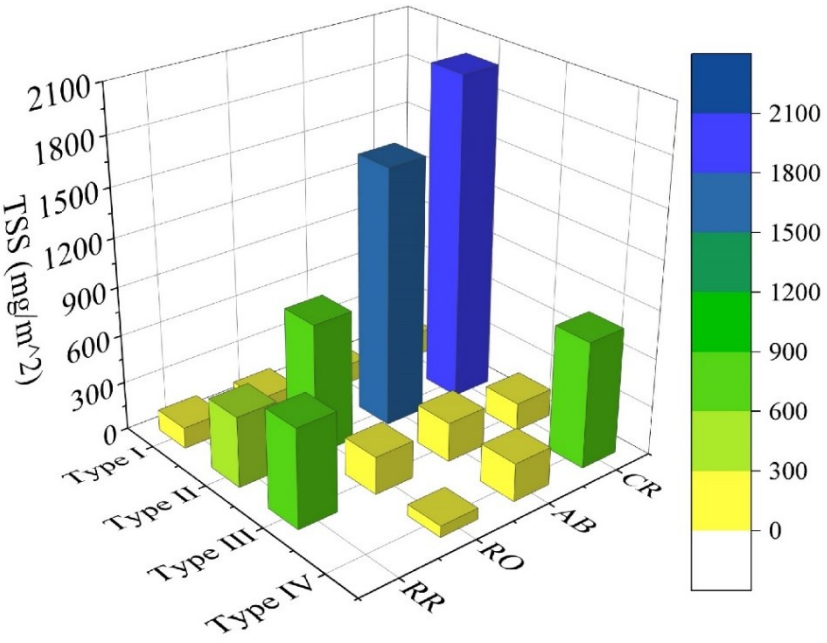
(a)



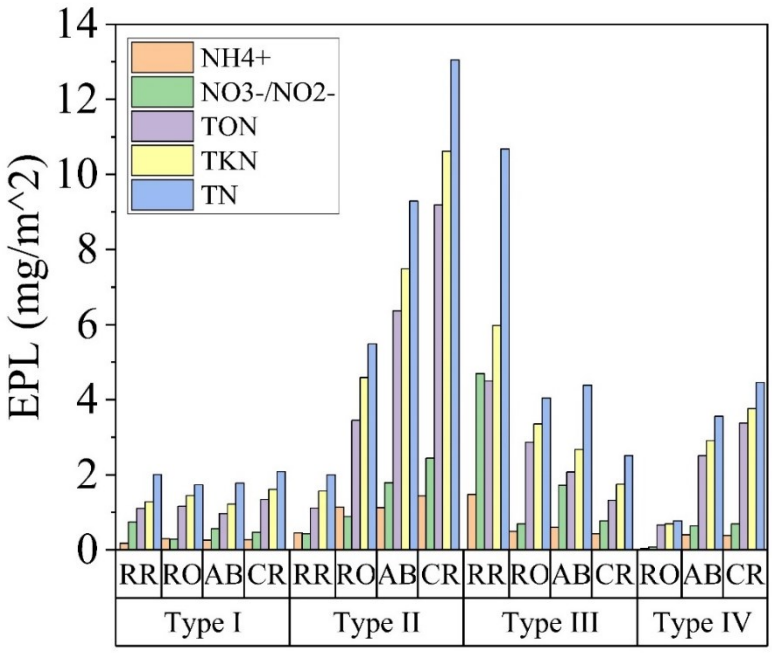
(b)


EMC (mg/L)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
TDP
TPP
TRP
NRP
TP
RR RO AB CR RR RO AB CR RR RO AB CR RO AB CR
Type I
Type II
Type III
Type IV

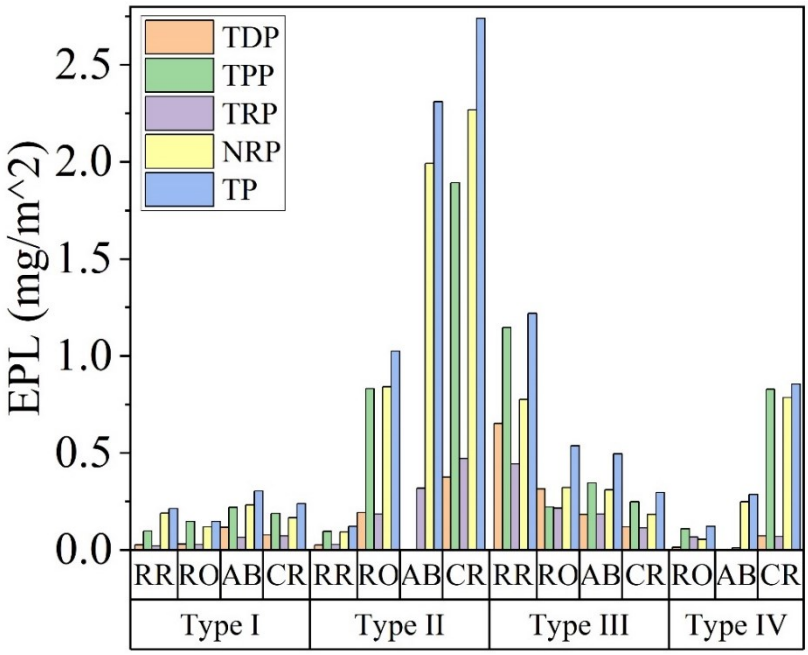
(c)


TSS (mg/m^2)
2100
1800
1500
1200
900
600
300
0
Type I
Type II
Type III
Type IV
RR
RO
AB
CR

(d)



(e)



(f)

Fig. 4-6. Median EMCs of (a) TSS; (b) nitrogen components; (c) phosphorus components and median EPLs of (d) TSS; (e) nitrogen components; (f) phosphorus components under different rainfall event types.

4.3.4 Seasonal effects

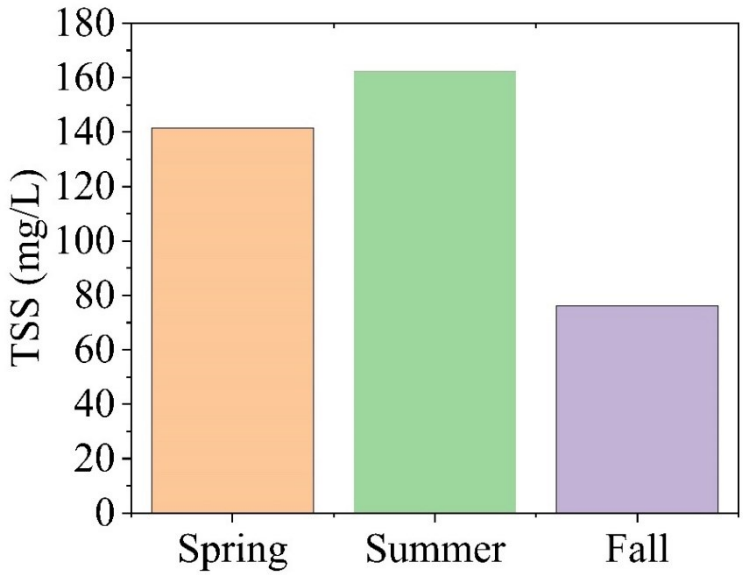
Apart from rainfall event types, seasons could also impact the EMCs and EPLs in runoff. Pollutants were assessed seasonally, by assessing them for three-month intervals that represent spring (March to May), summer (June to August), and fall (September to November) (Valtanen et al., 2014). In Fig. 4-7 the median EMCs and EPLs in the different seasons are presented. For TSS (Fig. 4-7a), the median EMC in fall (76 mg/L) was obviously lower than the spring and summer, as the season was generally characterized by drier conditions and reduced runoff. This observation was consistent with Smith et al. (2020), who found TSS concentrations were higher in the spring and summer than the fall in Ohio, USA. The higher median EMC (142 mg/L) in the spring could be attributed to the influence of rain events washing off pollutants that accumulated in winter season (Yang and Lusk 2018, Chen et al., 2018). The higher median EMC (162 mg/L) in the summer was likely due to frequent storm events that contributed to increased pollutant concentrations.

The EMCs for the nitrogen components are presented in Fig. 4-7b. The EMCs of NH4+ (0.39 mg/L) and NO_2^-/NO_3^- (0.57 mg/L) were highest in spring and then decreased in summer and further decreased in fall (0.17 mg/L). The high spring concentration may be related to lawn fertilizer used in the spring (Yang and Toor 2017). The low value in summer suggests that warmer temperatures in summer might facilitate their uptake by plants (IPCS 1986). TON and TKN peaked in summer, indicating possible organic matter decomposition and the subsequent release of organic nitrogen (Yang et al., 2018). The warm temperatures of summer have been shown to accelerate microbial activities, leading to a higher breakdown of organic matters (Barthélémy et al., 2022). Reflecting the patterns of its components, TN was highest in summer (1.62 mg/L), followed by spring (1.43mg/L) and then fall (1.16 mg/L). The EMCs of the phosphorus components are presented in Fig. 4-7c. TDP (0.08 mg/L) and TRP (0.08 mg/L) were higher in spring, while TPP and NRP were higher in spring and summer, indicating that particulate phosphorus was more prevalent in spring and summer. Following the trend of its components, TP was higher in spring (0.24 mg/L) and summer (0.23 mg/L), and lowest (0.12 mg/L) in fall, which is consistent with the findings of Pitt et al. (2018).

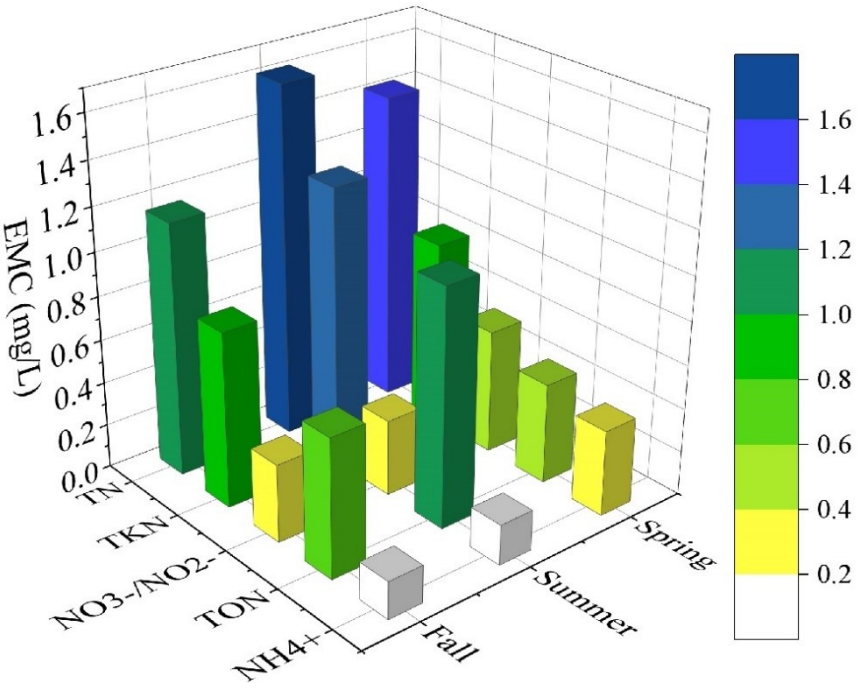
The EPLs of TSS (Fig. 4-7d) in summer, spring and fall were 472, 404 and 194 mg/m2, respectively which is similar to the pattern for EMCs (Fig. 4-7a). The EPLS of all nitrogen components (Fig. 4-7e) were lowest in the fall compared to the summer and spring. The highest

EPL values of NH4+ (1.16 mg/L), NO_2^-/ NO_3^- (2.00 mg/L) and TN (5.09 mg/L) were observed in the Spring, while the highest EPL values of TON (2.70 mg/L) occurred in the summer. EPLs of phosphorus components (Fig. 4-7f) exhibited higher values in spring and summer compared to the fall (except for TDP). The highest EPL values of TDP (0.19 mg/L), TPP (0.48 mg/L) and TRP (0.21 mg/L) were observed in the spring, while the highest EPL values of NRP (0.54 mg/L) were observed in the summer.

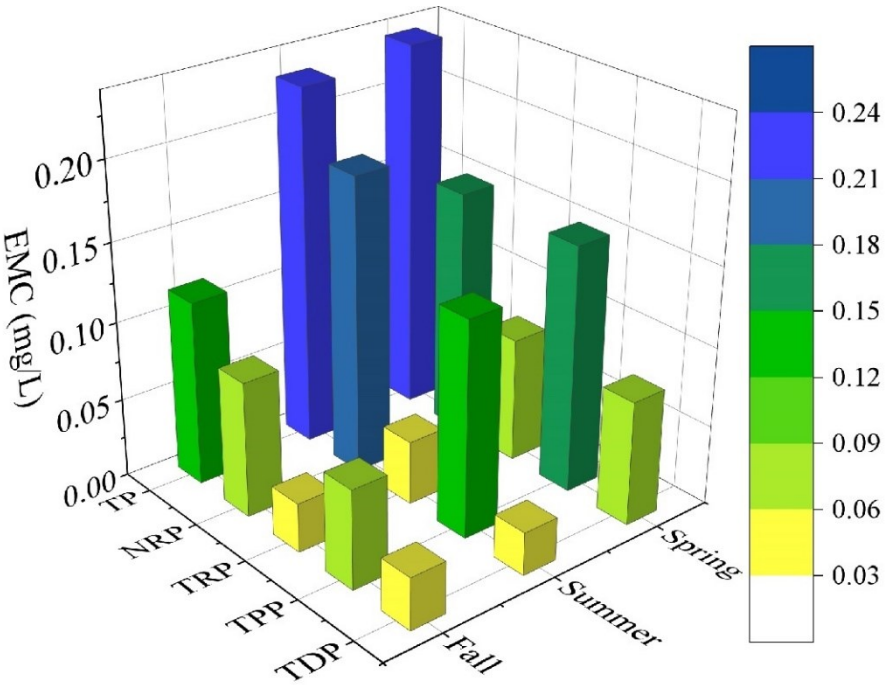
These results clearly indicate that seasonal variations impact both EMCs and EPLs in stormwater runoff. These variations were influenced by factors like storm events, anthropogenic activities and biogeochemical processes. A decreasing pattern summer-spring-fall was observed for EMC and EPL of TSS. For EMCs, a decreasing pattern spring-summer-fall occurred in 2 of 5 for Nitrogen components, and 4 of 5 for Phos. For EPLS, the decreasing pattern spring-summer-fall occurred in 3 of 5 for both nitrogen and phosphorus. EMC and EPL of TSS was lowest in the fall and EPLs in the fall were lowest all the time for both nitrogen and phosphorus. For EMCs, 3 out 5 nitrogen components and 4 out of 5 phosphorus components were lowest in the fall.



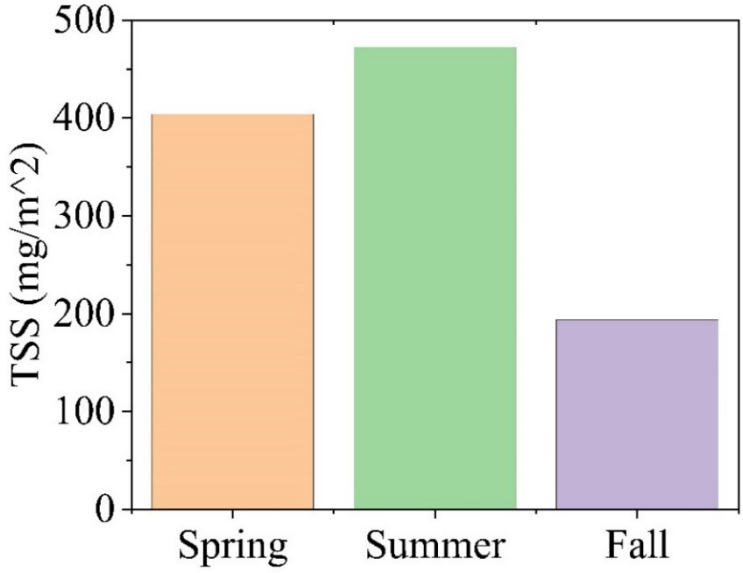
(a)


EMC (mg/L)
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
TN
TKN
NO3-/NO2-
TON
NH4+
Fall
Summer
Spring
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2

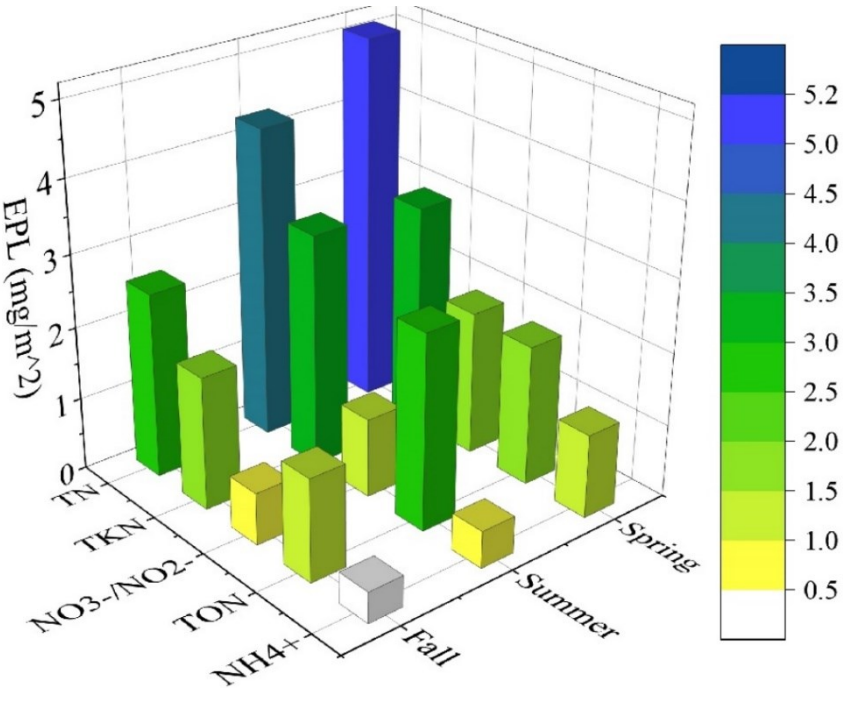
(b)


EMC (mg/L)
0.20
0.15
0.10
0.05
0.00
TP
NRP
TRP
TPP
TDP
Fall
Summer
Spring
0.24
0.21
0.18
0.15
0.12
0.09
0.06
0.03

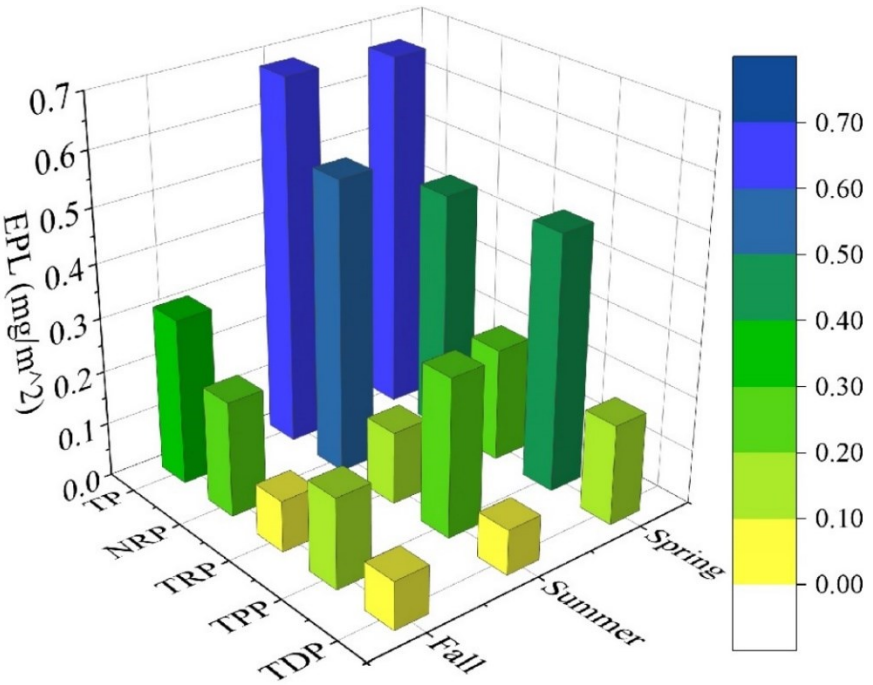
(c)


500
400
300
200
100
0
TSS (mg/m^2)
Spring
Summer
Fall

(d)


5
4
3
2
1
0
EPL (mg/m^2)
TN
TKN
NO3-/NO2-
TON
NH4+
Fall
Summer
Spring
5.2
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5

(e)



(f)

Fig. 4-7. Median EMCs of (a) TSS; (b) nitrogen components; (c) phosphorus components and median EPLs of (d) TSS; (e) nitrogen components; (f) phosphorus components in different seasons.

4.4 Conclusions

The joint impact of mixed land use and rainfall event types on urban stormwater quality was explored. Water quality data were collected and analyzed from four catchments with mixed land use types in Calgary, Alberta, Canada. The main findings and conclusions are as follows:

(1) The correlation analysis between various pollutant components in urban runoff revealed that the EMC of TSS was highly correlated with Non-Reactive Phosphorus (0.87) and Total Phosphorus (0.88), and moderately with Total Organic Nitrogen (0.53), Total Kjeldahl Nitrogen (0.56), and Total Nitrogen (0.54), suggesting these phosphorus and nitrogen forms may exist primarily in particulate form. Moreover, EPLs generally exhibited higher correlation coefficients compared to EMCs.

(2) Differences in EMCs across the four study catchments is evidence that mixed land use could influence the generation of stormwater pollutants. The presence of multi-family residential,

commercial, and institutional areas tended to elevate EMCs compared to single residential areas. The spatial configurations of mixed land uses could affect EMCs.

(3) Type II rainfall events, characterized by high intensity and short duration, tended to elevate TSS and nutrient concentrations, especially in complex land-use catchments. Type IV rainfall events, notable for long antecedent dry days, also exhibited increased pollutant concentrations, while Type III rainfall events, despite lower intensities, could result in higher EPLs due to their longer durations.

(4) Seasonal variations markedly influence the EMCs and EPLs of pollutants in runoff, as demonstrated by patterns in TSS and nutrient components. Spring typically exhibited elevated EMCs and EPLs due to pollutants accumulated in winter and fertilizer usage in spring, while summer saw varied impacts on different pollutants, often due to storm events and biogeochemical processes like organic matter decomposition and nutrient uptake by plants. Fall generally witnessed a decline in pollutant concentrations and loads, attributed to drier conditions and reduced runoff.

This research provided insights into understanding the complexities introduced by mixed land use and rainfall event types in managing urban stormwater quality, thereby contributing to the advancement of sustainable and resilient urban environments.

5. Particle size distribution of suspended sediments in urban stormwater runoff: effect of land uses and rainfall conditions

5.1 Introduction

Road deposited sediment (RDS), also known as surface particles or road dust, is the sediment that accumulates on paved surfaces such as roads and parking lots (Wang et al. 2020). RDS on the urban surface is a non-negligible stormwater pollution source and the primary carrier of other particulate pollutants such as particulate nutrients and heavy metals constituents (Zhao and Li 2013, Shen et al. 2016). Large amounts of various pollutants can be adsorbed by RDS and transported by urban runoff to receiving water bodies causing water quality deterioration during rainfall and snowmelt events (Yang and Lusk 2018). The particle size distribution (PSD) of suspended sediments in urban runoff has been considered an essential factor affecting the transport and fate of sediment and the associated pollutants (Kim and Sansalone 2008, Zhao et al. 2010). Coarser particles could be temporarily or permanently deposited in long pipes of drainage systems during small rainfall events (Chow et al. 2011). Many studies have found that finer particles have affinity for various and large amounts of pollutants (Vaze and Chiew 2004, Wijesiri et al. 2016, Wang et al. 2020). In addition, PSDs of suspended sediments in urban runoff are also necessary for the effective treatment of urban stormwater (Selbig et al. 2016). For example, stormwater wetlands remove coarse materials first, then finer particulates, and finally dissolved components (Kadlec 1999). Hence, characterizing PSD of suspended sediments in urban runoff is critical to understanding pollutant fate as well as designing efficient stormwater treatment measures.

Land use types can considerably impact PSD (Bian and Zhu 2009). For example, eroded soils from riverside parks are easily trapped by the ground, so riverside park soil can provide more fine particles than commercial areas (Bian and Zhu 2009). Road construction materials and industrial inputs can produce sediments with different particle sizes (Taiwo et al. 2014). Furthermore, winter road sand, street cleaning and traffic conditions can change PSD (Amato et al. 2010, Buonanno et al. 2011). Some studies have investigated PSD from different land use types. A study reported that the mean particle size of three impervious surface types in Beijing decreased in the following order: roads in residential areas > roofs > main traffic roads (Shen et al. 2016). Characterizing PSD from different land use types is helpful for watershed managers to design efficient stormwater treatment

from different land uses. However, collection and analysis of water samples are often time-consuming and expensive (Yan et al. 2022), so few long-term studies conducted research on multiple sites and land uses. Long-term studies can provide more reliable and comprehensive data that can be used for statistical analysis.

PSD of suspended sediments in runoff can also be impacted by rainfall characteristics. Rainfall characteristics has been shown to influence PSD during sediment build-up and wash-off processes (Wijesiri et al. 2016). During antecedent dry days, sediments of different sizes accumulate on urban surfaces, and the distribution will change with time, influenced by vehicle-induced turbulence and wind effect (Sabin et al. 2006, Gunawardena et al. 2013). It was found that finer particles were more easily mobilized by runoff than coarser particles (Miguntanna et al. 2013, Wijesiri et al. 2015). High rainfall intensities can wash off sediments with a broader particle size range, particularly for the transport limited case (Egodawatta et al. 2007). However, there are knowledge gaps in the study of the impacts of rainfall event types on PSD in runoff.

Pollutants can be accumulated in snow in cold regions, then released during snowmelt runoff and enter receiving water bodies (Chen et al. 2018). In early spring, the ecosystems of receiving water bodies are fragile and have poor self-purification capacities (Zhu et al. 2012). Hence, pollutants in snowmelt runoff could seriously threaten the receiving water body, e.g., in northern Sweden, pollutant concentrations and loads were significantly higher during snowmelt events compared to rainfall events (Westerlund and Viklander 2006). Westerlund and Viklander (2006) reported that the particle concentrations of suspended solids in road snowmelt runoff decrease as the particle size increases. Cristina et al. (2002) collected samples from the entire cross section of the snowbanks from 10 highway shoulder sites immediately before the snowbanks dissipated through melting and a subsequent rainfall runoff event that followed the snow dissipation. They reported that all sites had similar PSD and more than 50% of particles were larger than 1225 μm. These studies only collected samples from a single land use type during limited snowmelt events to analyze the PSD. Few researchers conducted field studies in multiple land uses during long periods and compared PSDs between snowmelt and rainfall events.

This study aims to examine the effects of land use and precipitation conditions on PSD in stormwater and snowmelt runoff. This research collected water samples in 15 urban areas in Calgary, Alberta, Canada, during 179 rainfall and 118 snowmelt events from 2016 to 2021 to

analyze PSD of suspended sediments in runoff. Specifically, there are four objectives of this study: (1) to characterize PSD of suspended sediments in runoff from different land uses; (2) to compare the results with previous literature to provide more comprehensive information about PSD from different land uses. (3) to study the effects of rainfall event types and seasons on PSD of suspended sediments in runoff during ice-free seasons; (4) to compare PSDs between snowmelt and rainfall events.

5.2 Materials and methodology

5.2.1 Description of the study sites

The study sites include 15 catchments located in Calgary, Alberta, Canada (Fig. 5-1). With a population exceeding one million people, Calgary is in a semi-arid, continental climate region with cold winters and mild to warm summers. Warm, dry Chinook winds routinely blow into the city during winter from the Rocky Mountains and can raise the temperature by 30 ℃ in a few hours, which can cause rapid snowmelt. The average annual precipitation of Calgary is 418 mm, with 22% of that being snow (Government of Canada).

The study sites have six land use types: gravel lane residential (GLR), paved residential (PR), commercial (CO), roads (RO), industrial (IN), and mixed land use (MLU) (Table 5-1). It should be noted that there will be roads in some of the other larger catchments as well. In this study, roads specifically refer to catchments consisting of only roads. Each land use had at least two sampling sites. GLR (Millrise and Mt. Brewster) have gravel lanes, while PR (Douglas Park and McKenzie Lake) only have paved lanes. RO include a highway and an arterial street. CO mainly consist of supermarkets, shopping centers, and parking lots without green space and gravel lanes. MLU consisted of residential areas with gravel lanes and commercial areas. IN are light industry and there is no heavy industry in Calgary. The study site areas range from 8.4 ha to 2619.4 ha, where RO and CO have smaller average areas (15.9 ha and 72 ha, respectively), and IN and MLU have larger average areas (422 ha and 1241 ha, respectively). Study sites with the same land use type are located relatively close to each other, which make them similar in terms of climate conditions.

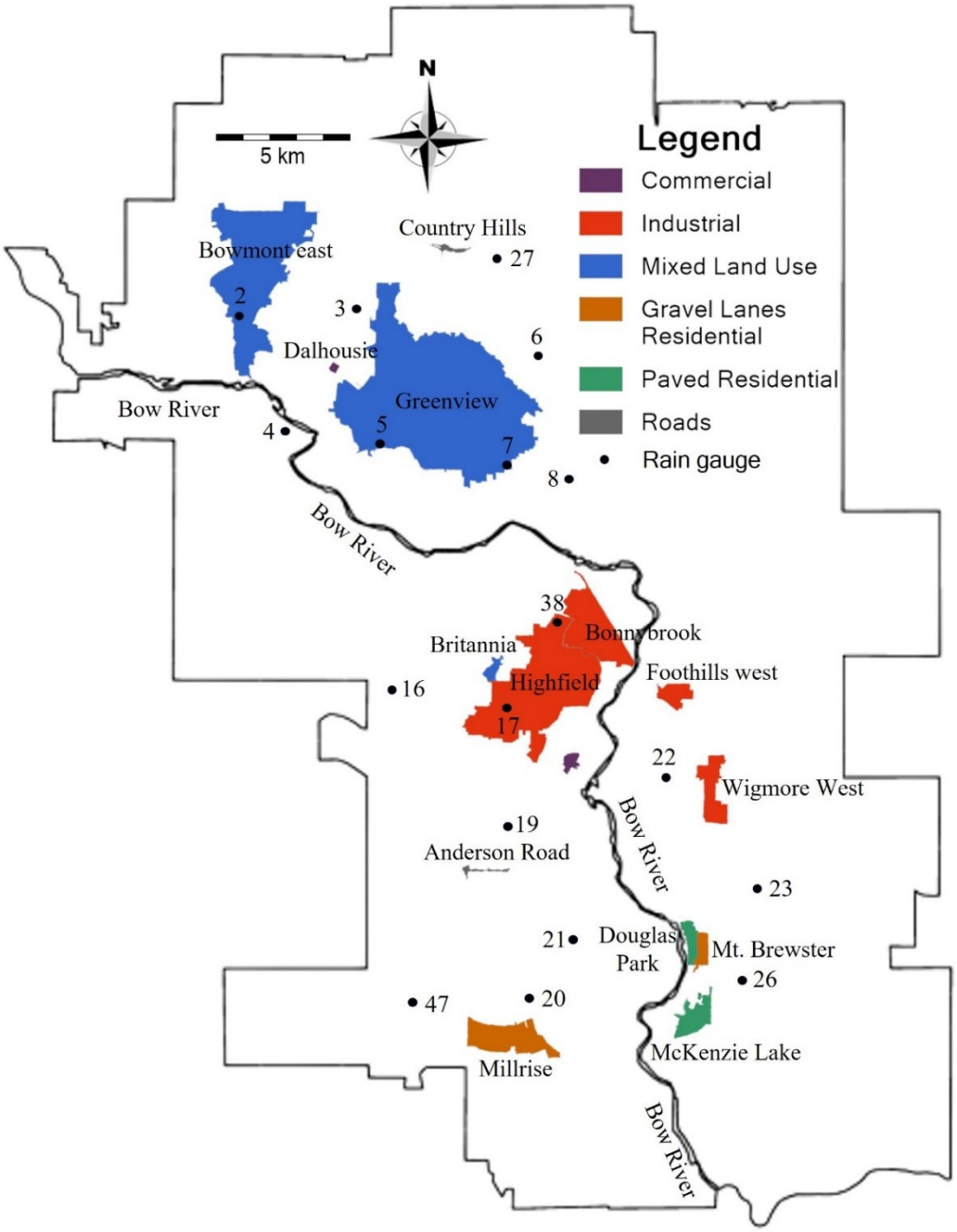


Fig. 5-1. Map of study areas and rain gauge sites in Calgary, Alberta, Canada: The numbers are the numbering of rain gauges.

5.2.2 Sample collection

Water samples were collected during and after rainfall and snowmelt events using an autosampler from a manhole in each study site. Autosamplers have be widely and effectively used to collected

rainfall-runoff samples to analyze suspended particulate matters (Furumai et al. 2002, Selbig et al. 2015, Winston and Hunt 2017). Therefore, the bedload typically was excluded when using autosamplers. The energy needed to move sediment through a pipe with a higher flow rate can change rapidly. This can result in sediment that was previously transported as bedload being transported as suspended load (Selbig and Bannerman 2011). In this study, since the sample intake orifices were fixed near the pipe floor, the particulate matter in water samples consisted mainly of suspended sediment but also included some particles larger than 500 μm. A flow-paced program was used to collect composite samples representative of the events. Composite samples were constructed of multiple individual samples (aliquots) taken over a specific volume of flow for PSD analysis. Given the limitations of the maximum collection capability of autosamplers and the desire to maximize the volume of water quality samples during a runoff event, different flow pacing programs were used for different events.

Table 5-1 summarizes the information on data collection. Samples were taken during 2016 - 2021 by Water Quality Services of the City of Calgary and Advisian (A Worley Group Company). Rainfall-runoff samples were collected at all sites, with the number of events ranging from 1 to 32, with a total of 179 composite samples. Snowmelt samples were taken at six sites using autosamplers, totaling 118 composite samples. On average, the collected samples covered 63% to 93% of the runoff hydrograph duration in the individual rainfall events at each site, which indicates that the composite samples were representative of the rainfall event.

Data from City of Calgary rain gauges (Fig. 5-1) located near the study sites were used to calculate rainfall characteristics, including average rainfall intensity, maximum rainfall intensity, total rainfall, rainfall duration, and antecedent dry days. To avoid errors using a single rain gauge, data from several rain gauges near the study site was used to calculate the average rainfall parameter values. For example, Rain Gauges 14, 15, and 18 are close to Douglas Park, so their data were used to calculate the average rainfall parameter values for Douglas Park. The rain gauge locations are plotted in Fig. 5-1 and the gauges used for each site are listed in Table 5-1.

Table 5-1. Summary of catchment characteristics and data collection.

Land use type	Site	Numbering of rain gauge	Catchment area (ha)	Sampling years	Number of samples		Average coverage of event (%)
					Rainfall	Snowmelt	
Paved Residential (PR)	Douglas Park	14, 15, 18	57.4	2018-21	32	24	90%
	McKenzie Lake	14, 15, 18	147.8	2018-21	31	19	88%
Gravel lane residential (GLR)	Millrise	16, 17	35.9	2018-21	22	21	92%
	Mt. Brewster	15, 18	48	2018-21	26	23	92%
Commercial (CO)	Heritage Meadows	12, 13	28.7	2018-20	25	25	93%
	Dalhousie	1, 5, 6	8.4	2017-21	6	/	74%
Roads (RO)	Country Hills	2, 3, 4	16.5	2017-21	4	/	89%
	Anderson Road	13	15.3	2017-21	3	6	/
Mixed land use (MLU)	Bowmont East	1, 2	1073.3	2019-21	11	/	/
	Greenview	4, 7, 8	2619.4	2016-18	5	/	75%
	Britannia	9, 10, 11	30.5	2020-21	6	/	90%
Industrial (IN)	Highfield	9, 11	1084.2	2016-17	3	/	63%
	Foothills west	12	78.1	2017	2	/	93%
	Bonnybrook	9, 11, 12	375.5	2017	2	/	89%
	Wigmore west	12	151.9	2020	1	/	/

Note: Average coverage of event (%) refers to the percent of the collected runoff to the runoff hydrograph. For example, 90% means that on average, the collected samples covered 90% of the runoff hydrograph duration for the entire sampling period.

5.2.3 Particle size analysis

The laboratory analytical method followed an approach developed by the U.S. Geological Survey (Selbig and Bannerman 2011). Approximately 500 g of water from each composite water sample was analyzed for PSD analysis. First, particles larger than 32 µm were wet sieved into five separate particle-size fractions through a series of stacked nylon-mesh sieves: ⩾ 500, 250-500, 125-250, 63-25, and 32-63 µm. Second, particles in each sieve were dried overnight and then weighted. Materials less than 32 µm were quantified using a Coulter counter into five separate particle-size fractions: 14-32, 8-14, 5-8, 2-5, and < 2 µm (Li et al. 2005, Anta et al. 2006, Westerlund and Viklander 2006). Finally, the mass of each particle-size fraction was measured and used to develop a PSD curve and the median particle size (D_{50}) was calculated for each sample.

5.2.4 Statistical analysis

Statistical tests were performed to assess whether PSDs show significant differences between land uses. Rainfall water samples were collected at all sites while snowmelt samples were only collected at six sites (see table 5-1), so the D_{50} of the rainfall-runoff was selected as the foundation for statistical analysis.

The normality of the D_{50} datasets were assessed using the Shapiro–Wilk (S-W) and Kolmogorov–Smirnov (K-S) test methods first before further statistical analysis. The S-W test was used if the sample size was less than 50 and the K-S when the sample size exceeded 50 (Das and Imon 2016). The null hypothesis of these two tests was that the D_{50} of each land use follows a normal distribution. The hypothesis was rejected if the p-value was less than the indicated significance level of $\alpha = 0.05$, and then the D_{50} data was log-transformed and tested again for its normality (i.e., tested to see if it was log-normal). An analysis of variance (ANOVA) using a significance level of $\alpha=0.05$ was performed to test whether there were significant differences between the means of D_{50} from the different land uses. When the p-value is less than $\alpha = 0.05$, it indicates that at least one land use differs from others.

Then multiple comparison tests were conducted to determine which land use differs from others. The Tukey honestly significant difference (HSD) test is an easy and frequently used multiple comparison technique that can not only indicate whether one group is significantly different from other groups but also assign one group into one or more subsets based on homogeneity of variance

(Abdi and Williams 2010). In this study, the results are presented with letters representing subsets. Each land use type could own one letter representing one subset. If land uses share a letter, it means differences between them were not statistically significant.

5.2.5 Rainfall event classification

To study the effects of rainfall event type on PSD, the rainfall events should be effectively clustered into different types representing the rainfall characteristics in the study areas. In our previous study (Yan et al. 2023), number of antecedent dry days (ADD), average rainfall intensity (AgI), and rainfall duration (RD) were found to be the most critical rainfall characteristics affecting the event mean concentrations (EMCs) of total suspended solids. Therefore, these three rainfall indices were selected in this study as the basis for dividing rainfall events into different clusters using the K-means clustering algorithm. This algorithm is a widely used unsupervised learning method to identify groups in a dataset (Choi and Beven 2007). K-means clustering method includes four key steps: (1) Select K random data points from the dataset as the initial clustering group centers; (2) Calculate the distance of each data point from all clustering group centers and assign it to its nearest clustering group; (3) Recalculate the new clustering group centers of all data points assigned to the clustering groups and; (4) Repeat the second and third steps until a partition is found in which data points in each clustering group are as close to each other as possible and as far away as possible from data points in other clustering groups. An elbow method was used to determine the best cluster number based on the sum of squared errors (SSE) (Yuan and Yang 2019).

5.3 Results and discussion

5.3.1 Statistical analysis results

The normality test for PR was performed by K-S test because the sample sizes is more than 50, and the S-W test was conducted for other land use types. The results in Table 5-2 show that the raw data were not normally distributed for gravel lane residential. Therefore, a log-transformation was applied for all land uses. The normal distribution tests show that the log-transformed data are normally distributed for all land uses (p-value > 0.05). Then the ANOVA tests were performed for these land uses, and the results show that the p-value (<0.001) is less than the indicated significance level of $\alpha=0.05$, which means that there was a significant difference between different sites for the mean D_{50} (marked as $< D_{50} >$ below).

Table 5-2. P-values from normality test results of raw D50 and log transformed D50 in land uses.

Land use	Raw D_{50}		Log D_{50}	
	S-W	K-S	S-W	K-S
PR	/	0.119	/	0.2
GLR	0.017	/	0.094	/
CO	0.655	/	0.684	/
RO	0.068	/	0.241	/
MLU	0.064	/	0.892	/
IN	0.479	/	0.606	/

The Tukey HSD test (Fig. 5-2) classified <D_{50}> into three subsets labeled A, B, and C, with subset relationships as <D_{50}> of A>B>C. The land uses RO and IN share the same subset C, indicating that their <D_{50}> values are not significantly different. On the other hand, PR and CO belong to subsets A and AB, respectively, which are different from RO and IN. Therefore, the <D_{50}> of PR and CO is significantly higher compared to RO and IN. The land use GLR falls into the subset ABC, which intersects with all other land uses, indicating no significant difference in <D_{50}> between GLR and the others. MLU belongs to subset BC, with PR being the only land use that does not intersect, making the <D_{50}> of MLU significantly lower than that of PR. In summary, the <D50> decreases in the following order: PR>CO> GLR>MLU>IN, RO.

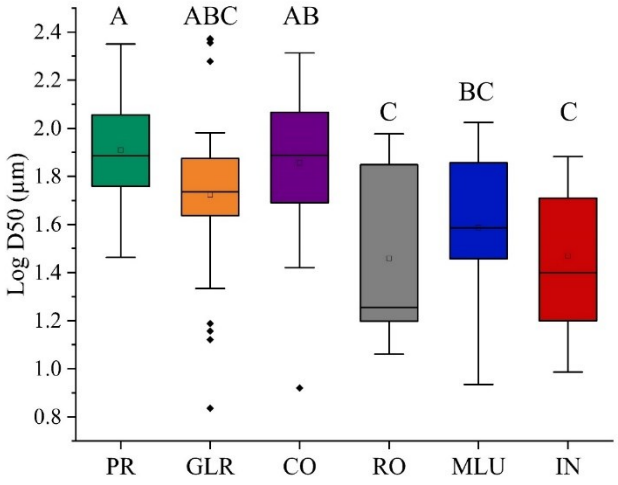


Fig. 5-2. Tukey HSD test results for land uses.

5.3.2 Rainfall event classification

A data matrix (179 × 3) consisting of 179 rainfall events and three rainfall indices (i.e., ADD, AgI, and RD) was used to classify the rainfall events. Fig. 5-3 shows the variation of SSE with the number of clusters. The SSE decreases with an increase in cluster number, with the magnitude of the decline slowing down after 4 clusters, indicating that the classification is not considerably improved by increasing the cluster number beyond 4 clusters. Therefore, the optimal number of clusters was determined to be 4. Fig. 5-4 displays how close each data point in one cluster is to the points in the neighboring clusters. The silhouette value of each object is a measure of how similar that object is to other objects in the same cluster. The silhouette value ranges from -1 to 1. A value of 1 indicates the data point is tightly clustered within its assigned cluster and well-separated from other clusters. A value near 0 denote overlapping clusters, and the worst value is -1. The average silhouette values are 0.8, 0.67, 0.52, and 0.83 for four Clusters, respectively. This indicates that the rainfall events can be successfully divided into four types. The four clusters were used to define rainfall event types, representing Type I, II, III, and IV, respectively.

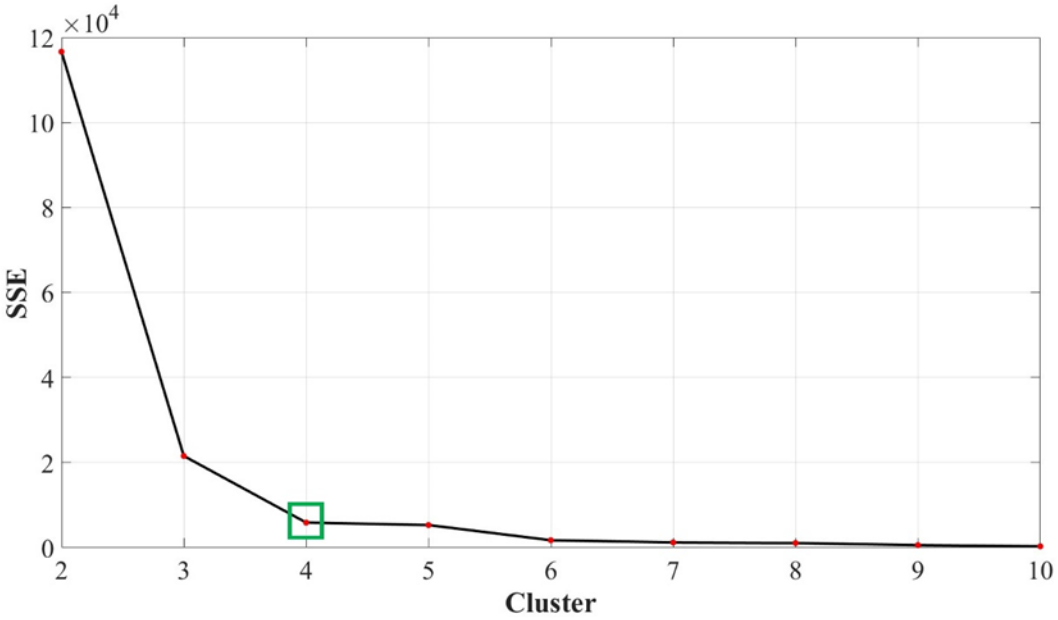


Fig. 5-3. Variation of the SSE index with the number of clusters.

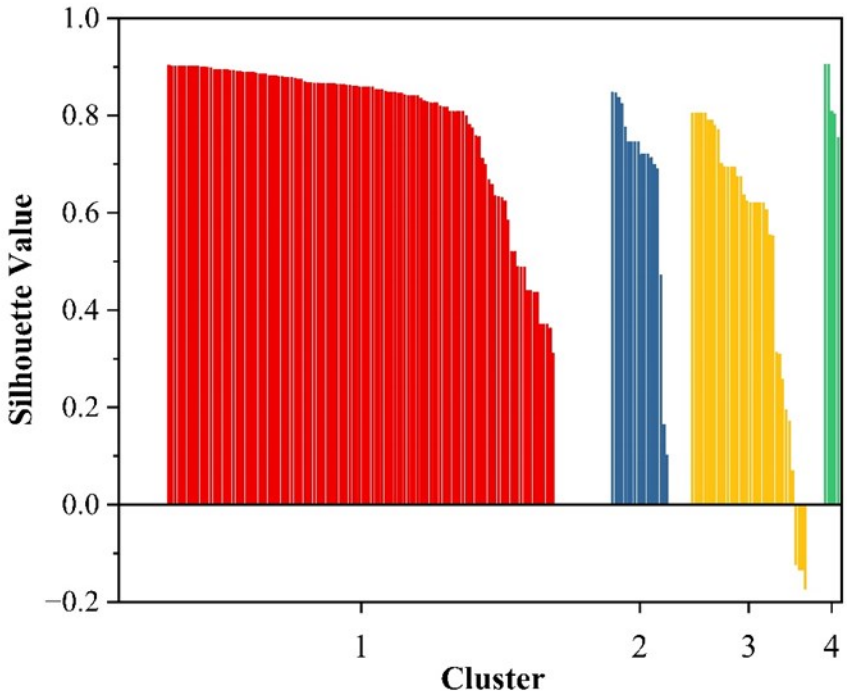


Fig. 5-4. Variation of the silhouette value with the number of clusters.

The mean value and range of rainfall parameters of the four rainfall event types are summarized in Table 5-3. The characteristics of the four types are as follows. 1) Type I rainfall events have a short rainfall duration (4.8 h), a low average rainfall intensity (2.5 mm/h), and a small number of antecedent dry days (4.3 days). Type I rainfall events represent small and frequent rainfall events in Calgary, amounting to 120 out of the 179 events. 2) Type II rainfall events have a short rainfall duration and a small number of antecedent dry days, while they have the highest average rainfall intensity (36.2 mm/h). These events have a high rainfall intensity with relatively short duration, consistent with the characteristics of typical convective rainfall events (Thyregod et al. 1999). Type II rainfall events amounted to 18 out of the 179 events. 3) Type III rainfall events have a small average rainfall intensity and a small number of antecedent dry days, but the rainfall duration is much longer (21.3 h). These events are relatively continuous and uniform in intensity, consistent with typical stratiform rainfall events (Löwe et al. 2016). Type III rainfall events amounted to 36 out of the 179 events. 4) Type IV rainfall events have the largest number of antecedent dry days (37.8 days) and a long rainfall duration (20.8 h). Type IV rainfall events amounted to 5 out of the 179 events.

Table 5-3. Summary of the four rainfall event types.

Rainfall event types		RD (h)	AgI (mm/h)	ADD (days)	Total Rainfall (mm)	Event number
Type I	Average	4.8	2.3	4.3	11.0	120
	Range	0.2-13.0	0.02-14.9	0.2-16.0		
Type II	Average	0.3	36.2	4.5	10.9	18
	Range	0.2-0.6	21.6-70.5	1.0-9.0		
Type III	Average	21.3	0.8	6.7	17.0	36
	Range	12.0-38.0	0.1-1.7	0.4-16.7		
Type IV	Average	20.8	0.5	37.8	10.4	5
	Range	13.0-32.0	0.1-0.9	32.1-50.0		

Table 5-4 summarizes the number of the recorded rainfall events for each of the different land use types. Type I rainfall events were captured for every land use type, while Type IV rainfall events were only captured for the residential areas. This does not mean that Type IV rainfall events can only occur in residential areas, but rather it reflects differences in the duration of the monitoring period at the individual study sites. For example, for the industrial land use areas, samples were only collected from 8 events from 2016 to 2017 and in 2020. This shorter duration resulted in only Type I and III rainfall events being observed for the industrial land use areas.

Table 5-4. Summary of number of recorded rainfall events for each of the different land use types.

Land use	Type I	Type II	Type III	Type IV
Paved residential	48	/	13	1
Gravel lane residential	31	1	13	4
Commercial	21	2	8	/
Roads	4	2	1	/
Mixed land use	9	13	/	/
Industrial	7	/	1	/

5.3.3 PSD for the different land uses

The particles in rainfall-runoff were classified into six types: clay (< 2 μm), silt (5-32 μm), very fine sand (32-125 μm), fine sand (125-250 μm), medium sand (250-500 μm), and coarse sand (>

500 μm). The PSD was presented in terms of the percentage of all particles in a water sample that are finer (by mass) than a specific particle size. Given that multiple samples were collected for each land use type, average PSDs for the runoff resulting from the rainfall events from different land uses are summarized in Fig. 5-5. The x-axis is the particle size, and the y-axis is the cumulated mass fraction finer than corresponding size. For example, the first point for RO means that there are 10.2% particles in runoff are less than 2 μm for roads. The last point of RO means that there are 92.2% particles are less than 500 μm for roads.

Fig. 5-5 shows that the mass fractions of particles less than 250 μm are over 75%, and the mass fractions of particles less than 63 μm are over 43%. Particles from 32 to 63 μm account for about 20%. Therefore, fine particles are the dominant particles of suspended solids in urban runoff in Calgary.

The percentage of particles less than 32 μm decreased in the following order: RO, MLU/IN, GLR, CO, and PR. Above this size the curves cross and the order changes slightly. Comparing the RO curve to the PR, 17% is finer than 32 μm for PR but 49% is finer for RO. These two roads are a highway and an arterial street with a high volume of traffic. Vehicle emissions, tire wear, and abrasion of the road could provide a large number of fine particles that are carried into the runoff during rainfall events (Müller et al. 2020). Sixty-one % of the particles in runoff from IN are less than 63 μm, which could be related to the presence of Oil-Grid separator. Coarse particulate matters were effectively removed by Oil-Grid separators (Tang et al. 2018). The removal efficiency for particles larger than 75 μm could reach more than 80% (ETV 2016). Therefore, greater proportion of fine particles were found in runoff from industrial land use.

GLR has finer particles than PR (33% vs. 17% for particle size less than 32 μm). Gravel lanes can provide a large number of particles of various sizes (Thenoux et al. 2007). Vehicles on the gravel lanes also grind to produce more fine particles. It was found that coarser particles on rougher surfaces need higher rainfall intensities that provide higher kinetic energy to mobilize compared to smoother surfaces (Zhao et al. 2018). Therefore, the rough gravel lanes and the low intensity of rainfall runoff characteristic of Calgary resulted in greater proportion of fine particles being mobilized compared to smoother paved roads. Greater proportion of coarse particles will be carried into runoff in smoother paved roads under the same rainfall conditions.

MLU tends to produce more fine particles which could be due to the larger area of the mixed land use and the grinding effect that occurs during sediment transport in long pipes, resulting in the production of more fine particulate matters (Garde. 1956). In addition, it was found that coarse materials are easily deposited at the upstream end of the storm trunk sewer (Tang et al. 2020), which could result in proportionally more fine suspended solids at the downstream end.

We compared our PSD with Calgary design manual (The City of Calgary 2011), ETV (ETV 2016) and NJDEP (NJDEP 2021). PSD curve of Calgary design manual is between the lower and upper envelopes of our study. For the particles finer than 150 µm, Calgary design manual has similar percent as commercial and mixed land use. NJDEP and ETV used the same PSDs where 90% particles are less than 250 µm.

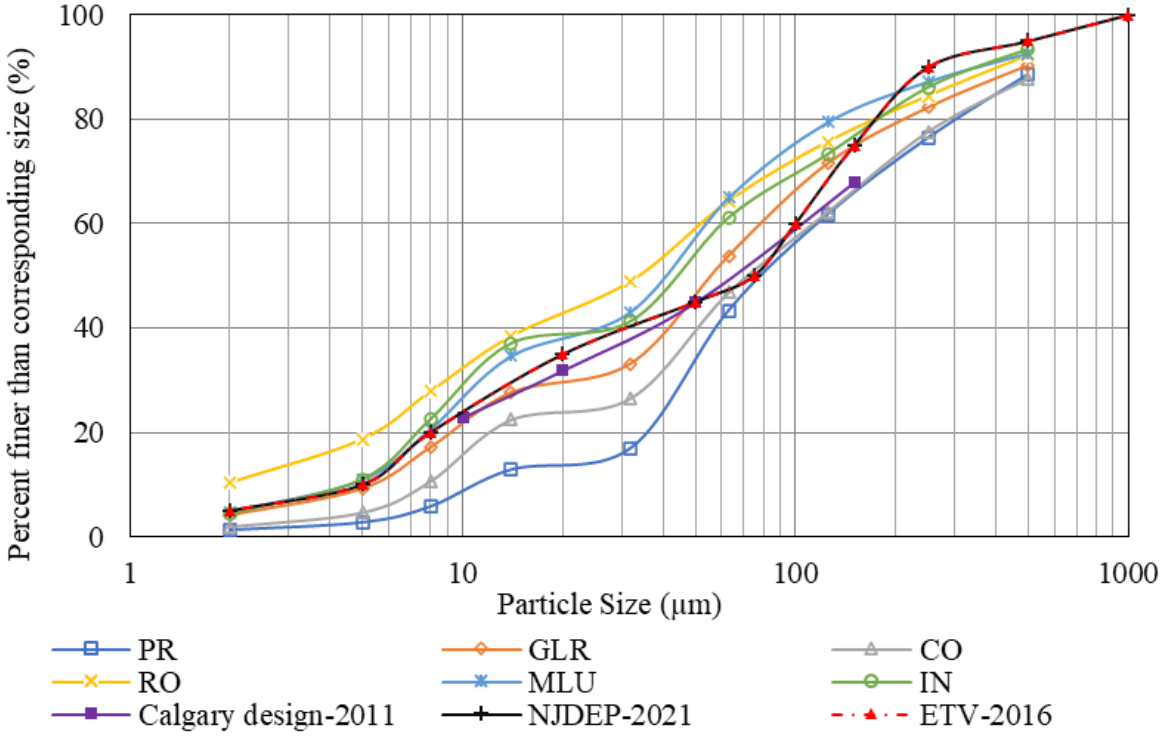


Fig. 5-5. PSDs from different land uses.

Table 5-5. Comparison of urban runoff PSDs with previous studies (modified from Kim (2008) and Charters (2015)).

Land use	Number of sites	Number of samples	Sample method	Analytical method	D_{50}	Sources
Highways	1	9	Autosampler	Wet sieve	<50	(Furumai et al. 2002)
	3	172	grab sample	Particle analyzer	2.7 – 7.1	(Li et al. 2006)
	2	8	Grab, runoff collection tank	Wet sieve; particle analyzer	136	(Kim and Sansalone 2008)
	4	20	Autosampler		74.6	(Winston and Hunt 2017)
	1	4	Autosampler	Wet sieve; particle analyzer	55	Present
Arterial streets	N/A	at least 12	Grab	Particle analyzer	< 100	(Drapper et al. 2000)
	1	48	Autosampler	Wet sieve; particle analyzer	95	(Selbig and Bannerman 2011)
	2	17	Autosampler	Wet sieve; particle analyzer	43	(Selbig 2015)
	4	N/A	domestic vacuum cleaner		80-350	(Borris et al. 2016)
	2	13	autosampler		38	(Winston and Hunt 2017)
	1	3	Grab	Particle analyzer	150*	(Zhao et al. 2022)
	1	3	Autosampler	Wet sieve; particle analyzer	17	Present
Residential	1	64	Autosampler		8	(House et al. 1993)
	1	N/A	Autosampler	Settling tests	< 4	(Greb and Bannerman 1997)
		46	A, B	Particle analyzer	9	(Burton Jr and Pitt 2001)

Continued Table 5-6. Comparison of urban runoff PSDs with previous studies (modified from Kim (2008) and Charters (2015)).

Land use	Number of sites	Number of samples	Sample method	Analytical method	D_{50}	Sources
Residential	4	68	Autosampler	Wet sieve; particle analyzer	250	(Selbig and Bannerman 2007)
	2	10	Grab	dry sieved	180-710	(Bathi et al. 2009)
	1	62			80	(Bian and Zhu 2009)
	1	12	Autosampler	Particle analyzer	81	(Charters et al. 2015)
	1	19	Autosampler	Wet sieve; particle analyzer	80	(Selbig 2015)
	4	198	Autosampler	Wet sieve; particle analyzer	80	Present
Commercial	1	51	Grab	dry sieved	63*	(Hoffman et al. 1984)
	1	5	Grab	dry sieved	300-450	(Bathi et al. 2009)
	1	62			108	(Bian and Zhu 2009)
	3	94	Autosampler	Wet sieve; particle analyzer	54	(Selbig and Bannerman 2011)
	3	56	Autosampler	Wet sieve; particle analyzer	86	Present
Industrial	1	76	Grab	dry sieved		(Hoffman et al. 1984)
	1	18	Grab	dry sieved	20-105*	(Jeong et al. 2020)
	4	5	Autosampler	Wet sieve; particle analyzer	46	Present

Continued Table 5-7. Comparison of urban runoff PSDs with previous studies (modified from Kim (2008) and Charters (2015)).

Land use	Number of sites	Number of samples	Sample method	Analytical method	D_{50}	Sources
Mix land use	1	4	Autosampler	Particle analyzer	31	(Anta et al. 2006)
	1	20	Autosampler	Wet sieve; particle analyzer	42	(Selbig and Bannerman 2011)
	1	10	Autosampler	Wet sieve; particle analyzer	95	(Selbig 2015)
	3	22	Autosampler	Wet sieve; particle analyzer	40	Present

A comparison urban runoff PSDs was made with previous studies (Table 5-5). In addition to PSD in runoff, a study of PSD of road deposited sediment was included in this comparison (Bian and Zhu 2009) because they collected sediment samples from various source areas. These studies were compared from different aspects including land use, number of sites, number of samples, sample method, analytical method and D_{50}. D_{50} varies among land uses. For instance, the D_{50} for residential ranged from 710 to less than 4 μm. Most studies have similar D_{50} (80 μm) as this study for residential areas. Mixed land use tends to generate finer particles less than 95 μm. The variation was related to the number of sites, sample method and analysis method.

The variability of PSD among rainfall events has been found in these studies, and the number of samples and the number of sampling points can affect the characterization of PSD. Therefore, more sampling numbers and rainfall classifications are needed to characterize the PSD accurately. The sampling method can also affect the PSDs in runoff. Automatic sampling is a common method in most studies and grab sampling also has been used in some studies. Although the automatic sampling method is convenient, it can only obtain the PSD of the suspended solids from specific locations. When the intake of sampling tube is near the bottom, it is susceptible to bedload, and when it is near the top, the sampling of PSD may be under sampled (Selbig 2011). For the analysis of the suspended fraction, a combination of wet sieve and particle analyzer is usually used, which can cover the entire range of particle sizes in runoff for suspended fraction. The particle analyzer measures the finer particles, while the wet sieve method measures the coarse particulate matters.

5.3.4 PSD for the different land uses under the same rainfall conditions

Fig. 5-5 compares the PSD from different land uses but ignores the possible effect of rainfall on the PSD. Fig. 5-6 compares the PSD from different land uses under the same rainfall event type. The solid line is the PSD containing all events, and the dotted line is the PSD under specific rainfall event type.

Compared to the original PSD from RO, the percentage of particles less than 63 μm increased from 64% to 73% under type I rainfall events (Fig. 5-6a). Among the other land uses, the changes in PSD were relatively minor. The percentage of particles less than 63 μm decreased in the following order: roads, mixed land use/industrial, gravel lane residential, commercial, and paved residential land uses. During type II rainfall events (Fig. 5-6b), the PSD curve of RO exhibited noticeable changes compared to the original PSD. The road had a straight and uniform PSD curve. However,

the PSD curves of other land use experienced some position shifts without obvious changes in shape and order.

During type III rainfall events (Fig. 5-6c), the PSD curve of RO exhibited obvious changes in shape. The proportion of particles smaller than 63 μm decreased from 64% to 38%. On the other hand, the PSDs of other land uses experienced some shifts in position, but the overall shape and order remained relatively stable. During type IV rainfall events (Fig. 5-6d), both the shape and position of PSDs from PR and GLR underwent obvious changes. The peak in the PSD from gravel lane residential disappeared, and the proportion of particles smaller than 63 μm increased from 54% to 76%. Similarly, paved residential also exhibited a similar trend, with the proportion of particles smaller than 63 μm increasing from 43% to 54%. However, it is worth noting that the particles from gravel lane residential remained finer compared to those from paved residential.

In summary, under the same rainfall conditions, the positions of PSDs from land uses other than roads showed slight changes under types I, II, and III rainfall events, but the overall order remained relatively consistent. However, the PSD from roads exhibited obvious changes under type III, and residential areas experienced dramatic changes under type IV rainfall events. This indicates that land use could indeed influence the PSDs and determine the order of the decreased fine particle fraction. Additionally, these findings reveal that rainfall event types could also impact PSDs from different land uses.

In this study, the rainfall events are assumed to be independent of each other. It should be noted that the individual sampling events may not be independent. Factors such as sediment build-up and wash-off can influence the sediment transport between events, which could impact interpretation of the statistical analysis. Collecting sediment samples on urban surface and incorporating a time-series analysis into the methodology would be a valuable direction for future research in this area.

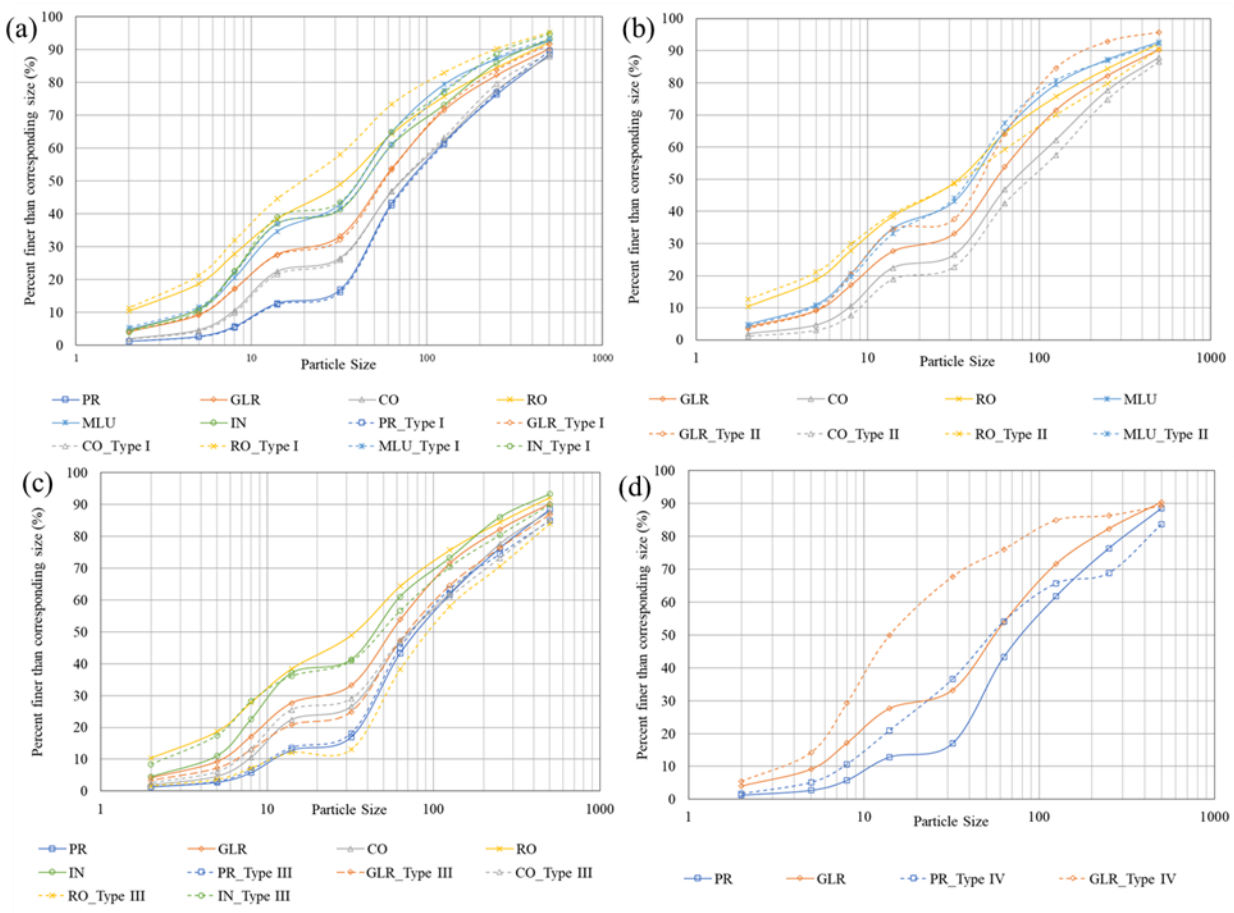


Fig. 5-6. PSDs from different land uses under (a) Type I rainfall events; (b) Type II rainfall events; (c) Type III rainfall events; (d) Type IV rainfall events.

5.3.5 Effects of rainfall event types on PSD

The aforementioned results demonstrate that rainfall event types could have varying impacts on the PSDs from different land uses. Therefore, this section evaluates to what extent the PSDs for the different land uses is a function of the rainfall event type. Fig. 5-7 (a-f) s compared PSD under different rainfall event types to evaluate the effects of rainfall. The PSD curves for PR, GLR, and RO showed obvious variability for the different rainfall event types.

In the case of PR (Fig. 5-7a), the availability of particulate matter is limited. The PSD in runoff is primarily influenced by the accumulation process of particulate matter during antecedent dry days (Vaze and Chiew 2002). Type I and III rainfall events have short antecedent dry days, resulting in similar PSD curves. While during long antecedent dry period under Type IV rainfall events with long rainfall duration, more particles could be accumulated on urban surface and finer particles are

more easily to be mobilized (Miguntanna et al. 2013, Wijesiri et al. 2015). Compared to type I and III rainfall events, type IV rainfall events demonstrate a notable increase in particles smaller than 32 μm, with the percentage rising from 16% to 37%.

For GLR (Fig. 5-7b), the availability of particulate matter is abundant due to the presence of gravel lanes, resulting PSD in runoff being primarily influenced by the wash-off process during rainfall events (Vaze and Chiew 2002). Type III rainfall events have a longer duration and higher total rainfall compared to Type II rainfall events (Table 5-3). Consequently, type III rainfall events exhibit a greater proportion of coarse particles than type II rainfall events (13% compared to 4% for particles larger than 500 μm). Additionally, during long antecedent dry period, the movement of vehicles on gravel lanes can generate significant amounts of dust (Thenoux 2007), which are susceptible to being entrained and transported by rainfall. This may be one explanation of why there is a obvious increase in the percentage of particles ranging from 8 to 32 μm during type IV rainfall events.

There is no discernible variation in the PSD from commercial areas under different rainfall event types (Fig. 5-7c). This could potentially be attributed to frequent cleaning activities in the commercial area, which help maintain the surface PSD close to its initial state. It is found that as a particle size weighted average, the removal efficiency of a vacuum sweeper (60 to 92%) was higher than that of a mechanical sweeper (20 to 30%) (Breault et al. 2005). Higher rain intensity could transport more large particles, while longer rainfall durations also contribute to this effect. Therefore, compared to type I rainfall events, type II and III rainfall events tend to exhibit a higher proportion of coarse particulate matter from RO (Fig. 5-7d).

The PSDs from MLU (Fig. 5-7e) showed similar results. This could be attributed to the representativity of rainfall data. The mixed land use areas in the study are larger than other land uses, and rainfall data from one to three nearby rain gauges were used to represent the rainfall conditions. However, heavy rainfall events such as Type II rainfall events tend to be localized, which means that they are unlikely to occur on a catchment-wide basis. More representative rainfall data need to be obtained by installing more rain gauges in larger catchments. PSDs from IN during Type I and III rainfall events exhibit similar patterns (Fig. 5-7f). The PSD curves intersect in the 8-14 μm range, with slightly more fine particles observed for Type III rainfall events and slightly more coarse particles for Type I rainfall events, although the difference is not

obvious. Besides the issue of rainfall representativeness, the presence of an oil-grit separator may also contribute to the resemblance of PSDs.

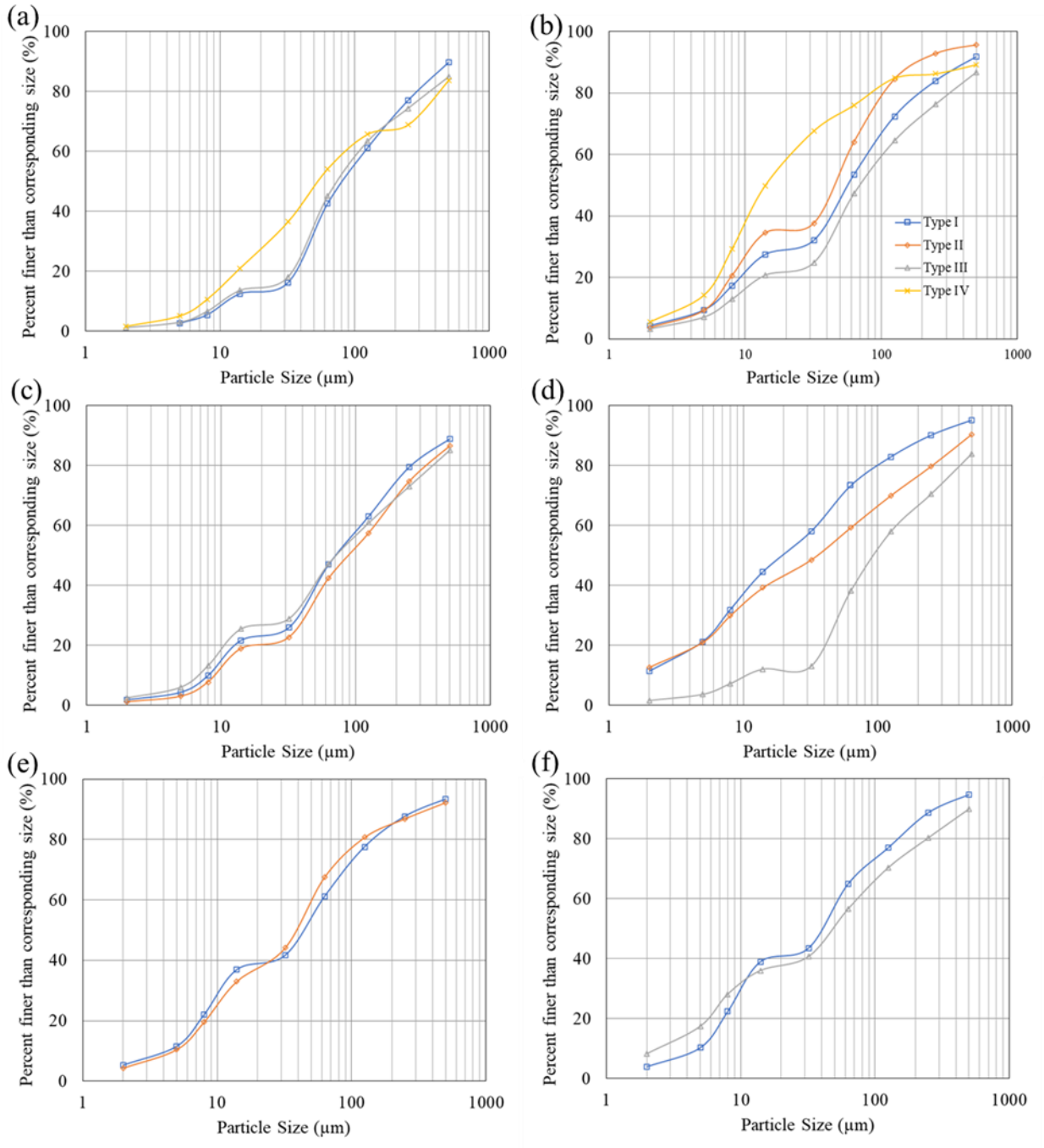


Fig. 5-7. PSDs under different rainfall event types from (a) PR; (b) GLR; (c) CO; (d) roads; (e) MLU; (f) IN.

Apart from rainfall event types, seasons could also impact the PSDs in runoff. Fig. 5-8 illustrates the PSDs under different seasons, with similar trends for different land uses. The PSD in spring exhibits the finest particles. Taking RO as an example (Fig. 5-8d), in spring, 62% of particles are smaller than 32 μm, compared to 48% in summer and 13% in fall. This observation could be attributed to the street sweeping activities in the spring. In Calgary, a significant street sweeping operation is conducted in the spring. Previous studies indicate that sweepers have high removal efficiency for particles larger than 63 μm but poor efficiency for finer particles (Amato et al. 2010). This leads to an increase in the proportion of fine particles on the streets, consequently resulting in a higher proportion of fine particles in the spring runoff. The increase of coarse particles in the fall may be due to the accumulation of leaf litter and the reduction of vegetation on greenspaces and landscapes. These factors make the soil more susceptible to erosion by rainfall runoff, resulting in a higher percentage of coarse particles in the runoff.

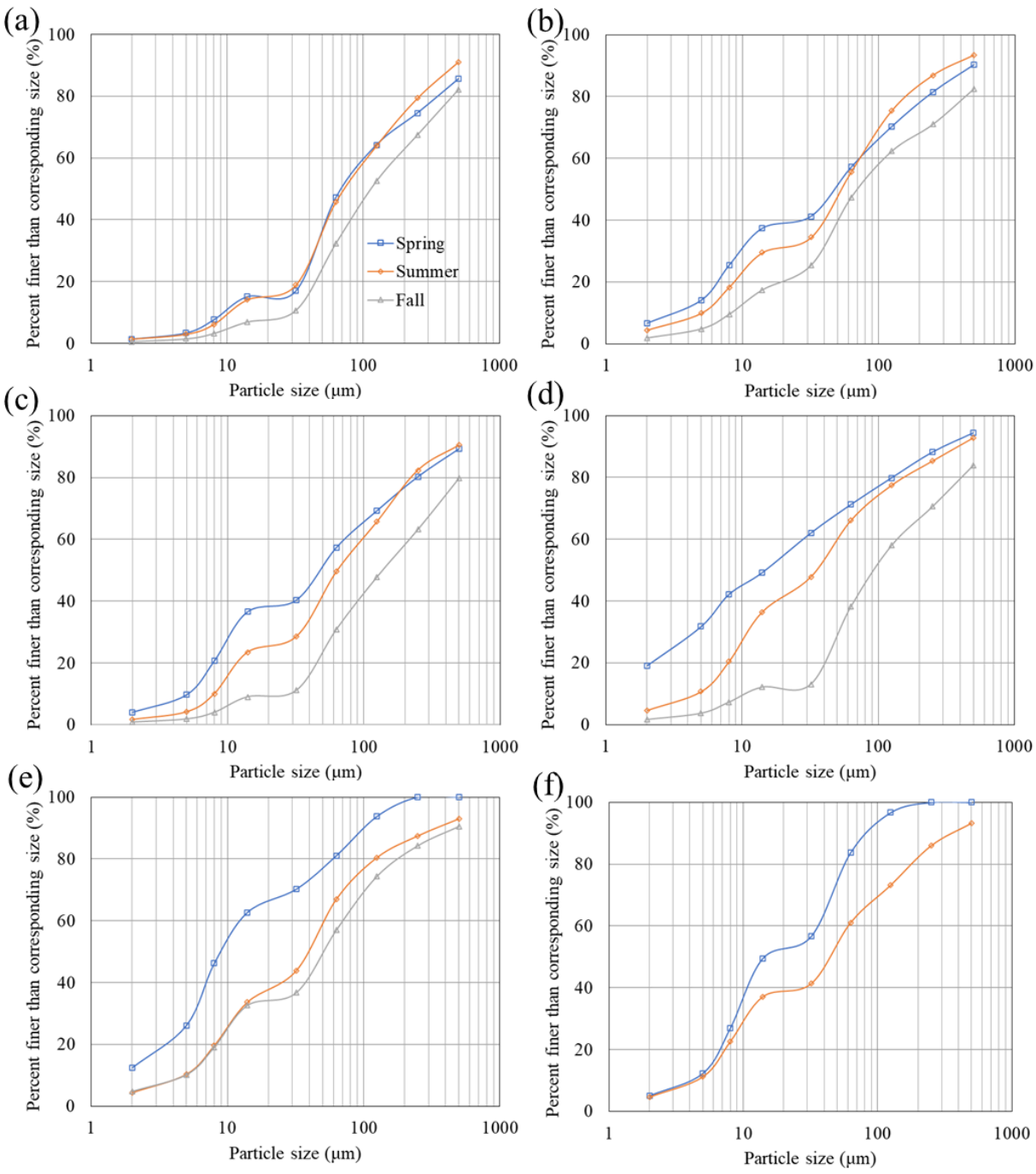


Fig. 5-8. PSD in different seasons from (a)PR; (b) GLR; (c) CO; (d) RO; (e) MLU; (f) IN.

5.3.6 Comparison of PSD between rainfall events and snowmelt events

Fig. 5-9. compares PSD curves for rain event and snowmelt from different land uses. The particle size increases in the order of roads, gravel lane residential, commercial, and paved residential areas.

Roads have the finest particles compared to other land uses except in the 8-14 um fraction. The snowmelt PSDs have similar trends as the rain event PSDs, i.e., the paved residential land use tends to have coarser particle sizes compared to the commercial and gravel lane residential land uses. The particles during snowmelt events are finer for the same land use compared to those during the rainfall events. This could be attributed to the fact that rainfall-runoff flows tend to be orders of magnitude larger than snowmelt flows, which therefore could only mobilize and carry finer particles.

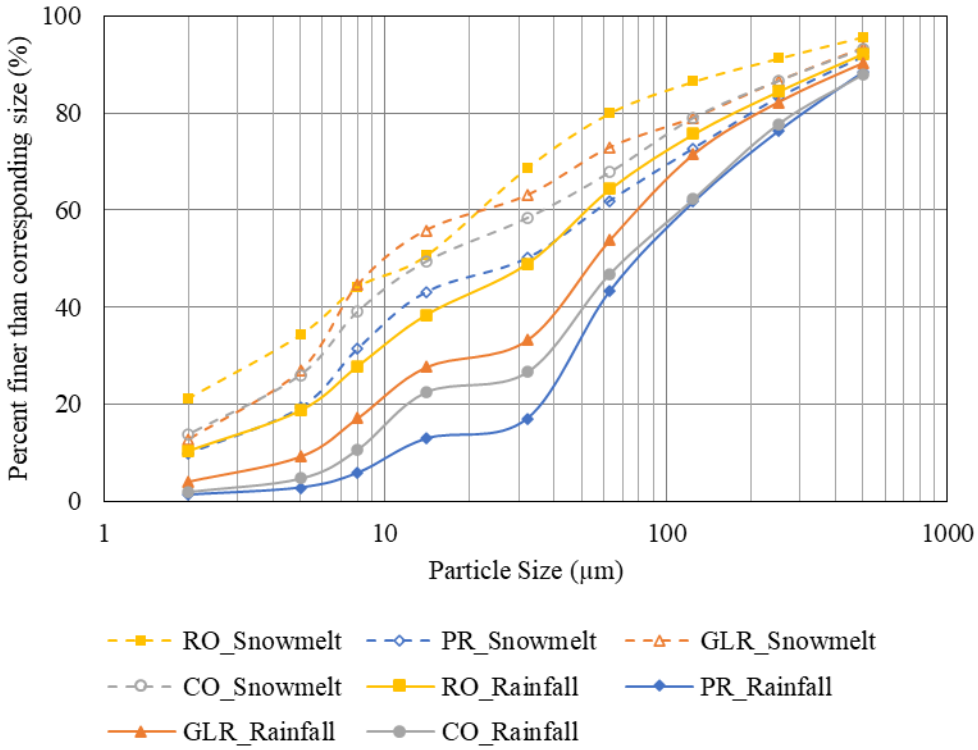


Fig. 5-9. Comparison of PSDs from different land uses during rainfall and snowmelt events.

5.4 Conclusions

This research is based on the findings from a long-term (2016 - 2021) water sampling program at multiple sites in a cold-region, semi-arid and summer rain dominated city to study the PSD from different land uses under different weather conditions. The main findings and conclusions are as follows:

(1) The mean D_{50} from roads and industrial areas are not significantly different, while the mean D_{50} from paved residential and commercial area are significantly higher compared to roads and

industrial areas. In all, the order of the mean D_{50} is paved residential>commercial>gravel lane residential>mixed land use>industrial, roads.

(2) Fine particles less than 125 μm are the dominant particles (over 75%) of suspended sediments in runoff in Calgary. The PSD curves of runoff from different land use showed obvious differences. Roads have largest percentage of particles finer than 32 μm (49%) due to vehicle emissions, tire wear, and abrasion of the road. Gravel lane residential areas have finer particle sizes than paved residential areas due to the grinding effect and the rough properties of gravel lanes.

(3) Rainfall event types could impact PSD in runoff from residential areas and roads. Typically, A long antecedent dry period tends to result in the accumulation of fine particles. A high rainfall intensity and long duration could wash off more coarse particles. The effects of rainfall event types could vary depending on land use types. The impact of rainfall event types for commercial and industrial areas is negligible due to the removal efficiency of street sweeping and oil-grid separators. The season is a factor in the PSD of suspended sediment in runoff. The PSD in spring exhibits the finest particles due to the street sweeping.

(4) The order of particle size during snowmelt events is similar to rainfall events. Particles in snowmelt are finer for the same land use than that during rainfall events because the rainfall-runoff flows are usually an order of magnitude larger than the snowmelt flows, which could only tend to mobilize and carry finer particles.

6. Prediction of urban stormwater quality in data-deficient areas using a semi-supervised machine learning framework

6.1 Introduction

Urbanization can introduce large number of pollutants into receiving water bodies through stormwater runoff, which can have detrimental effects on environment and public health (Li et al., 2016). Urban stormwater quality models are crucial tools for the planning, analytical assessment, and design of drainage system to manage stormwater quality (Gironás et al., 2010, Liu et al., 2013). Stormwater quality models may generally be classified into two primary modelling approaches: process-driven models such as PCSWMM, which are based on physical principles, and data-driven models that rely on data analysis (Wijesiri et al., 2020). PCSWMM, a dynamic rainfall-runoff simulation model, is widely used for managing urban stormwater quantity and quality. These models offer detailed simulations of the processes involved in urban catchments (Rossman, 2010). However, they come with the caveat of requiring extensive and detailed data, which encompasses not only meteorological and water quality metrics but also necessitates intricate knowledge of the catchment drainage systems (Fletcher et al., 2013). The need for comprehensive calibration and validation data, often scarce or unavailable, constrains their utility, especially in data-deficient regions (Chow et al., 2012).

Data-driven models, especially those utilizing machine learning (ML) algorithms, are based on statistical relationships derived from observed data to apply in environmental research (Granata et al., 2017, Ahmed et al., 2019, Behrouz et al., 2022). Data-driven models require processes-based studies to determine a parsimonious selection of parameters. Data-driven models have emerged as a simpler alternative of process-driven models due to their fewer extensive data requirements, relying predominantly on weather, land use and water quality data. For example, using six factors (rainfall depth, runoff volume, and antecedent dry days, drainage area, land use, percent of imperviousness) as input features, effective machine learning models were developed to predict total suspended solids (TSS) (Moeini et al., 2021). Data-driven models can capture complex, non-linear relationships inherent in environmental systems when sufficient training data are available (Xu et al., 2020). For instance, a machine learning approach was developed to predict pollutant event mean concentrations (EMC) in urban runoff using the National Stormwater Quality Database

(NSQD) that contained over 5000 stormwater samples from approximately 300 urban catchments (Behrouz et al., 2022). Results demonstrate that machine learning is a powerful approach for estimating EMCs from urban catchments and determined the characteristics that had the greatest effects on EMCs prediction.

Despite their advantages, ML models are not without limitations. Their predictive power is contingent on the quantity and quality of training data (same as process-driven models) (Zhi et al., 2021). Collecting stormwater quality data, is often time-consuming and expensive, leading to a lack of available stormwater quality data (McKenzie et al., 2013). The majority of urban catchments suffer from a dearth of comprehensive water quality datasets, which can lead to poor model performance and generalization in ML applications (Zhu et al., 2022). Furthermore, ML models often lack interpretability and transparency, making it difficult to understand how they make predictions and what factors influence their outputs (Rudin et al., 2019). This can limit the trust and acceptance of ML models by decision makers.

Semi-supervised machine learning could be a promising solution to the challenge of data scarcity in urban stormwater quality modeling (van Engelen and Hoos et al., 2020). Semi-supervised learning is a machine learning paradigm that utilizes both labeled and unlabeled data for model training (Berthelot et al., 2019). By effectively leveraging unlabeled data, it allows for improved learning and generalization, especially in contexts where obtaining labeled data is challenging. Therefore, by utilizing the abundance of rainfall data, which is more readily available and often of high quality, in conjunction with limited water quality data, semi-supervised models could be trained more effectively. In addition, random forest (RF) model, recently found favor machine learning approach, can provide insights into variable importance (Yang et al., 2023). The selection of input variables was typically based on the hypotheses that the influential variables of urban stormwater quality can also affect stormwater quality predictions. This interpretability is important in understanding which factors mostly influence stormwater quality predictions.

Therefore, this research aims to harness the strengths of semi-supervised machine learning based on random forest algorithm to develop a framework capable of predicting urban stormwater quality in regions where data are lacking and apply this approach in urban catchments in Calgary. Through the integration of ample rainfall data with sparse water quality measurements, we anticipate the formulation of a model that offers both accuracy and applicability, even in data-deficient areas.

The ensuing sections will detail the methodology employed in this study, the results and discussion thereof, and the conclusions drawn from our findings, thereby contributing to the ongoing endeavor of sustainable and effective urban stormwater management.

6.2 Methodology of semi-supervised machine learning framework

6.2.1 Model selection: random forest (RF)

To address the challenge of predicting urban stormwater quality, we selected the Random Forest (RF) model as the foundational machine learning approach, attributed to its benefits suitable for urban stormwater datasets. RF demonstrates resilience against overfitting, a common defect in machine learning models, especially when the dataset is limited (Díaz-Uriarte and Alvarez de Andrés, 2006). Overfitting can lead to models that perform well on training data but failed on testing data. The structure of RF, which involves bootstrapping data and building multiple decision trees, can avoids this issue, ensuring a robust generalization capability. Urban stormwater datasets often exhibit non-linear relationship due to the various interactions between different features. RF model can use the existing data and capture the complex nonlinear relationships between input and output data (Boulesteix et al., 2012). Despite being a machine learning model, RF can provide insights into variable importance (Grömping., 2015). This interpretability is important in understanding which factors mostly influence stormwater quality predictions.

Fig. 6-1 provides a comprehensive flowchart of the RF implementation in our study. This study utilized the dataset from the four catchments in Calgary, Alberta, Canada, as described in Section 4.2. The individual catchments, Rocky Ridge, Royal Oak, Cranston, and Auburn Bay recorded 15, 13, 18, and 11 events respectively, totaling 57 events. The RF model requires a labeled dataset, wherein each instance is paired with a corresponding output label (dependent variables). Labeled data includes both input features (independent variables) and output labels. In this study, the labeled dataset consists of input features and output labels from four catchments.

The input features encompass rainfall and catchment characteristics because they can significantly affect stormwater quality: Potentially suitable rainfall characteristics were selected that could impact urban stormwater quality. They include antecedent dry days (ADD), total rainfall amount (RT), average rainfall intensity (RI-avg), rainfall duration (RD), maximum rainfall intensity (RI-max), time to peak rainfall intensity (T-peak), and the initial 30-min total rainfall amount (RT-30).

Catchment characteristics are primarily categorical variables, emphasizing the physical and anthropogenic attributes of the catchments. They include catchment area (CA), imperviousness (Imp), and detailed land use and cover categories. Land use consists of single family residential (SFR), multi-family residential (MFR), commercial (CO), institutional (IN), open space (OS) and transportation (TR). Land cover consists of Road, Roof, Landscaped (LS) and Gravel.

The primary objective is to predict Event Mean Concentrations (EMC) for three critical water quality parameters: Total Suspended Solids (TSS), Total Nitrogen (TN), and Total Phosphorus (TP). These parameters are important indicators of water quality and have significant impacts on the environment.

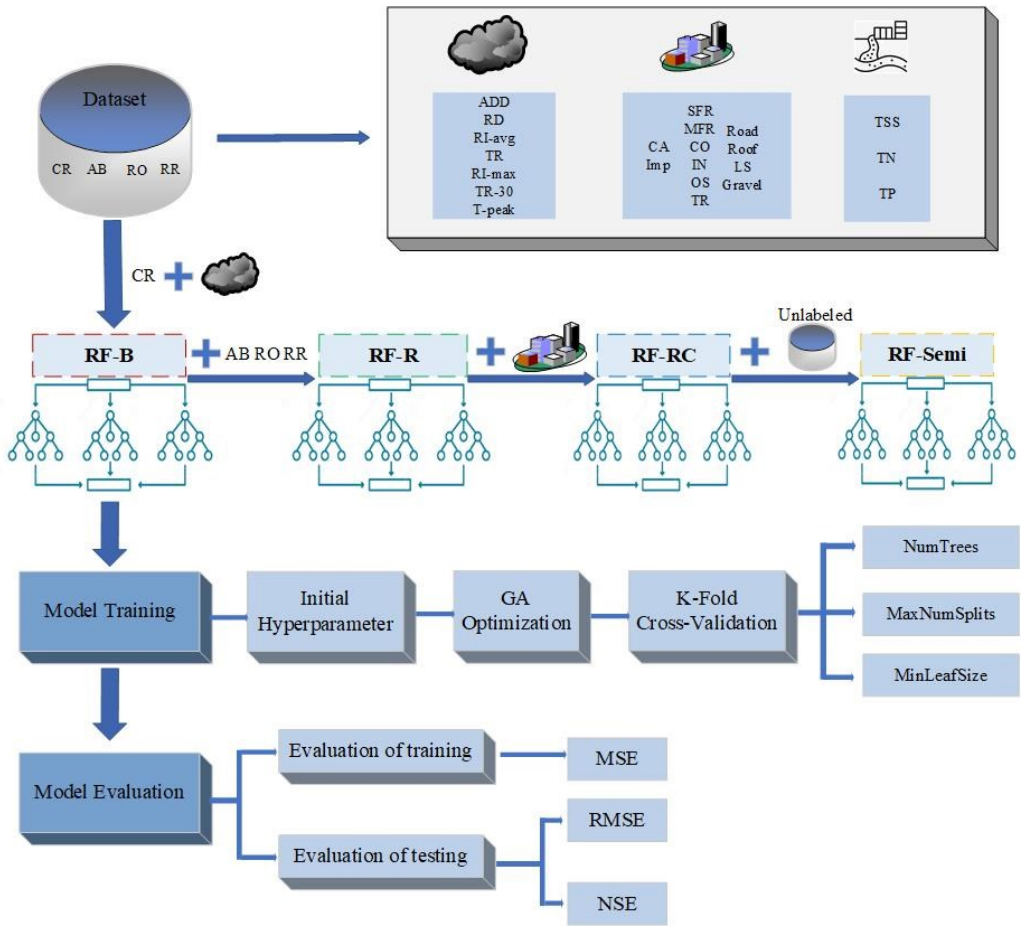


Fig. 6-1. Flowchart of random forest for four scenarios.

6.3.2 Base model development: RF-B

The initial stage involved developing a base RF model, termed RF-B. Given the data scarcity, RF-B model utilized only the Cranston catchment data, which had relatively more water samples. For

this model, only rainfall data was employed as the input features. The 2019 dataset was designated for training and validation, ensuring a recent and representative data subset. The RF model hyperparameters—namely the number of trees (NumTrees), maximum number of splits (MaxNumSplits), and minimum leaf size (MinLeafSize)—are crucial to its performance (Tyralis et al., 2019). NumTrees represents the total decision trees in the forest, influencing robustness and computational intensity (Probst and Boulesteix, 2017). MaxNumSplits defines the maximum splits in each tree, controlling the tree size and complexity, while MinLeafSize is the minimum number of observations contained in each leaf node in each decision tree.

To optimize these hyperparameters, a 5-fold Cross-Validation is embedded within a Genetic Algorithm (GA) framework based on 2019 dataset. GA iteratively searches for the optimal hyperparameter set, evaluating each set performance using the Cross-Validation average Root Mean Square Error (RMSE). This integrated approach ensures a comprehensive assessment of the hyperparameters, focusing both on model accuracy and ensuring that there are no signs of overfitting (Yang et al., 2023).

Upon determining the optimal hyperparameters, the RF models are retrained and validated on the complete 2019 dataset. The dataset was randomly divided into a training dataset (80%) and a validation dataset (20%). Following retraining, the model performance is assessed on the 2018 dataset, serving as an independent testing set. Testing dataset can evaluate the performance of the model on unseen data, i.e., the model generalization ability. This evaluation phase utilizes RMSE and Nash-Sutcliffe Efficiency (NSE) as index to quantify predictive accuracy and model efficacy.

6.3.3 Model enhancements: RF-R and RF-RC

Based on the RF-B model, we attempt to improve model performance through data and feature enhancement: The RF-R model extends the dataset by integrating data from all catchments and using rainfall data as input features. RF-RC, representing a more comprehensive approach, utilized data from all catchments and enriched the feature set by including catchment characteristics.

6.3.4 Semi-supervised RF model development (FR-Semi)

The success of semi-supervised learning depends on the accuracy of pseudo-labeling (Lee et al., 2013). Selection of pseudo-labels with high confidence as training targets helps ensure that the model learns from correct predictions rather than uncertain predictions (Zhou., 2021). In ensemble

methods like Random Forests, the variance in predictions across different models can serve as an indicator of confidence (Sohn et al., 2020). In this study, we employed the ensemble method for the selection of high-confidence pseudo-labels (Fig. 6-2). The ensemble was created by slightly varying the hyperparameters (number of trees, maximum depth, and minimum leaf size) within specified ranges around their optimal values, resulting in 27 unique RF-RC models. Each model independently predicted outcomes for the unlabeled dataset, and the coefficient of variation (CV) across these predictions was calculated for each sample to assess prediction dispersion. Samples with lower CV values, indicating higher agreement among models, were selected as high-confidence pseudo-labels. The average predictions of these selected samples formed the pseudo-labels, which were combined with the original labeled dataset to create an enriched training set. This enriched set was used to retrain the RF-RC model, aiming to improve its predictive accuracy and generalization capabilities in scenarios with limited labeled data.

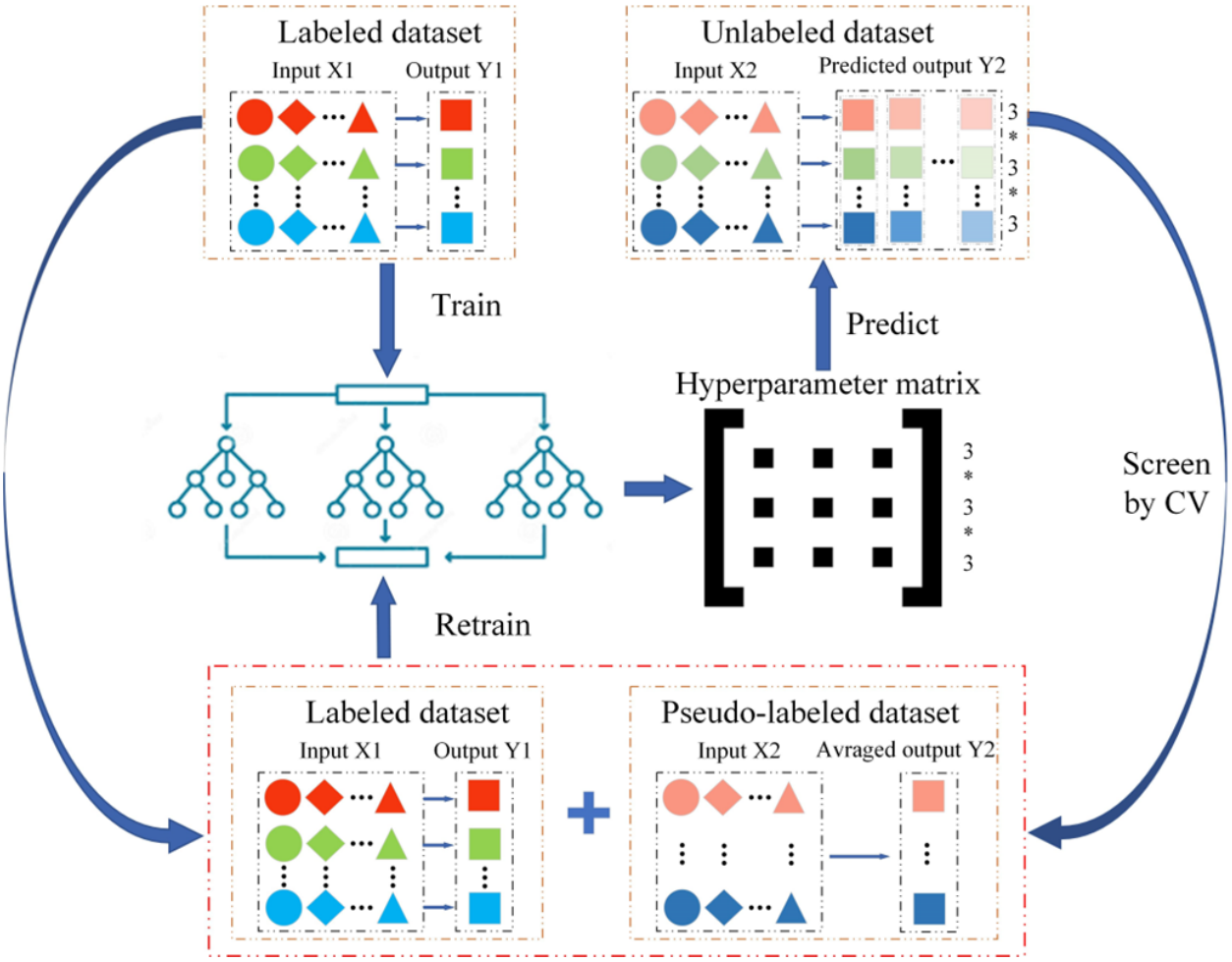


Fig. 6-2. Flowchart of semi-supervised random forest model.

6.3 Results and discussion

6.3.1 Selection of Pseudo-Labeling samples

The process of pseudo-labeling involves using the predictions of a model on an unlabeled dataset and incorporating those predictions back into the training process as if they were labeled data. Fig. 6-3. shows the boxplots of Coefficient of Variation for Pseudo-Labeling selection, which represent the variability of 27 RF-RC models in predicting TSS, TN and TN using unlabeled dataset. Each data point in the figure represents one unlabeled rainfall event. The upper extreme of the CV boxplot is set as the threshold for each parameter. Accordingly, the CV thresholds for TSS, TN, and TP have been set to 0.12, 0.06, and 0.09, respectively. After applying these thresholds, 55 out of the 65 unlabeled samples were retained for pseudo-labeling, indicating that most of the model predictions are consistent enough to be considered for retraining the model and hopefully improving its performance.

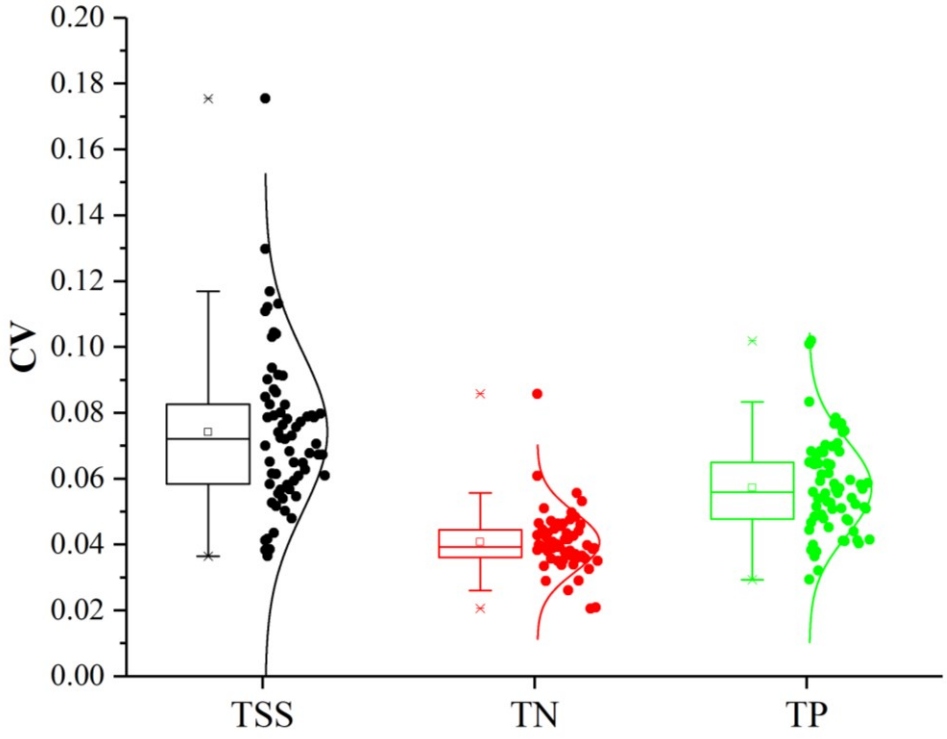


Fig. 6-3. Boxplots of Coefficient of Variation for Pseudo-Label Selection.

6.3.2 Performance of RF model (Compare with other studies)

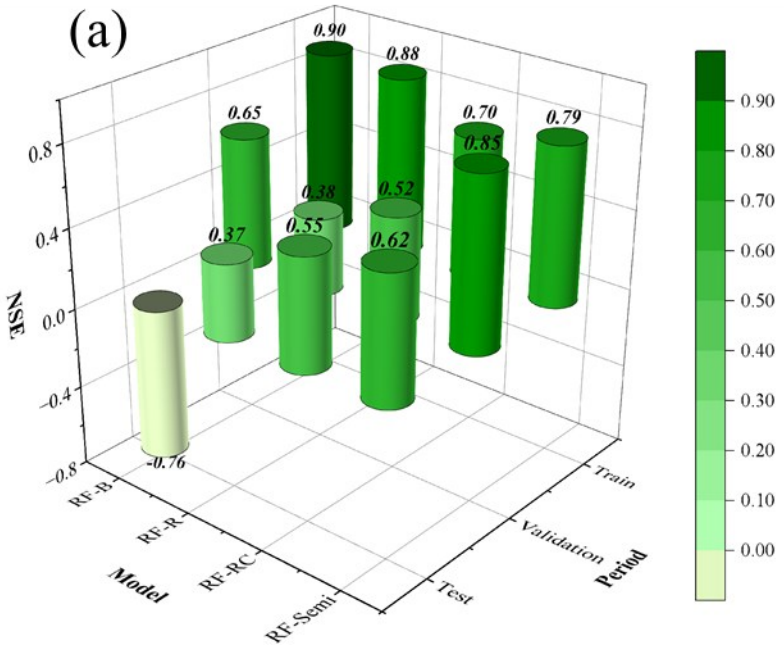
To evaluate the performance of the different RF models for TSS, TN and TP, the Nash-Sutcliffe NSE and RMSE values are shown in Fig. 6-4 and Fig. 6-5, respectively. The base model (RF-B) shows high NSE value (0.90) for TSS during training period (Fig. 6-4a), indicating good fit to the training data. However, the NSE for testing dataset is -0.76, which is worse, suggesting the model is not generalizing well to unseen data. The RMSE on the test dataset is 127 mg/L (Fig. 6-5a), reflecting the poor performance. With an NSE of -1.81 on the test and 0.15 on validation (Fig. 6-4b), the model performs worse than a simple mean predictor, and the high RMSE of 1.37 mg/L (Fig. 6-5b) on the test dataset confirms this. The NSE of TP on the test dataset is 0.54 (Fig. 6-4c) with an RMSE of 0.14 mg/L indicating better performance for TP. RF-B model performed better than RF-R and RF-RC model during training and validation period, which indicates that relying on limited data may lead to overfitting.

RF-R model shows improvement in the test dataset for TSS and TP compared to RF-B model. The NSE of TSS during testing period improves to 0.37 (Fig. 6-4a), and the RMSE decreases to 104 mg/L (Fig. 6-5a), showing better prediction capability than RF-B model. The NSE of TN is still negative at -0.21 (Fig. 6-4b) for the testing dataset, but the RMSE lowers to 0.57 mg/L (Fig. 6-5b), which is an improvement, although the model still struggles with TN prediction. With an NSE of 0.21 (Fig. 6-4c) for TP on the test dataset and an RMSE of 0.16 mg/L (Fig. 6-5c), the model is somewhat effective but not satisfactory. Overall, by integrating data from different catchments, the model predicting accuracy could be enhanced although the performance during training and validation period decreased.

RF-RC see slightly improvements in TSS and TP than RF-R model. RF-RC model shows a larger NSE of 0.55 (Fig. 6-4a) for TSS during testing period, indicating a better fit than RF-R. The NSE of TP for testing dataset increased from 0.21 to 0.46 (Fig. 6-4c). This indicates that consideration of catchment characteristics could improve the model capacity for generalization. TN, however, remains problematic, which might be due to the complex dynamics of nitrogen in urban stormwater that are not well captured by the model.

The RF-Semi model shows the best performance across all parameters, with high NSE values for the test and validation datasets, and the lowest RMSE values for the test dataset. The NSE of TSS for testing dataset achieves the best (0.62, Fig. 6-4a), showing better predictive accuracy. During

the testing period, the NSE and RMSE for TP reached 0.62 (Fig. 6-4c) and 0.11 mg/L (Fig. 6-5c), respectively, confirming the good predictive capability of the semi-supervised approach. The RF-Semi model has a more complex model structure due to additional data from pseudo-labeling, which provides a richer dataset for training (Berthelot et al., 2019). This indicates that incorporating pseudo-labeling has effectively improved the model's ability to predict unseen data, which allow the model to learn more generalized patterns. Although RF-Semi model performed well during training and validation periods for TN, NSE for testing period remain negative (-0.03), indicating a need for further model refinement for TN.



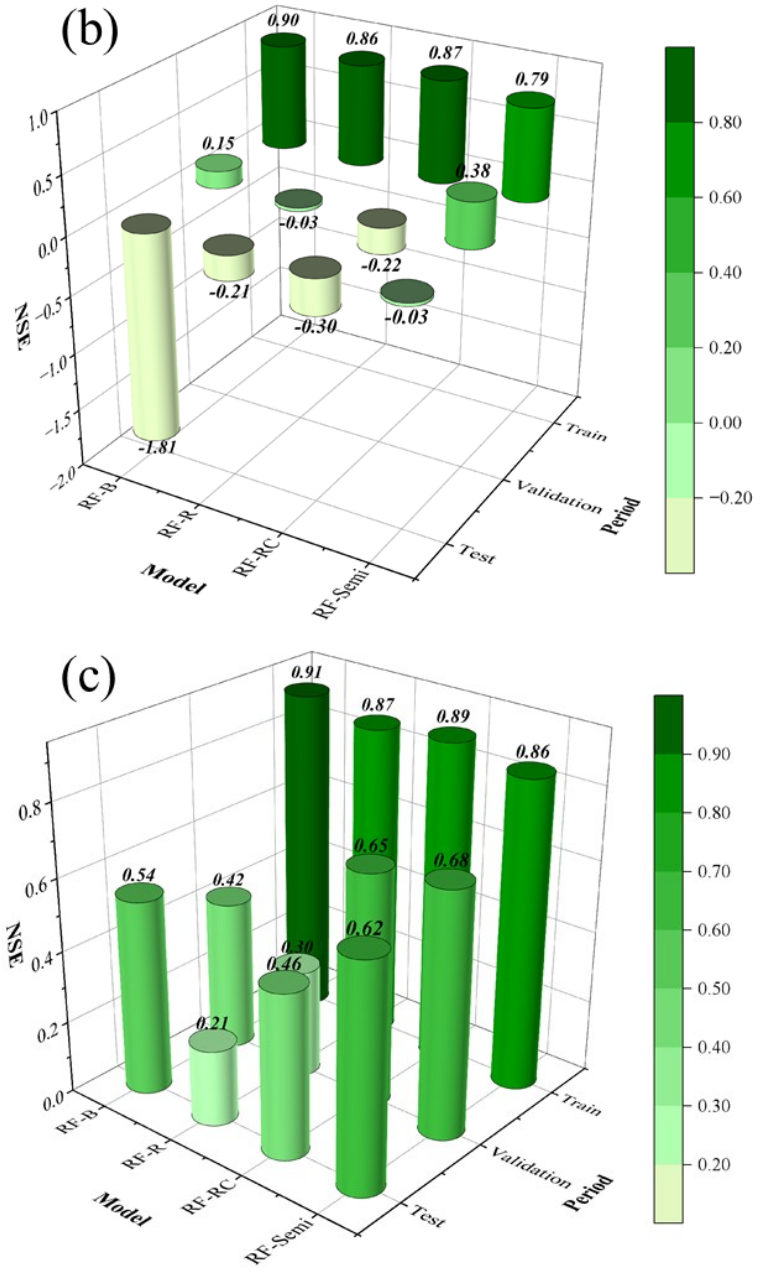
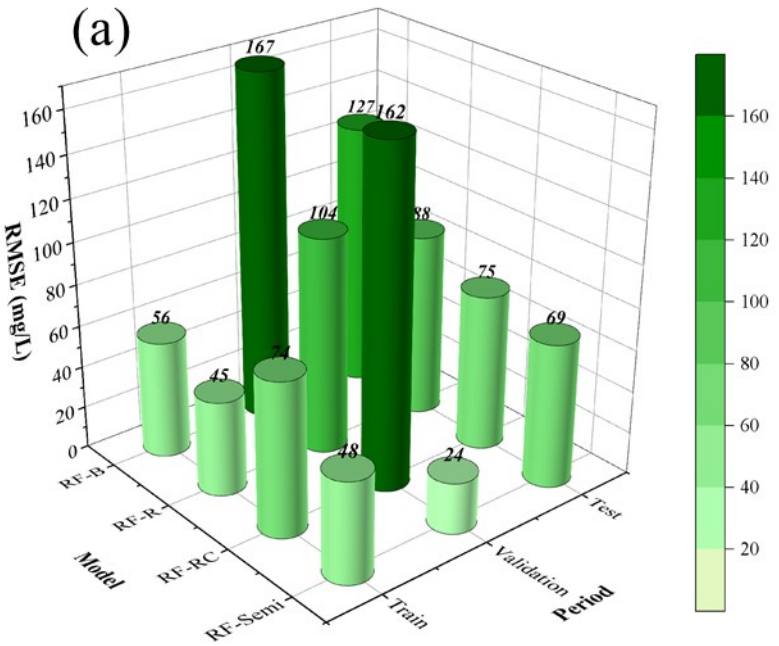
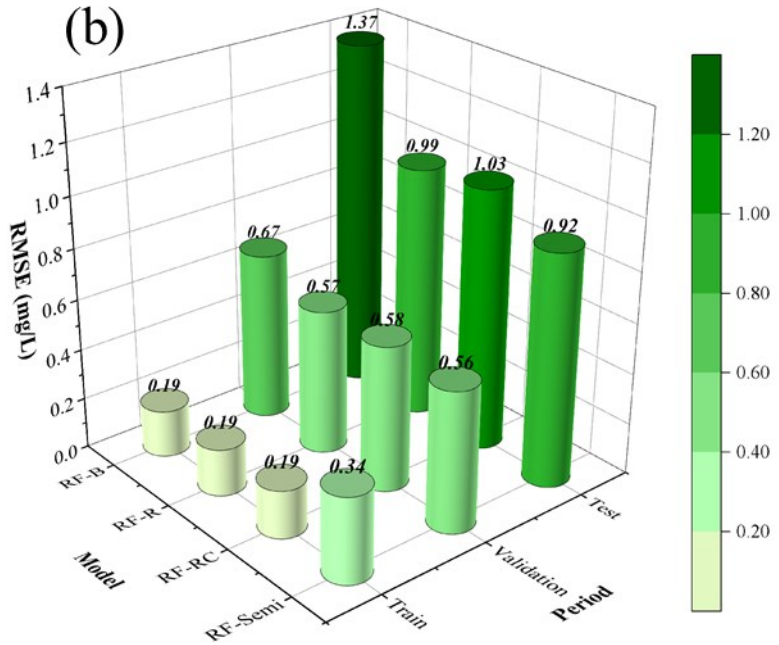


Fig. 6-4. NSE of RF for (a) TSS; (b) TN; (c) TP.


(a)
RMSE (mg/L)
0
20
40
60
80
100
120
140
160
167
127
162
104
88
75
69
56
45
74
48
24
RF-B
RF-R
RF-RC
RF-Semi
Model
Train
Validation
Test
Period


(b)
RMSE (mg/L)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.37
0.99
1.03
0.92
0.67
0.57
0.58
0.56
0.19
0.19
0.19
0.34
RF-B
RF-R
RF-RC
RF-Semi
Model
Train
Validation
Test
Period

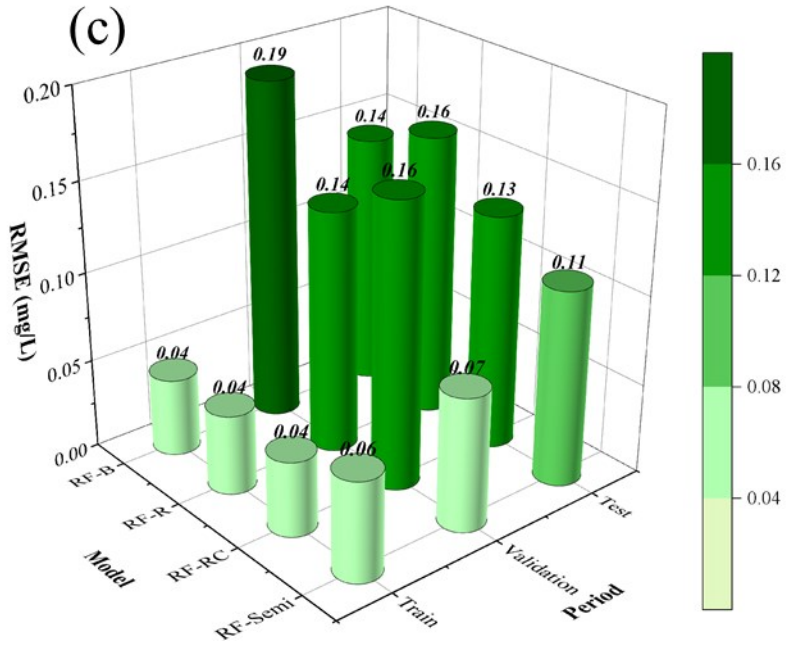


Fig. 6-5. RMSE of RF for (a) TSS; (b) TN; (c) TP.

6.3.3 Comparison with previous studies

Machine learning prediction have been widely applied in environmental research. Although water quality data is not readily available, recently some studies attempt to apply machine learning in prediction of stormwater and stream water quality. A comparison of model performance was made with recently published studies (Table 6-1).

Moeini et al., (2021) used six features (drainage area, land use, imperviousness, rainfall depth, runoff volume, and antecedent dry days) to train random forest model with TSS concentration as the target parameter. This study shows a comparable NSE (0.79 vs. 0.76) during training period and a slightly lower NSE (0.62 vs. 0.67) during testing period than Moeini et al. (2021) study. Considering the larger range of TSS concentration in their study, the RMSE for testing dataset is also comparable (69 mg/L vs. 150 mg/L). It is worth noting that Moeini et al. (2021) study used a larger dataset with 530 events from 7 states in the USA, which explains whey they were able to obtain good model performance despite using only six features as input variables.

Wang et al. (2021) utilized a dataset comprising 868 events from Texas, USA, to predict TP concentrations in stream using a random forest model. This study considered various factors that could explain the variation in TSS concentration, such as the characteristics of the land cover, the

climatic conditions, the soil properties, the terrain features, and the human population density. Their study achieved an NSE of 0.91 during training and 0.61 in testing, with an RMSE of 0.88 mg/L for the testing dataset. In contrast, our study achieved a slightly lower NSE (0.86) during training period but demonstrated a close NSE (0.62) during testing period. Our model exhibited enhanced precision in the testing period, with an RMSE of 0.11 mg/L for TP, which is substantially lower than that reported by Wang et al (2021). This is attributed that their study covered a broad TP concentration range (0-4 mg/L).

Behrouz et al. (2022) utilized NSQD dataset gathered from over 300 cities across the United States, encompassing more than 5000 events to train random forest models for predicting TSS, TN, and TP. In their study, the considered input variables include climatological characteristics (i.e., Antecedent Dry Period or ADP, Precipitation Depth or P, Duration or D, and Intensity or I) and catchment characteristics including land use-related parameters including Imperviousness or Imp, Saturated Hydraulic Conductivity or Ksat, and Available Water Capacity or AWC), and site-specific parameters including Slope (S), and Catchment Size (A). They reported NSE values of 0.79 (TSS), 0.81 (TN), and 0.79 (TP) during training period, with a obvious drop in the NSE to 0.3 for TSS, 0.35 for TN, and 0.33 for TP during testing period. This study presents similar NSEs for TSS (0.79) and TP (0.86) during training period, but substantially improved NSEs of 0.62 during testing period, indicating better generalization to unseen data. Conversely, the prediction of TN underperforms (NSE of -0.03 during testing period).

Kshirsagar et al. (2023) trained a Support Vector Regression (SVR) model to predict TSS in Pune City, India. The samples were collected for 22 storm events, which is only study that utilized less data than us. Rainfall and Antecedent Dry days (ADD) are the independent variables of SVR model. They reported a RMSE of 278 mg/L for testing dataset, which is larger than our RMSE of 69 mg/L for TSS. Despite the difference in data sizes and environmental conditions, our model demonstrates better accuracy.

Jalagam et al. (2023) developed a Random Forest model for TSS prediction in urban streams. Land use data and precipitation data were utilized as input variables. They reported RMSE of 59 mg/L for validation and a RMSE of 25 mg/L for testing dataset. Our study yielded a lower RMSE of 24 mg/L for validation and a slightly higher test RMSE of 69 mg/L for testing dataset.

Overall, the previous studies primarily utilized labeled datasets for water quality analysis, without incorporating unlabeled data. Their model performances were influenced by the size of the datasets, with larger datasets generally contributing to better model training and potentially improved prediction accuracy. Our semi-supervised approach that utilized both labeled and unlabeled data indicates that data utilization efficiency could also play critical roles in application of machine learning in data-deficient areas

Table 6-1. Comparison of model performance of machine learning with recent studies.

Model	Water quality parameters	Period	Evaluation index		Range of observed water quality parameters	Database	Sources
			NSE	RMSE			
RF	TSS	Train	0.76	110	0-1100 mg/L	530 events from 7 states in USA	(Moeini et al., 2021)
		Test	0.67	150			
RF	TP	Train	0.91	0.40	0-4 mg/L	868 datapoints from Texas in USA	(Wang et al., 2021)
		Test	0.61	0.88			
RF	TSS	Train	0.79			Over 5000 events from over 300 cities in USA	(Behrouz et al., 2022)
		Test	0.30				
	TN	Train	0.81				
		Test	0.35				
	TP	Train	0.79				
		Test	0.33				
SVR	TSS	Test		278	Mean: 927 mg/L	22 events from Pune City in India	(Kshirsagar et al., 2023)
RF	TSS	Validation		59	0-600 mg/L	Stream water quality from North Carolina, USA	(Jalagam et al., 2023)
		Test		25			
RF	TSS	Train	0.79	48	11-593 mg/L	57 events from Calgary in Canada	Present
		Validation	0.85	24			
		Test	0.62	69			
	TN	Train	0.79	0.34	0.41-3.74 mg/L		
		Validation	0.38	0.56			
		Test	-0.03	0.92			
	TP	Train	0.86	0.06	0.03-0.66 mg/L		
		Validation	0.68	0.07			
		Test	0.62	0.11			

Note: The empty space in this table indicates the variables are not mentioned from the sources. RF is Random forest model, SVR is support vector regression model.

6.3.4 Variable importance

The variable importance (Fig. 6-6) provides insight into the contribution of each input variable in the semi-supervised random forest model to predict EMCs of TSS, TN and TP.

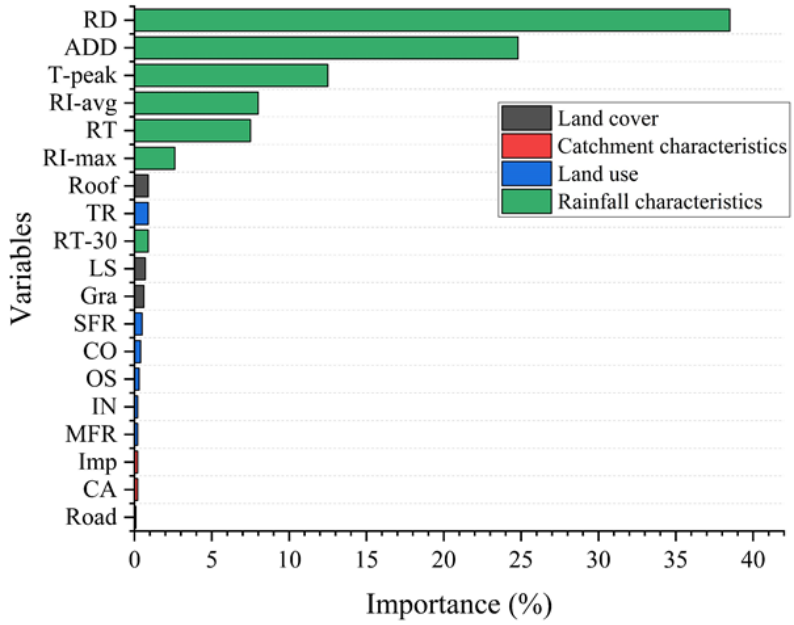
Rainfall duration (RD) is the most influential variable for prediction of EMCs of TSS, accounting for 38.5% of the importance, followed by antecedent dry days (ADD) at 24.8% (Fig. 6-6a). This suggests that both the length of the rainfall and the antecedent dry period impact TSS concentrations in stormwater, likely due to the accumulation and subsequent wash-off of suspended particles (Salim et al., 2019). Detailed land use/land cover categories like single-family residential (SFR) and transportation (TR) have lesser importance for predicting TSS.

The total rainfall amount (RT) has the highest importance (28.7%) for prediction of EMCs of TN (Fig. 6-6b), indicating the overall volume of rainfall contributes to nitrogen runoff. RD and ADD are also crucial, indicating that both the duration of the rainfall and the conditions prior to the storm are key factors affecting nitrogen transport (Murphy et al., 2015).

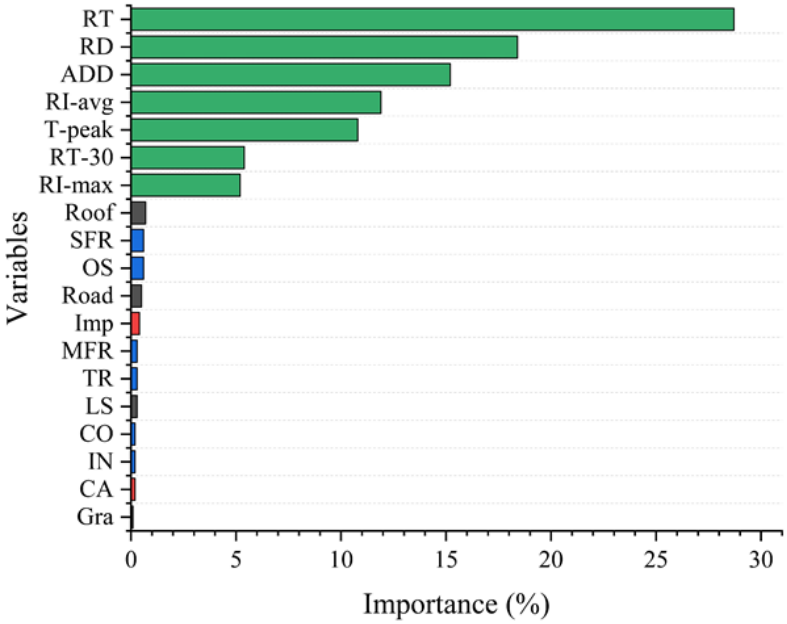
For TP prediction (Fig. 6-6c), RT again plays an important role (25.8%), followed by maximum rainfall intensity (RI-max) at 13.0% and average rainfall intensity (RI-avg) at 13.3%. This emphasizes that the intensity of the rainfall, along with the volume, is vital in predicting TP concentrations, as heavier rainfalls can lead to increased runoff and phosphorus loading (Li et al., 2015). Catchment area (CA) and Imperviousness (Imp) has a noticeable importance for TP, which is not as evident for TSS and TN, suggesting that the size of the catchment may influence phosphorus levels more than nitrogen or suspended solids. Behrouz et al., (2022) reported that land use-related characteristics were effective variables for predicting EMCs for TP.

In summary, rainfall characteristics are consistently important across all three pollutants, with varying focus on duration, amount, and intensity, indicating their crucial role in the mobilization and transport of pollutants. ADD also plays an important role in predicting pollutant levels, particularly for TSS and TN. The inclusion of catchment characteristics, such as prediction for TP, suggests that catchment-specific factors can also be influential and should be carefully considered in urban water quality modeling using machine learning.

(a)


RD
ADD
T-peak
RI-avg
RT
RI-max
Roof
TR
RT-30
LS
Gra
SFR
CO
OS
IN
MFR
Imp
CA
Road
Variables
Land cover
Catchment characteristics
Land use
Rainfall characteristics
0
5
10
15
20
25
30
35
40
Importance (%)

(b)


RT
RD
ADD
RI-avg
T-peak
RT-30
RI-max
Roof
SFR
OS
Road
Imp
MFR
TR
LS
CO
IN
CA
Gra
Variables
0
5
10
15
20
25
30
Importance (%)

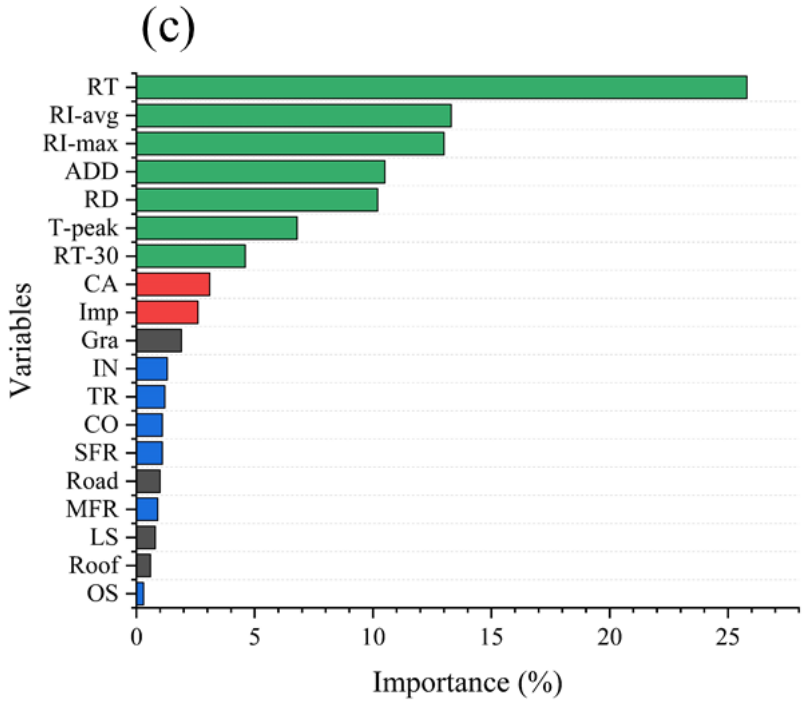


Fig. 6-6. Variable importance of RF-Semi model for predicting (a) TSS; (b) TN; (c) TP.

6.4 Conclusions

This paper proposed a semi-supervised machine learning (random forest) framework to predict urban stormwater quality in data-deficient areas. This approach was applied in four catchments in Calgary, Alberta, Canada to predict event mean concentration (EMC) of Total Suspended Solids (TSS), Total Nitrogen (TN) and Total Phosphorus (TP). The main findings and conclusions are as follows:

(1) The results show that machine learning is an effective tool for predicting EMCs of TSS and TP. Machine learning approach could capture the complex interplay of factors that influence stormwater quality and model the non-linear relationships.

(2) An over-reliance on limited datasets and the selection of rainfall as single input features could potentially lead to overfitting. When only rainfall characteristics in one catchment were selects as input features to train RF model, the model performed well for prediction of TSS during training period (NSE=0.90) but failed on testing dataset (NSE=-0.76). This limitation of overfitting could

obstruct the application of the ML models due to decreased predictive effectiveness on unseen data.

(3) The augmentation of training dataset with information from multiple catchments effectively improved the model predicting accuracy although the performance during training and validation period decreased. The introduction of catchment characteristics as input variables within the ML models could further enhance model capacity for generalization. This finding emphasizes the potential benefits of integrating catchment-specific data into predictive modeling.

(4) The application of pseudo-labeling as a semi-supervised learning technique has yielded positive results by enriching the training dataset. This approach utilizes the model predictions to generate labels for unlabeled data, thereby expanding the training set and providing additional information that could refine the model learning process. The semi-supervised machine learning strategy has proven to enhance the model predictive accuracy.

(5) The variable importance shows that rainfall characteristics are important variables for predicting TSS, TN and TP. Rainfall duration, total rainfall, and rainfall intensity played crucial role in the mobilization and transport of pollutants. ADD also plays an important role in predicting pollutant levels, particularly for TSS and TN.

In conclusion, while machine learning models are effective tools for stormwater quality prediction, the model performance is impacted by the availability and quality of the dataset. The integration of comprehensive datasets, including catchment characteristics, along with semi-supervised machine learning techniques such as pseudo-labeling, could improve the accuracy and applicability of predictive models in data-deficient areas.

7. Practical application in urban stormwater quality management and design

This thesis conducted a research based on the field monitoring data in Calgary to study the response of urban stormwater quality to rainfall characteristics and LULC and the improvement of performance of urban stormwater quality model. The findings have implications for effective stormwater management and design that account for the specific pollutant contributions of different land uses and the variability in rainfall characteristics. In addition, this research has potential to enhance stormwater quality modeling and prediction, which is crucial for effective urban water management and planning. This chapter will discuss the practical application of the research for Calgary from three aspects: water quality design event, particle size distribution and water quality modelling.

7.1 Water quality design event

7.1.1 General introduction of current approach in Calgary stormwater management and design manual

Selection of water quality design event is crucial when designing stormwater control practices (SCPs) or best management practices (BMPs) such as bioretention areas, wet ponds. Water quality design event is used generate pollutographs using computer water quality model. In Calgary, the Intensity-Duration-Frequency (IDF) curve is utilized for selection of design event. A Chicago distribution is the applied to formulate the synthetic design storm. A 1:100 year storm event is used in the design of the major system, while runoff from more frequent events should be controlled and treated to improve water quality of the runoff. Two key considerations pertaining to the analysis and design of SCPs include the following: 1) For Calgary, the water quality design event has a rainfall depth of 15 mm. 2) Typically, a duration of 1 hour is used to examine the operation of SCPs. 3) A 1:5 year event must be used to quantify any benefits of these SCPs.

However, the IDF curve derived from extreme rainfall intensities for a specific duration through annual maximum analysis, which does not reflect the average rainfall level of the year. The design event chosen should be responsible for generating the majority of pollutants from a catchment. This choice enhances the efficiency of the stormwater treatment system, as events with higher

pollutant production are targeted. It is not clear if the design event from IDF curve can contribute most of the pollutant generated from a catchment. In addition, the EMCs from the design event need to be estimated by computer water quality models (e.g., process-based models), which are relatively complex and tend to be less accurate than quantity models. Therefore, a simpler alternative of process-based models with good accurate is needed. The following sections will evaluate the contribution of different rainfall types on pollutant loads and provide an approach to estimating EMCs for design event.

7.1.2 Contribution of different rainfall event types on pollutant loads.

The pollutant loads from four catchments during four rainfall event types were estimated by the following formula:

$$Loads = \sum A \cdot P_i \cdot \alpha \cdot EMC_i \tag{7-1}$$

where A is the catchment area, P is the total precipitation, α is the rainfall-runoff coefficient, EMC is the event mean concentration, i is the number of target rainfall type events.

For example, to calculate TSS load during Type I events, the pollutant loads from the four catchments during all Type I events need to be added together. Fig. 7-1 shows the pollutant loads from four catchments during four rainfall event types. For TSS loads, although the number of Type II events only account for 16% (Table 2-3) of all rainfall events, they contributed to 60% of the total TSS loads. Type I events are the most frequent events (30 out of 57 events), while they only generated 14% of the total TSS loads. Similarly, Type II events are major contributors to TN and TP loads (42% and 55%, respectively).

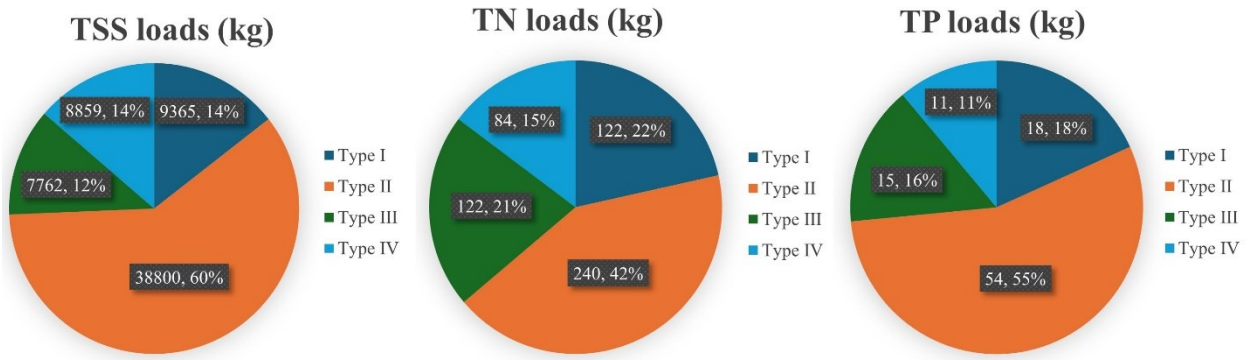


Fig. 7-1. Contribution of different rainfall event types on pollutant loads.

Fig. 7-2 shows the IDF curves (used in Calgary stormwater management and design) combined with observed four rainfall type events. The IDF curve is derived from Meteorological Service of Canada (MSC) rainfall data taken from the International Airport. Rainfall, which has been collected from 1947 to 1998 (48 years), has been analyzed using Gumbel distribution. As shown in Fig. 7-2, Type I and Type IV events are distributed below the 2-year IDF curve and 77% of Type III events (10 out of 13 events) are also distributed below 2-year IDF curves. Observing the distribution of Type II events in relation to the IDF curves, it appears that these events generally align with, or closely approximate the IDF curves (except for 5-year and 10-yearcurves). This pattern may be attributed to the relatively short monitoring period (only two years), which could limit the representativeness of the data for these specific return periods. The distribution of Type II events indicates that they can be somewhat representative of events within specific return periods. As shown in Fig. 7-1, Type II events can contribute most of the pollutant generated from catchments. Therefore, events from IDF curves could be suitable water quality design events for effective SCPs or BMPs design.

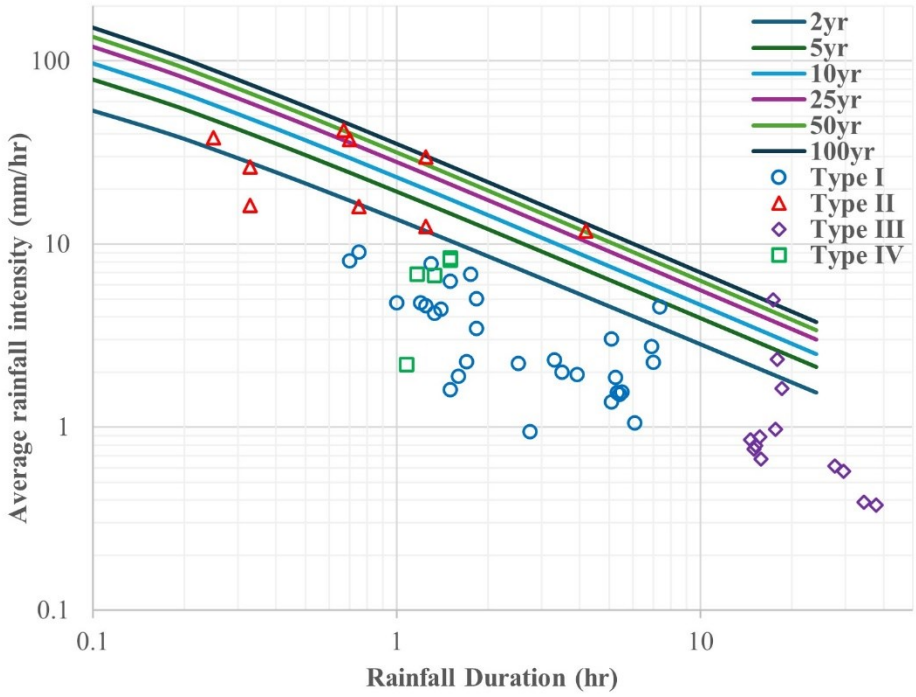


Fig. 7-2. IDF curves combined with observed rainfall events.

7.1.3 Estimation of EMCs for design event

To estimate the EMCs for water quality design event, the developed semi-supervised machine learning approach in Chapter 6 is selected as a simpler alternative tool of process-driven models.

Based on the requirements of Calgary stormwater management and design, a 1:5 year event is selected as water quality design event with a duration of 1 hour. From the adjusted MSC IDF Curve, the design event has a rainfall depth of 15 mm. The Chicago Hydrograph Method is used to create a rainfall distribution curve for the selected design event. The created hyetograph is shown in Fig. 7-3. Based on the Chicago distribution of the design event, the rainfall input features for semi-supervised machine learning approach are determined: average rainfall intensity: 19.40 mm/hr, total rainfall: 19.40 mm, rainfall duration: 1 hr, maximum rainfall intensity: 87.48 mm/hr, time to peak rainfall intensity: 0.3 hr, the initial 30-min total rainfall amount: 14.83 mm. The antecedent dry period is selected as the average value of Type II events: 6.6 day.

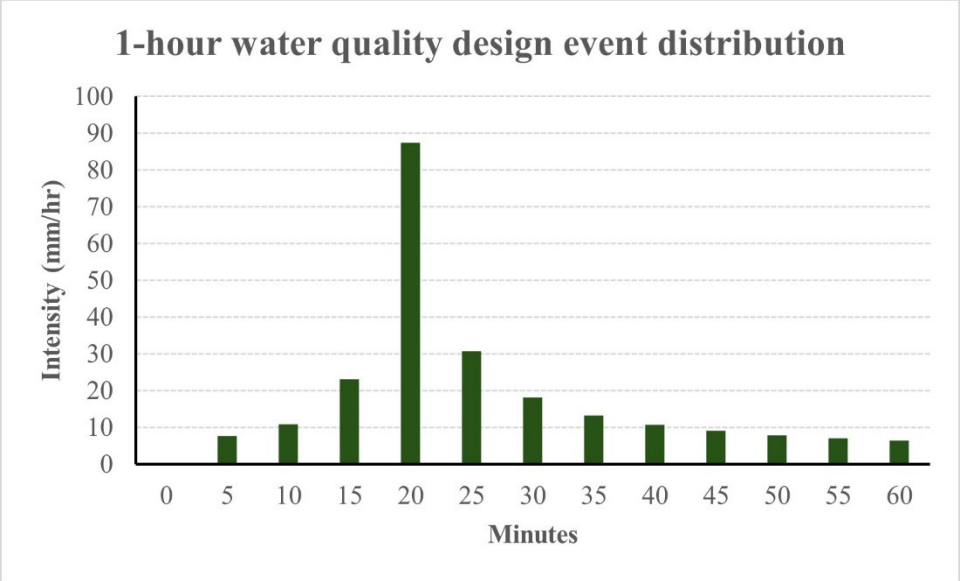


Fig. 7-3. Chicago distribution of the design event.

To predict the EMCs of TSS, TN and TP for the design event using machine learning approach, Royal Oak catchment is selected as an example, whose catchment characteristics are selected as catchment input features. The predicted EMCs of TSS, TN and TP are 226 mg/L, 1.71 mg/L and 0.27 mg/L (Table 7-1). To evaluate if the predicted EMC values are reasonable, two observed events (occurred on Royal Oak catchment) that are most likely close to the design event in IDF cures (Fig. 7-2) are selected. These two events are located at the 1:2 year IDF curve. The duration and average intensity of two events are 0.8, 16.0 and 1.3, 12.5, respectively. The average EMCs values of two observed events are shown in table 7-1. The relative error for TSS, TN and TP are less than 25%, which means that the predicted EMC values are slightly larger than the average of observed events. This is reasonable because the design event with higher rainfall intensity can carry more pollutants.

Table 7-1. Estimated EMCs by machine learning and average observed events.

EMC (mg/L)	ML	Average of observed events	Relative error (%)
TSS	226	181	25
TN	1.71	1.37	25
TP	0.27	0.23	17

7.2 Particle size distribution

The particle size distribution (PSD) of suspended sediments in urban runoff has been considered an important factor affecting the transport and fate of sediment and the associated pollutants. PSDs of suspended sediments in urban runoff are necessary for the effective treatment of urban stormwater. It is necessary to consider the PSD for soils and their settling velocities.

7.2.1 PSD curves for different land use types

To estimate the total suspended solids removal rate from wet ponds, wetlands, and other devices or technologies requiring sediment removal, City of Calgary provides Particle Size and Settling Velocities. The particle size distribution follows the New Jersey Department of Environmental Protection curve. This PSD curve was compared to the PSD curves derived from the water samples in 15 urban areas in Calgary, Alberta, Canada, during 179 rainfall and 118 snowmelt events from 2016 to 2021 (Fig. 7-4). PSD curve of Calgary design manual is between the lower and upper envelopes of our study. For the particles finer than 50 μm, Calgary design manual has similar percent as gravel lane residential land use. This indicates that the PSD curve of Calgary design manual can represent the average particle size distribution. However, when BMPs are designed for specific land use type, the PSD curves from Calgary field monitoring program are more suitable for effective design. The PSD curves in Chapter 5 can provide reference for effective BMPs design.

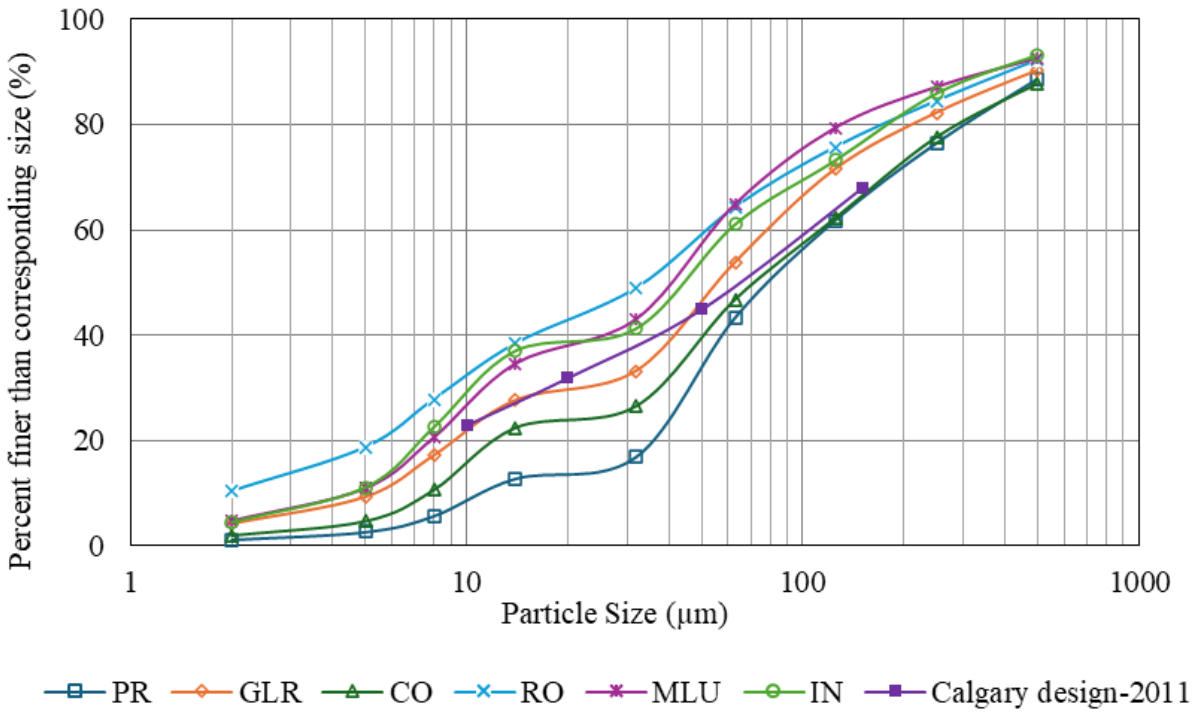


Fig. 7-4. Proportions of PSDs from different land uses.

7.2.2 Recommendation of TSS removal criteria for specifc conditions

Current stormwater quality criteria for the City of Calgary requires the removal of a minimum of 85% Total Suspended Solids (TSS) for particle sizes greater than, or equal to, 50 µm. Based on the PSD curves in Fig. 7-4, the proportion of particles sizes greater than, or equal to, 50 µm for paved residential land use is approximately 65%, which indicates most particles can be removed following the TSS removal criteria. However, only 40% particles in mixed land use are greater than, or equal to, 50 µm. Large proportion of particles will remain in runoff if BMPs follow the current removal criteria. Therefore, appropriately adjusting removal criteria for specific land use types can contribute to effective removal of TSS. In addition, this research found that PSDs of different land use types vary with rainfall event types. Type II events align with water quality design event, so PSDs from different land uses under Type II events (Table 7-2) are suitable for BMPs design.

In general, treatment ponds are effective in removing solids, but less effective for removing BODs, phosphorus, and nitrogen. Many studies have found that finer particles have affinity for various and large amounts of pollutants (e.g., phosphorus, and nitrogen). This research found that the EMC

of TSS was highly correlated with Non-Reactive Phosphorus (0.87) and Total Phosphorus (0.88), and moderately with Total Organic Nitrogen (0.53), Total Kjeldahl Nitrogen (0.56), and Total Nitrogen (0.54), suggesting these phosphorus and nitrogen forms may exist primarily in particulate form. Therefore, increasing forebay's TSS removal criteria could help reduce nitrogen and phosphorus concentrations through sedimentation.

Table 7-2. PSDs from different land uses under Type II events as accumulated volume fraction %.

Land Use	<500 (µm)	<250 (µm)	<125 (µm)	<63 (µm)	<32 (µm)	<14 (µm)	<8 (µm)	<5 (µm)	<2 (µm)
Gravel_lane Residential	95.7	92.8	84.5	64.1	37.6	34.5	20.6	9.3	3.6
Commercial	86.6	74.8	57.6	42.5	22.8	18.9	7.8	3.0	1.2
Roads	90.4	79.8	70.0	59.3	48.6	39.3	29.9	21.1	12.7
Mixed land use	92.2	86.9	80.7	67.5	44.1	33.1	19.6	10.5	4.4

7.3 Water quality modelling

Urban stormwater quality models are essential tools for effective stormwater management and design. This research attempts to improve water quality modelling performance through catchment delineation approach and model calibration approach. The research findings could provide some insight into model improvement.

7.3.1 Singel event vs continuous simulation

Stormwater computer models can be used to model a drainage system for either single or continuous rainfall events. Continuous simulation is required to model urban runoff water quality from all wet ponds and wetlands to estimate overall sediment removal rates (TSS). To achieve the total loading objectives for stormwater and wastewater, which were developed as part of the Total Loading Objective Assessment (TLOA) project by City of Calgary, a City-Wide Stormwater modelling work was conducted by Tetra Tech, Inc. The goal of the continuous model calibration is to obtain low relative error when all the calibration sites are considered. The TSS calibration generated results with an overall relative error for TSS loading estimation at 0.2%. However, the

continuous model may result in large error from the point view of events. The rainfall event type-based calibration approach established in Chapter 2 calibrate water quality parameter for events of a single rainfall type during a continuous simulation. This approach will provide the optimal water quality parameter set for different rainfall types. This calibration approach could take advantage of both continuous simulation and event-based simulation. The calibrated Build-up and Wash-off Coefficients for different rainfall types (Table 2-5) could be reference for related modelling work in Calgary.

7.3.2 Model parameter transferability

Collecting stormwater quality data, is often time-consuming and expensive, leading to a lack of available stormwater quality data for model calibration. Therefore, it is anticipated that the calibrated stormwater model could be applied in ungauged catchments to predict hydrographs and associated pollutant loads, mitigating the data shortfall. In the current Calgary design manual, little calibration information is available to verify the accuracy of the parameters. This research tested parameters transferability of land cover-based catchment delineation approach. The results indicates that the hydrologic-hydraulic parameters can be satisfactorily transferred from neighboring gauged catchments to similar ungauged catchments. However, the transferring of water quality parameters did not perform as satisfactory. Therefore, even in adjacent catchments with similar catchment characteristics, the transfer of water quality parameters should be done cautiously. Much more efforts should be taken to calibrate water quality model parameters that can be widely used in Calgary catchments.

8. General conclusions and recommendations for future research

8.1 General conclusions

This thesis studied the response of urban stormwater quality to rainfall characteristics and LULC and the improvement of performance of urban stormwater quality model from five major aspects. Detailed conclusions can be found in each of the preceding five major chapters. Specific conclusions for the study catchments in Calgary are summarized as follows:

In Chapter 2, the impact of rainfall characteristics on urban stormwater quality was studied using a data mining framework. The K-means method effectively clustered the rainfall events into four types that could represent the rainfall characteristics in the study areas. Compared to the traditional continuous simulation model, the relative error of the rainfall type-based (RTB) approach was reduced by 11.4% to 16.4% over the calibration period for the four catchments in Calgary. The calibrated stormwater quality parameters could be transferred to adjacent catchments with similar characteristics in Calgary.

In Chapter 3, the impacts of land cover on the simulation of stormwater runoff and pollutant loading were investigated. The land cover-based approach's better performance in hydrological simulation and its ability to be regionalized based on land cover types can be applied to similar urban catchments in Calgary. Transferring of hydrologic-hydraulic parameters from gauged to ungauged catchments in Calgary is feasible. In Calgary, roads and roofs were major contributors to urban runoff and pollutants, while green spaces became important contributor during large storm events.

In Chapter 4, the effects of mixed land use on urban stormwater quality under different rainfall event types was studied using the database from a two-year field monitoring program in Calgary. The correlation analysis showed that the event mean concentrations (EMCs) of total suspended solids (TSS) was positively correlated with most nutrient components, and event pollutant loads (EPLs) exhibited higher correlation compared to EMCs in Calgary. The differences in EMCs across four catchments indicated that mixed land use could influence the generation of stormwater pollutants in Calgary, with spatial configurations affecting EMCs. Seasonal variations were found in EMC and EPL, with higher values in the spring and summer compared to the fall in Calgary.

In Chapter 5, the effects of land use and rainfall conditions on the particle size distribution (PSD) of suspended sediments in stormwater and snowmelt runoff was investigated using dataset from a six-year water sampling program across 15 study sites in Calgary. The results revealed that the median particle size decreased in the order: paved residential, commercial, gravel lane residential, mixed land use, industrial, and roads. Fine particles less than 125 μm are the dominant particles (over 75%) of suspended sediments in runoff in Calgary. Paved roads have the largest percentage of particles finer than 32 μm (49%) due to vehicle emissions, tire wear, and road abrasion. Gravel lane residential areas have great proportion of fine particle sizes compared to paved residential areas due to the rough properties of gravel lanes. The season is a factor in the PSD of suspended sediment in runoff in Calgary. The PSD in spring exhibits the finest particles due to the street sweeping. Particles in snowmelt are finer for the same land use than that during rainfall events because the rainfall-runoff flows are usually an order of magnitude larger than the snowmelt flows.

In Chapter 6, a semi-supervised machine learning framework was proposed to predict urban stormwater quality in data-deficient areas. Semi-supervised machine learning using pseudo-labeling could effectively predict stormwater quality in Calgary, even with limited labeled data. Rainfall characteristics were more significant than catchment characteristics in predicting urban stormwater quality in Calgary.

In Chapter 7, the practical application of this research outcomes in urban stormwater quality management and design was explored in Calgary. It is important to consider specific rainfall event types and land use in stormwater management strategies. For instance, Type II events, though less frequent, contribute significantly to total pollutant loads, emphasizing the need for targeted strategies in urban stormwater management. PSD of suspended sediments is an important factor for effective stormwater treatment, and it is recommended to adjust TSS removal criteria based on specific land use types.

More general conclusions are summarized as follows:

The results revealed that antecedent dry days, average rainfall intensity, and rainfall duration were the most critical rainfall characteristics affecting the EMCs of TSS, total nitrogen (TN), and total phosphorus (TP), while total rainfall was found to be of negligible importance for the four catchments. The RTB calibration approach could improve PCSWMM model accuracy in simulating water quality model. The Land Cover-Based (LCB) approach can effectively evaluate

the contribution of different land covers to runoff and pollutant loads. Intense rainfall and long antecedent dry days could yield higher EMC and EPL, whereas long rainfall duration could generate lower EMC and higher EPL. The impact of rainfall event types on PSD in runoff could vary depending on land use types. A long antecedent dry period tends to result in the accumulation of fine particles. High rainfall intensity and long duration could wash off more coarse particles. Relying on limited dataset and input features may lead to overfitting, which affects model predictive effectiveness. The semi-supervised machine learning framework can be generalized for predicting stormwater quality in other urban catchments, emphasizing the importance of integrating datasets from multiple catchments and considering catchment characteristics for model generalization.

8.2 Recommendations for future research

Overall, this thesis contributes to the knowledge of response of urban stormwater quality to rainfall characteristics and LULC and the improvement of performance of urban stormwater quality model. However, there are still many aspects left to continued research in these broad areas. The followings are recommended for future study:

1. This research mainly based on the dataset from a two-year field monitoring program in four urban catchments in Calgary. To capture a broader range of weather patterns and their impact on stormwater quality, extending the monitoring program beyond two years could provide a more comprehensive dataset.
2. In this study, monitoring sites were set at the outlets of catchments to ensure that the collected data can represented the pollutant levels across the entire catchments. Understanding the spatial distribution of pollutants within urban catchments will benefit for identifying the specific contributions of different land use/land cover types to overall pollution form the point view of catchments. Therefore, it is recommended to set monitoring sites at different land use/land cover types within the same catchment to collect and analyze runoff samples more comprehensively.
3. While the current study utilized data from Calgary, it would be beneficial to include more geographically diverse catchments to enhance the model's robustness and applicability to various urban environments.

4. There are some sources of uncertainties in PCSWMM in this research including input uncertainty, parameter uncertainty and model structure uncertainty. For example, the rainfall data from the solo rain gauge in large catchment (e.g., Auburn Bay) may not represent the average rainfall level, which could introduce input uncertainty. The representative of pollutant build-up and wash-off process using four parameters in PCSWMM could bring model structure uncertainty. The uncertainty analysis can provide useful information for evaluating model reliability. Therefore, uncertainty quantification in urban stormwater quality modelling is recommended.
5. Monitoring water quality parameters is crucial for urban stormwater quality control. Monitoring large number of water quality parameters are often time-consuming and expensive. Selection of suitable parameters that can be reprehensive of other parameter could be a potential cost-saving strategy. For example, the strong correlation between TON and TKN suggest that monitoring TON might be a suitable representative parameter of TKN in the water quality monitoring.
6. PSD can affect pollutant transport and their fate in receiving water. Further research on PSD should be along with pollutant concentrations and loading.
7. Random forest algorithm is selected as foundational machine learning approach, while other machine learning approaches also have their distinct advantages. In the future work, comparing model performance between different approaches is recommended.

Bibliography

Abdi, H., Williams, L.J., 2010. Tukey's honestly significant difference (HSD) test. Encycl. Res. Des. 3, 1–5.

Ahilan, S., Guan, M., Wright, N., Sleigh, A., Allen, D., Arthur, S., Haynes, H., Krivtsov, V., 2019. Modelling the long-term suspended sedimentological effects on stormwater pond performance in an urban catchment. J. Hydrol. 571, 805–818.

Ahmed, A.N., Othman, F.B., Afan, H.A., Ibrahim, R.K., Fai, C.M., Hossain, M.S., Ehteram, M., Elshafie, A., 2019. Machine learning methods for better water quality prediction. J. Hydrol. 578, 124084.

Al Bakri, D., Rahman, S., Bowling, L., 2008. Sources and management of urban stormwater pollution in rural catchments, Australia. J. Hydrol. 356, 299–311.

Alin, A., 2010. Multicollinearity. Wiley interdisciplinary reviews: computational statistics 2(3), 370-374.

Amato, F., Querol, X., Johansson, C., Nagl, C., Alastuey, A., 2010. A review on the effectiveness of street sweeping, washing and dust suppressants as urban PM control methods. Sci. Total Environ. 408, 3070–3084.

Anta, J., Peña, E., Suárez, J., Cagiao, J., 2006. A BMP selection process based on the granulometry of runoff solids in a separate urban catchment. Water Sa. 32, 419–428.

Aryal, R., Vigneswaran, S., Kandasamy, J., Naidu, R., 2010. Urban stormwater quality and treatment. Korean J. Chem. Eng. 27, 1343–1359.

Ashley, R.M., Bertrand-Krajewski, J.-L., Hvitved-Jacobsen, T., Verbanck, M., 2004. Solids in sewers. IWA Publishing.

Bach, P.M., McCarthy, D.T., Deletic, A., 2010. Redefining the stormwater first flush phenomenon. Water Res. 44(8), 2487-2498.

Bannerman, R.T., Owens, D.W., Dodds, R., Hornewer, N.J., 1993. Sources of pollutants in Wisconsin stormwater. Water Sci. Technol. 28, 241–259.

Bárdossy, A., 2007. Calibration of hydrological model parameters for ungauged catchments. Hydrol. Earth Syst. Sci. 11, 703–710.

Barthélémy, N., Sarremejane, R., Datry, T., 2022. Aquatic organic matter decomposition in the terrestrial environments of an intermittent headwater stream. Aquat. Sci. 84, 45.

Bathi, J.R., Pitt, R., Clark, S.E., 2009. Associations of PAHs with size fractionated sediment particles. 2009 World Environmental and Water Resources Congress: Great Rivers. 1–10.

Beck, Scott M., Melissa R. McHale, and George R. Hess, 2016. Beyond Impervious: Urban Land-Cover Pattern Variation and Implications for Water Quality. Environ. Manage. 58, 15-30.

Behrouz, M.S., Yazdi, M.N., Sample, D.J., 2022. Using Random Forest, a machine learning approach to predict nitrogen, phosphorus, and sediment event mean concentrations in urban runoff. J. Environ. Manage. 317, 115412.

Behrouz, M.S., Yazdi, M.N., Sample, D.J., Scott, D., Owen Jr, J.S., 2022. What are the relevant sources and factors affecting event mean concentrations (EMCs) of nutrients and sediment in stormwater? Sci. Total Environ. 828, 154368.

Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C.A., 2019. Mixmatch: A holistic approach to semi-supervised learning. Adv. Neural Inf. Process. Syst. 32.

Bian, B., Zhu, W., 2009. Particle size distribution and pollutants in road-deposited sediments in different areas of Zhenjiang, China. Environ. Geochem. Health 31, 511–520.

Bonhomme, C., Petrucci, G., 2017. Should we trust build-up/wash-off water quality models at the scale of urban catchments? Water Res. 108, 422–431.

Borris, M., Österlund, H., Marsalek, J., Viklander, M., 2016. Contribution of coarse particles from road surfaces to dissolved and particle-bound heavy metal loads in runoff: A laboratory leaching study with synthetic stormwater. Sci. Total Environ. 573, 212–221.

Borris, M., Viklander, M., Gustafsson, A.M. and Marsalek, J., 2014. Modelling the effects of changes in rainfall event characteristics on TSS loads in urban runoff. Hydrol. Processes 28(4), 1787-1796.

Breault, R.F., Smith, K.P., Sorenson, J.R., 2005. Residential street-dirt accumulation rates and chemical composition, and removal efficiencies by mechanical-and vacuum-type sweepers, New Bedford, Massachusetts, 2003-04. Department of the Interior, U.S. Geological Survey, Reston, VA. 1–27.

Brezonik, P.L., Stadelmann, T.H., 2002. Analysis and predictive models of stormwater runoff volumes, loads, and pollutant concentrations from watersheds in the Twin Cities metropolitan area, Minnesota, USA. Water Res. 36, 1743–1757.

Burton Jr, G.A., Pitt, R., 2001. Stormwater effects handbook: A toolbox for watershed managers, scientists, and engineers. CRC/Lewis Publishers, Boca Raton, FL.

Cahn, M., Hartz, T., 2014. Load vs. concentration: implications for reaching water quality goals. Univ. Calif. Coop. Ext. Monterey Cty.

Carroll, S., Liu, A., Dawes, L., Hargreaves, M., Goonetilleke, A., 2013. Role of land use and seasonal factors in water quality degradations. Water Resour. Manag. 27, 3433–3440.

Charters, F. J., Cochrane, T. A., O'Sullivan, A. D., 2020. Predicting event-based sediment and heavy metal loads in untreated urban runoff from impermeable surfaces. Water. 12(4), 969.

Charters, F. J., Cochrane, T. A., O'Sullivan, A. D., 2021. The influence of urban surface type and characteristics on runoff water quality. Sci. Total Environ. 755, 142470.

Charters, F.J., Cochrane, T.A., O'Sullivan, A.D., 2015. Particle size distribution variance in untreated urban runoff and its implication on treatment selection. Water Res. 85, 337–345.

Chaudhary, S., Chua, L.H. and Kansal, A.,2022. Event mean concentration and first flush from residential catchments in different climate zones. Water Res., 118594.

Chen, L., Zhi, X., Shen, Z., Dai, Y., Aini, G., 2018. Comparison between snowmelt-runoff and rainfall-runoff nonpoint source pollution in a typical urban catchment in Beijing, China. Environ. Sci. Pollut. Res. 25, 2377–2388.

Choi, H.T., Beven, K., 2007. Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework. J. Hydrol. 332(3-4), 316-336.

Chow, M., Yusop, Z., Toriman, M., 2012. Modelling runoff quantity and quality in tropical urban catchments using Storm Water Management Model. Int. J. Environ. Sci. Technol. 9, 737–748.

Chow, M.F., Yusop, Z., Mohamed, M., 2011. Quality and first flush analysis of stormwater runoff from a tropical commercial catchment. Water Sci. Technol. 63(6), 1211-1216.

Connolly, R., Costantini, A., Loch, R., Garthe, R., 1999. Sediment generation from forest roads: bed and eroded sediment size distributions, and runoff management strategies. Soil Res. 37, 947–964.

Cristina, C., Tramonte, J., Sansalone, J.J., 2002. A granulometry-based selection methodology for separation of traffic-generated particles in urban highway snowmelt runoff. Water. Air. Soil Pollut. 136, 33–53.

Damle, C., Yalcin, A., 2007. Flood prediction using Time Series Data Mining. J. Hydrol. 333(2-4), 305-316.

Das, K.R., Imon, A., 2016. A brief review of tests for normality. Am. J. Theor. Appl. Stat. 5, 5–12.

Deng, W., Wang, G., 2017. A novel water quality data analysis framework based on time-series data mining. J. Environ. Manage. 196, 365-375.

Dongquan, Z., Jining, C., Haozheng, W., Qingyuan, T., Shangbing, C., Zheng, S., 2009. GIS-based urban rainfall-runoff modeling using an automatic catchment-discretization approach: a case study in Macau. Environ. Earth Sci. 59, 465–472.

Donigian, A., 2002. Watershed model calibration and validation: The HSPF experience, in: TMDLS Conference 2002. Water Environment Federation, pp. 44–73.

Dotto, C.B.S., Kleidorfer, M., Deletic, A., Rauch, W., McCarthy, D.T., 2014. Impacts of measured data uncertainty on urban stormwater models. J. Hydrol. 508, 28–42.

Drapper, D., Tomlinson, R., Williams, P., 2000. Pollutant concentrations in road runoff: Southeast Queensland case study. J. Environ. Eng. 126, 313–320.

Ebrahimian, A., Gulliver, J.S., Wilson, B.N., 2016. Effective impervious area for runoff in urban watersheds. Hydrol. Process. 30, 3717–3729.

Egodawatta, P., Thomas, E., Goonetilleke, A., 2007. Mathematical interpretation of pollutant wash-off from urban road surfaces using simulated rainfall. Water Res. 41(13), 3025-3031.

Elliott, A., Trowsdale, S.A., 2007. A review of models for low impact urban stormwater drainage. Environ. Model. Softw. 22, 394–405.

Elliott, A.H., Trowsdale, S.A., Wadhwa, S., 2009. Effect of aggregation of on-site storm-water control devices in an urban catchment model. J. Hydrol. Eng. 14, 975–983.

ETV 2016 Technology Verified: Downstream Defender. Toronto. Available online: http://etvcanada.ca/wp-content/uploads/2017/02/Canadian-ETV-VS-TFS-Hydro-International_September-2016.pdf.

Fletcher, T.D., Andrieu, H., Hamel, P., 2013. Understanding, management and modelling of urban hydrology and its consequences for receiving waters: A state of the art. Adv. Water Resour. 51, 261–279.

Furumai, H., Balmer, H., Boller, M., 2002. Dynamic behavior of suspended pollutants and particle size distribution in highway runoff. Water Sci. Technol. 46, 413–418.

Geiger, W., 1987. Flushing effects in combined sewer systems. Proceedings of the 4th International Conference Urban Drainage, Lausanne, Switzerland, pp. 40-46.

Ghosh, I., Hellweger, F.L., 2012. Effects of spatial resolution in urban hydrologic simulations. J. Hydrol. Eng. 17, 129–137.

Gironás, J., Roesner, L.A., Rossman, L.A., Davis, J., 2010. A new applications manual for the Storm Water Management Model(SWMM). Environ. Model. Softw. 25, 813–814.

Gong, Y., Liang, X., Li, X., Li, J., Fang, X., Song, R., 2016. Influence of rainfall characteristics on total suspended solids in urban runoff: A case study in Beijing, China. Water 8, 278.

Google (2023) Calgary, AB. Available at: https://www.google.com/maps.

Goonetilleke, A., Thomas, E., Ginn, S., Gilbert, D., 2005. Understanding the role of land use in urban stormwater quality management. J. Environ. Manage. 74(1), 31-42.

Gorgoglione, A., Bombardelli, F.A., Pitton, B.J., Oki, L.R., Haver, D.L., Young, T.M., 2019. Uncertainty in the parameterization of sediment build-up and wash-off processes in the simulation of sediment transport in urban areas. Environ. Model. Softw. 111, 170–181.

Government of Canada 1981–2010 Climate Normals & Averages. Available online: https://climate.weather.gc.ca/climate_normals/ (accessed on 16 January 2022).

Granata, F., Papirio, S., Esposito, G., Gargano, R., De Marinis, G., 2017. Machine learning algorithms for the forecasting of wastewater quality indicators. Water 9, 105.

Granato, D., Santos, J.S., Escher, G.B., Ferreira, B.L., Maggio, R.M., 2018. Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. Trends Food Sci. Technol. 72, 83-90.

Greb, S.R., Bannerman, R.T., 1997. Influence of particle size on wet pond effectiveness. Water Environ. Res. 69, 1134–1138.

Guse, B., Reusser, D.E., Fohrer, N., 2014. How to improve the representation of hydrological processes in SWAT for a lowland catchment - temporal analysis of parameter sensitivity and model performance. Hydrol. Processes 28(4), 2651-2670.

Guzman, C.B., Wang, R., Muellerklein, O., Smith, M., Eger, C.G., 2022. Comparing stormwater quality and watershed typologies across the United States: A machine learning approach. Water Res. 216, 118283.

Hawkins, D.M., 1980. Identification of outliers, Springer.

Helmreich, B., Hilliges, R., Schriewer, A., Horn, H., 2010. Runoff pollutants of a highly trafficked urban road–Correlation analysis and seasonal influences. Chemosphere 80(9), 991-997.

Hernández-Crespo, C., Fernández-Gonzalvo, M., Martín, M., Andrés-Doménech, I., 2019. Influence of rainfall intensity and pollution build-up levels on water quality and quantity response of permeable pavements. Sci. Total Environ. 684, 303–313.

Hoffman, E.J., Mills, G.L., Latimer, J.S., Quinn, J.G., 1984. Urban runoff as a source of polycyclic aromatic hydrocarbons to coastal waters. Environ. Sci. Technol. 18, 580–587.

Hossain, S., Hewa, G.A., Wella-Hewage, S., 2019. A Comparison of Continuous and Event-Based Rainfall–Runoff (RR) Modelling Using EPA-SWMM. Water 11(3).

House, L.B., Waschbusch, R.J., Hughes, P.E., 1993. Water quality of an urban wet detention pond in Madison, Wisconsin, 1987-88. U.S. Geological Survey, in cooperation with the Wisconsin Department of Natural Resources. 93-172.

James, W., L. A. Rossman, W. R. C. James., 2003. "User's Guide to PCSWMM." Computational Hydraulics International: Guelph, Ontario, Canada.

James, W., Rossman, L.A., James, W.R.C., 2010. User's guide to SWMM 5:[based on original USEPA SWMM documentation], CHI.

Jani, J., Yang, Y.-Y., Lusk, M.G., Toor, G.S., 2020. Composition of nitrogen in urban residential stormwater runoff: Concentrations, loads, and source characterization of nitrate and organic nitrogen. PLoS One 15, e0229715.

Janke, B.D., Finlay, J.C., Hobbie, S.E., Baker, L.A., Sterner, R.W., Nidzgorski, D., Wilson, B.N., 2014. Contrasting influences of stormflow and baseflow pathways on nitrogen and phosphorus export from an urban watershed. Biogeochemistry 121, 209–228.

Jeong, H., Choi, J.Y., Lee, J., Lim, J., Ra, K., 2020. Heavy metal pollution by road-deposited sediments and its contribution to total suspended solids in rainfall runoff from intensive industrial areas. Environ. Pollut. 265, 115028.

Jin, G., Xu, J., Mo, Y., Tang, H., Wei, T., Wang, Y.-G., Li, L., 2020. Response of sediments and phosphorus to catchment characteristics and human activities under different rainfall patterns with Bayesian Networks. J. Hydrol 584.

Kadlec, R.H., 1999. Chemical, physical and biological cycles in treatment wetlands. Water Sci. Technol. 40, 37–44.

Kellner, E., Hubbart, J.A., 2017. Improving understanding of mixed-land-use watershed suspended sediment regimes: Mechanistic progress through high-frequency sampling. Sci. Total Environ. 598, 228–238.

Kim, J.-Y., Sansalone, J.J., 2008. Event-based size distributions of particulate matter transported during urban rainfall-runoff events. Water Res. 42, 2756–2768.

Kim, L.H., Ko, S.O., Jeong, S., Yoon, J., 2007. Characteristics of washed-off pollutants and dynamic EMCs in parking lots and bridges during a storm. Sci. Total Environ. 376(1-3), 178-184.

Knighton, J., Lennon, E., Bastidas, L., White, E., 2016. Stormwater detention system parameter sensitivity and uncertainty analysis using SWMM. J. Hydrol. Eng. 21, 05016014.

Krebs, G., Kokkonen, T., Valtanen, M., Setälä, H., Koivusalo, H., 2014. Spatial resolution considerations for urban hydrological modelling. J. Hydrol. 512, 482–497.

Kshirsagar, M.P., Khare, K.C., 2023. Support Vector Regression Models of Stormwater Quality for a Mixed Urban Land Use. Hydrology 10, 66.

Kusiak, A., Verma, A., Wei, X., 2013. A data-mining approach to predict influent quality. Environ. Monit. Assess. 185(3), 2197-2210.

Lakshmi, G., Sudheer, K.P., 2021. Parameterization in hydrological models through clustering of the simulation time period and multi-objective optimization based calibration. Environ. Modell. Softw. 138.

Lan, T., Lin, K.R., Liu, Z.Y., He, Y.H., Xu, C.Y., Zhang, H.B., Chen, X.H., 2018. A Clustering Preprocessing Framework for the Subannual Calibration of a Hydrological Model Considering Climate-Land Surface Variations. Water Resour. Res. 54(12).

Leandro, J., Schumann, A., Pfister, A., 2016. A step towards considering the spatial heterogeneity of urban key features in urban hydrology flood modelling. J. Hydrol. 535, 356–365.

Lee, S.-W., Hwang, S.-J., Lee, S.-B., Hwang, H.-S., Sung, H.-C., 2009. Landscape ecological approach to the relationships of land use patterns in watersheds to water quality characteristics. Landsc. Urban Plan. 92, 80–89.

Li, A., Feng, M., Li, Y., Liu, Z., 2016. Application of outlier mining in insider identification based on boxplot method. Procedia Computer Science 91, 245-251.

Li, C., Liu, M., Hu, Y., Gong, J., Xu, Y., 2016. Modeling the Quality and Quantity of Runoff in a Highly Urbanized Catchment Using Storm Water Management Model. Pol. J. Environ. Stud. 25.

Li, D., Wan, J., Ma, Y., Wang, Y., Huang, M., Chen, Y., 2015. Stormwater runoff pollutant loading distributions and their correlation with rainfall and catchment characteristics in a rapidly industrialized city. PLoS One 10(3), e0118776.

Li, J., Liang, Z., Li, Y., Li, P., Jiang, C., 2018. Experimental study and simulation of phosphorus purification effects of bioretention systems on urban surface runoff. Plos One 13, e0196339.

Liu, A., 2015. Role of rainfall and catchment characteristics on urban stormwater quality. Book.

Liu, A., Egodawatta, P., Guan, Y., Goonetilleke, A., 2013. Influence of rainfall and catchment characteristics on urban stormwater quality. Sci. Total Environ. 444, 255-262.

Liu, A., Goonetilleke, A., Egodawatta, P., 2015. Stormwater Treatment Design. In Role of Rainfall and Catchment Characteristics on Urban Stormwater Quality (pp. 15-30). Springer, Singapore.

Liu, A., Goonetilleke, A., Egodawatta, P., 2012. Inadequacy of land use and impervious area fraction for determining urban stormwater quality. Water Resour. Manag. 26, 2259–2265.

Liu, A., Goonetilleke, A., Egodawatta, P., 2015. Role of rainfall and catchment characteristics on urban stormwater quality. Springer.

Liu, A., Gunawardana, C., Gunawardena, J., Egodawatta, P., Ayoko, G. A., Goonetilleke, A., 2016. Taxonomy of factors which influence heavy metal build-up on urban road surfaces. J. Hazard. Mater. 310, 20-29.

Liu, S., Ryu, D., Webb, J.A., Lintern, A., Waters, D., Guo, D., Western, A.W., 2018. Characterisation of spatial variability in water quality in the Great Barrier Reef catchments using multivariate statistical analysis. Mar. Pollut. Bull. 137, 137-151.

Löwe, R., Madsen, H., McSharry, P., 2016. Objective classification of rainfall in northern Europe for online operation of urban water systems based on clustering techniques. Water 8(3), 87.

Lucke, T., Drapper, D., Hornbuckle, A., 2018. Urban stormwater characterisation and nitrogen composition from lot-scale catchments—New management implications. Sci. Total Environ. 619, 65–71.

Maestre, A., Pitt, R.E., 2006. Identification of significant factors affecting stormwater quality using the national stormwater quality database. J. Water Manage. Modell.

Maestre, A., Pitt, R.E., Williamson, D., 2004. Nonparametric statistical tests comparing first flush and composite samples from the national stormwater quality database. J. Water Manage. Modell.

Mamun, A.A., Shams, S., Nuruzzaman, M., 2020. Review on uncertainty of the first-flush phenomenon in diffuse pollution control. Appl. Water Sci. 10(1), 1-10.

Mansfield, E.R., Helms, B.P., 1982. Detecting multicollinearity. The American Statistician 36(3a), 158-160.

McKenzie, E.R., Young, T.M., 2013. A novel fractionation approach for water constituents - distribution of storm event metals. Environ. Sci.: Processes Impacts 15(5), 1006-1016.

McLeod, S.M., Kells, J.A., Putz, G.J., 2006. Urban runoff quality characterization and load estimation in Saskatoon, Canada. J. Environ. Eng. 132, 1470–1481.

Meng, F., Sa, C., Liu, T., Luo, M., Liu, J., Tian, L., 2020. Improved model parameter transferability method for hydrological simulation with SWAT in Ungauged Mountainous Catchments. Sustainability 12, 3551.

Miguntanna, N.P., Goonetilleke, A., Egodowatta, P., Kokot, S., 2010. Understanding nutrient build-up on urban road surfaces. J. Environ. Sci. 22, 806–812.

Miguntanna, N.P., Liu, A., Egodawatta, P., Goonetilleke, A., 2013. Characterising nutrients wash-off for effective urban stormwater treatment design. J. Environ. Manage. 120, 61–67.

Moeini, M., Shojaeizadeh, A., Geza, M., 2021. Supervised machine learning for estimation of total suspended solids in urban watersheds. Water 13, 147.

Mukaka, M.M., 2012. A guide to appropriate use of correlation coefficient in medical research. Malawi Med. J. 24 (3), 69–71.

Müller, A., Österlund, H., Marsalek, J., Viklander, M., 2020. The pollution conveyed by urban runoff: A review of sources. Sci. Total Environ. 709, 136125.

Murphy, L.U., Cochrane, T.A., O'Sullivan, A., 2015. Build-up and wash-off dynamics of atmospherically derived Cu, Pb, Zn and TSS in stormwater runoff as a function of meteorological characteristics. Sci. Total Environ. 508, 206-213.

Muthusamy, M., Tait, S., Schellart, A., Beg, M.N.A., Carvalho, R.F., de Lima, J.L., 2018. Improving understanding of the underlying physical process of sediment wash-off from urban road surfaces. J. Hydrol. 557, 426–433.

Nelson, E.J., Booth, D.B., 2002. Sediment sources in an urbanizing, mixed land-use watershed. J. Hydrol. 264(1-4), 51-68.

Niazi, M., Nietch, C., Maghrebi, M., Jackson, N., Bennett, B.R., Tryby, M., Massoudieh, A., 2017. Storm water management model: Performance review and gap analysis. J. Sustain. Water Built Environ. 3, 04017002.

Niemi, T.J., Kokkonen, T., Sillanpää, N., Setälä, H., Koivusalo, H., 2019. Automated urban rainfall–runoff model generation with detailed land cover and flow routing. J. Hydrol. Eng. 24, 04019011.

NJDEP 2021: Stormwater Manufactured Treatment Device Archive and Expired Stormwater MTDs. New Jersey. Available online: https://dep.nj.gov/wp-content/uploads/stormwater/2021_hds_protocol_final_version.pdf

Oberts, G.L., 1986. Pollutants associated with sand and salt applied to roads in Minnesota 1. J. Am. Water Resour. Assoc. 22(3), 479-483.

Orouji, H., Bozorg Haddad, O., Fallah-Mehdipour, E., Mariño, M., 2013. Modeling of water quality parameters using data-driven models. J. Environ. Eng. 139(7), 947-957.

Oudin, L., Andréassian, V., Perrin, C., Michel, C., Le Moine, N., 2008. Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. Water Resour. Res. 44.

Palli, L., Rosadoni, G., Mori, F., Milighetti, O., Lubello, C., 2021. Presence and behavior of metals in public sewage: long-term modelling with EPA SWMM. Urban Water J. 18, 53–60.

Patil, S., Stieglitz, M., 2014. Modelling daily streamflow at ungauged catchments: what information is necessary? Hydrol. Process. 28, 1159–1169.

Patil, S.D., Stieglitz, M., 2015. Comparing spatial and temporal transferability of hydrological model parameters. J. Hydrol. 525, 409–417.

Paule-Mercado, M.C.A., Salim, I., Lee, B.-Y., Memon, S., Sajjad, R.U., Sukhbaatar, C., Lee, C.-H., 2018. Monitoring and quantification of stormwater runoff from mixed land use and land cover catchment in response to land development. Ecol. Indic. 93, 1112–1125.

Perera, T., McGree, J., Egodawatta, P., Jinadasa, K., Goonetilleke, A., 2021. Catchment based estimation of pollutant event mean concentration (EMC) and implications for first flush assessment. J. Environ. Manage. 279, 111737.

Petrucci, G., Bonhomme, C., 2014. The dilemma of spatial representation for urban hydrology semi-distributed modelling: Trade-offs among complexity, calibration and geographical data. J. Hydrol. 517, 997–1007.

Pitt, R., 2001. Stormwater effects handbook: A toolbox for watershed managers, scientists, and engineers, CRC Press.

Pitt, R., Maestre, A., Clary, J., 2018. The national stormwater quality database (NSQD), version 4.02. Dep. Civ. Environ. Eng.

Pitt, R.E., Williamson, D., Voorhees, J., Clark, S., 2005. Review of historical street dust and dirt accumulation and washoff data. J. Water Manage. Modell.

Rai, P.K., Chahar, B., Dhanya, C., 2017. GIS-based SWMM model for simulating the catchment response to flood events. Hydrol. Res. 48(2), 384-394.

Reichwaldt, E.S., Ghadouani, A., 2012. Effects of rainfall patterns on toxic cyanobacterial blooms in a changing climate: between simplistic scenarios and complex dynamics. Water Res. 46, 1372–1393.

Rio, M., Salles, C., Cernesson, F., Marchand, P., Tournoud, M.-G., 2020. An original urban land cover representation and its effects on rain event-based runoff and TSS modelling. J. Hydrol. 586, 124865.

River, M., Richardson, C.J., 2018. Particle size distribution predicts particulate phosphorus removal. Ambio 47, 124–133.

Rossman, L.A., Huber, W.C., 2016. Storm water management model reference manual volume III—Water quality. US EPA National Risk Management Research Laboratory: Cincinnati, OH, USA.

Rossman, L.A., 2010. Storm water management model user's manual.

Roy, A.H., Wenger, S.J., Fletcher, T.D., Walsh, C.J., Ladson, A.R., Shuster, W.D., Thurston, H.W., Brown, R.R., 2008. Impediments and solutions to sustainable, watershed-scale urban stormwater management: lessons from Australia and the United States. Environ. Manage. 42, 344–359.

Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. 1, 206–215.

Saber, A., James, D.E., Hayes, D.F., 2019. Estimation of water quality profiles in deep lakes based on easily measurable constituents at the water surface using artificial neural networks coupled with stationary wavelet transform. Sci. Total Environ. 694, 133690.

Salim, I., Paule-Mercado, M.C., Sajjad, R.U., Memon, S.A., Lee, B.-Y., Sukhbaatar, C., Lee, C.-H., 2019. Trend analysis of rainfall characteristics and its impact on stormwater runoff quality from urban and agricultural catchment. Membr. Water Treat 10(1), 45-55.

Santhi, C., Arnold, J.G., Williams, J.R., Dugas, W.A., Srinivasan, R., Hauck, L.M., 2001. Validation of the swat model on a large rwer basin with point and nonpoint sources 1. J. Am. Water Resour. Assoc. 37(5), 1169-1188.

Selbig, W.R., 2015. Characterizing the distribution of particles in urban stormwater: advancements through improved sampling technology. Urban Water J. 12, 111–119.

Selbig, W.R., Bannerman, R.T., 2007. Evaluation of street sweeping as a stormwater-quality-management tool in three residential basins in Madison, Wisconsin. U.S. Geological Survey, Middleton, Wisconsin, Water Resource Investigations Report. 2007-5156.

Selbig, W.R., Bannerman, R.T., 2011. Characterizing the size distribution of particles in urban stormwater by use of fixed-point sample-collection methods. U.S. Geological Survey Open File Report. 2011-1052.

Selbig, W.R., Fienen, M.N., Horwatich, J.A., Bannerman, R.T., 2016. The effect of particle size distribution on the design of urban stormwater control measures. Water. 8, 17.

Shen, Z., Liu, J., Aini, G., Gong, Y., 2016. A comparative study of the grain-size distribution of surface dust and stormwater runoff quality on typical urban roads and roofs in Beijing, China. Environ. Sci. Pollut. Res. 23, 2693–2704.

Simpson, I.M., Winston, R.J., Brooker, M.R., 2022. Effects of land use, climate, and imperviousness on urban stormwater quality: A meta-analysis. Sci. Total Environ. 809, 152206.

Simpson, I.M., Winston, R.J., Dorsey, J.D., 2023. Monitoring the effects of urban and forested land uses on runoff quality: Implications for improved stormwater management. Sci. Total Environ. 862, 160827.

Smith, J.S., Winston, R.J., Tirpak, R.A., Wituszynski, D.M., Boening, K.M., Martin, J.F., 2020. The seasonality of nutrients and sediment in residential stormwater runoff: Implications for nutrient-sensitive waters. J. Environ. Manage. 276, 111248.

Smiti, A., 2020. A critical overview of outlier detection methods. Computer Science Review 38.

Sönmez, İ., Kömüşcü, A.Ü., 2011. Reclassification of rainfall regions of Turkey by K-means methodology and their temporal variability in relation to North Atlantic Oscillation (NAO). Theor. Appl. Climatol. 106(3), 499-510.

Soonthornnonda, P., Christensen, E. R., 2008. A load model based on antecedent dry periods for pollutants in stormwater. Water Environ. Res. 80(2), 162-171.

Steinley, D., Brusco, M.J., 2007. Initializing K-means batch clustering: A critical evaluation of several techniques. J. Classif. 24(1), 99-121.

Sun, N., Hall, M., Hong, B., Zhang, L., 2014. Impact of SWMM catchment discretization: Case study in Syracuse, New York. J. Hydrol. Eng. 19, 223–234.

Taiwo, A.M., Beddows, D.C., Shi, Z., Harrison, R.M., 2014. Mass and number size distributions of particulate matter components: Comparison of an industrial site and an urban background site. Sci. Total Environ. 475, 29–38.

Tang, S., Jiang, J., Zheng, Y., Hong, Y., Chung, E.S., Shamseldin, A.Y., Wei, Y., Wang, X., 2021. Robustness analysis of storm water quality modelling with LID infrastructures from natural event-based field monitoring. Sci. Total Environ. 753, 142007.

Tang, Y., Zhu, D.Z., Rajaratnam, N., van Duin, B., 2020. Sediment depositions in a submerged storm sewer pipe. J. Environ. Eng. 146, 04020118.

Tang, Y., Zhu, D.Z., van Duin, B., 2018. Note on sediment removal efficiency in oil–grit separators. Water Sci. Technol. 2017, 729–735.

Taylor, G.D., Fletcher, T.D., Wong, T.H., Breen, P.F., Duncan, H.P., 2005. Nitrogen composition in urban runoff—implications for stormwater management. Water Res. 39, 1982–1989.

The City of Calgary 2011 Stormwater management and design manual. The City of Calgary, Alberta, Canada. Available online: http://www.calgary.ca.login.ezproxy.library.ualberta.ca/PDA/pd/Documents/urban_development/bulletins/2011-stormwater-management-and-Design.pdf .

Thenoux, G., Bellolio, J.P., Halles, F., 2007. Development of a methodology for measurement of vehicle dust generation on unpaved roads. Transp. Res. Rec. 1989, 299–304.

Thyregod, P., Carstensen, J., Madsen, H., Arnbjerg-Nielsen, K., 1999. Integer valued autoregressive models for tipping bucket rainfall measurements. Environmetrics Off. J. Int. Environmetrics Soc. 10, 395–411.

Tscheikner-Gratl, F., Bellos, V., Schellart, A., Moreno-Rodenas, A., Muthusamy, M., Langeveld, J., Clemens, F., Benedetti, L., Rico-Ramirez, M.A., de Carvalho, R.F., Breuer, L., Shucksmith, J., Heuvelink, G.B.M., Tait, S., 2019. Recent insights on uncertainties present in integrated catchment water quality modelling. Water Res. 150, 368-379.

Tu, M.-C., Smith, P., 2018. Modeling pollutant buildup and washoff parameters for SWMM based on land use in a semiarid urban watershed. Water, Air, Soil Pollut. 229(4), 1-15.

Tuomela, C., Sillanpää, N., Koivusalo, H., 2019. Assessment of stormwater pollutant loads and source area contributions with storm water management model (SWMM). Journal of Environ. Manage. 233, 719-727.

Valtanen, M., Sillanpää, N., Setälä, H., 2014. The effects of urbanization on runoff pollutant concentrations, loadings and their seasonal patterns under cold climate. Water. Air. Soil Pollut. 225, 1–16.

Van Engelen, J.E., Hoos, H.H., 2020. A survey on semi-supervised learning. Mach. Learn. 109, 373–440.

Vaze, J., Chiew, F.H.S., 2003. Comparative evaluation of urban storm water quality models. Water Resour. Res. 39(10).

Vaze, J., Chiew, F.H., 2002. Experimental study of pollutant accumulation on an urban road surface. Urban Water. 4, 379–389.

Vaze, J., Chiew, F.H., 2004. Nutrient loads associated with different sediment sizes in urban stormwater and surface pollutants. J. Environ. Eng. 130, 391–396.

Wang, J., Huang, J.J., Li, J., 2020. Characterization of the pollutant build-up processes and concentration/mass load in road deposited sediments over a long dry period. Sci. Total Environ. 718, 137282.

Wang, J., Qin, M., Huang, T., Tu, N., Li, B., 2021. Particle size distribution and pollutant dissolution characteristics of road-deposited sediment in different land-use districts: a case study of Beijing. Environ. Sci. Pollut. Res. 28, 38497–38505.

Wang, K.-H., Altunkaynak, A., 2012. Comparative Case Study of Rainfall-Runoff Modeling between SWMM and Fuzzy Logic Approach. J. Hydrol. Eng. 17(2), 283-291.

Westerlund, C., Viklander, M., 2006. Particles and associated metals in road runoff during snowmelt and rainfall. Sci. Total Environ. 362, 143–156.

Wijesiri, B., Bandala, E., Liu, A., Goonetilleke, A., 2020. A framework for stormwater quality modelling under the effects of climate change to enhance reuse. Sustainability 12, 10463.

Wijesiri, B., Deilami, K., Goonetilleke, A., 2018. Evaluating the relationship between temporal changes in land use and resulting water quality. Environ. Pollut. 234, 480–486.

Wijesiri, B., Egodawatta, P., McGree, J., Goonetilleke, A., 2015. Influence of pollutant build-up on variability in wash-off from urban road surfaces. Sci. Total Environ. 527-528, 344-350.

Wijesiri, B., Egodawatta, P., McGree, J., Goonetilleke, A., 2016. Assessing uncertainty in pollutant build-up and wash-off processes. Environ. Pollut. 212, 48-56.

Wijesiri, B., Egodawatta, P., McGree, J., Goonetilleke, A., 2016. Understanding the uncertainty associated with particle-bound pollutant build-up and wash-off: A critical review. Water Res. 101, 582–596.

Winston, R., Hunt, W., 2017. Characterizing runoff from roads: Particle size distributions, nutrients, and gross solids. J. Environ. Eng. 143, 04016074.

Wu, Q., Liu, S., Cai, Y., Li, X., Jiang, Y., 2017. Improvement of hydrological model calibration by selecting multiple parameter ranges. Hydrol. Earth Syst. Sci. 21, 393–407.

Xu, T., Coco, G., Neale, M., 2020. A predictive model of recreational water quality based on adaptive synthetic sampling algorithms and machine learning. Water Res. 177, 115788.

Yan, H., Fernandez, A., Zhu, D.Z., Zhang, W., Loewen, M.R., van Duin, B., Chen, L., Mahmood, K., Zhao, S., Jia, H., 2022. Land cover based simulation of urban stormwater runoff and pollutant loading. J. Environ. Manage. 303, 114147.

Yan, H., Zhu, D.Z., Loewen, M.R., Zhang, W., Liang, S., Ahmed, S., van Duin, B., Mahmood, K., Zhao, S., 2023. Impact of rainfall characteristics on urban stormwater quality using data mining framework. Sci. Total Environ. 862, 160689.

Yang, Y., Zhu, D.Z., Loewen, M.R., Ahmed, S.S., Zhang, W., Yan, H., van Duin, B., Mahmood, K., 2023. Evaluation of pollutant removal efficiency of urban stormwater wet ponds and the application of machine learning algorithms. Sci. Total Environ. 167119.

Yang, Y.-Y., Lusk, M.G., 2018. Nutrients in urban stormwater runoff: Current state of the science and potential mitigation options. Curr. Pollut. Rep. 4, 112–127.

Yang, Y.-Y., Toor, G.S., 2017. Sources and mechanisms of nitrate and orthophosphate transport in urban stormwater runoff from residential catchments. Water Res. 112, 176–184.

Yazdi, M.N., Sample, D.J., Scott, D., Wang, X., Ketabchy, M., 2021. The effects of land use characteristics on urban stormwater quality and watershed pollutant loads. Sci. Total Environ. 773, 145358.

Yazdi, M.N., Scott, D., Sample, D.J., Wang, X., 2021. Efficacy of a retention pond in treating stormwater nutrients and sediment. J. Clean. Prod. 290, 125787.

Yen, H., Wang, X., Fontane, D.G., Harmel, R.D., Arabi, M., 2014. A framework for propagation of uncertainty contributed by parameterization, input data, model structure, and calibration/validation data in watershed modeling. Environ. Model. Softw. 54, 211–221.

Yuan, C., Yang, H., 2019. Research on K-value selection method of K-means clustering algorithm. J. Multidiscip. Sci. J. 2, 226–235.

Zhang, H., Huang, G.H., Wang, D., Zhang, X., 2011. Multi-period calibration of a semi-distributed hydrological model based on hydroclimatic clustering. Adv. Water Resour. 34(10), 1292-1303.

Zhao, H., Jiang, Q., Ma, Y., Xie, W., Li, X., Yin, C., 2018. Influence of urban surface roughness on build-up and wash-off dynamics of road-deposited sediment. Environ. Pollut. 243, 1226–1234.

Zhao, H., Li, X., 2013. Understanding the relationship between heavy metals in road-deposited sediments and washoff particles in urban stormwater using simulated rainfall. J. Hazard. Mater. 246, 267–276.

Zhao, H., Li, X., Wang, X., Tian, D., 2010. Grain size distribution of road-deposited sediment and its contribution to heavy metal pollution in urban runoff in Beijing, China. J. Hazard. Mater. 183, 203–210.

Zhao, H., Ma, Y., Fang, J., Hu, L., Li, X., 2022. Particle size distribution and total suspended solid concentrations in urban surface runoff. Sci. Total Environ. 815, 152533.

Zhi, W., Feng, D., Tsai, W.-P., Sterle, G., Harpold, A., Shen, C., Li, L., 2021. From hydrometeorology to river water quality: can a deep learning model predict dissolved oxygen at the continental scale? Environ. Sci. Technol. 55, 2357–2368.

Zhou, L., Shen, G., Li, C., Chen, T., Li, S., Brown, R., 2021. Impacts of land covers on stormwater runoff and urban development: A land use and parcel based regression approach. Land Use Policy 103, 105280.

Zhu, H., Xu, Y., Yan, B., Guan, J., 2012. Snowmelt runoff: a new focus of urban nonpoint source pollution. Int. J. Environ. Res. Public. Health 9, 4333–4345.

Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., Wu, B., Ye, L., 2022. A review of the application of machine learning in water quality evaluation. Eco-Environ. Health.

Zivkovich, B.R., Mays, D.C., 2018. Predicting nonpoint stormwater runoff quality from land use. PLoS One 13, e0196782.

Appendix A. Summary of rainfall events.

Table A1. Summary of rainfall events in 2018 and 2019 including EMC values for TSS, TN and TP.

Catchment	Date	RD (h)	RI-max (mm/h)	RI-avg (mm/h)	RT (mm)	ADD (days)	RT-30 (mm)	T-peak (h)	TSS (mg/L)	TN (mg/L)	TP (mg/L)
RO	6/28/2018	3.9	4.8	1.9	7.6	5.4	1.0	0.3	72	1.55	0.04
	7/24/2018	5.1	9.6	1.4	7.0	4.9	0.6	1.9	116	2.08	0.25
	8/24/2018	5.3	14.4	1.9	9.8	4.2	3.4	0.9	96	3.13	0.10
	8/26/2018	5.4	7.2	1.5	8.2	2.2	1.0	1.9	22	0.76	0.07
	5/2/2019	2.5	7.2	2.2	5.6	17.0	0.8	1.0	1328	3.89	0.89
	6/8/2019	1.3	12	4.2	5.6	30.1	3.8	0.3	636	0.76	0.48
	6/19/2019	0.8	72	16.0	12.0	3.2	10.8	0.3	148	1.10	0.20
	6/20/2019	17.3	52.8	5.0	85.8	1.6	0.4	5.0	115	0.41	0.14
	6/27/2019	18.0	64.8	2.3	41.4	2.2	2.2	14.0	215	1.63	0.25
	7/3/2019	18.5	7.2	1.6	30.0	6.0	0.8	13.7	11	0.91	0.08
	7/16/2019	1.8	43.2	6.9	12.0	2.5	1.2	1.3	77	0.99	0.06
	8/6/2019	1.1	4.8	2.2	3.6	13.2	0.4	2.0	190	2.09	0.33
	8/16/2019	1.5	45.6	6.3	9.4	5.0	7.0	0.2	136	0.61	0.13
	8/22/2019	1.3	12	4.6	5.8	6.4	1.8	0.9	130	1.15	0.16
	9/1/2019	0.3	88.8	38.0	9.6	7.2	9.6	0.2	247	1.35	0.32
RR	6/23/2018	7.3	21.6	4.5	33.2	6.6	4.0	0.5	144	1.21	0.18
	6/28/2018	3.5	4.8	2.0	7.0	5.4	1.0	1.0	13	0.96	0.03
	7/10/2018	2.8	4.8	0.9	2.6	5.7	0.4	2.5	40	1.57	0.12

	7/23/2018	0.7	24	8.1	5.4	4.9	5.2	0.3	38	0.82	0.11
	8/24/2018	1.0	14.4	7.6	7.6	11.8	3.4	0.9	50	3.65	0.11
	8/26/2018	5.3	7.2	1.5	8.2	2.2	1.0	1.9	28	1.07	0.06
	9/10/2018	1.5	4.8	1.6	2.4	4.6	0.4	1.3	22	0.95	0.06
	9/20/2018	6.1	4.8	1.1	6.4	4.6	0.8	0.6	15	1.43	0.06
	9/26/2018	15.0	7.2	0.8	11.4	2.8	1.2	11.8	17	0.66	0.04
	5/2/2019	1.2	12	4.1	4.8	4.3	0.8	0.8	562	3.47	1.13
	6/6/2019	17.7	7.2	1.0	17.2	12.4	1.0	0.8	11	0.82	0.06
	6/8/2019	1.4	14.4	4.4	5.2	1.2	4.0	0.3	99	1.58	0.16
	6/27/2019	17.9	55.2	2.3	42.0	2.2	2.2	14.0	129	1.95	0.24
	7/16/2019	1.3	26.4	7.8	10.4	2.5	2.8	1.1	34	1.29	0.06
	8/6/2019	2.6	4.8	1.1	2.8	17.0	1.2	0.3	20	0.77	0.05
	8/16/2019	1.8	40.8	5.0	9.2	1.5	6.6	0.3	46	0.63	0.05
	9/1/2019	0.3	72	26.4	8.8	1.5	8.8	0.3	187	0.85	0.05
	7/23/2018	0.7	153.6	42.0	28.0	13.1	27.4	0.2	457	2.66	0.66
	8/24/2018	1.5	19.2	8.1	12.2	11.8	4.2	0.9	202	3.08	0.25
	8/26/2018	5.5	4.8	1.6	8.6	2.3	1.0	0.3	37	1.56	0.35
	9/26/2018	15.8	2.4	0.7	10.6	2.7	0.4	0.1	87	1.34	0.17
AB	10/3/2018	1.8	7.2	3.5	6.4	4.5	2.2	0.3	120	1.78	0.18
	5/17/2019	37.8	2.4	0.4	14.2	7.2	0.4	0.1	108	1.65	0.18
	6/20/2019	15.2	7.2	0.8	12.0	1.7	0.4	13.8	51	0.83	0.11
	6/23/2019	0.8	19.2	9.1	6.8	2.1	6.4	0.1	207	1.10	0.32
	7/16/2019	1.6	24	4.6	7.4	1.7	5.4	0.4	181	1.62	0.22

	9/9/2019	29.6	4.8	0.6	17.0	8.4	0.6	9.9	23	0.98	0.10
	9/19/2019	7.0	9.6	2.3	15.8	8.4	1.0	0.6	48	1.06	0.14
CR	6/22/2018	4.2	91.2	11.8	49.4	5.8	12.0	0.8	169	1.99	0.39
	6/28/2018	2.5	4.8	2.2	5.6	3.5	1.2	0.3	121	3.17	0.09
	7/10/2018	0.3	48	16.20	7.6	7.6	5.4	0.3	345	3.74	0.58
	7/23/2018	0.7	153.6	37.3	28.0	9.8	27.4	0.2	326	1.85	0.38
	8/24/2018	1.5	19.2	8.4	12.6	22.0	4.4	0.9	238	3.17	0.46
	9/27/2018	14.6	4.8	0.9	12.4	2.8	1.2	0.2	72	1.55	0.04
	5/17/2019	34.5	4.8	0.4	13.4	7.1	0.4	7.2	41	1.18	0.11
	6/19/2019	1.3	19.2	6.8	9.0	10.4	3.2	0.5	588	2.36	0.58
	6/20/2019	15.7	4.8	0.9	13.8	1.7	0.4	13.0	126	0.93	0.22
	6/23/2019	1.3	7.2	4.20	6.6	2.1	0.4	0.6	313	1.03	0.34
	6/27/2019	1.3	38.4	29.92	44.2	3.2	1.6	17.0	593	2.5	0.61
	7/3/2019	1.0	16.8	4.8	4.8	1.5	3.2	0.2	158	2.09	0.22
	8/6/2019	1.2	48	6.9	8.0	16.7	7.4	0.3	500	2.13	0.43
	8/16/2019	1.2	26.4	4.8	5.6	2.8	4.4	0.2	182	1.16	0.23
	8/22/2019	3.3	16.8	2.3	7.6	6.4	2.8	0.2	92	1.85	0.13
	8/31/2019	6.9	9.6	2.8	19.0	6.5	1.8	2.7	44	0.72	0.09
	9/9/2019	27.7	4.8	0.6	17.0	8.4	0.6	9.9	28	1.06	0.11
	9/19/2019	5.1	9.6	3.0	15.4	8.5	3.4	0.6	48	0.90	0.14

Note: RD is the rainfall duration, RI-max is the max rainfall intensity, RI-avg is the average rainfall intensity, RT is the total rainfall amount, ADD is the number of antecedent dry days RT-30 is the initial 30-min total rainfall amount, and T-peak is the time to peak rainfall intensity.

Appendix B. Boxplots of measured EMCs for TSS, TN and TP from four catchments.

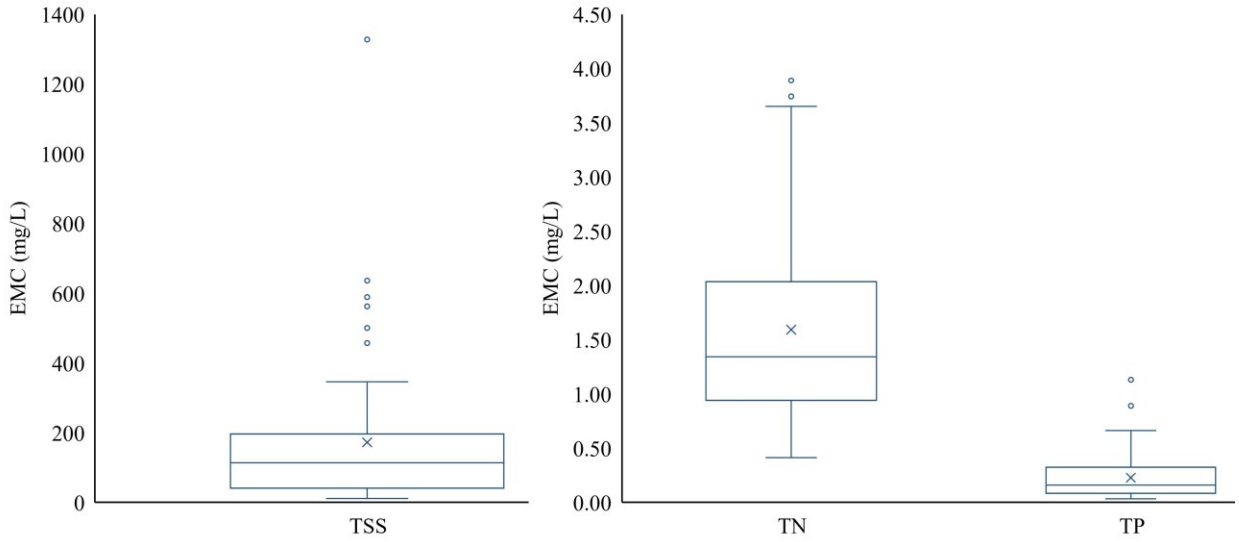


Fig. B1. Boxplots of measured EMCs for TSS, TN and TP from four catchments (central bar = Q_2; hinges = Q_1 and Q_3; empty circle = outliers; cross = average).

Appendix C. Results of K-means clustering.

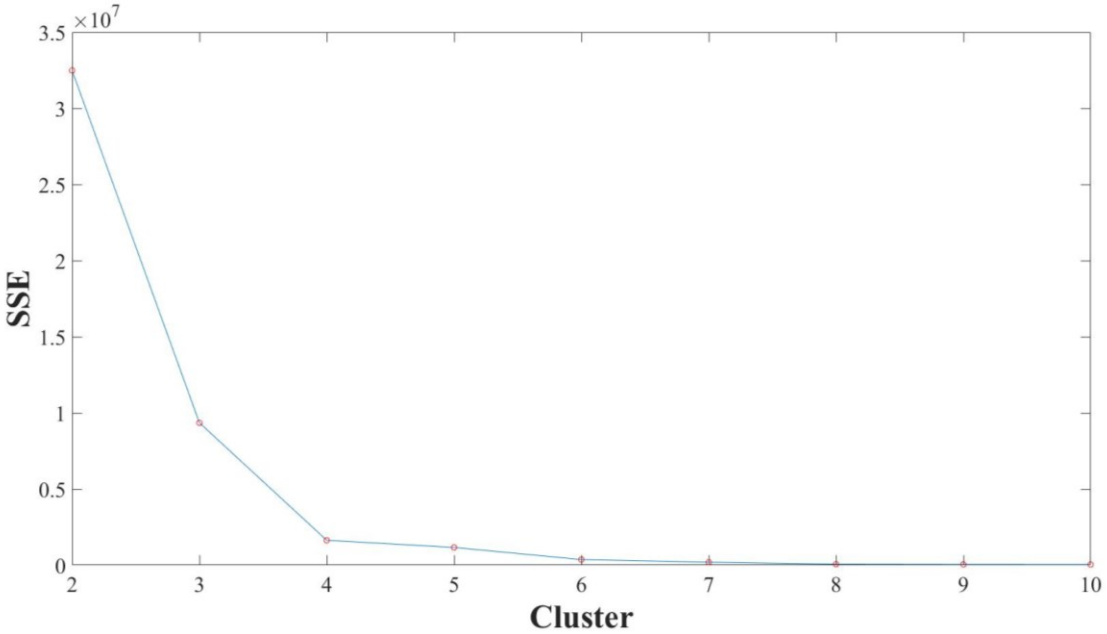


Fig. C1. Variation of the *SSE* index with the number of clusters.

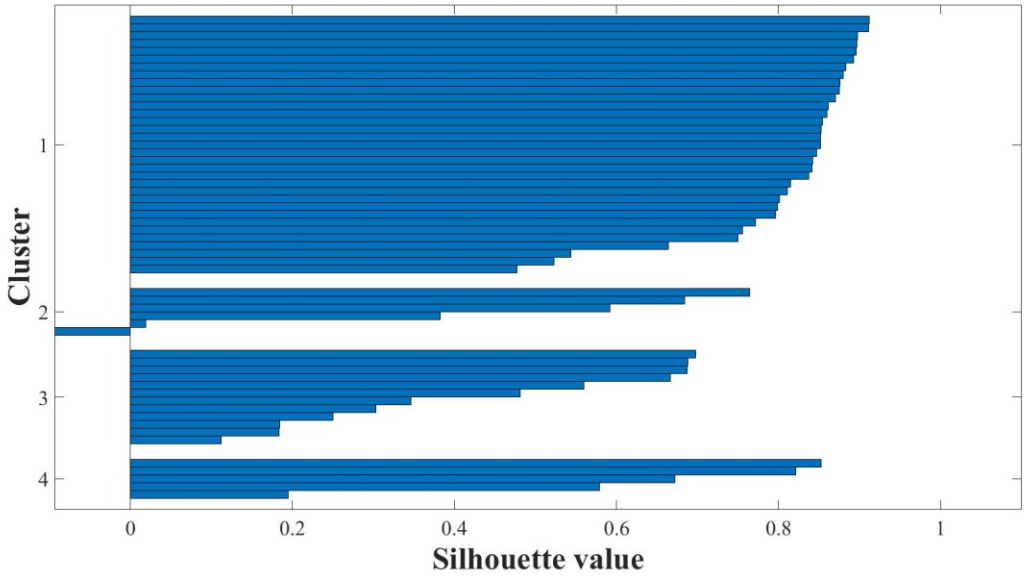
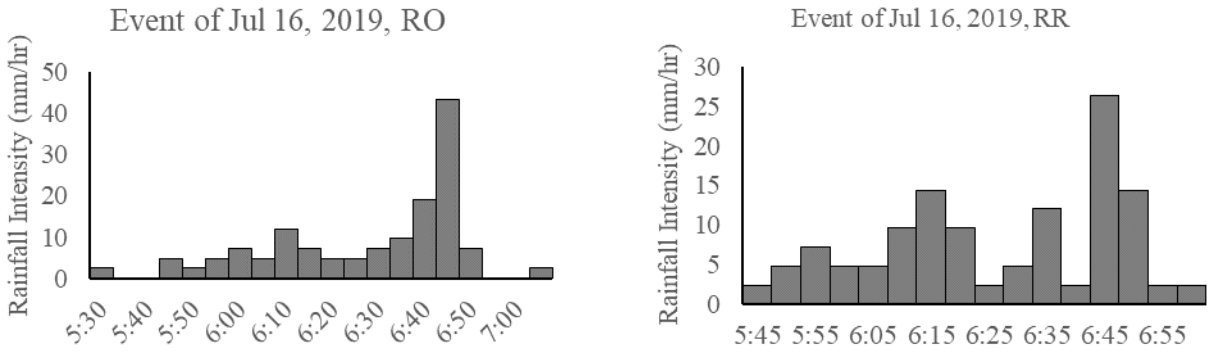


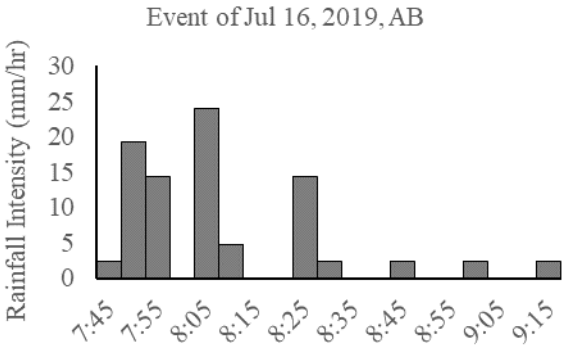
Fig. C2. Variation of the silhouette value with the number of clusters.

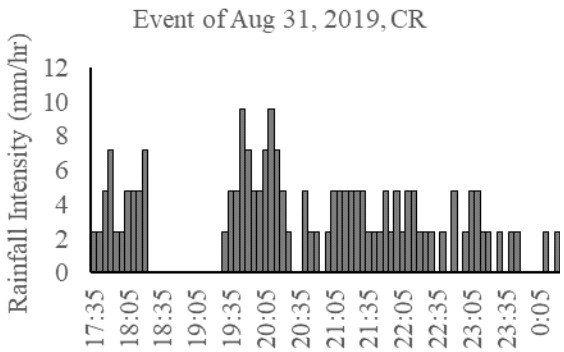
Appendix D. Examples of rainfall histograms for four rainfall event types.

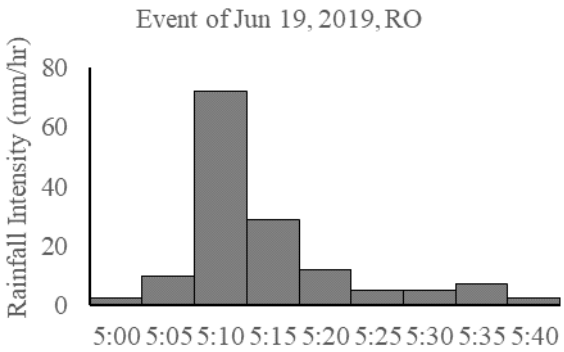
a

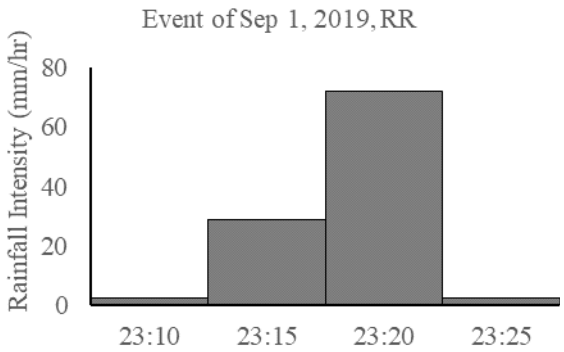


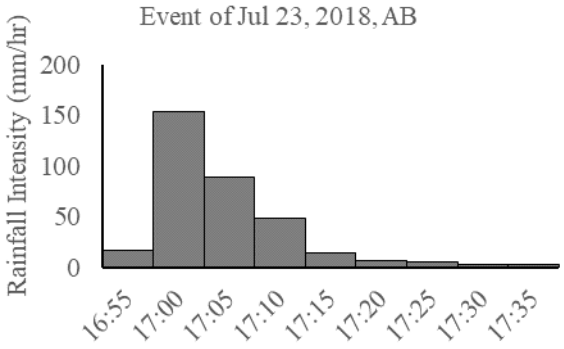
b

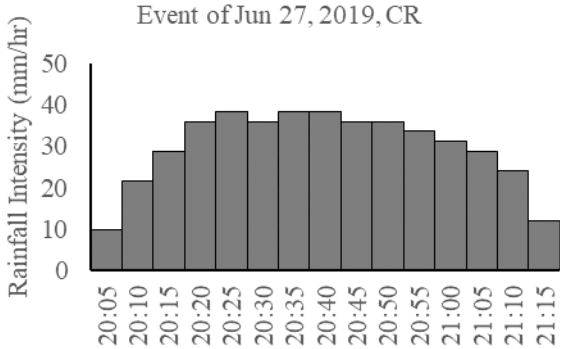

Event of Jul 16, 2019, AB
Rainfall Intensity (mm/hr)
30
25
20
15
10
5
0
7:45
7:55
8:05
8:15
8:25
8:35
8:45
8:55
9:05
9:15


Event of Aug 31, 2019, CR
Rainfall Intensity (mm/hr)
12
10
8
6
4
2
0
17:35
18:05
18:35
19:05
19:35
20:05
20:35
21:05
21:35
22:05
22:35
23:05
23:35
0:05

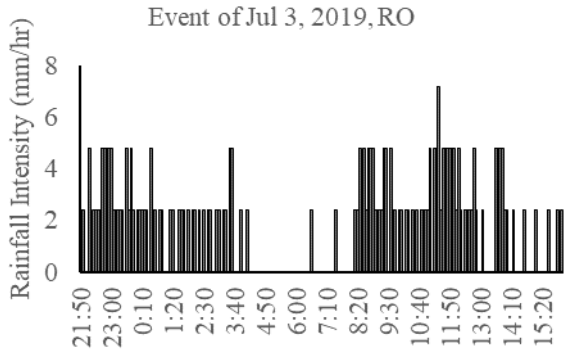

Event of Jun 19, 2019, RO
Rainfall Intensity (mm/hr)
80
60
40
20
0
5:00 5:05 5:10 5:15 5:20 5:25 5:30 5:35 5:40

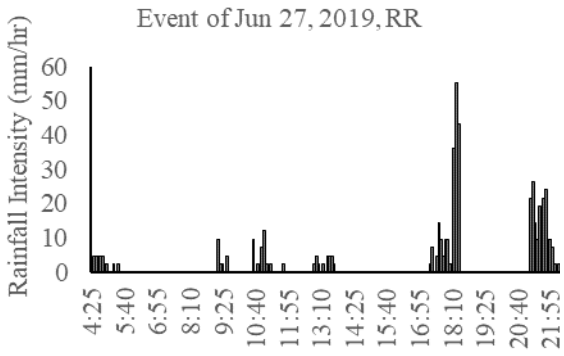

Event of Sep 1, 2019, RR
Rainfall Intensity (mm/hr)
80
60
40
20
0
23:10
23:15
23:20
23:25

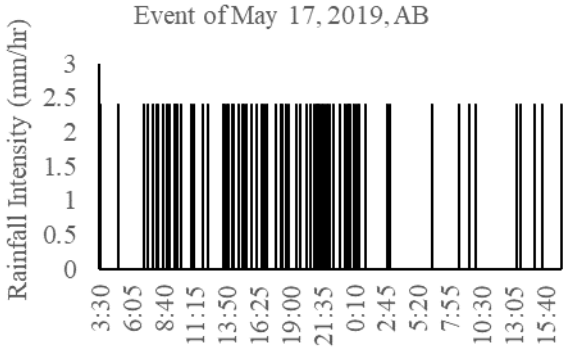

Event of Jul 23, 2018, AB
Rainfall Intensity (mm/hr)
200
150
100
50
0
16:55
17:00
17:05
17:10
17:15
17:20
17:25
17:30
17:35

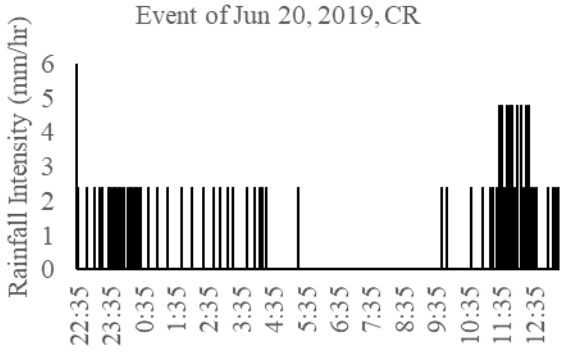

Event of Jun 27, 2019, CR
Rainfall Intensity (mm/hr)
50
40
30
20
10
0
20:05
20:10
20:15
20:20
20:25
20:30
20:35
20:40
20:45
20:50
20:55
21:00
21:05
21:10
21:15

c

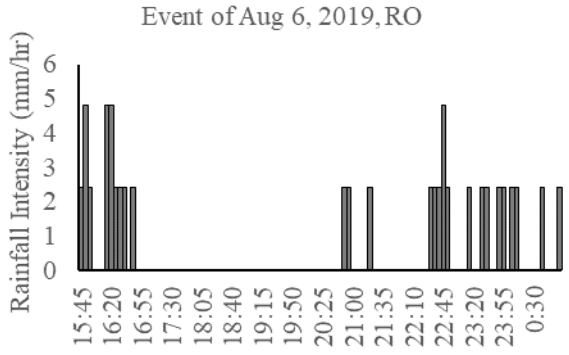

Event of Jul 3, 2019, RO
Rainfall Intensity (mm/hr)
0
2
4
6
8
21:50
23:00
0:10
1:20
2:30
3:40
4:50
6:00
7:10
8:20
9:30
10:40
11:50
13:00
14:10
15:20

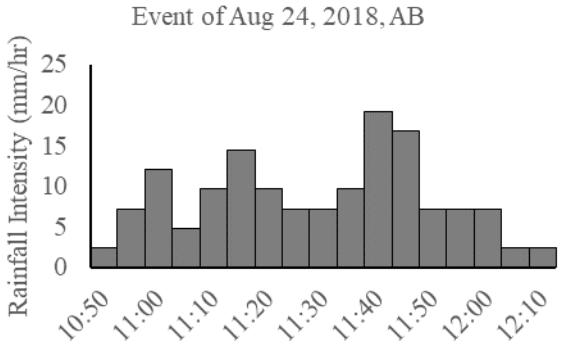

Event of Jun 27, 2019, RR
Rainfall Intensity (mm/hr)
0
10
20
30
40
50
60
4:25
5:40
6:55
8:10
9:25
10:40
11:55
13:10
14:25
15:40
16:55
18:10
19:25
20:40
21:55


Event of May 17, 2019, AB
Rainfall Intensity (mm/hr)
0
0.5
1
1.5
2
2.5
3
3:30
6:05
8:40
11:15
13:50
16:25
19:00
21:35
0:10
2:45
5:20
7:55
10:30
13:05
15:40


Event of Jun 20, 2019, CR
Rainfall Intensity (mm/hr)
0
1
2
3
4
5
6
22:35
23:35
0:35
1:35
2:35
3:35
4:35
5:35
6:35
7:35
8:35
9:35
10:35
11:35
12:35

d


Event of Aug 6, 2019, RO
Rainfall Intensity (mm/hr)
0
1
2
3
4
5
6
15:45
16:20
16:55
17:30
18:05
18:40
19:15
19:50
20:25
21:00
21:35
22:10
22:45
23:20
23:55
0:30


Event of Aug 24, 2018, AB
Rainfall Intensity (mm/hr)
0
5
10
15
20
25
10:50
11:00
11:10
11:20
11:30
11:40
11:50
12:00
12:10

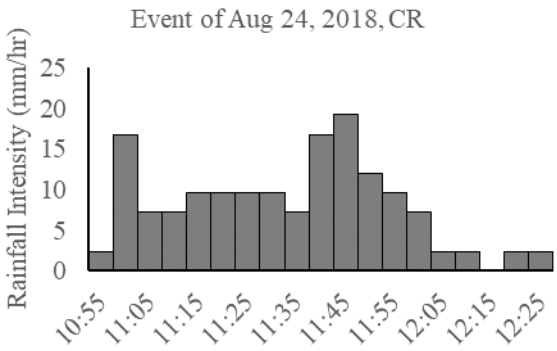
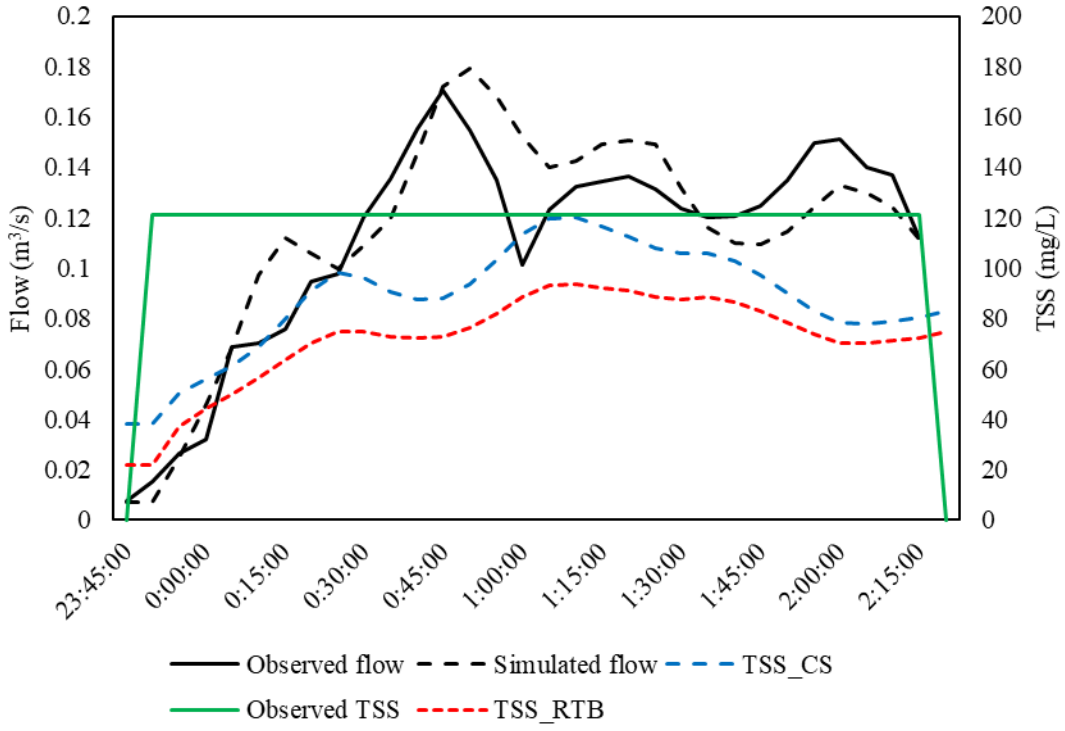


Fig. D1. Examples of rainfall histograms for four rainfall event types: (a) Type I; (b) Type II; (c)Type III; (d) Type IV

Appendix E. Hydrological and water quality modeling results of RTB and CS approaches in Cranston catchment.



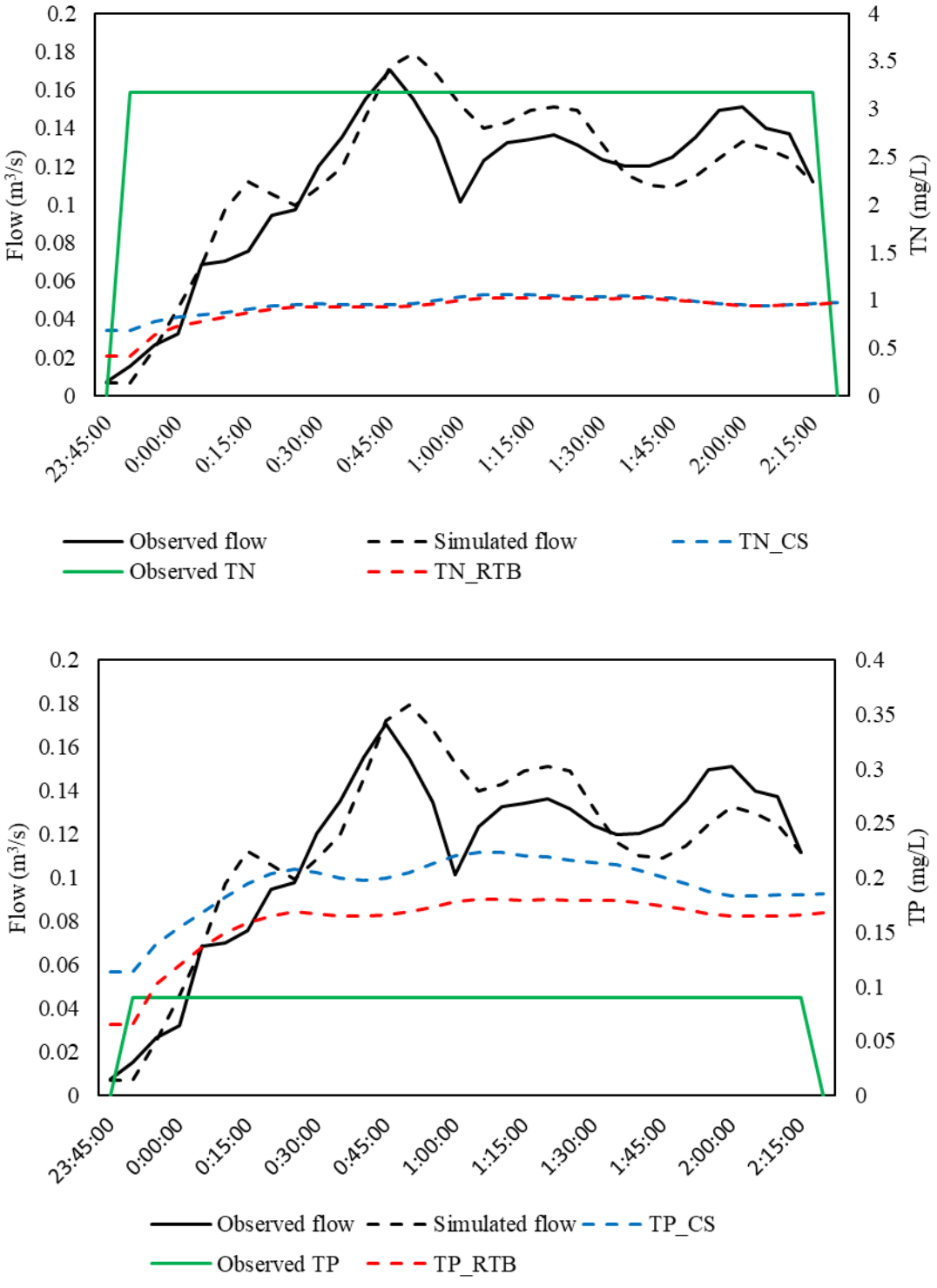
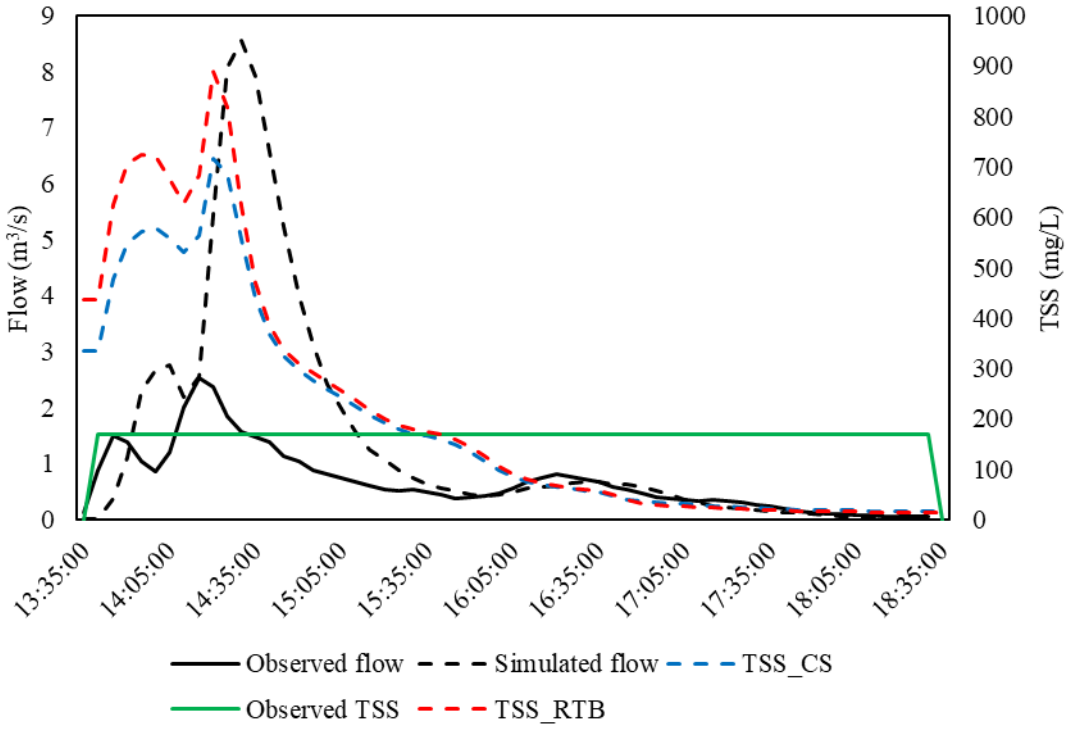
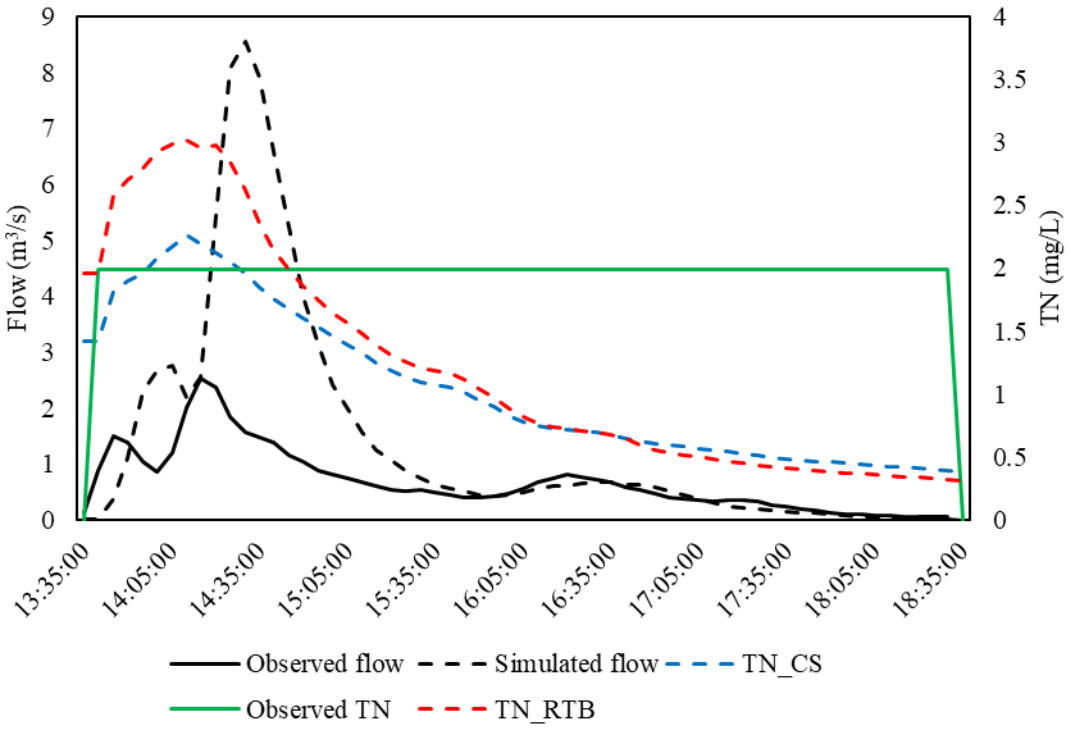


Fig. E1. Simulation results during a validation event on 6-28, 2018.


Flow (m³/s)
TSS (mg/L)
0
1
2
3
4
5
6
7
8
9
0
100
200
300
400
500
600
700
800
900
1000
13:35:00
14:05:00
14:35:00
15:05:00
15:35:00
16:05:00
16:35:00
17:05:00
17:35:00
18:05:00
18:35:00
Observed flow
Simulated flow
TSS_CS
Observed TSS
TSS_RTB


Flow (m³/s)
TN (mg/L)
0
1
2
3
4
5
6
7
8
9
0
0.5
1
1.5
2
2.5
3
3.5
4
13:35:00
14:05:00
14:35:00
15:05:00
15:35:00
16:05:00
16:35:00
17:05:00
17:35:00
18:05:00
18:35:00
Observed flow
Simulated flow
TN_CS
Observed TN
TN_RTB

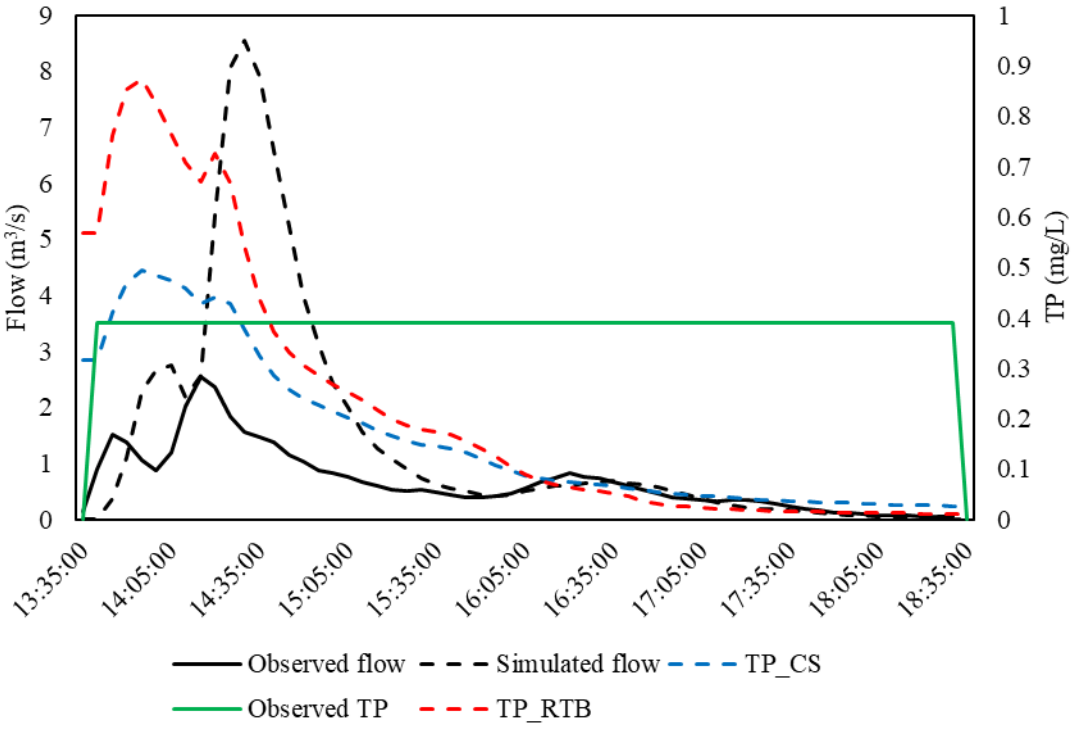
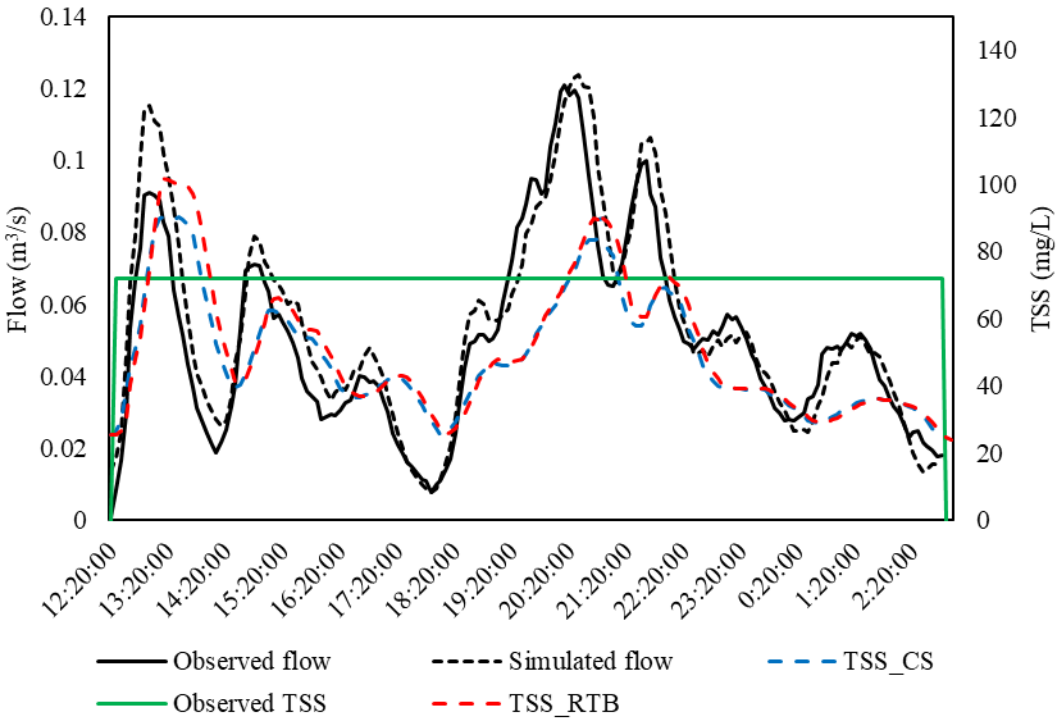


Fig. E2. Simulation results during a validation event on 6-22, 2018.



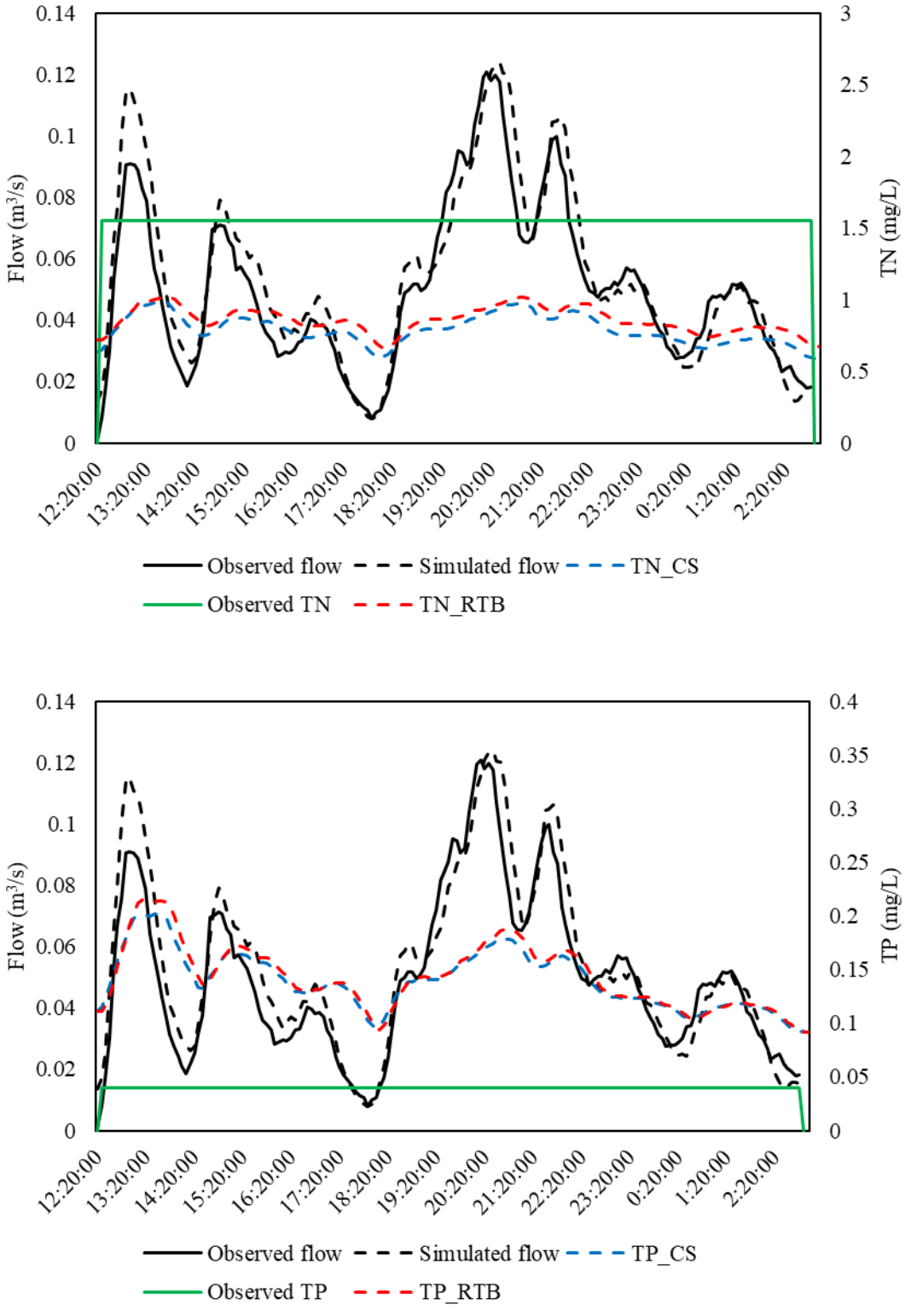
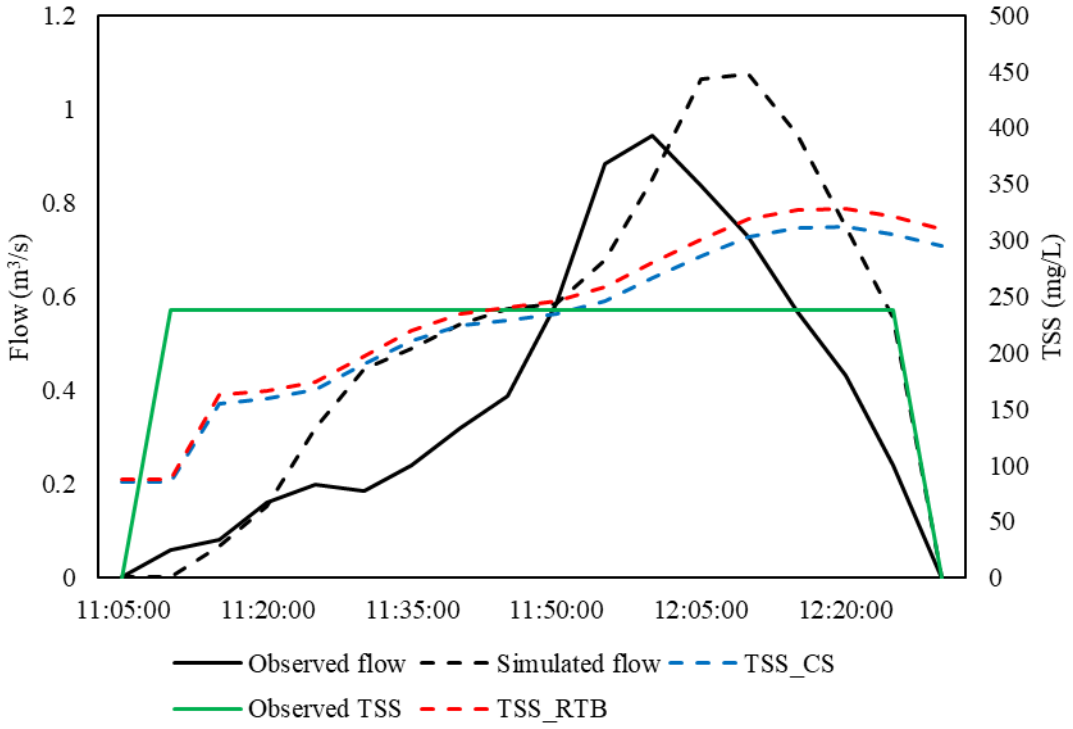
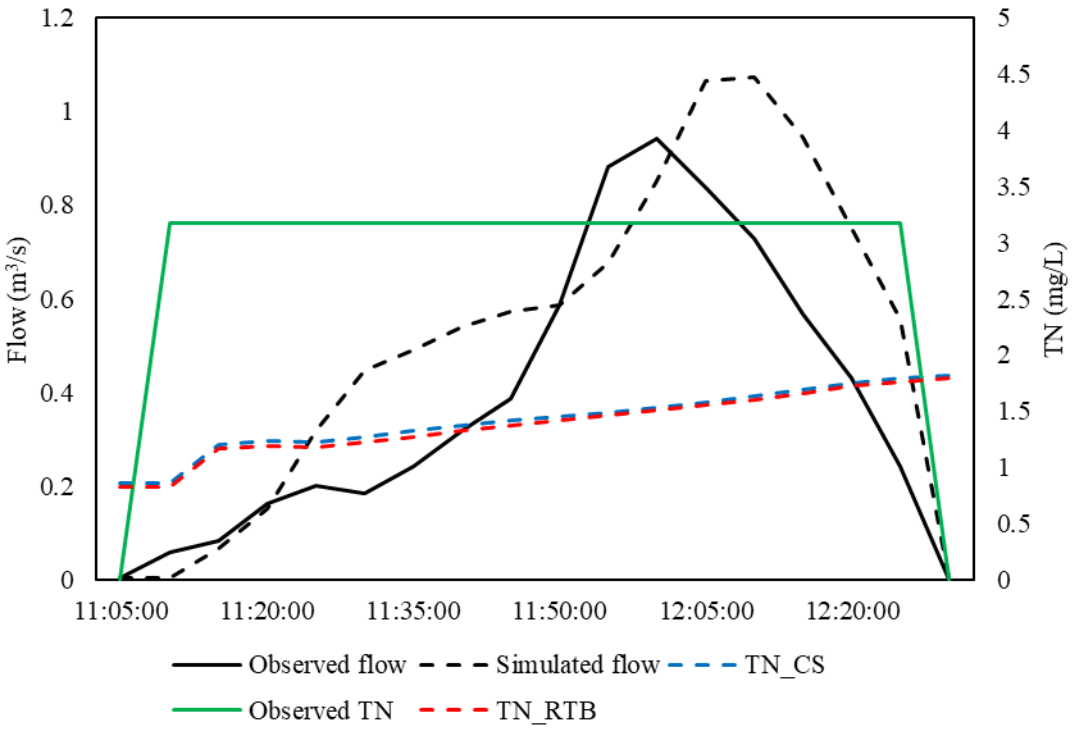


Fig. E3. Simulation results during a validation event on 9-27, 2018.


1.2
1
0.8
0.6
0.4
0.2
0
Flow (m3/s)
500
450
400
350
300
250
200
150
100
50
0
TSS (mg/L)
11:05:00
11:20:00
11:35:00
11:50:00
12:05:00
12:20:00
Observed flow
Simulated flow
TSS_CS
Observed TSS
TSS_RTB


1.2
1
0.8
0.6
0.4
0.2
0
Flow (m3/s)
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
TN (mg/L)
11:05:00
11:20:00
11:35:00
11:50:00
12:05:00
12:20:00
Observed flow
Simulated flow
TN_CS
Observed TN
TN_RTB

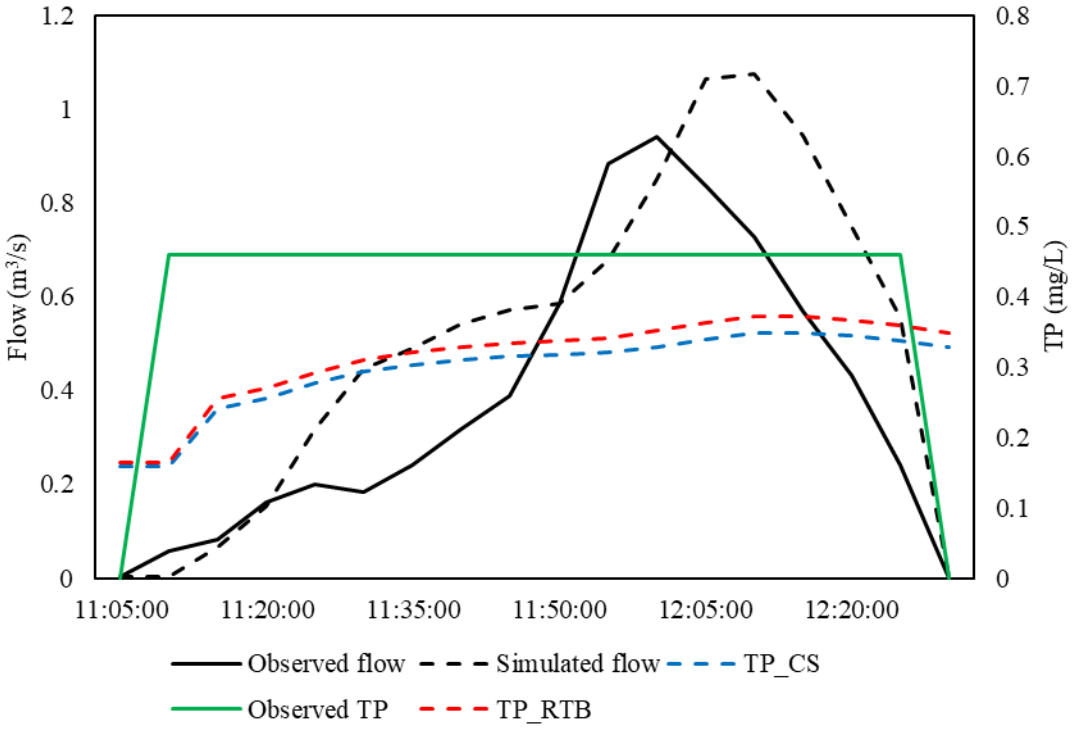
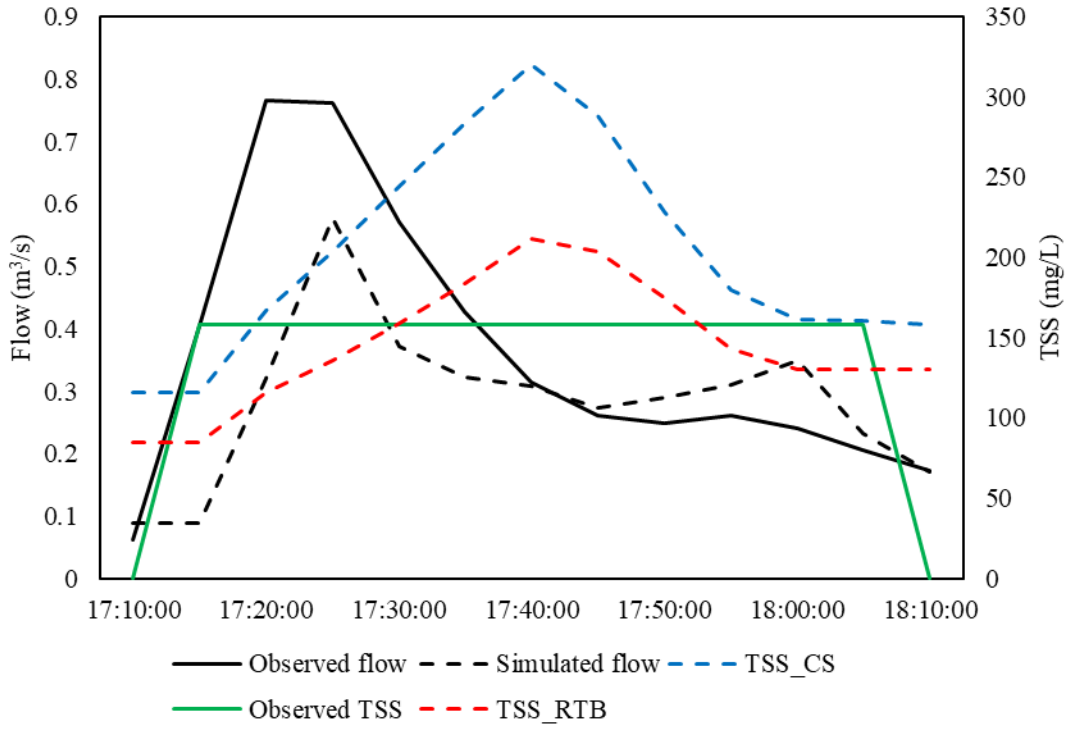


Fig. E4. Simulation results during a validation event on 8-24, 2018.



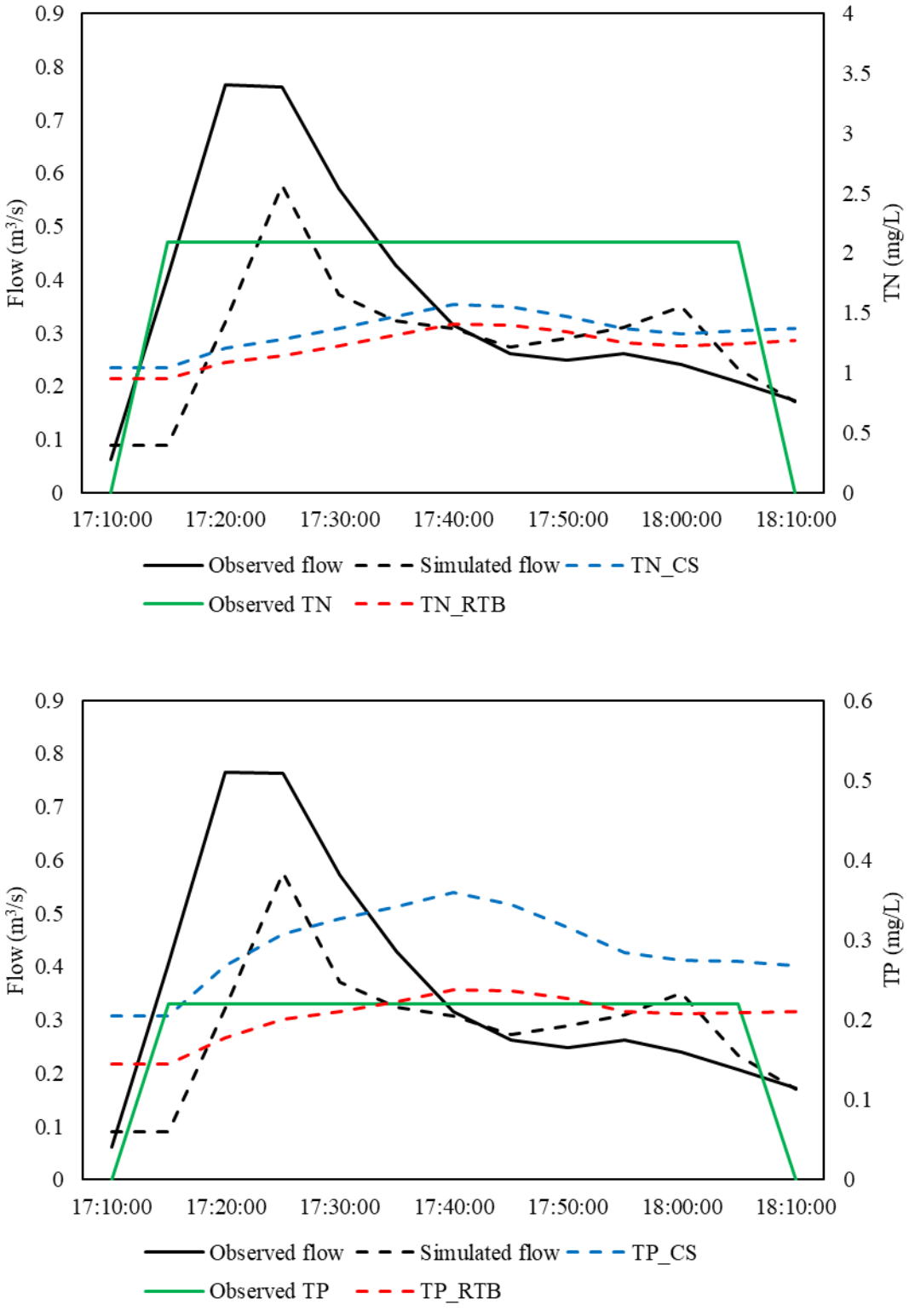
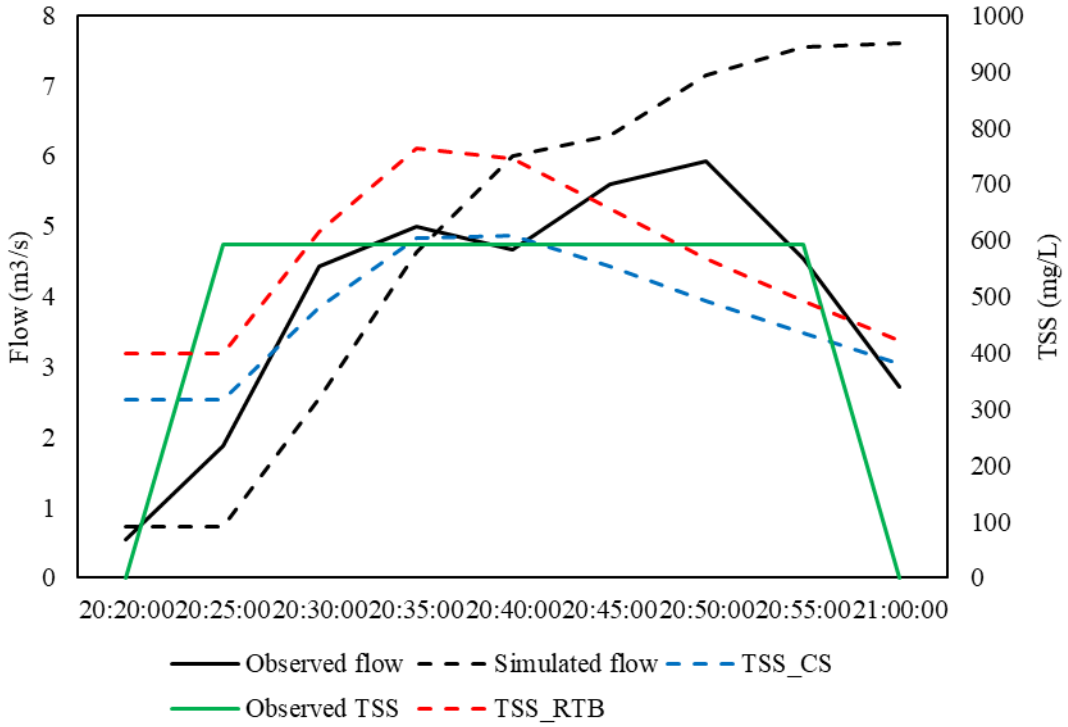
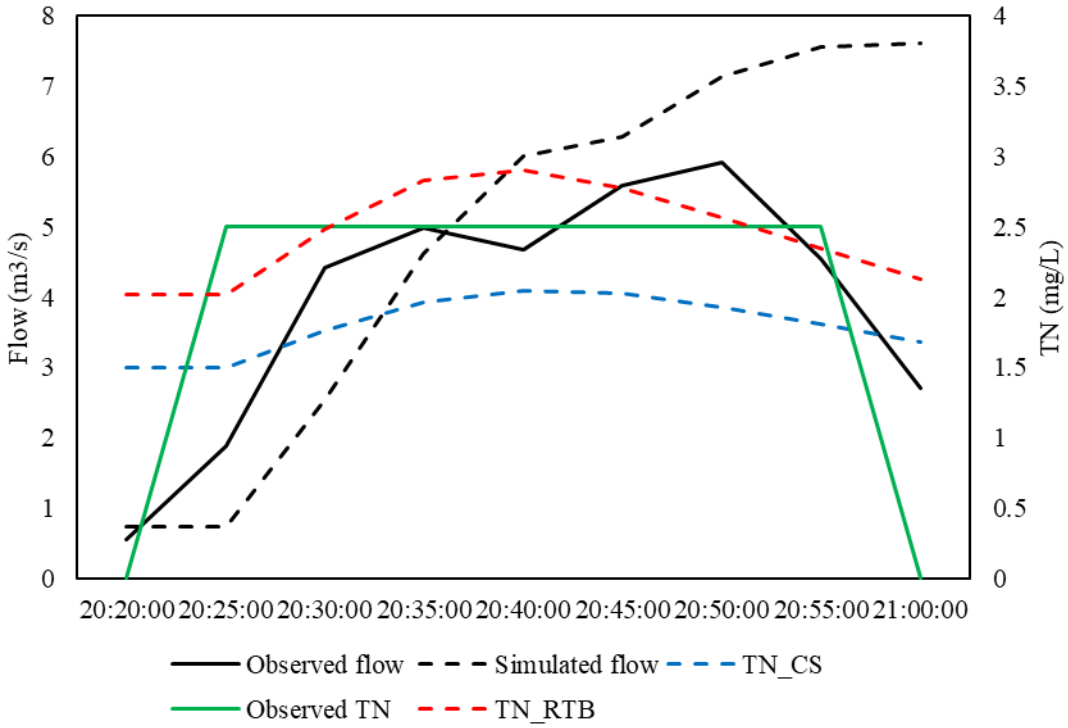


Fig. E5. Simulation results during a calibration event on 7-3, 2019.


Flow (m3/s)
TSS (mg/L)
0
1
2
3
4
5
6
7
8
0
100
200
300
400
500
600
700
800
900
1000
20:20:00
20:25:00
20:30:00
20:35:00
20:40:00
20:45:00
20:50:00
20:55:00
21:00:00
Observed flow
Simulated flow
TSS_CS
Observed TSS
TSS_RTB


Flow (m3/s)
TN (mg/L)
0
1
2
3
4
5
6
7
8
0
0.5
1
1.5
2
2.5
3
3.5
4
20:20:00
20:25:00
20:30:00
20:35:00
20:40:00
20:45:00
20:50:00
20:55:00
21:00:00
Observed flow
Simulated flow
TN_CS
Observed TN
TN_RTB

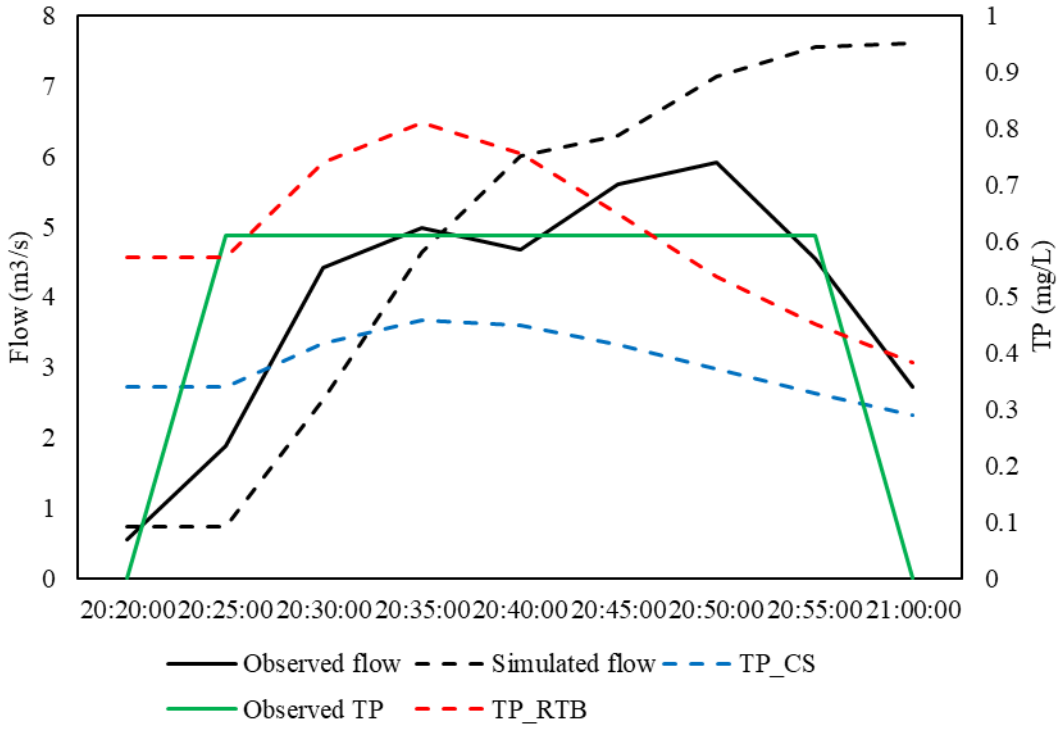
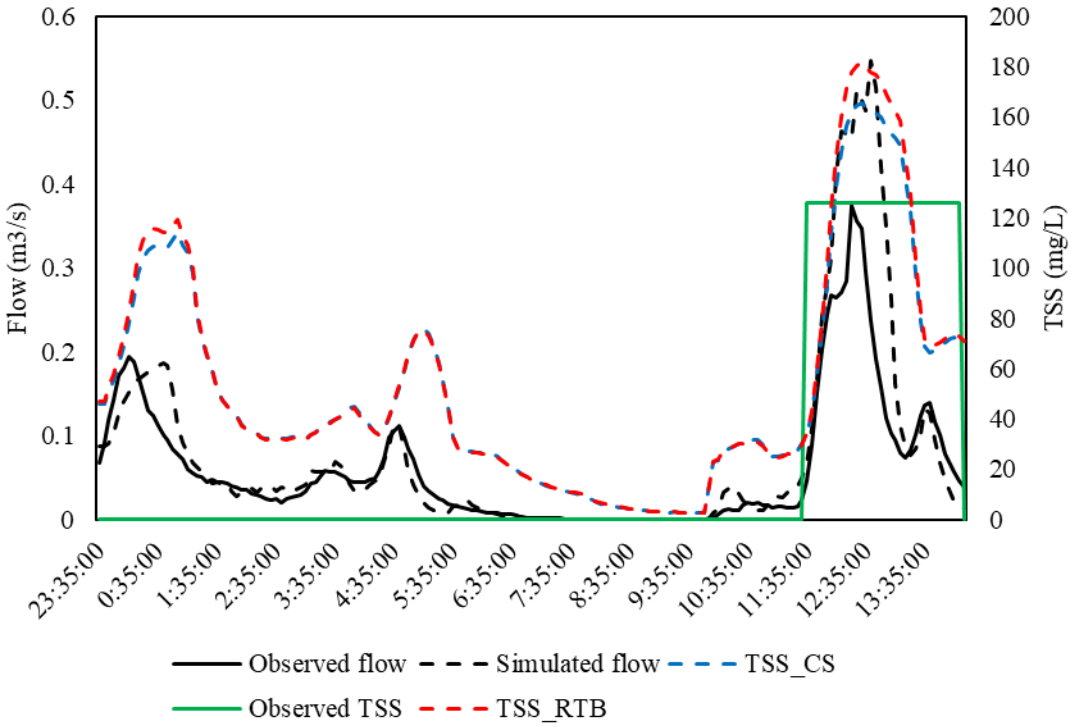
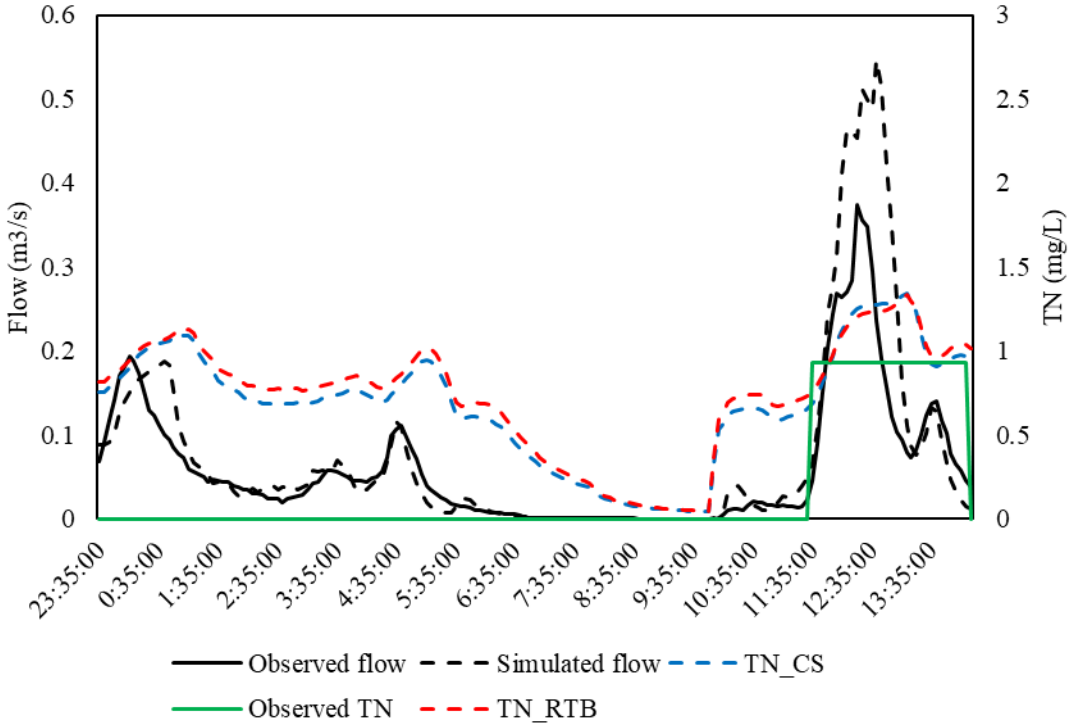


Fig. E6. Simulation results during a calibration event on 6-27, 2019.





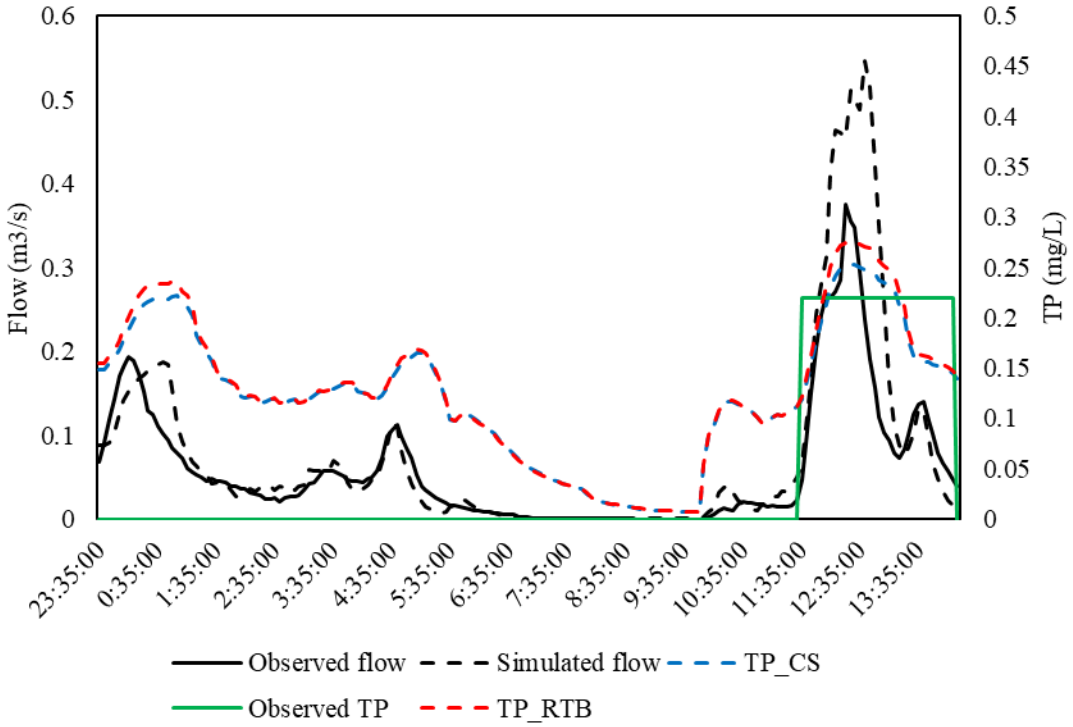
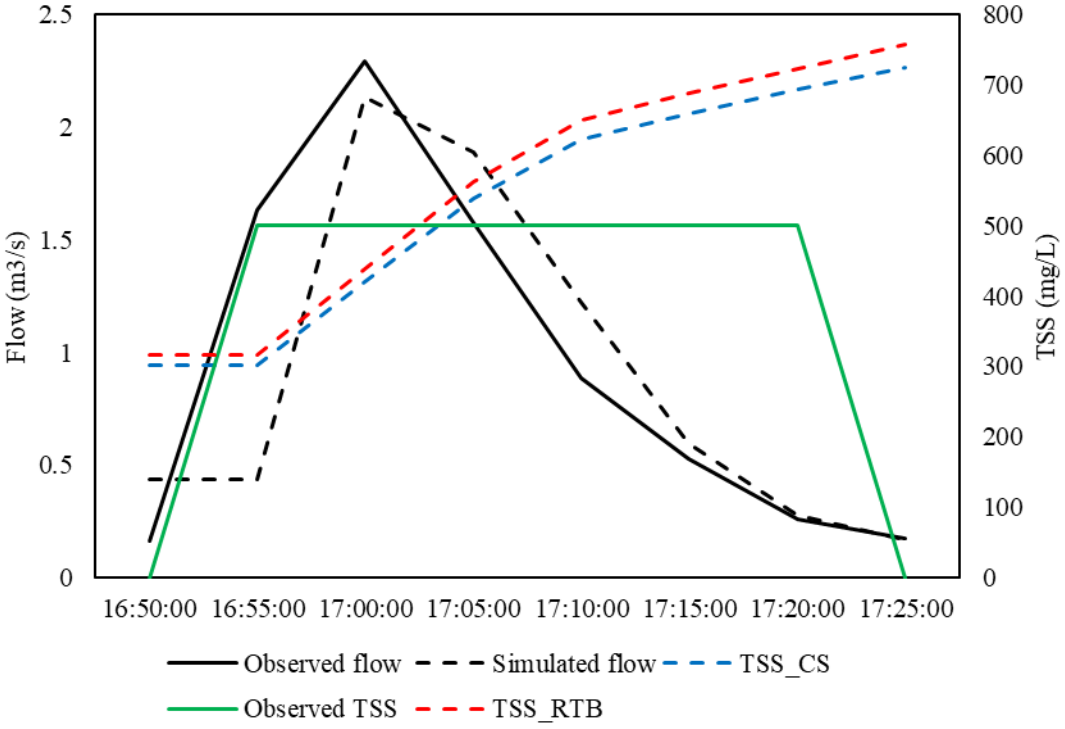
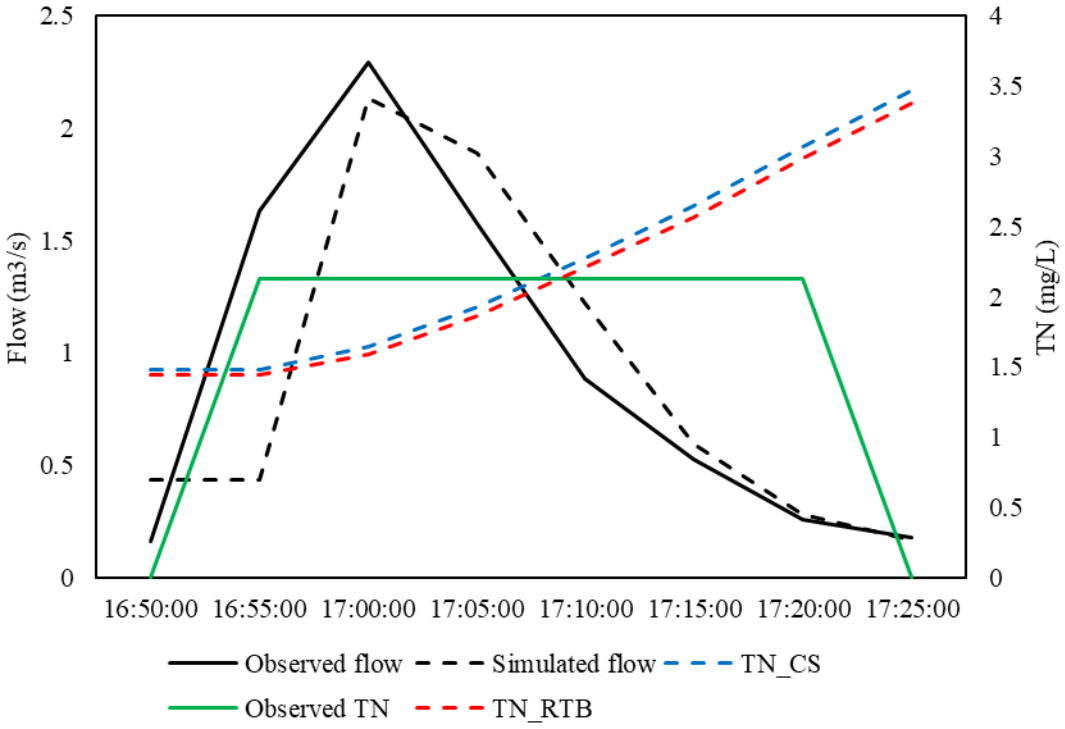


Fig. E7. Simulation results during a calibration event on 6-20, 2019.


2.5
2
1.5
1
0.5
0
Flow (m3/s)
800
700
600
500
400
300
200
100
0
TSS (mg/L)
16:50:00 16:55:00 17:00:00 17:05:00 17:10:00 17:15:00 17:20:00 17:25:00
Observed flow
Simulated flow
TSS_CS
Observed TSS
TSS_RTB


2.5
2
1.5
1
0.5
0
Flow (m3/s)
4
3.5
3
2.5
2
1.5
1
0.5
0
TN (mg/L)
16:50:00 16:55:00 17:00:00 17:05:00 17:10:00 17:15:00 17:20:00 17:25:00
Observed flow
Simulated flow
TN_CS
Observed TN
TN_RTB

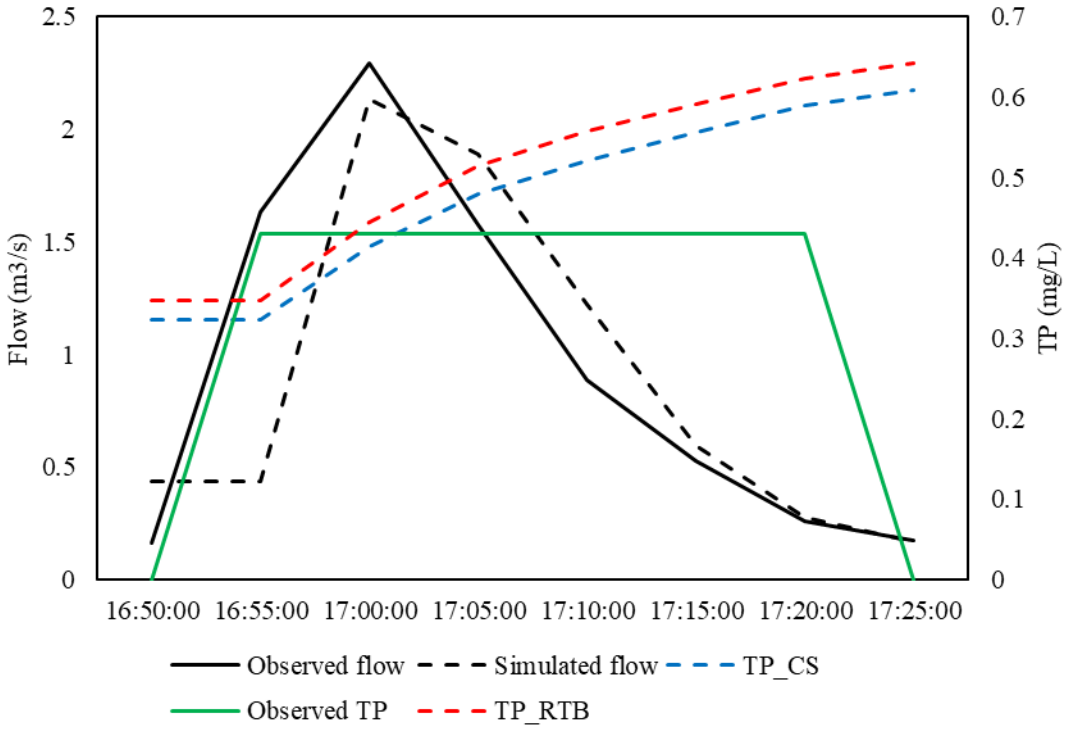


Fig. E8. Simulation results during a calibration event on 8-6, 2019.

Appendix F. Hydrological modeling results of LCB and WDT approaches.

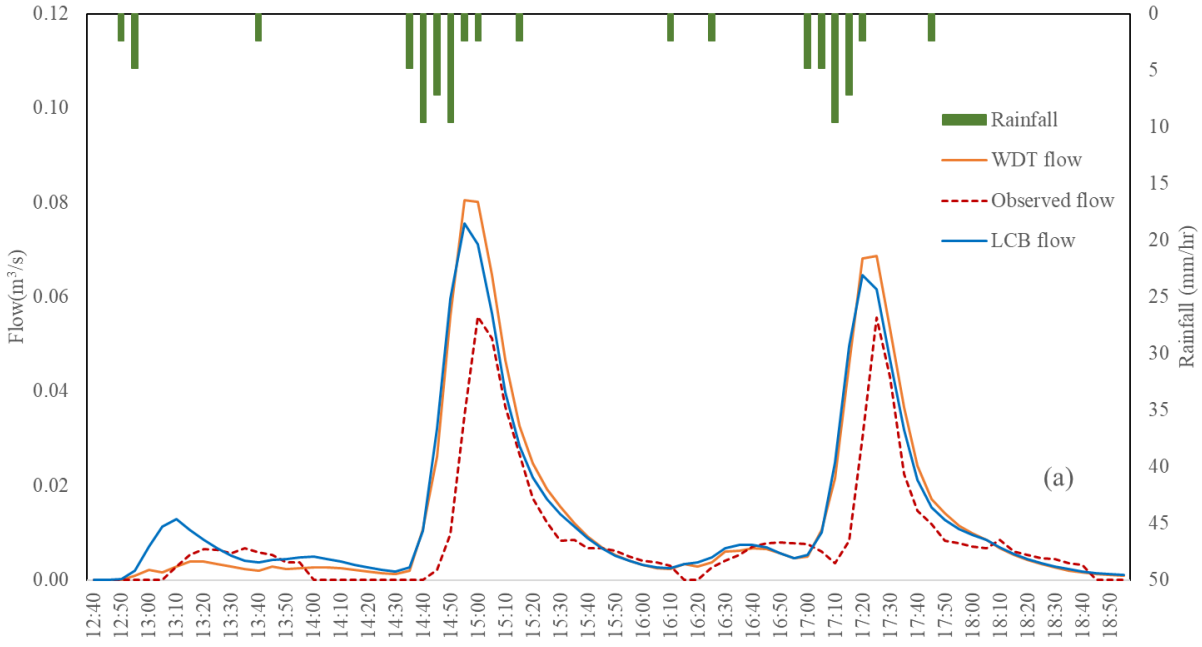


Fig. F1. Simulation results during an event on July 24, 2018, for RO catchment.

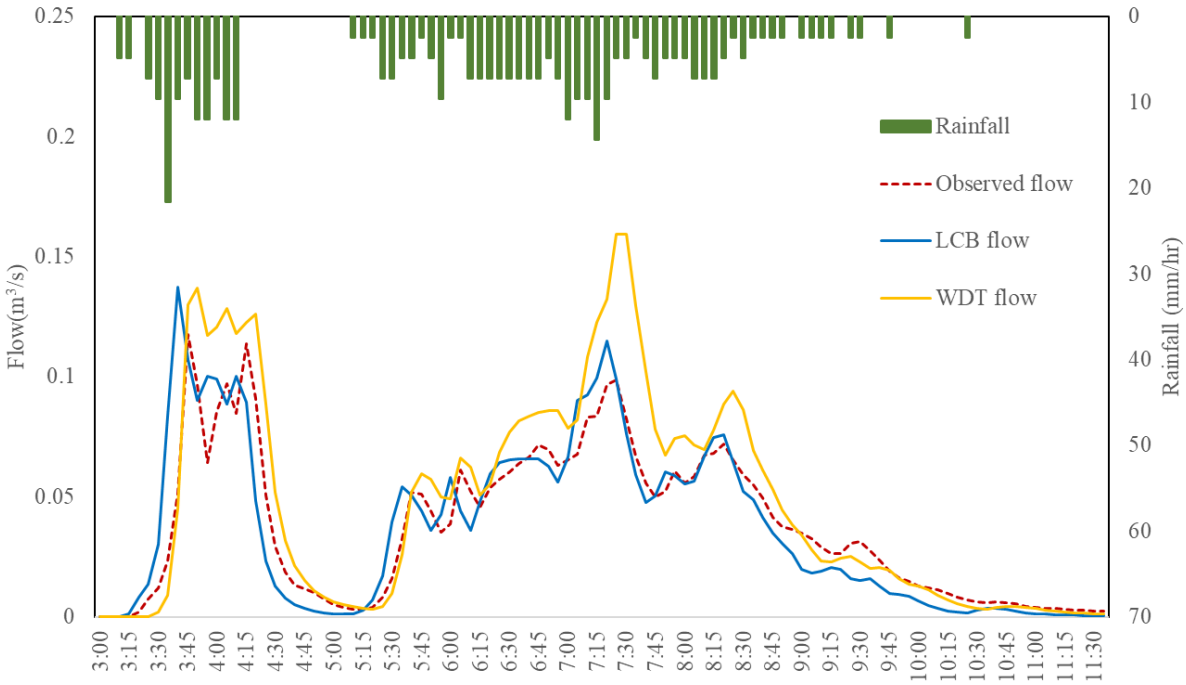


Fig. F2. Simulation results during an event on June 23, 2018, for RR catchment.

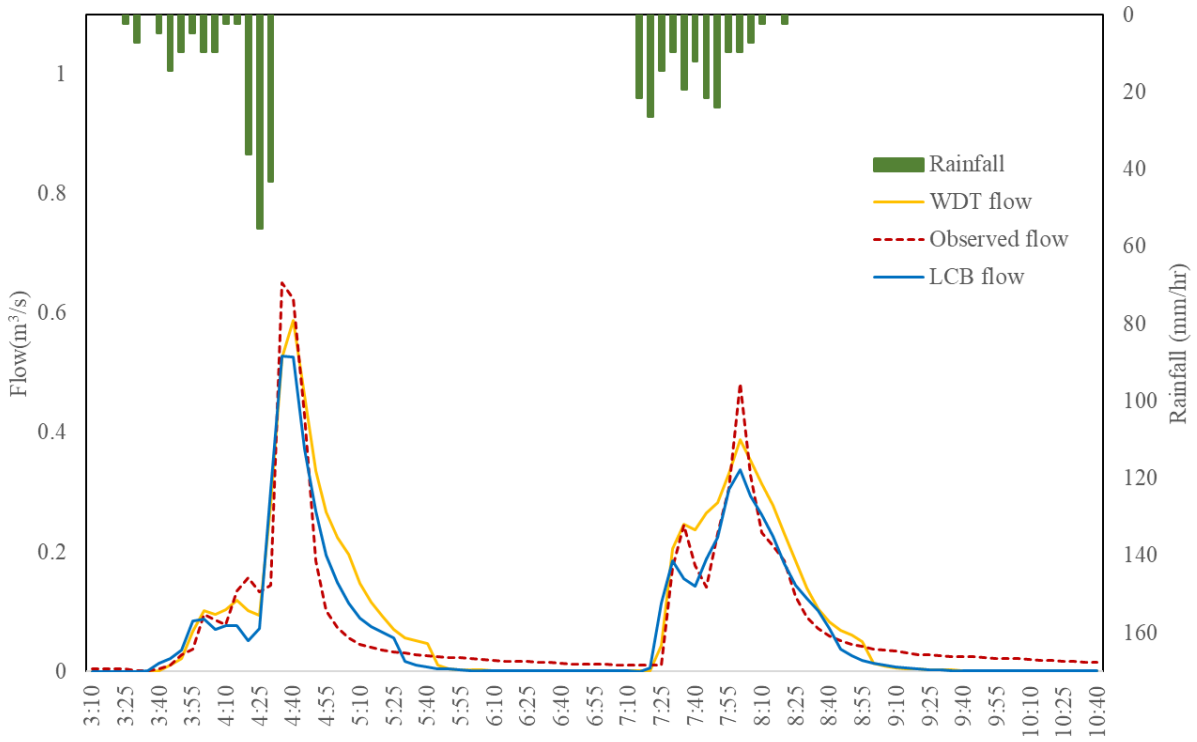


Fig. F3. Simulation results during an event on June 27, 2019, for RR catchment.

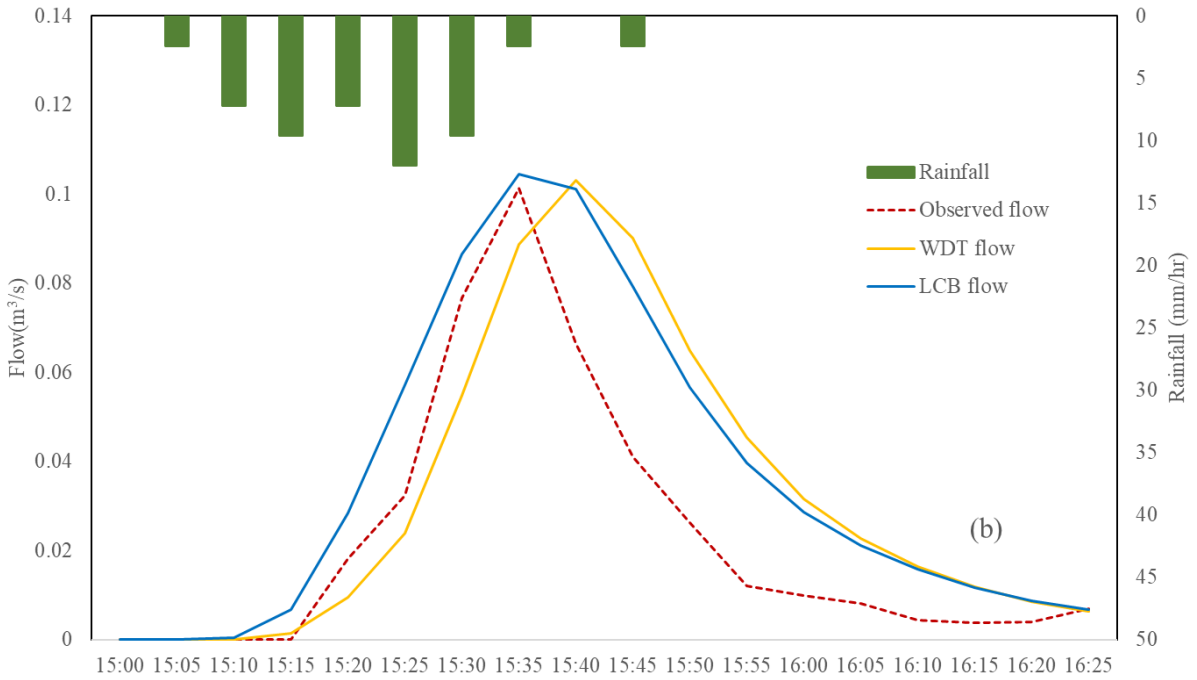


Fig. F4. Simulation results during an event on July 6, 2019, for RO catchment.