

Performance Evaluation of Low Impact Development Practices Using Linear Regression

Marija Eric^{1*}, James Li¹ and Darko Joksimovic¹

¹Department of Civil Engineering, Ryerson University, 350 Victoria Street, Toronto, Ontario, M5B 2K3, Canada.

Authors' contributions

This work was carried out in collaboration between all authors. Author ME designed the study, performed the statistical analysis, wrote the protocol, and wrote the manuscript. Authors JL and DJ managed the analyses of the study and contributed to the design of the study. Author DJ provided guidance and support for technical issues. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BJECC/2015/11578

Original Research Article

Received 3rd April 2014
Accepted 21st June 2014
Published 5th June 2015

ABSTRACT

Aims: To develop a modelling methodology for evaluating the cumulative stormwater performance of Low Impact Development technologies on a watershed basis to address stormwater impacts of urban development.

Study Design: A method is presented to perform hydrological modelling on large watersheds. Hydrological modelling simulations and linear regression analyses of a small sample of randomly selected lots were performed to generate results which were extrapolated to the entire watershed.

Place and Duration of Study: Department of Civil Engineering, Ryerson University, between September 2010 and September 2012.

Methodology: Urban hydrological response units were developed by using the K-means cluster analysis procedure to group 6926 lot parcels amenable to the residential rain barrel Low Impact Development practice into clusters. Two versions of a Microsoft Excel macro were developed to run simulations for thousands of lots simultaneously before and after Low Impact Development implementation to determine the total runoff produced by all lots for both cases. The results of computer modelling all lots were compared with the results from developing calculation methods to be used after computer modelling subsets of lots. Two calculation methods based on clustering lots to form urban hydrological response units were developed.

A random sample of 5 % of all lots was then extracted from 6616 lots amenable to the porous pavement Low Impact Development. Stepwise linear regression and linear regression were performed on the random sample for each case of no Low Impact Development and with Low Impact Development. Regression equations were used to extrapolate results from the sample to

*Corresponding author: Email: meric@ryerson.ca

the entire data set to determine the total runoff volume produced by each set of lots.

Results: Results from the cluster-based calculation methods developed as applied to residential rain barrels were unsatisfactory since they did not approximate the output values from modelling all lots using software. The alternative method applied to porous pavement Low Impact Development implementation, entailing stepwise linear regression and linear regression, produced 945,382.97 m³ and 747,380.13 m³ of total runoff respectively. These values closely approximated corresponding values generated by the modelling software of 937,088.58 m³ and 746,462.40 m³.

Conclusion: The formation of urban hydrological response units may be unnecessary for hydrological modelling Low Impact Development technologies for large watersheds. Hydrological characteristics for only a small, randomly selected subset of all lots can be used to determine total runoff volume produced by all lots in the watershed before and after Low Impact Development implementation.

Keywords: Low impact development (LID); stormwater management; urban hydrologic response unit (UHRU); hydrologic simulations; linear regression.

1. INTRODUCTION

Stormwater or rainwater runoff is now viewed as a resource instead of as just a nuisance. Runoff is generated by impervious surfaces associated with land development and urbanization such as concrete, that prevent rainfall or water from infiltrating into the ground [1,2]. Numerous environmental effects result by preventing rainfall from infiltrating such as decreases in groundwater recharge volume, base flows and time of concentration as well as lower water tables. Other effects include increased flood flows, stream erosion and water contamination [2,3,4]. The deterioration of urban streams with respect to habitat and water quality has been recorded throughout the world [1,2,4].

The aim of conventional stormwater management techniques is to remove water from a site as fast as possible and store it as a larger volume at an off-site, downstream facility such as a detention pond or an infiltration basin. Conventional techniques may have controlled peak discharge rates but have not improved issues with respect to increased runoff volume [1,2]. Low Impact Development practices (LID) are devices or techniques that mitigate stormwater impacts of urban development. Most LIDs are lot-based practices including bioretention cells, greenroofs, rainfall harvesting, porous pavement, dry wells, and grass swales. The United States Environmental Protection Agency (US EPA) has recognized LID as a leading planning approach for runoff management [5].

Modelling of the hydrologic performance of LID requires detailed accounting of hydrologic components at the lot level which may require

extensive resources when modelling thousands of lots for large watersheds. A significant amount of lot-level detail is lost, however, when modelling an entire area for LID implementation. An innovative approach based on a conceptual model must be devised to evaluate the entire region [2]. A common approach for simulating the cumulative performance of LID for large watersheds is to develop urban hydrologic response units (UHRUs or urban HRUs) and extrapolate the results of the performance of the UHRUs to all lots.

UHRUs are urban drainage areas or lots that exhibit similar runoff generating mechanisms as a result of similar hydrological characteristics such as level of imperviousness, slope, lot area, and soil type [6]. By applying the UHRU concept, large study areas with similar hydrological characteristics can be evaluated efficiently. A systematic approach is required to group drainage areas or lots into UHRUs based on common hydrological characteristics. Appropriate hydrologic models such as the US EPA Stormwater Management Model (SWMM) can be used to simulate runoff volume from various UHRUs.

Although the UHRU approach enhances the efficiency of modelling large watersheds, it requires large data sets as well as lengthy algorithms for preparing the data to disaggregate the study area into units or lot clusters sharing similar hydrological characteristics. Many simulations must be performed using hydrological modelling software to generate runoff volume from each UHRU before extrapolating it to the entire watershed. A small computer program can be developed to work in tandem with hydrological modelling software in

order to perform hydrological modelling for thousands of lots amenable to LID implementation concurrently. The total runoff volume of all lots can then be summed directly from the modelled output runoff volume of each lot both before and after LID implementation.

If a linear regression is performed on a random selection of 5% of all lots, the regression equation for the random sample can be used to extrapolate the runoff volume (m^3) generated by all lots to find the total runoff volume for the entire watershed. By running a regression and extrapolating the results, UHRUs or lot clusters do not have to be formed and so the UHRU concept does not have to be applied.

A random sample of 5% of all lots was first used to determine whether a relatively small sample of lots could effectively replace hydrological modelling each and every lot. More work is needed to determine whether other small samples are also effective and whether there is a certain percentage of lots that would allow the method to be applicable to all types of LIDs.

The concept of hydrologic response unit is not new and has been applied in hydrologic studies before [6,7,8,9]. According to the United States Department of Agriculture (1972), the traditional definition of HRUs is based on soil type and land use [7,10]. Hydrological simulations require meteorological input data including temperature, precipitation and solar radiation [7]. Unconventional hydrological attributes such as average slope and percentage imperviousness were used to delineate HRUs in the previous case study by Li et al. [6] as well as in this study. HRU is based on a linear assumption and is appropriate for runoff volume simulation.

A study by England and Stephenson [11] presented a technique for isolating relatively homogeneous areas or units of rangeland watersheds based on soil properties, geologic, climatic, and topographic features. The HRUs that were formed may be used in computations of watershed performance, as experimental units in field studies, and as units for the application of conservation management practices. Prior to this study, England and Holtan [12] determined that HRUs were internally homogeneous enough to be used as computational units in mathematical models simulating the hydrologic performance of agricultural watersheds.

Although the standard version of the Soil and Water Assessment Tool (SWAT) uses the traditional method for defining the area of each HRU, HRUs provided the conceptual framework in a study investigating the SWAT model. A modified version of the SWAT model divided sub-basins into HRUs by intersecting soil topographic index (STI) and land use rasters or Geographic Information Systems (GIS) shapefiles [7]. Different GIS layers were intersected with one another in the current study and the distributions of hydrological attributes over the study area were determined.

In a study by Kessler et al. [13], a micro-model was developed and applied at three different spatial scales to a developing urban neighbourhood to investigate the effects of spatial resolution on the hydrological modelling process. The model was also used to investigate the effects of urban development on infiltration and runoff. The intermediate spatial scale consisted of a cluster of residential lots and their immediate vicinity. Each cluster was a separate subcatchment which had a similar composition in terms of land use and can be regarded as a UHRU [13]. In the current work, the K-means cluster analysis procedure was used to form clusters of urban residential lots that can also be regarded as UHRUs.

The study by Kessler et al. [13] reached a similar conclusion to the study by Li et al. [6] which provides the foundation for the current work. The conclusion is 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 [6,13]. However, a more detailed model was deemed to be generally necessary for carrying out simulations of hydrological processes at the lot scale [13].

In another study, a detailed stormwater model was developed for a suburban catchment by taking each individual property and road section between catch basins as a separate source area containing a hypothetical stormwater device. The catchment was subdivided into different levels of aggregation from 810 source areas to the extreme scenario of a single source area and a single aggregated device. The effects of performing hydrological modelling for each level of aggregation were investigated by examining key summary measures such as flow and water quality. Based upon certain conditions and assumptions, it was concluded that the

aggregation of on-site devices and associated source areas did not significantly impact key summary measures under investigation [9]. The findings of the study support the final method proposed in the current study of extrapolating hydrological modelling results from individual lot parcels to an entire watershed.

The findings also support the method developed in the case study by Li et al. [6] of extrapolating hydrological modelling results of lot-based LID practices to a watershed and neighbouring municipalities using a spreadsheet model. The current research is an extension of this previous case study. First, an attempt was made to refine the development of UHRUs as carried out in the case study and then to find a more efficient way of determining the hydrological response of a watershed to LID implementation.

The focus of both studies was on the performance of LID technology systems on

uncontrolled areas, pre-defined areas to which conventional stormwater practices have never been applied and where no opportunities for implementing stormwater management ponds exist [6]. The study area for the current research entails residential lots amenable to LID implementation in uncontrolled areas within the City of Barrie, a city situated in the Lake Simcoe Watershed in southern Ontario, Canada. Uncontrolled areas within the entire Lake Simcoe Watershed are shown on the map in Fig. 1.

The insert map in the lower right-hand corner of Fig. 1 illustrates the location of the Lake Simcoe Watershed within the Great Lakes region. The City of Barrie has a population of just over 143,000 and is approximately a one-hour drive north from the City of Toronto [2,14]. The population density per km² is 1753.6 and the dominant household building type is single detached [15].

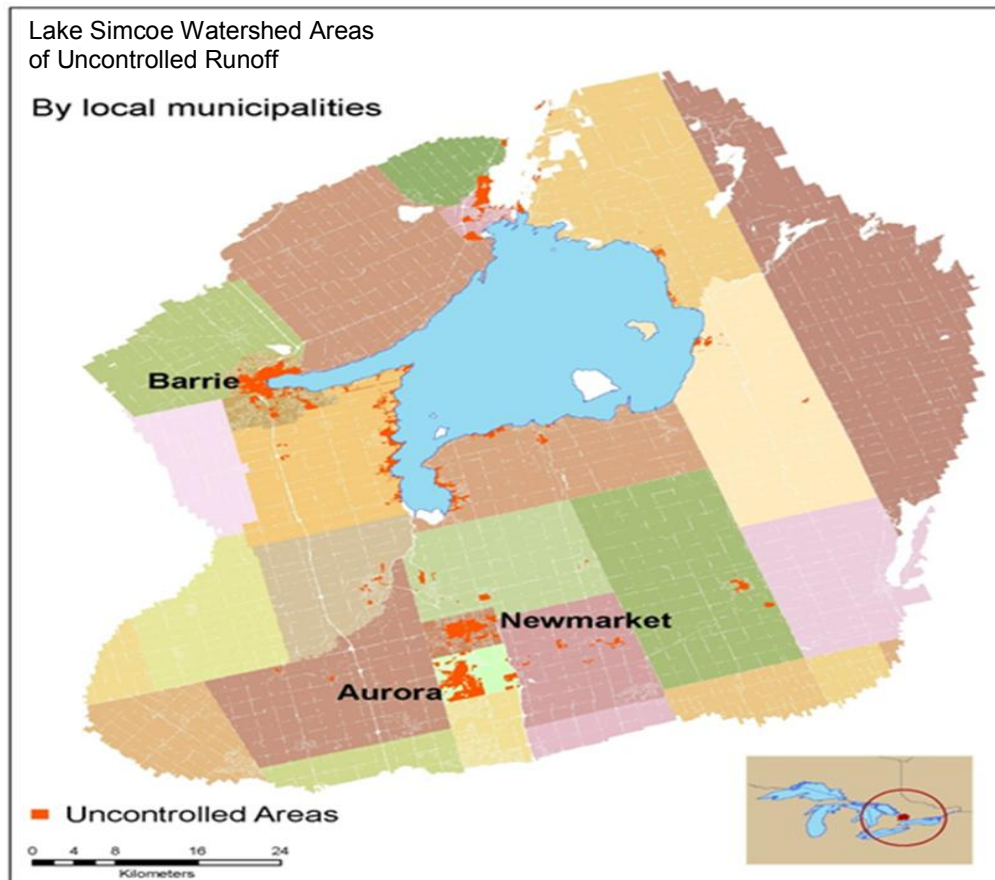


Fig. 1. Map of uncontrolled areas within the Lake Simcoe Watershed in Southern Ontario, Canada [6]

The Lake Simcoe Watershed contains provincially significant wetland, woodland, and agricultural areas [6]. Approximately half of the land in the Lake Simcoe Watershed is agricultural and 35% is woodlands and wetlands [6]. The Lake Simcoe Watershed entertains seasonal populations such as tourists and recreational users including boaters, anglers, and cottagers.

As the permanent population within the watershed grows, development within the region must comply with policies specified in acts such as the *Ontario Water Resources Act* (RSO 1990, c.O.40) and the *Clean Water Act* (S.O. 2006, c.22). These policies apply to the quality and quantity of sewage and stormwater with respect to potable and non-potable water resources [6].

The *Lake Simcoe Protection Act* was enacted which allowed the development of the Lake Simcoe Protection Plan, an action plan for achieving water quality and quantity targets [6]. As a result, municipalities within the watershed were required to develop master plans entailing a comparative review of current practices, new technologies, and retrofit opportunities to determine the best course of action for optimizing stormwater management efficiencies [6]. In the study by Li et al. [6], the suitability and effect of implementing LID technologies within the uncontrolled stormwater areas was evaluated.

The hydrological modelling method that was developed in the current work is generalizable. It is only limited by the data and the application. It can potentially be applied to regions outside of southern Ontario if criteria are defined which determine whether it is appropriate for the application. For example, data on infrastructure, hydrological, and land use information is required for LID selection and hydrologic analysis [2].

2. METHODOLOGY

2.1 UHRU Development

The UHRU concept has been applied in hydrologic modelling of large watersheds [6,9,13,16]. Before grouping together lots with similar hydrological characteristics, a common database with hydrologic attributes (e.g. lot area, lot width, percent imperviousness, parking area, driveway area, building area, percent slope, soil type) should be compiled. Attributes were

selected for experimentation based on characteristics intrinsic to the modelling software, previous experience from the case study by Li et al. [6], and the suitability and availability of data [2]. Five major attributes selected for further analyses and examination were:

1. Lot area (m²),
2. Lot width (m),
3. Average slope (%),
4. Percent imperviousness (building area (%), parking area (%), and driveway area (%)), and
5. Soil type (categorical value).

Depending upon the grouping techniques used, normalization of some attributes may be necessary. After comparing various clustering methods offered by different software programs, the K-means cluster analysis procedure offered by IBM SPSS was selected as the classification method. The procedure uses a commonly used method of measuring the dissimilarity or distance between two data objects, the Euclidean distance. The Euclidean distance is a dissimilarity measure that is sensitive to the differences in magnitudes or scales of input variables [17]. The K-means clustering method was selected mainly because of its ability to provide detailed, usable output in a convenient form regarding the final cluster centre for each cluster and the distance to the final cluster centre for each lot or case. It is also capable of analyzing large data files and maintaining the categorical values assigned to soil type.

Four data standardization techniques were applied across the data set to standardize input data variables to dimensionless data before performing clustering analysis. Each technique used the following general equation:

$$x_{ij} = \frac{x_{ij}^* - L_j}{M_j}$$

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 general equation varied according to L_j and M_j as shown in Table 1. The "Range 2" data standardization technique presented in Table 1 refers to the second of two techniques involving division by the range, R_j^* , of the input variable for each case.

Table 1. Location measure, L_j , and scale measure, M_j , for each data standardization technique investigated

| Data standardization technique | L_j | M_j |
|--------------------------------|---|---|
| USTD | 0 | σ_i^* |
| Sum | 0 | $\sum_{i=1}^n x_{ij}^*$ |
| Range 2 | $\min x_{ij}^* \quad (1 \leq i \leq n)$ | R_i^* |
| Maximum | 0 | $\max x_{ij}^* \quad (1 \leq i \leq n)$ |

The mean (x_{avg}^), range (R_i^*), and standard deviation (σ_i^*) of the j^{th} variable were calculated according to the conventional mathematical equations for these variables [17]

Recommendations were also included within the statistical software used to perform clustering analysis, IBM SPSS, to standardize all input variables to the same scale before running the program. Since soil type was input as a categorical variable, it was left unstandardized to maintain consistent numerical values in the analyses output. Histograms plotted before and after standardization demonstrated that frequency distributions for each variable remained unchanged.

Some studies have recommended data standardization whereas others have suggested that it may not be advisable. A simulation study by Cooper and Milligan [18] examined the standardization problem and presented results for eight standardization strategies. It was concluded that approaches which standardize by division by the range of the variable gave consistently superior recovery of the underlying data structure. This conclusion lent further support to using the data standardization technique involving the range of the input variable, known as Range 2 in this research.

After assigning lots or cases to clusters using the Euclidean distance measure, the K-Means Cluster Analysis procedure updates the locations of the cluster centres based on mean values of cases in each cluster [2]. The final outcome after this cyclical procedure should be clusters of relatively homogeneous groups of cases or lots based on the selected attributes. One of two methods for classifying lot parcels using the K-Means Cluster Analysis Procedure was then selected: (1) updating cluster centres iteratively or (2) classifying only. The iterative option is a more detailed procedure involving more decisions and steps. Both methods were used to perform cluster analyses for data sets derived from each of the four standardization techniques. After reviewing statistical output reports and analysing cluster diagrams which were

subsequently developed, the method which produced the most accurate results was selected [2].

Cluster diagrams are colour-coded outlines of the clusters or UHRUs created from the clustering analysis procedure. Since real spatial data were used as input in all of the methods and procedures explored, cluster diagrams were developed to visualize visible hydrological characteristics and lot properties such as lot area (m^2) and imperviousness (%). ArcMap software was used to develop the diagrams which were overlaid on current orthophotos or aerial maps of the City of Barrie. Lots within each cluster and between clusters were compared for similarities and dissimilarities of visible hydrological characteristics [2].

In Fig. 2, the distinction between cluster groups due in part to soil type is shown for a group of 10 clusters formed using the Range 2 technique and the classifying only option. The curvy, diagonal yellow line drawn on the figure illustrates the partitioning of the lots into different clusters based on soil type. The relatively large property outlined in blue, situated in the area containing sandy loam of Fig. 2, is a multi-unit dwelling that was assigned to a different cluster than the neighbouring single-family unit lots outlined in green. The placement of the multi-unit dwelling into another cluster was therefore based on other characteristics such as lot area and level of imperviousness.

Following a visual analysis and inspection of the cluster diagrams, the data standardization technique and classification method most closely resembling reality as shown in the aerial map was selected [2].

Microsoft Excel macros were developed by using the Visual Basic for Applications (VBA) programming language to run batches of input

data using SWMM software. Macros were developed to simultaneously perform hydrological modelling simulations for thousands of residential urban lots. Two versions of the macro were developed for each scenario of no LID implementation and LID implementation. The first type of LID to be modelled on all of the lots was the rain barrel (RH) for residential properties. The macros were first used to simulate total runoff (mm) for all of the residential lots with RH and without RH. The same procedure was then repeated for the porous pavement (PP) singular LID.

2.2 Calculation of Total Runoff (m^3)

After performing hydrological modelling on all lots using the two versions of the macro and developing lot clusters using the K-means cluster analysis procedure, two cluster-based methods were developed to calculate the total runoff (m^3) volume of all lots: (1) the Minimum Distance Method and (2) the Random Sampling Method. The first method, the Minimum Distance Method, uses the lot with the minimum distance to the cluster centre or with a distance of zero to the cluster centre as the final cluster centre. After the lot clusters have been formed and hydrological modelling output has been generated, the total runoff (mm) produced by the final cluster centre was extrapolated to the rest of the lots within the cluster. First, the lot area (m^2) for each lot within the cluster 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 final cluster centre. The total runoff (m^3) of each cluster within the group of

clusters was then summed to yield the overall total runoff (m^3) for all lots [2]. The procedure for the Minimum Distance Method is illustrated in Fig. 3.

The Minimum Distance Method for calculating the total runoff (m^3) for each group of clusters was applied to both the non-LID scenario and the LID scenario as was the Random Sampling Method.

For The Random Sampling Method, a random percentage sample of lots was extracted from each cluster within a group of clusters. For example, 5% of lots were randomly selected from each cluster of a group of three clusters (3-cluster group). The total runoff (mm) generated by each lot from hydrological modelling was considered as the dependent variable and the other lot characteristics used when developing clusters or UHRUs were considered as the independent variables. All variables were left in unstandardized form for performing stepwise linear regression on each cluster using IBM SPSS [2].

Stepwise linear regression was performed on each cluster and the “best-performing” linear regression equation was selected for each cluster based on statistical diagnostics. The selected regression equation was used to calculate runoff (mm), the dependent variable, for each lot within the cluster. The total runoff (m^3) for each cluster was then summed for all clusters to yield total runoff (m^3) for all lots or the entire cluster group [2]. Fig. 4 presents a general overview of the Random Sampling Method.



Fig. 2. Clusters from a 10-cluster group developed by using the Range 2 data standardization technique and selecting the classifying only option [2]

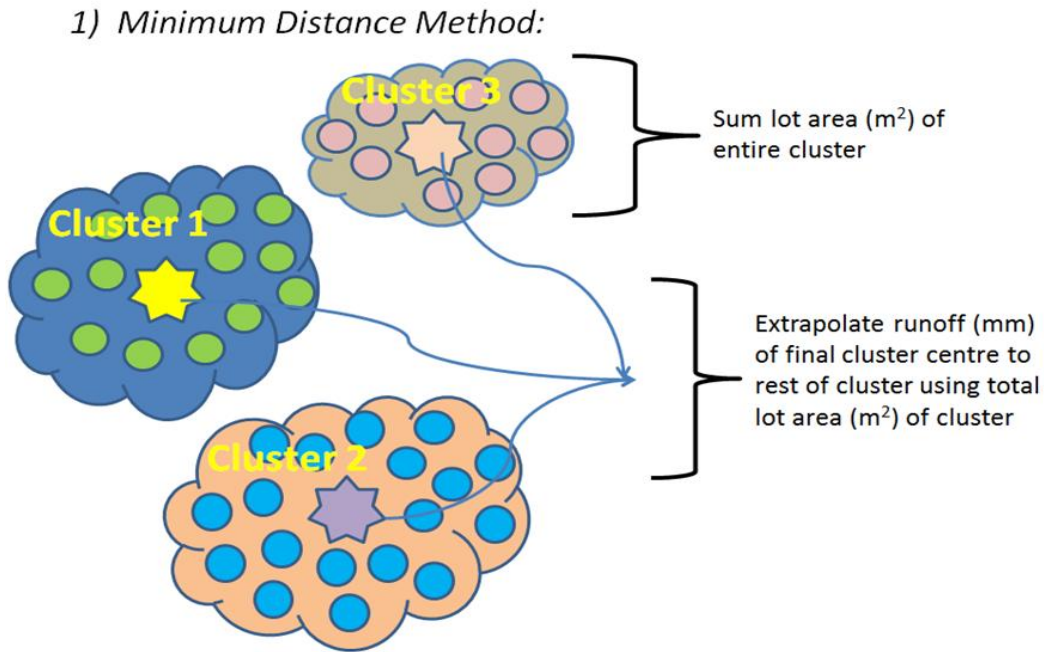


Fig. 3. Overall procedure for calculating total runoff (m^3) of all lots using the minimum distance method [2]

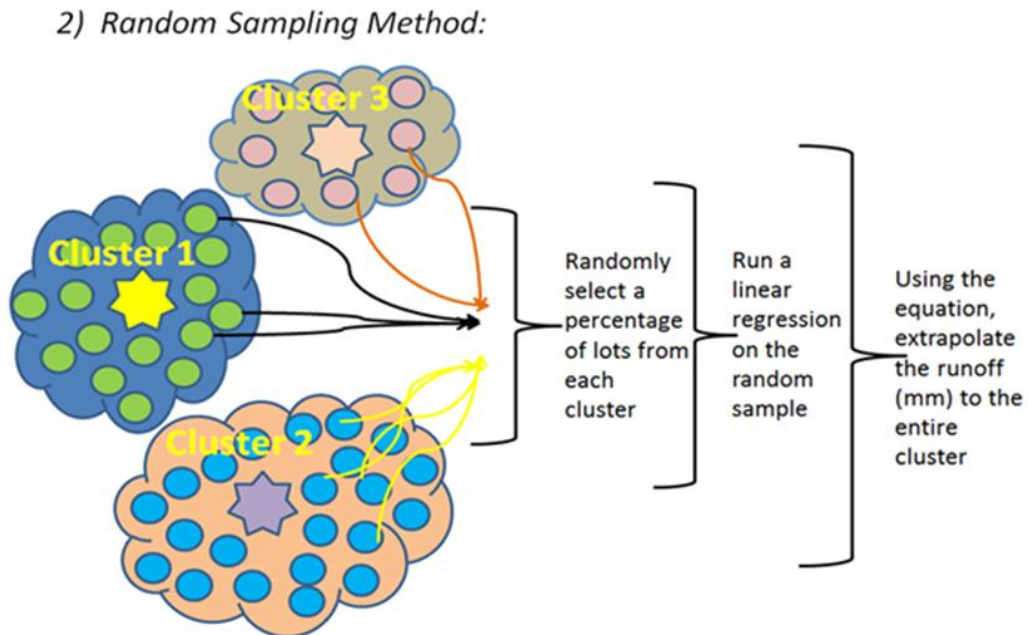


Fig. 4. General overview of the procedure for calculating total runoff (m^3) of all lots using the random sampling method [2]

After evaluating the results from the two cluster-based methods, another method was developed to simplify the process and tested using the porous pavement LID. The two versions of the Excel macro were modified and tailored to model residential properties amenable to

implementation of porous pavement LIDs. A random sample of 5% was selected from all lots in the data sets for both scenarios of no LID and LID implementation after performing hydrological modelling [2].

A stepwise linear regression was performed on the 5% random sample of lots without PP. A best-performing regression equation from the random sample was again selected based on statistical diagnostic output to extrapolate and sum total runoff (mm) for all lots in the data set. The dependent variable was total runoff (mm) from the hydrological modelling output and the independent variables were the hydrological characteristics used in clustering, just as in the previous two methods discussed [2].

Similarly, a linear regression was performed on a random sample of 5% of all lots from the “with PP” data set to generate a regression equation used to extrapolate and sum total runoff (mm) for all lots in the entire data set. The selection of a best-performing regression equation was not required since only one equation was provided in the regression analysis output. The total runoff (mm) of all lots amenable to PP implementation was calculated for both situations of with and without LID implementation [2].

3. RESULTS AND DISCUSSION

3.1 Cluster-Based Calculation Methods for RH LID Implementation

Output produced by the K-means cluster analysis procedure was reviewed in conjunction with cluster diagrams to determine which combination of data standardization technique and classification method most closely matched the actual distribution of hydrological characteristics. For example, a cluster run producing three clusters should have had a fair amount of lots in each cluster because it was highly unlikely that a lot within the study area would be so unique as to warrant its own cluster. An ANOVA table and a

matrix showing the Euclidean distance between final cluster centres are other examples of clustering output that were analyzed. Table 2 presents a matrix illustrating the dissimilarities between final cluster centres for a group of five clusters. Since a greater distance between two points leads to greater dissimilarities between final cluster centres, Table 2 demonstrates that clusters two and one are the most dissimilar. These matrices were used to confirm that degrees of dissimilarity between clusters existed [2].

Colour-coded cluster diagrams were placed on top of aerial orthophotos of the study area to verify whether lots within the same cluster resembled each other in terms of visible hydrological characteristics such as lot area (m^2) and percentage imperviousness. After reviewing clustering output and cluster diagrams, the data standardization technique involving division by range, Range 2, and the classification method option of classifying only were selected for performing clustering analyses. The “true” total runoff (m^3) values for the two scenarios under study were taken to be the benchmark output values (m^3) derived from modelling all of the lots. The end values from employing the methods developed for calculating total runoff (m^3) were compared to the benchmark output values (m^3) to determine which method produced the most accurate results [2].

Although some groups of clusters approximated the benchmark output value (m^3) for both cluster-based calculation methods developed, there was no observable pattern or trend for either method for both scenarios of RH implementation. For the Minimum Distance Method, only one group of clusters, the 15-cluster group, approximated the benchmark output value for the initial scenario of no RH LID. Three groups of clusters (15-cluster group, 20-cluster group, 25-cluster group) approximated the benchmark output value (m^3) for the scenario of RH LID implementation as shown in Fig. 5 [2].

Table 2. A matrix illustrating dissimilarities between final cluster centres [2]

| 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 | |

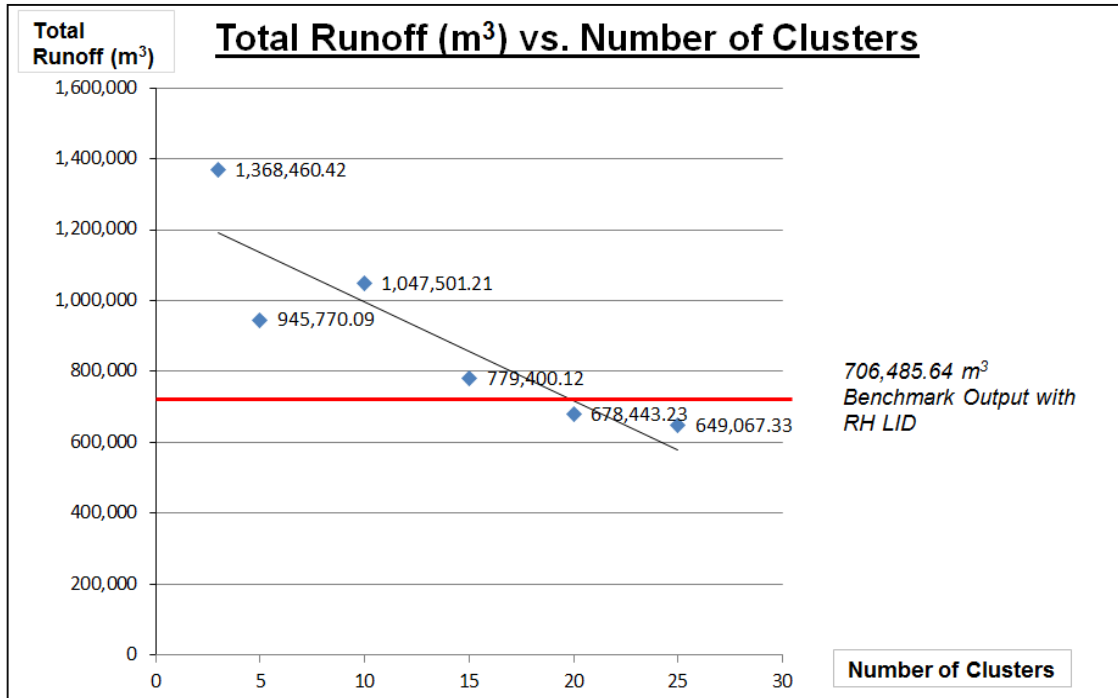


Fig. 5. Total runoff (m³) versus number of clusters within a cluster group for residential lots with the RH LID [2]

A possible explanation for the lack of observable pattern is the inherent arbitrariness of the K-means cluster analysis procedure that may result in some groups of clusters approximating the benchmark output more closely at times than others. Some groups of clusters may be swayed by unusually high or low values for certain characteristics when compared with the values of those same characteristics for the final cluster centre. The final cluster centre is therefore not always a useful representation of the values within the cluster [2]. The results, however, still indicate the potential for modelling LIDs efficiently over an entire watershed by demonstrating that only a small subset of the database is required to be modelled. Modelling results can then be extrapolated to the rest of the lots in the database [6].

Similar results were obtained by the Random Sampling Method. The only random sample to closely approximate the benchmark output was a random sample of 7 % of lots extracted from each cluster of a 3-cluster group. The approximation of the benchmark output value by the random sample of 7 % of all lots from each cluster of a 3-cluster group is shown in Fig. 6 for the situation excluding RH LID implementation [2].

A random sample of 90% of all lots was required from each cluster of a 3-cluster group to approximate the benchmark output for the situation involving RH LID implementation. Such results are almost equivalent to modelling every lot and provide little benefit to LID modelling based on the UHRU concept. Stepwise linear regression results for both situations, however, were very high ($R^2 \approx 1$) confirming that variation in total runoff (mm) is accounted for by the selected independent variables [2].

As a result of highly accurate regression results, the selected regression equation seemed to be replicating or reproducing actual total runoff (mm) results for each lot in the case of LID implementation. The method became tantamount to adding the total runoff (mm) results for each lot which is why a high amount, 90 %, was required to reach the benchmark value. The low percentage required for the no LID case may have been due to slightly less accurate regression equations. This led to an overestimation of total runoff (mm) for some lots when the regression equation was used to extrapolate total runoff (mm) to other lots. The benchmark output value therefore required a smaller amount of lots to be approximated.

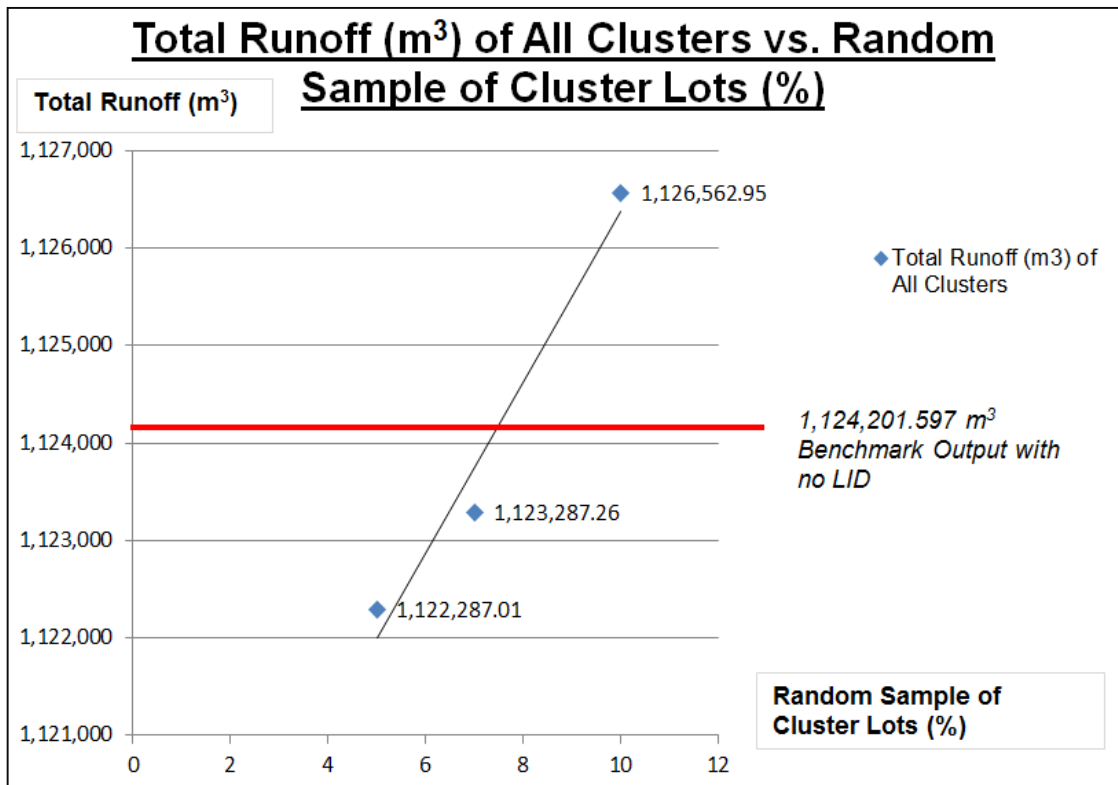


Fig. 6. Approximation of benchmark output by a random sample of 7 % of clustered lots from each cluster of a 3-cluster group with no RH LID implementation [2]

The results of both cluster-based methods provide no clear trends or patterns that can be generalized to produce more efficient methods for developing and modelling UHRUs for LID technology application. The results do confirm, however, that lot-level detail can still be captured without modelling every lot. They also indicate that the UHRU method may also be unnecessary when performing hydrological modelling over a large watershed area [2].

3.2 Random Sampling of All Lots for PP Implementation

High statistical scores from regression analyses for the Random Sampling Method and the application of regression equations to lot clusters, demonstrated that regression equations based on a subset of lots could be used to closely approximate total runoff (mm) values for all lots. When a 5 % random sample was extracted from all lots for both scenarios, regression equations generated by both the stepwise linear and linear regression analyses

for the random sample were able to almost reproduce exact total runoff (mm) values for each lot in the random sample. The extrapolation of the results from the 5 % random sample to all lots using the regression equation approximated the benchmark output very closely for both scenarios as illustrated in Table 3.

The results from extracting a random sample of 5 % of all lots to the entire data set suggest that a small random sample may be sufficient when extrapolating modelling results of a small number of lots to an entire watershed using linear regression equations. Since clusters or UHRUs do not have to be developed, this can be viewed as a more efficient approach than an approach based on an UHRU concept. Linear regression may be preferable to stepwise linear regression because there is no opportunity for the modeler to select a regression equation that is not considered as optimal. More work is needed to confirm the results for other types of LID practices and other percentage values of random samples.

Table 3. Total runoff (m³) values for no PP LID and PP LID implementation using a 5 % random sample from all lots

| | No PP LID (stepwise linear regression) | With PP LID (linear regression) |
|---|---|--|
| Benchmark total runoff (m ³) | 937,088.58 | 746,462.40 |
| Total runoff (m ³) of a 5 % random sample from all lots | 945,382.97 | 747,380.13 |

4. CONCLUSION

The results of this study suggest that it may be unnecessary to develop HRUs or UHRUs when performing hydrological modelling on large watersheds for the evaluation of LID performance. An alternative approach is to first perform hydrological modelling on a small random sample of lots, in this case 5 % of approximately 6600 lots amenable to PP LID implementation. A regression equation can then be generated by performing stepwise linear regression or linear regression on the random sample to extrapolate values for the dependent variable, total runoff (mm), to the entire sample. The sum of all total runoff (mm) values for the random sample of lots will closely approximate the total benchmark output value for total runoff (mm) produced by modelling all lots in the data set using software.

This alternative approach can be used for each condition of with and without PP LID implementation to determine the benefits of LID implementation. The approach alleviates the need for modelling thousands of lots and/or obtaining an extensive data set. It can provide a more efficient and practical method of hydrological modelling LID implementation for municipalities or organizations with limited resources. However, the findings are limited to case study data and depend on the distribution of lot characteristics. Since the approach is based on case study data, it may not be applicable to every data set or every region.

ACKNOWLEDGEMENTS

The authors would like to acknowledge support from the Natural Sciences and Engineering Research Council of Canada and the Lake Simcoe Region Conservation Authority.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Gilroy KL, McCuen RH. Spatio-temporal effects of low impact development practices. *J Hydrol (Amst)*. 2009;367:228-236.
2. Eric M. Modelling low impact development potential with urban hydrological response units. Master's Thesis, Ryerson University; 2012.
3. Dietz ME. Low impact development practices: a review of current research and recommendations for future directions. *Water Air Soil Pollut*. 2007;186:351-363.
4. Elliott AH, Trowsdale SA. A review of models for low impact urban stormwater drainage. *Environ Model Softw*. 2007; 22:394-405.
5. Carmon N, Shamir U. Water-sensitive planning: integrating water considerations into urban and regional planning. *Water Environ J*. 2010;24:181-191.
6. Li J, Banting D, Joksimovic D, Eric M, Fan C, Hahn K, Lawson S, Mirzajani M. 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; 2010.
7. Cowan DM, Easton ZM, Fuka DR, Schneiderman EM, Steenhuis TS, Walter MT. Re-conceptualizing the soil and water assessment tool (SWAT) model to predict runoff from variable source areas. *J Hydrol (Amst)*. 2008;348:279-291.
8. Andrieu H, Morena F, Rodriguez F. A distributed hydrological model for urbanized areas – Model Development and application to case studies. *J Hydrol (Amst)*. 2008;351:268-287.
9. Elliott AH, Trowsdale SA, Wadhwa S. Effect of aggregation of on-site stormwater control devices in an urban catchment model. *J Hydrol Eng*. 2009; 14(9):975-983.

10. United States Department of Agriculture - Soil Conservation Service (USDA-SCS). Part 630 Hydrology. National Engineering Handbook. 1972;Section 4, Chapter 10.
11. England CB, Stephenson GR. Response Units for Evaluating the Hydrologic Performance of Rangeland Watersheds. 1970;11:89-97.
12. England CB, Holtan HN. Geographic grouping of soils in watershed engineering. Journal of Hydrology. 1969;7:217-225.
13. Kessler A, Kronaveter L, Shamir U. Water-sensitive urban planning: modeling on-site infiltration. Journal of Water Resources Planning and Management. 2001;127(2): 78-88.
14. The City of Barrie. The City of Barrie 2012 Accessibility Plan. City of Barrie, Ontario, Canada; 2012.
15. City of Barrie. Fast Facts; 2013. Accessed 3 June 2014. Available:<http://www.barrie.ca/Doing%20Business/profile/Pages/FastFacts.aspx>
16. Carrillo G, Sawicz K, Sivapalan M, Troch PA, Wagener T. Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA. Hydrology and Earth System Sciences Discussions. 2011;8:4495-4534.
17. Gan G, Ma C, Wu J. Data clustering theory, algorithms, and applications. Society for Industrial and Applied Mathematics, Philadelphia, PA; 2007.
18. Cooper MC, Milligan GW. A study of standardization of variables in cluster analysis. Journal of Classification. 1988; 5:181-204.

© 2015 Eric et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.