Theses and dissertations

1-1-2012

Modelling Low Impact Development Potential With Urban Hydrological Response Units

Marija Eric
Ryerson University

Follow this and additional works at: http://digitalcommons.ryerson.ca/dissertations

Part of the Environmental Sciences Commons

MODELLING LOW IMPACT DEVELOPMENT POTENTIAL WITH URBAN

HYDROLOGICAL RESPONSE UNITS

by

Marija Eric, Bachelor of Engineering, Ryerson University, 2010

A thesis presented to Ryerson University

in partial fulfillment of the requirements for the degree of Master of Applied Science in the

program of Environmental Engineering

Toronto, Ontario, Canada, 2012

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

Modelling Low Impact Development Potential with Urban Hydrological Response Units

Master of Applied Science 2012

Marija Eric

Environmental Engineering

Ryerson University

Abstract

The purpose of this thesis is to develop a methodology for hydrological modelling the performance of Low Impact Development technologies using an Urban Hydrological Response Unit approach. The K-Means Cluster Analysis procedure was carried out to create clusters of lot parcels which represented the Urban Hydrological Response Units. Different sampling methods were used to select lots from each of the clusters to model before and after Low Impact Development implementation. The runoff response (m^3) of an approximate final cluster centre was used to calculate the total runoff (m^3) of each cluster. After adding the total runoff (m^3) for a group of 15 clusters, the benchmark runoff value (m^3) from modelling all lots was closely approached with and without Low Impact Development. A random sample of 7 % and 90 % of lots from each cluster for a group of three clusters closely approached the benchmark runoff value (m^3) for both no Low Impact Development and Low Impact Development respectively.

Acknowledgements

Table of Contents

List of Tables

List of Figures

List of Appendices

1. Introduction

1.1 Background

In ancient times, rainwater was stored near homes and used. Excess rainwater that flows off of impervious surfaces such as streets and sidewalks is termed runoff. The Romans were the first group of people to devise advanced techniques for eliminating urban runoff. During this period, runoff began to be viewed as a nuisance that causes flooding which results in damage and inconvenience. The belief was held that runoff should be removed from developed areas in a rapid and effective manner (Carmon and Shamir 2010).

At present, runoff can be viewed as a resource and not simply as a nuisance. Runoff quality is the subject of a guide published by Engineers Australia, National Committee for Water Engineering (2006). A relatively recent and severe drought in Australia caused stormwater management, with respect to water quantity and quality, to become a national issue. Several countries have undertaken research efforts contemporaneously to develop new approaches to stormwater management, although there has been little communication between each other (Carmon and Shamir 2010).

In the United States of America (USA), the main objective for stormwater management since the 1970s has been to protect water quality in streams and lakes from overland flow pollution. New approaches have been suggested and tested to promote such objectives (Carmon and Shamir 2010). Examples of these new approaches include limiting imperviousness, compact development and increased housing density, and low impact development (LID).

According to Carmon and Shamir (2010), LID was first developed by Prince George's County in 1999 in Maryland, USA. In essence, LIDs are devices or techniques which are used to mimic pre-development conditions. Common examples include rain barrels and downspout disconnections as shown in Figures 1 and 2:

Source: Google Images 2012

Figure 1. A rain barrel or rainwater harvesting LID.

Source: Google Images 2012

Figure 2. A disconnected downspout LID.

The volume of runoff leaving the LID unit after each storm or rain event must be the same as it was before development. Therefore, the pre-development hydrological response of the area serviced by the LID unit must be maintained. Another objective of LID implementation is pollutant loading reduction (Carmon and Shamir 2010).

The United States Environmental Protection Agency (US EPA) has recognized LID as a leading planning approach for runoff management. In general, LIDs can be divided into three categories: (1) urban land use practices, (2) land cover design, and (3) constructed facilities. The first category, urban land use practices, includes common examples such as higher-density development which can serve to lower runoff per housing unit as well as pollution loads. Combining different types of land use such as housing, employment, and services reduces the size of parking lots and the amount of area occupied by roads and sidewalks. This type of land use practice is known as mixed land use and can also be considered as a LID since the purpose is to decrease the amount of impervious area (Carmon and Shamir 2010).

The second category, land cover design, entails transforming urban runoff from a nuisance into a water resource mainly by infiltrating clean urban runoff into the ground. Urban runoff from small and medium rainstorm events, which comprise 90 % of all rainfall events, can be used directly for irrigation and landscaping. It can also be used indirectly through recharge into the groundwater (Carmon and Shamir 2010). Non-structural measures such as different layouts of roads and buildings to minimize imperviousness and maximize the use of pervious soils and vegetation are also examples of land cover design strategies (Elliott and Trowsdale 2007).

The third LID category is constructed facilities and includes previously mentioned common devices and techniques such as rain barrels in addition to others like pervious pavement, swales, filter strips and bioretention devices (Carmon and Shamir 2010). LID places special emphasis

on on-site small-scale control of stormwater sources for which there are many design guidelines available (Elliott and Trowsdale 2007). Constructed LIDs can be roughly classified into three categories: (1) point structures such as parking areas covered with pervious pavement, (2) linear structures such as swales along a road, and (3) local reservoirs which provide volumetric space for detention, retention, infiltration, and wetlands. It is important that the size and design of the local reservoir fit into the built area without endangering residents (Carmon and Shamir 2010).

In order to assess the effectiveness of LID constructed facilities, hydrological modelling can be performed. For example, the runoff generated by the LID constructed facility can be determined before and after implementation. If the LID constructed facility or technology performs successfully, the runoff generated after implementation should be less than the runoff generated before implementation.

There are various approaches towards the hydrological modelling of LID technologies. Lot-based approaches apply LID technologies on individual lot parcels. Multiple-lot-based approaches focus on stormwater issues over parcels of land. Lastly, right-of-way-based approaches manage the transport of stormwater from one point to another (Li et al. 2010). In this research study, a lot-based approach was used to apply the rain barrel LID technology on individual lot parcels and hydrological modelling was performed.

Growth in land development and urbanization has led to an increase in impervious surfaces such as concrete that prevent water from infiltrating (Gilroy and McCuen 2009). By preventing infiltration, environmental changes ensue such as decreases in the volume of groundwater recharge and lower water tables. Other changes include decreases in base flows during dry periods, decreases in time of concentration, and a diminishment in water quality (Dietz 2007).

Consequently, stream channels have deteriorated and the conveyance of pollutants has increased (Gilroy and McCuen 2009). Other results of urbanization are increased flood flows and stream erosion along with water contamination from suspended sediments, hydrocarbons, nutrients, heavy metals, and pathogens. A decline in habitat and water quality of urban streams has been recorded on a global scale (Elliott and Trowsdale 2007).

Conventionally, stormwater management methods have strived to remove water from a site as quickly as possible and then store it as a larger volume at an off-site, downstream facility such as a detention pond or an infiltration basin in order to control the peak discharge. Although these methods may have controlled the downstream peak discharge rates, they have not improved the issues surrounding increased runoff volume. As a result, first-order streams and aquatic life within them usually disappear (Gilroy and McCuen 2009).

LIDs offer a way to restore the hydrologic cycle and return some of the rainwater back to the environment by infiltration. This contrasts with conventional or traditional stormwater management techniques such as downstream ponds, which do not address the hydrologic cycle. Traditional stormwater management efficiently captures, conveys, and treats rainfall-induced runoff that is generated on impervious surfaces. A popular example in many urban centres is the combined sewer system which conveys household sewage and stormwater runoff to water pollution control facilities for treatment. However, when runoff exceeds the available system capacity combined sewer overflows (CSOs) occur as direct discharges to water bodies. Conventional approaches to CSO abatement can be effective but are often expensive and difficult to site (Alfredo 2007).

LIDs have therefore been suggested as an alternative to traditional stormwater management. LID methods attempt to control rainfall on-site, preferably near the source and are sometimes

referred to as source controls (Gilroy and McCuen 2009). LID devices are designed to detain, store, infiltrate, or treat urban runoff (Elliott and Trowsdale 2007). The ultimate goal of LID implementation is to deliver post-development characteristics of lot runoff which do not differ significantly from pre-development conditions (Gilroy and McCuen 2009).

1.2 Problem Definition

The conventional approach towards evaluating the benefits of LID implementation has been to aggregate or lump together their characteristics such as storage volume or increased infiltration capacity in some way in a large scale model. An evaluation of the benefits of LID implementation can extend all the way up to a watershed scale. The relative scales used in measuring the performance of LID implementation are illustrated in Figure 3:
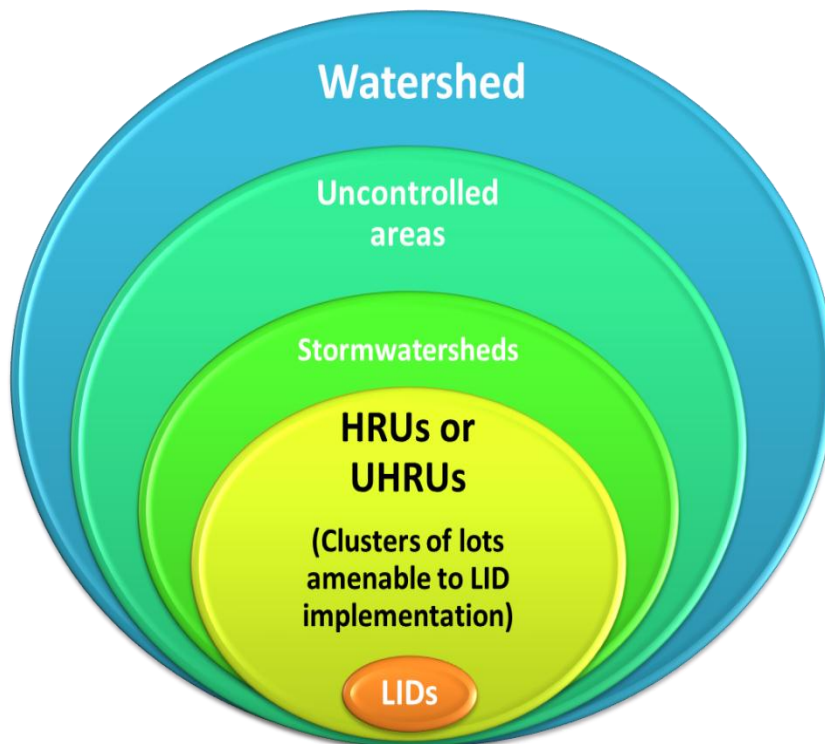
Source: Li et al. 2010

Figure 3. Scales used in measuring the performance of LID implementation.

As Figure 3 illustrates, the LID unit is usually orders of magnitude smaller in size than the watershed. This presents a challenge for representing LID units in watershed models. In a study by Li et al. (2010), hydrological modelling was performed on lot-based LID practices. Lots which were amenable to a particular LID combination were grouped together and taken to be a Hydrologic Response Unit (HRU). The results from modelling a HRU were extrapolated to the entire watershed using a spreadsheet model, and then to neighbouring municipalities (Li et al. 2010). A watershed is defined as the land that contributes runoff to a given site and can be used interchangeably with other terms such as basin, catchment, and drainage area (James and Wurbs 2002).

The current research also presents an approach which is focused on evaluating the benefits of LID implementation at the lot-level. However, the HRUs were constructed using a clustering analysis procedure and the results were not extrapolated to an entire watershed. The proposed approach involves the concept of Urban Hydrological Response Units (UHRUs) which are urban HRUs. UHRUs and HRUs are drainage areas that exhibit similar runoff generating mechanisms as a result of similar hydrological characteristics such as slope, lot area, and soil type (Li et al. 2010). UHRUs are HRUs located in urban centres. Drainage areas which comprise an UHRU also contain similar hydrological characteristics such as level of imperviousness and slope. By applying the HRU or UHRU concept, large study areas with similar hydrological characteristics can be evaluated more efficiently.

The research presented in this thesis is an extension of the previous study by Li et al. (2010), in which the results from evaluating areas with similar hydrological characteristics or HRUs were extrapolated to other areas that shared these characteristics. This method was used to evaluate the performance of LID technology systems on pre-defined areas referred to as uncontrolled

areas. Uncontrolled areas are areas which have never been subjected to conventional stormwater management practices and where there are no opportunities for the implementation of stormwater management ponds. The study concluded that LID technology systems can be simulated or modelled at the lot-level and then extrapolated to the larger scale of the watershed-level (Li et al. 2010).

In order to assess the feasibility and effectiveness of LID implementation, the performance of LIDs was simulated by conducting hydrological modelling. Hydrological modelling can be conducted on large areas such as watersheds in different ways. A conceptual model can be used to evaluate the entire area or every aspect or feature of the watershed can be modelled individually. The drawback to evaluating an entire area is that it has the potential for the loss of a great amount of lot-level detail. However, the notion of modelling every aspect or feature of a watershed individually is very impractical. It is therefore necessary to devise an innovative approach based on a conceptual model to evaluate the entire region. Furthermore, there is lack of guidance for modellers and engineers as to how to evaluate the benefits of LID implementation.

1.3 Objective

The overall objective of this research is to develop a process to model the stormwater performance of LIDs using an UHRU concept. The study area was first disaggregated into units which share similar hydrological characteristics such as slope, soil type and % imperviousness by performing cluster analyses in order to identify the UHRUs. The performance of the LIDs was measured by using runoff volume (m^3). By running a small computer program in conjunction with hydrological modelling software, hydrological modelling was performed simultaneously for thousands of lots.

8

1.4 Scope

The hydrological modelling method that was developed is generalized and is only limited by the data and the application. The data which was supplied for this research pertained to Southern Ontario. The study area was the uncontrolled areas in the City of Barrie. The City of Barrie is an urban centre of about 140,000 people located within the Lake Simcoe Watershed in Southern Ontario, Canada (The City of Barrie, 2012). The uncontrolled areas within the City of Barrie are shown on the map in Figure 4:
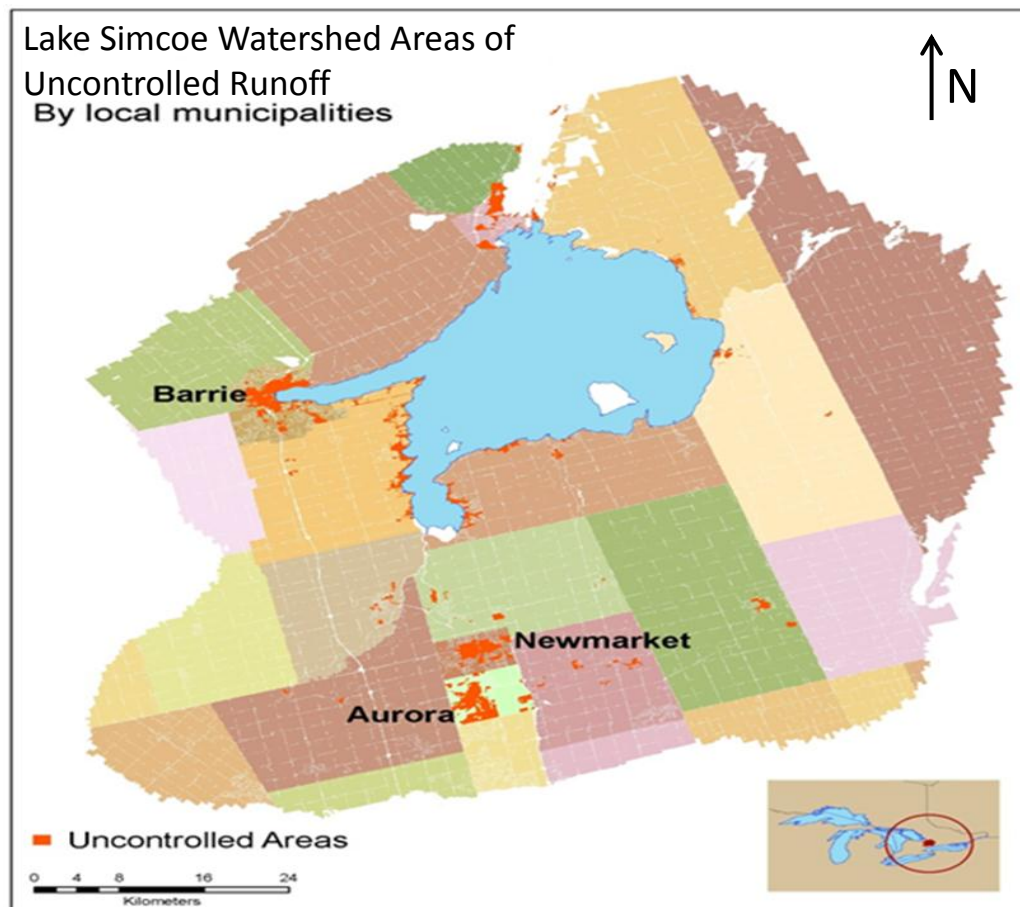
Source: Li et al. 2010

Figure 4. Map of uncontrolled areas within the City of Barrie.

The map in Figure 4 also shows the placement of the City of Barrie within the Lake Simcoe Watershed.

The hydrological modelling method that was developed can be applied to other regions outside of Southern Ontario. However, certain criteria have to be defined to determine whether the methodology is appropriate for any application. Data pertaining to the infrastructure, hydrological, and land use information is required for LID selection and hydrologic analysis. The lot characteristics dictate the cluster analysis and the distribution of these characteristics will determine the sampling procedure.

The data which was used in this study was obtained from the study of Li et al. (2010) which was also conducted for a part of Southern Ontario and the assumptions which were developed related to this area. However, if the study area changes the assumptions may change which could affect the validity of the modelling method. While the other study examined 7 individual LIDs and 17 LID combinations, this study will only examine the rain barrel LID that is specifically designed for residential lots.

There were also some issues encountered with the data used in the study by Li et al. (2010) that posed challenges in this research. This is because the data used in the current study was obtained from the previous study as already processed data. A common limitation of both the previous study and the current study is the availability of data. The previous study involved many preliminary Geographic Information Systems (GIS) processes which depend on accurate and abundant spatial data for the study area. However, various issues with respect to the GIS data were encountered. For example, there were sometimes gaps in the data and imperfect registration of buildings on lots. Table 1 illustrates the issues with respect to the availability of data for the previous study:

Table 1. Issues with respect to the availability of GIS data from the previous study.

Layers	Barrie	Newmarket	East Gwillimbury	Aurora
Parcels	Comprehensive (from LSRCA/Teranet)			
Parking	Completed	Present	None	None
Driveways	Present	None	None	None
Buildings	Comprehensive (except attributes)	Comprehensive (except attributes)	None	None
Land Use	Satisfactory (from LSRCA and DMTI)			Comprehensive
Sidewalks	Present	Require restructuring (lines)	None	Incomplete, lines
Soils	Complete (Hydrographic Classes and Depths, from LSRCA)			
Roads	Completed (lines)	Require restructuring (lines)	Present (lines)	Present (lines)
Storm Drainage System	Present (lines)	Comprehensive	Present	Present
Ditches	Present	None	None	None
Parks	Present	Present	from LSRCA	Present
DEM/DTM	from LSRCA			

Source: Li et al. 2010

According to Table 1, the City of Barrie presented the least amount of data issues. Table 1 also illustrates the importance of having data that is suitable to the objectives of the research. Therefore, one of the limitations of the current research study is the requirement for an extensive data set that can be used to identify areas with similar hydrological characteristics for residential lots.

Another limitation imposed by the data was the measurements of the lengths of the lots in the City of Barrie. Data for the lengths of the lots were provided along with the areas of the lots and these two measurements were used to calculate width values for all of the residential lots. The width of the lot is an important parameter for hydrological modelling. The software that was used for hydrological modelling was the United States Environmental Protection Agency Storm

Water Management Model (US EPA SWMM). The version of US EPA SWMM software that was used was Version 5.0.022 with Low Impact Development (LID) Controls.

Using the GIS software known as ArcMap, it was determined that the given lengths of the lots were logically inconsistent. Consequently, some of the derived widths were also somewhat inaccurate. The width parameter required for the software is not always the actual width of the lot. For the purposes of SWMM software, the width parameter that is input into the software is referred to as the characteristic width of the overland flow path for sheet flow and should be recorded in feet or meters (Rossman 2010).

The characteristic width is defined as the subcatchment's area divided by the length of the longest overland flow path that water can travel. The resulting answer might not coincide with the conventional width of the lot. Close inspection of the lot is required to determine the longest overland flowpath of runoff for the lot. If there are several such paths then the average of their lengths is used to compute the width. For urban catchments, the typical overland flow length is the length from the back of the representative lot to the centre of the street (Davis et al. 2010).

Nonetheless, some lots may drain to the rear of the lot. They may also be blocked by buildings or other sturctures while others may not have a rectangular or square shape. Therefore for the purposes of this research, all of the lots were assumed to be rectangular or square and all of the derived widths were assumed to be the characteristic widths required for the hydrological modelling process. The following chapters include a literature review, further details regarding the case study, a description of the methodology, the results and discussion, and the conclusions and recommendations.

2. Literature Review

2.1 LID Trends

New urban water management approaches have been introduced in the past 25 years in an attempt to deliver improved environmental, economic, social, and cultural outcomes. These approaches are usually referred to as LID although other acronyms are sometimes used. Examples of other acronyms are SUDS (sustainable urban drainage systems), WSUD (water sensitive urban design), and LIUDD (low impact urban design and development) which is popular in New Zealand (Elliott and Trowsdale 2007). Another acronym which is often used interchangeably with LID is BMP (best management practices).

An example of a relatively new urban water management approach is water-sensitive planning (WSP). The aim of WSP is to promote sustainable development and construction in a cost-effective way and with the participation of the surrounding community. The WSP approach was established for newly developed areas (greenfields). However the approach is still applicable to infill and redevelopment projects (brownfields), as well as fully developed urban areas (Carmon and Shamir 2010).

The WSP approach emphasizes that stormwater management should always take into account the context of the whole relevant catchment area. Both the flows upstream and the flows downstream of the area under planning should be considered as well as local conditions. WSP advocates that three aspects of stormwater management should be minimized referred to as the 3Ms, at each scale involved beginning from a single lot all the way up to the entire city. The first minimization criterion involves minimizing the difference in runoff volume leaving the post-development and pre-development areas. The second minimization criterion involves

minimizing the difference in discharge leaving the post-development and pre-development areas. The third minimization criterion is to minimize the pollutant load in the runoff leaving the post-development area. By following the 3Ms, the natural hydrological balance is maintained (Carmon and Shamir 2010). The focus of this research was on the first minimization criterion as the runoff volume before and after LID implementation was compared.

An implication of the WSP approach is the requirement for a change in work habits and work paradigms. Hence, three work paradigms have been suggested by the WSP approach. The first paradigm declares that the role of water resources should be integral throughout an entire planning project. All professionals involved in the project should work together regardless of their different backgrounds. The second paradigm asserts that the selection process for the location and spatial layout of any planned development should begin with the natural hydro-geographical structure. The natural hydro-geographical structure encompasses hydrological characteristics such as slopes, soil types, and water bodies and streams. Some of these hydrological characteristics were a part of the study in this thesis and formed the starting point of investigation. The second paradigm was therefore followed in this research. Lastly, the third paradigm requires the treatment of stormwater runoff as a resource instead of as a nuisance (Carmon and Shamir 2010).

Nevertheless, the WSP approach developed by Carmon and Shamir (2010) was a work in progress for several reasons. One reason for example was that the addition of other fields to WSP, such as urban wastewater management, was still pending (Carmon and Shamir 2010). Other studies have lent support to some of the suggestions put forth by the WSP approach. For instance, it has been shown that LID practices are effective, in general, at preserving the natural hydrologic function of a site and retaining pollutants when applied in developed areas (Dietz

14

2007). The retention of pollutants confirms that the third minimization criterion for stormwater management has been followed. Moreover, the preservation of the natural hydrologic function of the site indicates that all three minimization criteria have been applied. It can therefore be demonstrated that the implementation of LID practices follows the stormwater management approach of WSP.

It has, however, been suggested that it is necessary for long-term studies of all LID practices to be conducted since few studies exist on how these technologies perform for long periods of time (Dietz 2007). The time period for this research study was only eight months, from April 1 to November 1 which is considered to be the rainy season (Li et al. 2010). A more thorough investigation of the factors that influence the efficacy of LIDs should also be undertaken. The optimal number of LIDs to implement and the placement of LIDs should be investigated further. Only one LID was implemented in this study and the placement of the LID was assumed to be the same for each simulation. Other factors such as the effects of return period and land use types should also be explored (Gilroy and McCuen 2007). The focus of this study was on residential properties, therefore the land use type was not varied. However, the influences of other hydrological factors which influence the efficacy of LIDs were investigated such as average slope (%) and soil type.

Another possible area of research is the ability of LID systems to retain and destroy viruses and bacteria. Additionally, investigations with respect to the minimization of phosphorus export by bioretention and green roof LIDs are also needed. Bioretention, green roof, and grass swale LIDs have all been shown to export phosphorus. For bioretention and green roofs, this is possibly due to the high phosphorus levels already present in the soil. It can also be due to the fertilization of turf or planted areas which are possible causes for grass swales as well. The

15

export of phosphorous can lead to excess loading of total phosphorus (TP) to surfaces waters which may exacerbate an already existing problem. Therefore, different media mixtures for LIDs which export phosphorus can be examined (Dietz 2007).

Since an initial US EPA literature review was published in 2000, advances have been made with respect to individual LID practices and new questions have been raised. Common questions address the suitability of LID for all sites, groundwater contamination, and winter performance of LID practices. Nonetheless, the pervious pavement and bioretention LIDs have demonstrated to work effectively in cold climates with frost in the ground. The ability of LID systems to work in cold climates underscores the importance of proper design and installation (Dietz 2007).

However, conditions exist in which it is not appropriate to use an LID practice that relies on infiltration. For instance, areas that are already highly contaminated may risk even more contamination from LID practices. Examples of such areas are gas stations, recycling centres, or brownfield areas. Traditional stormwater management practices and pavement may also be more suitable for areas with steep slopes, shallow (< 3 ft.) depth to bedrock, or seasonal high water table. It is unusual for an entire area to be characterized by such conditions which renders LID practices as widely applicable (Dietz 2007).

There is some contradiction with respect to the growth in popularity of LID practices. In one study, it was found that the progression towards adopting LID practices has been considered slow although there has been increasing awareness and knowledge of sustainable development. A lack of LID design tools such as available software was thought to be a possible reason for the slow progression (Elliott and Trowsdale 2007). However, in 2007 the US EPA described the increased use of LID as a popular movement by governmental and private-sector agencies (Carmon and Shamir 2010).

Another contradiction found in LID studies concerns cost. In the initial literature review published by the US EPA in 2000, cost savings of 15 % to 80 % were estimated for LID implementation relative to conventional design in all cases examined except one (Carmon and Shamir 2010). However, it has also been claimed that the installation costs of LID technologies are generally more expensive than conventional stormwater infrastructure although they can be more cost-effective on a volumetric basis for storing stormwater in the landscape (Alfredo 2007).

2.2 LIDs and the UHRU Concept

One study concluded that more effective and available tools could aid in the efficiency of the design and application of LID, as well as demonstrate outcomes which could be used for education and policy development (Elliot and Trowsdale 2007). The UHRU approach proposed in this thesis can perhaps contribute to education and policy development with respect to LIDs by facilitating the hydrological modelling process.

The HRU or UHRU concept is not brand new and was introduced around 1972 (Cowan et al. 2008). It has been applied and explored in various hydrological studies. In this study, different GIS layers were intersected with one another and the distributions of hydrological attributes over the study area were determined. The UHRUs were then constructed using clustering analysis to group lots with similar hydrological attributes into the same cluster which then became an UHRU.

Similarly, HRUs formed the conceptual framework in a study which investigated the Soil and Water Assessment Tool (SWAT) model. In this study, a modified version of the SWAT model divided sub-basins into HRUs by intersecting soil topographic index (STI) data and land use rasters (GIS shapefiles) as illustrated in Figure 5:
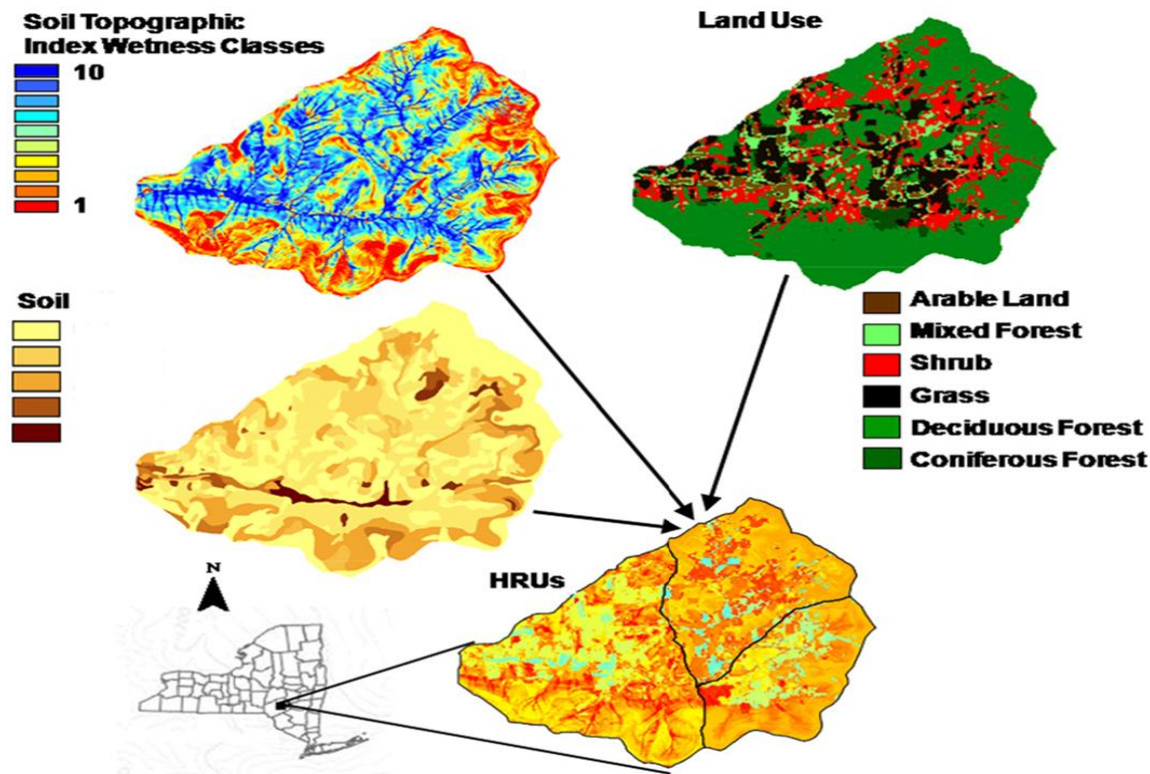
Source: Cowan et al. 2008

Figure 5. Intersection of STI and land use rasters to form HRUs.

The SWAT models under investigation were applied to a sub-watershed in upstate New York to compare the predictions made by the model regarding responses such as surface runoff, shallowly perched water table depth, and stream phosphorus loads against direct measures. The underlying spatial distribution of the data was ignored as SWAT lumped the parameters in the data into HRUs (Cowan et al. 2008). The standard version of SWAT uses soil type and land use type to define the area of each HRU which is the conventional method of defining HRUs. The delineation of HRUs using unconventional hydrological attributes such as average slope (%) and imperviousness (%) is an area that was explored by this thesis.

The UHRU concept was applied in a study that investigated the details pertaining to the circulation of rainwater within urban areas. This is an area of research that requires further

18

investigation since the majority of studies on this topic usually only address the runoff on impervious surfaces. This research gap, however, was not addressed by the current study. In general, the major flow component is the runoff of rainwater from impervious surfaces for design purposes (Andrieu 2008).

A distributed hydrological model was developed and adapted to urban areas in order to provide a more comprehensive account of the hydrological mechanisms of urban catchment areas. The model was tested at two different scales: (1) a small and homogeneous urban catchment and a (2) medium-sized and more heterogeneous urban catchment. An urban hydrological element (UHE) was the smallest component of the first scale and was composed of a cadastral parcel and its corresponding adjacent street segments. Map layers in an urban databank were used to describe the geometrical characteristics of UHEs such as surface area, the impervious fraction, the vegetation fraction, and slope. In Figure 6, a 2D representation of the UHE is shown and a map which illustrates the different layers of an urban databank is shown in Figure 7.
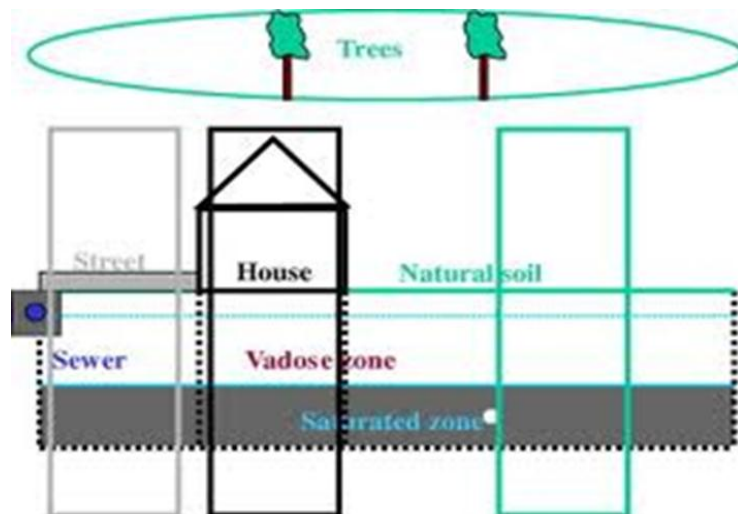
Source: Andrieu et al. 2008

Figure 6. A 2D diagram of the UHE including three vertical profiles.

19

Cadastral parcel		Street segment
Building		Street section
Trees		Sewer system segment
		Elevation point
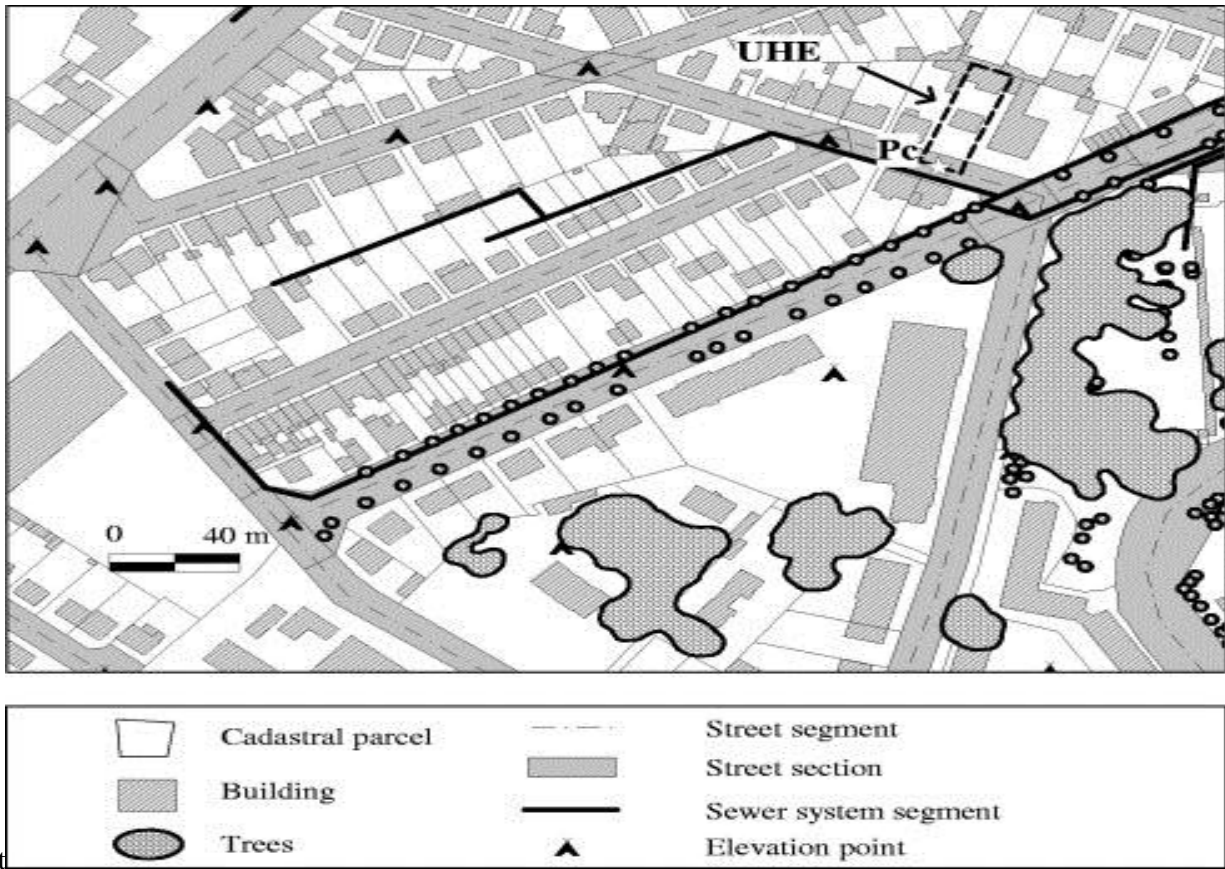
Source: Andrieu et al. 2008

Figure 7. A map representing the different layers of an urban databank.

The small and homogeneous urban catchment can be likened to a UHRU. The urban catchment was viewed as being comprised of UHEs which were connected to the catchment outlet by a runoff branching structure (RBS). The map layers in the urban databank were used to describe the hydrological characteristics of the UHEs in a similar manner to how the GIS layers in the current research were used to describe and form the UHRUs. The major finding of the study was the significance of urban soil and soil-atmosphere interaction on hydrological mechanisms of urban areas when performing hydrological modelling over longer time periods such as seasons and years (Andrieu 2008).

20

In a study by Elliott et al. (2009), an approach to modelling the cumulative catchment-scale effects of a few LID devices was investigated. The approach entailed the disintegration of a catchment to the scale of a land parcel or finer and the modelling of each LID device separately. The study focused on examining the effects of aggregation on predictions of water quantity and quality for detention, infiltration, and bioretention devices. The simulations were based on total suspended solids which was the representative contaminant studied (Elliott et al. 2009). This study further demonstrates the application of the HRU or UHRU concept.

A detailed stormwater model was devised for a suburban catchment in which each individual property and road section between catchbasins was represented as a separate source area. Each source area contained a hypothetical stormwater device and can be regarded as a HRU. The catchment was initially subdivided into 810 source areas for the detailed level of aggregation. The aggregation increased to various levels until the extreme scenario of a single source area and a single aggregated device was reached.

Hydrological modelling was conducted for each level of aggregation and the effects of aggregation on key summary measures of flow and water quality were examined. The study concluded that the aggregation of on-site devices and associated source areas did not impact predictions of mean flow, baseflow, and water quality at the outlet of a catchment a great deal. This conclusion was made using the conditions and assumptions which the study was based upon (Elliott et al. 2009). The conclusions drawn from this study support the method of extrapolating the results of modelling HRUs to an entire watershed (Li et al. 2010).

2.3 UHRUs and Clustering Analysis

In another study which was carried out by Kessler et al. (2001), the effect of spatial resolution on the hydrological modelling process was examined. The study entailed a hydrological micro-model that was developed and applied to a neighbourhood which was typical of urban development in that region. The model was used to analyze the effects of urban development on infiltration and runoff. It was also used to evaluate a number of practices which were designed to enhance on-site infiltration such as infiltration trenches and grassed swales. Drawbacks to the micro-approach included the potential for increased flooding within the yard itself, the flooding or wetting of basements, and the reduction in beneficial watercourse flows downstream (Kessler et al. 2001).

The basic approach focused on the residential lots of an urban neighbourhood just as the focus of the current study was on residential lots in a similar kind of area. However in the study by Kessler et al. (2001), the residential lots were combined with relatively small public areas that were dotted among them as well as the adjacent sidewalks and roads. The hydrological model was applied at three different spatial scales which were termed micro, mezzo, and macro. The micro, mezzo, and macro spatial scales were defined respectively as: (1) the individual residential lot, (2) the cluster of residential lots and their immediate vicinity, and (3) a whole neighbourhood (Kessler et al. 2001).

The mezzo spatial scale was comprised of six subcatchments which included three clusters, two separate public areas, and a street area. Each cluster has a similar composition in terms of land use pattern and can therefore be regarded as a UHRU. The arrangement of the lots within each cluster is shown in Figure 8:
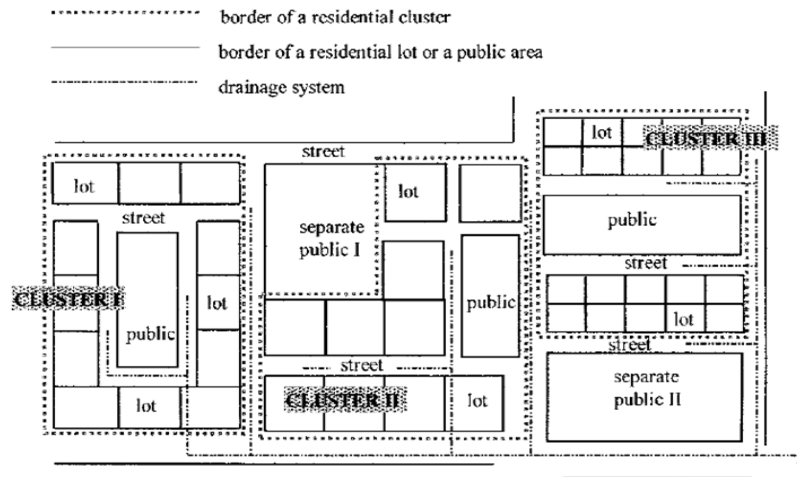
FIG. 2.   Synthetic Urban Neighborhood

Source: Kessler et al. 2001

Figure 8. The composition of the clusters used in the mezzo scale analysis.

Although the clusters illustrated by Figure 8 were not formed using a clustering algorithm, the

mezzo scale analysis demonstrates the applicability of the clustering concept to form UHRUs.

There were numerous conclusions which were reached by the development of the hydrological

micro-model. One conclusion was that the hydrological response computed at the lot scale can

be extrapolated to yield the response of an entire neighbourhood with reasonable accuracy by

adding the responses of individual units. This conclusion is similar to one of the conclusions

reached in the case study regarding the Lake Simcoe Watershed (Li et al. 2010). It was also

concluded that draining the roof area through an infiltration trench of appropriate size can be

almost as effective as connecting it to infiltration strips. The infiltration trench was found to be

quite effective where the upper foil layer is compacted by the building process while the soil that

remains below has higher conductivity. In general, a more detailed model was deemed to be

23

necessary in order to carry out the simulations of hydrological processes at the lot scale (Kessler et al. 2001).

HRUs or UHRUs were previously defined as areas that contain similar hydrological characteristics such as average slope (%) and soil type. In this research, the K-Means Cluster Analysis Procedure was used as the classification method to group together lots with similar hydrological characteristics. In other research, a Bayesian clustering scheme was used to separate 280 catchments located in the Eastern United States into nine classes. The nine classes were analyzed with respect to their hydrologic, as well as climatic and landscape attributes. The study was conducted to determine the extent to which widely available hydrologic information (precipitation-temperature-streamflow) can be used to create a first order grouping of hydrologically similar catchments (Carrillo et al. 2011).

Based on empirical evidence, the study developed two hypotheses. The second hypothesis described spatial proximity as a good first indicator of hydrologic similarity because of the strong control climate exerts on catchment function and because it varies slowly in space. In the end, eleven clusters were identified which showed some degree of connectivity in terms of spatial distribution. This suggests that spatial proximity is a good indicator of similarity in hydrological attributes which is likely due to the slow changes in climate and landscape characteristics throughout space (Carrillo et al. 2011). The lots within the study area of the current research must therefore contain similar hydrological attributes since they are located very closely to one another.

The study by Carrillo et al. (2011) defines cluster analysis as "the process of grouping similar entities (catchments) according to one or more chosen similarity measures (signatures), while concurrently separating those that are different" (p. 4506). The experiment used the fuzzy

partitioning Bayesian mixture clustering algorithm in which a "good" cluster analysis was considered to be one that shows relatively good separation between the clusters.

The K-Means Cluster Analysis procedure, which was used in this research, uses the Euclidean distance as a similarity or dissimilarity measure. The Euclidean distance is a commonly used method to measure the dissimilarity or distance between two data objects (Gan et al. 2007). The Euclidean distance between two vectors $x = (x_1, x_2, \ldots, x_p)^T$ and $y = (y_1, y_2, \ldots, y_p)^T$ is calculated as follows in Equation 1:

$$d_{euc}(x, y) = (\textstyle\sum_{j=1}^{p}(x_j - y_j)^2)^{1/2} \ (1)$$

Using Equation 1, cases or lots are assigned to clusters and the cluster centres are calculated as the mean of all of the cases in the cluster.

The Euclidean distance is a dissimilarity measure that is sensitive to the differences in the magnitudes or scales of the input variables. It therefore seems necessary to standardize the input data variables before clustering analysis is performed. Some studies have recommended data standardization whereas others have suggested that it may not be advisable. Consequently, a simulation study by Cooper and Milligan (1988) examined the standardization problem by generating a variety of data structures. The inter-cluster spacing and the scales of the input variables of the data structures were varied. The data sets were examined in four different types of error environments and the results for eight standardization strategies were presented. The study concluded that approaches which standardize by dividing the variable by the range gave consistently superior recovery of the underlying data structure (Cooper and Milligan 1988). This finding supported the use of the data standardization technique which used the range of the variable, "Range 2" that was selected after a visual analysis of the cluster output was performed.

3. **Case Study: Evaluation of LIDs in the Lake Simcoe Region**

The current study attempts to enhance the HRU approach developed by Li et al. (2010). The approach taken in the case study entailed the application of the HRU concept to evaluate the implementation of LIDs and combinations of LIDs in the Lake Simcoe Watershed. A histogram was created based on the available data for lots which were amenable to each LID practice or combination of LID practices. The histogram created for lots amenable to the application of bioretention LIDs in residential areas is illustrated in Figure 9. A bioretention or rain garden is a shallow depression with plants that is designed to retain or detain stormwater before it is infiltrated or discharged downstream (Acomb and Clark 2008).
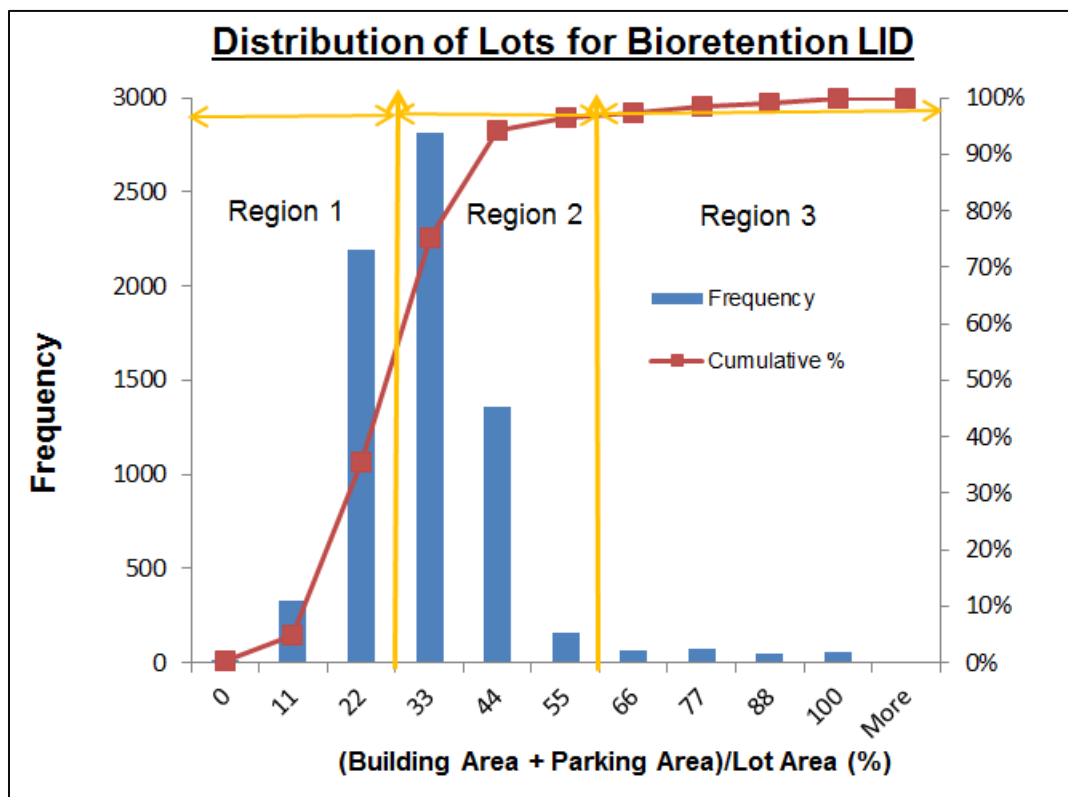
Source: adapted from Li et al. 2010

Figure 9. Distribution of residential lots amenable to the bioretention LID.

Histograms such as the one in Figure 9 were used to select representative lots for modelling the lot-based LIDs using SWMM software. The x-axis was divided into three regions based on the percentage value (0 % to 33 %, 34 % to 66 %, and 67 % to 100 %) of criteria such as (building area + parking area)/lot area (%) or imperviousness (%). One lot was selected from each of the three regions for hydrological modelling. SWMM was used to model the representative lots for the scenarios of both with and without LID implementation. Performance curves and associated equations were then developed based on the data derived from the hydrological modelling procedure. The performance curves plotted the percentage reduction in runoff versus the criterion used to develop the histogram (Li et al. 2010). The performance curve for the bioretention LID is illustrated in Figure 10:

Performance Curve for Biorention LID

$y = 0.2023x^{0.9714}$
$R^2 = 0.9988$

Runoff Reduction (%) vs (Building Area +Parking Area)/Lot Area (%)
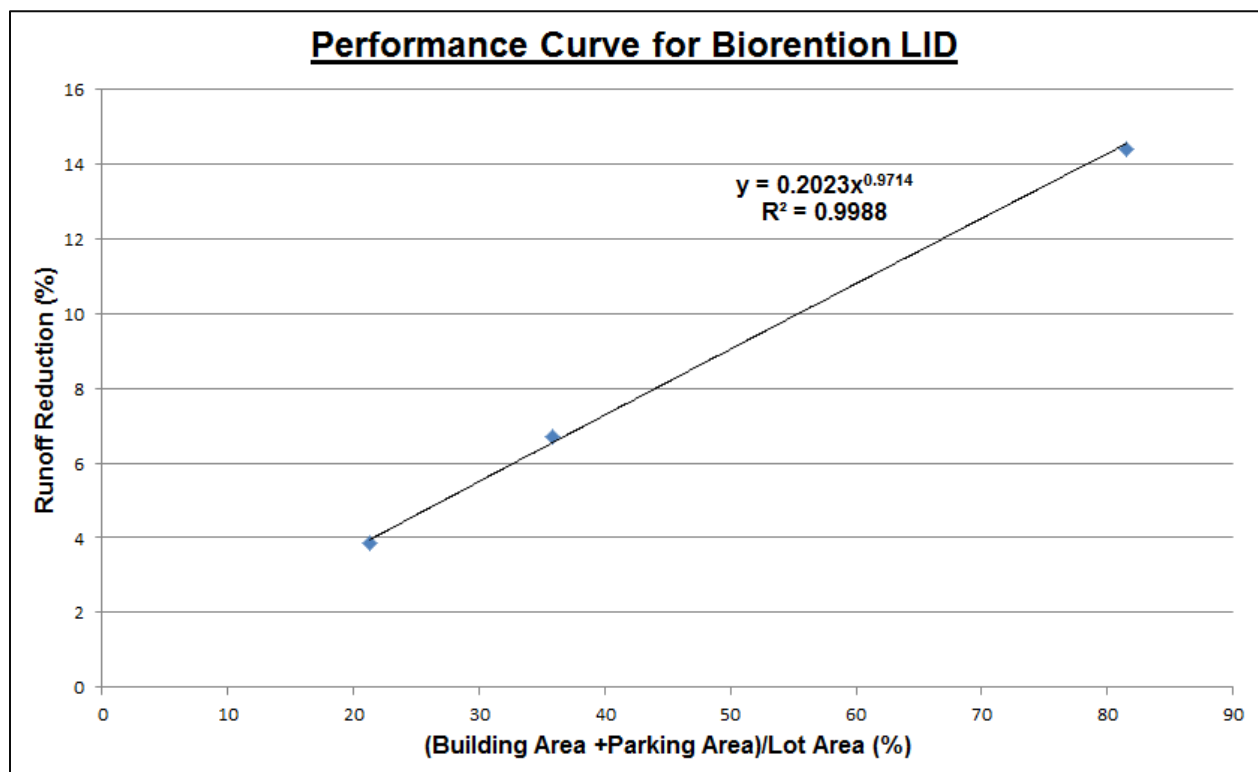
Source: Li et al. 2010

Figure 10. Performance curve for the bioretention LID in residential areas.

The graph demonstrates that as the percentage of impervious area increases, the percentage of reduction in runoff also increases which is logical since a higher proportion of the runoff will be directed to the bioretention LID (Li et al. 2010).

The equation stated on the graph in Figure 10 is the function for volumetric runoff reduction (%) of the bioretention LID. The R^2 value beside the function is very close to the value of one. The relationship provided by the equation and graph is therefore a very "good fit" to the underlying data. The function was used in the procedure to aggregate the runoff results and extrapolate the results to other municipalities based on land use. The area with respect to each land use category was summed for each stormwatershed within each municipality. The runoff reduction was then calculated for each type of LID treatment in each municipality. Thus, the identification of HRUs allowed larger areas which were composed of units with similar hydrological characteristics to be evaluated more efficiently (Li et al. 2010).

The arbitrary nature of selecting the lots to be modelled from each of the three intervals on the histogram provided the impetus for further investigation. The performance curves were generated using only a few metrics such as percentage impervious area (% ImpArea) or building size/lot size (%). The previous study therefore had limitations and suggested opportunities to become more comprehensive. The current study addresses some of these limitations while striving to refine the process of the development of UHRUs. The current study is derived from the previous study with the focus being urban residential lots in the City of Barrie as opposed to the entire Lake Simcoe Watershed which involved lots with different types of land use.

Fewer LIDs were examined in the current study than in the case study. In the current study, only the rain barrel or rainwater harvesting (RH) LID was studied. The sizing assumptions for the residential RH LID practice were based on those from the case study and are listed as follows:

- A storage layer height of 1300 mm for residential areas.

- All of the captured water is directed to a pervious surface.

- The overflows are directed to a storm sewer.

- The LID is modelled as a residential "rain barrel" LID module in SWMM-LID.

- A drain coefficient of 25.4 mm/hr.

- A drain exponent of 0.5.

- A drain offset height of 0 mm.

- A drain delay of 6 hours.

These assumptions were developed in the previous case study to select suitable lots for the RH LID. The selected lots from the case study were used as input data for the RH LID in this study. The assumptions developed were also used for modelling the RH LID using SWMM (Li et al. 2010).

Lot-based LIDs can be combined to produce combinations of LIDs on a particular lot. The following seven different lot-based LID practices were investigated in the case study:

1. soakaway pit (SP),

2. dry well (DW),

3. rainwater harvesting (rain barrel) or RH,

4. downspout disconnection (DD),

5. green roof (GR),

6. bioretention cell (BR), and

7. porous pavement (PP).

The seven LID technologies were then combined in various ways to yield seventeen combinations of LIDs. For example, the green roof, downspout disconnection, and bioretention cell technologies were combined to yield the GR + DD + BR LID combination as illustrated in Figure 11:
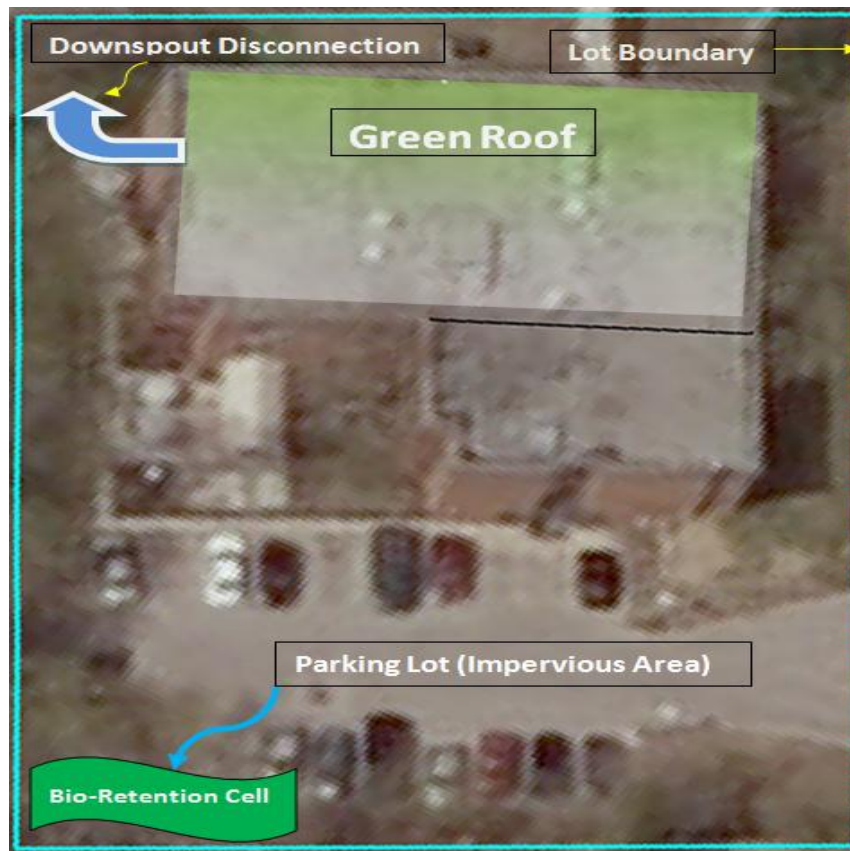
Source: Li et al. 2010

Figure 11. A combination of 3 LIDs that formed the GR + DD + BR LID combination.

Assumptions regarding the connectivity of each of the LID devices were made. For example, it was assumed that the GR LID occupied only 75 % of the roof and that the runoff from the roof drained to pervious area. It was also assumed that the BR LID treated parking and driveway runoff only (Li et al. 2010).

4. Methodology

The methodology is comprised of the following six major stages, each of which will be discussed in turn:

1. Development of a Common Database

2. Data Standardization

3. K-Means Cluster Analysis

4. Development of Cluster Diagrams

5. Running the VBA Program for SWMM

6. Calculation of Total Runoff

The overall sequence of the six major stages listed above is illustrated by the following flowchart:

Figure 12. The overall sequence of the six major stages of the methodology.
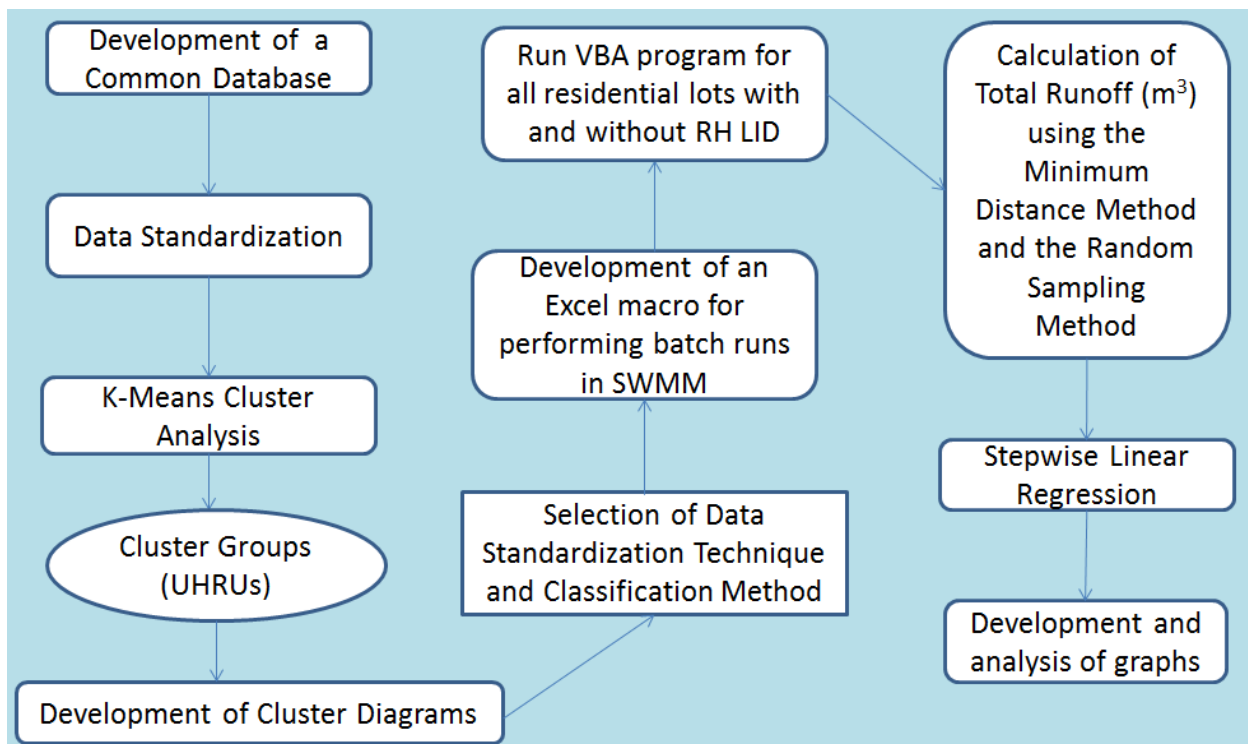
4.1 Development of a Common Database

The first major stage in the methodology was to compile a common database to store all of the data necessary for hydrological modelling. This began with familiarizing oneself with the given data in order to determine which data was most up-to-date and relevant to the potential experiment. Most of the given data for this study came in the form of GIS shapefiles. After determining which shapefiles stored the most recent and potentially useful data, a common database was developed.

Before developing the common database, however, the attributes to be used for hydrological modelling were selected. Data pertaining to the selected attributes were stored for each lot in list form. When conducting hydrological modelling, data for the group of attributes selected were used as input for each lot while all other input variables remained in a default state. The group of attributes selected was based on intrinsic characteristics of the modelling software as well as the availability of data.

For example, slope was selected as an attribute because the hydrological modelling software uses Manning's equation to compute overland flow rate which involves a slope parameter. Soil type was selected as an attribute because this parameter affects infiltration capacity. The selection of attributes was also based on experience from the case study by Li et al. (2010) and the suitability of available data. In total, the five major attributes listed below were selected for experimentation:

1. Lot area (m^2)
2. Lot width (m)
3. Imperviousness (%)

4. Average slope (%)

5. Soil type (categorical value)

The imperviousness category was further decomposed into parking area (%), driveway area (%), and building area (%) to produce a total of seven attributes.

After obtaining the GIS shapefiles with the pertinent data, the attribute tables were exported to Microsoft Excel spreadsheets. The data from the spreadsheets were then synthesized into a common spreadsheet and edited by removing extraneous or erroneous data such as duplicate records. Filtering techniques were used to trim the data in various ways, for example eliminating parcels with lot areas (m^2) less than 100 m^2. Some preliminary data analysis was then performed on the attributes under investigation in order to determine the distributions, for example by generating histograms for each attribute.

4.2 Data Standardization

The approach suggested in this experiment utilized the K-Means Cluster Analysis procedure to group lots with similar hydrological properties using IBM SPSS Statistics 20 (IBM SPSS) software. One of the assumptions of using this procedure with IBM SPSS software is that all input variables be measured on the same scale. The results may be misleading if the input variables are measured on different scales - for example, one variable is expressed in per cent and another is expressed in metres squared (m^2). The scaling of variables is therefore an important consideration (International Business Machines [IBM], 2011).

The attributes included in the input data are often associated with a dimension, for example lot width was associated with the dimension metres. At first, some of the input data was re-expressed in a different form to create dimensionless data. By converting all of the input data

to dimensionless form, the variables would have been measured on the same scale. Some types of data can be made dimensionless by simply being re-expressed in a different form. For example, variables expressed in percentage values such as 97 % impervious area can be re-expressed in decimal form as 0.97 of impervious area. However, some data such as lot area (m^2) cannot simply be re-expressed in a different form or be easily measured by another scale. Data standardization techniques must therefore be applied to convert such data into dimensionless form.

The result of data standardization techniques is dimensionless data in which all knowledge of the location and scale of the original data may be lost. In general, there are two approaches with respect to the standardization of variables: (1) global standardization and (2) within-cluster standardization (Gan et al. 2007). The first approach standardizes the variables across all elements in the data set and the second refers to the standardization that occurs within clusters on each variable. The second approach is not possible because it requires that clusters be determined before standardization. Since variables cannot be standardized directly within clusters in cluster analysis, other methods must be used (Gan et al. 2007).

The global standardization approach was therefore used to standardize all of the variables under investigation in the common database using various data standardization techniques. A data standardization technique was applied across all of the relevant variables in the data set to produce a standardized data set. The data set produced by a data standardization technique was used as input for a cluster analysis run. The output from performing a cluster analysis run was studied in order to determine the data standardization technique which should be selected.

This method was repeated for each of the four different data standardization techniques which were examined. Each technique used the following general equation:

$$x_{ij} = \frac{x^*_{ij} - L_j}{M_j} \quad (2)$$

where x_{ij} denotes the standardized value, x^*_{ij} is the original data value, L_j is the location measure, and M_j is the scale measure. The mean (x^*_{avg}), range (R^*_j), and standard deviation (σ^*_j) of the jth variable were calculated according to the conventional mathematical equations for these variables (Gan et al. 2007).

Equation 2 varied according to the location measure, L_j, and the scale measure, M_j, for each data standardization technique. The weighted uncorrected standard deviation (USTD) technique is similar to the conventional z-score transformation. The location measures and the scale measures for each data standardization technique investigated are listed in Table 2:

Table 2. Location and scale measures for the four data standardization techniques investigated.

Data Standardization Technique	Lj	Mj
USTD	0	σ^*_j
Sum	0	$\sum_{i=1}^{n} x^*_{ij}$
Range 2	min x^*_{ij} ($1 \leq i \leq n$)	R^*_j
Maximum	0	max x^*_{ij} ($1 \leq i \leq n$)

Source: adapted from Gan et al. 2007

The Range 2 data standardization technique refers to the second of two techniques which involve division by the range of the input variable. Both techniques produce identical clustering analysis results, therefore only the second method involving division by the range of the data was investigated. The new data columns with the standardized input variables were placed beside the original, unstandardized data columns in a main spreadsheet. In this experiment, there were four data columns placed alongside each of the original data columns. The variable for soil type, soil categorical value, was not standardized. It was left in its original numerical form so that it would

be treated as a categorical variable by the software. If it was standardized, each soil category would have been represented by varying decimal values which would have caused the software to treat the categorical variable for soil type as a continuous variable.

Therefore, the distinct soil categorical values would not have been maintained for the final cluster centre lots. Lots would have also been assigned to clusters with varying decimal values for the soil category, resulting in cluster analysis output that is not very clear. Since the categorical variable for soil type was left unstandardized, the remaining six variables that were standardized are listed below:

1. Percentage average slope (decimal form)

2. Lot area (m^2)

3. Percentage parking area (decimal form)

4. Percentage driveway area (decimal form)

5. Percentage building area (decimal form)

6. Lot width (m)

All of the above variables were considered as continuous or scale variables by the cluster analysis software. Histograms were plotted for each continuous variable before and after standardization to determine whether there were any changes in in the distribution. For categorical variables, such as soil type, a bar chart was created to describe the distribution.

4.3 K-Means Cluster Analysis

The cluster analyses were performed using the K-Means Cluster Analysis procedure provided by IBM SPSS software. In general, the final outcome of the K-Means Cluster Analysis procedure is relatively homogeneous groups of cases based on selected attributes developed by using an

algorithm that can handle a large number of cases (IBM, 2011). The selected attributes were the

hydrological attributes of each property which acted as the input data. In this experiment, the

input data were the six standardized continuous variables in addition to the unstandardized soil

categorical variable. If the K-Means Cluster Analysis procedure was effective at grouping

together lots with similar hydrological characteristics, the lots contained in the clusters which

were formed as a result of the cluster analysis should have been similar to one another. The

overall process of the K-Means Cluster Analysis procedure is cyclical as illustrated in the

flowchart in Figure 13:
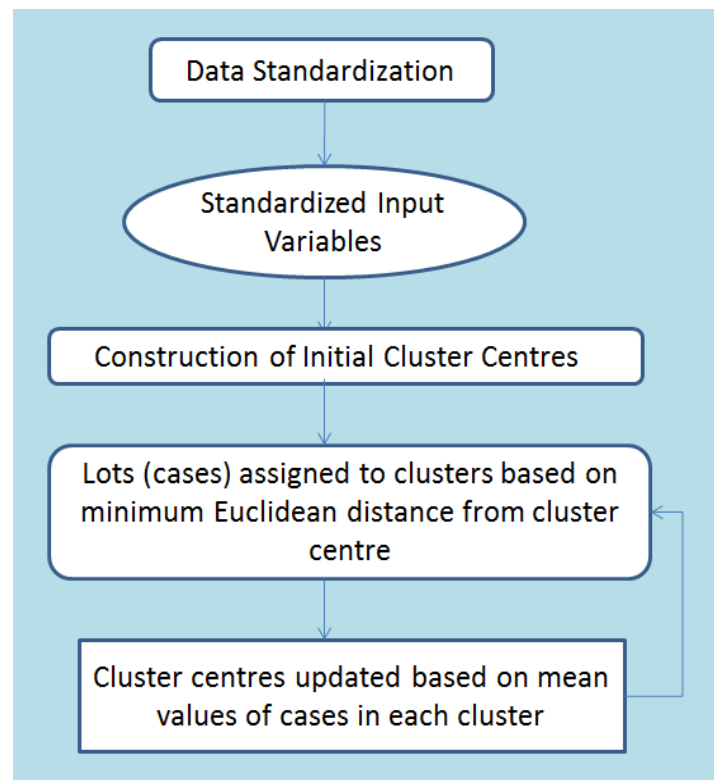
Figure 13. Overall cyclical process of the K-Means Cluster Analysis procedure.

To run the K-Means Cluster Analysis procedure using IBM SPSS software, the "K-Means

Cluster…" option was selected under the Classify menu.   After adding the input variables, any

information associated with the variables was confirmed to be correct. For example, it was

confirmed that categorical variables such as soil type were measured as nominal variables and that continuous variables such as average slope (%) were measured as scale variables. The assessment number, sometimes known as the roll number, is the unique identification number for each lot and was used to label each lot. Each lot was considered as a separate case. The number of total cases was therefore equal to the number of total lots in the common database.

The number of clusters to be created was then specified. The number of clusters has to be at least two and must not be greater than the number of cases in the common database (IBM, 2011). A group of three clusters was created in order to examine the cluster analysis output results of a small group of clusters. The number of clusters within a cluster group was then increased to examine the cluster analysis output results of larger cluster groups. The second group contained five clusters and the number of clusters within a group increased by five so as to provide a steady increase of even increments. To summarize, the number of clusters created was three, five, ten, fifteen, twenty, and twenty-five.

After specifying the number of clusters, one of two methods used for classifying cases was selected - either updating cluster centres iteratively or classifying only. Both methods were used to perform cluster analysis on each group of clusters that was created in order to determine which method produced the most accurate results. The accuracy of the results was estimated afterwards with the development of cluster diagrams.

The method entitled Iterate and Classify (ITCL) involves an iterative process for generating final cluster centres. There are a few options which are specific to the ITCL method. The first option, Maximum Iterations, refers to the maximum number of iterations that one cluster analysis will perform. Using the ITCL option, the number of maximum iterations is sometimes set to 20 as a

standard threshold. The procedure therefore stops after 20 iterations are performed regardless of what is specified for the second option, Convergence Criterion.

The Convergence Criterion determines when iterations stop by representing a proportion of the minimum distance between initial cluster centres at which point the iterations must stop. For example, if the Convergence Criterion equals 0.02 then iterations stop when a complete iteration does not move any of the cluster centres by a distance of more than 2 % of the smallest distance between any initial cluster centres. Since the convergence criterion is a proportion, it must be greater than 0 but not greater than 1 (IBM, 2011).

The Maximum Iterations option ultimately overrides the Convergence Criterion option and was selected for this clustering analysis experiment. A value for the Convergence Criterion was therefore not specified. If the iterations stopped because the maximum number of iterations was performed and the iterations failed to converge, the number of maximum iterations was increased until the iterations converged. The number of maximum iterations must be increased until convergence is achieved otherwise the solution may be unstable (IBM, 2011). The output report generated by the software program indicated whether convergence was achieved by showcasing the change in cluster centres. The change in cluster centres should be zero by the time the final iterations are reached. A final statement was also included stating whether convergence was achieved or iterations failed to converge.

Another option that is available under the ITCL method is the Use Running Means option. If this option is selected, cluster centres are updated after each lot or case is assigned to a cluster otherwise new cluster centres are calculated after all cases have been assigned. The Use Running Means option was not selected in this experiment. This is because the selection of this option makes the resulting solution potentially dependent on case order, regardless of how initial

cluster centres are chosen (IBM, 2011). Additionally, initial cluster centres were not specified beforehand although this can be done if this information is known.

The remaining options that were selected to run the cluster analyses were common to both methods of classifying cases – iteratively and classifying only. The steps for performing a cluster analysis while using the Classifying Only method are almost the same as those for the iterative method. The differences are that when using the Classifying Only method, the Classify Only option is selected and the options that are associated with the iterative process are left in the default state.

For both methods of classifying cases, there are two options available with respect to missing values: (1) Excludes Cases Listwise and (2) Excludes Cases Pairwise. The Exclude Cases Listwise option excludes cases with missing values for any clustering variable from the analysis. The Excludes Cases Pairwise option assigns cases to clusters based on distances that are computed from all clustering variables with nonmissing values. The Exclude Cases Listwise option was selected when using both methods for classifying cases. This is because each case is a distinct lot or property, therefore cases or data rows with missing input variables were not considered valuable to the analysis. Nonetheless, the entire data set was checked for blank values before submitting the data for clustering analysis and was usually free from such errors.

Other options which were selected when using either method for classifying cases pertained to the final output report of the cluster analysis. The output report displayed the statistics which were selected to be included in the report. All of the available statistical options which were selected included the initial cluster centres, an ANOVA table, and cluster information for each case. This final report contained a Cluster Membership table which listed each case along with its corresponding assessment number, the number of the cluster to which it belongs, and the

distance to the final cluster centre. After running the cluster analyses, the output reports were compared for each of the four data standardization techniques and for each of the two methods for classifying cases. The output reports were generated in the form of spreadsheets and the values within these spreadsheets were compared.

4.4 Development of Cluster Diagrams

Certain hydrological characteristics or lot properties such as lot area (m^2) and % imperviousness are usually visible using orthophotos. The clustering analyses used real spatial data as input. The output generated from the cluster analyses was therefore compared by visually determining whether lots within one cluster were similar to one another and dissimilar to lots in other clusters. Using ArcMap software, this was achieved by the development of cluster diagrams which illustrate the placement of lots within each cluster that is created. An orthophoto was then added beneath the layer of clusters and the lots within each cluster were visually inspected to see if they exhibit similar hydrological characteristics.

The first step in the visual inspection process was to gather all of the cluster output data into an Excel spreadsheet for each of the four data standardization techniques. The smallest, first four cluster groups were used to generate cluster diagrams for each type of classifying method and are listed as follows:

1. 3-cluster group
2. 5-cluster group
3. 10-cluster group
4. 15-cluster group

A separate cluster diagram was generated for each cluster group according to the method used for classifying cases. This resulted in a total of eight cluster maps for each data standardization technique.

Four Excel workbooks were developed, one for each data standardization technique. The data contained in each workbook was used as input data for developing cluster diagrams in ArcMap. An Excel worksheet or spreadsheet cannot be directly displayed as a map layer ArcMap. In order for the data in a worksheet to be displayed as a map layer, the data in the worksheet must be transformed into a shapefile or other GIS format. An attribute table, which is the underlying data table for a shapefile, was linked to the cluster output data worksheet.

A shapefile which displays all of the lots located in the City of Barrie, even those excluded from the cluster analysis, was used as the main source of GIS data. The attribute table of this shapefile was joined to the data from the worksheet containing the Cluster Membership table by using the assessment number. After the join was validated to be successful, it was accepted only if the number of matched records was equivalent to the number of records involved in the cluster analysis.

The attribute table of the existing shapefile was then modified to contain only the records that were matched. The data in the modified attribute table was exported to become a new shapefile. The different clusters in the new shapefile were separated into distinct categories by importing the previous layer containing the matched records into the new shapefile. The properties of the new shapefile were selected and the symbology definition was imported from another layer in the map or from a layer file. The clusters in the cluster diagram were distinguished by altering the colour of each lot boundary that belonged to a particular cluster while the interior of the lot was left transparent in order to see the underlying orthophoto. A consistent colour scheme was

used for each cluster diagram which allowed the lot to be seen after adding the orthophoto from Bing Maps, which is a feature that is embedded in ArcMap.

4.4.1 Selection of Data Standardization Technique and Classification Method

The colour-coded cluster diagrams were overlaid on a current orthophoto or aerial map of the City of Barrie which was provided by Bing within ArcMap. The lots within each cluster were visually inspected for similarities and dissimilarities with respect to visible hydrological attributes such as imperviousness (%). The visual inspection process was carried out by zooming in and out using the Zoom tools.

Some hydrological attributes are more visible than others. For example, the sizes of the lots between clusters can be easily compared since the coloured outline of each lot highlights this attribute, however other attributes such as soil type are not visible. The similarity or dissimilarity between lots with respect to hydrological attributes which are not visible was confirmed by adding a shapefile layer on top of the cluster diagram. For example, a soils layer was added on top of the cluster diagrams to see if lots within the same cluster contained the same soil type.

The similarity or dissimilarity of lots within clusters and of lots between clusters was determined by visual analysis. An assessment regarding the ability of the cluster diagrams to reflect reality was then made. Since the input data for the cluster analyses varied according to the data standardization technique used, the cluster diagrams generated from different data standardization techniques were compared. The standardized data set which produced a cluster diagram that most closely resembled reality, as it appeared in the orthophoto, was selected to be used in subsequent analyses. While examining the various cluster diagrams, observations were

made as to which method of classifying cases appeared to produce the most accurate cluster diagram and this method of classification was selected to be used consistently afterwards.

4.5 Running the VBA Program for SWMM

4.5.1 Batch SWMM Runs Using Excel Macros

A Microsoft Excel macro was developed by using the Visual Basic for Applications (VBA) programming language. The objective of developing the macro or VBA program was to run batches of input data using SWMM software. The data from the spreadsheet within the macro was fed into the SWMM software which performed a hydrological modelling simulation or run and generated SWMM output. The VBA code therefore linked the SWMM input data to the software, activated the software to perform a hydrological modelling run, and directed the placement of the output files to the hard drive of the computer for storage and analysis.

Output files from the cluster analyses were exported in the form of a spreadsheet which listed each lot along with corresponding information such as the cluster number and the assessment number. The assessment number for each lot is the unique lot identifier which was used to trace the lots from the cluster analysis output to the common database. The data required for modelling each lot could then be extracted into a spreadsheet within the VBA program workbook. The number of lots within each cluster that was created was used to estimate the accuracy of the cluster analysis output. For example if a cluster within a 3-cluster group only contained one lot, the result was deemed to be suspicious since it would be difficult to find this kind of distinctive lot in reality. Suspicious clustering output was confirmed using the orthophoto. After being checked against the orthophoto, final clusters which were relatively well-balanced in terms of the number of lots within each cluster were accepted.

Preliminary exercises using the VBA program highlighted weaknesses in the VBA code which were edited to make the program run more efficiently. For example, the data required for modelling all 7871 residential lots was placed into a spreadsheet of the workbook containing the macro. The macro was run and a SWMM output file was generated which listed the output values for each lot. Each lot was treated as a separate subcatchment by SWMM software. A comparison of the SWMM output generated by some of the preliminary investigations indicated that the results were not sufficiently accurate. For instance, the infiltration (mm) was zero for several lots which indicated that further editing of the VBA code was necessary. The overall procedure for developing the VBA code was cyclical in nature and is illustrated in Figure 14:
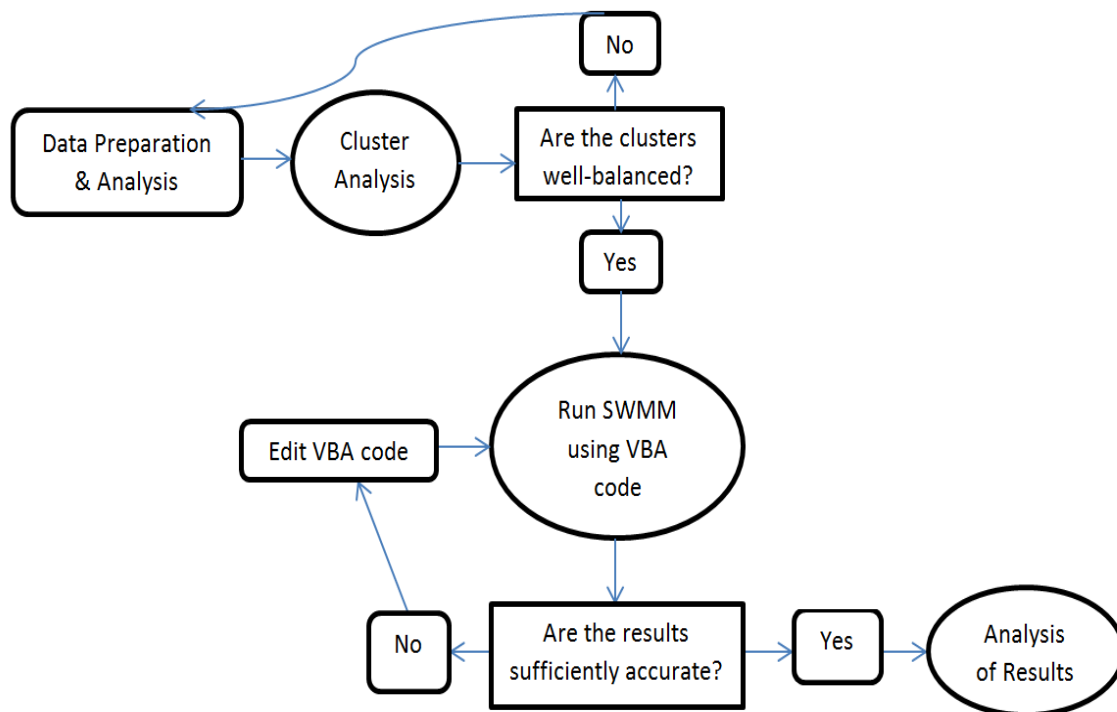
Figure 14. Flowchart of the overall procedure for developing the VBA code.

4.5.2 Running the VBA Program for All Residential Lots without LIDs

The input data used to run the VBA Program for SWMM is required to be in unstandardized form. The input data was extracted from the common database into the Excel spreadsheet of the macro using the VLOOKUP function. The VBA program was first created for lots that had not yet implemented a LID. The columns of input data corresponded to the input data required to run SWMM without LID implementation. The following list contains the data columns which were required for the scenario of no LID implementation:

1. Assessment Number

2. Lot Area (ha)

3. Imperviousness (%)

4. Parcel Width (m)

5. Average Slope (%)

6. Initial Deficit

7. Hydraulic Conductivity (mm/hr.)

8. Suction Head (mm)

The above data for all residential lots was placed into the spreadsheet of the macro. At the end of a run, three files were generated: (1) an input file, (2) an output file, and (3) a report file. The report file was imported into an Excel spreadsheet to calculate the total runoff (m^3) from all lots. The Total Runoff (mm) from each lot was listed in the report file and was first summed for all lots after being imported into the Excel spreadsheet. The Lot Area (m^2) was then added to the spreadsheet using the VLOOKUP function or a Merge Tables Wizard in order to calculate the Total Runoff (m^3) for each lot which was then summed. The VBA macro essentially creates a

SWMM input file to connect the input data with SWMM to perform batch runs. The main page of code can be seen in Appendix A.1 for the case of no LID implementation.

4.5.3 Running the VBA Program for All Residential Lots with Rain Barrel LIDs

The code for the VBA program was modified to accommodate the implementation of the rain barrel LID on lot properties. A "For" loop was added beneath the LID usage section in order to activate the implementation of the LID. Additional data pertaining to the LID was also placed in the spreadsheet. The modified section of the VBA code can be seen in Appendix A.2. The additional data columns required for the implementation of the RH LID are:

1. LID Control (type of LID)

2. Area of Each LID Unit (m^2)

3. Width

4. Initial Saturation

5. Per cent of Impervious Area Treated

6. To Pervious

The width parameter refers to the width of the outflow face of each identical LID unit (in ft. or m) and does not apply to the RH LID because it does not use overland flow to convey surface runoff off of the unit. The initial saturation column is applicable to LIDs with a storage zone because it represents the degree to which the storage zone is initially filled with water (Rossman 2010). The RH LID does not have a storage zone so both the width and initial saturation columns were input as zero values for all lots. The per cent of impervious area treated is defined as, "the per cent of the impervious portion of the subcatchment's non-LID area whose runoff is treated by the LID practice" (Rossman 2010, p.217). In this experiment the per cent of

impervious area treated was equal to the percentage of area occupied by the roof which was the variable building area (%). The "to pervious" column was selected because the outflow from the LID was returned onto the subcatchment's pervious area instead of going to the subcatchment's outlet. A value of one was listed for each lot in this column so that it was selected for each hydrological modelling run.

Approximately 950 of the 7871 residential lots were not amenable to the implementation of the rain barrel LID. The remaining lots which were amenable to the implementation of the RH LID were either extracted or merged from the common database. The assessment numbers for these lots were copied into the spreadsheet for the VBA program and the relevant data columns were either calculated or filled. The VBA Program was run and the output files were checked to ensure that code was free from error.

4.5 Calculation of Total Runoff (m^3)

4.5.1 The Minimum Distance Method

The Cluster Membership table illustrated in the final report produced by a cluster analysis listed the minimum distance of each lot to the final cluster centre. The lot with the minimum distance to the final cluster centre was used as an approximation of the final cluster centre. If the minimum distance to the final cluster centre was zero, then the lot itself was the final cluster centre and was regarded as the embodiment of cluster properties.

The Cluster Membership table was sorted in ascending order by the number of the cluster. The list of data was segmented according to cluster number and each cluster was sorted in ascending

order again based on the distance of the lot with respect to the final cluster centre. The lot with the minimum distance to the final cluster centre appeared at the top of the list and was highlighted. The highlighted lots with the minimum distance to the final cluster centre were grouped together and relevant data was extracted for each lot.

The Total Runoff (mm) produced by the lot with the minimum distance to the final cluster centre was extrapolated to the rest of the lots within the cluster. First, the lot area (m^2) was summed to yield the total lot area (m^2) for each cluster. The total lot area (m^2) for each cluster was then weighted by the total runoff (mm) of the lot with the minimum distance to the final cluster centre, after total runoff (mm) was converted to total runoff (m). The total runoff (m^3) of each cluster within the cluster group was then added to yield the overall total runoff (m^3) for the entire cluster group. This method of calculating total runoff (m^3) for each cluster group was applied to both the non-LID scenario and the LID scenario. Lastly, the total runoff (m^3) for each cluster group was plotted against the number of clusters in the cluster group. The procedure for the Minimum Distance Method is illustrated in Figure 15:

1) Minimum Distance Method:

Sum lot area (m²) of entire cluster

Extrapolate runoff (mm) of final cluster centre to rest of cluster using total lot area (m²) of cluster
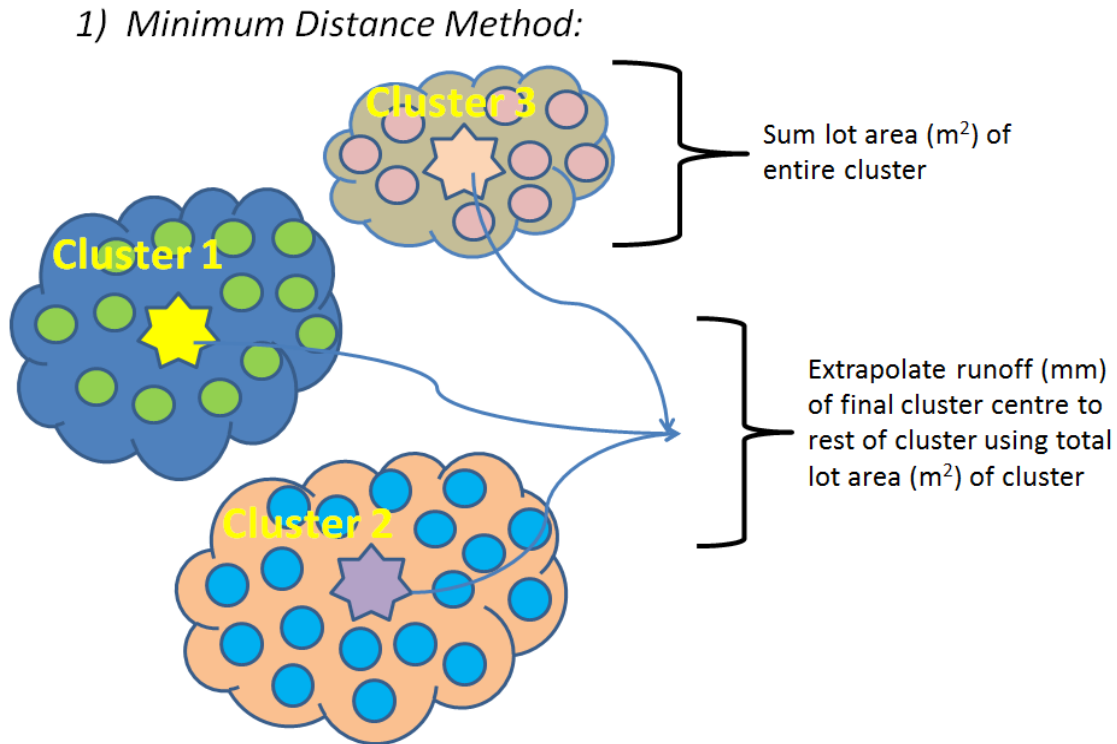
Figure 15. Procedure for calculating total runoff (m³) using the Minimum Distance Method.

4.5.2 The Random Sampling Method

Another method which was used to calculate the overall total runoff (m³) involved taking a random sample of lots from each cluster. An Excel tool known as Random Sorter was used to generate a certain percentage of random lots from each cluster within a cluster group. The first trial, for example, entailed randomly selecting 5 % of the lots from each cluster within the 3-cluster group. Tables were created which contained the input data for the random samples that were required to perform linear regression in IBM SPSS. Each lot in these tables was accompanied by the following unstandardized data:

1. Total Runoff (mm)

2. Soil Categorical Value

3. Average Slope (%)

4. Lot Area (m^2)

5. Parcel Width (m)

6. Parking Area (%)

7. Driveway Area (%)

8. Building Area (%)

The Total Runoff (mm) was the dependent variable whereas the rest of the variables were the independent variables. When linear regression was performed for the no LID case, the data for Total Runoff (mm) with no LID were used. When performing linear regression for the case with the RH LID, the Total Runoff with LID (mm) data column was used. Using IBM SPSS software, the Linear Regression option was selected in conjunction with the Stepwise method.

Since the Stepwise Linear Regression method was used, the output report listed the coefficients for each proposed model of the linear regression equation. The Model Summary table listed the R value, the R Square value (R^2), and the Adjusted R (Adjusted R^2) square for each proposed model. The model with an R Square value closest to 1 or the highest R Square value among all of the models was generally selected to be used to calculate the runoff (mm) for all lots within a cluster because this suggested a strong relationship between the dependent variable and the independent variables. However, other output values were also considered before selecting the most suitable linear regression equation.

For example, the Coefficients table contained three columns which were listed under the Correlations section. If the partial and part correlations dropped sharply in absolute value from the zero-order correlation for an input variable, this was an indication that there may have been a problem with respect to multicollinearity (IBM, 2011). It also indicated that much of the

51

variance for the dependent variable that was explained by that particular input variable was also explained by other input variables. As a result the input variable may have been redundant, therefore a model without that particular input variable could be selected.

There was also a Collinearity Statistics section of the Coefficients table. In this section, the term Tolerance is defined as, "the percentage of the variance in a given independent variable that cannot be explained by the other independent variables" (IBM, 2011). Small variances or coefficients of determination, which are considered to be around 0.1 to 0.3, demonstrated that approximately 70 % to 90 % of the variance of a given independent variable can be explained by other independent variables. Therefore, if the tolerances are close to zero there is high multicollinearity and the standard error of the regression coefficients will be inflated. The column beside the Tolerance column was named VIF which stands for Variance Inflation Factor. A VIF greater than 2 is usually considered problematic (IBM, 2011).

In the Collinearity Diagnostics table, there was an Eigenvalue column. Eigenvalues that are close to zero are an indication that the independent variables are highly intercorrelated and that small changes in the data values may lead to large changes in the estimates of the coefficients. The Condition Index column, next to the Eigenvalue column, listed condition indices which are calculated as the square root of the ratio of the largest eigenvalue to each successive eigenvalue. If the condition indices are greater than 15, this indicates a possible problem with collinearity and if they are greater than 30 then this suggests a serious problem with collinearity. Collinearity problems can potentially be fixed by rerunning the regression using standardized independent variables (IBM, 2011).

The indicators presented in the output reports suggested that some of the models proposed by the stepwise linear regression method may have had a problem with collinearity while others were

free from such a problem. The model that exhibited the fewest problems with collinearity and that was associated with the highest R^2 value or with an Adjusted R^2 value that was closest to 1, was selected as the best-performing regression model for a cluster. The coefficients from this model were used in the regression equation to produce the best-performing regression equation.

Using the best-performing regression equation, the runoff (mm) was calculated for each lot within a cluster as the dependent variable or the y-value. The Total Runoff (mm) for the entire cluster was then calculated, followed by the Total Runoff (m^3) produced by each cluster within the cluster group. The Total Runoff (m^3) for each cluster was then summed to yield the overall Total Runoff (m^3) for the cluster group. Lastly, a plot of the Total Runoff (m^3) produced by each cluster group was developed. A general overview of the Random Sampling Method is given in Figure 16:
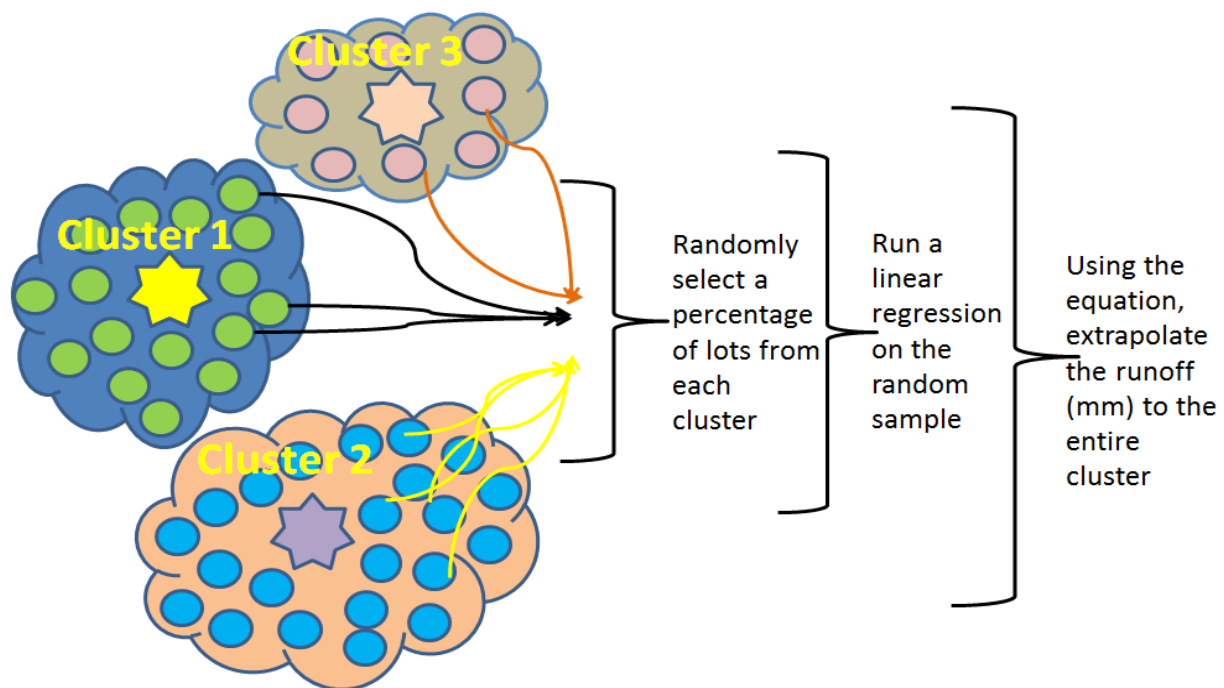
Figure 16. Procedure for calculating total runoff (m^3) using the Random Sampling Method.

4.5.3 The Generation of Random Samples from Both Data Sets

In the previous section, random samples were taken from clusters which were developed by using the K-Means Cluster Analysis procedure. The development of clusters is a method used to construct UHRUs since each cluster is composed of lots with similar hydrological conditions. In order to determine whether the development of clusters or UHRUs was necessary when applying the Random Sampling Method, random samples were drawn from the entire data sets of both the non-LID case and the LID case. The Random Sorter tool which was added to Excel was used to randomly extract percentages of lots from both data sets. The data set for the non-LID case contained 7,871 residential lots without any LID implementation while the data set for the LID case contained 6,926 lots with RH LID implementation. The runoff per lot (m^3) which was derived from the SWMM output of each data set was used to add the runoff (m^3) from each lot in the random sample to calculate the total runoff (m^3) of the entire sample. The results were then plotted on a graph.

5. Results and Discussion

5.1 Preliminary Data Exploration

Data for the common database was synthesized by first reviewing and scanning all of the given data for the City of Barrie. As previously mentioned, this data was derived from the case study by Li et al. (2010) which formed the basis for this research. The common database was first developed for all lots within the City of Barrie, regardless of land use category. For example, it included residential lots, commercial lots, and lots dedicated to open areas. Since approximately 87 % of the lots in the initial database belonged to the residential land use category, the focus of the current research shifted to residential lots within the City of Barrie. All types of residential properties were a part of the common database such as single-family homes, town homes, and apartment buildings.

The availability of data was critical with respect to the selection of hydrological conditions or characteristics to be studied. Most of the data for the City of Barrie was available in the form of GIS shapefiles. Once a certain lot-based hydrological characteristic was selected to be studied, the attribute table for that characteristic was exported to an Excel spreadsheet. Data for each hydrological characteristic from different Excel spreadsheets were merged to form the common database. The composition of the common database is illustrated in Figure 17:
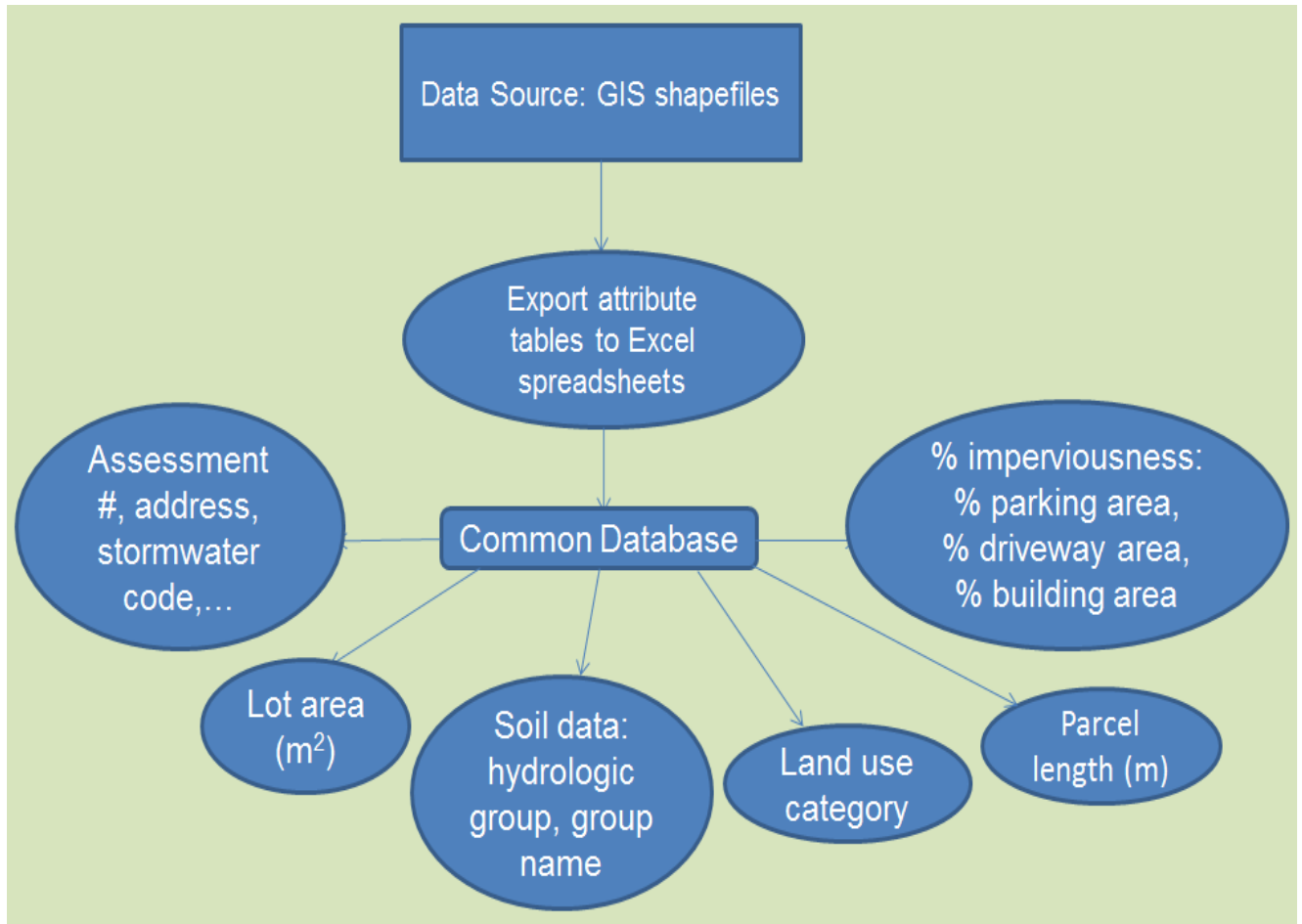
Figure 17. Composition of the common database.

After the selection of the attributes to be investigated and the development of the common database, preliminary data exploration revealed the structure of the distributions of the attributes under investigation. Histograms were created for each of the six continuous variables, both before and after data standardization. An example of the distribution of the percentage building area continuous variable (decimal form) before data standardization is shown in Figure 18.
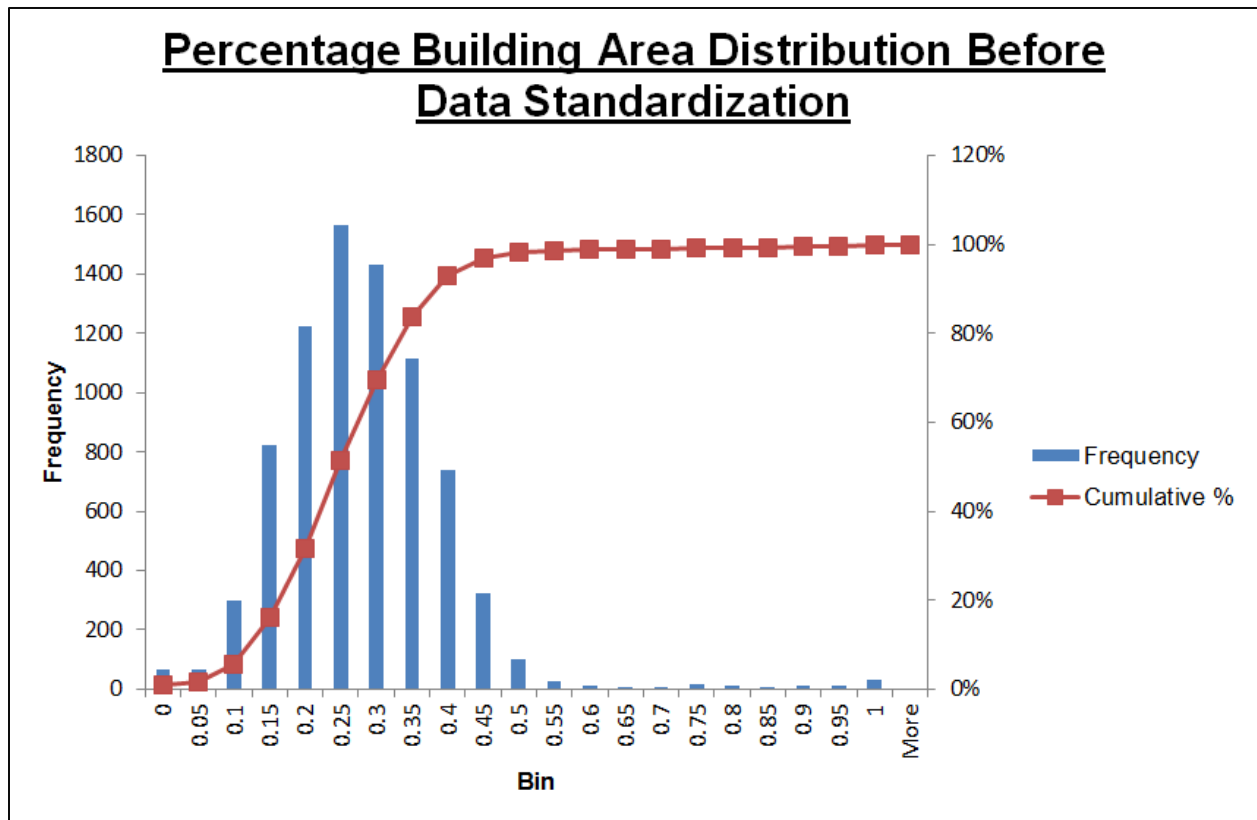
Figure 18. Distribution of the percentage building area data before data standardization.

Figure 18 illustrates how the frequency of lots with certain percentages of building areas exhibits a bell-shaped or normal distribution that is slightly skewed to the right. The positively skewed distribution remains the same after applying the Range 2 data standardization technique as shown in Figure 19. A similar example of the histograms generated before and after the four data standardization techniques were applied is illustrated in Appendix A.3 for the percentage average slope (decimal form) continuous variable. The shape of the distribution before data standardization remained unchanged after each technique was applied, although data standardization techniques can potentially normalize non-normal data distributions.
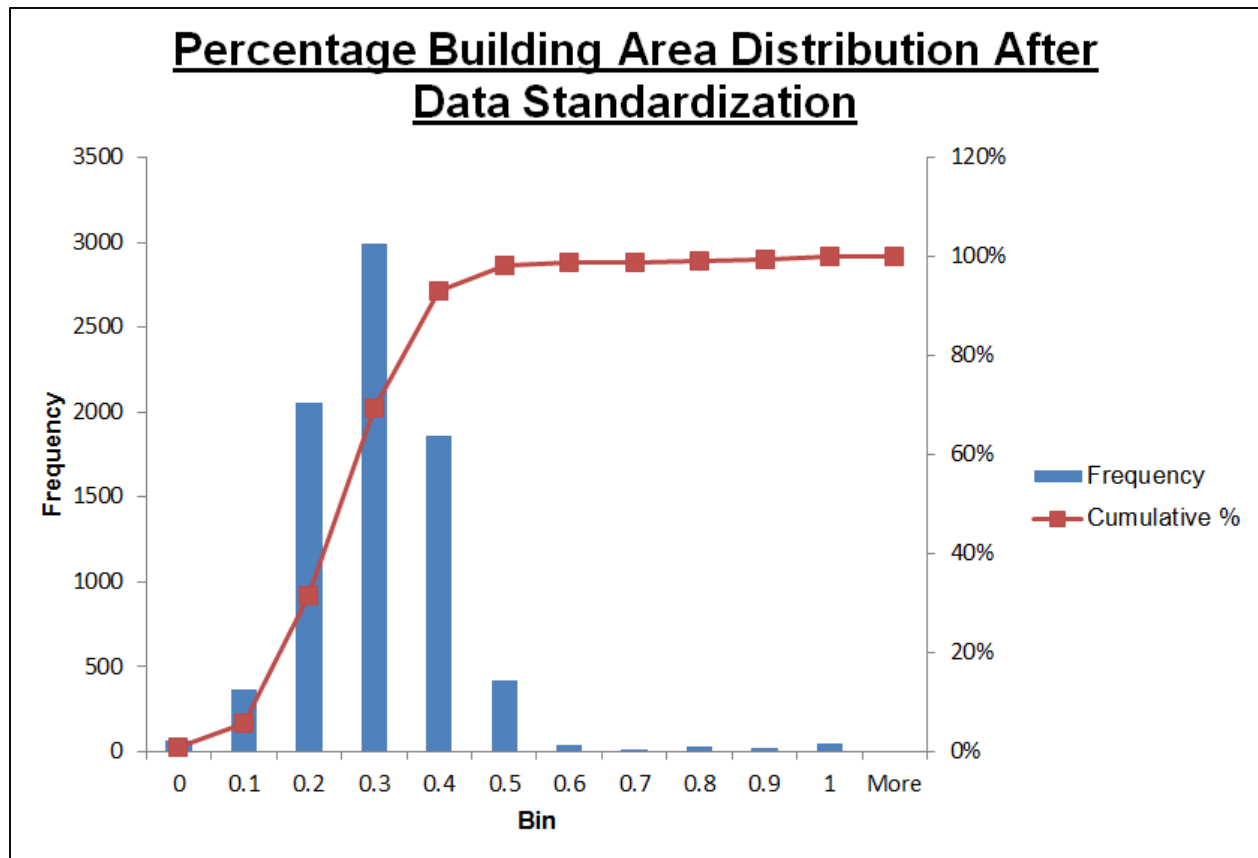
Figure 19. Distribution of the percentage building area data after the Range 2 data standardization technique was applied.

When comparing Figures 18 and 19, it is clear that the shape of the distribution is maintained even though the bin intervals change. The width of the bin interval in Figure 18 is 0.05 whereas the width of the bin interval in Figure 19 is 0.1. The width of the bin interval was adjusted after each data standardization technique was applied in order to illustrate the shape of the distribution more clearly.

In a previous preliminary analysis of the raw data, histograms with varying bin widths were plotted for each variable. The range of histograms created demonstrated that the distribution for each variable is unchanged by modifying the bin width. However, some distributions are more clearly visible using certain bin widths. The bin widths used to create the histograms for the

various data standardization techniques were modified until a histogram with a clearly visible distribution was created.

The shape of the histogram did not change after applying any of the data standardization techniques for all of the continuous variables, regardless of the technique used. This demonstrated that applying data standardization techniques did not modify or normalize the distribution of the data used in this experiment. Only some of the histograms illustrated distributions which approximated the bell-shaped or normal distribution. The majority of the distributions, however, were either negatively skewed or positively skewed. The rest of the distributions were asymmetric and non-normal.

A common data standardization technique which was first used entailed calculating the z-score which is a conventional and popular data standardization technique. This technique was ineffective for the given data set since many of the standardized values were negative which was unacceptable for certain variables. For example, negative lot areas (m^2) were unacceptable because such a concept is unrealistic. Therefore, even if the z-score technique could have been effective at scaling all of the input variables the results were not usable.

In addition, z-scores are based on normal theory and are appropriate for quantitative variables with normal distributions. Variables with skewed distributions should be avoided when the standardization technique is the z-score because the distribution of z scores has the same shape as that of the original data. If the distribution of the standardized data is the same as that of the original data, then problems with the data such as non-normality or skewness cannot be resolved by the data standardization technique that was applied (IBM, 2011).

Since the distributions resulting from the data standardization techniques investigated in this experiment remained the same after standardization, the techniques did not normalize the data or correct for skewness. However, normal or symmetric distributions are not a requirement for the K-Means Cluster Analysis procedure in IBM SPSS because the scaling of the variables is the main requirement. The scaling of the variables was accomplished by using the Range 2 data standardization technique (IBM, 2011). There was also no requirement for independence between input variables which is why related variables such as lot area (m^2) and lot width (m) could be included together in the analysis.

5.2 K-Means Cluster Analysis Output

The output reports generated by the K-Means Cluster Analysis procedure list the output from each cluster run in the form of tables. One of the first tables shown in the output for the Classify Only method, which was the classification method that was selected, is the Initial Cluster Centres table. The Initial Cluster Centres table lists the first cluster centre or "centroid" values for each input variable that the algorithm determines for each cluster. For example, the Initial Cluster Centres table for the 5-cluster group using the Classify Only method is shown in Table 3:

Table 3. Table of Initial Cluster Centres for the 5-cluster group using the Classify Only method.

Initial Cluster Centres					
	Cluster				
	1	2	3	4	5
Soil Categorical Value	1	3	1	2	2
%Avg_Slope - RANGE2	.10	.11	.11	.48	.95
Lot_Area (m^2) - RANGE2	.44	.41	.03	.84	.01
% Parking_Area - RANGE2	.83	.51	0.00	0.00	0.00
% Driveway_Area - RANGE2	0.00	0.00	0.00	.05	.17
% Building_Area - RANGE2	.08	.08	.10	.04	.36
Parcel Width (m) - RANGE2	.70	.68	.19	.97	.09

*The input variable names reflect the titles of the standardized data columns in the input data file.

The construction of initial cluster centres is the first step in the K-Means Cluster Analysis procedure that is automatically carried out by the software. The initial cluster centres are the first representative values for each input variable for each of the k clusters selected which should contain well-spaced observations. After constructing the initial cluster centres, the procedure assigns cases to clusters based on the minimum Euclidean distance from the cluster centres. The algorithm then updates the locations of the cluster centres based on the mean value of cases in each cluster. This sequence of steps is repeated until the point when another re-distribution of cases would make the clusters more internally variable or externally similar (IBM, 2011).

The next table contained in the output report for the Classify Only method is the Cluster Membership table. The Cluster Membership table is followed by the Final Cluster Centres table which is similar to the Initial Cluster Centres table except that it lists the values for the input variables of the final cluster centres that were constructed by the clustering procedure. The final cluster centres are computed as the mean for each variable within each final cluster and reflect the characteristics of the typical case for each cluster (IBM, 2011). The values for each variable within each final cluster may be hypothetical and may not be actual data values that belong to a certain lot.

The table named Distances between Final Cluster Centres is a matrix which shows the Euclidean distances between the final cluster centres. The greater the distance between two points, the greater the dissimilarities between the final cluster centres. Table 4 demonstrates that clusters 2 and 1 are the most dissimilar clusters within the 5-cluster group:

Table 4. A matrix showcasing the dissimilarities between the final cluster centres.

Distances between Final Cluster Centres					
Cluster	1	2	3	4	5
1		2.03	.55	1.13	1.13
2	2.03		1.98	1.15	.99
3	.55	1.98		1.18	.99
4	1.13	1.15	1.18		.64
5	1.13	.99	.99	.64	

An ANOVA table is also shown in the output report to indicate which variables contribute most to the solution according to the F values because variables with large F values provide the greatest separation between clusters.

However, caution is advised in the report to interpret the F values (F tests) in the ANOVA table for descriptive purposes only because the clusters have been chosen to maximize the differences between cases in different clusters. The observed significance levels, located beside the F values, are not corrected for this and cannot be interpreted as tests of the hypothesis that the cluster means are equal (IBM, 2011). It was observed that no single variable consistently provided the greatest separation between clusters. In general, the largest F values from the cluster analysis output reports for all of the residential lots varied between the soil categorical value, the lot area (m^2), and the parking area (%).

The final table in the output report is the most important for the purposes of this research. This table lists the number of cases in each cluster and whether all of the cases were accounted for in a cluster analysis. If some cases were listed as missing, the cluster analysis was re-run. An example of this table for the 5-cluster group for all residential lots is shown as Table 5:

Table 5. Table showing the number of cases in each cluster for the 5-cluster group.

Number of Cases in each Cluster		
Cluster	1	242
	2	497
	3	4048
	4	51
	5	3033
Valid		7871
Missing		0

The orthophoto for the study area was scanned and lots with extreme values for lot area (m^2) were eliminated. Most of the study area was composed of relatively homogeneous residential areas with a few more heterogeneous residential areas located near or in the downtown core. It seemed highly unlikely that there were very distinct or unusual lots which merited their very own cluster. Therefore, the number of cases in each cluster should have been fairly well-balanced in order to reflect reality more accurately.

For example, if one or more clusters contained only a single lot or a few lots then the results would have been considered questionable because it was unlikely that there was such a distinctive kind of lot in reality. If such a cluster appeared, the questionable lot was traced to its position on the orthophoto to validate its distinctiveness. It was often found that there was no apparent justification for the lot or for the group of few lots to be in a separate cluster.

Hence, the final table in each cluster output was used as an approximate measure regarding the effectiveness and accuracy of the overall cluster analysis. The more evenly balanced the final number of cases in each cluster, the more likely it was that the final clusters created were representative of reality. This gave an indication as to which data standardization technique was the most suitable for this experiment. It also indicated whether the number of final clusters

created and the method used to classify cases, either iteratively or just by classifying, was the most representative of reality. The exploratory nature of cluster analysis makes it difficult to determine the "best" number of clusters until all of the analyses have been run and all of the solutions have been examined (IBM, 2011). The subsequent translation of the cluster analyses output into cluster diagrams helped to draw conclusions with respect to which data standardization technique, number of clusters to be formed, and classifying method was most effective.

5.3 Visual Analysis of Cluster Diagrams

The cluster diagrams which were developed were added in ArcMap on top of the orthophoto provided by Bing Maps to visually determine which data standardization technique and method of classifying cases was the most reflective of reality. The diagrams for the smaller cluster groups were used in the beginning of the analysis to observe any noticeable differences as a result of the data standardization techniques.

For example, the cluster diagrams for the 3-cluster group visibly segmented the clusters into three groups according to soil type. This result was consistent for all data standardization techniques while using the Classify Only method. The only exception was the result from applying the USTD technique because one soil group was seen to have a dominating presence on the cluster diagram. A shapefile which illustrated the three major soil types in the City of Barrie was compared with the cluster diagrams. A screen shot of the shapefile showing the location of the three soil types in the City of Barrie is shown in Figure 20. The uncoloured areas are areas in the City of Barrie for which data was not available.
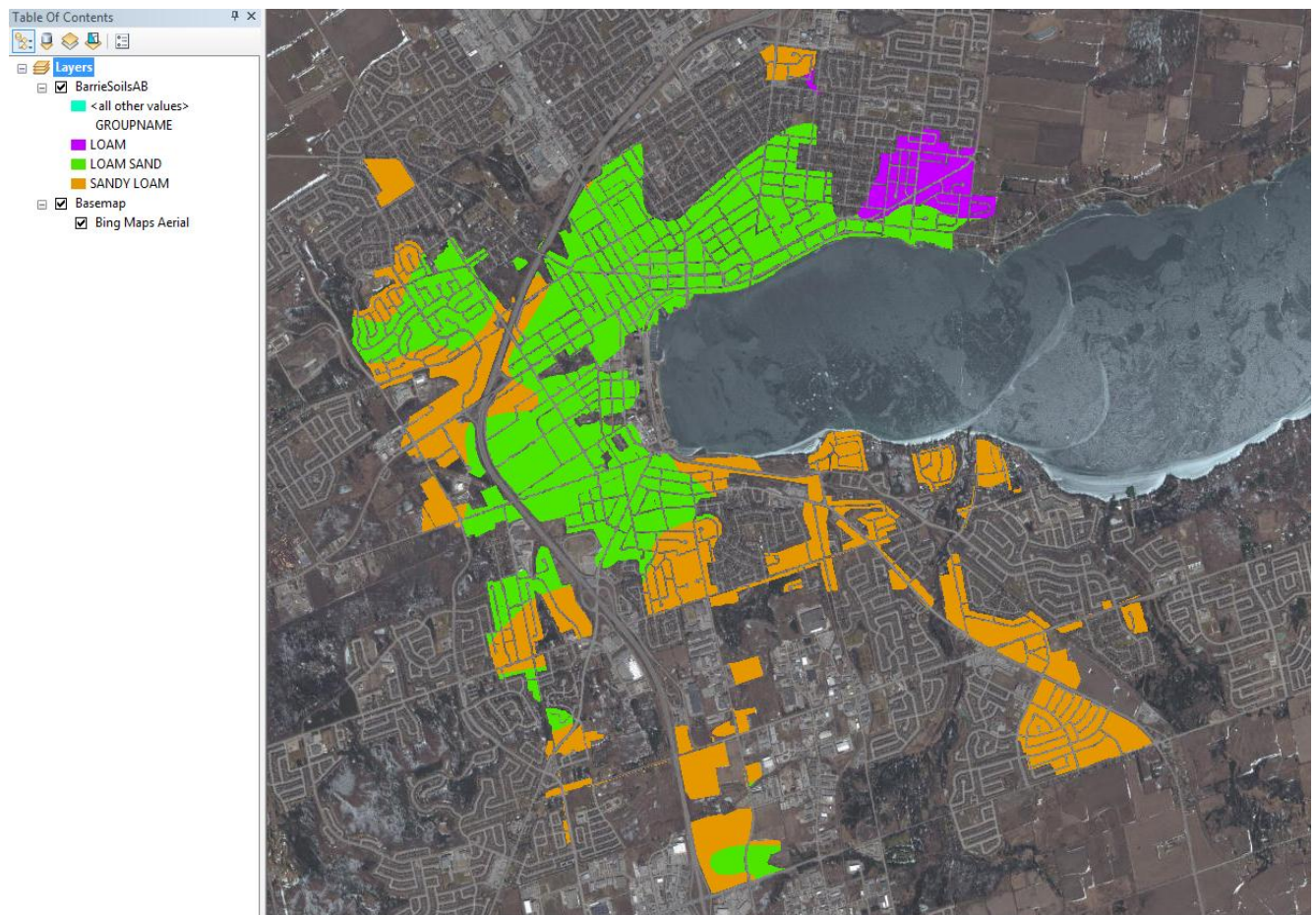
Figure 20. Shapefile of available data for the three different soil types in the City of Barrie.

A legend explaining the type of soil for each colour is shown on the left-hand side in the Table of

Contents in Figure 20. The three different soils are situated within relatively distinct locations

within the City of Barrie with almost no overlap, although it was discovered that some lots

contained more than one soil type in the initial data. This problem was addressed in the GIS

pre-processing portion of the case study. The distinction between the locations of the soil groups

was maintained by the SUM, the Maximum, and the Range 2 data standardization techniques.

For example, the cluster diagram for the 3-cluster group using the Range 2 technique and the

Classifying Only method is shown in Figure 21.

Figure 21. Cluster diagram for the 3-cluster group using the Range 2 technique and the Classifying Only method.

When comparing Figure 21 and Figure 20, cluster 3 (light blue lot boundaries in Figure 21) seems to be mainly composed of lots with loamy sand (green lots in Figure 20). Cluster 2 (orange lot boundaries in Figure 21) is mainly composed of lots with loam soil (purple lots in Figure 20). Cluster 1 (purple lot boundaries in Figure 21) is mainly composed of sandy loam soil (orange lots in Figure 20). The three clusters within the cluster diagram for the 3-cluster group clearly correspond with the three types of soil in Barrie. Another example is the 10-cluster group which was also developed using the Range 2 technique and the Classifying Only method.

Figure 22 takes a close-up look at how soil type contributed to distinguishing the lots of the

10-cluster group.

Figure 22. A close-up look at some clusters from the 10-cluster group developed by the Range 2
technique and the Classifying Only method.

The curvy, diagonal yellow line across Figure 22 divides the area into two regions with different

types of soil. The lots to the left and above the yellow line contain sandy loam soil whereas the

lots to the right and below the yellow line contain loamy sand soil. The colours of the lot

boundaries are different on both sides of the yellow line which indicates that the clusters were at

least partly distinguished by soil type. Therefore, soil type contributed to the creation of clusters even in a larger group of clusters such as the 10-cluster group. There is more than one cluster on each side of the yellow "divider" line. In the region with sandy loam, for example, the relatively large property outlined in light blue was placed into a separate cluster although it contained the same type of soil as the neighbouring single residential lots outlined in green. This property was placed into a separate cluster because the area of the lot and the level of imperviousness were different from the single-family homes as a result of being a multi-unit dwelling.

This distinction between the three soil groups, however, was not maintained when the USTD data standardization technique was applied using the Classifying Only method as seen in Figure 23.
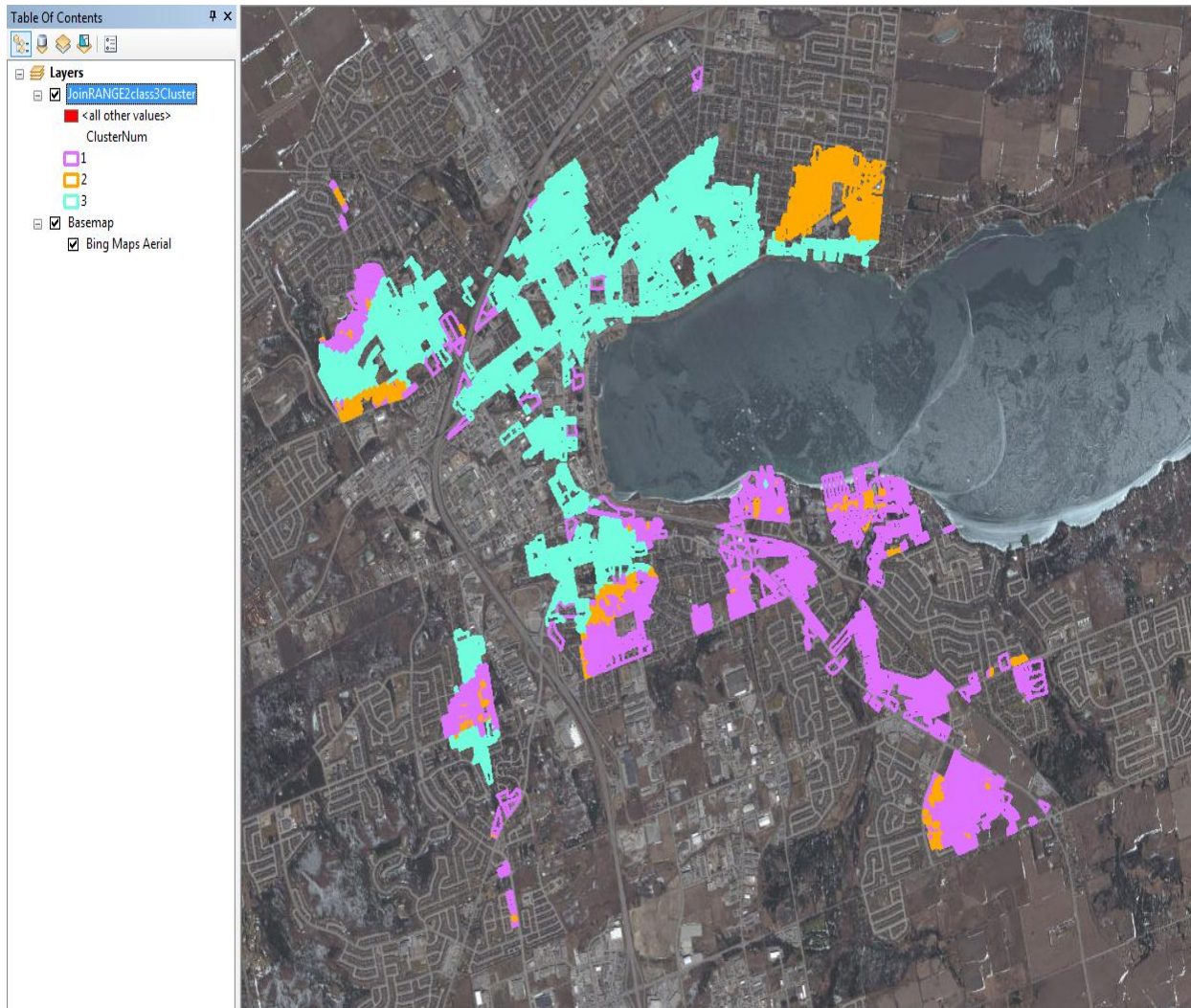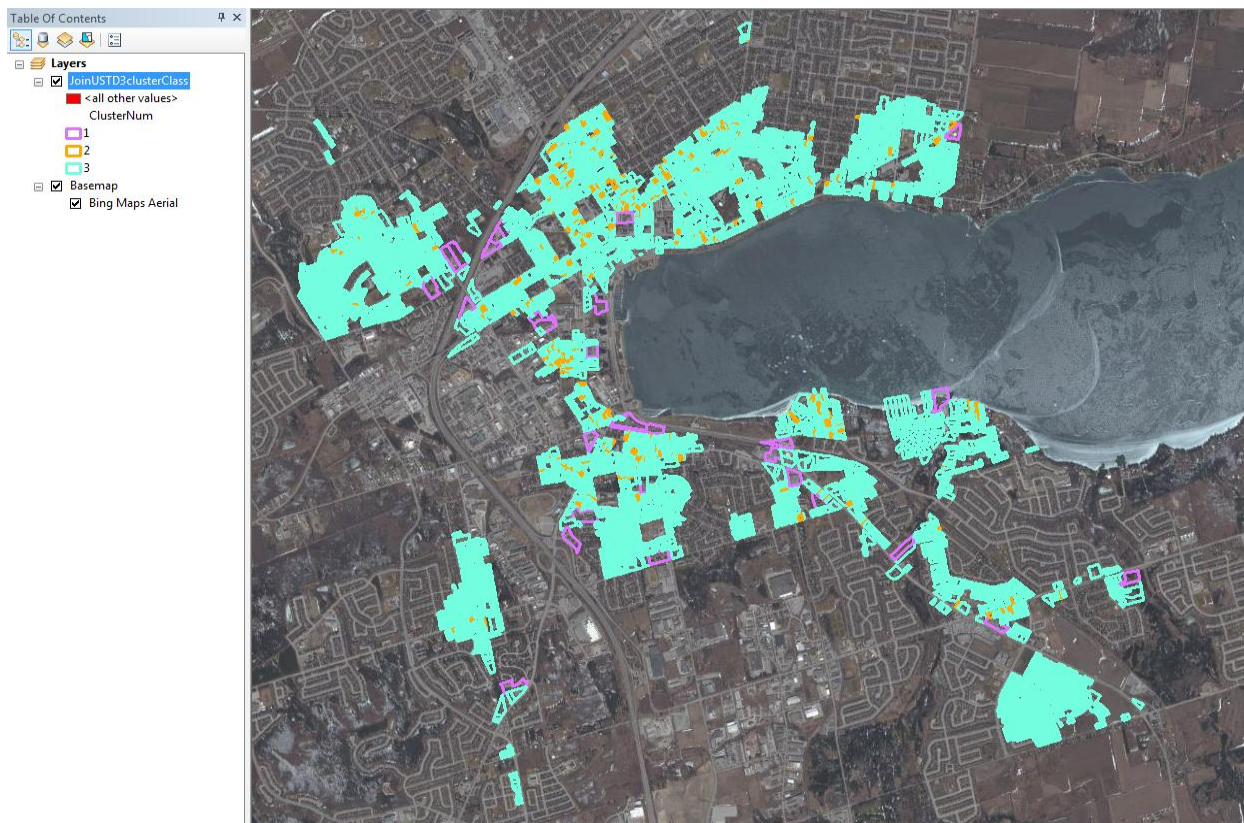

Figure 23. Cluster diagram for the 3-cluster group using the USTD technique and the Classifying Only method.

Cluster 3 in Figure 23 (light blue lot boundaries) contains the most lots which suggest a very uneven cluster structure. Therefore, the cluster diagram is mostly comprised of lots belonging to the third cluster. The composition of cluster groups does not mimic the spatial pattern exhibited by soil category. This implies that the USTD data standardization technique did not develop clusters of lots which were reflective of all of the existing hydrological characteristics, particularly soil type. Consequently, the USTD method was removed from further investigation.

The remaining three data standardization techniques resulted in cluster diagrams with similar spatial distributions of clusters for all of the different versions created. In order to detect the differences between the diagrams and determine which technique most closely represented reality, comparisons were made at the lot-level using various zooming techniques. The cluster diagrams generated based on data standardized by the Range 2 technique appeared to exhibit a spatial distribution of lots which reflected reality most closely. Lots assigned to a particular cluster group appeared to be more alike to one another than in cluster diagrams based on data standardized by the other two techniques. Based on this visual analysis, the Range 2 data standardization technique was determined to produce cluster groups which were most representative of the hydrological attributes in the study area.

After selecting the data standardization technique as Range 2, the differences between cluster diagrams between the two different methods of classifying cases were studied. A cluster diagram was produced for both the Classifying Only method and the ITCL method for each cluster group that was investigated for each of the data standardization techniques. This resulted in a total of eight cluster diagrams for each data standardization technique. It was expected that the cluster diagrams generated from using the ITCL method would produce a closer representation of reality due to the iterative nature of the technique. However, the difference

between the cluster diagrams produced by these two methods of classifying cases was not significantly noticeable.

The diagrams produced by the clusters generated by using the Classify Only method appeared to more closely represent the study area that was shown by the underlying orthophoto. For example, the cluster diagrams for the 10-cluster group of the Range 2 technique revealed different results when using either method. The cluster diagram produced by using the ITCL method is illustrated in the screen shot of Figure 24.
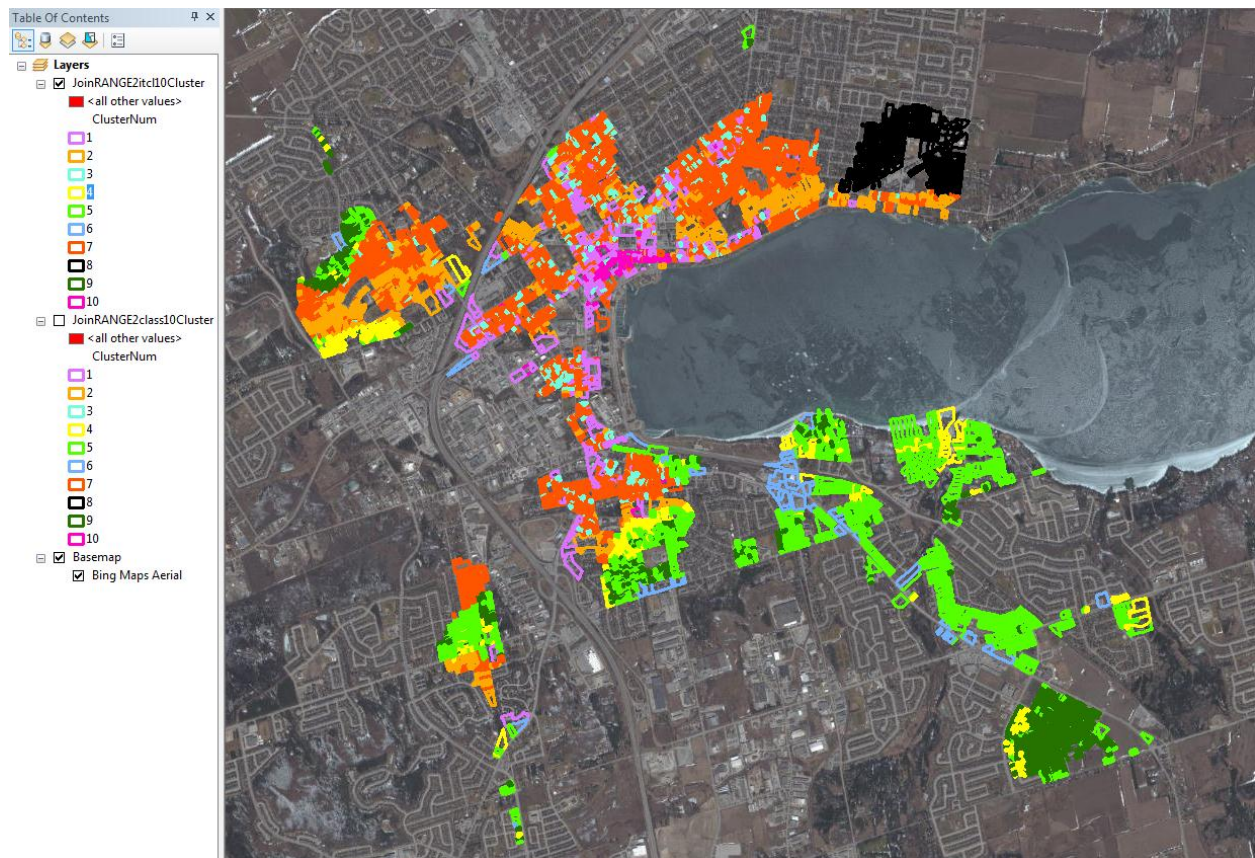
Figure 24. Cluster diagram for the 10-cluster group of the Range 2 method using the ITCL method.

The cluster diagram in Figure 24 shows a lot of variation with respect to the distribution of clusters. Neighbouring lots that have the same lot area (m^2) and that are similar with respect to

70

other attributes were placed into different clusters. After zooming in on such lots, it was difficult

to determine why they had been placed into different clusters by the clustering algorithm. This

result is similar for many of the other diagrams derived from the ITCL method. The cluster

diagrams produced by the ITCL method appear to create too much variability in the spatial

distribution of lots belonging to different clusters.

The cluster diagram derived from the Classify Only method, however, illustrates much less

variability in the spatial distribution of lots belonging to different clusters as illustrated in
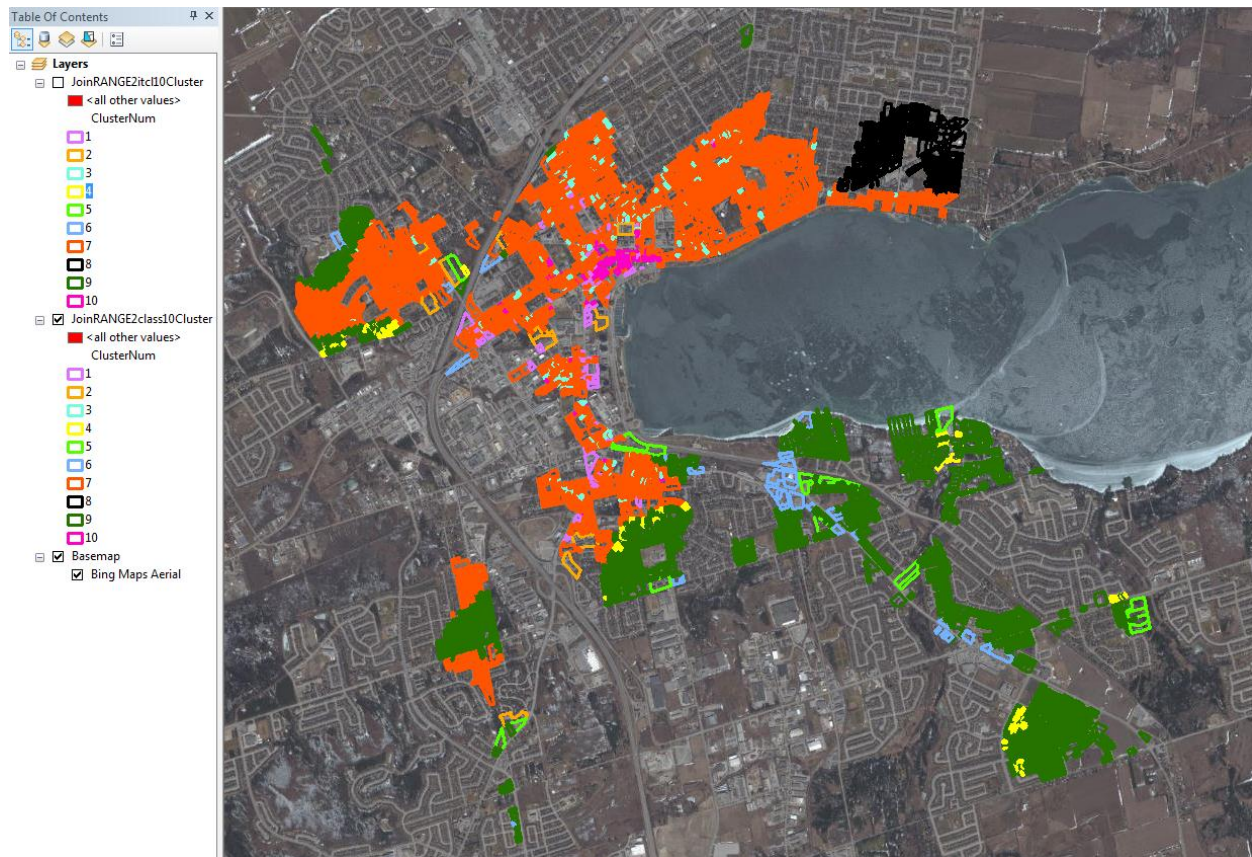
Figure 25:

Figure 25. Cluster diagram for the 10-cluster group of the Range 2 technique using the Classify Only
method.

A close inspection of the lots within the different clusters of the cluster diagram in Figure 25, using the Zoom and Pan Tools, revealed that lots within one cluster were fairly similar to one another and fairly dissimilar to lots contained in other clusters. For example, visible hydrological characteristics such as the lot area (m^2) and the imperviousness (%) were similar. The variation in the spatial distribution of clusters was accounted for by the visual similarity of the lots within a cluster even if the lots were scattered around and not in close proximity to one another. It was also apparent that the cluster arrangement in Figure 25 is based on the soil distribution shown in Figure 20 much more than the cluster arrangement shown in Figure 24. The examination of both visible characteristics such as lot area (m^2) and others which are not visible such as soil type support the selection of the Classify Only method as the method for classifying cases and the Range 2 data standardization technique.

5.4 Analysis of SWMM Output

The report file produced by SWMM software after performing a hydrological modelling simulation presents a few ways to confirm that the simulation was successful before accepting the results. Some of the ways apply to both the non-LID and LID cases, while others only apply to the LID case since there is more information contained in the report file. One way of confirming the results for both the non-LID and the LID situation is to look at the continuity error that is stated in the beginning of the report file after each run. An acceptable continuity error was considered to be within ± 10 % because this indicated that there were no significant problems with the modelling process. It was also considered preferable that the continuity error be as close to zero as possible. If the continuity error met these criteria, then the results in the report file were accepted.

Another way to assure that the simulation was successful for both the non-LID and the LID case entailed checking that the Runoff Coefficient in the Subcatchment Runoff Summary table was less than one. Additionally, the Surface Runoff in the Runoff Quantity Continuity section of the report is stated both as a volume (ha –m) and as a depth (mm). After using the lot area (m^2) for each lot in the Subcatchment Runoff Summary section to convert the Total Runoff (mm) for each lot to Total Runoff (m^3), the overall Total Runoff (m^3) for all of the lots was calculated. The calculated overall total value was confirmed to be equivalent to the Surface Runoff (ha-m) stated in the Runoff Quantity Continuity section of the report.

Furthermore, the SWMM output for the LID scenario was confirmed to be correct by checking that the Per cent Error for each lot listed within the LID Performance Summary table was less than ± 1 %. The Per cent Error is calculated as:

Pcnt. Error = ((Total Inflow (mm) − Drain Outflow (mm))/ (Total Inflow (mm))*100 (3)

If the Per cent Error is negative then the amount of water flowing out of the LID control is greater than the amount of water flowing into the LID control. This result is acceptable as long as the negative percentage value is not greater than 1 % in absolute value.

Therefore, the simulation was confirmed to be successful for both the LID case and the non-LID case before accepting the final results by reviewing and checking the report file. The input file produced by SWMM was also confirmed to be correct by checking that the input data matched the data contained in the common database and that selections such as flow units were correct. After accepting the results in both the input file and the report file, the Total Runoff (mm) for each lot was plotted against each of the input variables under investigation for all of the 7871 lots without a LID. The plots are illustrated in Appendix A.4, aside from the two plots illustrated in

Figure 26 and Figure 27. The plot of runoff per lot (mm) versus soil categorical value is not

illustrated because the soil categorical value remains constant for each runoff per lot (mm) value.
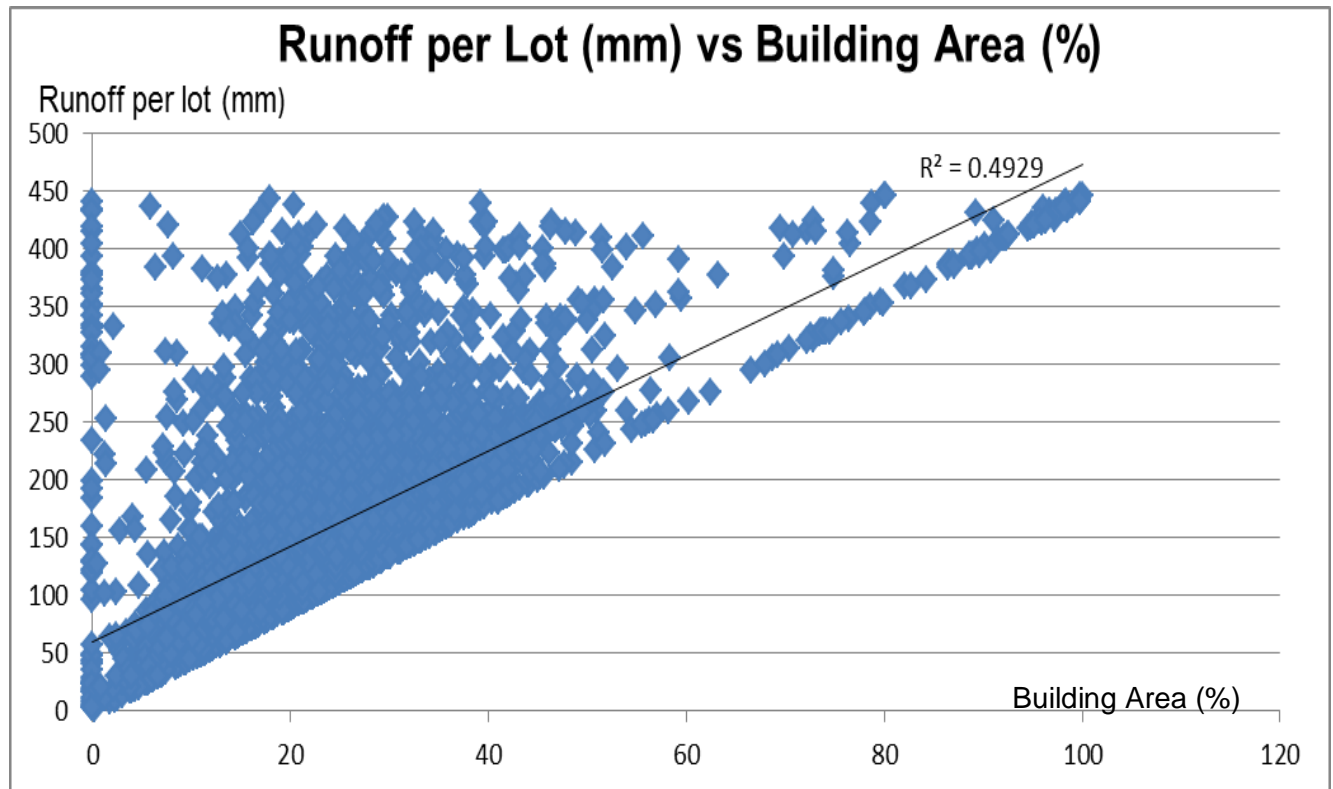
Figure 26. The relationship between runoff per lot (mm) and building area (%) for the no LID case.

The graph in Figure 26 illustrates the weak linear relationship between runoff per lot (mm) and

building area (%). The linear relationship is described as weak because $R^2 = 0.50$. Attempts

were made to fit other Excel trendlines to achieve the highest R^2 value, however other

relationships such as exponential, for example, also proved to be weak.

The Total Runoff (mm) is the output of the hydrological modelling run listed for each lot which

can also be described as runoff per lot (mm). The linear relationship in Figure 26 shows that in

general, as the percentage of the building area increases on the property the runoff per lot (mm)

also increases. An increase in the percentage of building area increases the percentage of

imperviousness. Therefore, as there is more impervious area there is more runoff because less rainwater can seep into the ground.

The Total Runoff (mm) value stated in the report file was used to plot the graphs as opposed to the calculated total runoff (m^3) because the Total Runoff (mm) is already normalized for lot area (m^2). Therefore, the size of the lot does not impact the Total Runoff (mm) value unlike total runoff (m^3) which includes the lot area (m^2) value and could have affected the results of the graphs. The points on the graph assume a triangular shape with a lower diagonal boundary and a boundary on the y-axis. Similar bounded triangular shapes are visible on the two graphs which plot the remaining components of imperviousness (%) in Appendix A.4: parking area (%) and driveway area (%). However, the remaining two graphs in Appendix A.4 for average slope (%) and parcel width (m) have no distinctive pattern or shape to indicate the type of relationship between each variable and runoff per lot (mm).

A possible explanation for the similar bounded triangular shapes seen on the graphs involving building area (%), parking area (%), and driveway area (%) is the connection between the data points as a result of being components of imperviousness (%). The values for runoff per lot (mm) or the y-axis values are the same for each plot while the x-axis values or the percentage impervious area values change. The boundary on the y-axis is due to many lots having 0 % of a particular type of impervious area; however runoff was still generated as a result of having a large percentage of another type of impervious area. The diagonal boundary is formed by some of the lots which were mainly composed of one type of impervious area. The diagonal boundary represents the minimum runoff generated as a result of having a certain percentage of imperviousness.

Since the input values used to run SWMM were varied for thousands of lots, the types of graphs plotted in Figures 26, 27 and Appendix A.4 can be viewed as sensitivity analyses with respect to the input variable under question. These graphs can be viewed as sensitivity analyses because the change in output or Total Runoff (mm) is determined with respect to changes in the input variables. Although the relationship between the output variable and the input variable is somewhat clear in Figure 26, this was not always the case. The relationship between runoff per lot (mm) and lot area (m^2) for the no LID scenario is not immediately clear in Figure 27:
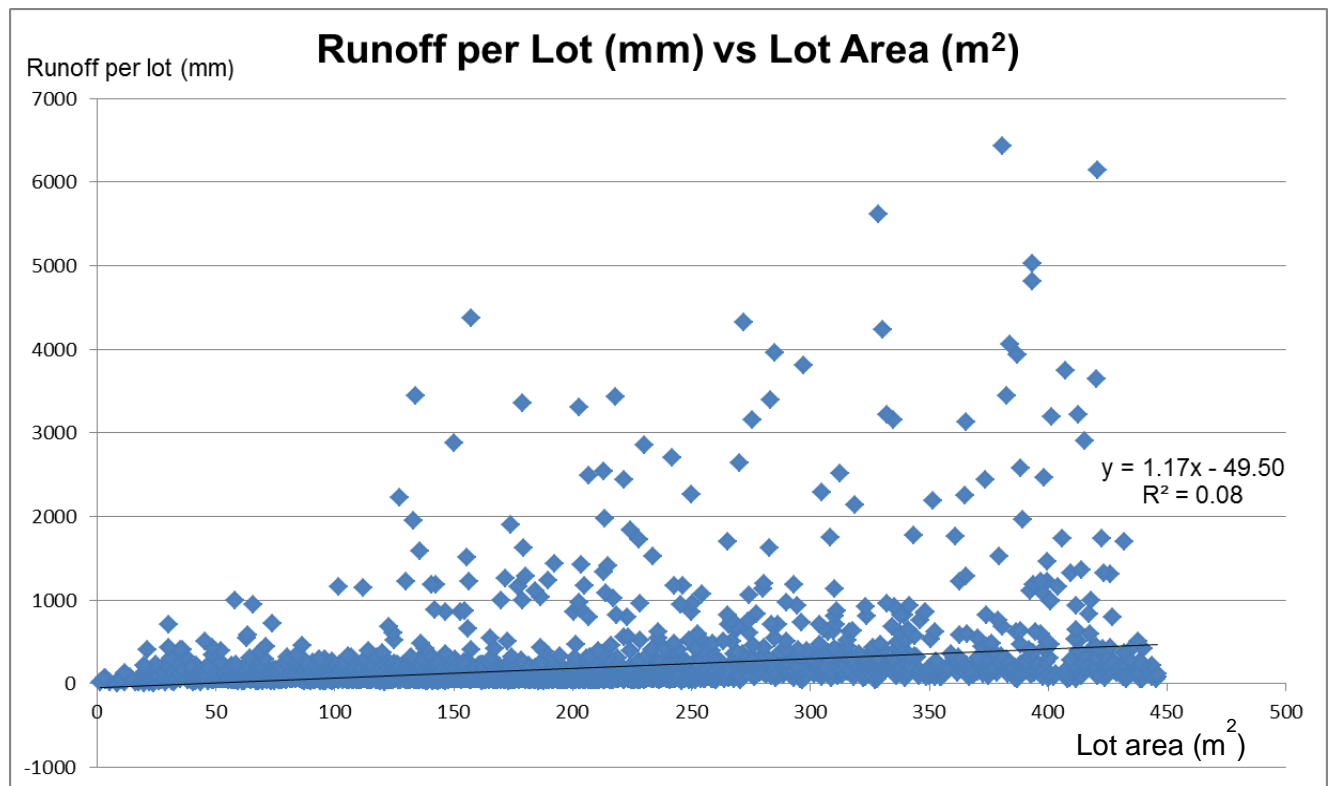
Figure 27. Runoff per lot (mm) plotted against lot area (m^2) for the no LID case.

If only the location of the data points is observed, the linear relationship in Figure 27 is less obvious than in Figure 26. The graphical line drawn indicates a very weak but existent linear relationship that is demonstrated by a low R^2 value of approximately 0.1. As the lot area (m^2) increases, the runoff per lot (mm) increases only slightly. This is logical depending upon the

76

composition and shape of the lot. For instance, if the level of imperviousness remains constant as the lot area (m^2) increases then the runoff per lot (mm) should decrease. If the lot becomes wider as the lot area (m^2) increases, then the runoff per lot (mm) should increase because the runoff has less of an opportunity to infiltrate into the ground. Therefore, the output or Total Runoff (mm) of each lot should be less sensitive to changes in lot area (m^2) than to changes in building area (%). The lower sensitivity level is illustrated by the flatter line in Figure 27 than in Figure 26. The relationship between the output variable to the input variables has the potential to change after all of the input variables are combined into one equation using linear regression.

5.5 Calculation of Total Runoff (m^3) Results

5.5.1 Results Using the Minimum Distance Method

The total runoff (m^3) for each group of clusters was calculated and plotted against the number of clusters for both the case with the RH LID and the case without the RH LID. The total runoff (m^3) was calculated by extrapolating the Total Runoff (mm) produced by the lot with the minimum distance to the final cluster centre to the rest of the lots in the cluster. The results for the case with no LID or all available residential lots are shown in Figure 28:
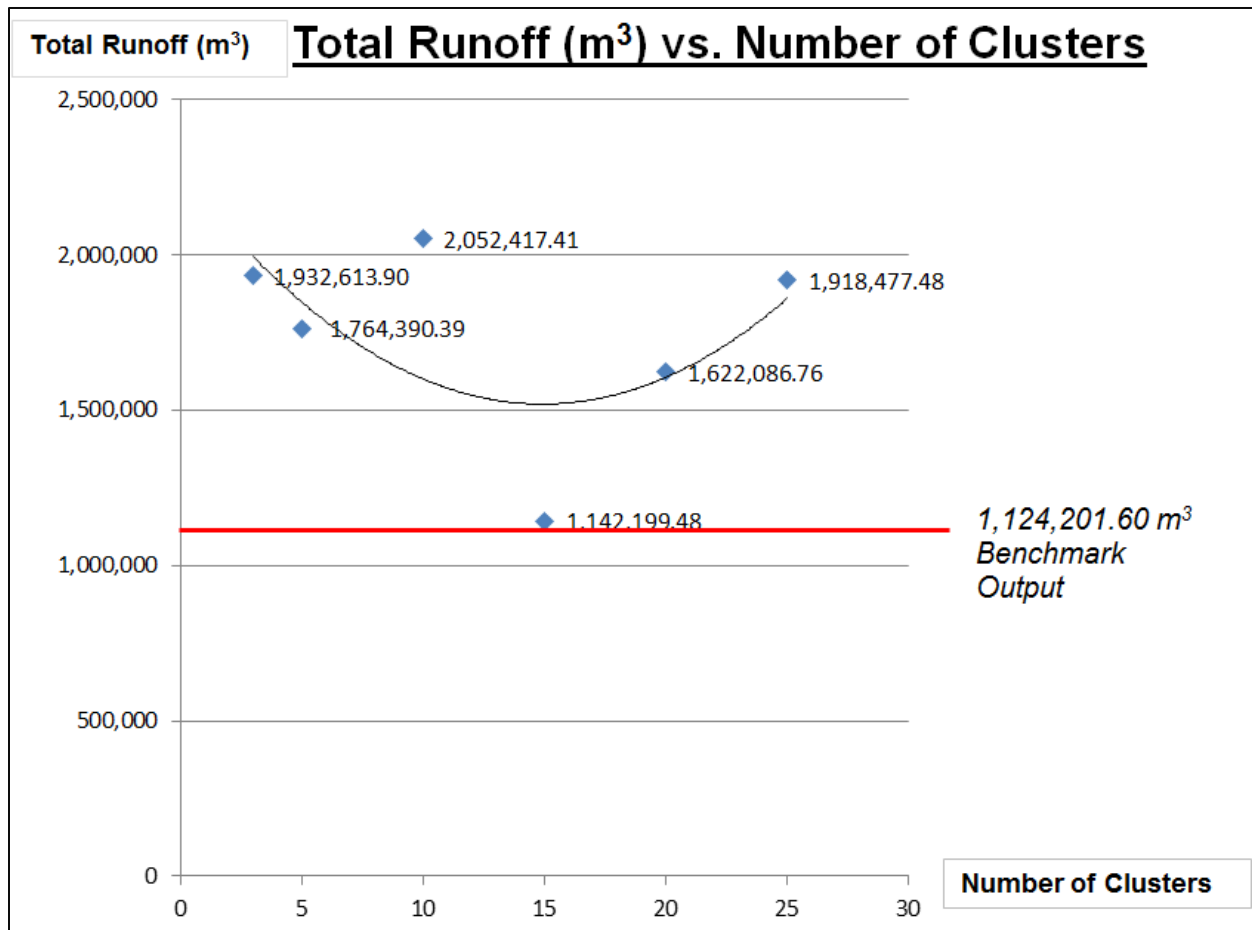
Figure 28. Total runoff (m^3) versus number of clusters for the no LID case.

The benchmark output after running all residential lots with no LID is shown in Figure 28 as the red line. This output is the summation of the total runoff (m^3) that was calculated for all lots using the final SWMM output. It is considered to be the "true" total runoff (m^3) or the "truth" for the no LID case to which all other runoff (m^3) output values can be compared.

The data points in Figure 28 illustrate the total runoff (m^3) for each cluster group as calculated by the Minimum Distance Method which used an approximation to the final cluster centre. In this experiment, the minimum distance to the final cluster centre within a cluster group was very often zero. Therefore, the lot which had a distance to the final cluster centre of zero was used to approximate the final cluster centre.

The results in Figure 28 demonstrate that 15 clusters provide the closest total runoff (m^3) value to the benchmark output value. Therefore, instead of performing hydrological modelling on each lot for all 7871 residential lots, the total runoff (mm) from 15 final cluster centres can be used to approximate the benchmark or true output value. The other cluster groups provide too high of an approximation for the benchmark output and form a U-shaped curve.

However, there is no consistent pattern or relationship between the total runoff (m^3) and the number of clusters developed for the Minimum Distance Method. This is because the output produced by each cluster within a cluster group varies according to the lots selected. The different lots selected are dependent on the clustering algorithm which assigns lots to groups based on the Euclidean distance dissimilarity measure. Due to the arbitrary nature of the clustering process, some groups of clusters may approximate the benchmark output more closely at times than others. However, the results give a rough indication as to the number of clusters or UHRUs that may be required to attain the benchmark output.

The total runoff (m^3) of each cluster that was calculated using the Minimum Distance Method was compared against the total runoff (m^3) of each cluster that was derived directly from the output of the SWMM run. The total runoff (m^3) of each cluster using the Minimum Distance Method was calculated based on the Total Runoff (mm) of the lot with the minimum distance to the final cluster centre and the total lot area (m^2) for the cluster. This output is referred to as the calculated total runoff (m^3) of a cluster. The modelled total runoff (m^3) refers to the total runoff (m^3) for each cluster which was calculated using the Total Runoff (mm) output values for each lot in the SWMM report file. The Total Runoff (mm) was converted to total runoff (m^3) using the area of each lot (m^2) and then summed for the entire cluster. The calculated total

runoff (m^3) is plotted against the modelled total runoff (m^3) for the 10-cluster group from the no LID case in Figure 29:
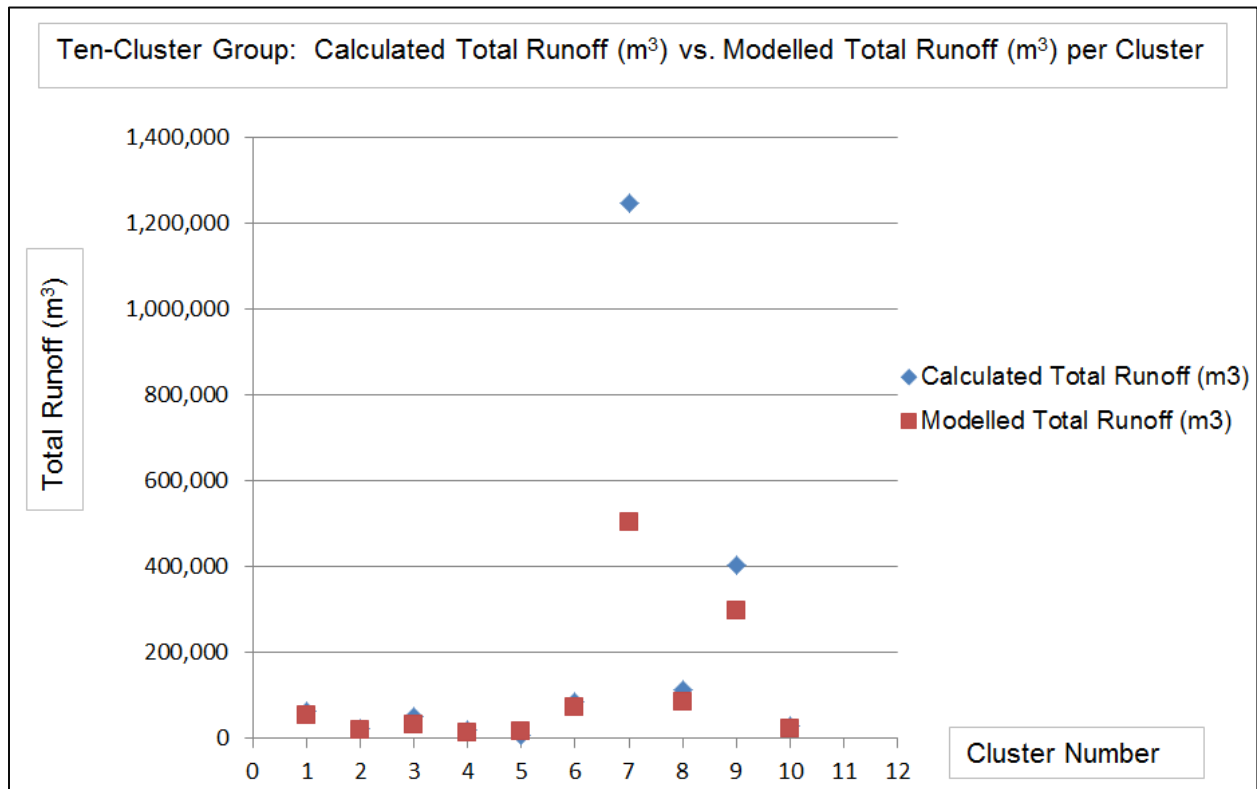
Figure 29. Calculated total runoff (m^3) versus modelled total runoff (m^3) for each cluster group.

The calculated total runoff (m^3) is the same as the modelled total runoff (m^3) for each cluster within the 10-cluster group except for clusters number 7 and 9. The method of extrapolating the results of the final cluster centre lot to the entire cluster is therefore effective at capturing an approximate output value for a cluster. Clusters number 7 and 9 both exhibit lower modelled total runoff (m^3) values than calculated total runoff (m^3) values. Extrapolation of the total runoff (m^3) produced by the lot with the minimum distance to the final cluster centre lot overestimated the actual total runoff (m^3). This outcome suggests that the final cluster centre was not effective in its representation of the entire cluster and was probably swayed by some

unusually high or low values for some of the hydrological characteristics studied. However, it does not imply that there are errors in the data or outliers because all of the data is real, therefore lots with unusually high or low values are expected although some lots with extreme values were removed.

The case involving the implementation of the RH LID on each lot produced similar results with respect to the Minimum Distance Method. The lot with the minimum distance to the final cluster centre was again used to extrapolate the total runoff (mm) output to all of the lots within the cluster. The total runoff (m^3) produced by each cluster group was plotted against the number of clusters within a cluster group to produce Figure 30:
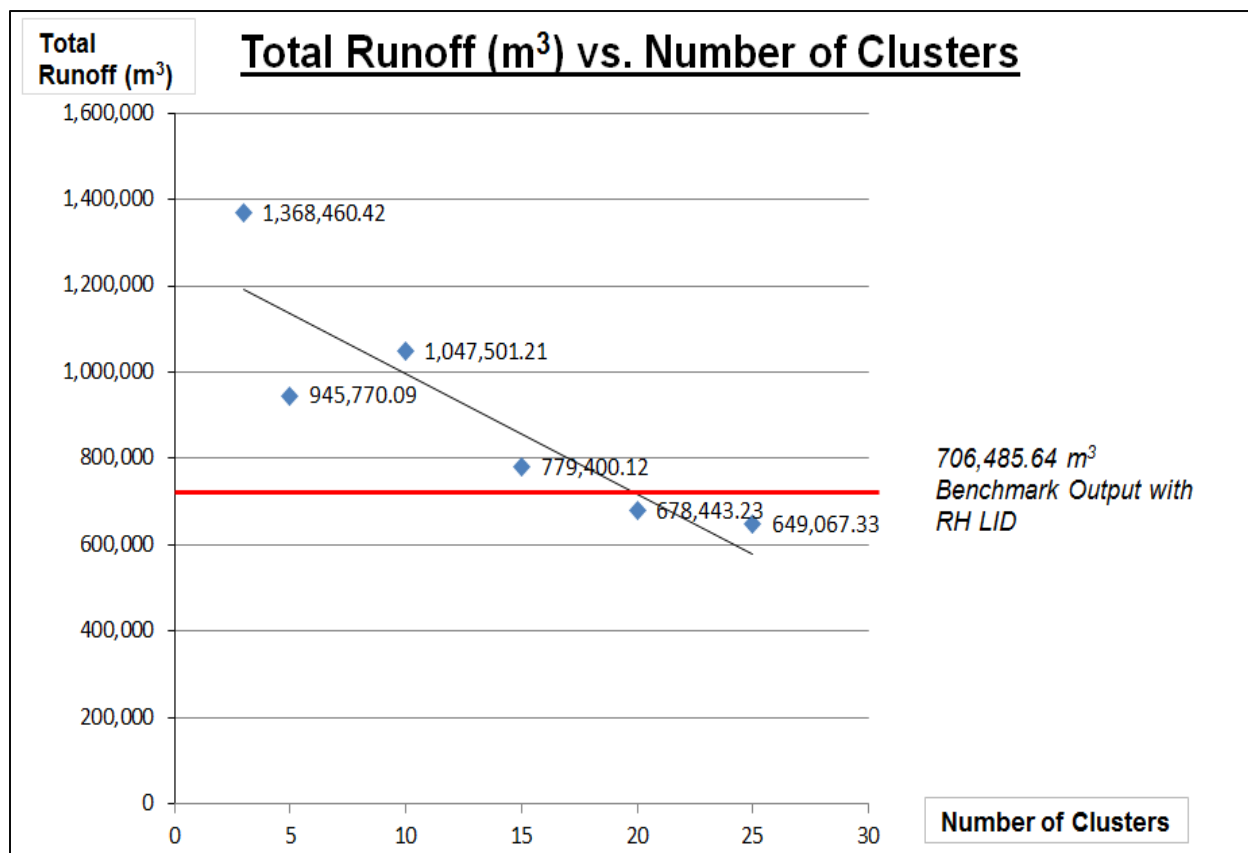

Figure 30. Total runoff (m^3) versus number of clusters for residential lots with the RH LID.

In Figure 30, the benchmark output from running all of the lots with LID implementation is 706,485.64 m^3. The benchmark output value is 417,715.957 m^3 less than in the no LID situation which has a benchmark output value of 1,124,201.597 m^3 of runoff. However, about 1,000 less lots were modelled in the LID scenario than in the no LID scenario because some lots were not amenable to RH LID implementation. For the lots which were modelled in the RH LID case, the total runoff (m^3) before LID implementation was 947,278.95 m^3 and the total runoff (m^3) after LID implementation was the benchmark output of 706,485.64 m^3. The reduction in runoff as a result of LID implementation was 240,793.31 m^3 or 25.42 %. The RH LID was effective at reducing runoff which was the main purpose of LID implementation in this experiment.

In Figure 30, the 20-cluster group most closely approximates the benchmark output for the "with LID" situation, although both the 15-cluster group and the 25-cluster group also yield a close approximation. In the "with LID" situation, there is a linear, decreasing relationship between the total runoff (m^3) produced and the number of clusters. As the total number of clusters increases, the total runoff (m^3) decreases. The relationship is fairly strong with an R^2 value of approximately 0.78, however, it cannot be concluded that the relationship is linear until larger groups of clusters are studied.

Furthermore, the decrease in total runoff (m^3) with respect to the number of clusters is not completely consistent. For example, the 5-cluster group produces a smaller total runoff (m^3) than the 10-cluster group. Therefore, the decreasing relationship may not be generalizable to every situation with a RH LID. It may be just the result of the particular selection of lots for the clusters within these cluster groups. The total runoff (m^3) also seems to be continually decreasing past the point of the benchmark output. If the decreasing relationship continues, then

the total runoff (m^3) generated by larger groups of clusters will deviate much further from the

true benchmark value.

The calculated total runoff (m^3) value and the modelled total runoff (m^3) value for each cluster

within a cluster group were also compared with one another for the condition of having a

RH LID. The results are illustrated in Figure 31 for each cluster of the 15-cluster group for the

RH LID scenario:
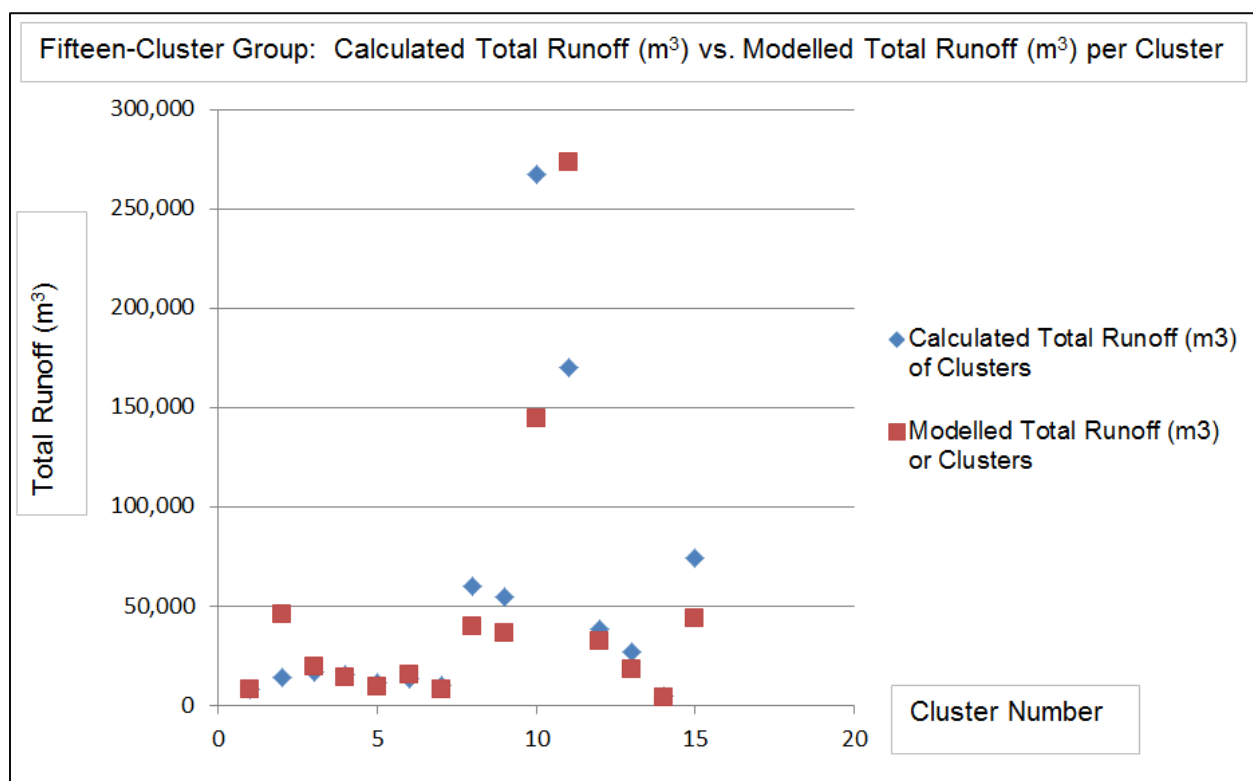
Figure 31. Calculated total runoff (m^3) versus modelled total runoff (m^3) for the 15-cluster group of the
RH LID scenario.

Figure 31 illustrates that the calculated total runoff (m^3) and the modelled total runoff (m^3) are

nearly the same for almost all of the cluster groups. However, they are not always equal and

sometimes the modelled total runoff (m^3) is higher in value and sometimes the calculated total

runoff (m^3) is higher in value. The final cluster centre represents the mean of all of the input

variables for the lots within the cluster. The difference in total runoff (m^3) between the calculated and modelled total runoff (m^3) values is due to some lots having input variables that are above or beneath the mean value of the input variable in the cluster. Therefore, the final cluster centre is not always a useful representation of all of the cases in the cluster. Since the calculated and modelled values are the same or quite close for the majority of the clusters, the vast majority of lots within a cluster seem to be well-represented by the final cluster centre. The final cluster centre is therefore a fairly accurate representation of the mean values pertaining to the lots within the cluster and can be used to approximate the total runoff (m^3) output of the entire cluster. The total runoff (m^3) is also well-explained by the input variables used to form the cluster.

5.5.2 Results Using the Random Sampling Method

The random sampling method involves arbitrarily selecting a sample of lots from each cluster and performing a linear regression to derive an expression for the entire cluster. The unstandardized input variables were used for both the dependent and the independent variables. The expression or regression equation was used to calculate the output or runoff (mm) for each lot in a cluster. The total runoff (m^3) was then calculated for each lot and summed for the entire cluster. Lastly, the total runoff (m^3) for all of the clusters in a group of clusters was summed.

The method was first used by randomly selecting 5 % of the lots from each cluster within the 3-cluster group for all of the residential lots without LIDs. A linear regression was then run on each of the three samples to estimate the output or total runoff (mm) for each lot. A stepwise method of linear regression was used; therefore more than one equation model was usually generated for each cluster. The equation model was selected based on the criteria described in

section 5.5.2 of the Methodology. The three regression equation models in Table 6 were selected

for the 3-cluster group with a random selection of 5 % of all residential lots for the non-LID

situation.

Table 6. Regression equations used for the 3-cluster group with 5 % random lot selection for the no LID case.

Cluster Number	Model Selected	Regression Equation
1	2	y = 0.606 + 4.281(Parking Area (%)) + 4.660(Building Area (%))
2	3	y = 9.872 + 4.377(Building Area (%)) + 4.438(Parking Area (%)) + 4.158(Driveway Area (%))
3	4	y = 0.506 + 4.435 (Building Area (%)) + 4.365(Parking Area (%)) + 4.418(Driveway Area (%)) + 0.162 (Average Slope (%))

The dependent variable, y, is total runoff (mm). The y-value was converted to total runoff (m^3)

and then summed for all of the lots within a cluster. The total runoff (m^3) for all of the clusters

was then summed. The total runoff (m^3) for the 3-cluster group was 1,122,287.01 m^3 which

closely approached the benchmark output of 1,124,201.60 m^3 but did not attain the value exactly.

A greater proportion of each cluster was expected to yield a greater total runoff that might attain

the benchmark output value exactly. Therefore 10 % of the lots from each cluster were

randomly selected and the random sampling method was repeated for the 3-cluster group,

followed by the performance of linear regression for each cluster. The total runoff (m^3) for all

three clusters after running a linear regression with 10 % of randomly selected lots from each

cluster was 1,126,562.95 m^3 which was slightly above the benchmark output.

Since a 5 % random sample produced a value that was too low and a 10 % random sample produced a value that was too high, the correct percentage to use for random sampling of lots was thought to be somewhere in between those values. By linear interpolation, a value of 7 % was calculated to produce a total runoff (m^3) value that approximates the benchmark value very closely. The random sampling method was repeated with a 7 % random lot selection from each cluster within the 3-cluster group. A linear regression was then run and a final output value for all three clusters was calculated to be 1,123,287.26 m^3 which most closely approximated the benchmark value compared with the other random samples for the 3-cluster group. Therefore, a 7 % random sample of lots from each cluster within a 3-cluster group almost exactly yields the benchmark output from running all 7871 residential lots using SWMM as illustrated in Figure 32.
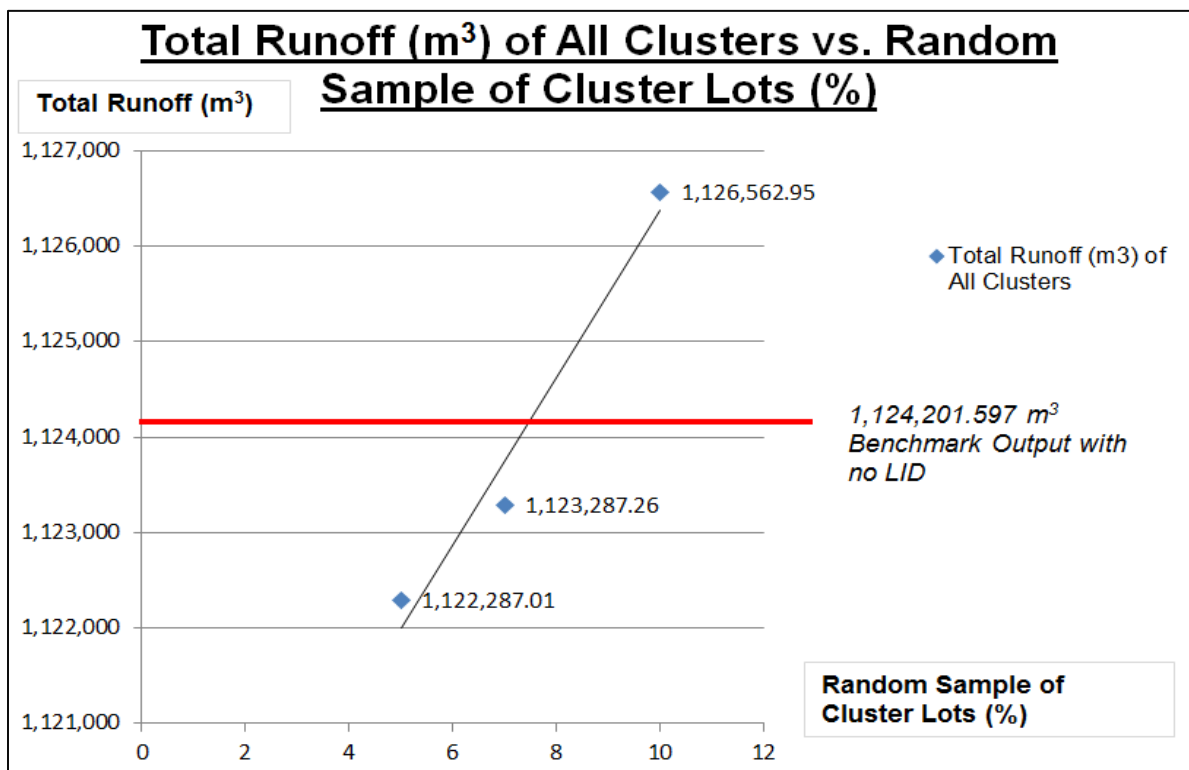

Figure 32. Total runoff (m^3) of the 3-cluster group versus percentage of lots selected randomly.

The same random sample percentages of 5 %, 7 %, and 10 % were repeated for cluster groups that had more clusters in order to see if another solution was possible. The results were much poorer than those for the 3-cluster group. The Random Sampling Method using the 5-cluster group is ineffective for the sampling percentages of 5 %, 7 %, and 10 %. When this approach was applied to the 10-cluster group for all residential lots without a LID, similar results were found. Therefore in order to approximate the benchmark output value for all residential lots without a LID, 7 % of the lots should be randomly selected from each cluster of the 3-cluster group.

The same result was not found, however, for the regression analysis involving the lots with the implementation of the RH LID. The result from taking a 5 % random sample from the 3-cluster group was much lower than the benchmark output for all lots with the RH LID of 705,929.07 m^3. The percentage of the random sample was then increased to 25 % in an attempt to increase the runoff (m^3) output from the 3-cluster group. However, there was little increase in the volume of total runoff (m^3). This prompted a further increase in the percentage of random sample selected to 50 % which resulted in an increase in the total runoff (m^3) volume to 367,332.35 m^3 which was still only about half of the benchmark output value. The regression equations in Table 7 were used in the trial involving a 50 % random sample from the 3-cluster group.

Table 7. Regression equations for a 50 % random sample from the 3-cluster group for the LID case.

Cluster Number	Model Selected	Linear Regression Equation
1	3	y = 33.82 + 3.39(Parking Area (%)) + 3.28(Driveway Area (%)) + 1.74 (Building Area (%))
2	3	y = -11.50 + 3.19(Parking Area (%)) + 2.19(Building Area (%)) + 29.40(Soil Categorical Value)
3	2	y = 42.05 + 3.64(Parking Area (%)) + 0.72(Building Area (%))

The same percentages of random samples were then taken from the 5-cluster group and the 10-cluster group in order to see whether cluster group runoff (m^3) would be closer to the benchmark runoff (m^3).

The solution with the first, closest approximation to the benchmark output value was the 50 % sample from the 3-cluster group for the case which implemented the RH LID. A 90 % sample was then taken from each cluster within the 3-cluster group and the LID benchmark output was attained almost exactly with an output value of 711,727.45 m^3. By linear interpolation, the benchmark output value should be attained with a random sample percentage of approximately 89 %. However, after running a regression on the 89 % random sample for the 3-cluster group, an output value of 638,833.68 m^3 was attained which was not as close to the benchmark output value as the 90 % random sample result. Therefore, in order to approximate the benchmark output value for the rain barrel LID case, a 90 % random sample from the 3-cluster group can be used as illustrated in Figure 33:
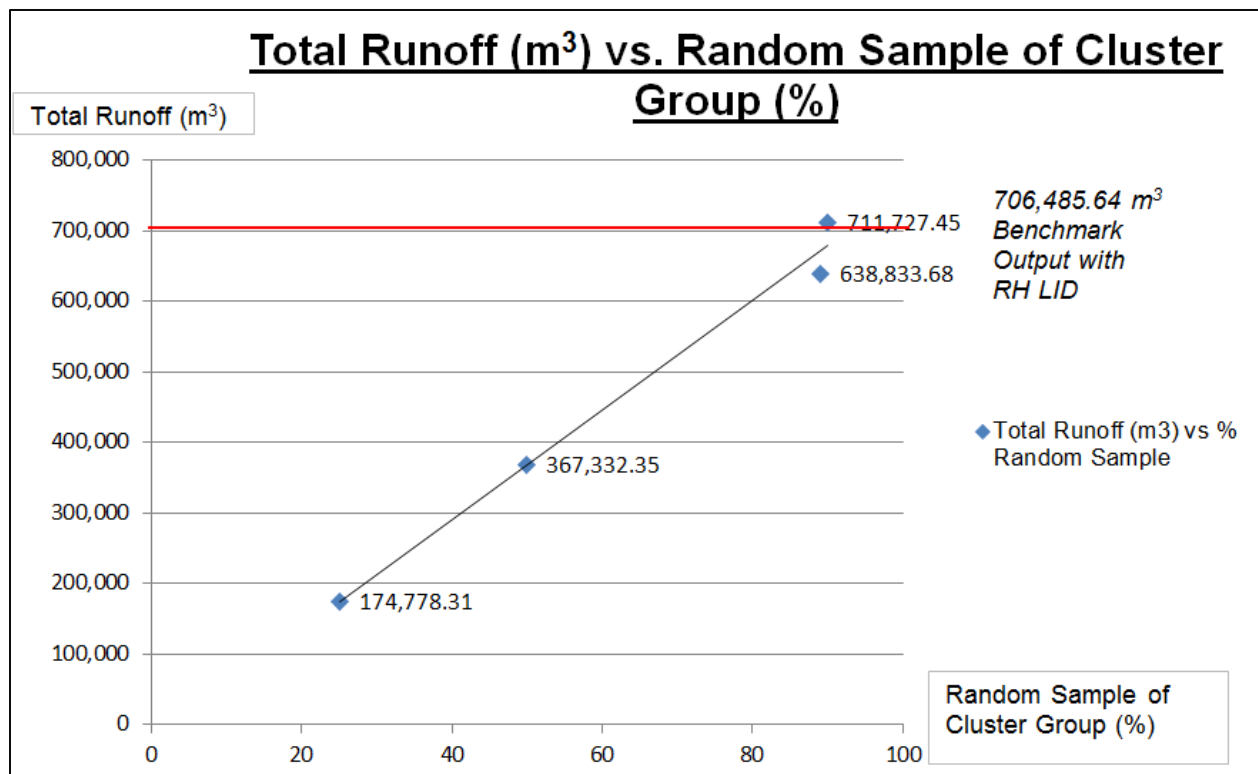
Figure 33. Illustration of the random sample that most closely approximates the benchmark output value for the RH LID case.

Nevertheless, linear regression models could not be generated for every data set within a group of clusters since some clusters contained too few lots or data rows. When small random samples of lots were taken from each cluster, such as 5 % random samples, the input data set for linear regression only contained a few lots. In cases such as these, the linear regression software was often unable to derive regression equations for the cluster. As a result, output from that particular cluster was not included in the total runoff (m³) output for the group of clusters which decreased the final output value.

This problem rarely occurred for the case of non-LID residential lots but was quite common for the case of RH LID residential lots, especially when the percentage of lots selected from each cluster was small. For example, the 5 % random sample from the 10-cluster group of the LID

case yielded samples of clusters for which regression equations could not be derived. The reason that the software could not develop equation models for those particular clusters was probably due to the fact that they each contained only one lot. In general, the more input data that is supplied to the regression software, the better the chances are of successfully performing a linear regression.

Nonetheless, the inability to derive a regression equation may not have had a major impact on the overall total runoff (m^3) result for the clusters of the LID scenario with low percentages of random samples. This is because each random sample was not contributing enough total runoff (m^3) so that the benchmark output could be closely reached. The omission of one or a few very small random samples would therefore not have altered the final outcome to a great extent.

5.5.3 Results of Random Samples Drawn from the Non-LID Case and the LID Case

Samples of lots were randomly selected from the data sets of both the non-LID case and the LID case. For the non-LID case, percentage random samples were drawn in an attempt to attain the benchmark output of 1,124,201.60 m^3 of total runoff. Figure 34 demonstrates that the percentage random samples approach 100 % as the total runoff (m^3) values move towards the benchmark output value.
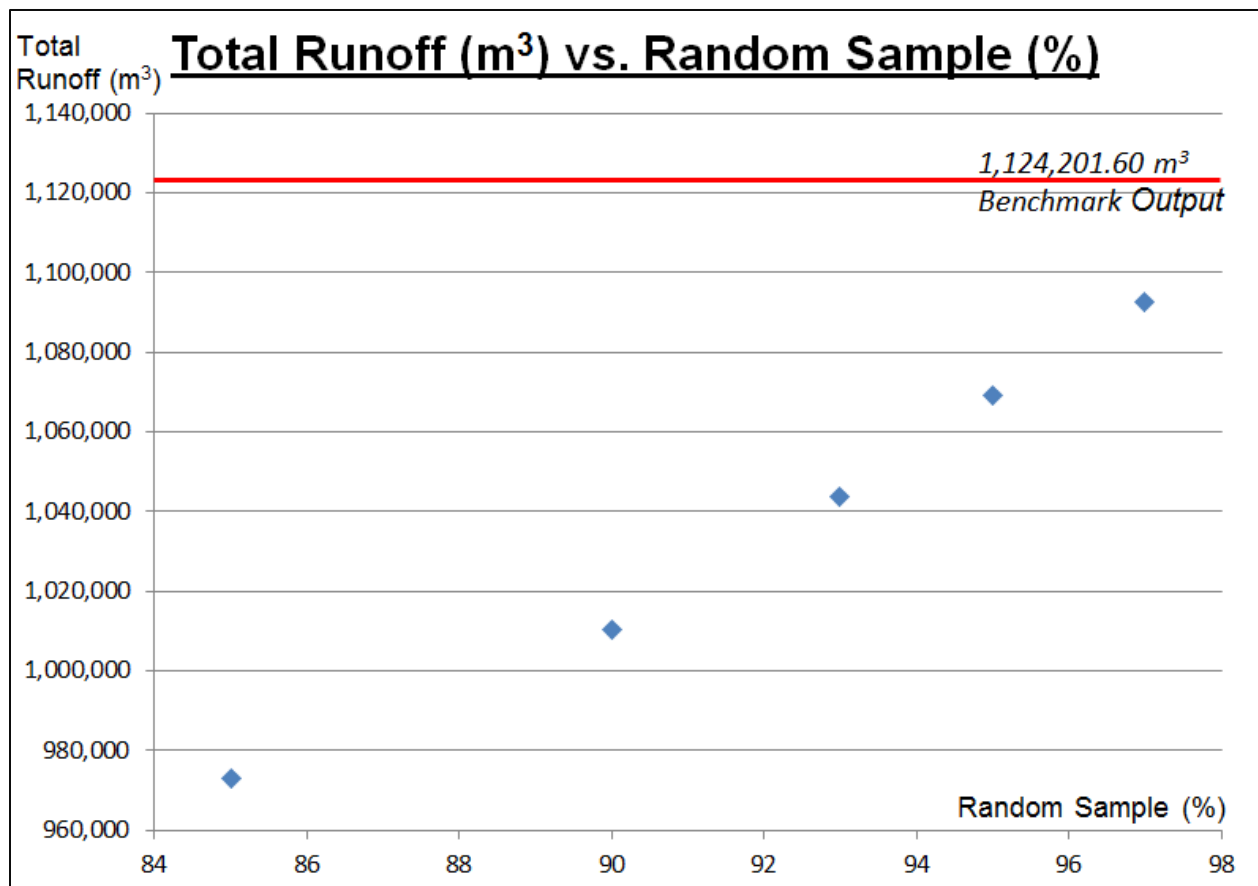
Figure 34. Total runoff (m^3) versus percentage random sample for the non-LID case.

Therefore, all of the lots or nearly all of the lots in the non-LID case must be included in the sample in order to approximate or attain the benchmark output. As a result, taking a percentage random sample from the entire data set for the non-LID case is unnecessary.

In this analysis, the same result was found for the LID case and the non-LID case. Percentage random samples which approached the entire data set produced an approximate benchmark output value for all lots with RH LID implementation as shown in Figure 35:
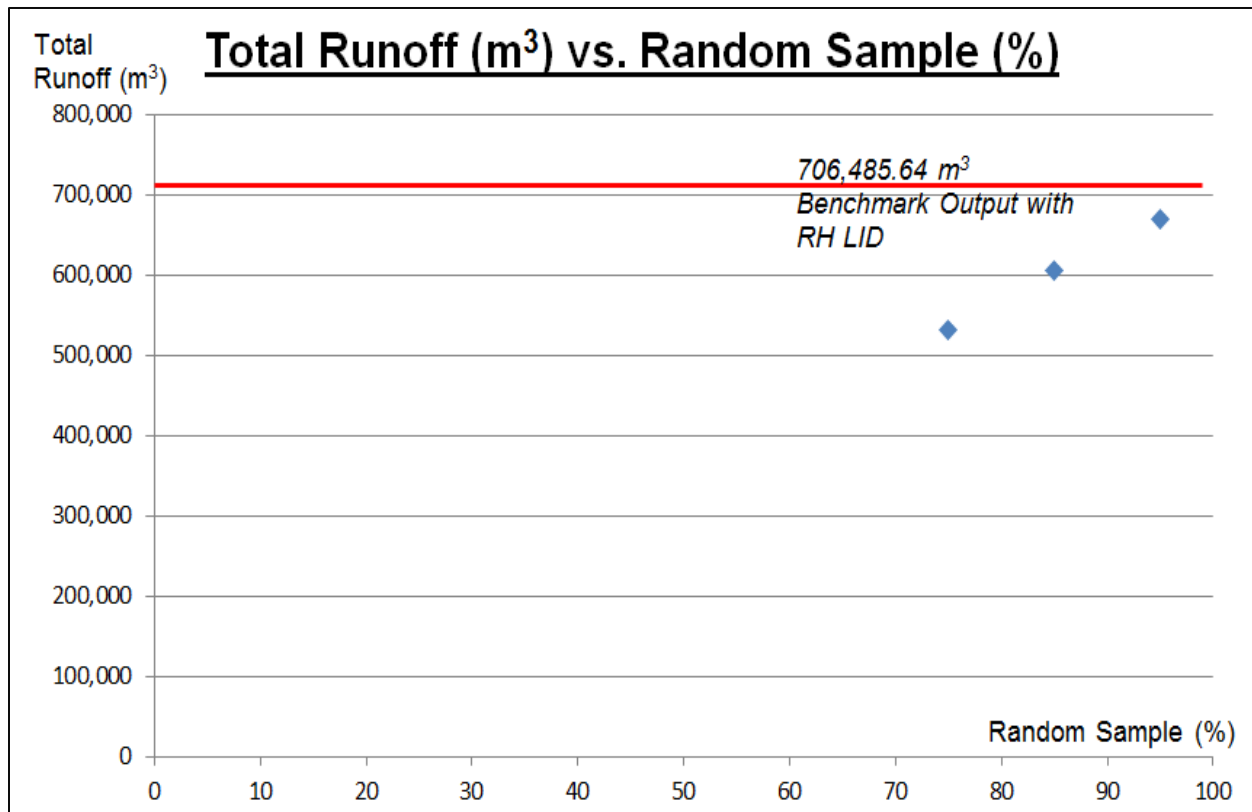
Figure 35. Total runoff (m^3) versus percentage random sample for the LID case.

Therefore, this method of calculating total runoff (m^3) does not offer any additional advantages over running all of the lots in the data sets for both the LID case and the non-LID case. Randomly selecting a percentage of lots from the data sets for both the non-LID case and the LID case is almost equivalent to running all of the lots in each data set.

6. **Conclusions and Recommendations**

The output from the cluster analyses initially exhibited an unbalanced cluster structure in which most of the cases were grouped together in the first cluster, regardless of how many clusters were formed or which data set was analysed. Hence, it is evident that standardizing the input variables so that they are not associated with different units of measurement is a very important step when performing K-Means Cluster Analysis. However, the distribution of each input data variable was not affected by applying a data standardization technique.

The investigation into the standardization of the input variables revealed that the Range 2 technique is the most effective at clustering data which agreed with the literature found on this topic. This was further demonstrated by the cluster diagrams which were developed using the standardized data of the Range 2 technique. The clusters formed in these cluster diagrams generally agreed with the distribution of the visible hydrological attributes shown on the underlying orthophoto. There was no significant difference between using the Classifying Only method and using the ITCL method for classifying cases with the K-Means clustering algorithm. However, the Classifying Only method seemed to be more reflective of the visible hydrological characteristics as seen in the orthophoto.

Therefore, the Classifying Only method was found to be the most effective at creating clusters which were representative of the underlying hydrological conditions when using data which was standardized by the Range 2 data standardization technique. Nonetheless, visual analysis of the cluster diagrams was a subjective process. Different people may come to different conclusions when undertaking this process. Future research should strive to develop a less subjective way of selecting a method for classifying cases. Since the literature review provided sufficient evidence to support the use of the Range 2 data standardization technique and the technique provided

satisfactory clustering analysis results, the investigation of different techniques is unnecessary. It is recommended this step be omitted in similar future investigations.

The Classifying Only method was selected in this experiment. However, the ITCL method was not investigated in an in-depth manner. There are more options available when using the ITCL method than when using the Classifying Only method. These options should be studied in further detail before it is confirmed that the Classifying Only method produces more satisfactory clustering analysis results. For example, the initial cluster centres can be supplied by the researcher when using the ITCL method instead of allowing them to be automatically generated by the software.

Furthermore, the focus of this study was on residential properties. Properties associated with other types of land use should also be examined. Land use category can even be incorporated as an input variable when performing the clustering analysis and subsequent regression analysis. During the clustering analysis, land use category would be considered as a categorical variable much like the soil categorical variable. The effect of land use category on runoff response can therefore be studied.

The Minimum Distance Method was used to approximate the benchmark output. The results from using this method suggested that the selection of the lot with the minimum distance to the final cluster centre for a 15-cluster group is sufficient to achieve the benchmark output for both the LID and non-LID case. The benchmark output can be approached even more closely for the LID case if the same method is applied to the 20-cluster group.

Therefore if data for a representative lot from each UHRU can be obtained, then it is not necessary to run all lots simultaneously using hydrological modelling software. It is not difficult

or time-consuming to perform thousands of simulations given the continuous improvements in software. However, it is not always easy to obtain all of the necessary data. The task also becomes more onerous after LID implementation because detailed information such as the placement of the LID for each lot may vary, which increases the level of difficulty of running repetitive simulations. It is recommended that the entire experiment be repeated again while using a more realistic data set that contains fewer variables.

The calculation of total runoff (m^3) using the Random Sampling Method suggests that by taking a random sample of 7 % of the lots from each cluster within a 3-cluster group, the benchmark output can be achieved in the non-LID case. The same result was not found for the LID case. In the LID case, a large percentage of randomly selected lots must be selected (90 %) in order to closely approximate the benchmark output.

A lack of data is a possible explanation for the unsatisfactory results of the LID case since more data was required to perform linear regression effectively. Many linear regression models could not be developed for the LID case due to an insufficient amount of data. Linear regression generally requires a certain amount of data to assure that it can be carried out. The approach proposed in this research may therefore be limited to data sets of at least 7,500 lots or more for both the no LID case and the LID case.

Thus, more work is required to confirm the results for the LID case since it was expected that the final results should be more similar to the non-LID case. LIDs represent a transformation process which implies that the total runoff (m^3) from running lot simulations should stabilize as the number of lot simulations increases. Therefore, it should not be necessary to run all of the lot simulations in order to determine the final total output for all of the lots under study.

However when random percentage samples were taken from all of the lots for both the LID and non-LID cases, the process was almost equivalent to running all of the lots using the VBA program. This suggests that there is still some value to the clustering approach to form UHRUs since the results from taking random percentage samples from each cluster can be advantageous when compared with performing hydrological modelling on all of the lots.

The rain barrel LID investigated is a roof-based LID; therefore a pavement-based LID such as bioretention should be investigated as well. Since width was considered to be an important parameter, a more accurate way of determining the width of each lot should be developed. For example, a macro could be developed to somehow determine width as defined according to SWMM software. Other macros can be developed to automate some of the processes using Excel spreadsheets, such as with the Minimum Distance Method. This would be particularly useful for working with larger groups of clusters. Thus, there are various aspects of this research thesis which can be further explored.

Appendices

A.1 VBA Code Without LID Implementation

```vba
Sub SWMM_FILE_GENERATION()


For Each Worksheet In Worksheets
Dim SWMM As String
Dim i As Integer
Dim NoOfLots As Integer
Dim SheetName As String
NoOfLots = Worksheet.Range("A2", Worksheet.Range("A2").End(xlDown)).Rows.Count
SheetName = Worksheet.Name


  SWMM = "[TITLE]"
  SWMM = SWMM & vbCrLf
  SWMM = SWMM & vbCrLf & "[OPTIONS]"
  SWMM = SWMM & vbCrLf & "FLOW_UNITS          LPS"
  SWMM = SWMM & vbCrLf & "INFILTRATION        GREEN_AMPT"
  SWMM = SWMM & vbCrLf & "FLOW_ROUTING        KINWAVE"
  SWMM = SWMM & vbCrLf & "START_DATE          04/01/1985"
  SWMM = SWMM & vbCrLf & "START_TIME          00:00:00"
  SWMM = SWMM & vbCrLf & "REPORT_START_DATE   04/01/1985"
  SWMM = SWMM & vbCrLf & "REPORT_START_TIME   00:00:00"
  SWMM = SWMM & vbCrLf & "END_DATE            11/01/1985"
  SWMM = SWMM & vbCrLf & "END_TIME            00:00:00"
  SWMM = SWMM & vbCrLf & "SWEEP_START         01/01"
  SWMM = SWMM & vbCrLf & "SWEEP_END           12/31"
  SWMM = SWMM & vbCrLf & "DRY_DAYS            0"
  SWMM = SWMM & vbCrLf & "REPORT_STEP         01:00:00"
  SWMM = SWMM & vbCrLf & "WET_STEP            00:15:00"
  SWMM = SWMM & vbCrLf & "DRY_STEP            01:00:00"
```

```
SWMM = SWMM & vbCrLf & "ROUTING_STEP        0:01:00 "

SWMM = SWMM & vbCrLf & "ALLOW_PONDING       NO"

SWMM = SWMM & vbCrLf & "INERTIAL_DAMPING    PARTIAL"

SWMM = SWMM & vbCrLf & "VARIABLE_STEP       0.75"

SWMM = SWMM & vbCrLf & "LENGTHENING_STEP    0"

SWMM = SWMM & vbCrLf & "MIN_SURFAREA        0"

SWMM = SWMM & vbCrLf & "NORMAL_FLOW_LIMITED  BOTH"

SWMM = SWMM & vbCrLf & "SKIP_STEADY_STATE   NO"

SWMM = SWMM & vbCrLf & "FORCE_MAIN_EQUATION  H-W"

SWMM = SWMM & vbCrLf & "LINK_OFFSETS        DEPTH"

SWMM = SWMM & vbCrLf & "MIN_SLOPE           0"

SWMM = SWMM & vbCrLf

SWMM = SWMM & vbCrLf & "[EVAPORATION]"

SWMM = SWMM & vbCrLf & ";;Type      Parameters"

SWMM = SWMM & vbCrLf & ";;---------- ----------"

SWMM = SWMM & vbCrLf & "MONTHLY    0    0    0.01864 0.9428 2.519 3.766 4.129 3.684
2.580 1.223 0.3149 0  "

SWMM = SWMM & vbCrLf & "DRY_ONLY    NO"

SWMM = SWMM & vbCrLf

SWMM = SWMM & vbCrLf & "[RAINGAGES]"

SWMM = SWMM & vbCrLf & ";;           Rain    Time Snow  Data "

SWMM = SWMM & vbCrLf & ";;Name       Type    Intrvl Catch  Source "

SWMM = SWMM & vbCrLf & ";;-------------- --------- ------ ------ ----------"

SWMM = SWMM & vbCrLf & "Gage1        INTENSITY 1:00  1.0   TIMESERIES Rain"

SWMM = SWMM & vbCrLf

SWMM = SWMM & vbCrLf & "[SUBCATCHMENTS]"

SWMM = SWMM & vbCrLf & ";;                          Total  Pcnt.    Pcnt.  Curb
Snow "

SWMM = SWMM & vbCrLf & ";;Name        Raingage    Outlet      Area   Imperv Width
Slope   Length  Pack"

SWMM = SWMM & vbCrLf & ";;-------------- ---------------- ---------------- -------- -------- -------- -------
- -------- --------"

For i = 2 To (NoOfLots + 1)
```

SWMM = SWMM & vbCrLf & Worksheet.Cells(i, 1) & "   Gage1   Out1   " & Round(Worksheet.Cells(i, 4).Value, 3) & _

   "   " & Round(Worksheet.Cells(i, 6).Value, 3) & "   " & Round(Worksheet.Cells(i, 7).Value, 3) & "   " & Worksheet.Cells(i, 8).Value & "   0   "

  Next

  SWMM = SWMM & vbCrLf

  SWMM = SWMM & vbCrLf & "[SUBAREAS]"

  SWMM = SWMM & vbCrLf & ";;Subcatchment N-Imperv N-Perv   S-Imperv S-Perv   PctZero RouteTo   PctRouted "

  SWMM = SWMM & vbCrLf & ";;-------------- ---------- ---------- ---------- ---------- ---------- ---------- ----------"

  For i = 2 To (NoOfLots + 1)

    SWMM = SWMM & vbCrLf & Worksheet.Cells(i, 1) & "   0.015   .24   1.524   7.62   25   OUTLET   "

  Next i

  SWMM = SWMM & vbCrLf

  SWMM = SWMM & vbCrLf & "[INFILTRATION]"

  SWMM = SWMM & vbCrLf & ";;Subcatchment Suction   HydCon   IMDmax   "

  SWMM = SWMM & vbCrLf & ";;-------------- ---------- ---------- ----------"

  For i = 2 To (NoOfLots + 1)

    SWMM = SWMM & vbCrLf & Worksheet.Cells(i, 1) & "   " & Worksheet.Cells(i, 12).Value & "   " & Worksheet.Cells(i, 11).Value & "   " & Worksheet.Cells(i, 10).Value

  Next i

  SWMM = SWMM & vbCrLf

  SWMM = SWMM & vbCrLf & "[LID_CONTROLS]"

  SWMM = SWMM & vbCrLf & ";;   Type/Layer Parameters"

  SWMM = SWMM & vbCrLf & ";;-------------- ---------- ----------"

  SWMM = SWMM & vbCrLf & "RH-CGI   RB   "

  SWMM = SWMM & vbCrLf & "RH-CGI   STORAGE 2000   0.75   10   0 "

  SWMM = SWMM & vbCrLf & "RH-CGI   DRAIN   25.4   0.5   0   6   "

  SWMM = SWMM & vbCrLf

  SWMM = SWMM & vbCrLf & "RH-Res   RB   "

  SWMM = SWMM & vbCrLf & "RH-Res   STORAGE 1300   0.75   10   0"

SWMM = SWMM & vbCrLf & "RH-Res   DRAIN   25.4   0.5   0   6   "

SWMM = SWMM & vbCrLf

SWMM = SWMM & vbCrLf & "GR   BC   "

SWMM = SWMM & vbCrLf & "GR   SURFACE 0   0   0.1   2   5   "

SWMM = SWMM & vbCrLf & "GR   SOIL   95   0.25   0.2   0.1   12.7   10.0
88.9   "

SWMM = SWMM & vbCrLf & "GR   STORAGE 300   0.75   0   0   "

SWMM = SWMM & vbCrLf & "GR   DRAIN   25.4   0.5   0   6   "

SWMM = SWMM & vbCrLf

SWMM = SWMM & vbCrLf & "DD   RB   "

SWMM = SWMM & vbCrLf & "DD   STORAGE 1   0.75   10   0 "

SWMM = SWMM & vbCrLf & "DD   DRAIN   75   0   0   0 "

SWMM = SWMM & vbCrLf

SWMM = SWMM & vbCrLf & "DW   IT   "

SWMM = SWMM & vbCrLf & "DW   SURFACE 0   0   0.06   2   5   "

SWMM = SWMM & vbCrLf & "DW   STORAGE 914.4   0.4   254   0   "

SWMM = SWMM & vbCrLf & "DW   DRAIN   0   0.5   0   6   "

SWMM = SWMM & vbCrLf

SWMM = SWMM & vbCrLf & "SP   IT   "

SWMM = SWMM & vbCrLf & "SP   SURFACE 0   0   0.024   5   5   "

SWMM = SWMM & vbCrLf & "SP   STORAGE 1500   0.4   254   0   "

SWMM = SWMM & vbCrLf & "SP   DRAIN   0   0.5   0   6   "

SWMM = SWMM & vbCrLf

SWMM = SWMM & vbCrLf & "PP   PP   "

SWMM = SWMM & vbCrLf & "PP   SURFACE 0   0   0.024   2   5   "

SWMM = SWMM & vbCrLf & "PP   PAVEMENT 125   0.165   0   2540   0   "

SWMM = SWMM & vbCrLf & "PP   STORAGE 300   0.4   254   0   "

SWMM = SWMM & vbCrLf & "PP   DRAIN   0   0.5   0   6   "

SWMM = SWMM & vbCrLf

SWMM = SWMM & vbCrLf & "BR   BC   "

SWMM = SWMM & vbCrLf & "BR   SURFACE 250   0   0.24   1   5   "

SWMM = SWMM & vbCrLf & "BR   SOIL   450   0.44   0.08   0.04   75   10   55.88   "

SWMM = SWMM & vbCrLf & "BR   STORAGE   300   0.625   254   0   "

SWMM = SWMM & vbCrLf & "BR   DRAIN   0   1   12.7   6   "

SWMM = SWMM & vbCrLf

SWMM = SWMM & vbCrLf & "[LID_USAGE]"

SWMM = SWMM & vbCrLf & ";;Subcatchment   LID Process   Number   Area   Width   InitSatur   FromImprv   ToPerv   Report File"

SWMM = SWMM & vbCrLf & ";;-------------- ---------------- ------- ---------- ---------- ---------- ---------- ---------- -----------"

SWMM = SWMM & vbCrLf

SWMM = SWMM & vbCrLf & "[OUTFALLS]"

SWMM = SWMM & vbCrLf & ";;   Invert   Outfall   Stage/Table   Tide"

SWMM = SWMM & vbCrLf & ";;Name   Elev.   Type   Time Series   Gate"

SWMM = SWMM & vbCrLf & ";;-------------- ---------- ---------- ---------------- ----"

SWMM = SWMM & vbCrLf & "Out1   0   FREE   NO"

SWMM = SWMM & vbCrLf

SWMM = SWMM & vbCrLf & "[TIMESERIES]"

SWMM = SWMM & vbCrLf & ";;Name   Date   Time   Value"

SWMM = SWMM & vbCrLf & ";;-------------   ---------- ---------- ----------"

SWMM = SWMM & vbCrLf & "Rain   FILE   " & Chr(34) & "C:\Temp1\BarrieWPCC-1985met.dat" & Chr(34)

SWMM = SWMM & vbCrLf & "[REPORT]"

SWMM = SWMM & vbCrLf & "INPUT   NO"

SWMM = SWMM & vbCrLf & "CONTROLS   NO"

SWMM = SWMM & vbCrLf

SWMM = SWMM & vbCrLf & "[TAGS]"

SWMM = SWMM & vbCrLf

SWMM = SWMM & vbCrLf & "[MAP]"

SWMM = SWMM & vbCrLf & "DIMENSIONS 0.000 0.000 10000.000 10000.000"

SWMM = SWMM & vbCrLf & "Units   None"

```
SWMM = SWMM & vbCrLf

SWMM = SWMM & vbCrLf & "[COORDINATES]"

SWMM = SWMM & vbCrLf & ";;Node           X-Coord          Y-Coord          "

SWMM = SWMM & vbCrLf & ";;-------------- ----------------- -----------------"

SWMM = SWMM & vbCrLf & "Out1          1018.676       7640.068        "

SWMM = SWMM & vbCrLf

SWMM = SWMM & vbCrLf & "[VERTICES]"

SWMM = SWMM & vbCrLf & ";;Link           X-Coord          Y-Coord          "

SWMM = SWMM & vbCrLf & ";;-------------- ----------------- -----------------"

SWMM = SWMM & vbCrLf

SWMM = SWMM & vbCrLf & "[Polygons]"

SWMM = SWMM & vbCrLf & ";;Subcatchment   X-Coord          Y-Coord   "

SWMM = SWMM & vbCrLf & ";;-------------- ----------------- -----------------"

For i = 2 To (NoOfLots + 1)

    SWMM = SWMM & vbCrLf & Worksheet.Cells(i, 1) & "           560.272        8132.428        "

    SWMM = SWMM & vbCrLf & Worksheet.Cells(i, 1) & "           -271.647        8132.428        "

    SWMM = SWMM & vbCrLf & Worksheet.Cells(i, 1) & "           -271.647        8998.302       "

    SWMM = SWMM & vbCrLf & Worksheet.Cells(i, 1) & "           560.272        8998.302        "

Next i

SWMM = SWMM & vbCrLf

SWMM = SWMM & vbCrLf & "[SYMBOLS]"

SWMM = SWMM & vbCrLf & ";;Gage          X-Coord          Y-Coord        "

SWMM = SWMM & vbCrLf & ";;-------------- ----------------- -----------------"

SWMM = SWMM & vbCrLf & "Gage1          -865.874       9388.795        "


'Debug.Print SWMM
'declare and initiate required objects
Dim fs As FileSystemObject
Dim ts As TextStream
Set fs = New FileSystemObject
```

```vba
'To write

Dim BasicFileName As String

BasicFileName = "C:\Temp1\" & SheetName


'" & Left(Sheets(1).Cells(i, 2), 4) & "_" & Sheets(1).Cells(i, 1)


Set ts = fs.OpenTextFile(BasicFileName & ".inp", ForWriting, True)

ts.WriteLine SWMM

ts.Close


'clear memory used by FSO objects

Set ts = Nothing

Set fs = Nothing


SWMM = ""


'a = RunSwmmDll("C:\Temp1\Comm_1.inp", "c:\Temp1\aaa.rpt", "c:\Temp1\a.out")

A = RunSwmmDll(BasicFileName & ".inp", BasicFileName & ".rpt", BasicFileName & ".out")



MsgBox "Complete " & SheetName


Next


End Sub
```

A.2 Modified Section of VBA Code for RH LID Implementation

```
SWMM = SWMM & vbCrLf

   SWMM = SWMM & vbCrLf & "[LID_USAGE]"

   SWMM = SWMM & vbCrLf & ";;Subcatchment   LID Process     Number  Area      Width
InitSatur  FromImprv  ToPerv     Report File"

   SWMM = SWMM & vbCrLf & ";;-------------- ---------------- ------- ---------- ---------- ---------- ----------
---------- -----------"

  For i = 2 To (NoOfLots + 1)

     SWMM = SWMM & vbCrLf & Round(Worksheet.Cells(i, 1).Value, 3) & "    " & Worksheet.Cells(i,
13).Value & "   1   " & Round(Worksheet.Cells(i, 14).Value, 3) & "    " & Round(Worksheet.Cells(i,
15).Value, 3) & _

     "    " & Round(Worksheet.Cells(i, 16).Value, 3) & "    " & Round(Worksheet.Cells(i, 17).Value, 3)
& "    " & Round(Worksheet.Cells(i, 18).Value, 3)


  Next
```
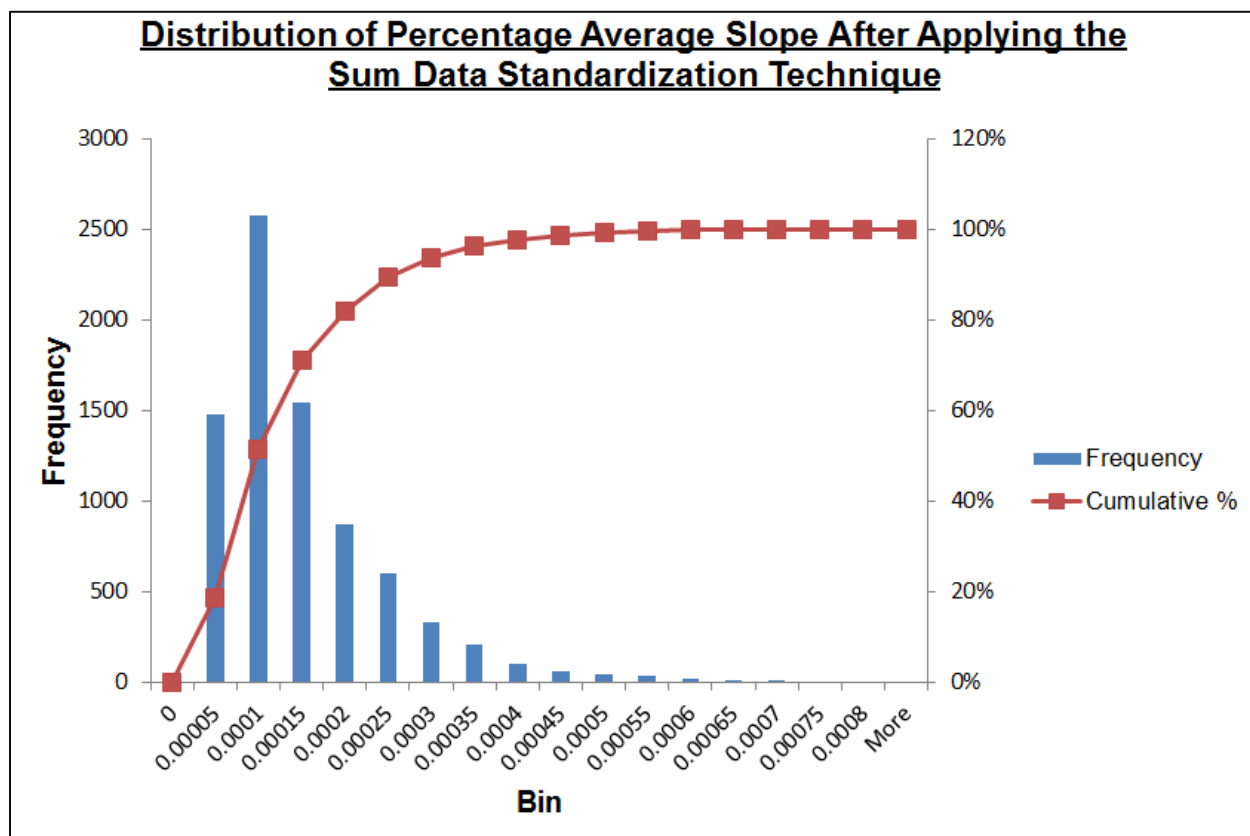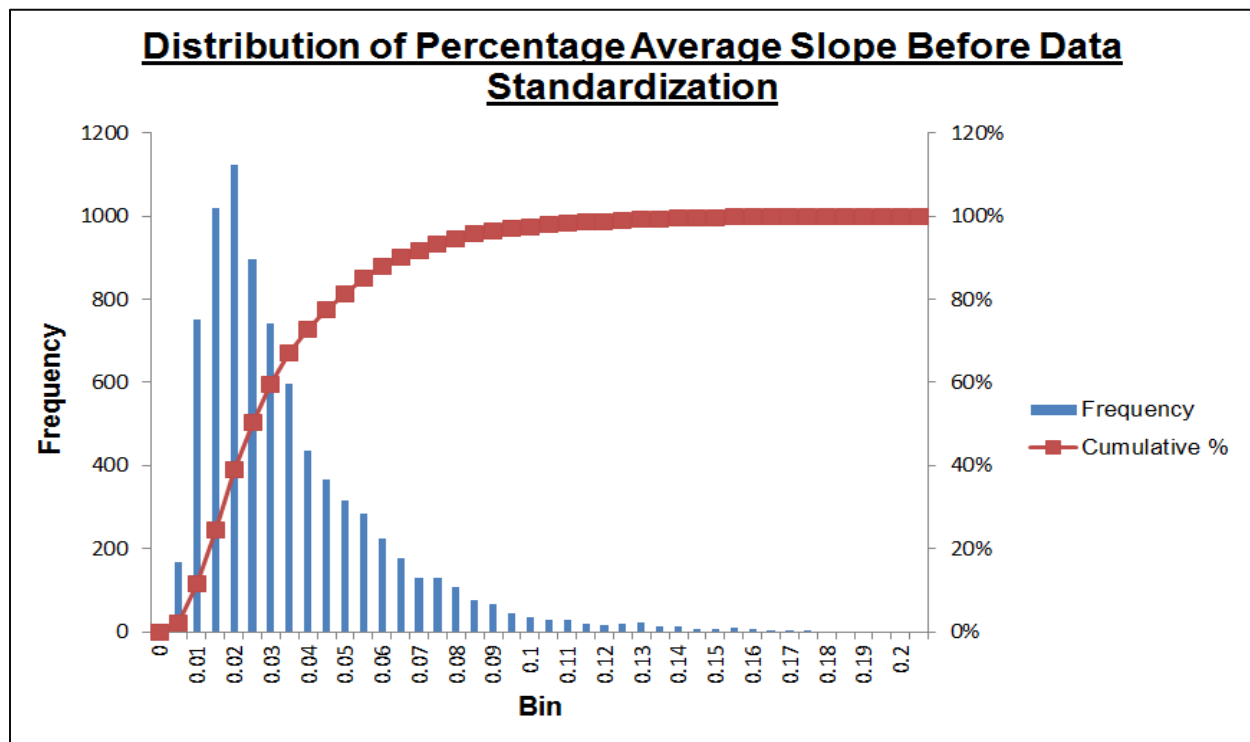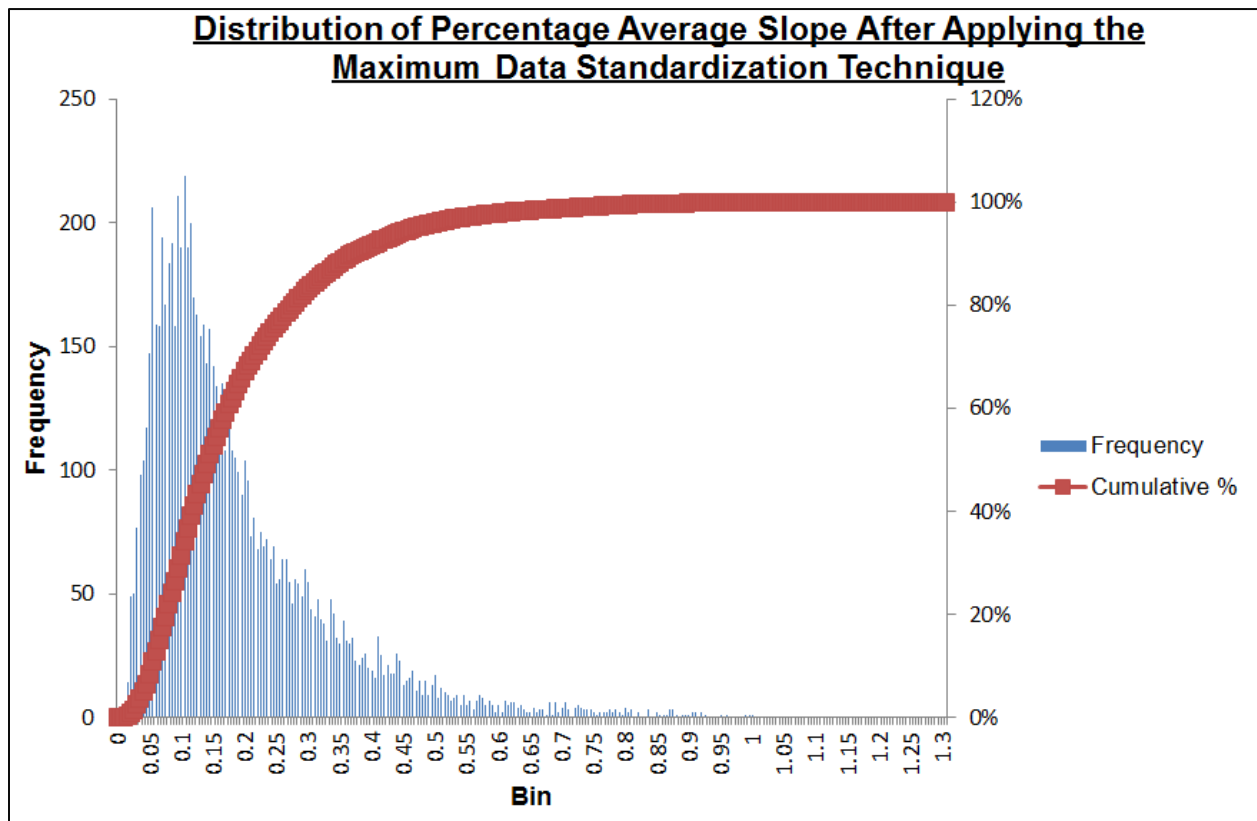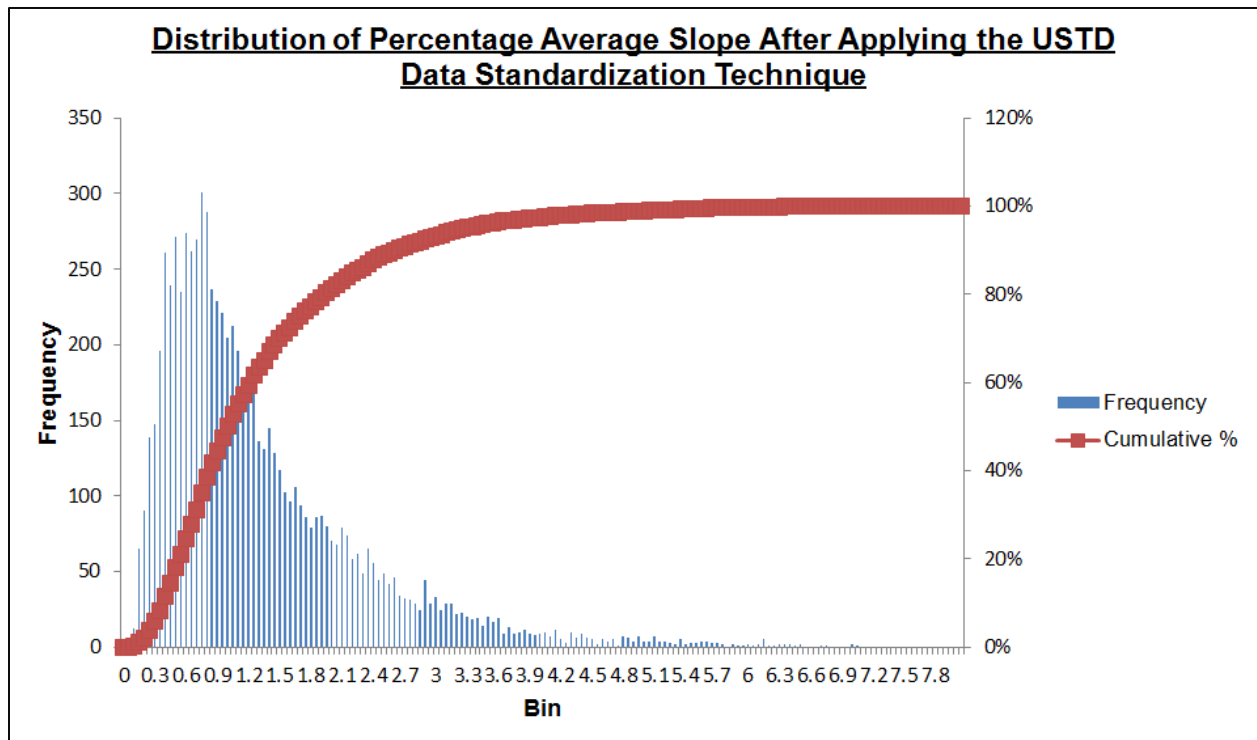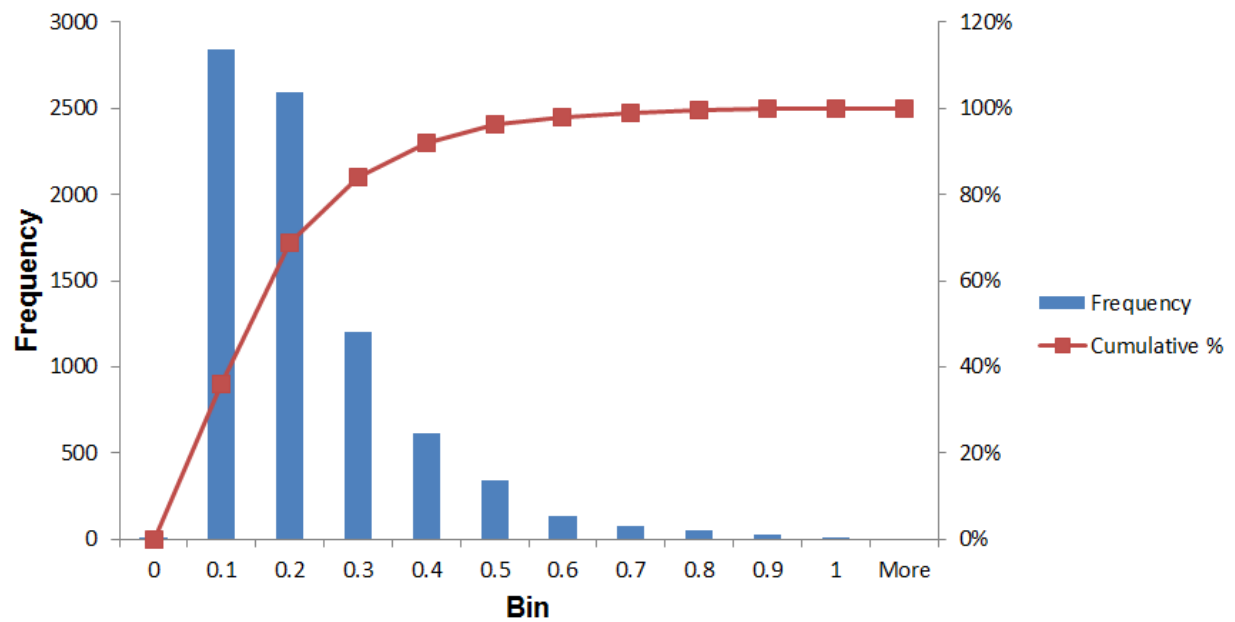
A.3 Distribution of Percentage Average Slope

Distribution of Percentage Average Slope After Applying the USTD Data Standardization Technique

Distribution of Percentage Average Slope After Applying the Maximum Data Standardization Technique
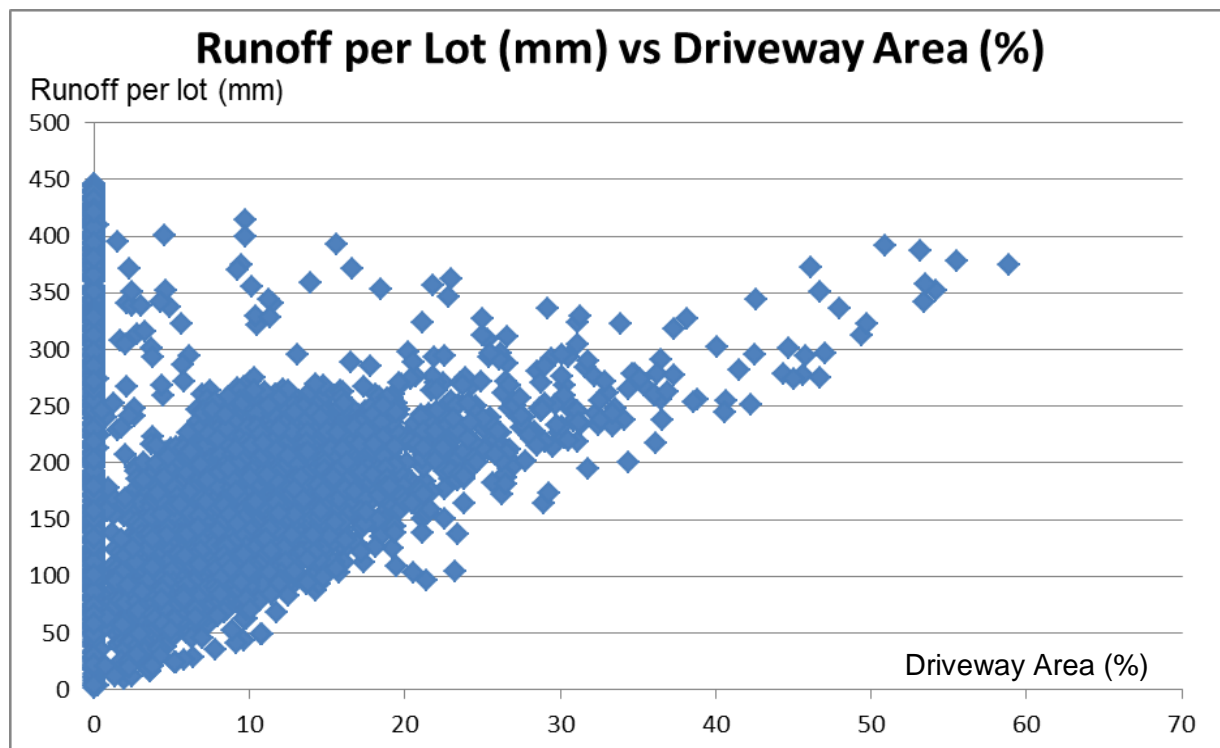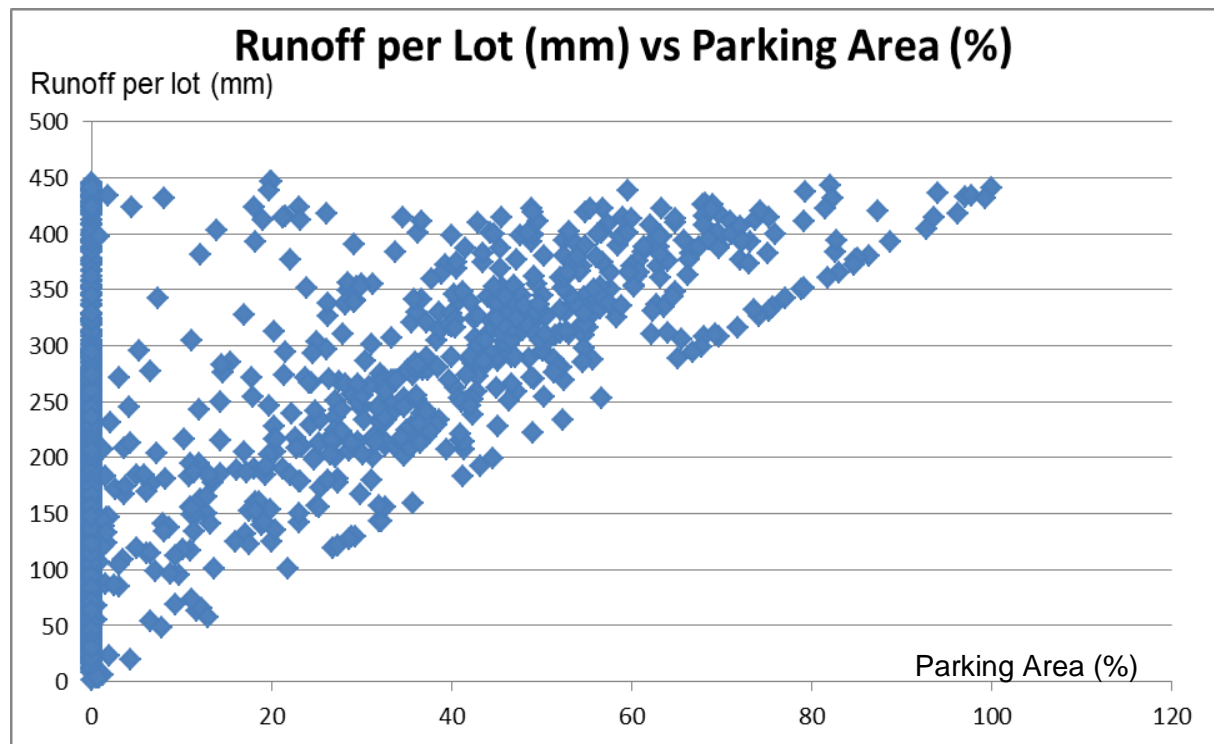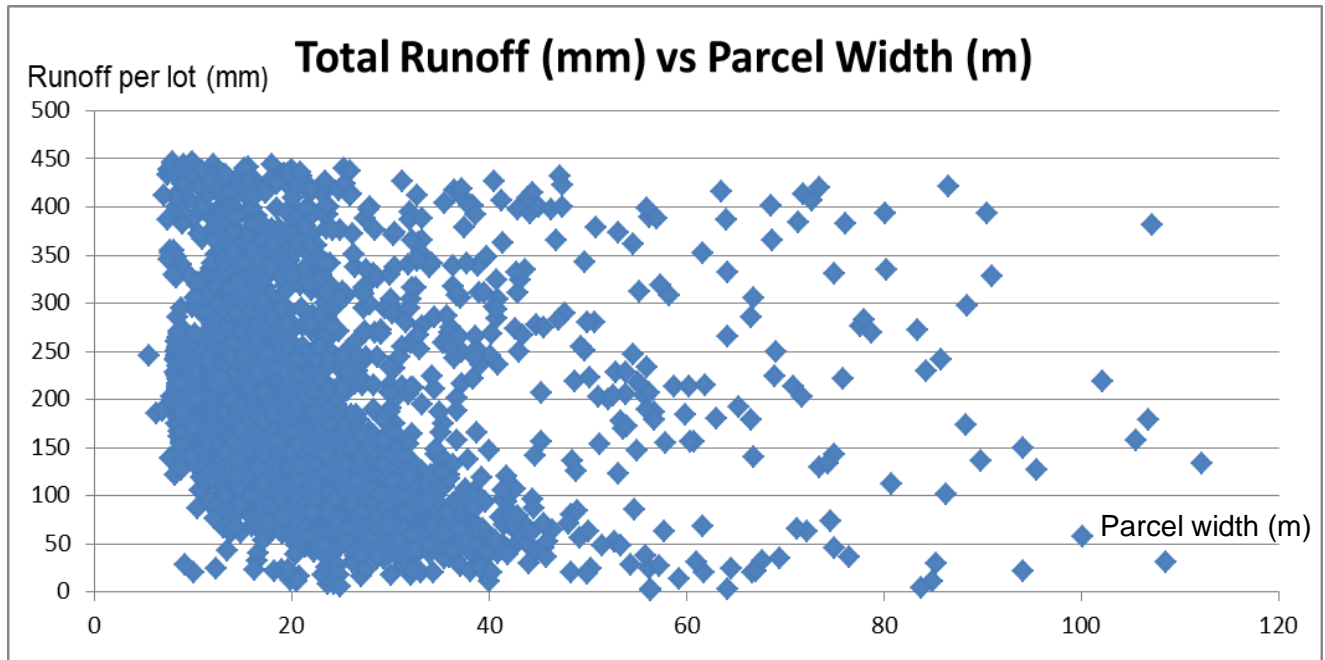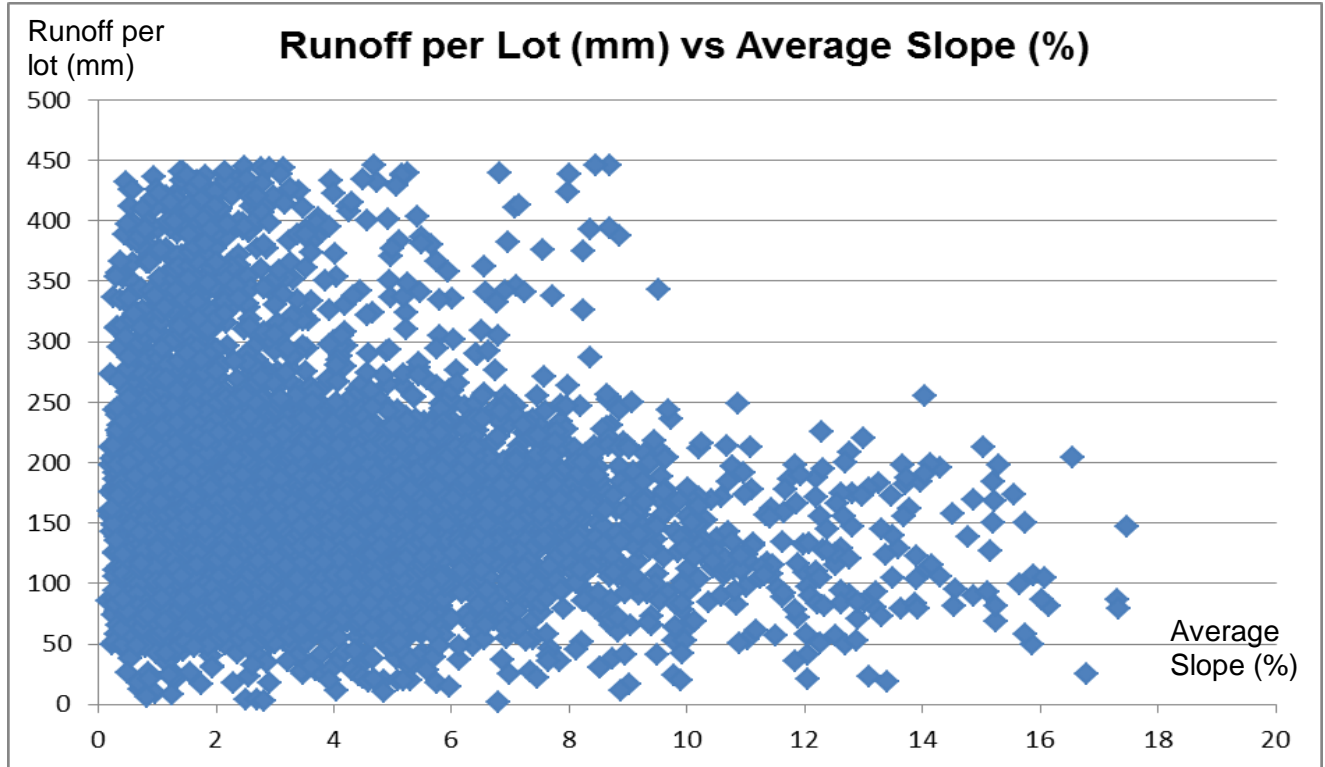
Distribution of Percentage Average Slope After Applying the Range 2 Data Standardization Technique

A.4 Relationships between Total Runoff (mm) and Each Input Variable

Runoff per Lot (mm) vs Parking Area (%)

Runoff per Lot (mm) vs Driveway Area (%)

Runoff per Lot (mm) vs Average Slope (%)

Runoff per lot (mm)

Average Slope (%)

Total Runoff (mm) vs Parcel Width (m)

Runoff per lot (mm)

Parcel width (m)

References

Acomb, G., and Clark, M. (2008). *Florida Field Guide to Low Impact Development*. University of Florida – Program for Resource Efficient Communities, 1-4.

Alfredo, K., Arye, M., Behr, C., Montalto, F., Walsh, M., and Wolf, M. (2007). "Rapid assessment of the cost-effectiveness of low impact development for CSO control" [Electronic version]. *Landscape and Urban Planning,* 82, 117-131.

Andrieu, H., Morena, F., and Rodriguez, F. (2008). "A distributed hydrological model for urbanized areas – Model Development and application to case studies." *Journal of Hydrology*, 351, 268-287.

Carmon, N., and Shamir, U. (2010). "Water-sensitive planning: integrating water considerations into urban and regional planning." *Water and Environment Journal*, 24, 181-191.

Carrillo, G., Sawicz, K., Sivapalan, M., Troch, P.A., and Wagener, T. (2011). "Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA." *Hydrology and Earth System Sciences Discussions*, 8, 4495-4534.

Cooper, M.C., and Milligan, G.W. (1988). "A Study of Standardization of Variables in Cluster Analysis." *Journal of Classification*, 5, 181-204.

Cowan, D.M., Easton, Z.M., Fuka, D.R., Schneiderman, E.M., Steenhuis, T.S., and Walter, M.T. (2008). "Re-conceptualizing the soil and water assessment tool (SWAT) model to predict runoff from variable source areas." *Journal of Hydrology*, 348, 279-291.

Davis, J., Gironas, J., and Roesner, L.A. (2009). "Stormwater Management Model Applications Manual." United States Environmental Protection Agency, Fort Collins, Colorado.

Dietz, M.E. (2007). "Low Impact Development Practices: A Review of Current Research and Recommendations for Future Directions" [Electronic Version]. *Water Air Soil Pollution,* 186, 351-363.

Elliott, A.H., and Trowsdale, S.A. (2007). "A review of models for low impact urban stormwater drainage" [Electronic version]. *Environmental Modelling & Software*, 22, 394-405.

Elliott, A.H., Trowsdale, S.A., and Wadhwa, S. (2009). "Effect of Aggregation of On-Site Storm-Water Control Devices in an Urban Catchment Model." *Journal of Hydrologic Engineering,* 14(9), 975-983.

Gan, G., Ma, C. and Wu, J. (2007). *Data Clustering Theory, Algorithms, and Applications.* Society for Industrial and Applied Mathematics, Philadelphia, PA.

Gilroy, K.L., and McCuen, R.H. (2009). "Spatio-temporal effects of low impact development Practices." *Journal of Hydrology*, 367, 228-236.

International Business Machines [IBM]. (2011). "K-Means Cluster Analysis." <www.ibm.com> (2011-2012).

International Business Machines [IBM]. (2011). "Linear Regression." <www.ibm.com> (2011-2012).

James, W.P., and Wurbs, R.A. (2002). *Water Resources Engineering*. Prentice Hall, New Jersey.

Kessler, A., Kronaveter, L., and Shamir, U. (2001). "Water-Sensitive Urban Planning: Modeling On-Site Infiltration". *Journal of Water Resources Planning and Management,* 127(2), 78-88.

Li, J., Banting, D., Joksimovic, D., Eric, M., Fan, C., Hahn, K., Lawson, S., and Mirzajani, M. (2010). "Evaluation of Low Impact Development Stormwater Technologies for the Uncontrolled Urban Areas in the Lake Simcoe Regions Final Report." *Report*, Lake Simcoe Region Conservation Authority, Ontario, Canada.

Rossman, L.A. (2010). "Storm Water Management User's Manual Version 5.0." United States Environmental Protection Agency, Cincinnati, Ohio.

The City of Barrie. (2012). *The City of Barrie 2012 Accessibility Plan*. City of Barrie, Ontario, Canada.