DERIVATION OF FLOOD FREQUENCY DISTRIBUTION EQUATIONS

DATA ANALYSIS OF RAINFALL EVENT CHARACTERISTICS AND DERIVATION OF FLOOD FREQUENCY DISTRIBUTION EQUATIONS FOR URBAN STORMWATER MANAGEMENT PURPOSES

by

SONIA HASSINI, B.Eng., M.Eng.

A Thesis Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements for the Degree
Doctor of Philosophy in Civil Engineering
McMaster University

DOCTOR OF PHILOSOPHY (2017) McMaster University
(Civil Engineering) Hamilton, Ontario

TITLE: Data analysis of rainfall event characteristics and derivation of flood frequency distribution equations for urban stormwater management purposes

AUTHOR: Sonia Hassini, B.Eng. (Ecole Nationale d'Ingénieurs de Tunis, ENIT), M.Eng. (University of Waterloo)

SUPERVISOR: Professor Yiping Guo

NUMBER OF PAGES: xi, 227

LAY ABSTRACT

Urban stormwater management aims at mitigating the adverse impacts of urbanization. Hydrological models are used in support of stormwater management planning and design. The analytical probabilistic stormwater management model (APSWM) is a promising tool for planning and design analysis. The purpose of this thesis is to further develop APSWM in order to make it more reliable and accurate. First, a clear procedure for rainfall data analysis as required by APSWM is provided. Second, a new APSWM is derived incorporating other runoff temporal-distribution patterns. Finally, the possibility of soil layer saturation while it is still raining is added to the model. All the models developed in this thesis are tested and compared to methods used in engineering practice, reasonable results were obtained.

ABSTRACT

Urban stormwater management aims at mitigating the adverse impacts of urbanization. Hydrological models are used in support of stormwater management planning and design. There are three main approaches that can be applied for this modeling purpose: (1) continuous simulation approach which is accurate but time-consuming; (2) design storm approach, which is widely used and its accuracy highly depends on the selected antecedent moisture conditions and temporal distribution of design storms; and (3) the analytical probabilistic approach which is recently developed and still not used in practice. Although it is time-effective and it can produce results as accurate as the other two approaches; the analytical probabilistic approach requires further developments in order to make it more reliable and accurate. For this purpose, three subtopics are investigated in this thesis. (1) Rainfall data analysis as required by the analytical probabilistic approach with emphasis on testing the exponentiality of rainfall event duration, volume and interevent time (i.e., time separating it from its preceding rainfall event). A goodness-of-fit testing procedure that is suitable for this kind of data analysis was proposed. (2) Derivation of new analytical probabilistic models for peak discharge rate incorporating trapezoidal and triangular hydrograph shapes in order to include all possible catchment's responses. And (3) the infiltration process is assumed to continue until the end of the rainfall event; however, the soil may get saturated earlier and the excess amount would

contribute to the runoff volume which may have adverse impact if not taken into consideration. Thus, in addition to the infiltration process, the saturation excess runoff is also included and new models for flood frequencies are developed. All the models developed in this thesis are tested and compared to methods used in practice, reasonable results were obtained.

ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisor, Dr. Yiping Guo, for his commitment, patience, understanding and guidance. In particular, he has been exceptionally quick in providing feedback on my work which helped me stay on track.

I would also like to thank my other three supervisory committee members, Dr. Paulin Coulibaly, Dr. Shui Feng, and Dr. Syed Moin for their commitment and invaluable comments and suggestions.

Special thanks to all my dear teachers from elementary school to graduate school. I am also indebted to my relatives, friends and those who have been a source of enlightenment.

I would also like to thank my Parents Bayya and Sahbi for raising me with a love of learning and for their continuous support. They have taught me to never give up on my dreams. My mother instilled in me the respect for education. I learned to be an independent thinker thanks to my father's support and encouragements.

Thanks are also due to my siblings for their unconditional love. They have been a

source of joy that has always boosted my spirits.

Last but not least, I am greatly appreciative for my family's support; my husband Elkafi for his complete support, patience, encouragement, and motivation and my kids, Tesneem, Ahmed, Sara, Zeinab, and Sumaya, for providing me with the joy and inspiration that has lifted me during tough times. Many of my peers and friends have been wondering about how I was able to manage my PhD thesis while at the same time raising five beautiful kids? While the challenges are different, raising the kids has definitely been a source of positive energy for me.

TABLE OF CONTENTS

DECLARATION OF ACADEMIC ACHIEVEMENT

I, Sonia Hassini, declare this thesis to be my own work, under the supervision of Professor Yiping Guo. I completed all the research work included within this thesis.

This thesis was prepared in accordance with the regulations provided by the school of graduate studies, McMaster University for a sandwich thesis. It consists of three co-authored papers presented in chapters 2 through 4. The literature review was conducted by S. Hassini, derivation of equations and comparisons were carried out by S. Hassini in consultation with Dr. Y. Guo. The three papers were written by S. Hassini and edited by Dr. Y. Guo.

CHAPTER 1

Introduction

1.1. Stormwater management

Within a catchment, the existence of urban facilities such as paved roads, buildings and parking lots reduces pervious areas, which results in an increase of stormwater volume and peak discharge rate, rapid catchment response and alteration of natural water courses. Moreover, urbanization is usually accompanied by an inevitable increase of pollutants such as grease, heavy metals and nutrients. Thus, urban stormwater increases flooding risks and transports pollutants to the receiving water systems resulting in the disruption of habitats and ecosystem balance. In order to mitigate urban floods and reduce the adverse impacts of stormwater and accompanying pollutants, it is essential to manage urban stormwater (Nnadi et al., 1999). Besides non-structural (soft) controls such as zoning control and public education, structural stormwater management facilities are required.

Structural facilities are classified into three categories: (1) lot level controls, such as rear- yard storage and parking lot storage, collecting stormwater from individual or multiple lots within small drainage areas; (2) conveyance controls, such as ditches and

sewers, conveying stormwater to other facilities or locations; and (3) end-of-pipe controls, such as ponds and infiltration basins, receiving stormwater from the conveyance systems and then discharging it to the receiving waters after settlement of suspended solids. In order to meet the multiple criteria for water quantity, water quality, water balance and erosion management, the three types of controls have to be integrated (Ontario Ministry of the Environment, 2003).

Quantity controls limit flooding and its effects and help to preserve and redirect stormwater as required. Quality controls tend to protect the quality of downstream receiving waters and underlying groundwaters by removing maximum possible amounts of suspended solids. The design of stormwater management practices requires mainly the frequency distributions of runoff volume and peak discharge. So far there is no, one-step, accurate and time-effective method that can be used to determine the required design information.

1.2. Hydrologic models

Hydrologic models are required for analysis, design, long-term runoff-volume forecasting, future urbanization impact investigations, real-time flood forecasting, etc. There are different types of hydrologic models developed by independent or related research based on the need for a model and enhancement of existing models. There are two main types of models, which are material and formal. A material model is a simplified structure that represents the prototype with respect to its main properties. For instance, an experimental watershed is a material model. In hydrology, a formal

model or mathematical model is a set of mathematical equations that describes one or more phases of the hydrologic cycle. The details of the hydrological cycles at the location of interest usually depend on the regional climate (arid, semi-arid or humid), the catchment characteristics (size, relief, soil etc.) and the season. Consequently, a hydrologic model that is suitable for a catchment may not work for another one. A hydrologic model may describe every aspect of the hydrologic cycle (complete model) such as continuous simulation model for rainfall-runoff transformation or represent only part of the hydrological cycle (partial model) such as the infiltration process (Hydrocomp, 2007).

For catchment modeling, mathematical models are widely used because of their flexibility, availability and cost-effectiveness, as compared to material models. The development and implementation of mathematical models are highly related to the advancement of computer features and programming (Ponce, 1989). Normally, the accuracy level of mathematical models increases with its degree of complexity, however a balance between accuracy and simplicity is required. Calibration and verification of mathematical models are important steps for a model to be accepted; however due to the lack of reliable data, generally, stormwater models are not fully calibrated or verified (Adams and Papa, 2000).

1.2.1. Hydrological model classifications

There are different classifications of mathematical catchment models where the main ones are presented as follows.

a. Physical, conceptual and parametric models

According to Ponce (1989), mathematical catchment models are mainly classified into three categories of mathematical models: physical, conceptual and parametric. A physical model is governed by laws of physical processes and it is described by equations of mathematical physics. In general, it is represented by differential equations such as the Kinematic Wave-Routing technique. Although physical models provide the best details of physical processes, their applicability is limited due to the complexities of the physical phenomena. A conceptual model is a representation of a simplified physical system where empirical components are included. For instance, a watershed can be represented by reservoirs and channels as in the software package HBV (Hydrologiska Byråns Vattenbalansavdelning; Hydrological Bureau Water balance-section) originally developed by the Swedish Meteorological and Hydrological Institute (SMHI-HBV-model, 2007). Mathematically, conceptual models are generally described by ordinary differential equations or algebraic equations such as the Green-Ampt infiltration model (McGuen, 1989). In engineering hydrology, conceptual models are widely used as a result of the difficulties encountered in using physical models. A parametric model, also called an empirical model, is usually described by algebraic equations with empirical parameters. The rational method that is used for peak discharge estimation is an example of parametric models (Ponce, 1989).

b. Probabilistic and deterministic models

Probabilistic models are generated by laws of chance and used for hydrologic time-series analysis, where one or more variables are random. In hydrology, probabilistic models often involve the determination of the probability distribution of equaling or exceeding an event (Kottegoda, 1980). For instance, flood frequency analysis can be achieved by a probabilistic model such as the Gumbel distribution model (Linseley et al., 1982). In contrast to probabilistic models, deterministic models have no random input or output variables. Most of the existing stormwater models are deterministic such as Storm Water Management Models (SWMM) developed by the US Environmental Protection Agency (USEPA).

c. Lumped and distributed models

For a lumped model (lumped-parameter model), catchment characteristics and rainfall data are considered constant over the catchment space. As an example, the Analytical Probabilistic Stormwater Model (APSWM) developed by Guo and Adams (1998a, b; 1999a, b) is a lumped model. For a distributed model, the catchment is divided into a large number of small sub-areas where the sum of the sub-areas' responses, which are separately simulated, constitutes the entire catchment response. Theoretically, distributed models are more accurate than lumped models; however, if detailed inputs are not available they may not give better results than lumped models. In practice distributed models are too time-consuming to construct; they are however

efficient for investigating the consequences of heterogeneity in a catchment (Linseley et al., 1982). A model representing a watershed divided into sub-watersheds, each has representative lumped parameters, is considered a distributed model. For example, the Hydrological Simulation Program-Fortran (HSPF) developed by the US Environmental Protection Agency (USEPA) is a distributed model (Hydrocomp, 2007).

d. Event and continuous models

An event model is based on rainfall-runoff event transformations. It requires initial conditions, which are used as input data and they can be assumed or estimated from catchment characteristics (Hydrocomp, 2007). Event models focus on infiltration and surface runoff processes to evaluate the direct runoff of the catchment of interest (Ponce, 1989). The accuracy of event models depends on the initial conditions, in addition to the selected rainfall-runoff transformation method (Hydrocomp, 2007). For instance, HEC-1 developed by the Hydrologic Engineering Center, US Army Corps of Engineers is an event model (Ponce, 1989).

A continuous model takes into account all the phases of the hydrologic cycle, usually it considers the three runoff components: surface flow, interflow and groundwater flow. It requires long series of rainfall records and produces flow rates and conditions over the entire period of estimation. Consequently, continuous models determine soil moisture conditions during both wet and dry periods, which imply that they estimate the initial conditions required for runoff events. For instance, SWMM is

a continuous model (Ponce, 1989; Hydrocomp, 2007).

e. Numerical and analytical models

Mathematical models are classified as analytical or numerical depending on the type of their solutions (Ponce, 1989). Numerical models produce numerical results. On the other hand, analytical models generate analytical solutions. Analytical models are extremely useful for planning stage analysis where a number of alternative designs can be easily evaluated (Chen and Adams, 2005).

1.2.2. Design flow estimation techniques

The objective of design analysis is to provide hydrologic information which is crucial for the planning and design of urban stormwater practices (Guo and Zhuge, 2008). Peak discharge rates of specific return periods are used for the design of flood control facilities, runoff volumes of specific return periods are used for quality control practices and the complete discharge hydrographs are used to check the function of control practices. Design flows can be predicted from either measured or estimated flow data.

a. Flood frequency analysis

The procedures of flood frequency analysis are, first, calculation of statistical data (means, standard deviations, etc.) from annual series of measured flow data. Then, the obtained statistical information along with a probability distribution is used to represent discharges as a function of return periods or exceedance probabilities. The

commonly used probability distributions for flood frequency are Normal, Log-Normal, Log-Pearson Type III and Gumbel distributions. The empirical relative frequency relations (i.e., plotting position formulas) are also used to determine flood frequency distributions (Wurbs and James, 2002).

Although this technique is simple, its applicability is limited due to two main reasons: (1) it requires a long series of recorded flow data, which are usually, short or not available; (2) construction of stormwater management practices and fast rate of urbanization affect runoff characteristics; consequently, predevelopment flows are not suitable to design postdevelopment structures (Marsalek and Watt, 1984).

Regional frequency analysis can be used to overcome the first drawback of the flood frequency analysis method by using the data available from surrounding sites. However, its applicability is limited due to site heterogeneity, in addition to the second drawback of flood frequency analysis (Quader and Guo, 2006).

b. Design storm approach

The design storm approach was first developed to replace flood frequency analysis where runoff records are short or unavailable (Linsley et al., 1982). The concept is frequently used for the planning and design of urban drainage systems and other small projects such as urban stormwater management practices (Linsley et al., 1982; Guo and Zhuge, 2008).

The design storm approach is used to simulate the direct runoff volume and peak based on a single design storm and a rainfall-runoff transformation model. A design

storm is a selected hyetograph, i.e., temporal rainfall distribution, defined by a number of characteristics such as return period, and duration, which are used as inputs for a rainfall-runoff transformation model. The rational method and unit hydrograph are the most widely used rainfall-runoff transformation models (Hromadka, 1997). Although both the design storm definition and the transformation model are critical for the accuracy of the design storm approach, there were more criticisms about the selection of design storm characteristics.

Linsley et al. (1982) and Adams and Howard (1986) reported that the design storm approach is based on the assumption that the rainfall event and the corresponding runoff have the same return period; however, this assumption is rarely true. Generally, the frequency of the storm depth is assigned to the entire storm; although other characteristics such as duration, temporal distribution, etc. may have different return periods and also affect the runoff volume and peak. For example, two rainfall events having the same rainfall volume and different durations may not produce the same runoff volume or discharge peak. As a result, the use of the design storm approach is not always appropriate. However, in linear systems it can be assumed that rainfall and runoff have equal frequencies (Vaes et al., 2001).

Moreover, the soil's antecedent moisture conditions may affect the hydrograph as well; the same hyetograph with identical catchment and different antecedent conditions produce two different hydrographs. Thus, the antecedent moisture conditions should be properly included to get more realistic results (Adams and

Howard, 1986). Since the design storm concept is based on a single event, the antecedent soil condition during the inter-event dry periods cannot be taken into account (Nnadi et al., 1999). If rainfall depth and flood's frequencies are considered equal, proper design storm and antecedent conditions are required in order to obtain acceptable results (Packman and Kidd, 1980). Urbonas (1979) reported that in semi-arid regions or mainly urbanized catchments, the moisture conditions may have an insignificant impact on the flow peaks. Despite its drawbacks, design storm approach can estimate design flow peaks with acceptable levels of accuracy if the design storm characteristics and antecedent conditions are properly selected (Guo and Zhuge, 2008).

Guo and Zhuge (2008) reported that most of the studies carried out to ensure the applicability of design storms, with emphasis on their limitations, were useful for peak flow estimation and not suitable for runoff volume estimation. Thus, the design storm approach may be adequate for the design of urban stormwater quantity controls but not appropriate for the design of urban stormwater quality control practices or major projects (Urbonas, 1979). Conventional urban design storms (where hyetographs are specified but antecedent conditions and computational method are not specified) are reliable for catchments that are largely impervious. However, for the design of storage and quality control practices and for largely pervious catchments more advanced specialized approaches are required (Watt et al., 1986).

In hydrologic engineering, the design storm approach is still the most widely used

design technique due to its simplicity and the unavailability of other simple and feasible alternatives (Benjamin et al., 1999; Guo and Zhuge, 2008).

c. Continuous simulation approach

The rainfall input of the continuous simulation approach is the observed long-term continuous rainfall data at the catchment of interest or surrounding areas for a consistent time step (daily, hourly, 15 min etc.). Usually, daily time increment is used, however, the shorter the time step, the more accurate the results will be (Nnadi et al., 1999). Flood frequency analysis has to be applied to the model outputs in order to determine the design flows.

The use of long series of actual rainfall data allows the model to include temporal rainfall variations. Moreover, the continuous simulation approach is able to add in the antecedent moisture conditions. Even the soil condition during inter-event dry periods can be accounted if the Horton equation, or similar models for infiltration, is used (Nnadi et al., 1999).

On the other hand, the continuous simulation approach is too time-consuming for design purposes (Quader and Guo, 2006; Vaes et al., 2001); although the presence of computers and software packages played a great role in reducing the complexity and processing time of continuous simulations. In fact, the predevelopment flows are not adequate for postdevelopment conditions, thus different trials are required which increase processing times (Marsalek and Watt, 1984; Packman and Kidd, 1980). Consequently, the applicability of the continuous simulation concept is limited due to

time constraints and lack of rainfall data (Quader and Guo, 2006).

The design storm approach is the first approach developed for the design of water management structures and assessment of post-development hydrologic impacts. Even after, the appearance of the continuous simulation approach, the design storm approach is still the most widely used in practice. The continuous simulation and design storm approaches can be used together to reach an optimum combination of time consumption and results' accuracy (Packman and Kidd, 1980). For example, suitable observed flows are not available, synthetic flows can be estimated using continuous simulation approach. These synthetic flows can be used to find the design storm and antecedent conditions that produce comparable flows, based on flood frequency analysis.

d. Analytical probabilistic approach

The analytical probabilistic approach was recently developed to estimate flood frequency distributions for urban catchments. It can produce frequency distributions of flood peaks and runoff volumes to be used for the design of urban stormwater quantity and quality control practices. The recognition of the downsides of the widely used design technique - the design storm approach and the lack of simple and accurate urban stormwater management design techniques were some of the reasons behind the development of the analytical probabilistic approach.

For each return period, the design storm concept produces a peak flow rate from a single rainfall event, assuming that rainfall event and its resulting hydrograph have the

same return period (Adams and Howard, 1986). However, the analytical probabilistic models, based on the derived probability distribution theory, use the probability density functions (pdfs) of rainfall event characteristics along with a rainfall-runoff transformation model to estimate the probability density functions of runoff characteristics (Guo and Adams 1998a, b; 1999a, b).

The derived probability distribution theory as discussed in Benjamin and Cornell (1970) states that the probability distribution functions of a dependent random variable can be determined from the probability distribution functions of independent random variables through the functional relationship between the independent and dependent random variables. It was first used by Eagleson (1972) to estimate the frequency of peak stream flows using pdfs of rainfall intensity and duration, and the Kinematic Wave formula, which represents a functional relationship between peak stream flows and rainfall characteristics.

For the development of the runoff characteristics' pdfs, the rainfall data considered are event-based. The long-term series of rainfall records can be divided into series of single rainfall events. The interevent time definition (IETD), the minimum time of dry period between two successive events, is a criterion used to separate events. Each rainfall event is characterized by its duration (t), volume (v) and interevent time (b), time separating the event to its preceding event, which has to be greater or equal to IETD. In order to facilitate the derivation of the runoff event volume and peak discharge pdfs, rainfall event characteristics are assumed to follow exponential

distributions and rainfall event duration and volume are assumed to be independent.

In the following, the analytical probabilistic and design storm flood frequency estimation approaches are reviewed more in-depth because the former is the focus of this study and the latter is used to test the performance of the newly developed models in this thesis.

1.3. Design storm approach

A design storm is characterized by its rainfall depth, duration, return period (annual exceedance probability or recurrence interval), temporal distribution, areal distribution and computational time step (Wurbs and James, 2002). More details about the important characteristics and their relationships are presented in the following paragraphs.

1.3.1. Rainfall depth

A design storm depth or volume can be estimated from the rainfall data of the watershed of interest or from its surrounding regions. A storm with a specific depth can be presented by different combinations of storm duration and intensity; consequently, it is difficult to construct a realistic storm (Nnadi et al., 1999). Rainfall depth or average intensity over a specific duration is considered instead of peak instantaneous intensity because rainfall event duration is more important for rainfall-runoff transformation (Wurbs and James, 2002).

1.3.2. Design storm duration

Design storm duration is an artificial duration that does not usually coincide with the real rainfall event duration (Cheng et al., 2003) and it is usually selected arbitrarily (Wurbs and James, 2002). Levy and McCuen (1999) reported the following points: (1) there are very few investigations about the selection of design storm duration despite its significant effect on the estimated runoff peak. (2) for most of the hydrologic designs, the selected storm durations are either 24 hours or the catchment time of concentration; (3) generally, the time of concentration that depends on the drainage area may have an effect on the time to hydrograph-peak but not on the storm duration; (4) for a specific return period, the runoff volume and peak increase with the storm duration. Levy and McCuen (1999) estimated design-storm durations for six Maryland watersheds based on the rainfall events that caused the maximum annual discharge, they found that 24-hour storm duration is suitable and reasonable to be used for watersheds with area 5 - 130 km^2. However, for areas less than 5 km^2, the storm duration can be assumed to be the same as the corresponding time of concentration.

1.3.3. Return period

The return period is selected depending on the study's objective and level of safety. For example, a return period equal to 50 years may be suitable for the design of a highway culvert (Wurbs and James, 2002). The rainfall event is characterized by its volume, duration, intensity, time distribution and areal distribution. Each characteristic

may have its own return period, so it is difficult to assign a return period to a rainfall event. The rainfall volume frequency is generally the only one considered (Adams and Howard, 1986; Linsley et al., 1982).

1.3.4. Temporal distribution

The temporal distribution of actual rainfall events is highly random, however, that of the design storm is generally synthetic that does not necessarily coincide with the real one (Urbonas, 1979). The temporal distribution of the design storm is usually considered balanced or nested (Wurbs and James, 2002). The first design storm used was rectangular, which was applied with the rational method to produce the runoff peak (Adams and Howard, 1986). Then other design storm shapes were developed such as triangular storm (Yen and Chow, 1980) and linear/exponential rise and exponential decay or exponential rise and linear decay for early-peaking storms and late-peaking storms, respectively (Watt et al., 1986). These two shapes were developed using generated mathematical models with parameters fitted to the observed data. In addition to their simplicity and applicability on a regional basis, they have less smoothing effects of averaging than the rectangular shape (Watt et al., 1986).

1.3.5. Spatial distribution

The actual storm characteristics are spatially variable, depending on the area of the site in question. Wurbs and James (2002) mentioned that point rainfall measurements

are applicable for areas up to 26 km^2; for larger watersheds, areal adjustments are required. The spatial distribution for large watersheds complicates analyses (Wurbs and James, 2002). On the other hand, the missing information about the storm distributions over space and time affects the rainfall-runoff transformation (Urbonas, 1979).

1.3.6. Intensity-duration-frequency relationship

Intensity-duration-frequency (IDF) curves express the relationship between rainfall event characteristics, i.e., average intensity and duration, for different return periods. IDF curves are produced based on rainfall frequency analysis, or regional analysis if local rainfall data are not available. The common rainfall durations used in IDF curves are 5, 15 and 30 minutes and 1, 3, 6, 12, 18, 24, 48 and 72 hours. The set of return periods used for the development of IDF curves are 2, 5, 10, 25, 50 and 100 years. The most widely used IDF relationships for drainage and stormwater systems are curves, 6 curves in one plot, each curve represents the mean intensity as a function of rainfall duration (Wurbs and James, 2002). IDF curves are widely used for the delineation of design storms.

1.3.7. Design storm delineation

Design storms can be either extracted from the observed data or based on the IDF curves (Adams and Howard, 1986)

a. Design storm delineation based on IDF curves

For a given return period and storm duration, the average value of rainfall intensity is extracted from the Intensity-Duration-Frequency (IDF) curves. Then a geometric form is adopted to describe rainfall pattern such as the rectangular, triangular or linear/exponential storm shapes.

Chicago method (Keifer and Chu, 1997) and the Frequency Based Hypothetical Storm (USACE, 2000) are other IDF-based techniques to generate design storms. However, instead of using a single IDF point, they are based on the entire values representing the intensity versus the duration for a particular return period (Prodanovic and Simonovic, 2004). Since IDF curves do not represent the real rainfall data, methods based on IDF curves produce rainfall events with unrealistic characteristics (Watt et al., 1986).

b. Design storm delineation based on observed data

An actual rainfall event is transformed to a standardized profile - a dimensionless curve of cumulative fraction of total precipitation versus the cumulative fraction of storm time as in the Huff and Soil Conservation Service (SCS) methods. These methods can be used for hydrologic analysis even when return period exceeds 100 years, contrary to IDF-based methods. The main weakness of standardized profile-based methods is the use of temporal smoothing due to the uncertainties of rainfall events' definition and physical variability (Prodanovic and Simonovic, 2004).

1.4. Analytical probabilistic approach

The late seventies and early eighties was the start for the development of analytical probabilistic models for stormwater management. The majority of these models do not explicitly include the soil infiltration capacities and the catchment imperviousness (Guo and Adams, 1998a). For more details about the analytical probabilistic models developed prior to 1998, refer to Guo (1998).

Guo and Adams (1998a, b and 1999a, b) developed closed-form analytical expressions to estimate the probability distributions of runoff event volume and peak discharge rate with and without a detention pond. In these models, which are called Analytical Probabilistic Storm Water Models or APSWM, the soil infiltration capacities and the catchment imperviousness were explicitly considered. In addition, Guo and Adams (1998a, b and 1999a, b) carried out comparisons between APSWM results and continuous simulation (SWMM) results, for a catchment in Toronto, Canada, to test the performance of APSWM and verify the assumptions considered for its development.

The first comparison between analytical, design storm and continuous simulation approaches was carried out by Guo (2001), for a test catchment in Chicago, US. It was found that the three approaches can produce similar results for the design of detention ponds and estimation of flow peaks. However, proper selections of design storm durations and hyetographs are required so that the design storm approach can provide similar results to those of analytical and continuous simulation approaches.

Rivera et al. (2005) reported that the degree of dependency between rainfall event volume and duration is related to the region of interest. They found that APSWM does not produce accurate results where rainfall volume and duration are highly dependent. They proposed alternative models that include this dependency by using average intensity and duration as independent variables.

Chen and Adams (2005) developed closed-form analytical models for evaluating storm water runoff control performance of storage/treatment practices. They used two rainfall-runoff volume transformations: (1) a model that is based on a runoff coefficient and (2) a simplified version of the model proposed by Guo and Adams (1998a), which explicitly use the infiltration term and rate of urbanization. The second models were found to give better results. Behera et al. (2006) developed analytical probabilistic models for water quality.

Quader and Guo (2006) reported that the main differences between analytical and design storm approaches are rainfall input, catchment conditions and treatment of the catchment time of concentration. For the design storm approach, the inputs are design storm duration and hyetograph, the catchment can be divided into subcatchments and the time of concentration can be variable. However, for APSWM, the inputs are rainfall statistics, the catchment parameters and inputs are lumped and the time of concentration is constant. In that paper, APSWM was applied for an actual design case (in Kingston, Canada) in order to compare between APSWM and design storm results and to study the impact of the main differences on the results. Despite these

differences, acceptable results were found.

Chen and Adams (2007) pointed out that the analytical probabilistic models can be developed with different levels of complexity. Two analytical probabilistic models, with and without infiltration term, for the assessment of the urban stormwater runoff volumes were developed. Both models were found to produce acceptable results compared to continuous simulation (SWMM), for two test catchments in the city of Toronto, Canada. It was found that the infiltration term can be omitted if the catchment of interest is largely impervious.

Guo and Zhuge (2008) expanded APSWM by including an analytical probabilistic approach for flood routing through channel reaches and detention ponds. APSWM produced results similar to the results of the design storm approach. They also studied the influence of some design-storms' characteristics - durations and temporal distribution - on the estimation of flow peaks and highlighted the potential problems associated with the application of the design storm approach.

Bacchi et al. (2008) proposed a semi-probabilistic model for storage facility design, rainfall event characteristics are assumed to follow Weibull distributions. Balistrocchi et al. (2009) applied the analytical probabilistic approach to assess the quality of the sewer tank system, where rainfall event characteristics follow Weibull distributions. Guo et al. (2009) added into APSWM the Muskingum-Cunge method for reach routing. The curve-number method for infiltration calculations and areal reduction for large watershed were incorporated into APSWM by Guo and Dai (2009).

Guo and Markus (2011) integrated Clark's unit hydrograph into APSWM. With each new expansion, APSWM was applied under different conditions and was found to provide comparable results to either design storm or continuous simulation approach.

Zhang and Guo (2012a) developed analytical probabilistic models for evaluating the hydrologic performance of green roof systems. Analytical probabilistic models estimating the long-term average storm-water capture efficiency of rain gardens and bio-retention systems were also developed by Zhang and Guo (2012b and 2014, respectively). Guo et al. (2014) revised analytical probabilistic stormwater models for regular catchments so that they can be used for green roofs with irrigation systems.

Previously, only infiltration excess runoff generation was considered; surface runoff from pervious areas occurs after the soil's infiltration capacity is exceeded. Guo and Adams (1998a) incorporated Horton's model in APSWM for estimating the maximum possible infiltration losses. During the Hortonian runoff process, infiltration starts at an initial rate and then decreases exponentially over time to reach an equilibrium for the rest of the event regardless of how long the event lasts. For some events, the soil layer may get saturated before the end of the rainfall event; the maximum allowable infiltration volume will be reached while it is still raining. As a result, some saturation excess runoff may occur during an event, which may have an important effect on the rainfall-runoff transformation process. The saturation excess runoff generation process was recently incorporated into SWMM (Rossman, 2010). Guo et al. (2012) added in APSWM, for the first time, the saturation excess runoff

volume calculations and approximated analytical model to estimate runoff volume frequencies.

1.5. Thesis structure

The objective of this PhD research is to further develop APSWM, which is a promising approach that can provide accurate results in a cost- and time-effective way. For this purpose, three areas were investigated and the results obtained are illustrated in chapters 2, 3, and 4 of this thesis. Chapters 2 and 3 are published already in the journal of hydrological engineering and journal of hydrology, respectively. Chapter 4 was submitted for publication in the Journal of Hydrology in October 2017. Since chapters 2, 3, and 4 are three independent papers, each chapter has its own abstract, introduction, conclusions, and references. Tables and figures are presented at the end of each chapter. Chapter 1, i.e., this chapter, provides an overall introduction which includes an overall literature review. Chapter 5 provides an overall summary, conclusions, and recommendations for future expansions and improvement of the analytical probabilistic approach. References listed at the end of this thesis include only the references cited in chapters 1 and 5.

In most of the existing analytical probabilistic models, the rainfall event characteristics are assumed to follow exponential distributions. This assumption is usually verified visually only; no rigorous statistical analysis was performed. This issue is addressed in chapter 2 in order to provide a more detailed rainfall data analysis that is useful for analytical probabilistic rainfall-runoff transformation

models. More attention is paid to goodness-of-fit tests that are suitable to test the exponentiality of rainfall event characteristics.

For the development of APSWM, the event runoff hydrograph was assumed to be triangular, which may result in an overestimated peak discharge. According to Ponce (1989), the shape of a runoff hydrograph depends on the catchment time of concentration: (1) If the storm duration is less than the time of concentration, the hydrograph can also be assumed to be trapezoidal with lower and upper bases equal to $(t + t_c)$ and $(t_c - t)$, respectively; the catchment response is called subconcentrated. (2) If the storm duration (t) is equal to the time of concentration (t_c), the catchment response is called concentrated and the hydrograph can be assumed triangular. (3) If the rainfall duration is greater than the time of concentration, the catchment response is called superconcentrated and the runoff shape can be considered trapezoidal with lower and upper bases equal to $(t + t_c)$ and $(t - t_c)$, respectively. This proposition of different catchment responses by Ponce (1989) is used in chapter 3 to develop new analytical equations for the frequency distribution of peak discharge rates.

The soil saturation excess was first considered by Guo et al. (2012), where the frequency distribution of runoff volume was derived incorporating some simplifying assumptions. Taking into account the saturation excess runoff volume as estimated by Guo et al. (2012) and the catchment responses proposed by ponce (1989), a more accurate and complete analytical probabilistic model of peak discharge rate is developed in chapter 4.

CHAPTER 2

Exponentiality Test Procedures for Large Samples of Rainfall Event Characteristics

The content of this chapter is the manuscript text published under the following citation:

Hassini, S. and Guo, Y. (2016). "Exponentiality Test Procedures for Large Samples of Rainfall Event Characteristics." *Journal of Hydrologic Engineering, 21(4).*

https://doi.org/10.1061/(ASCE)HE.1943-5584.0001352

Exponentiality Test Procedures for Large samples of Rainfall Event Characteristics

Sonia Hassini and Yiping Guo

Abstract: The main purpose of this paper is to examine and recommend procedures that can be used to statistically test the exponentiality of large sample data of rainfall event volume, duration and inter-event time. Based on literature review and initial analysis, the Poisson and chi-square goodness-of-fit tests were selected first. Some misconceptions about parameter estimators and degrees of freedom associated with the use of the chi-square goodness-of-fit tests were then clarified. Using rainfall data from seven stations in the north-central region of the United States, the choice of the event volume threshold and the minimum inter-event time for separating continuous rainfall data into individual events were examined in detail. Findings from this study suggest that the Poisson test can be used for testing the exponentiality of inter-event times and for examining the statistical independence of consecutive rainfall events. The use of the minimum chi-square estimator combined with the chi-square goodness-of-fit test is recommended for rainfall event volume and duration. An equation which can be used to determine the appropriate number of bins for grouping sample data when conducting the chi-square goodness-of-fit tests is also proposed.

Keywords: Rainfall event; Exponentiality; Large sample size; Poisson test; Chi-square test; Minimum chi-square estimator; Degree of freedom.

2.1. Introduction

In the absence of observed flow data, which is often the case in urban stormwater management practice, rainfall data in some different forms serve as one of the major input data for rainfall-runoff models used routinely for the estimation of flood frequencies. There are three approaches to determining the flood peaks of different exceedance frequencies from a catchment without observed flow data: (1) the design storm approach, (2) the continuous simulation approach, and (3) the analytical probabilistic approach. A design storm has a specific depth and a pre-selected duration. The specific depth of a design storm has a desired exceedance frequency. With the use of a catchment rainfall-runoff model, peak flow resulting from the input of a design storm is assumed to have the same exceedance frequency as the input design storm. Design storms with different exceedance frequencies are constructed based on results from statistical analysis of historical rainfall data, focusing only on large rainfall events. Flood peaks of different exceedance frequencies can be estimated by running the catchment rainfall-runoff model with the input of design storms of different exceedance frequencies. This constitutes the design storm approach.

In the continuous simulation modeling of the catchment rainfall-runoff processes for stormwater management purposes, rainfall-runoff transformation calculations are often performed using as input the recorded continuous rainfall data over a long period of time (e.g., 50 years or more) at a location of interest. Frequency

analysis on the output flow values is conducted subsequently to estimate the flood peaks of different exceedance frequencies. This continuous simulation approach does not require the statistical analysis of the historical rainfall data but conducts similar statistical analysis on each set of the simulation results. It is therefore much more time-consuming than the design storm approach. The analytical probabilistic approach was developed to overcome some of the shortcomings of both the design storm approach and the continuous simulation approach. Using the analytical probabilistic approach, the probability distributions of runoff characteristics (peak flow and runoff volume) are derived directly from the probability distributions of rainfall characteristics (Eagleson 1972; Howard 1976; Adams et al. 1986; Guo and Adams 1998a; Guo 2001; Bacchi et al. 2008; Balistrocchi et al. 2009). Thus the input rainfall data has to be represented by some theoretical probability distributions, where the exponential distribution is the most widely used theoretical distribution (Eagleson 1972, 1978; Howard 1976; Adams et al. 1986; Guo and Adams 1998a; Adams et al. 2000; Guo 2001; Guo and Baetz 2007; Zhang and Guo 2012a, b). Also, the exponential distribution is the only theoretical distribution that would result in closed-form analytical solutions from the application of the analytical probabilistic approach.

The advantages of using exponential distributions to represent local rainfall event characteristics for urban stormwater management purposes have been recognized for a long time. For example, early in 1986, the US Environmental Protection Agency (USEPA) recommended the possible use of exponential

distributions for the design of stormwater quality control detention ponds (USEPA 1986). Driscoll et al. (1989) analyzed rainfall data and reported parameter values of exponential distributions fitted to stations across the US. Wanielista and Yousef (1993) explained the possible use and listed the exponential distribution parameter values for rainfall stations across the US; while Adams and Papa (2000) listed the exponential distribution parameter values for rainfall stations across Canada. These are all done because of the simplicity and usefulness of exponential distributions. However, for a specific location of interest, proper procedures that can be used to statistically test the exponentiality of its historical rainfall data have not been investigated in detail and are therefore a necessary and worthwhile subject of study. Once the exponentiality of rainfall event characteristics is established for a specific location, the use of the analytical probabilistic approach in that location would expedite dramatically the modeling processes required for stormwater management planning and design.

The detailed rainfall hyetograph observed at a point changes from event to event. Nevertheless, the main characteristics of a point-observed rainfall event and its temporal relation to other neighboring events can be very well represented by its rainfall event volume, rainfall event duration, and inter-event time which is the dry time before it occurred. The rainfall event volumes of all the historically observed and statistically independent individual events can be viewed as realizations of a random variable; the rainfall event durations of all the historically observed and statistically

independent individual events can be viewed as realizations of another random variable; the same for all the individual inter-event times. Statistically, these historically observed rainfall event volume, duration, and inter-event time data are all left bounded by zero and often highly skewed to the right. They contain frequent low values and infrequent high values. For stormwater management purposes, these low and high values are both of interest because they affect either stormwater quantity or quality or both. That is why some investigations have been conducted to see if rainfall event characteristics of all possible events occurring at a location of interest can be represented by some suitable theoretical probability distributions. For instance, the US Environmental Protection Agency (1986) prepared a report for urban runoff quality control assuming that rainfall event characteristics follow either exponential or Gamma distributions. Guo and Adams (1998a, b) developed analytical probabilistic stormwater management models (APSWM) using the following exponential probability density functions to represent rainfall event characteristics of all possible rainfall events (including small, medium, and large events):

$$f_V(v) = \zeta e^{-\zeta v}, \qquad v \geq 0 \tag{1}$$

$$f_T(t) = \lambda e^{-\lambda v}, \qquad t \geq 0 \tag{2}$$

$$f_B(b) = \psi e^{-\psi b}, \qquad b \geq 0 \tag{3}$$

In Eqs. (1) - (3), V (mm), T (h) and B (h) are the continuous random variables of

rainfall event volume, duration, and inter-event time, respectively; v, t, and b are the corresponding specific values (or individual observations) of these random variables. The symbols ζ, λ, and ψ are the distribution parameters for rainfall event volume, duration, and inter-event time, respectively.

It has been found that APSWM can be more accurate as compared to the design storm approach and computationally more efficient as compared to both the design storm and continuous simulation approaches (Guo 2001). The exponential distributions expressed in Eqs. (1) – (3) are the simplest theoretical distributions and may not be widely suitable for all locations. For example, for some rainfall stations in Italy, Bacchi et al. (2008) and Balistrocchi et al. (2009) found that Weibull distribution is more appropriate for rainfall event volume.

For the whole set of all possible rainfall events, the exponentiality of rainfall event characteristics was first proposed and tested by Eagleson (1972). Several earlier studies found that rainfall event characteristics often follow exponential distributions (e.g., Eagleson 1972, 1978; Howard 1976; Adams et al. 1986; Guo and Adams 1998a; Guo 2001). The method of moments or the method of maximum likelihood along with visual comparisons of theoretical and observed histograms was mainly used in the past to fit observed rainfall data to exponential distributions. Although rigorous statistical techniques such as goodness-of-fit (GOF) tests may provide greater credibility of a theoretical model, they remain under-employed when fitting theoretical distributions to rainfall event characteristics of all possible events. This is probably due to the fact

that the sample sizes are often large when dealing with the whole set of all possible rainfall events and high rejection rates are often associated with large samples. In previous studies where statistical testing tools were used for rainfall event characteristics, the sample sizes were still not that large or only the distribution of the inter-event time was tested. For instance, Restrepo and Eagleson (1982) found that inter-event times follow exponential distributions for a wide range of geographic areas, using Poisson process testing with record lengths varying between 0.75 and 27 years. Guo and Baetz (2007) applied both Poisson process and Kolmogorov–Smirnov (K-S) tests to examine the exponentiality of rainfall event characteristics, but only summer rainfall data were used with the largest sample size of 795.

Recently, the volume-based hydrology has emerged as a new practice for stormwater management; it is partially adopted by various local jurisdictions in the US such as Philadelphia, PA and Phoenix, AZ (Reese 2009). The target of using volume-based hydrology is to reduce the post-development runoff volumes, instead of flow peaks, to pre-development levels. Reese (2009) advocates the shift to a volume-based approach in urban hydrology in order to better control stormwater quality and quantity. More details about the volume-based approach and its application for green infrastructure design can be found in Reese et al. (2010). In order to use runoff volume as the stormwater control criterion, it is also necessary to fit suitable probability distributions to rainfall event characteristics of all possible events, not just the extreme large events as in the construction of design storms for flood control

purposes.

The purpose of this paper is to provide some practice-oriented guidelines to statistically test the exponentiality of rainfall event characteristics of a specified set of rainfall data. These guidelines are especially suitable for cases with extremely large samples as often encountered in analyzing all possible rainfall events. More insights about graphical comparison versus formal statistical tests are given. Long-term recorded hourly rainfall data of selected stations from the North Central Region of the US are used as examples. This selection also serves the purpose of getting a general idea, but not to generalize, about the exponentiality of rainfall event characteristics of this region.

2.2. Rainfall Data and Method of Analysis

2.2.1. Area and Sources of Data

The area of study consists of seven States, one station per State (Table 1). Stations with the longest continuous hourly precipitation data were chosen. The number of years of record is 53 (1949-2001) and the percentage of coverage is 100% (i.e., 0% of data is missing) for all stations except for the Huron Airport station in South Dakota and the Dodge City Regional Airport station in Kansas, where the coverage is 99%. The selected states are adjacent to each other and located in the west part of the north central region of the USA.

This study focused on rainfall only and snow months were excluded based on the daily minimum temperatures. According to Ruffner and Bair (1978) and the National Climatic Data Center (2011), the months of study (i.e., rainfall months) are from March through November for the Springfield Regional Airport station in Missouri and from April through October for the rest of the stations.

2.2.2. Data Transformation

In order to statistically analyze rainfall event characteristics, a continuous rainfall series has to be discretized into individual rainfall events first. For small urban catchments where runoff is mainly generated from impervious areas and the routing effect of detention ponds is accounted for separately from the catchment itself, the input of statistically independent rainfall events would almost always result in statistically independent runoff events from the catchment as long as the catchment's time of concentration is shorter than the dry periods between rainfall events. The identification of statistically independent rainfall events (and therefore statistically independent runoff events) can be done by either considering the catchment size and other related physical characteristics or relying only on the rainfall data (Bonta and Rao 1988). In this study we focus on the rainfall data ensuring the statistical independence between rainfall events. A minimum inter-event time, referred to as the inter-event time definition - IETD (Guo and Adams 1998a; Adams and Papa 2000), also known as the minimum number of dry hours [U. S. Environmental Protection

Agency (USEPA) 1986] or the critical duration (Bonta and Rao 1988), can be used as the criterion to segregate continuous rainfall series into separate events. If the dry time separating two rainfall episodes is less than the selected IETD, then these two rainfall episodes belong to the same rainfall event; otherwise they are parts of two consecutive rainfall events. Each separated rainfall event is characterized by its volume (v, also known as event depth), duration (t), and the inter-event time (b) which is the dry time separating the event of interest from its preceding event.

Several methods were implemented in the past to assist in determining the most suitable IETD; for example, determination based on autocorrelation, rank correlation, and Poisson process testing (also known as the exponential method). The rank correlation and the Poisson process testing are the most objective methods for ensuring the statistical independence between events (Bonta and Rao 1988). The Poisson process testing and analysis (more details are given in the following section) were selected because they result in independent events and exponential inter-event time distributions (Restrepo and Eagleson 1982). Poisson process analysis was also selected because of its relative simplicity and practical usefulness. It is recommended by Bonta and Rao (1988) because it uses actual dry-period data to estimate the most suitable IETD as well.

Detailed IETD specification is discussed later; however, because of the intended applications in urban stormwater management, IETD is set to (1) a minimum of 6 hours to ensure the separation of runoff events generated from small urban catchments

as a result of consecutive rainfall events, and (2) a maximum of 12 hours to avoid the combination of rainfall episodes that occur too far apart (e.g., 12~13 hours apart) into a single rainfall event. The 12-hour maximum IETD is set because urban catchments usually have a time of concentration less than 12 hours and point-observed rainfall episodes 12 hours (half-a-day) or longer apart usually result from unrelated meteorological processes and should therefore be treated as separate events. However, for mainly larger catchments where the time of concentration is much longer, longer IETDs may still be considered to ensure the segregation of resulting runoff events.

2.2.3. Poisson Processes and Tests

The Poisson process is a counting process where the number of arrivals within a finite interval of time follows a Poisson distribution. One of the characteristics of the Poisson process is that the arrivals of events are relatively rare. According to the law of rare events, the total number of rare events within a finite interval of time must be, approximately, Poisson distributed (Cameron and Trivedi 2013). One other important Poisson process property is that the inter-arrival times between the relatively rare events are independent and identically distributed exponential random variables (Ross 2007). Because of these desirable properties, the Poisson process has been used to describe rainstorm arrivals (Waymire and Gupta 1981a, b, c; Restrepo and Eagleson 1982; Keim and Cruise 1998; Guo and Baetz 2007; Zhang and Guo 2012a, b).

The Poisson test was first proposed by Cunnane (1979) and applied by other researchers such as Ashkar and Rousselle (1987) and Cruise and Arora (1990) in partial duration series modeling, mainly dealing with flood peaks. This test is based on the equi-dispersion (i.e., equality of variance and mean) characteristic of the Poisson distribution. In details, the ratio R known as the index of dispersion and defined by

$$R = \frac{Var[m]}{E[m]} \tag{4}$$

will approach unity if the annual number of events is Poisson admissible. In Eq. (4), m is the number of events per year; $Var[m]$ and $E[m]$ are the variance and expectation of m, respectively.

To statistically test whether the annual number of events is a Poisson random variable, Cunnane (1979) suggested the use of the Fischer dispersion test statistic which is given by

$$d = (M - 1)R \tag{5}$$

where M is the number of years of record. The statistic d is χ^2-distributed with $(M - 1)$ degrees of freedom.

Restrepo and Eagleson (1982) applied the above-described Poisson test and focused on the distribution of inter-arrival times of rainfall events, known in Poisson series as the waiting time distribution (Ross 2007). They suggested the use of Poisson test as a sufficient statistical technique for the determination of suitable IETDs to divide rainfall series into statistically independent storms, under the assumption that the average rainfall event duration is negligible compared to the average inter-event

time. It was also applied by Bonta and Rao (1988), Guo and Baetz (2007) and Zhang and Guo (2012a, b). They all assumed that if the annual number of rainfall events is Poisson-distributed then the rainfall events are independent and the inter-event times are exponentially distributed.

It is worth mentioning that there are other tools such as the Allan factor that can be used to test the Poisson assumption (Serinaldi 2013). Similarly to the index of dispersion, the Allan factor is a variance to mean ratio; here, the Allan variance is used instead of the classical variance. The Allan variance is expressed in terms of the variability of successive counts. As the same for the index of dispersion, the Allan factor is equal to unity for homogeneous Poisson processes (Serinaldi 2013). Serinaldi (2013) studied the relationship between the Allan factor and the index of dispersion and their effectiveness for testing the Poisson hypothesis. It was found that (a) the performance of these tools depends on the application; (b) the index of dispersion is preferred when equi-dispersion is the aim of the test; and (c) the Allan factor is generally more unbiased when the events considered in the study are very rare such as extreme storms, which occur very infrequently. In this study the index of dispersion is selected since the tested data contain all the events that occur annually, about 20~35 events per year.

As for the data or criteria required to determine a suitable IETD, Bonta and Rao (1988) highlighted some important points to consider. First, a minimum of 10 years of record is recommended for an adequate estimate of the suitable IETD. Second,

rainfall measuring interval has a minimal effect on IETD estimation which implies that hourly precipitation data are sufficient. Finally, the suitable IETD depends greatly on the change of weather from year-to-year, seasons of a year and location. Thus the suitable IETD should be determined for each set of data and station of interest when it is possible.

To avoid over- and under-dispersions (i.e., the variance exceeds the mean or vice-versa, respectively), a two-tailed Poisson test was used in this study. The significance level is chosen to be 10%, thus a *p-value* between 5% and 95% results in an accepted hypothesis.

2.2.4. Goodness-of-Fit Tests

Probabilistic goodness-of-fit (GOF) test is an important tool for assessing the closeness of a theoretical model in representing the frequency distribution of observations (Ott 1995). Choosing a suitable GOF test is, however, a tedious task, especially when dealing with large sample sizes. A sample size is considered large if it satisfies the central limit theorem, which states that the average of a large enough sample of a random variable will follow approximately a normal distribution regardless of the distribution followed by the random variable itself. In practice, generally, the minimum sample size that can be viewed as sufficiently large is 30 or less for random variables following symmetrical distributions and 200 for random variables following highly skewed distributions (Clarke and Cooke

1978). Nevertheless, the larger the sample size, the better it represents the population. In this study, the sample sizes depend greatly on the IETD selected and vary between 1185 and 4178.

For large samples, almost any model used to describe the distribution of observed data is rejected. However, for small samples different models may be equally accepted to represent the same set of data (Bentler and Bonett 1980). On one hand, large sample sizes increase the power of GOF tests and give better parameter estimates. On the other hand, large sample sizes increase the sensitivity of GOF tests and small differences between theoretical and observed distributions may be detected as statistically significant. Moreover, the number of anomalies in data caused by measurement and data transformation errors increases with the increase of sample size, which in return increases the rejection rate of the tested hypothesis. It is extremely difficult to fit a model with significance level higher than 5% for large samples (Bentler and Bonett 1980). Our own initial analysis showed that when GOF tests such as Kolmogorov–Smirnov (K-S) and Anderson-Darling (A-D) tests are applied to rainfall event characteristics, the rejection rate is very high due to the large sizes of samples (p-values are zeros). In order to minimize this effect and still take advantage of large sample sizes, statistical analysis of grouped data is preferred.

There are two types of statistical tests: parametric versus nonparametric. Parametric tests have more statistical power than their counterparts. Fortunately, non-parametric tests are only slightly less powerful than the parametric ones when the

sample size is large. On the other hand, parametric tests are less robust due to the required assumptions; and these assumptions may be violated easily in the time series of hydrological events (Keim and Cruise 1998). But based on the central limit theorem, parametric tests work well with large samples even if the population is non-Gaussian. Therefore for large samples it does not really matter whether the test is parametric or nonparametric.

Since we have large samples, the nonparametric and widely used chi-square GOF test is selected to test whether a sample fits an exponential distribution. The Chi-square test requires the sample data to be grouped into individual bins, making it less sensitive to small differences. However, the chi-square GOF test results are very sensitive to the number and width of bins; which are often selected with some arbitrariness. There are different rules suggested by various researchers that may be followed for estimating the most suitable number of bins (k) for a sample of size n (e.g., Sturges 1926, Mann and Wald 1942, Williams 1950, Cochran 1952, White et al. 2009). The different rules produce widely different estimates and there is no consensus as to which rule is the best for specific cases. It is therefore necessary for us to investigate which rule to follow for our specific data. It is a common practice to set the bin widths so that the bins are equally spaced or occurrence of events within individual bins is equally probable. For example, Cochran (1952) believes that a moderate number of bins with equal widths should be used when fitting a continuous distribution to observed data. We started by using the following equation as the rule

for setting the number (k) of equi-width bins:

$$2^k \leq n \leq 2^{k+1} \tag{6}$$

This rule was suggested by an experienced statistician (Narayanaswamy Balakrishnan, McMaster University, personal communication, February 22, 2013) based on his own practical experience. Eq. (6) can be rewritten as:

$$log_2(n) - 1 \leq k \leq log_2(n) \tag{7}$$

The above equation is close to the one below which was suggested by Sturges (1926):

$$k = log_2(n) + 1 \tag{8}$$

After several trials using different rules to select the number and width of bins, equal width and Eq. (9), which is a combination of Eqs. (7) and (8), were found appropriate and were applied as the rule in this study:

$$log_2(n) - 1 \leq k \leq log_2(n) + 1 \tag{9}$$

Also as a rule of thumb, each bin should contain no less than five observations; if a bin has less than five observations it is regrouped to its adjacent bin.

The value of the chi-square test statistic is

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{10}$$

where k is the number of bins, each with at least 5 observations. O_i and E_i are the observed and expected frequencies of occurrence of data bounded by bin i,

respectively; and

$$E_i = n[F(x_2) - F(x_1)] \tag{11}$$

where $F(x_1)$ and $F(x_2)$ are the theoretical cumulative distribution function (CDF) values evaluated at x_1 and x_2, respectively; and x_1 and x_2 are the lower and upper limits of bin i, respectively.

The statistic χ^2 in Eq. (10) follows a chi-square distribution with $(k-1-s)$ degrees of freedom, where s is the number of distribution parameters estimated from the sample. For exponential distributions tested in this study, s is equal to one. Chernoff and Lehmann (1954) illustrated that the number of degrees of freedom is equal to $(k-1-s)$ only if the estimators for the parameters are efficient (i.e., with minimum variance); otherwise, it lies between $(k-1-s)$ and $(k-1)$. Originally, when using the chi-square GOF test, the minimum chi-square (MCS) estimator was suggested as an efficient estimator for the parameters of the theoretical distributions (Harris and Kanji 1983). However, the maximum likelihood estimator (MLE) is widely used for almost all cases (Harris and Kanji 1983). For exponential distributions, based on either the MLE or the method of moments (MOM), which is also commonly used, the value of the one parameter is equal to the reciprocal of the sample mean. Although the accuracy of this estimate increases with the sample size, the MCS estimator is still recommended by many researchers (Berkson 1980; Harris and Kanji 1983).

In this study, the issues of parameter estimation and degrees of freedom (df) are

specifically investigated. In order to have a better look on the results, we performed the chi-square GOF tests where the one distribution parameter is estimated from the samples using both MOM and MCS and the df is taken as $(k-1-s)$, i.e., $(k-2)$. In addition, we tested the exponentiality of random variables v and t with parameters estimated using MOM but df taken as $(k-1)$, assuming that MOM is not efficient when used with the chi-square GOF test.

The one parameter exponential distribution is sensitive to the more frequent events which makes it relatively inflexible to fit for some of our observed data (Cruise and Arora 1990). Guo and Adams (1998a) proposed to neglect rainfall events with volumes less than a volume threshold (v_t), which do not usually produce runoff, in order to improve the goodness of fit. These extremely small events have minor significance and may even be a result of measurement errors (Bacchi et al. 2008). They may be lost in the rainfall-runoff transformations because of initial abstractions resulting from interception and depression storages (Bacchi et al. 2008; Balistrocchi et al. 2009). For an urban environment, 1 to 5 mm was considered as a suitable volume threshold (Balistrocchi et al. 2009). Other researchers (Guo and Adams 1998a; Guo and Baetz 2007; Zhang and Guo 2012a, b) also neglected small events with volumes less than 5 mm and reached satisfactory results. The threshold value affects the inter-arrival times and their probability distributions as well (Beguería 2005). With the adoption of a volume threshold, both the duration of the neglected event and its preceding inter-event time are added to the next inter-event time, and the number of

events decreases.

2.2.5. Procedure of Analysis

Hourly rainfall data from each of the stations listed in Table 1 are subjected to the following steps of analysis:

(1) Select an IETD (6 - 12 h, if it is not specified for a particular application) and select a volume threshold v_t (0 - 5 mm).

(2) Separate the continuous rainfall record into individual events according to the selected IETD and then remove rainfall events with volumes less than v_t.

(3) Apply Poisson test as described above to test the independence of the remaining events and the exponentiality of inter-event time b.

(4) Apply chi-square GOF test for event volume v.

(5) Apply chi-square GOF test for rainfall event duration t.

During testing (i.e., steps 3 - 5), the passage to the next step requires positive results from the current step. If the exponential distribution hypothesis is rejected, the analysis is restarted again from step 1 with a different combination of IETD (if it is not specified for particular application purposes) and v_t. IETD and v_t are increased gradually and separately. If the distribution hypotheses are rejected for all pairs of possible IETD and v_t values as prescribed in step 1, the pair closest to passing is recorded. In addition, the exponential distribution may not be the best model to fit the

data, thus it is necessary to check other possible distributions such as gamma, Weibull and lognormal.

The above procedure is designed for the separation of a set of rainfall series into independent events and for testing to see if exponential distributions can represent the frequency distributions of rainfall event characteristics. The resulting distribution models for rainfall event characteristics can be used with the analytical probabilistic models developed for stormwater management planning and design purposes (Adams and Papa 2000); similar to using the entire set of rainfall data with continuous simulation models or individual design storms as with the design storm approach. Although not widely applied yet, the resulting distribution models for rainfall event characteristics can also be used to provide frequency estimates of rainfall events of different magnitudes. Reliable and robust frequency estimates may be made after validating the fitted exponential models using, e.g., split-sample, jackknife or bootstrap methods (McCuen 2003). To illustrate an example of validation tests, split-sample tests are applied to rainfall event volume data for the seven stations. For each station, the sample data is divided randomly into two sets: training and testing sets. The training set is used to fit an exponential model and estimate its parameter; the testing set is used to see if it follows the fitted model.

2.3. Results and Discussion

2.3.1. Poisson Tests

After analyzing the effects of IETD and v_t on the Poisson test results, we noticed that the *p-value* increases with the increase of IETD for a fixed v_t. The perfect *p-value* for the Poisson test is 0.5; *p-values* below or above 0.5 imply that there is, respectively, over- or under-dispersion. For a fixed IETD and increasing v_t , the *p-value* does not change monotonically, which makes testing results unpredictable. It is therefore not advisable to primarily rely on v_t to achieve desired results. We noticed that once the Poisson test results are positive for a specific IETD value, they become unlikely to be rejected when changing v_t, although the exact *p-values* differ. However, when the exponential distribution hypothesis about the inter-event time is rejected with a *p-value* close to the significance level, it is more likely to be accepted if we change v_t a little bit. Thus it is recommended that the most appropriate IETD should be found first with zero v_t and then different v_t values may be tried to reach an acceptable or optimum result.

Table 2 contains part of the Poisson test results for cases without neglecting small events (i.e., $v_t = 0$). It can be concluded that the Poisson tests may be accepted with IETDs varying from 6 to 12 hours for three stations; i.e., Fargo, ND, Huron, SD and Doge, KS. IETD ranging from 8 to 12 and 10 to 12 hours were found suitable for Springfield, MO and Des Moines, IA, respectively. For St. Cloud, MN and North Platte, NE, 12-hour IETD is needed to ensure acceptable tests. As these *p-values*

account for all the events (i.e., v_t = 0), they may improve when increasing v_t. It is also noticeable from Table 2 that IETD has an effect on the average annual (excluding snow months) number of events (θ) and the statistics of rainfall event characteristics $\bar{v}$, $\bar{t}$, and $\bar{b}$, which are the averages of v, t, and b, respectively. These effects differ from station to station; on average, θ decreases by 6 to 14 events, $\bar{v}$, $\bar{t}$ and $\bar{b}$ increase by 1 - 2.5 mm, 1.7 - 3.1 hours and 10.3 - 15.5 hours, respectively, when IETD changes from 6 to 12 hours. The average event duration $\bar{t}$ is the most sensitive one to IETD; by changing the IETD from 6 to 12 hours, $\bar{t}$ changed by 35 to 53% and the relative changes of other statistics are from 10 to 20%.

The volume threshold v_t has a great effect on θ, $\bar{v}$, $\bar{t}$, and $\bar{b}$. The increase of v_t causes θ to decrease, since events with volumes less than v_t are eliminated; consequently $\bar{v}$ increases. The increase of v_t also results in longer $\bar{b}$, because the durations and the inter-event times of the omitted events are added to the inter-event times of their neighboring events. Since our goal is to reach exponentiality for rainfall event characteristics whenever possible, IETD and v_t values with accepted Poisson and GOF tests are reported in Table 3. In order to obtain acceptable chi-square GOF tests, the IETD and v_t values reported in Table 3 were increased slightly from those just for achieving acceptable Poisson tests. By comparing Tables 2 (IETD = 12 h) and 3, we noticed that the *p-values* improved when v_t is increased for all the stations except Doge, KS and Springfield, MO. When we increased v_t from 0 to 3.05 mm the *p-values* of St. Cloud, MN changed positively and the *p-values* of Doge, KS changed

negatively, both of them have the same IETD of 12 hours. This implies that the effect of v_t on testing results varies from station to station, which complicates the decision process when dealing with different locations.

The statistics given in Table 3 corresponds to the final combinations of IETD and v_t shown in Table 3 that resulted in acceptable or optimum results. These statistics, compared to those in Table 2 (IETD = 12 h), confirm the great effect of v_t on θ, $\bar{v}$, $\bar{t}$, and $\bar{b}$. The average annual number of events per station decreased on average by 40 and 43% for a v_t of 2 and 3 mm, respectively. When v_t is set as 4 mm for Fargo, ND and 5 mm for Springfield, MO, θ decreased by 56 and 45%, respectively. It can be seen that a large number of rainfall events are small and eliminating them improves the goodness of fit. The elimination of these small events resulted in an increase of $\bar{v}$, $\bar{t}$, and $\bar{b}$ by about 40 - 100, 20 - 65 and 45 - 115%, respectively. The percentages of change of θ, $\bar{v}$, $\bar{t}$, and $\bar{b}$ differ from station to station.

As the stations are from the same region, the statistics are not that diverse except for some noticeable peaks. For instance, Des Moines, IA has the largest θ (Table 3) because of its lowest IETD and lowest v_t (Table 3). It is also noticeable that Springfield's θ, $\bar{v}$ and $\bar{t}$ (Table 3) are larger than those of the other stations which have the same IETD (12 hours). The large rainfall event statistics at Springfield are mainly due to its highest v_t and the additional data of the extra two months; March and November, as compared to the other stations. Fargo's $\bar{b}$ is the highest (Table 3) which can be partially explained by its high v_t (Table 3).

Generally, wide ranges of IETD and v_t values resulted in Poisson-distributed annual number of events; i.e., statistically independent rainfall events and exponential inter-event times. These wide IETD and v_t ranges were narrowed due to the need to fit the data of rainfall event volume and duration to exponential distributions. The data of the region of study tend to have a suitable IETD of about 12 hours except for Des Moines, IA, which requires an IETD of 10 hours. The volume threshold varies between 2 and 5 mm. The results are statistically significant with *p-values* remarkably higher than the significance level of 5% for all the stations except for North Platte, NE (Table 3), which is still considered acceptable.

For volume thresholds less than 5 mm, similar suitable IETD estimates were found by several investigators for other locations in the USA. Guo (2001) found that an IETD of 12 hours is suitable for Chicago, Illinois, based on comparison of theoretical and empirical cumulative frequency curves. Using Poisson and K-S tests, Guo and Baetz (2007) fitted exponential distributions to rainfall event characteristics for the summer months with an IETD of 10 and 12 hours for Phoenix, Arizona and Chicago, Illinois, respectively. With Poisson tests and histogram comparisons, Zhang and Guo (2012a, b) found that an IETD of 8 hours for Detroit, Michigan and Atlanta, Georgia and an IETD of 12 hours for Flagstaff, Arizona produced acceptable results.

2.3.2. Chi-square GOF tests

For all the stations, the values of the exponential distribution parameter for rainfall event volume (ζ) estimated by MCS are slightly higher than those estimated by MOM. A proportional increase of the *p-value* also occurs when ζ is estimated by MCS rather than MOM assuming a df of $(k-2)$, the lowest such increase is 46.1% (Table 4). Similarly, the estimates of the exponential distribution parameter for rainfall event duration (λ) also slightly increased for all the stations, except North Platte, when using the MCS estimator rather than the MOM ($df = k-2$); the corresponding *p-values* all increased with the highest increase of 143.3% (Table 4).

Using the same values of ζ and λ estimated by MOM but using $(k-1)$ as the df, we noticed an increase of *p-values* by up to 122% for rainfall event volume and 74.5% for rainfall event duration as compared to those also from MOM but using $(k-2)$ as the df (Table 4). These *p-values* are still generally less than those corresponding to the MCS estimator for rainfall event volume (Table 4). However, for rainfall event duration, the estimated *p-values* with MOM and a df of $(k-1)$ are generally slightly higher than those corresponding to the MCS estimator (Table 4).

For chi-square GOF tests, the optimum *p-value* is one; with a fixed sample size, the higher the *p-value*, the better the fit. For a significance level of 0.05, referring to Table 4, the exponentiality hypothesis of rainfall event volume is accepted for (1) only one station (Des Moines, IA) based on MOM with $df = k-2$, (2) three more stations when using MOM with $df = k-1$, and (3) all the stations except St. Cloud,

MN, but its *p-value* is close to the significance level when using MCS. The exponentiality of the rainfall event duration is accepted for (1) three stations using MOM with $df = k - 2$, (2) all the stations when using MOM with $df = k - 1$, and (3) all except Doge, KS (*p-value* = 0.046) when using MCS (Table 4).

Generally both ζ and λ values estimated by MOM are lower than their counterparts estimated using MCS, with ζ more affected than λ. Although the differences between MOM and MCS parameter estimates are small, the changes of *p-values* are relatively large. Since the parameter estimates have a recognizable effect on the *p-values*, an efficient and suitable parameter estimator is highly recommended. In addition, the *p-values* of MOM with a df of $(k - 1)$ are closer to those from the MCS as compared to the *p-values* of MOM with a df of $(k - 2)$. These results suggest that the MOM is not a good estimator to be used with the chi-square GOF test, especially for rainfall event volume. If MOM is used, the loss of one degree of freedom due to parameter estimation from observed data should be neglected. The MCS estimator is obviously the preferred parameter estimator.

We also noticed that higher values of IETD would more likely result in accepted chi-square GOF tests, particularly when certain small events are omitted. This can be explained by two facts: (1) the size of the sample becomes smaller, and (2) the statistical independence between rainfall events is more likely achieved when longer IETDs are used (Bonta and Rao 1988). This is confirmed in this study (Poisson test results) for an IETD up to 12 hours. In fact, positive dependence between observations

causes GOF tests that assume independence to more likely reject a true hypothesis about the distribution of these observations (D'Agostino and Stephens 1986, chapter 3).

2.3.3. Validation Tests, Alternative Distributions and Seasonal Frequency of rainfall Events

For the split-sample validation tests of rainfall event volumes, the parameter value is assumed to be the inverse of the average volume. The *p-values* are determined assuming that the degree of freedom is equal to the number of bins minus one (i.e., k - 1). The *p-values* (Table 5) show that the validation tests are positive for all the stations except the MN and MO stations; where either the training or the testing set is not exponential. As rainfall event volume was found to be the most likely rainfall event characteristic that cannot be fitted to exponential distributions, the split-sample testing results seem to suggest that the recommended statistical testing procedures are largely reliable for the different locations and sample sizes.

The goodness of fit of rainfall event volume and duration are also tested for different distributions: exponential, gamma, lognormal and Weibull using the chi-square test; K-S and A-D tests produced zero *p-values* as mentioned previously for the exponential distributions. The chi-square tests used in this part of analysis are conducted with the following procedures: (1) the number of bins are estimated using the Sturges formula, (2) the bins are of equi-intervals, each interval contains no less

than five elements, (3) the degrees of freedom are set to be equal to the number of bins minus 1, i.e., $(k-1)$, and (4) the least squares estimation (LSE) is used to estimate the parameters of the 2-parameter Weibull distributions, while the method of moments (MOM) is used to estimate the parameters of the 1-parameter exponential and 2-parameter gamma distributions, and for lognormal distribution, maximum likelihood estimation (MLE) is used. These procedures are slightly different from those used early in this paper which results in different *p-values* when testing the exponential goodness of fit. This is acceptable here since these procedures are used to only get a quick and general idea about the fitting of different distributions.

The above-described GOF tests are applied to the seven stations and the results (Table 6) show that the lognormal and gamma distributions fit better than the exponential distributions for rainfall event volume. For rainfall event duration (Table 6), the gamma distribution is superior followed by the exponential distribution, and the lognormal distribution is always rejected. The Weibull distribution is rejected for both rainfall event volume and duration for all the stations (Table 6). These results are approximate because of the use of the same number of degrees of freedom $(df = k - 1)$ to test the GOF of 1-parameter distributions (the exponential distribution) and 2-parameter distributions (the gamma, lognormal and Weibull distributions). The use of the same *df* is in favor of the 2-parameter distributions, and as shown above the *df* has a great effect on the *p-values*. Nevertheless, some of the exponential GOF test results are close to those of the best-fit distributions. Moreover, when the gamma distribution

is the best fit with a shape parameter close to one, this gamma distribution is very close to and can be approximated by an exponential distribution. What we did find is that, for rainfall event volume, the shape parameters of the gamma distributions fitted for the stations studied here are all fairly close to one, with the smallest being 0.93 and the largest being 1.2. Exponential distributions are special case gamma distributions with a shape parameter of unity. In fact, exponential distributions are also special case Weibull distributions, however, similar results between exponential and Weibull was not found here. This is because the least squares parameter estimation for Weibull distributions resulted in shape parameters ranging from 1.33 to 1.59, much farther away from the required unity. The Weibull distributions are therefore always rejected. The above-described results may serve as an additional evidence supporting the consideration of exponential distribution as an alternative distribution for rainfall event characteristics.

To minimize the effect of the removal of small rainfall events on the goodness of fit, a location parameter equaling to v_t can be considered. To check the effect of a location parameter on GOF test results, the 2-parameter exponential, 3-parameter gamma, 3-parameter Weibull and 3-parameter lognormal distributions were tested; the maximum likelihood parameter estimation is used to determine the parameter values of all the distributions, the degrees of freedom (df) is taken as the number of bins (k) minus one ($df = k - 1$) and k is estimated using Sturges' rule. As shown in Table 7, based on the Chi-square GOF test results, only the 3-parameter Gamma is not

rejected; for some stations, the 3-parameter Weibull distribution is also not rejected if the K-S test results are used. Table 7 shows that adding a location parameter to the exponential distribution does not improve its goodness of fit, whereas adding a location parameter to the Gamma and Weibull distributions may improve their goodness of fit. Compared to the results reported in Table 6, it can be concluded that the simple exponential distribution without a location parameter is indeed a good candidate.

For the region of study, the meteorological seasons are spring from March 1 to May 31, summer from June 1 to August 31 and fall from September 1 to November 30, the rest are considered as winter which is not included in this study. Generally, within the region of study, the winter season is longer than normal. That is why for all the stations, the spring and fall seasons are each one month short except for the Springfield station. The seasonal and monthly relative frequencies of occurrence of rainfall events are obtained for all the stations (Tables 8 and 9). The monthly relative frequency of occurrence is calculated as the number of rainfall events occurred within a specific month divided by the total number of events. These monthly relative frequencies are then averaged over the months comprising each season to obtain the corresponding seasonal relative frequency. This makes the seasonal relative frequencies directly comparable for stations which do not have the same number of months per season.

Table 8 shows that, generally, the frequency of occurrence of rainfall events is higher in the summer followed by spring and fall. Table 9 shows that although there are monthly differences in the frequency of occurrence of rainfall events, these differences are not significant enough (e.g., more than an order of magnitude) as to warrant separate statistical analysis for different months.

2.3.4. Graphical Comparison of Cumulative Distribution Curves

Graphs are widely used in water resources studies as decision tools. For the purpose of testing the trustworthiness of graphical techniques for examining the goodness-of-fit, the observed empirical and the fitted exponential cumulative distribution functions (CDF) were plotted together. In particular, the aim is to see if we can visually reach conclusions similar to those from the GOF tests. We focused on the rainfall event volume only, since its GOF results showed more variations than those of rainfall event duration. Two stations with the most representative results were selected: Springfield, MO and Huron, SD. The corresponding *p-value* of the former station is the lowest when the MOM with $df = k - 2$ was used as the estimator and improved the most when the MCS estimator was used. The *p-value* of the latter station is close to the significance level when MOM with $df = k - 2$ was used, but not enough for the exponentiality hypothesis to be accepted. When the MCS was used for the latter station (i.e., Huron, SD), its *p-value* passed the significance level although it did not increase that much as compared to the other stations (Table 4).

Both the MOM and MCS parameter estimates found in the preceding section were used; the theoretical CDFs resulting from different parameter estimates are graphed separately for clarity. For a better visual inspection, the number of bins was increased from the numbers used for calculating the chi-square test statistics by narrowing the bin width. The bin width in the graphs was initially selected to be 2 mm, it was then gradually increased when the number of data points per bin is less than five. The empirical distribution function values were calculated using the Weibull plotting position formula.

For the two selected stations, the slight misfit that appears in Figs. 4 through 7 for volumes less than or equal to 5 mm is partly due to the elimination of small events. The smaller the value of the volume threshold, the better the fit between theoretical and empirical CDFs for small rainfall event volumes (i.e., volumes less or equal to v_t). However, it is not a matter of concern since these events have minimal hydrologic effects. For the Springfield, MO station, the use of MOM (Fig. 1 a) shows that the exponential distribution does not fit very well the observed data, especially for the volume range of 30 to 75 mm. However when the MCS method is used (Fig. 1 b), the goodness-of-fit improved and visually, without any doubt, we can accept the exponentiality hypothesis. These conclusions are very similar to those resulting from the statistical GOF tests as shown in Table 4. The plot of empirical and theoretical CDFs of Huron, SD, using MOM (Fig. 1 c) illustrates that there is only a slight disagreement between the two curves, which occurs within the volume interval of 10

to 20 mm. In this case it is not very clear whether to accept or reject the exponentiality hypothesis; it depends on the investigator and the purpose of study, in particular the volume's range of interest. Fig. 1 d, where MCS was used as the estimator for the distribution parameter, shows a better agreement between the empirical and theoretical curves, which can be considered as a good fit. The results of chi-square GOF tests (Table 4) for the same station gave similar conclusions as those drawn from the graphs. If the significance level is fixed at 0.05, formal chi-square GOF test would clearly reject the exponentiality hypothesis for rainfall event volume when MOM is used, although the *p-value* (0.038) is close to the fixed significance level. However, the significance level was selected more or less arbitrarily, therefore there are some uncertainties associated with both formal GOF tests and graphical comparisons especially when the test statistic falls in the critical zone between rejection and acceptance. Common sense with related scientific information may ultimately be used to deal with ambiguity.

2.4. Summary and Conclusions

Rainfall event characteristics of all possible rainfall events are often assumed to be exponentially distributed. Although there are several viable exponentiality goodness-of-fit tests, they remain under-applied for rainfall event characteristics mainly due to the almost guaranteed rejection associated with large sample sizes when all possible rainfall events are considered. Visual inspection of distribution curves is

widely used to evaluate the exponentiality of rainfall event characteristics. No clear and reliable guidelines are available about the application of rigorous statistical tests to assist in fitting large rainfall data samples to exponential distributions.

Using rainfall data from seven stations of the North-Central region of the USA, we conducted rigorous statistical tests for the exponentiality of rainfall event volume, duration, and inter-event time. The Poisson test was used for the exponentiality of inter-event times. It can also be used to evaluate the statistical independence of consecutive rainfall events. Selection of different IETDs affects the statistical independence of consecutive rainfall events, the Poisson test can therefore assist in the selection of appropriate IETDs. The chi-square goodness-of-fit (GOF) test was used for testing the exponentiality of rainfall event volume and duration. We paid more attention to the degrees of freedom and the parameter estimators. Commonly, the exponential distribution parameter is estimated using MOM or MLE and it is considered to cause a loss of one degree of freedom (thus $df = k - 2$). However, this assertion is true only when the parameter estimate from MOM coincides with that from the MCS estimator (Chernoff and Lehmann 1954).

The rainfall event volume (v) distribution was found to be the most difficult one to fit exponentially. We had to omit rainfall events with small volumes (i.e., volume threshold $v_t \leq 5$ mm) to improve the goodness-of-fit of v, which also generally improved the fitness for t and b. With the proper selection of IETD and v_t, the v, t and b distributions for the seven stations in the North-Central region of the USA are

accepted to be exponential. We found that the suitable ranges of IETDs and v_t's are 10~12 h and 2~5 mm, respectively. Thus for another station in the region, to save time, we may start with 10~12 h as its suitable IETD. Based on the Poisson tests, the statistically or practically acceptable range of IETD may be longer than our initially specified maximum (i.e., 12 h). Longer than 12-hour IETDs may be appropriate for large watersheds but not suitable for small urban watersheds. When the exponentiality of rainfall event volume and duration is also of interest, an IETD value closer to the optimum based on Poisson tests may not be the suitable one. For example, for the Des Moines data, an IETD of 12 h resulted in better Poisson distributed annual numbers of events than an IETD of 10 h but the 10-hour IETD was found to give better exponentiality of rainfall event volume and duration. The increase of the IETD and v_t tends to raise, generally, the acceptability of the exponentiality hypotheses; nevertheless, only practically reasonable ranges of IETD and v_t should be considered. Moreover, if v_t and IETD are too large, v and b can no longer be approximated as going from 0 to ∞ since the rainfall event volume sample actually has $v \geq v_t$ and the inter-event time sample actually has $b \geq$ IETD. In addition, large values of IETD and v_t may result in combining together or neglecting too many statistically independent and hydrologically significant events.

As for parameter estimation, our results show that the values of the exponential parameter estimated by the MOM are slightly smaller than those estimated by the MCS method. The exponentiality hypothesis is generally rejected when MOM is used

with a df of $(k-2)$. However it was accepted for the majority of the stations of this study when the MCS estimator was employed. When the MOM is applied taking the df as $(k-1)$, the chi-square GOF test *p-values* are generally somewhere between those using the other two methods. These results suggest that the MOM is not an efficient estimator to be used with chi-square GOF tests and confirm that the df lies between $k-2$ and $k-1$, depending on the closeness between the parameter values estimated by the MOM and MCS methods.

As pointed out by D'Agostino and Stephens (1986), there is no unique and well-defined best procedure to follow in order to select and properly apply GOF tests; recommendations are often based on the individual's understanding and view of the specific problem. In order to fit exponential distributions to rainfall event data, based on the results presented in this paper and our previous experience related to other GOF tests, we recommend the use of (i) Poisson tests for ensuring the independence of events and the exponentiality of inter-event times, and (ii) chi-square GOF tests for the exponentiality of event volume and duration. When the chi-square GOF test is used, Eq. (9) may be used for the determination of the appropriate number of equal-width bins that the sample data should be grouped into. We also recommend that the MCS estimator be used to determine the distribution parameter values for rainfall event volume and duration. If the MOM is applied, there should not be a loss of one degree of freedom, especially for rainfall event volume.

Surprisingly, graphical comparisons resulted in similar conclusions as compared

to the chi-square GOF tests; although the numbers of bins used in both methods were different by more than 100%. The chi-square GOF tests would definitely reject the exponentiality hypotheses if the same number of bins as used in graphical comparisons were used. These findings imply that, firstly, the formula used to estimate the number of bins for the chi-square GOF test works well. Secondly, if the chi-square GOF results are positive so will be the graphical comparison results. Thirdly, if visual inspection results are positive, the chi-square GOF results will also be positive or close to acceptance. Finally, visual inspections of CDF curves are very helpful and can be used to detect ranges of variables with bad fittings.

Although quantitative GOF tests enable us to learn more from the data and draw conclusions on a more quantitative basis, the overemphasized statistical significance may over-shadow the practical acceptability of the tested distribution models, especially for cases with large samples (Bentler and Bonett 1980). When a hypothesis is rejected but the *p-value* is not too far from the selected significance level, it is highly recommended that other supplementary information be used to decide on the adoption of models. Such supplementary information may be acquired by investigating the subsequent effects resulting from accepting or rejecting a fitted model. The examination of the effect of a small change of the distribution parameter values on the associated hydrologic results may enrich our common sense about acceptable fittings. Rigorous GOF test techniques and common sense are complementary to each other and none of them is supposed to substitute the other, just

as Cormack (1971, p. 164) noted "The techniques are the horses: common sense holds the reins". More in-depth validation studies are also required before using the fitted models for any specific purposes.

The single parameter exponential distribution is the simplest theoretical distribution that can be used to represent rainfall event characteristics. Exponentiality is desirable not only for its simplicity but more importantly for its application in the analytical probabilistic approach for urban stormwater management. However, for a specific location, exponentiality should not be the only aim of rainfall event frequency analysis. Other distributions such as gamma, Weibull and lognormal may fit rainfall event volume or duration better than exponential distributions. In this initial study, we focused only on exponential distributions; other alternative distributions may be looked at more closely in future studies.

Acknowledgements: Prof. N. Balakrishnan of the Department of Mathematics and Statistics, McMaster University is greatly acknowledged for his useful suggestions and fruitful guidance. The authors thank the anonymous reviewers for their comments and suggestions. This work was supported by the Natural Sciences and Engineering Research Council of Canada.

Notation

The following symbols are used in this paper:

B = continuous random variable of rainfall inter-event time

b = rainfall inter-event time;

$\bar{b}$ = average rainfall inter-event time;

d = Fischer dispersion test statistic;

df = degrees of freedom;

$E[.]$ = mean acronym;

E_i = expected frequencies of bin i;

$F(.)$ = theoretical cumulative distribution function (CDF);

$f(.)$ = probability density function;

IETD = inter-event time definition;

k = number of bins;

M = number of years of record;

m = number of events per year;

n = sample size, i.e., total number of events;

O_i = observed frequencies of bin i;

R = variance-mean ratio;

s = number of parameters estimated from the data;

T = continuous random variable of rainfall event duration;

t = rainfall event duration;

$\bar{t}$ = average rainfall event duration;

V = continuous random variable of rainfall event volume;

v = rainfall event volume;

$\bar{v}$ = average rainfall event volume;

$Var[.]$ = variance acronym;

v_t = volume threshold;

x_1 = lower limit of bin i;

x_2 = upper limit of bin i;

ζ = exponential distribution parameter for rainfall event volume;

λ = exponential distribution parameter for rainfall event duration;

ψ = exponential distribution parameter for rainfall inter-event time;

References

Adams, B. J., Fraser, H. G., Howard, C. D. D., and Hanafy, M. S. (1986). "Meteorologic data analysis for drainage system design." *Journal of Environmental Engineering*, ASCE, 112(5), 827–848.

Adams, B. J. and Papa, F. (2000). *Urban Stormwater Management Planning with Analytical Probabilistic Models*, John Wiley & Sons, Inc., New York, NY, USA.

Ashkar, F., and Rousselle, J. (1987). "Partial duration series modeling under the assumption of a Poissonian flood count." *Journal of Hydrology*, 90(1), 135-144.

Bacchi, B., Balistrocchi, M., and Grossi, G. (2008). "Proposal of a semi-probabilistic approach for storage facility design." *Urban Water Journal*, 5(3), 195-208.

Balistrocchi, M., Grossi, G., and Bacchi, B. (2009). "An analytical probabilistic model of the quality efficiency of a sewer tank." *Water Resources Research*, 45(12), W12420.

Bentler, P. M., and Bonett, D. G. (1980). "Significance tests and goodness of fit in the analysis of covariance structures." *Psychological Bulletin*, 88(3), 588-606.

Berkson, J. (1980). "Minimum chi-square, not maximum likelihood!." *The Annals of Statistics*, 8(3), 457-487.

Beguería, S. (2005). "Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value." *Journal of Hydrology*, 303(1), 215-230.

Bonta, J. V., and Rao, A. R. (1988). "Factors affecting the identification of independent storm events." *Journal of Hydrology,* 98(3), 275-293.

Cameron, A. C., and Trivedi, P. (2013). *Regression analysis of count data (Vol. 53)*, Cambridge University Press, New York.

Chernoff, H., and Lehmann, E. L. (1954). "The use of maximum likelihood estimates in χ^2 tests for goodness of fit." *The Annals of Mathematical Statistics,* 25(3), 579-586.

Clarke, G. M., and Cooke, D. 1. (1978). *A basic course in statistics*, E. Arnold, London.

Cochran, W. G. (1952). "The χ^2 test of goodness of fit." *The Annals of Mathematical Statistics*, 315-345.

Cormack, R. M. (1971). *The statistical argument*, Oliver and Boyd, Edinburgh.

Cruise, J. F., and Arora, K. (1990). "A hydroclimatic application strategy for the Poisson partial duration model." *Water Resources Bulletin,* 26(3), 431-442.

Cunnane, C. (1979). "A note on the Poisson assumption in partial duration series models." *Water Resources Research*, 15(2), 489–494.

D'Agostino, R. B., and Stephens, M. A. (1986). *Goodness-of-fit techniques*, M. Dekker, New York.

Driscoll, E. D., Palhegui, G. E., Strecker, E. W., and Shelley, P. E. (1989). *Analysis of Storm Event Characteristics for Selected Rainfall Gauges throughout the United States*. Report to the U. S. Environmental Protection Agency, Washington, D. C.

Eagleson, P. S. (1972). "Dynamics of flood frequency." *Water Resources Research*, 8(4), 878–898.

Eagleson, P. S. (1978). "Climate, soil, and vegetation, 2, the distribution of annual precipitation derived from observed storm sequences." *Water Resources Research*, 14(5), 713–721.

Guo, Y. (2001). "Hydrologic design of urban flood control detention ponds." *Journal of Hydrologic Engineering*, 6(6), 472–479.

Guo, Y., and Adams, B. J. (1998a). "Hydrologic analysis of urban catchments with event-based probabilistic models: 1. Runoff volume." *Water Resources Research*, 34(12), 3421-3431.

Guo, Y., and Adams, B. J. (1998b). "Hydrologic analysis of urban catchments with

event-based probabilistic models: 1. Peak discharge rate." *Water Resources Research*, 34(12), 3433-3443.

Guo, Y., and Baetz, B. W. (2007). "Sizing of rainwater storage units for green building applications." *Journal of Hydrologic Engineering*, 12(2), 197– 205.

Harris, R. R., and Kanji, G. K. (1983). "On the use of minimum chi-square estimation." *The Statistician*, 32(4), 379-394.

Howard, C. D. D. (1976). "Theory of storage and treatment plant overflows." *Journal of the Environmental Engineering Division,* ASCE, 102(EE4), 709–722.

Keim, B. D., and Cruise, J. F. (1998). "A technique to measure trends in the frequency of discrete random events." *Journal of Climate*, 11(5), 848-855.

Mann, H. B., and Wald, A. (1942). "On the choice of the number of class intervals in the application of the chi-square test." *The Annals of Mathematical Statistics*,13(3), 306-317.

McCuen, R. H. (2003). *Modeling hydrologic change: Statistical methods*, Lewis Publishers, Boca Raton.

National Climatic Data Center (2011). "U. S. Climate Normals (1971-2000)." ⟨http://ggweather.com/normals/index71.htm⟩ (Sept. 2011).

Ott, W. (1995). *Environmental statistics and data analysis*, Lewis Publishers, Boca Raton.

Reese, A. J. (2009). "Volume-based hydrology."

⟨http://foresternetwork.com/daily/water/stormwater-management/volume-based

hydrology/⟩ (Jan. 2014).

Reese, A. J., Jawdy, C. M., Parker, J. M. (2010). "Green infrastructure and storm depth retention criteria."⟨http://foresternetwork.com/daily/water/stormwater-management/green-infrastructure-and-storm-depth-retention-criteria/⟩ (Jan. 2014).

Restrepo-Posada, P. J., and Eagleson, P. S. (1982). "Identification of independent rainstorms." *Journal of Hydrology*, 55(1), 303-319.

Ross, S. M. (2007). *Introduction to probability models*, Academic Press, New York.

Ruffner, J. A., and Bair, F. E. (1978). *Climates of the States, with current tables of normals 1941-1970 and means and extremes to 1975*, Gale Research Co., Detroit.

Serinaldi, F. (2013). "On the relationship between the index of dispersion and Allan factor and their power for testing the Poisson assumption." *Stochastic Environmental Research and Risk Assessment*, *27*(7), 1773-1782.

Sturges, H. A. (1926). "The choice of a class interval." *Journal of the American Statistical Association*, 21(153), 65-66.

U. S. Environmental Protection Agency (USEPA) (1986). *Methodology for analysis of detention basins for control of urban runoff quality*. Report EPA, 440, 5-87, Washington, D. C.

Wanielista, M. P., and Yousef, Y. A. (1993), *Stormwater Management*, John Wiley & Sons, Inc., New York, NY.

Waymire, E., and Gupta, V. K. (1981a). "The mathematical structure of rainfall representations: 1. A review of the stochastic rainfall models." *Water Resources Research*, 17(5), 1261-1272.

Waymire, E., and Gupta, V. K. (1981b). "The mathematical structure of rainfall representations: 2. A review of the theory of point process." *Water Resources Research*, 17(5), 1273-1285.

Waymire, E., and Gupta, V. K. (1981c). "The mathematical structure of rainfall representations: 3. Some applications of the point process theory to rainfall processes." *Water Resources Research*, 17(5), 1287-1294.

White, L. F., Bonetti, M., and Pagano, M. (2009). "The choice of the number of bins for the M statistic." *Computational statistics & data analysis*, 53(10), 3640-3649.

Williams Jr, C. A. (1950). "The choice of the number and width of classes for the chi-square test of goodness of fit." *Journal of the American Statistical Association*, 45(249), 77-86.

Zhang, S., and Guo, Y. (2012a). "Analytical Probabilistic Model for Evaluating the Hydrologic Performance of Green Roofs." *Journal of Hydrologic Engineering*, 18(1), 19-28.

Zhang, S., and Guo, Y. (2012b). "Explicit Equation for Estimating Storm-Water Capture Efficiency of Rain Gardens." *Journal of Hydrologic Engineering*, 18(12), 1739-1748.

Table1. Geographical Information of Selected Stations

Station ID	Station Name	County	State	Latitude	Longitude	Elevation (m)
MN7294	St. Cloud Municipal Airport	Sherburne	Minnesota	N45:32:33	W094:03:08	307.5
IA2203	Des Moines Airport	Polk	Iowa	N41:32:02	W093:39:11	291.7
NE6065	North Platte Regional Airport	Lincoln	Nebraska	N41:07:17	W100:40:10	846.7
ND2859	Fargo WSO Airport	Cass	North Dakota	N46:55:31	W096:48:40	274.3
SD4127	Huron Airport	Beadle	South Dakota	N44:23:53	W098:13:23	390.1
KS2164	Dodge City Regional Airport	Ford	Kansas	N37:46:07	W099:58:04	787.0
MO7976	Springfield Regional Airport	Greene	Missouri	N37:14:23	W093:23:23	383.7

Table 2. Poisson Test Results and Rainfall Event Statistics for IETD = 6 and 12 h (v_t= 0 mm)

Station	θ	$\bar{v}$(mm)	$\bar{t}$ (h)	$\bar{b}$ (h)	*p-value*
St. Cloud, MN	63.1 (53.8)[a]	8.75 (10.27)	5.69 (8.08)	72.93 (84.41)	0.0000 (0.0668)
Des Moines, IA[b]	65.5 (54.5)	9.64 (11.59)	5.65 (8.45)	71.06 (84.03)	0.0003 (0.1955)
North Platte, NE	54.3 (48.1)	7.66 (8.66)	4.82 (6.52)	85.49 (95.78)	0.0002 (0.070)
Fargo, ND	58.8 (50.5)	7.06 (8.22)	5.08 (7.26)	80.03 (92.06)	0.0836 (0.7596)
Huron, SD	56.5 (49.6)	7.27 (8.28)	5.18 (7.03)	82.48 (93.05)	0.2554 (0.7861)
Doge, KS	49.7 (42.8)	8.53 (9.90)	4.46 (6.48)	94.64 (108.81)	0.1045 (0.4202)
Springfield, MO[c]	78.8 (64.4)	11.2 (13.71)	5.86 (8.98)	76.43 (91.97)	0.0073 (0.3749)

[a] Values inside parentheses correspond to IETD = 12 h

[b] Acceptable result starts at IETD = 10 h with *p-value* = 0.105

[c]Acceptable result starts at IETD = 8 h with *p-value* = 0.120

Table 3. Poisson Test Results and Rainfall Event Statistics for the Selected IETD and v_t Values

Station	IETD (h)	v_t (mm)	R^*	*p-value*	θ	$\bar{v}$ (mm)	$\bar{t}$ (h)	$\bar{b}$ (h)
St. Cloud, MN	12	3.05	1.254	0.103	31.8	16.5	11.5	139.0
Des Moines, IA	10	2.03	1.217	0.135	36.2	16.5	10.0	121.4
North Platte, NE	12	3.05	1.291	0.077	26.4	14.8	9.6	170.7
Fargo, ND	12	4.06	1.013	0.447	22.4	16.8	12.0	199.1
Huron, SD	12	2.03	1.046	0.384	29.1	13.5	10.2	156.4
Doge, KS	12	3.05	0.942	0.592	24.6	16.4	9.1	182.0
Springfield, MO	12	5.08	1.219	0.134	35.5	23.4	13.3	166.8

* Note: R values which are calculated using Eq. (4) are given here to compare with the corresponding *p-values*. The *p-values* are used for decision on hypotheses.

Table 4. Chi-Square GOF Test Results for Rainfall Event Volume and Duration

Station	Rainfall event volume				Rainfall event duration			
	MOM		MCS		MOM		MCS	
	ζ	*p-value**	ζ	*p-value*	λ	*p-value*	λ	*p-value*
St. Cloud, MN	0.061	0.025 (0.043)	0.062	0.039	0.087	0.043 (0.067)	0.091	0.104
Des Moines, IA	0.060	0.077 (0.122)	0.062	0.112	0.100	0.041 (0.067)	0.102	0.052
North Platte, NE	0.068	0.039 (0.065)	0.070	0.076	0.104	0.063 (0.098)	0.104	0.063
Fargo, ND	0.060	0.023 (0.043)	0.063	0.080	0.083	0.290 (0.393)	0.084	0.304
Huron, SD	0.074	0.038 (0.064)	0.077	0.066	0.098	0.545 (0.661)	0.100	0.576
Doge, KS	0.061	0.039 (0.066)	0.064	0.117	0.109	0.044 (0.078)	0.110	0.046
Springfield, MO	0.043	0.002 (0.004)	0.046	0.058	0.075	0.032 (0.053)	0.077	0.053

* These are the *p-values* correspond to MOM with $df = k - 2$; the values inside parentheses are the *p-values* correspond to MOM with $df = k - 1$

Table 5. Split Sample Validation Test *p-values*

Data\Station	MN	IA	NE	ND	SD	KS	MO
Training set	0.2956	0.698	0.1964	0.4178	0.7149	0.1269	0.004
Testing set	0.039	0.2813	0.1861	0.297	0.4044	0.6979	0.3606

Table 6. Chi-square GOF Test *p-values* of Rainfall Event Volume and Duration Fitted to Different Families of Distributions*

	Exponential	Gamma	Lognormal	Weibull
St. Cloud, MN	0.076 (0.018)	0.008 (0.295)	0.186 (0.000)	0.000 (0.000)
Des Moines, IA	0.247 (0.981)	0.445 (0.960)	0.038 (0.000)	0.000 (0.000)
North Platte, NE	0.025(0.010)	0.011 (0.020)	0.028 (0.000)	0.000 (0.000)
Fargo, ND	0.000 (0.483)	0.000 (0.681)	0.006 (0.003)	0.000 (0.000)
Huron, SD	0.027 (0.657)	0.120 (0.672)	0.358 (0.001)	0.000 (0.000)
Doge, KS	0.025 (0.010)	0.011 (0.020)	0.028 (0.000)	0.000 (0.000)
Springfield, MO	0.002 (0.052)	0.000 (0.014)	0.037 (0.000)	0.000 (0.000)

* Note: The *p-values* of GOF tests of rainfall event duration are the ones inside parentheses.

Table 7: Chi-square GOF test *p-values* of rainfall event volume fitted to distributions with location parameters

Station	Location parameter	2-parameter Exponential	3-parameter Gamma	3-parameter Lognormal	3-parameter Weibull (K-S)*
St. Cloud, MN	3	0.000	0.586	0.000	0.000 (0.416)
Des Moines, IA	2	0.000	0.870	0.000	0.000
North Platte, NE	3	0.000	0.410	0.000	0.015 (0.070)
Fargo, ND	4	0.000	0.150	0.000	0.000
Huron, SD	2	0.000	0.800	0.000	0.000
Doge, KS	3	0.000	0.252	0.000	0.000 (0.084)
Springfield, MO	5	0.000	0.237	0.000	0.000 (0.062)

*The numbers in parentheses are K-S GOF test *p-values*; only non-zero *p-values* are reported.

Table 8: Seasonal relative frequency of occurrence of rainfall events*

	MN	IA	NE	ND	SD	KS	MO
Spring	0.13	0.15	0.16	0.13	0.15	0.15	0.12
Summer	0.17	0.16	0.17	0.17	0.17	0.17	0.11
Fall	0.11	0.12	0.09	0.11	0.10	0.09	0.10

*Note: The seasonal relative frequency of occurrence is expressed as the average of the monthly relative frequency of occurrence over the months comprising each individual season.

Table 9: Monthly relative frequency of occurrence of rainfall events

	MN	IA	NE	ND	SD	KS	MO
MAR							0.11
APR	0.11	0.11	0.12	0.11	0.14	0.12	0.12
MAY	0.16	0.18	0.19	0.16	0.16	0.19	0.14
JUN	0.19	0.17	0.19	0.21	0.19	0.19	0.14
JUL	0.16	0.16	0.19	0.17	0.17	0.18	0.10
AUG	0.16	0.15	0.13	0.14	0.13	0.15	0.10
SEP	0.13	0.13	0.11	0.12	0.12	0.10	0.11
OCT	0.09	0.11	0.07	0.09	0.09	0.08	0.09
NOV							0.09

Figure Captions

Fig. 1. Comparison of empirical and fitted exponential distributions of rainfall event volume: (a) Springfield, MO (IETD = 12 h, v_t = 5.08 mm) using MOM, (b) Springfield, MO (IETD = 12 h, v_t = 5.08 mm) using MCS, (c) Huron, SD (IETD = 12 h, v_t = 2.03 mm) using MOM and (d) Huron, SD (IETD = 12 h, v_t = 2.03 mm) using MCS

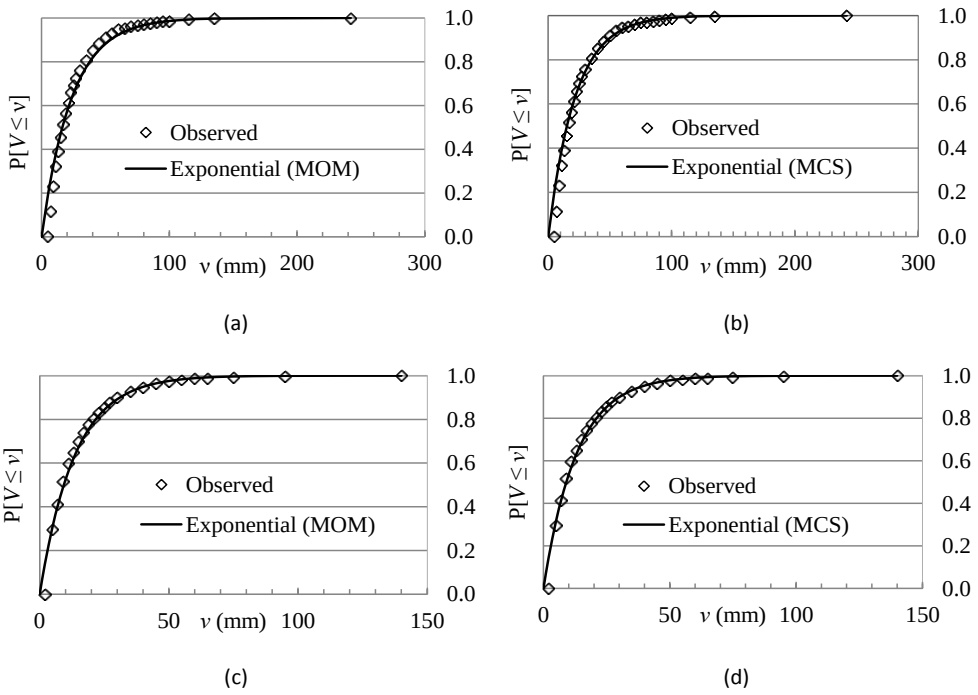

Fig. 1. Comparison of empirical and fitted exponential distributions of rainfall event volume: (a) Springfield, MO (IETD = 12 h, v_t = 5.08 mm) using MOM, (b) Springfield, MO (IETD = 12 h, v_t = 5.08 mm) using MCS, (c) Huron, SD (IETD = 12 h, v_t = 2.03 mm) using MOM and (d) Huron, SD (IETD = 12 h, v_t = 2.03 mm) using MCS

CHAPTER 3

Derived Flood Frequency Distributions Considering Individual Event Hydrograph Shapes

The content of this chapter is the manuscript text published under the following citation:

Hassini, S. and Guo, Y. (2017). "Derived flood frequency distributions considering individual event hydrograph shapes." *Journal of Hydrology*, *547*, 296-308.

https://doi.org/10.1016/j.jhydrol.2017.02.003

Derived Flood Frequency Distributions Considering Individual Event Hydrograph Shapes

Sonia Hassini and Yiping Guo

Abstract: Derived in this paper is the frequency distribution of the peak discharge rate of a random runoff event from a small urban catchment. The derivation follows the derived probability distribution procedure and incorporates a catchment rainfall-runoff model with approximating shapes for individual runoff event hydrographs. In the past, only simple triangular runoff event hydrograph shapes were used, in this study approximating runoff event hydrograph shapes better representing all the possibilities are considered. The resulting closed-form mathematical equations are converted to the commonly required flood frequency distributions for use in urban stormwater management studies. The analytically determined peak discharge rates of different return periods for a wide range of hypothetical catchment conditions were compared to those determined from design storm modeling. The newly derived equations generated results that are closer to those from design storm modeling and provide a better alternative for use in urban stormwater management studies.

Keywords: Rainfall event analysis; Runoff event analysis; Flood frequency distribution; Urban catchments; Urban stormwater management.

3.1. Introduction

Peak-discharge rates of different return periods from urban catchments are frequently required for the planning and design of urban stormwater management facilities. The accurate sizing of these facilities highly depends on the accuracy of these peak-discharge rates. Due to the lack of observed flow data, the required peak-discharge frequency information is often obtained from relevant rainfall data together with rainfall-runoff transformation models.

Both the design storm and the continuous simulation approaches may be used to determine the exceedence frequencies of peak discharges. The design storm approach assumes that the selected design storm and its resulting peak runoff rate have the same return period. This assumption makes the design storm approach simple to apply but with a possible sacrifice on accuracy (Adams and Howard, 1986). This is because the return period of the design storm and the resulting peak runoff rate may approximately equal to each other and there is no way to accurately calculate the exact return period of the peak runoff rate resulting from an input design storm. The continuous simulation approach uses a long and continuous rainfall record as input and transforms it into a flow rate series, a frequency analysis is then conducted to obtain the flow rates of desired exceedence frequencies. For locations where long-term rainfall records are not available, they may be synthetically generated based on limited observed rainfall data (Grimaldi et al., 2012a, b; 2013). The use of long series of actual rainfall data allows the model to include more complete meteorological

conditions, properly model the soil conditions during inter-event dry periods, and therefore better establish the actual antecedent moisture conditions for individual rainfall events (Nnadi et al., 1999). The application of the continuous simulation approach is limited due to time constraints and lack of rainfall or runoff data (Quader and Guo, 2006). Instead, the design storm approach is widely applied in practice due to its simplicity. The unavailability of other simple yet accurate alternative approaches is the other reason that the design storm approach is almost universally applied in practice (Levy and McCuen, 1999; Guo and Zhuge, 2008).

In the recent decades there appeared a new approach referred to as the analytical probabilistic approach (APA). This new approach was developed to overcome the shortcomings of the design storm and continuous simulation approaches. The APA includes some of the features of the design storm and continuous simulation approaches. The APA is event-based in terms of rainfall-runoff transformation which is similar to the design storm approach; however the APA also uses long rainfall data similar to the continuous simulation approach. The APA starts by identifying and analyzing actual individual rainfall events from a long-term continuous rainfall record. The APA, like the other two approaches, requires some kind of frequency analysis but with different procedures. First the long continuous recorded rainfall data is segregated into individual rainfall events where each event is characterized by its volume (v), duration (t), and the dry time (referred to as the interevent time, b) that separates it from the preceding rainfall event. Then similar to

the design storm approach, the frequency analysis employed by the APA is performed with the input rainfall data; however, the actual individual rainfall event's characteristics such as a rainfall event's total volume v (mm), total duration t (h) and the interevent time b (h) are subjects for frequency analysis in the APA. For the development of design storms, however, the subjects of frequency analysis are rainfall amounts fallen within pre-selected durations. For the continuous simulation approach, the frequency analysis is performed on the modelled runoff series in order to obtain the peak-discharge frequency distributions. For the APA, the probability distributions of runoff characteristics (peak flow and runoff volume) are derived directly from the probability distributions of the input rainfall event characteristics (Eagleson 1972; Howard 1976; Adams et al., 1986; Guo and Adams, 1998a, b and 1999a, b; Guo, 2001; Bacchi et al., 2008; Balistrocchi et al., 2009). The APA is able to generate, in a more straightforward manner, information about prediction uncertainty, either because of parameter uncertainty, forcing uncertainty, or both. It is worth mentioning that there is another alternative approach to design storm and continuous simulation modeling which is recently developed by Mejía et al. (2014). This approach uses stochastic models and has similarities to the APA approach. One similarity is that they both use probability density functions to characterize rainfall depths. Mejía et al. (2014) used stochastic models to generate flow duration curves for urbanized watersheds, but similar stochastic approaches may be developed for peak flow frequencies. Up till now, the only approach used in actual engineering practice for the

determination of the frequency distribution of peak discharge rates for small urban catchments is almost always the design storm approach. Ever since the beginning of the application of the analytical approach, it has been improving little by little due to the challenging mathematical derivation that is involved in its development.

For the derivation of the mathematical equations comprising the APA, the rainfall event characteristics (i.e. rainfall event volume, duration and interevent time) are generally assumed to follow exponential distributions (Eagleson 1972, 1978; Howard 1976; Adams et al., 1986; Guo and Adams, 1998a, b and 1999a, b; Adams and Papa, 2000; Guo 2001; Guo and Baetz, 2007; Zhang and Guo, 2013a, b). For many North American locations, exponential distributions were found to be appropriate for describing their rainfall characteristics (Adams and Papa, 2000; Wanielista and Yousef, 1993). In addition, the rainfall event volume and duration are assumed to be statistically independent. The APA is only applicable where the aforementioned assumptions are valid. Hassini and Guo (2016) provided a detailed procedure for testing the exponentiality of a location's rainfall event characteristics. The input rainfall statistics and catchment conditions together with the resulting closed-form mathematical equations describing the peak discharge and runoff volume frequency distributions are referred to as the analytical probabilistic models. These models are developed for stormwater management planning and design purposes (Adams and Papa, 2000); similar to continuous simulation and design storm models.

The APA was first used by Eagleson (1972) to estimate the frequency of peak

stream flows using the probability density functions (pdfs) of rainfall intensity and duration, and the kinematic wave formula, which represents a functional relationship between peak stream flows and rainfall characteristics. Then it was employed for stormwater management purposes by many other researchers (e.g., Adams et al., 1986; Guo and Adams, 1998a, b and 1999a, b; Adams and Papa, 2000; Guo, 2001; Quader and Guo, 2006; Guo and Baetz, 2007; Chen and Adams, 2005 and 2007; Bacchi et al., 2008; Balistrocchi et al., 2009; Zhang and Guo, 2013a, b). To estimate peak-discharge frequency distributions, Guo and Adams (1998b) developed an analytical probabilistic stormwater management model assuming that each individual runoff event hydrograph can be approximated as a triangle, the resulting model is referred to as the APSWM(Tri) model in this paper. For the development of APSWM(Tri), a hydrograph's total volume and duration is calculated first, peak discharge rate of that hydrograph is then calculated based on the assumed approximating shape of the hydrograph. Since most hydrographs have a broad peak area than a triangular hydrograph, the assumption of triangular hydrographs may result in an overestimation of peak discharge rates, especially for larger catchments and/or longer rainfall events. Although the capability of APSWM(Tri) was later expanded to include explicit channel flow routing (Guo et al., 2009) and different methods for rainfall loss calculations (Guo and Markus, 2011), the simple assumption of triangular hydrograph is always used. In order to further develop the analytical probabilistic models and increase their accuracy, other alternatives of estimating the peak-discharge rate based

on more accurate hydrograph shapes for individual runoff events should be considered.

Ponce (1989) illustrated that event hydrographs may take three possible shapes depending on the catchment's time of concentration and rainfall event duration (details of these shapes will be illustrated in the next section); two of the three possible shapes can be approximated as trapezoidal and the other can be approximated as triangular. Triangular or trapezoidal hydrograph is not used as a shape of a unit hydrograph for numerical hydrologic modeling purposes; instead it is used to approximate individual event hydrographs, this is needed in order to use the derived probability distribution approach for estimating flood frequencies. The objective of this study is to derive and verify new closed-form mathematical equations describing the peak-discharge rate frequency distributions assuming that the runoff event hydrographs take the three possible shapes as suggested by Ponce (1989). The resulting analytical probabilistic model is referred to as the APSWM(Tra). The APSWM(Tri) model that was developed earlier is simplified in the approximation of individual event hydrograph shapes, the APSWM(Tra) model which is developed in this paper improves greatly on that. To verify the performance of APSWM(Tra), it is vital to investigate the estimates of peak-discharge rate frequency distributions for as many as possible different catchment conditions. Third-party reliable estimates of peak-discharge rate frequency distributions are also required for this purpose. Despite its drawbacks, the widely used design storm approach can provide results with

acceptable levels of accuracy if the design storm characteristics and antecedent catchment conditions are selected properly (Guo, 2001; Guo and Zhuge, 2008). Thus the design storm results will be used in this paper as a reference to compare the two analytical models APSWM(Tra) and APSWM(Tri).

3.2. Peak Discharge Rate of a Runoff Event Based on Approximating Hydrograph Shapes

3.2.1. Runoff Event Volume

In the development of APSWM(Tri), the rainfall-runoff transformation is completed on an event-by-event basis. The total volume v_r (mm) of a runoff event (referred to as runoff event volume) is equal to the total rainfall volume of the input rainfall event (v) minus all the hydrologic losses that occur during the event. The hydrologic losses considered for the development of the APSWM(Tri) are interception, depression storage, and infiltration losses. Interception and depression storage losses are combined together and referred to as the depression storage losses which are assumed to be fully available and would be filled before the occurrence of any runoff for each rainfall event (Guo and Adams, 1998a).

The infiltration losses are the amount of rainfall seeped into the soils which may later become interflow into streams and lakes or percolate into deeper aquifers (Ponce, 1989). For a fair comparison later between APSWM(Tra) and APSWM(Tri), here the

infiltration losses are also estimated by the model proposed by Guo and Adams (1998a). For the development of APSWM(Tri), Guo and Adams (1998a) adopted the Horton equation and simplified it to obtain, for a given rainfall event with duration t, the maximum possible infiltration loss (S_{mf}) as:

$$S_{mf} = S_{iw} + f_c t \quad (1)$$

In Eq. (1), S_{iw} is the initial soil wetting infiltration volume (mm) and f_c is the ultimate infiltration capacity of the soil (mm/h). Guo and Adams (1998a) provided more details on how Eq. (1) is obtained and how S_{iw} may be estimated.

To estimate the runoff event volume, an urban catchment is first divided into pervious and impervious areas. The catchment runoff volume is then the sum of the runoff volumes from the two areas. Each area is assumed to contribute to the total runoff only after its rainfall losses are satisfied. For pervious areas, rainfall losses include infiltration and depression (including interception) amounts; for impervious areas only depression storage losses (including interception losses) are considered. The total runoff volume expressed in the form of depth of water over the catchment area is the sum of the area-weighted runoff volumes from the pervious and impervious portions of the catchment.

The impervious portion's rainfall losses are usually less than the pervious portion's losses. Consequently, for an urban catchment, there will be no runoff until the impervious areas' depression storage is satisfied. Then, only impervious areas will generate runoff when the rainfall volume is less than the sum of the pervious area's

initial and infiltration losses. The pervious area's initial losses (denoted as S_{il}) is the sum of the area's depression storage S_{dp} and initial soil wetting infiltration volume S_{iw}. When the pervious area's initial losses and infiltration losses are both satisfied, both impervious and pervious areas will contribute runoff (Guo and Adams, 1998a). The total volume of a runoff event resulting from a rainfall event with volume v and duration t is

$$v_r = \begin{cases} 0, & v \le S_{di} \\ h(v - S_{di}), & S_{di} < v \le S_{il} + f_c t \\ v - S_d - f_c(1-h)t, & v > S_{il} + f_c t \end{cases} \quad (2)$$

where v_r is referred to as the runoff event volume (mm); $S_{il} = S_{dp} + S_{iw}$ and $S_d = hS_{di} + (1-h)S_{il}$; h is the hardened (impervious) surface fraction of the catchment; S_{il} is the pervious area initial losses (mm); S_{di} is the impervious area depression storage (mm); S_{dp} is the pervious area depression storage (mm); and S_d is referred to as the area-weighted depression storage of the impervious areas and the initial losses of the pervious areas (mm). Detailed derivations of Eq. (2) can be found in Guo and Adams (1998a).

3.2.2. Peak Discharge Rate of a Runoff Event

For the determination of the peak discharge rate of a runoff event, Guo and Adams (1998b) assumed that the runoff event hydrograph has a triangular shape with a base (i.e., runoff event duration) equaling the sum of the input rainfall event duration (t) and the catchment time of concentration (t_c). Where the catchment time of concentration is defined as the time required for runoff to travel from the most distant

upstream point (in terms of travel time) of the catchment to the catchment's outlet. A catchment's time of concentration is assumed to be constant and independent of rainfall event characteristics.

According to Ponce (1989), the shape of a runoff event hydrograph is controlled mainly by the relative magnitude of the catchment's time of concentration and the input storm duration; the following three cases are possible. First, if the storm duration is less than the time of concentration, the hydrograph can be assumed to be trapezoidal with the lower and upper bases equaling $(t + t_c)$ and $(t_c - t)$, respectively; this type of catchment response is called sub-concentrated (Fig. 1, a). Second, if $t = t_c$, the catchment response is called concentrated and the hydrograph can be assumed to be triangular (Fig. 1, b). Third, if $t > t_c$, the catchment response is called super-concentrated and the hydrograph shape can be considered trapezoidal with the lower and upper bases equaling $(t + t_c)$ and $(t - t_c)$, respectively (Fig. 1, c). Based on the above-described simplifications, the peak discharge rate (Q_p) of a runoff event from an urban catchment can be geometrically determined as

$$Q_p = \begin{cases} \frac{v_r}{t_c}, & t \le t_c \\ \frac{v_r}{t}, & t > t_c \end{cases} \tag{3}$$

Substituting the runoff event volume (v_r) as expressed in Eq. (2) into Eq. (3), the analytical expression of Q_p as a function of the rainfall event volume v and duration t is

$$Q_p = \begin{cases} 0, & v \le S_{di} \\ \dfrac{h(v - S_{di})}{t_c}, & S_{di} < v \le S_{il} + f_c t \text{ and } t \le t_c \\ \dfrac{v - S_d - f_c(1-h)t}{t_c}, & v > S_{il} + f_c t \text{ and } t \le t_c \\ \dfrac{h(v - S_{di})}{t}, & S_{di} < v \le S_{il} + f_c t \text{ and } t > t_c \\ \dfrac{v - S_d - f_c(1-h)t}{t}, & v > S_{il} + f_c t \text{ and } t > t_c \end{cases} \quad (4)$$

3.3. Derivation of the Peak Discharge Exceedence Probabilities

The derived probability distribution theory (Adams and Papa, 2000) is used to obtain the probability distribution function of a random variable that is dependent on other random variables with known probability distributions. The peak discharge rate Q_p as expressed in Eq. (4) is a random variable that is dependent on two other random variables, i.e., the rainfall event volume (v) and duration (t). The random variables v and t are assumed to be statistically independent and have the following marginal probability density functions (Eagleson, 1972; Howard, 1976):

$$f(v) = \zeta e^{-\zeta v}, \qquad v \ge 0 \quad (5)$$

$$f(t) = \lambda e^{-\lambda t}, \qquad t \ge 0 \quad (6)$$

In Eqs. (5) and (6), ζ and λ are the distribution parameters for rainfall event volume and duration, respectively.

Since v and t are assumed to be independent, their joint probability density function is simply the product of their marginal probability density functions.

According to Eq. (4) and making use of the joint probability density function of v and t, the probability of having $Q_p = 0$ per rainfall event can be found as follows:

$$P[Q_p = 0] = \int_0^{S_{di}} \int_0^{\infty} \zeta e^{-\zeta v} \lambda e^{-\lambda t} dt\, dv = 1 - e^{-\zeta S_{di}} \tag{7}$$

In the above integration, t goes from 0 to ∞ and v goes from 0 to S_{di}; the region on the v-t plane covered by $t > 0$ and $0 < v \leq S_{di}$ is the so-called region of integration in the derived probability distribution theory. In the Appendix, similar sub-regions of integration are delineated for non-zero Q_p values and the exceedence probabilities for non-zero Q_p values are derived.

For summarizing the above and the additional derivation results presented in the Appendix, the following short-hand notation is introduced to simplify expressions:

$$g_1(q_p) = \exp\left(-\frac{\zeta q_p t_c}{h} - \zeta S_{di}\right)$$

$$g_2(q_p) = \frac{\zeta q_p}{\lambda h + \zeta q_p} \exp\left(-\lambda t_c - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right)$$

$$g_3(q_p) = \frac{(1-h)\lambda\zeta(q_p - h f_c)}{(\lambda h + \zeta q_p)(\lambda + \zeta q_p + (1-h)\zeta f_c)} \exp\left(\frac{-\zeta S_{il} q_p + h\zeta f_c S_{di} - h\lambda S_{dd}}{q_p - h f_c}\right)$$

$$g_4(q_p) = \frac{\lambda}{\lambda + (1-h)\zeta f_c} \exp(-\zeta q_p t_c - \zeta S_d)$$

$$g_5(q_p) = \frac{(1-h)\zeta f_c}{\lambda + (1-h)\zeta f_c} \exp\left(\frac{-f_c \zeta q_p t_c - h\zeta f_c S_{di} - \lambda q_p t_c + h\lambda S_{dd}}{h f_c}\right)$$

$$g_6(q_p) = \frac{\lambda\zeta q_p}{(\lambda + \zeta q_p + (1-h)\zeta f_c)(\lambda + (1-h)\zeta f_c)} \exp(-\lambda t_c - \zeta q_p t_c - (1-h)\zeta f_c t_c - \zeta S_d)$$

Using the above notation and summarizing all the derivation results, the peak

discharge rate exceedence probability per rainfall event can be expressed as follows:

For $f_c \leq \frac{S_{dd}}{t_c}$ or Type I catchments,

$$P[Q_p > q_p] = \begin{cases} g_1(q_p) - g_2(q_p), & q_p \leq hf_c \\ g_1(q_p) - g_2(q_p) + g_3(q_p), & hf_c < q_p \leq \frac{hS_{dd}}{t_c} \\ g_4(q_p) - g_2(q_p) + g_3(q_p) + g_5(q_p), & \frac{hS_{dd}}{t_c} < q_p \leq hf_c + \frac{hS_{dd}}{t_c} \\ g_4(q_p) - g_6(q_p), & q_p > hf_c + \frac{hS_{dd}}{t_c} \end{cases} \quad (8)$$

For $f_c > \frac{S_{dd}}{t_c}$ or Type II catchments,

$$P[Q_p > q_p] = \begin{cases} g_1(q_p) - g_2(q_p), & q_p \leq \frac{hS_{dd}}{t_c} \\ g_4(q_p) - g_2(q_p) + g_5(q_p), & \frac{hS_{dd}}{t_c} < q_p \leq hf_c \\ g_4(q_p) - g_2(q_p) + g_3(q_p) + g_5(q_p), & hf_c < q_p \leq hf_c + \frac{hS_{dd}}{t_c} \\ g_4(q_p) - g_6(q_p), & q_p > hf_c + \frac{hS_{dd}}{t_c} \end{cases} \quad (9)$$

3.4. Annual Peak-Discharge Exceedence Probability and Flood Frequency Distribution

In flood control design, annual exceedence probability or return period is often used instead of exceedence probability per rainfall event. By definition, return period (T_R, in years) is the inverse of the exceedence probability per year. The exceedence probabilities determined using Eqs. (8) and (9) [along with the expressions $g_1(q_p)$

through $g_6(q_p)$] are exceedence probabilities per rainfall event, which must be multiplied by the average number of events per year (θ) in order to obtain exceedence probabilities per year. The return period of a given peak discharge rate q_p can be calculated as

$$T_R = \frac{1}{\theta P[Q_p > q_p]} \tag{10}$$

Thus, Eqs. (8), (9), and (10) constitute a derived flood frequency distribution model, which is the basis of the aforementioned analytical probabilistic stormwater model APSWM(Tra).

3.5. Comparison of Analytical Probabilistic and Design Storm Modeling Results

3.5.1. Input Rainfall Data and hypothetical Catchments

In order to verify that APSWM(Tra) produces more accurate results than the original analytical probabilistic stormwater model APSWM(Tri), the two analytical probabilistic models' results were compared against each other, and results from the design storm approach were used as a reference. The design storm approach is widely used in engineering practice and its results are accepted as accurate enough and used for the design of stormwater control measures. For a given catchment, the two analytical probabilistic models require exactly the same inputs. However, some of the

design storm approach's inputs are different from those of the analytical probabilistic models, thus it is necessary to extract equivalent inputs for the design storm approach from those of the two analytical probabilistic models. One rainfall station was selected for rainfall input for both approaches and the same or equivalently the same catchment parameters were used for both approaches in modeling one specific catchment.

For this study, the rainfall station of St. Cloud Municipal Airport in Sherburne, Minnesota was selected. It is located in the north central region of the United States with a latitude, longitude, and elevation of N45:32:33, W094:03:08, and 307.5m, respectively. The selected years of record are from 1949 to 2001 (53 years) with no missing data. Although rainfall data are available beyond 2001, only the years of record from 1949 to 2001 were selected to match with the rainfall record used in the development of the IDF tables available on the National Oceanic and Atmospheric Administration (NOAA)'s website (National Weather Service, 2015), which are used in this study together with the design storm approach. Snow months are not included in this study; according to Ruffner and Bair (1978) and the National Climatic Data Center's website (National Climatic Data Center, 2011), St. Cloud's rainfall months are from April through October.

For the selected station, Hassini and Guo (2016) performed a statistical analysis on the hourly rainfall data and found that the rainfall event characteristics follow closely exponential distributions described by Eqs. (5) and (6). The required

rainfall inputs for the analytical probabilistic models were found to be $\lambda = 0.0868\ \mathrm{h}^{-1}$, $\zeta = 0.0605\ \mathrm{mm}^{-1}$, and $\theta = 31.8\ \mathrm{events/year}$ (Hassini and Guo 2016). The scatter-plot of rainfall event volume versus rainfall event duration is shown in Fig. 2.

The Kendall rank correlation coefficient between rainfall event volume and duration was found to be 0.31. Both the scatter-plot and the Kendall coefficient suggest that only weak correlations exist and it might be acceptable to assume that *v* and *t* are statistically independent. Rivera et al. (2005) compared different locations and reported that the degree of dependence between rainfall event volume and duration is location specific. They found that APSWM(Tri) would produce reasonably accurate results for locations where rainfall event volume and duration are statistically independent but not so accurate results where rainfall event volume and duration are highly dependent. Thus if the independence assumption is false, the final analytical probabilistic results will be inaccurate as compared to those from the design storm models which does not require similar assumptions.

A design storm is characterized by rainfall depth, duration, return period (or annual exceedence probability), temporal distribution and computational time steps (Wurbs and James, 2002). The design storm duration is an artificially selected duration that does not necessarily coincide with the duration of a real rainfall event (Cheng et al., 2003; Wurbs and James, 2002). Levy and McCuen (1999) suggested that for most of the hydrologic designs, the selected design storm duration should be

either 24 hours or the catchment time of concentration. Since catchments with different times of concentration will be tested here, 24-hour duration design storms will be used in this study.

Once the duration of the design storm is selected, the rainfall data required for the design storm approach are the rainfall depths of different return periods. These data can be obtained by statistically analyzing either the partial-duration series (PDS) or annual maximum series (AMS) obtained from the original rainfall data series. Both PDS and AMS analysis results are available on the website of the NOAA's National Weather Service (National Weather Service 2015). Since PDS is more appropriate and reliable, the PDS results are used in this study. The commonly used return periods for design purposes (i.e. 2, 5, 10, 25, 50, and 100 years) are included in this study. The 24-h partial-duration precipitation depths of the St. Cloud station based on rainfall data covering the period of 1949-2001 are shown in Table 1.

The temporal distribution of design storms was assumed to be the SCS Type 2 (SCS stands for the Soil Conservation Service which is now renamed as the Natural Resources Conservation Service, however, the methodologies that the SCS developed are still referred to by the old name) since the rainfall station used here is located where the SCS Type 2 distribution is applicable. SCS design storm hyetographs are specified such that the shorter and more intense rainfalls are nested within the total 24-h durations, and therefore the 24-h SCS hyetographs are appropriate for both small and large watersheds.

Hypothetical catchments are set up with two different types of soils, clay and sand; and different combinations of levels of imperviousness and times of concentration. All catchments have the same area of 0.18 km^2 since peak discharge rates are all expressed in the unit of mm/h and therefore catchment areas have no effect on the value of the peak discharge rates. The constant infiltration rate or the final (equilibrium) infiltration capacity f_c is 0.25 and 25 mm/h for clay and sandy soils, respectively (Viessman and Lewis, 2003). Hicks (1944) illustrated that the depression storage is likely to be 2.5 and 5 mm for clay and sandy soils respectively; and these values are used as the depression storage in this study as well. The initial soil wetting infiltration was found to be approximately 2 and 15 mm for clay and sand, respectively (Akan and Houghtalen, 2003). The impervious area depression storage is assumed to be 0 mm since it is neglected in HEC-HMS models (Scharffenberg and Fleming, 2009) and HEC-HMS was used for design storm modeling. Four different values of the time of concentration t_c (0.5, 1.5, 3 and 6 h) and two levels of imperviousness (100 and 35%) were considered.

The analytical probabilistic models transform input rainfall to output runoff using Eq. (2) on an event-by-event basis, while routing of this runoff through the catchment to determine the peak discharge rate is accomplished using Eq. (4). The catchment time of concentration t_c is the only parameter used to represent the runoff routing effect of a catchment. For numerically modeling the catchment runoff-routing required by the design storm approach, the SCS and Clark's unit hydrograph (UH)

methods were applied using the software HEC-HMS. These two methods are chosen because they are the most widely used methods and require the input of a parameter(s) similar to t_c. In the SCS UH method, the catchment lag time t_{lag} is required and it is commonly known to be about $0.6\ t_c$, where t_c is the catchment time of concentration which is also required as an input to the analytical probabilistic models. Clark's UH method requires a storage coefficient R as well as a time of concentration. The storage coefficient is used to represent the routing effect of a hypothetical linear reservoir located at the downstream end of a catchment. As shown in Guo and Zhuge (2008), adding a linear reservoir with a storage coefficient R to a catchment is equivalent to adding $2R$ to the catchment's time of concentration for runoff routing purposes. In order to minimize the differences between the analytical probabilistic and design storm models in representing the catchment's time of concentration, for each hypothetical catchment, the lowest possible R value accepted by HEC-HMS (0.1 h) was used and a time of concentration (t_c') equaling to the corresponding analytical probabilistic model's time of concentration minus $2R$ (i.e. $t_c' = t_c - 2R = t_c - 0.2$) is used as the time of concentration for the Clark's UH method. This way, for catchment runoff routing calculations, the HEC-HMS design storm and analytical probabilistic models are essentially equivalent for each hypothetical catchment.

Since the analytical probabilistic models are developed for small catchments where baseflow is usually equal to zero, baseflow is therefore not included in the analytical probabilistic models; neither in the design storm models. Canopy storage is

combined with depression storage in all models. In HEC-HMS, the initial loss and constant loss rate method was selected as the catchment hydrologic loss calculation method because it has the same structure as the infiltration loss model [Eq. (1)] used in the analytical probabilistic models. Adding the pervious area depression storage losses (S_{dp}) to the maximum possible infiltration losses [S_{mf} as expressed in Eq. (1)], the total loss is equal to ($S_{il} + f_c t$), where $S_{il} = S_{dp} + S_{iw}$ and stands for the initial losses as defined in the analytical probabilistic models. Therefore, for each catchment, S_{il} and f_c as used in the analytical probabilistic models are equated to the initial loss and the constant loss rate, respectively, as used in the design storm models. This way, for hydrologic loss calculations, the HEC-HMS design storm and analytical probabilistic models are essentially equivalent as well. A calculation time step of 5 minutes was chosen for HEC-HMS.

3.5.2. Results and Discussion

A total of three catchment groups were modeled using HEC-HMS and the analytical probabilistic models: (1) 35% impervious catchments with clay soils, (2) 35% impervious catchments with sandy soils, and (3) 100% impervious catchments. The t_c for each of the above groups of catchments ranges from 0.5 to 6 hours. Typical comparisons of results are shown in Fig. 3 for 100% impervious catchments with four different t_c (0.5, 1.5, 3 and 6 h) values and Fig. 4 for 35% impervious catchments with two different types of soils (sand and clay) and two different values of t_c (1.5 and 3 h) for each type of soils. Catchments with 100% imperviousness are purposely selected

so that rainfall loss calculation results from HEC-HMS and the analytical probabilistic models are the same, the difference in flood frequency results may be attributed entirely to the difference in runoff routing calculations. Catchments with 35% imperviousness and different types of soil are selected to test the competence of the developed models when the calculation of infiltration amounts is included.

Fig. 3 shows that, when compared with design storm modeling results, APSWM(Tri) always over-estimates flood peaks while APSWM(Tra) always produce flood peaks closer to design storm results.

For 35% impervious catchments, Fig. 4 together with other cases that are not plotted show similar comparisons. More quantitative comparisons are provided in the following paragraphs.

Using design storm modeling, for a 100% impervious catchment, with no calibration of parameter values, Clark's UH method produces results that are on average 23% higher than those produced by the SCS UH method; where this difference, generally, slightly increases when the time of concentration decreases. For a 35% impervious catchment, similar comparisons of Clark's and SCS UH results were found for the clay soil cases; while for the sandy soil cases, the average difference between Clark's and SCS UH results increases to 26%, which is slightly higher as compared to the cases when the catchments are 100% impervious. The differences between the Clark's and SCS UH methods are consistent across different return periods. The differences in results are mainly due to the fact that Clark's UH

has a sharp peak than that of the SCS UH, which resulted in higher peak discharges from the same input storms. These comparisons demonstrate that the design storm approach itself may produce different results depending on the choice of the synthetic UHs, and the difference in results increases with the soil infiltration capacity. Many jurisdictions do not specify which synthetic UH should be used within their jurisdictions. The comparisons presented here help illustrate the possible range of differences if different UHs are used. This is also why design storm results from one UH method alone should not be used to judge the performance of a new modeling approach.

Quantitative summaries of the comparisons of design storm versus APSWM(Tri) results are given in this and the following paragraphs. For a 100% impervious catchment, on average, the APSWM(Tri) results are higher than those from the SCS UH by 56% and higher than Clark's UH results by 27%. This illustrates that assuming a triangular shape for all runoff event hydrographs does result in some overestimation of the flood peaks by the APSWM(Tri) models. It was also found that the differences in results decrease when t_c increases. This is because for longer times of concentration, the event hydrograph shape exerts less influence on the peak discharges whereas the total volume of runoff has a strong influence on the peak discharges; given that the total volume of runoff for each event estimated by the two different modeling approaches are closer to each other. The differences in results are generally not very consistent across different return periods for $t_c = 0.5$ h. This is

because results from both methods are not very accurate since extremely short but intense rainfall periods produce flood peaks for catchments with extremely short t_c, while these extremely short but intense rainfall periods cannot be represented well by either the design storms or the analytical probabilistic approach. When $t_c = 1.5$ or 3 h, these differences are more consistent across different return periods except for the return period of 2 years. This is because 2-year return period peak discharges are often caused by relatively small rainfall events and for these small rainfall events, the difference in peak discharges caused by different UHs and event hydrograph shapes can be more significant as compared to that for larger rainfall events which produce the higher return period peak discharges. When $t_c = 6$ h, differences in results have better consistency across all different return periods of interest. This demonstrates again that for longer t_c, flood peaks depend more on the longer rainfall events with larger volumes and both the design storm method and the analytical probabilistic approach can represent the frequency distributions of these rainfall events similarly well, the differences in results are caused only by the differences in how the different models model the catchment rainfall-runoff transformation.

For a 35% impervious catchment with clay soils, on average, APSWM(Tri) results are higher than SCS UH results by 47% and higher than Clark's UH results by 19%. Similar to the cases with 100% impervious catchments, the differences in results are consistent across different return periods for the two tested t_c (1.5 and 3 h) groups except for the return period of 2 years. The differences in results also decrease when

t_c increases. For a 35% impervious catchment with sandy soils, APSWM(Tri) results are higher than SCS UH results by 40% and higher than Clark's UH results by 12%. Unlike the cases of 100% impervious catchments and 35% impervious catchments with clay soils, for cases with sandy soils, the differences in results increase more significantly when t_c increases. This may be caused by the difference in rainfall loss calculations, between the different approaches, for extremely permeable soils.

Quantitative summaries of the comparisons of design storm versus APSWM(Tra) results are given in this paragraph. For a 100% impervious catchment, on average, APSWM(Tra) results are higher than SCS UH results by 9%; the differences in results decrease with the increase of t_c. However, APSWM(Tra) results are less than Clark's UH results by an average of 11%. Similar comparisons were found for 35% impervious catchments; where the differences in results also increase with the increase of the soil's infiltration capacities. Generally the differences in results are consistent across different return periods; the main inconsistency appears for the 2-year return period and is more apparent for cases with sandy soils. This is because relatively small rainfall volumes producing the 2-year return period peak discharges and differences in infiltration calculations for sandy soils further complicates the comparisons. The differences in results between APSWM(Tra) and design storm modeling are much less than those between APSWM(Tri) and design storm modeling. The differences in results between APSWM(Tra) and design storm modeling using any one of the two UH methods are even less than the differences in

results from design storm modeling itself using the two different UH methods. Although design storm results from one UH method alone should not be used to judge the performance of a new approach, combination of the comparisons summarized in this and the above paragraphs suggest that APSWM(Tra) provides results closer to those from the design storm approach and is therefore preferable to use in practice than APSMW(Tri).

On average, APSWM(Tra) results are less than APSWM(Tri) results by 30% for 100% impervious catchments, and the difference increases with the increase of t_c. This is because longer t_c would result in more cases with trapezoidal event hydrographs and therefore increase the difference between a model which uniformly adopts triangular event hydrograph shapes and a model which takes into account the possibilities of trapezoidal event hydrographs. Generally, the difference in results of the two APSWM models are consistent for different return periods, the differences only change by 1 to 2% when return period changes. Similar results were observed for 35% impervious catchments; however, the differences in the results of the two APSWM models increased to 34% and some inconsistencies appeared at the low return periods of 2 and 5 years for cases with sandy soils. The larger difference for 35% impervious catchments is caused by the additional difference resulting from infiltration calculations. The 30 ~ 34% difference between APSWM(Tra) and APSWM(Tri) results is larger than the average difference of 23 ~ 26% between design storm results using two different synthetic UHs; justifying the improvement from

APSWM(Tri) to APSWM(Tra).

3.6. Conclusions

The analytical probabilistic stormwater models [e.g. APSWM(Tra) developed in this paper or APSWM(Tri) developed earlier] provide essentially a new approach for obtaining the flood frequency distributions of small catchments where observed flow data are not available. This approach does not require numeric modeling; for any small unban catchments of interest, their flood frequency distribution results are physically-based and can be expressed directly in closed-form mathematical equations. Before this work, in the development of analytical probabilistic stormwater models, individual event hydrographs are approximated as triangles to simplify the derivations. However, in reality, individual hydrograph shapes depend on the catchment time of concentration and rainfall event duration and are not always approximately triangular especially for storms with long durations. In this work, all possible approximating event hydrograph shapes were taken into consideration to derive a new set of equations; the derivation is much more challenging but the results obtained were proven to be much more accurate.

Detailed derivations of APSWM(Tra) [i.e., Eqs. (8) and (9)] are provided in this paper. The results from APSWM(Tra) are compared to results from the design storm approach using two different runoff-routing methods (Clark's and SCS UHs) as well as those from the previously developed analytical probabilistic stormwater model

APSWM(Tri). Real rainfall data and hypothetical catchments with different times of concentration, percentages of imperviousness and soil types are used for testing purposes. The use of real rainfall data ensures equivalent input rainfall information for both the analytical and design storm approaches; and the use of different hypothetical catchments provides versatilities in catchment conditions. Overall, the newly derived analytical probabilistic model APSWM(Tra) performed well as compared to the design storm approach using the Clark's and SCS UHs for catchment runoff routing. The same or very similar comparison results were observed for a wide range of catchment conditions (i.e., response time, imperviousness and soil types). These comparison results demonstrate that, for the rainfall station studied in this paper, simplifying assumptions adopted for the development of APSWM(Tra) are acceptable and APSWM(Tra) provides a reliable alternative to the design storm approach.

Although APSWM(Tra) is intended for use in urban stormwater management studies where future land-use conditions may need to be considered and observed flow data are usually not available, more comparisons preferably using observed flow data to estimate flood frequency distributions at different locations are required to verify further the new analytical probabilistic model. For this additional verification purposes, long-term observed flows from stabilized small urban catchments are required for the estimation of flood frequency distributions. For any one location of interest, such observed flow data are very limited and difficult to find. That is why in

the future APSWM(Tra) may be applied at different locations with results compared to those from both the design storm approach and observed flow data.

Appendix A: Detailed Derivation of the Peak Discharge Exceedence Probabilities

3.A.1. Regions of Integration

Eq. (7) shows the derivation of $P[Q_p = 0]$ using the derived probability distribution theory. Similarly, for a given non-zero peak discharge rate q_p, based on Eq. (4), the exceedence probability $P[Q_p > q_p]$ consists of four parts: (1) $P_1[Q_p > q_p]$ which is the exceedence probability with $S_{di} < v \leq S_{il} + f_c t$ and $t \leq t_c$, (2) $P_2[Q_p > q_p]$ which is the exceedence probability with $v > S_{il} + f_c t$ and $t \leq t_c$, (3) $P_3[Q_p > q_p]$ which is the exceedence probability with $S_{di} < v \leq S_{il} + f_c t$ and $t > t_c$, and (4) $P_4[Q_p > q_p]$ which is the exceedence probability with $v > S_{il} + f_c t$ and $t > t_c$. Thus, for $q_p > 0$,

$$P[Q_p > q_p] = P_1[Q_p > q_p] + P_2[Q_p > q_p] + P_3[Q_p > q_p] + P_4[Q_p > q_p] \quad \text{(A.1)}$$

According to the above definitions of P_1, P_2, P_3, and P_4, their corresponding sub-regions of integrations (referred to as R_1, R_2, R_3, and R_4, respectively) are mutually exclusive. R_1 is defined by

$$\begin{cases} \dfrac{h(v - S_{di})}{t_c} > q_p \\ S_{di} < v \leq S_{il} + f_c t \\ t \leq t_c \end{cases} \xrightarrow{yields} \begin{cases} \dfrac{q_p t_c}{h} + S_{di} < v \leq S_{il} + f_c t \\ t \leq t_c \end{cases}$$

Let L_1 be the line defined by $v = f_c t + S_{il}$ and L_2 be the line defined by $v = \frac{q_p t_c}{h} + S_{di}$ (both L_1 and L_2, and all other lines that will be defined later are plotted with v on the vertical axis and t on the horizontal axis); then R_1 is the area above L_2 but below L_1, and also on the left side of the vertical line defined by $t = t_c$.

R_2 is defined by

$$\begin{cases} \dfrac{v - S_d - f_c(1-h)t}{t_c} > q_p \\ v > S_{il} + f_c t \\ t \leq t_c \end{cases} \xrightarrow{yields} \begin{cases} v > q_p t_c + S_d + (1-h) f_c t \\ v > S_{il} + f_c t \\ t \leq t_c \end{cases}$$

Let L_3 be the line defined by $v = f_c(1-h)t + t_c q_p + S_d$, then R_2 is the area above L_1 and L_3, and also on the left side of the vertical line $t = t_c$.

R_3 is defined by

$$\begin{cases} \dfrac{h(v - S_{di})}{t} > q_p \\ S_{di} < v \leq S_{il} + f_c t \\ t > t_c \end{cases} \xrightarrow{yields} \begin{cases} \dfrac{q_p t}{h} + S_{di} < v \leq S_{il} + f_c t \\ t > t_c \end{cases}$$

Let L_4 be the line defined by $v = \frac{q_p}{h} t + S_{di}$, then R_3 is the area above L_4 but below L_1, and also on the right side of the vertical line $t = t_c$.

R_4 is defined by

$$\begin{cases} \dfrac{v - S_d - f_c(1-h)t}{t} > q_p \\ v > S_{il} + f_c t \\ t > t_c \end{cases} \xrightarrow{yields} \begin{cases} v > \left[q_p + (1-h)f_c\right]t + S_d \\ v > S_{il} + f_c t \\ t > t_c \end{cases}$$

Let L_5 be the line defined by $v = \left[f_c(1-h) + q_p\right]t + S_d$, then R_4 is the area above both L_1 and L_5, and also on the right side of the vertical line $t = t_c$.

In summary, R_1 and R_2 are on the left side of the vertical line $t = t_c$ and bounded by lines L_1, L_2, and L_3; R_3 and R_4 are on the right side of $t = t_c$ and bounded by lines L_1, L_4, and L_5. In the following, the relative location of the five non-vertical lines are examined and for simplicity, R_1 and R_2 which are bounded by the segments of lines L_1, L_2, and L_3 with $0 < t \leq t_c$ are combined together; and R_3 and R_4 which are bounded by the segments of lines L_1, L_4, and L_5 with $t > t_c$ are also combined together.

3.A.2. Delineation of the Exact Sub-regions of Integration

The exact sub-regions of integrations depend on the magnitudes of the intercepts of lines L_1, L_2, and L_3, the v-values of these lines intersecting with the vertical line $t = t_c$ and the slopes of lines L_1, L_4, and L_5. Let I_1, I_2, and I_3 be the intercepts of the lines L_1, L_2, and L_3, respectively; let I_{1c}, I_{2c}, and I_{3c} be the v-values of the intersection points between the vertical line $t = t_c$ and lines L_1, L_2, and L_3, respectively. Since lines L_2 and L_3 intersects L_4 and L_5, respectively, at $t = t_c$; the v-values of the intersection points between lines L_4 and L_5 and the vertical line $t = t_c$ are also I_{2c} and I_{3c}, respectively. Let S_1, S_2, S_3, S_4 and S_5 be the slopes of lines L_1, L_2,

L_3, L_4, and L_5, respectively. The relative magnitudes of the slopes and intercepts of these lines depend on the values of h, S_{il}, S_{di}, t_c, f_c, and q_p and can be deduced as follows.

Since $I_1 - I_2 = S_{il} - \left(\frac{q_p t_c}{h} + S_{di}\right) = S_{dd} - \frac{q_p t_c}{h}$ where $S_{dd} = S_{il} - S_{di}$, then $I_1 \geq I_2$ if $q_p \leq \frac{hS_{dd}}{t_c}$ and vice versa. Since $I_1 - I_3 = S_{il} - \left(q_p t_c + S_d\right) = hS_{dd} - q_p t_c$ thus $I_1 \geq I_3$ if $q_p \leq \frac{hS_{dd}}{t_c}$ and vice versa. Since $I_2 - I_3 = \frac{q_p t_c}{h} + S_{di} - \left(q_p t_c + S_d\right) = \frac{(1-h)q_p t_c}{h} - (1-h)S_{dd}$ thus $I_2 \leq I_3$ if $q_p \leq \frac{hS_{dd}}{t_c}$ and vice versa. Consequently, $I_1 \geq I_3 \geq I_2$ if $q_p \leq \frac{hS_{dd}}{t_c}$ and $I_2 > I_3 > I_1$ if $q_p > \frac{hS_{dd}}{t_c}$.

To determine the relative locations of lines L_1, L_2 and L_3, it is also necessary to know the relative magnitudes of their slopes (S_1, S_2, and S_3) or their v-values when intersecting with line $t = t_c$ (i.e., I_{1c}, I_{2c} and I_{3c}). The relative magnitudes of I_{1c}, I_{2c}, and I_{3c} are investigated since it is also required for the delineation of the sub-regions of integration when $t > t_c$. Since $I_{1c} - I_{2c} = f_c t_c + S_{il} - \left(\frac{q_p t_c}{h} + S_{di}\right) = f_c t_c + S_{dd} - \frac{q_p t_c}{h}$ thus $I_{1c} \geq I_{2c}$ when $q_p \leq hf_c + \frac{hS_{dd}}{t_c}$ and vice versa. Since $I_{1c} - I_{3c} = f_c t_c + S_{il} - \left((1-h)f_c t_c + q_p t_c + S_d\right) = hf_c t_c + hS_{dd} - q_p t_c$ thus $I_{1c} \geq I_{3c}$ if $q_p \leq hf_c + \frac{hS_{dd}}{t_c}$ and vice versa. Since $I_{2c} - I_{3c} = \frac{q_p t_c}{h} + S_{di} - \left((1-h)f_c t_c + q_p t_c + S_d\right) = \frac{(1-h)q_p t_c}{h} - (1-h)f_c t_c - (1-h)S_{dd}$ thus $I_{2c} \leq I_{3c}$ if $q_p \leq hf_c + \frac{hS_{dd}}{t_c}$ and vice versa. Consequently, $I_{1c} \geq I_{3c} \geq I_{2c}$ if $q_p \leq hf_c + \frac{hS_{dd}}{t_c}$ and $I_{2c} > I_{3c} > I_{1c}$ if $q_p > hf_c + \frac{hS_{dd}}{t_c}$.

The relative locations of lines L_1, L_4, and L_5 with $t > t_c$ can be determined by examining the relative magnitudes of the slopes S_1, S_4, and S_5 along with the already

determined relative magnitudes of I_{1c}, I_{2c}, and I_{3c}. Since $S_1 - S_4 = f_c - \frac{q_p}{h}$ thus $S_1 \geq S_4$ if $q_p \leq hf_c$ and vice versa. Since $S_1 - S_5 = f_c - \left(q_p + (1-h)f_c\right) = hf_c - q_p$ thus $S_1 \geq S_5$ if $q_p \leq hf_c$ and vice versa. Since $S_4 - S_5 = \frac{q_p}{h} - \left(q_p + (1-h)f_c\right) = \frac{(1-h)}{h}q_p - (1-h)f_c$ thus $S_4 \leq S_5$ if $q_p \leq hf_c$ and vice versa. Consequently, $S_1 \geq S_5 \geq S_4$ if $q_p \leq hf_c$ and $S_4 > S_5 > S_1$ if $q_p > hf_c$.

The total $P[Q_p > q_p]$ can be determined by summing up the four exceedence probabilities defined by the four sub-regions of integrations. This summation has to be carried out separately according to (a) whether q_p is less than $\frac{hS_{dd}}{t_c}$ or not, (b) whether q_p is less than $(hf_c + \frac{hS_{dd}}{t_c})$ or not, and (c) whether q_p is less than hf_c or not. Since the objective here is to obtain $P[Q_p > q_p]$ for any possible q_p value, all possible q_p values need to be divided into intervals so that within each interval q_p that satisfies all the above-listed conditions is clear and certain. The endpoints of the q_p intervals as described above for the determination of the relative magnitudes of the intercepts (I_1, I_2, I_3, I_{1c}, I_{2c}, and I_{3c}) and slopes (S_1, S_4, and S_5) are $\frac{hS_{dd}}{t_c}$, $hf_c + \frac{hS_{dd}}{t_c}$, and hf_c. The relative magnitudes between these three endpoints will become clear once the relative magnitude between f_c and $\frac{S_{dd}}{t_c}$ is known. For a given catchment, its f_c value may be less than (or equal to) or greater than $\frac{S_{dd}}{t_c}$. Thus, for subsequent derivation purposes, Type I catchments are defined as those with $f_c \leq \frac{S_{dd}}{t_c}$ and Type II catchments are those with $f_c > \frac{S_{dd}}{t_c}$. Table A.1 summarizes the relative magnitudes of the slopes and intercepts necessary to delineate the sub-regions of integration for the two types of catchments.

For ease of reference, the four q_p intervals as defined in Table A.1 are referred to as Intervals 1, 2, 3, and 4. It can be seen from the above that by dividing the possible non-zero q_p values into four consecutive intervals and all possible catchments into two types, the relative locations of the five lines can be determined and the exact sub-regions of integration can be delineated.

3.A.3. Configuration of the Regions of Integration and Derivation of the Peak Discharge Exceedence Probability

The region of integration described below is the overall region of integration required for the determination of $P[Q_p > q_p]$, which is the union of sub-regions R_1 through R_4, where R_1, R_2, R_3, and R_4 are mutually exclusive sub-regions required for the determination of $P_1[Q_p > q_p]$, $P_2[Q_p > q_p]$, $P_3[Q_p > q_p]$, and $P_4[Q_p > q_p]$, respectively. Here $P_1[Q_p > q_p]$, $P_2[Q_p > q_p]$, $P_3[Q_p > q_p]$, and $P_4[Q_p > q_p]$ are parts of $P[Q_p > q_p]$. For each type of catchment, for the determination of $P[Q_p > q_p]$, there are four different regions of integration depending on whether q_p is in Intervals 1, 2, 3, or 4. This is because, as shown in Table A.1, depending on whether q_p is in Intervals 1, 2, 3, or 4, the relative magnitudes of the intercepts and slopes of the five lines forming the regions of integration are different. These four regions of integration are different for the four different q_p value intervals.

For the two types of catchments, despite the difference on the endpoints of the q_p intervals, if q_p falls into Interval 1, 3, or 4, the relative magnitudes of the intercepts

and slopes of the five lines are the same for both types of catchments. Thus, there are a total of five different regions (i.e., overall regions of integration for determining $P[Q_p > q_p]$) of integration for the two types of catchments. The configurations of these regions of integrations are shown in Figs. A.1 through A.5 and they serve as a basis for the derivation of the exceedence probability of interest. In the following, each overall region of integration is described together with the derivation for the required exceedence probability.

The first overall region of integration configuration is applicable to both types of catchments with q_p taking on values from Interval 1. It can be seen from Table A.1 that Type I catchments when $0 < q_p \leq hf_c$ and Type II catchments when $0 < q_p \leq \frac{hS_{dd}}{t_c}$ have the same relative magnitudes of the slopes and intercepts of the lines forming the region of integration (Fig. A.1). For $0 < t \leq t_c$, the lines L_1, L_2, and L_3 do not intersect because $I_1 \geq I_3 \geq I_2$ and $I_{1c} \geq I_{3c} \geq I_{2c}$ and as shown in Fig. A.1, the combination of sub-regions R_1 and R_2 can be simply represented as $0 < t \leq t_c$ and $v > \frac{q_p t_c}{h} + S_{di}$; thus

$$P_1[Q_p > q_p] + P_2[Q_p > q_p] = \int_0^{t_c} \int_{\frac{q_p t_c}{h} + S_{di}}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dv dt$$

Carrying out the above integrations, the following is obtained:

$$P_1[Q_p > q_p] + P_2[Q_p > q_p] = \exp\left(-\frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) - \exp\left(-\lambda t_c - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) \quad \text{(A.2)}$$

For $t > t_c$, the lines L_1, L_4, and L_5 are also shown in Fig. A.1, and the sub-regions R_3

and R_4 combined together can be simply represented as $t > t_c$ and $v > \frac{q_p t_c}{h} + S_{di}$, thus

$$P_3[Q_p > q_p] + P_4[Q_p > q_p] = \int_{t_c}^{\infty} \int_{\frac{q_p t}{h}+S_{di}}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dv dt$$

Carrying out the above integrations, the following is obtained:

$$P_3[Q_p > q_p] + P_4[Q_p > q_p] = \frac{\lambda h}{\lambda h + \zeta q_p} \exp\left(-\lambda t_c - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) \quad \text{(A.3)}$$

The total exceedence probability is the sum of the right-hand-sides (RHSs) of Eqs. (9) and (10), i.e.,

$$P[Q_p > q_p] = \exp\left(-\frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) - \frac{\zeta q_p}{\lambda h + \zeta q_p} \exp\left(-\lambda t_c - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) \quad \text{(A.4)}$$

The second overall region of integration configuration is applicable to Type I catchments when $hf_c < q_p \leq \frac{hS_{dd}}{t_c}$ (Fig. A.2). The part of this region when $0 < t \leq t_c$ is the same as in the first overall region of integration; consequently, the exceedence probability $P_1[Q_p > q_p] + P_2[Q_p > q_p]$ is the same as expressed in Eq. (A.2). For $t > t_c$, the relative magnitudes of the slopes and intercepts of lines L_1, L_4, and L_5 (Table A.1) result in the intersection of these three lines. By equaling the RHSs of the equations representing each pair of lines out of the total three lines, it was found that the three lines intersect at a common point with co-ordinates ($t_{145} = \frac{hS_{dd}}{q_p - hf_c}$, $v_{145} = \frac{S_{il} q_p - hf_c S_{di}}{q_p - hf_c}$).

As shown in Fig. 3, the combination of sub-regions R_3 and R_4 is the area above L_4 for

$t_c < t \leq t_{145}$ plus the area above L_5 for $t > t_{145}$; thus

$$P_3[Q_p > q_p] + P_4[Q_p > q_p] = \int_{t_c}^{t_{145}} \int_{\frac{q_p t}{h}+S_{di}}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt + \int_{t_{145}}^{\infty} \int_{[f_c(1-h)+q_p]t+s_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt$$

Carrying out the above integrations, the following is obtained:

$$P_3[Q_p > q_p] + P_4[Q_p > q_p] = \frac{\lambda h}{\lambda h + \zeta q_p} \exp\left(-\lambda t_c - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) + \frac{(1-h)\lambda\zeta(q_p - hf_c)}{(\lambda h + \zeta q_p)(\lambda + \zeta q_p + (1-h)\zeta f_c)} \exp\left(\frac{-\zeta S_{il} q_p + h\zeta f_c S_{di} - h\lambda S_{dd}}{q_p - hf_c}\right) \quad \text{(A.5)}$$

The total exceedence probability is the sum of the RHSs of Eqs. (A.2) and (A.5), i.e.,

$$P[Q_p > q_p] = \exp\left(-\frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) - \frac{\zeta q_p}{\lambda h + \zeta q_p} \exp\left(-\lambda t_c - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) + \frac{(1-h)\lambda\zeta(q_p - hf_c)}{(\lambda h + \zeta q_p)(\lambda + \zeta q_p + (1-h)\zeta f_c)} \exp\left(\frac{-\zeta S_{il} q_p + h\zeta f_c S_{di} - h\lambda S_{dd}}{q_p - hf_c}\right) \quad \text{(A.6)}$$

The third overall region of integration configuration is applicable for Type II catchments when $\frac{hs_{dd}}{t_c} < q_p \leq f_c h$ (Fig. A.3). For $0 < t \leq t_c$, the relative magnitudes of the slopes and intercepts of the three lines are shown in Table 1. The segments of lines L_1, L_2, and L_3 with $0 < t \leq t_c$ intersect each other because $I_2 \geq I_3 \geq I_1$ but $I_{1c} \geq I_{3c} \geq I_{2c}$. They were found to intersect at a common point with co-ordinates $(t_{123} = \frac{q_p t_c - hS_{dd}}{hf_c}, v_{123} = \frac{q_p t_c}{h} + S_{di})$. As shown in Fig. A.3, the combination of sub-regions R_1 and R_2 is the area above L_3 when $0 < t \leq t_{123}$ plus the area above L_2 when $t_{123} < t \leq t_c$; thus

$$P_1[Q_p > q_p] + P_2[Q_p > q_p] = \int_0^{t_{123}} \int_{f_c(1-h)t+q_p t_c+S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt$$

$$+ \int_{t_{123}}^{t_c} \int_{\frac{q_p t_c}{h}+S_{di}}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt$$

Carrying out the above integrations, the following is obtained:

$$P_1[Q_p > q_p] + P_2[Q_p > q_p] = \frac{\lambda}{\lambda + (1-h)\zeta f_c} \exp(-\zeta q_p t_c - \zeta S_d)$$

$$- \exp\left(-\lambda t_c - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right)$$

$$+ \frac{(1-h)\zeta f_c}{\lambda + (1-h)\zeta f_c} \exp\left(\frac{-f_c \zeta q_p t_c - h\zeta f_c S_{di} - \lambda q_p t_c + h\lambda S_{dd}}{h f_c}\right) \quad \text{(A.7)}$$

For $t > t_c$, the relative magnitudes of the slopes and intercepts of lines L_1, L_4, and L_5 are the same as in the first overall region of integration (Table A.1); consequently the same exceedence probability is produced (i.e. Eq. (A.3)). The total exceedence probability is the sum of the RHSs of Eqs. (A.3) and (A.7), which is:

$$P[Q_p > q_p]$$

$$= \frac{\lambda}{\lambda + (1-h)\zeta f_c} \exp(-\zeta q_p t_c - \zeta S_d)$$

$$+ \frac{(1-h)\zeta f_c}{\lambda + (1-h)\zeta f_c} \exp\left(\frac{-f_c \zeta q_p t_c - h\zeta f_c S_{di} - \lambda q_p t_c + h\lambda S_{dd}}{h f_c}\right)$$

$$- \frac{\zeta q_p}{\lambda h + \zeta q_p} \exp\left(-\lambda t_c - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) \quad \text{(A.8)}$$

The fourth overall region of integration configuration is applicable for Type I catchments when $\frac{hS_{dd}}{t_c} < q_p \leq hf_c + \frac{hS_{dd}}{t_c}$ and Type II catchments when $hf_c < q_p \leq hf_c + \frac{hS_{dd}}{t_c}$, they both have the same relative magnitudes of the slopes and intercepts of

the five lines (Table A.1) forming the regions of integration (Fig. A.4).

The first part where $0 < t \leq t_c$ is equivalent to the first part of the third overall region of integration configuration, therefore Eq. (A.7) is applicable here as well. The portion of Fig. A.4 with $t > t_c$ is the same as that of Fig. A.2 for the second overall region of integration configuration, therefore Eq. (A.5) is applicable here too. Thus, the total exceedence probability for the fourth overall region of integration configuration is the sum of the RHSs of Eqs. (A.5) and (A.7); which is

$$P[Q_p > q_p] = \frac{\lambda}{\lambda + (1-h)\zeta f_c}\exp(-\zeta q_p t_c - \zeta S_d)$$
$$-\frac{\zeta q_p}{\lambda h + \zeta q_p}\exp\left(-\lambda t_c - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right)$$
$$+\frac{(1-h)\zeta f_c}{\lambda + (1-h)\zeta f_c}\exp\left(\frac{-f_c\zeta q_p t_c - h\zeta f_c S_{di} - \lambda q_p t_c + h\lambda S_{dd}}{h f_c}\right)$$
$$+\frac{(1-h)\lambda\zeta(q_p - hf_c)}{(\lambda h + \zeta q_p)(\lambda + \zeta q_p + (1-h)\zeta f_c)}\exp\left(\frac{-\zeta S_{il} q_p + h\zeta f_c S_{di} - h\lambda S_{dd}}{q_p - hf_c}\right) \quad \text{(A.9)}$$

The fifth overall region of integration configuration is applicable for both Type I and Type II catchments with $q_p > hf_c + \frac{hS_{dd}}{t_c}$, they both have the same relative magnitudes of the slopes and intercepts of the lines (Table A.1) forming the regions of integration (Fig. A.5). The segments of lines L_1, L_2, and L_3 with $0 < t \leq t_c$ do not intersect. As shown in Fig. A.5, sub-region R_1 does not exist and R_2 is just the area above L_3 with $0 < t \leq t_c$; thus

$$P_1[Q_p > q_p] + P_2[Q_p > q_p]$$

$$= \int_0^{t_c} \int_{f_c(1-h)t+q_p t_c+S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dv d$$

$$= \frac{\lambda}{\lambda + (1-h)\zeta f_c} \exp\left(-\zeta q_p t_c - \zeta S_d\right) [1$$

$$- \exp(-\lambda t_c - (1-h)\zeta f_c t_c)] \quad \text{(A.10)}$$

For $t > t_c$, according to the relative magnitudes of the slopes and intercepts of lines L_1, L_4, and L_5 shown in Table 1, lines L_1, L_4, and L_5 do not intersect (Fig. A.5), sub-region R_3 does not exist and R_4 is just the area above L_5 with $t > t_c$; thus

$$P_3[Q_p > q_p] + P_4[Q_p > q_p]$$

$$= \int_{t_c}^{\infty} \int_{(f_c(1-h)+q_p)t+s_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dv dt$$

$$= \frac{\lambda}{\lambda + \zeta q_p + (1-h)\zeta f_c} \exp\left(-\lambda t_c - \zeta q_p t_c - (1-h)\zeta f_c t_c - \zeta S_d\right) \quad \text{(A.11)}$$

The total exceedence probability is the sum of the RHSs of Eqs. (A.10) and (A.11), which is

$$P[Q_p > q_p] = \frac{\lambda}{\lambda + (1-h)\zeta f_c} \exp\left(-\zeta q_p t_c - \zeta S_d\right)$$

$$- \frac{\lambda\zeta q_p}{(\lambda + \zeta q_p + (1-h)\zeta f_c)(\lambda + (1-h)\zeta f_c)} \exp\left(-\lambda t_c\right.$$

$$\left. - \zeta q_p t_c - (1-h)\zeta f_c t_c - \zeta S_d\right) \quad \text{(A.12)}$$

Eqs. (A.4), (A.6), (A.8), (A.10), and (A.12) provide analytically the exceedence probability of peak discharge per rainfall event.

Acknowledgements: The authors thank Dr. Salvatore Grimaldi and the anonymous

reviewers for their comments and suggestions. This work was supported by the Natural Sciences and Engineering Research Council of Canada. This work was also partially completed within the framework of the Panta Rhei Research Initiative of the International Association of Hydrological Sciences.

References

Adams, B. J., Fraser, H. G., Howard, C. D. D., and Hanafy, M. S. (1986). "Meteorologic data analysis for drainage system design." *Journal of Environmental Engineering*, ASCE, 112(5), 827–848.

Adams, B. J., and Howard, C. D. (1986). "Design storm pathology." *Canadian Water Resources Journal*, *11*(3), 49-55.

Adams, B. J. and Papa, F. (2000). *Urban Stormwater Management Planning with Analytical Probabilistic Models*, John Wiley & Sons, Inc., New York, NY, USA.

Akan, A. O., & Houghtalen, R. J. (2003). *Urban Hydrology, Hydraulics, and Stormwater Quality: engineering applications and computer modeling*. John Wiley & Sons, Inc., New York, NY, USA.

Bacchi, B., Balistrocchi, M., and Grossi, G. (2008). "Proposal of a semi-probabilistic approach for storage facility design." *Urban Water Journal*, 5(3), 195-208.

Balistrocchi, M., Grossi, G., and Bacchi, B. (2009). "An analytical probabilistic model of the quality efficiency of a sewer tank." *Water Resources Research*, 45(12), W12420.

Chen, J., and Adams, B. J., (2005). "Urban storm water control evaluation with analytical probabilistic models." *Journal of water resources planning and management*, 131(5), 362-374.

Chen, J., and Adams, B. J. (2007). "Development of analytical models for estimation of urban stormwater runoff." *Journal of hydrology*, 336(3), 458-469.

Cheng, K., Wai, C., Cheng, Y., and Yeh, H. (2003). "Effect of spatial variation characteristics on contouring of design storm depth." *Hydrological processes*, 17(9), 1755-1769.

Eagleson, P. S. (1972). "Dynamics of flood frequency." *Water Resources Research*, 8(4), 878–898.

Eagleson, P. S. (1978). "Climate, soil, and vegetation, 2, the distribution of annual precipitation derived from observed storm sequences." *Water Resources Research*, 14(5), 713–721.

Grimaldi, S., Petroselli, A., Serinaldi, F. (2012). "A continuous simulation model for design-hydrograph estimation in small and ungauged watersheds." *Hydrological Sciences Journal*, 57(6), 1035-1051. doi: 10.1080/02626667.2012.702214

Grimaldi, S., Petroselli, A., Serinaldi, F. (2012). "Design hydrograph estimation in small and ungauged watersheds: continuous simulation method versus event-based approach." *Hydrological Processes*, 26, 3124-3134. doi: 10.1002/hyp.8384

Grimaldi, S., Petroselli, A., Arcangeletti, E., Nardi, F. (2013). "Flood mapping in ungauged basins using fully continuous hydrologic-hydraulic modelling."

Journal of Hydrology, 487, 39-47. http://dx.doi.org/10.1016/j.hydrol.2013.02.023.

Guo, Y. (2001). "Hydrologic design of urban flood control detention ponds." *Journal of Hydrologic Engineering*, 6(6), 472–479.

Guo, Y., and Adams, B. J. (1998a). "Hydrologic analysis of urban catchments with event-based probabilistic models: 1. Runoff volume." *Water Resources Research*, 34(12), 3421-3431.

Guo, Y., and Adams, B. J. (1998b). "Hydrologic analysis of urban catchments with event-based probabilistic models: 2. Peak discharge rate." *Water Resources Research*, 34(12), 3433-3443.

Guo, Y., and Adams, B. J. (1999a). "Analysis of detention ponds for storm water quality control." *Water Resources Research,* 35(8), 2447-2456.

Guo, Y., and Adams, B. J. (1999b). "An analytical probabilistic approach to sizing flood control detention facilities." *Water Resources Research,* 35(8), 2457-2468.

Guo, Y., and Baetz, B. W. (2007). "Sizing of rainwater storage units for green building applications." *Journal of Hydrologic Engineering*, 12(2), 197– 205.

Guo, Y., Hansen, D. and Li, C. (2009). "Probabilistic approach to estimating the effects of channel reaches on flood frequencies." *Water Resources Research,* 45, W08404, doi:10.1029/2008WR007387.

Guo, Y. and Markus, M. (2011). "An Analytical Probabilistic Approach for Estimating Design Floods of Small Watersheds." *Journal of Hydrologic Engineering*, Vol.

16, No. 11, 847-857, doi:10.1061/(ASCE)HE.1943-5584.0000380.

Guo, Y., and Zhuge, Z. (2008). "Analytical probabilistic flood routing for urban stormwater management purposes." *Canadian Journal of Civil Engineering*, *35*(5), 487-499.

Hassini, S. and Guo, Y. (2016). "Exponentiality Test Procedures for Large Samples of Rainfall Event Characteristics." *Journal of Hydrologic Engineering, 21(4),* 04016003, doi:10.1061/(ASCE)HE.193-5584.0001352.

Hicks, W. I. (1944). "A method of computing urban runoff." *Transactions of the American Society of Civil Engineers*, *109*(1), 1217-1253.

Howard, C. D. D. (1976). "Theory of storage and treatment plant overflows." *Journal of the Environmental Engineering Division,* ASCE, 102(EE4), 709–722.

Levy, B., and McCuen, R. (1999). "Assessment of storm duration for hydrologic design." *Journal of hydrologic engineering*, 4(3), 209-213.

Mejía, A., E. Daly, F. Rossel, T. Jovanovic, and J. Gironás (2014). "A stochastic model of streamflow for urbanized basins." *Water Resour. Res., 50, 1984–2001, doi:10.1002/2013WR014834.*

Nnadi, F. N., Kline, F. X., Wary, H. L., and Wanielista, M. P. (1999). "Comparison of design storm concepts using continuous simulation with short duration storms." *Journal of the American Water Resources Association*, 35(1), 61-72.

National Climatic Data Center (2011). "U. S. Climate Normals (1971-2000)." *Golden Gate Weather Services.* <http://ggweather.com/normals/index71.htm>

(September 2011)

NOAA's National Weather Service (2015). "The Precipitation Frequency Data Server" <http://hdsc.nws.noaa.gov/hdsc/pfds/> (April 2015).

Quader, A., and Guo, Y. (2006). "Peak discharge estimation using analytical probabilistic and design storm approaches." *Journal of hydrologic engineering*, 11(1), 46-54.

Ponce, V. M. (1989). *Engineering Hydrology: Principles and Practices*, Prentice Hall, New Jersey, USA.

Rivera, P., Gironas J., Montt, J. P., and Fernandez, B. (2005). "An analytical model for hydrologic analysis in urban watersheds." *10 th International Conference on Urban Drainage, Copenhagen, Denmark, 21-26 August*.

Ruffner, J. A., and Bair, F. E. (1978). *Climates of the States, with current tables of normals 1941-1970 and means and extremes to 1975*, Gale Research Co., Detroit, MI, USA.

Scharffenberg, W., and Fleming, M. (2009). *HEC-HMS Hydrologic Modeling System-User's Manual (Version 3.4)*. US Army Corps of Engineers-Hydrologic Engineering Center. USA.

Viessman Jr, W., & Lewis, G. L. (2003). *Introduction to Hydrology, Fifth Edition.* Pearson Education, Inc., Upper Saddle River, NJ, 187-190.

Wanielista, M. P., and Yousef, Y. A. (1993). *Stormwater Management*, John Wiley & Sons, Inc., New York, NY.

Wurbs, R. A., and James, P. W. (2002). *Water Resources Engineering*. Prentice-Hall, Inc., NJ, USA.

Zhang, S., and Guo, Y. (2013a). "Analytical Probabilistic Model for Evaluating the Hydrologic Performance of Green Roofs." *Journal of Hydrologic Engineering*, 18(1), 19-28.

Zhang, S., and Guo, Y. (2013b). "Explicit Equation for Estimating Storm-Water Capture Efficiency of Rain Gardens." *Journal of Hydrologic Engineering*, 18(12), 1739-1748.

Table 1. Partial-duration precipitation depths of the St. Cloud station for the design storm duration of 24 hours

Return Period (years)	2	5	10	25	50	100
Depth (mm)	69	85	100	123	142	161

Table A. 1. Relative magnitudes of the slopes and intercepts of the five lines necessary for the delineation of the sub-regions of integration

Type I Catchments with $f_c \leq \frac{S_{dd}}{t_c}$			
$0 < q_p \leq hf_c$ (Interval 1)	$hf_c < q_p \leq \frac{hS_{dd}}{t_c}$ (Interval 2)	$\frac{hS_{dd}}{t_c} < q_p \leq hf_c + \frac{hS_{dd}}{f_c}$ (Interval 3)	$q_p > hf_c + \frac{hS_{dd}}{t_c}$ (Interval 4)
$I_1 \geq I_3 \geq I_2$ $I_{1c} \geq I_{3c} \geq I_{2c}$ $S_1 \geq S_5 \geq S_4$	$I_1 \geq I_3 \geq I_2$ $I_{1c} \geq I_{3c} \geq I_{2c}$ $S_4 \geq S_5 \geq S_1$	$I_2 \geq I_3 \geq I_1$ $I_{1c} \geq I_{3c} \geq I_{2c}$ $S_4 \geq S_5 \geq S_1$	$I_2 \geq I_3 \geq I_1$ $I_{2c} \geq I_{3c} \geq I_{1c}$ $S_4 \geq S_5 \geq S_1$
Type II Catchments with $f_c > \frac{S_{dd}}{t_c}$			
$0 < q_p \leq \frac{hS_{dd}}{t_c}$ (Interval 1)	$\frac{hS_{dd}}{t_c} < q_p \leq hf_c$ (Interval 2)	$hf_c < q_p \leq hf_c + \frac{hS_{dd}}{t_c}$ (Interval 3)	$q_p > hf_c + \frac{hS_{dd}}{t_c}$ (Interval 4)
$I_1 \geq I_3 \geq I_2$ $I_{1c} \geq I_{3c} \geq I_{2c}$ $S_1 \geq S_5 \geq S_4$	$I_2 \geq I_3 \geq I_1$ $I_{1c} \geq I_{3c} \geq I_{2c}$ $S_1 \geq S_5 \geq S_4$	$I_2 \geq I_3 \geq I_1$ $I_{1c} \geq I_{3c} \geq I_{2c}$ $S_4 \geq S_5 \geq S_1$	$I_2 \geq I_3 \geq I_1$ $I_{2c} \geq I_{3c} \geq I_{1c}$ $S_4 \geq S_5 \geq S_1$

Figure Captions

Fig. 1. Individual-storm hydrograph shape: isosceles trapezoids when $(a)\ t < t_c$ or $(c)\ t > t_c$ and isosceles triangle when $(b)\ t = t_c$

Fig. 2. Scatter plot of rainfall event volume and duration

Fig. 3. Comparison of analytical and design storm modeling results for 100% impervious catchments with times of concentration of (a) 0.5 h, (b) 1.5 h, (c) 3 h, and (d) 6 h.

Fig. 4. Comparison of analytical and design storm modeling results for 35% impervious catchments with (a) $t_c = 1.5$ h, clay soil; (b) $t_c = 1.5$ h, sandy soil; (c) $t_c = 3$ h, clay soil; and (d) $t_c = 3$ h, sandy soil

Fig. A.1. Region of integration for Type I catchments with $q_p \leq f_c h$ and Type II catchments with $q_p \leq \frac{hs_{dd}}{t_c}$

Fig. A.2. Region of integration for Type I catchments with $hf_c < q_p \leq \frac{hs_{dd}}{t_c}$

Fig. A.3. Region of integration for Type II catchments with $\frac{hs_{dd}}{t_c} < q_p \leq hf_c$

Fig. A.4. Region of integration for Type I catchments with $\frac{hs_{dd}}{t_c} < q_p \leq \frac{hs_{dd}}{t_c} + hf_c$ and Type II catchments with $hf_c < q_p \leq \frac{hs_{dd}}{t_c} + hf_c$

Fig. A.5. Region of integration for both types of catchments with $q_p > \frac{hs_{dd}}{t_c} + hf_c$

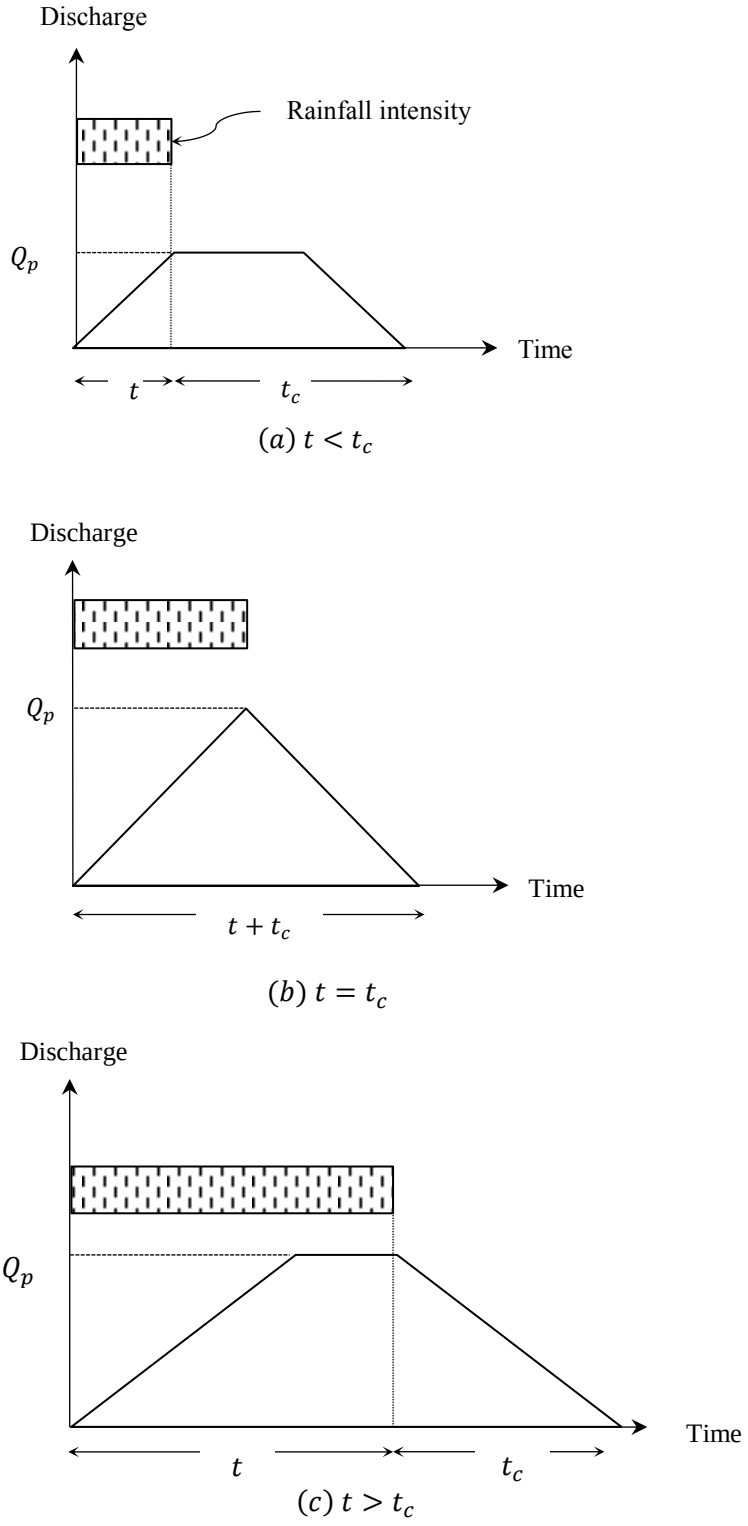

Fig. 1. Individual-storm hydrograph shape: isosceles trapezoids when $(a)\ t < t_c$ or $(c)\ t > t_c$ and isosceles triangle when $(b)\ t = t_c$

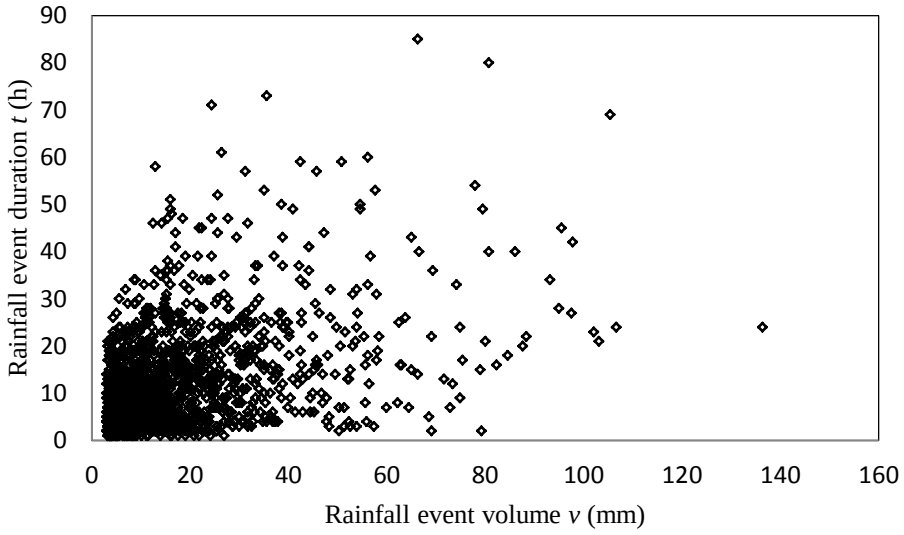

Fig. 2. Scatter plot of rainfall event volume and duration

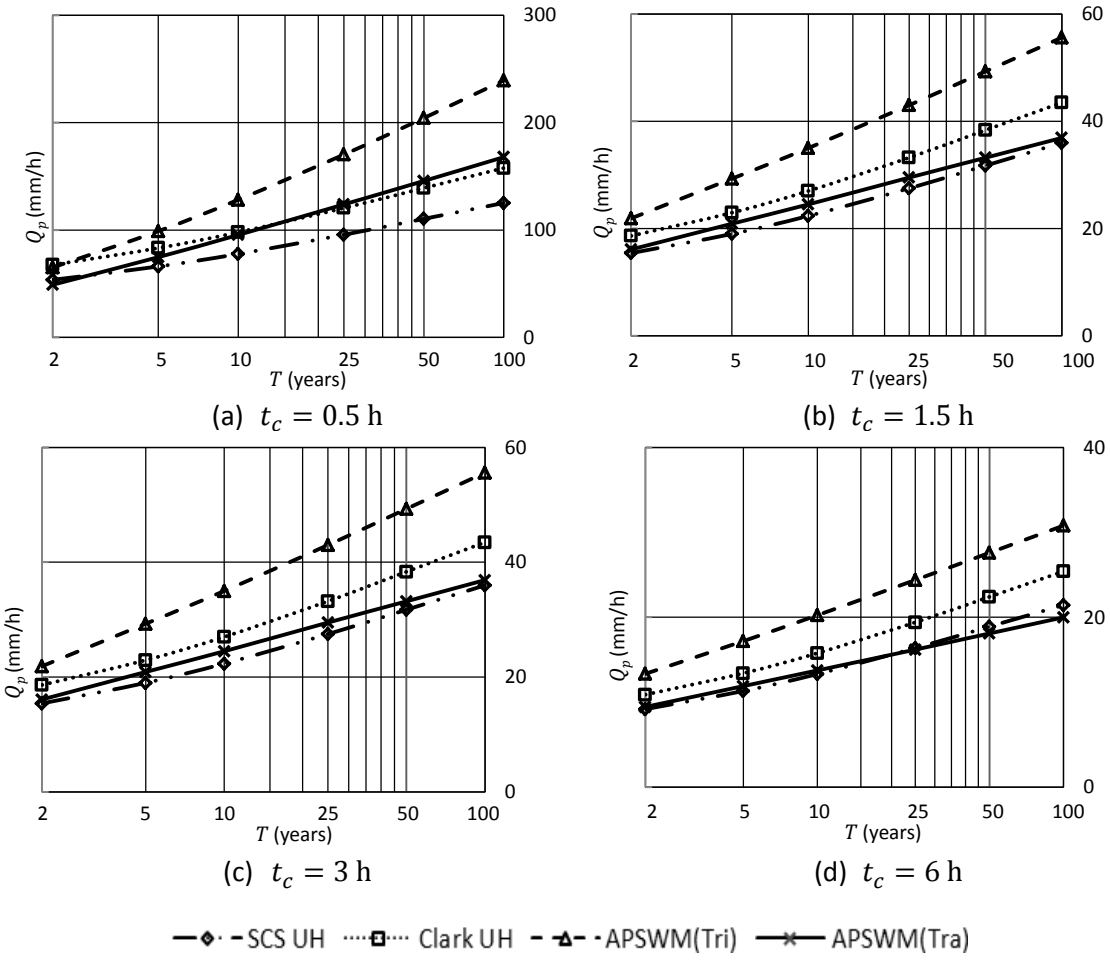

Fig. 3. Comparison of analytical and design storm modeling results for 100% impervious catchments with times of concentration of (a) 0.5 h, (b) 1.5 h, (c) 3 h, and (d) 6 h.

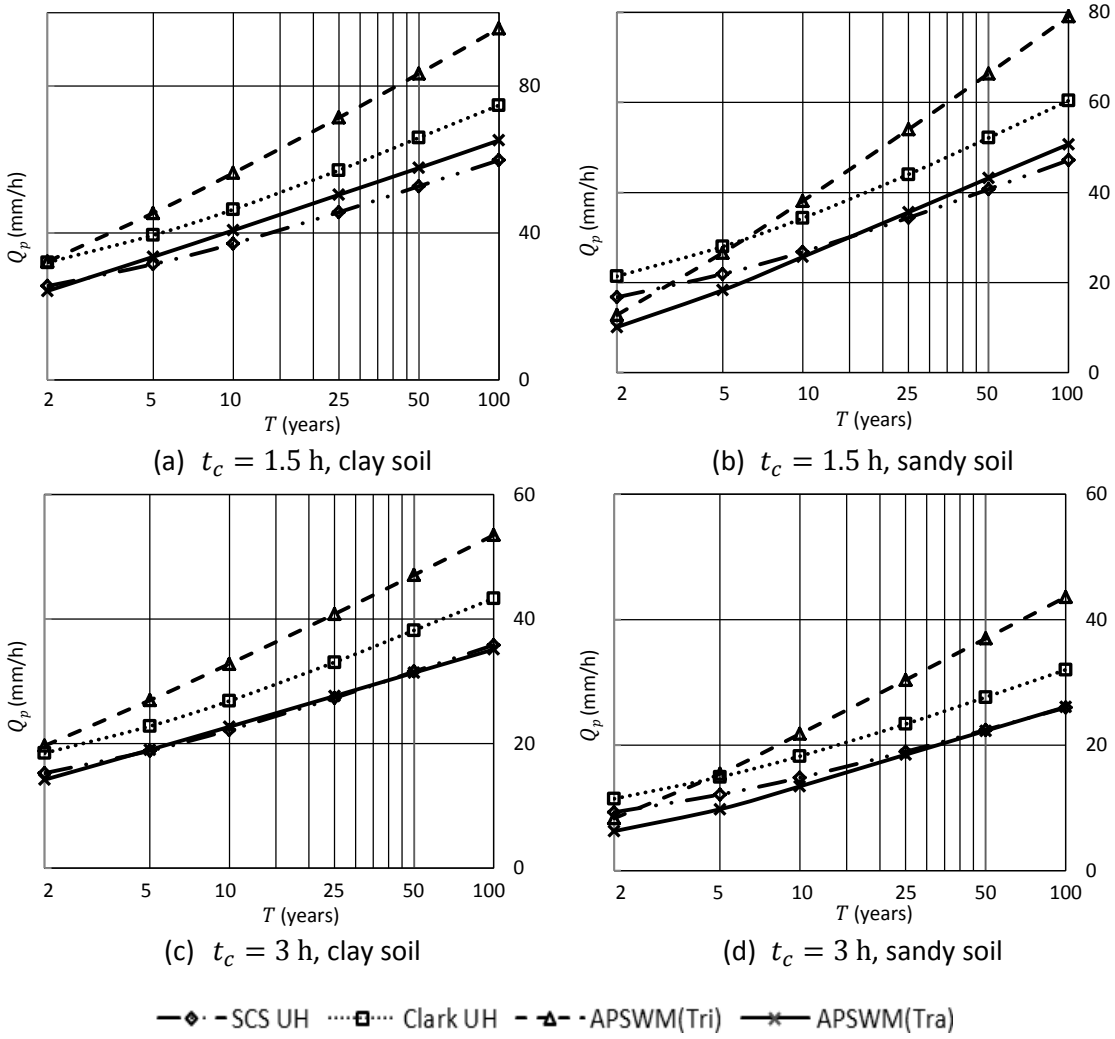

Fig. 4. Comparison of analytical and design storm modeling results for 35% impervious catchments with (a) $t_c = 1.5$ h, clay soil; (b) $t_c = 1.5$ h, sandy soil; (c) $t_c = 3$ h, clay soil; and (d) $t_c = 3$ h, sandy soil

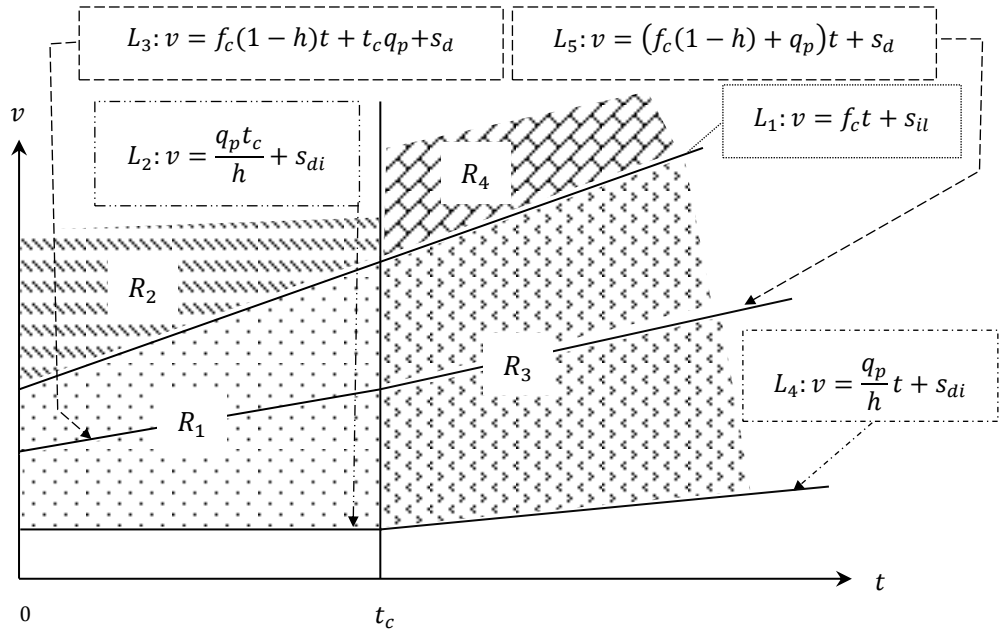

Fig. A.1. Region of integration for Type I catchments with $q_p \leq f_c h$ and Type II catchments with $q_p \leq \frac{h s_{dd}}{t_c}$

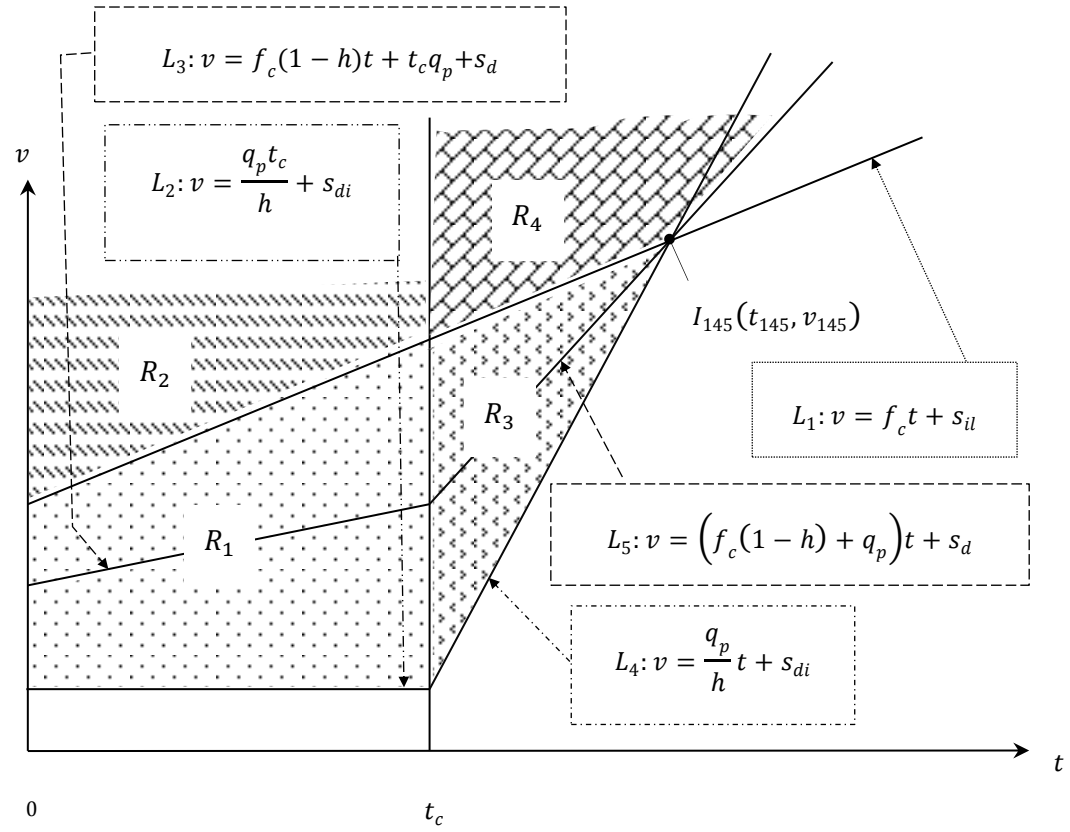

Fig. A.2. Region of integration for Type I catchments with $hf_c < q_p \leq \frac{hs_{dd}}{t_c}$

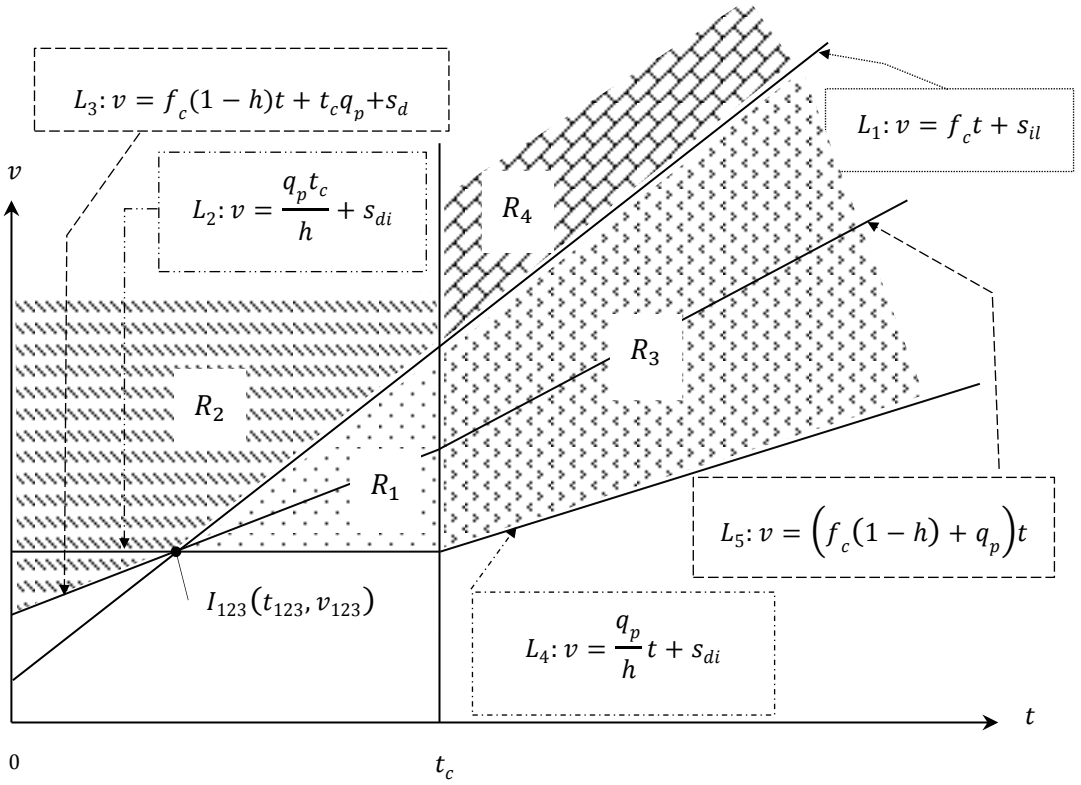

Fig. A.3. Region of integration for Type II catchments with $\frac{hs_{dd}}{t_c} < q_p \leq hf_c$

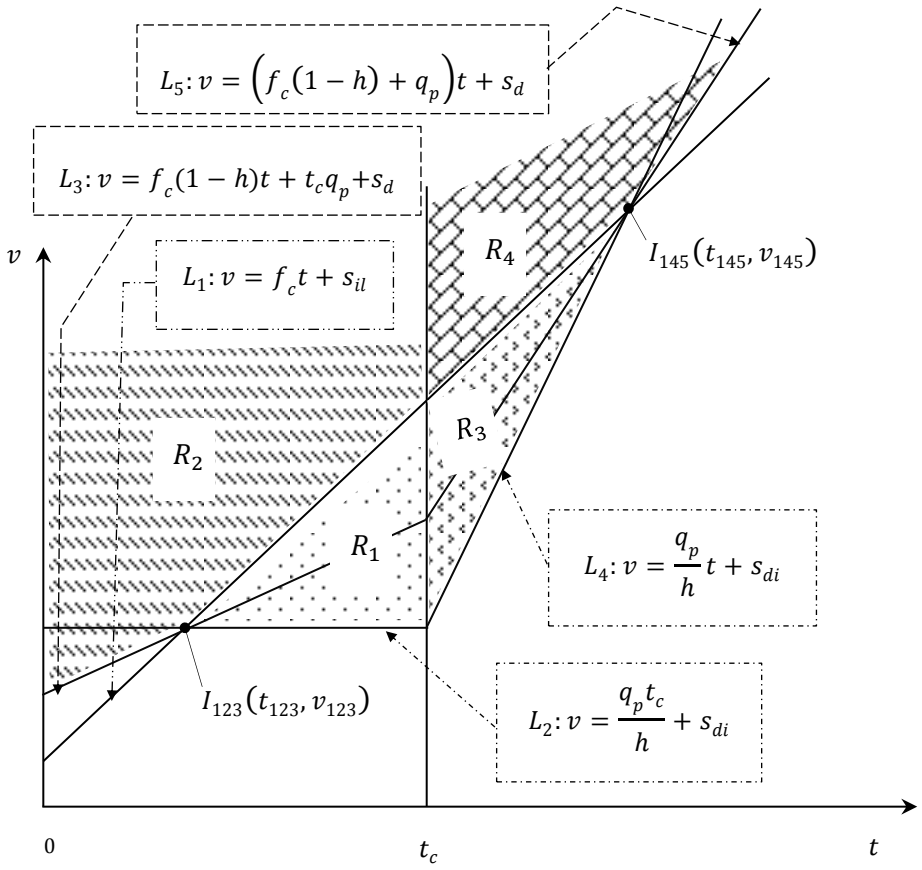

Fig. A.4. Region of integration for Type I catchments with $\frac{hS_{dd}}{t_c} < q_p \leq \frac{hS_{dd}}{t_c} + hf_c$ and Type II catchments with $hf_c < q_p \leq \frac{hS_{dd}}{t_c} + hf_c$

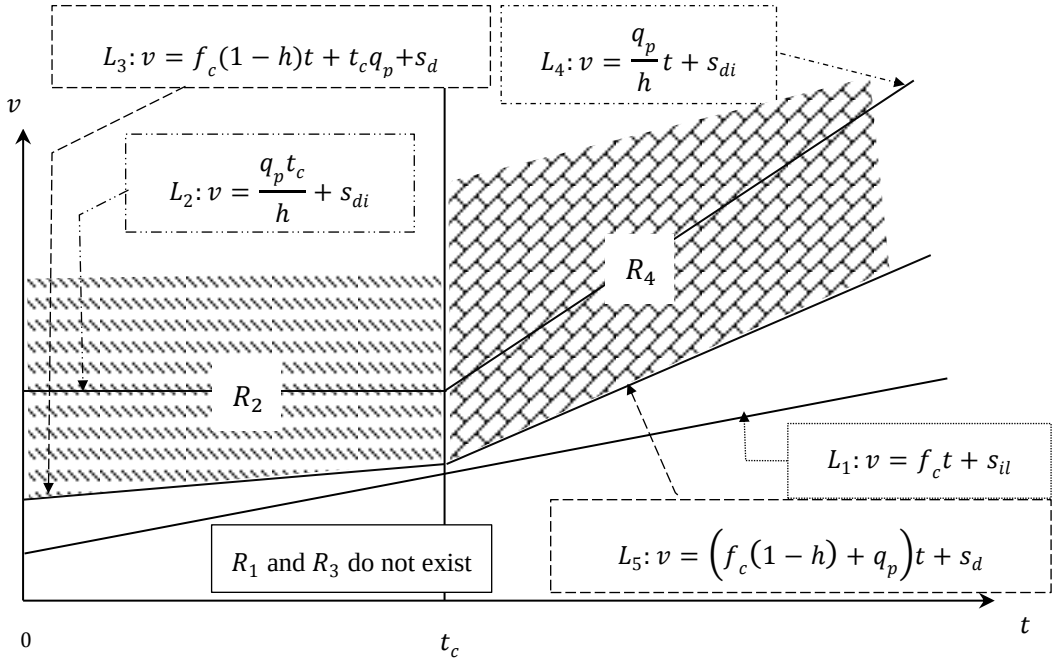

Fig. A.5. Region of integration for both types of catchments with $q_p > \frac{hS_{dd}}{t_c} + hf_c$

CHAPTER 4

Derived Urban Flood Frequency Models Accounting Saturation-Excess Runoff Generation

The content of this chapter is the manuscript is submitted to the Journal of Hydrology (October 2017).

Derived Urban Flood Frequency Models Accounting Saturation-Excess Runoff Generation

Sonia Hassini and Yiping Guo

Abstract: Derived flood frequency models developed specifically for urban stormwater management purposes are referred to as analytical probabilistic stormwater models. Although several analytical probabilistic stormwater models have been developed since the early 1970s, further improvements are still possible and needed for our changing urban catchments. Besides their reliable accuracy, the analytical probabilistic models use closed-form mathematical equations which facilitate quick and comprehensive analysis of different design alternatives. In order to expand the capability of the analytical probabilistic models that are mainly used for urban stormwater management purposes, new flood peak discharge expressions are derived considering saturation-excess in addition to infiltration-excess runoff and various possible hygrograph shapes. These expressions along with the frequency distributions of rainfall event volume and duration are used to develop a new derived flood frequency model suitable for catchments where saturation-excess runoff generation is possible. Rainfall data from several stations in the Midwest region of the United States and a real catchment in Hamilton, Ontario, Canada are used to verify the reliability of the new model.

Keywords: Peak discharge frequency distribution; Urban stormwater management; Analytical Probabilistic models; Saturation excess; Infiltration excess.

4.1. Introduction

Urban stormwater management practices are vital for the mitigation of the impact of polluted runoff on water bodies and reduction of urban stormwater strains on municipal infrastructures. Both runoff volume and peak discharge frequencies are needed for the planning, analysis, and design of stormwater management facilities. In practice, these frequencies are usually estimated using the design storm and/or continuous simulation approaches. The design storm approach is based on a set of individual rainfall events and does not model the inter-event conditions of the site of interest; however it eases the approach's applications in practice. The continuous simulation approach uses long rainfall records directly which makes it more accurate, however, its application for individual design cases is too time-consuming. Thus a more robust approach that can analytically estimate the runoff volume and peak-discharge frequencies would be useful. Although there is a promising analytical probabilistic approach (referred to as the APA in this paper) that has been under investigation for some decades, it is still under-applied in practice and requires more improvements.

The APA is based on the derived probability distribution theory which states that the probability distribution of a dependent variable (e.g. runoff event peak-discharge) can be determined analytically if the probability distributions of the

independent variables (e.g. rainfall event volume and duration) and the dependent-independent relationship (e.g. rainfall-runoff transformation) are known (Benjamin and Cornell, 1970). The derived probability distribution theory was first applied in water resources research by Eagleson (1972) to estimate the frequency of peak stream flows using the probability distribution functions (pdfs) of rainfall intensity and duration, and the kinematic wave formula, which represents the functional relationship between peak stream flows and rainfall characteristics. Then the same theory was applied in urban stormwater management analysis by many other researchers (e.g., Adams et al., 1986; Guo and Adams, 1998a, 1998b, 1999a, 1999b; Adams and Papa, 2000; Guo, 2001; Quader and Guo, 2006; Guo and Baetz, 2007; Chen and Adams, 2005 and 2007; Bacchi et al., 2008; Balistrocchi et al., 2009; Zhang and Guo, 2013a, b; Guo et al., 2012; Hassini and Guo, 2017).

The application of the derived probability distribution theory in the urban stormwater management domain results in some kind of APAs for the planning, analysis, and design of stormwater management facilities. It requires that a long-term continuous rainfall record be separated into actual individual rainfall events based on a selected minimum dry time between events known as the inter-event time definition (IETD). Each two consecutive rainfall periods are considered as separate rainfall events only if the dry time between them is greater than the IETD; for small urban catchments, the suitable IETD varies between 6 and 12 hours (Guo and Adams, 1998a; Guo and Baetz, 2007). Each rainfall event is characterized by its volume (v),

duration (t), and the dry time (referred to as the interevent time, b) that separates it from the preceding rainfall event. In order to develop the APA for urban stormwater management, a statistical analysis of the separated rainfall events is needed first to fit theoretical frequency distributions to rainfall event characteristics. Once this is done for a location of interest, the statistics of actual individual rainfall-event characteristics are the only necessary rainfall input data required by the APA for stormwater management analysis.

Guo and Adams (1998a, 1998b, 1999a, 1999b) developed closed-form analytical expressions to estimate the probability distributions of runoff event volume and peak discharge rate with and without a detention pond. In these analytical expressions, which are collectively referred to as the Analytical Probabilistic Storm Water Models (APSWM), the soil infiltration capacities and the catchment imperviousness are explicitly included. Guo and Adams (1998a, 1998b, 1999a, 1999b) found that APSWM produces similar results as compared to continuous storm water management model (SWMM) simulation results, for a set of hypothetical catchments with different physical characteristics (e.g. imperviousness and soil types) and rainfall data from the Pearson International Airport in Toronto, Canada. The first comparison between analytical, design storm and continuous simulation approaches was carried out by Guo (2001), for a test catchment in Chicago, Illinois, the United States. The three approaches were applied for the design of detention ponds and estimation of flood peaks. It was found that as long as proper selections of design storm durations

and hyetographs are made, the design storm approach can provide similar results to those of the analytical and continuous simulation approaches. In order to check the effect of the assumption of the independence between rainfall event volume and duration, Rivera et al. (2005) applied APSWM for two locations in the United States where the correlation between the rainfall event volume and duration is very low for one location and very high for the other. It was found that APSWM results are accurate when rainfall volume and duration are not highly correlated. Chen and Adams (2005) developed closed-form analytical models for evaluating stormwater runoff control performance of storage/treatment facilities. They used two rainfall-runoff volume transformations: (1) a model that is based on a runoff coefficient and (2) a simplified version of the model proposed by Guo and Adams (1998a), which explicitly uses the infiltration term and degree of urbanization. The second model was found to give better results.

The APSWM inputs for a catchment of interest, including rainfall statistics and catchment parameters, are catchment-wise constant or lumped. However in the design storm and continuous simulation approaches, the catchment can be divided into sub-catchments where inputs may be different. Quader and Guo (2006) applied APSWM for an actual design case in Kingston, Canada and compared it with the design storm approach where the catchment was divided into sub-catchments with different times of concentration. Despite APSWM's lumped treatment of catchments, comparable results were found. Chen and Adams (2007) demonstrated that the analytical probabilistic

models can be developed with different levels of complexity. They therefore developed two analytical probabilistic models, with and without infiltration term, for the assessment of urban stormwater runoff volumes. Both models were found to produce acceptable results compared to continuous simulation using the SWMM model for two test catchments in the City of Toronto, Canada. It was found that the infiltration term can be omitted if the catchment of interest is largely impervious.

Guo and Zhuge (2008) expanded the APSWM by including a simplified method for flood routing through channel reaches and detention ponds. Similar results were found compared to the results of the design storm approach. Then the Muskingum-Cunge routing method was added to the APSWM by Guo et al. (2009). The curve-number method for infiltration calculations and areal reduction for large watersheds were incorporated into the APSWM by Guo and Dai (2009). Guo and Markus (2011) integrated Clark's unit hydrograph into the APSWM. With each new expansion, the APSWM was applied under different conditions and was found to provide comparable results to either the design storm or continuous simulation approach.

Up until 2012, only infiltration excess runoff was considered in the APSWM; surface runoff from pervious areas occurs when the soil's infiltration capacity is exceeded by incoming rainfall's intensity. Guo and Adams (1998a) incorporated Horton's model in the APSWM to estimate the maximum possible infiltration loss volume which includes the initial soil wetting infiltration volume. During a Hortonian

infiltration process, infiltration starts with an initial rate then decreases exponentially over time reaching quickly to a constant rate for the rest of the event. However soils may be completely saturated before the end of a rainfall event; infiltration may stop while it is still raining. As a result, a saturation-excess runoff may occur during an event, which may have an important effect on the rainfall-runoff transformation process of some catchments. The saturation excess runoff generation process was recently incorporated into the SWMM model (Rossman, 2010). Guo et al. (2012) added into APSWM, for the first time, saturation-excess runoff volume calculations and derived an analytical model to estimate runoff volume frequencies. Guo et al. (2012) found that the saturation parameter representing the completely saturated conditions does have an effect on the runoff volume for some catchments with pervious areas. In addition, in urban stormwater management, the use of low impact development practices (LIDs) such as green roofs and bio-retention facilities has significantly increased. LIDs together with their serviced impervious areas may be considered as pervious areas, however, because of their limited size; they often provide limited maximum possible infiltration volume during a rainfall event. Thus incorporating the saturation parameter in runoff calculations is required for many urban catchments.

In essence, APSWM is an event-based approach where individual rainfall and runoff events are considered; each event has its own characteristics. If a rainfall event's volume exceeds the sum of all possible losses, the rainfall event produces a

runoff event characterized by runoff duration, volume, and peak discharge rate. The runoff event's duration is equal to the causal rainfall event's duration plus the time for a drop of rain fallen on the farthest point of the catchment (timewise) to reach the catchment's outlet; defined as the catchment's time of concentration. The runoff event volume is directly estimated as its causal rainfall event volume minus all the losses. For estimating the runoff peak discharge rate of an individual event, the hydrograph shape in addition to the runoff event volume and duration of the runoff event needs to be known. Thus further assumption about the individual hydrograph's shape is required. In the past, in order to facilitate the derivations of the peak discharge pdfs in the APSWM, individual hydrograph's shape was assumed to be triangular. Recently, Hassini and Guo (2017) derived the peak discharge pdfs assuming three possible shapes of runoff event hydrographs considering the magnitude of the rainfall event duration as compared to that of the catchment's time of concentration. Hassini and Guo (2017) found that their models produced better results than the models incorporating the triangular hydrograph assumption.

Although their derivations can be challenging, analytical probabilistic models are computationally efficient; once they are developed, they can easily be implemented in a spreadsheet or programmed as a handy software. Re-derivation of new equations is only necessary for further improvements or implementation of different calculation methods. In this paper, new pdfs of peak discharge rate are derived considering different hydrograph shapes (trapezoids and triangles) and

saturation-excess in addition to infiltration-excess runoff generations. The models developed in this study are referred to as $APSWM_{tis}$, an acronym for analytical probabilistic storm water models using trapezoidal hydrographs and including both infiltration- and saturation-excess runoff generations. Rainfall data of seven stations located in the Midwest region of the United States are used to verify $APSWM_{tis}$. The rainfall data analysis for these stations was completed by Hassini and Guo (2016); long rainfall records were separated into independent events and the fitted exponential probability distributions for rainfall event volume, duration, and interevent time were statistically tested.

With proper selection of duration and hyetograph for individual design storms, the design storm approach can produce reasonably accurate results of peak discharge rates (Guo 2001). Since peak discharge rates are the focus of this study, the design storm approach can be used for comparison. In practice, so far, SWMM is perhaps the only hydrologic model that includes the saturation excess process and is widely used in engineering practice. In addition, Catchment's parameters required by $APSWM_{tis}$ are similar to those required by SWMM. Thus design storm SWMM modelling results are used herein to verify the accuracy and reliability of $APSWM_{tis}$. For this study a subcatchment of the Davis Creek in Hamilton, Ontario, Canada, was selected as a test catchment. The SWMM model for this catchment was calibrated in an early study by Miao (2016).

4.2. Runoff Event Volume

In the development of APSWM, Guo and Adams (1998a) estimated the total volume v_r (mm) of a runoff event as the rainfall volume of the input rainfall event (v) minus the interception, depression storage and the infiltration losses that occur during the event. The interception and depression losses are combined together and referred to as the depression storage losses in APSWM, they can occur on both pervious and impervious areas. Guo and Adams (1998a) used the Horton equation to estimate infiltration losses, which are assumed to occur only on pervious areas. In order to ease the derivation of the analytical probabilistic models, Guo and Adams (1998a) simplified the Horton infiltration equation as

$$S_{mf} = S_{iw} + f_c t \quad (1)$$

where S_{mf} is the maximum possible infiltration loss for a given rainfall event with duration t (h), S_{iw} is the initial soil wetting infiltration volume (mm) and f_c is the ultimate infiltration capacity of the soil (mm/h). The initial soil wetting is assumed to include the additional infiltration losses at the beginning of a rainfall event when the infiltration rates are greater than f_c. APSWM is an event-based model that does calculations on an event-by-event basis and not on a time-step-by-time-step basis as in continuous simulation models. For each rainfall event, the runoff volume is the rainfall event volume minus all the rainfall losses. Generally, an urban catchment consists of pervious and impervious parts, each part is assumed to generate runoff only after its rainfall losses are satisfied. Then the runoff volume is estimated as the

area-weighted runoff coming from the impervious and pervious areas. The impervious area is assumed to contribute first since its rainfall losses are usually less than those lost on pervious areas. As a result, for each rainfall event with volume v and duration t, the resulting total runoff volume is

$$v_r = \begin{cases} 0, & v \le S_{di} \\ h(v - S_{di}), & S_{di} < v \le S_i \\ v - S_d - f_c(1-h)t, & v > S_{il} + f_c t \end{cases} \quad (2)$$

where v_r is the runoff event volume (mm); h is the impervious surface fraction of the catchment; S_{il} is the pervious area initial losses (mm); i.e., $S_{il} = S_{dp} + S_{iw}$ where S_{dp} is the pervious area depression storage (mm); S_{di} is the impervious area depression storage (mm), and S_d is the area-weighted depression storage of the impervious and pervious areas (mm), i.e. $S_d = hS_{di} + (1-h)S_{il}$. Detailed derivations of rainfall infiltration losses and runoff volume equations are provided in Guo and Adams (1998a).

In Guo and Adams (1998a), the soil is assumed to remain unsaturated and absorbing water at a rate equal to f_c during the entire rainfall duration. However, this assumption does not always hold true; the soil may become saturated before the end of the rainfall event and infiltration may cease and remaining rainfall will all contribute to the surface runoff. In order to account for this saturation-excess runoff, Guo et al. (2012) re-derived the analytical equations for runoff volume. For the pervious areas of a catchment, Guo et al. (2012) added another pervious area parameter - the maximum infiltration volume possible per rainfall event S_s - to account for saturation-excess

runoff. For each rainfall event, at the beginning, the soil is assumed to be relatively dry and at the end the soil may remain unsaturated (Scenario 1) or become saturated (Scenario 2). For Scenario 1, the soil remains unsaturated because the rainfall event volume is less than the pervious area depression storage plus the soil's maximum possible infiltration volume (i.e., $v < S_{dp} + S_s$) or the soil's maximum possible infiltration volume is greater than the total amount of infiltration that can occur during the rainfall which is equal to the initial soil wetting infiltration amount plus the amount of infiltration that can occur during the event at a constant rate f_c (i.e.; $S_s > S_{iw} + f_c t$). For Scenario 2, saturation will occur only if the opposites of the two conditions in Scenario 1 are both satisfied. In other words the soil will be saturated if $v \geq S_{dp} + S_s$ and $S_s \leq S_{iw} + f_c t$. In both scenarios some runoff may occur.

Let S_m be the maximum possible of amount of water that can infiltrate into the soil after the initial soil wetting infiltration amount is satisfied already, i.e., $S_m = S_s - S_{iw}$. For each rainfall event, after subtracting the rainfall losses and combining the resulting runoff volume from the pervious and impervious catchment areas, Guo et al. (2012) determined the total runoff volume (v_r) as follows. For Scenario 1 where $t < \frac{S_m}{f_c}$ or $v < S_{il} + S_m$, since the soil is not saturated no saturation-excess runoff will occur and only infiltration excess runoff may occur, the total runoff volume is the same as given by Eq. (2). For Scenario 2 where $t \geq \frac{S_m}{f_c}$ and $v \geq S_{il} + S_m$, the soil is saturated before the end of the rainfall event and

$$v_r = v - S_d - (1 - h)S_m \tag{3}$$

In this paper, in order to make the derivations more straightforward, the two scenarios are divided further into four mutually exclusive domains of v and t. Let t_s be the time required for the soil to reach saturation after the initial soil wetting infiltration is completed, i.e., $t_s = \frac{S_m}{f_c}$. Hereinafter, t_s is referred to as the saturation time, which may be less or greater than the rainfall event duration t. Logically, there are three conditions on the variables of v and t that can lead to Scenario 1: (a) $t < t_s$ and $v < S_{il} + S_m$, (b) $t < t_s$ and $v \geq S_{il} + S_m$, or (c) $t \geq t_s$ and $v < S_{il} + S_m$. However, Scenario 2 is only led by one condition, i.e., (d) $t \geq t_s$ and $v \geq S_{il} + S_m$. Then rainfall events can be classified into two categories based on their durations: (A) rainfall events with durations less than the saturation time ($t < t_s$), no matter how much the rainfall event volume is, the soil will remain unsaturated (conditions a and b); (B) $t \geq t_s$, which includes condition (c) where the soil remains unsaturated due to insufficient rainfall volume and condition (d) where soil saturation is attained. For each rainfall event, the corresponding runoff volume is as given in Eq. (2) if $t < t_s$. Otherwise, if $t \geq t_s$, thus $f_c t + S_{il} \geq S_m + S_{il}$, the runoff event volume can be calculated as follows:

$$v_r = \begin{cases} 0, & v \leq S_{di} \\ h(v - S_{di}), & S_{di} < v \leq S_{il} + S_m \\ v - S_d - (1-h)S_m, & v > S_{il} + S_m \end{cases} \quad (4)$$

Combining Eqs. (2) and (4) and listing separately the additional requirement for each applicable condition, the runoff event volume that results from a rainfall event can be expressed in a combined general way as:

$$v_r = \begin{cases} 0, & v \le S_{di} \\ h(v - S_{di}), & S_{di} < v < S_{il} + f_c t \text{ and } t < t_s \\ v - S_d - f_c(1-h)t, & v > S_{il} + f_c t \text{ and } t < t_s \\ h(v - S_{di}), & S_{di} < v < S_{il} + S_m \text{ and } t \ge t_s \\ v - S_d - (1-h)S_m, & v \ge S_{il} + S_m \text{ and } t \ge t_s \end{cases} \quad (5)$$

4.3. Peak Discharge Rate of a Runoff Event

Knowing the total volume [Eq. (5)] and the detailed shape of the hydrograph of a runoff event, the corresponding peak discharge rate can be estimated. An approximate hydrograph shape may be assumed in order to analytically determine the peak-discharge rate. In the past, a triangular shape was adopted by Guo and Adams (1998b) and recently a more general trapezoidal shape was incorporated by Hassini and Guo (2017). For both shapes the base which represents the runoff event duration is equal to the sum of the input rainfall event duration (t) and the catchment time of concentration (t_c). The catchment's time of concentration is assumed to be constant and solely depends on the catchment's characteristics. Both assumptions, i.e., smooth triangular or trapezoidal hydrograph shapes and constant catchment's time of concentration, are necessary for the derivation of the analytical expressions. These two assumptions have been used in other different studies and were found to be acceptable in practice.

The trapezoidal shape is more general since it actually does consider both trapezoidal

and triangular hydrograph shapes according to the conditions of individual events. Hassini and Guo (2017) found that the trapezoidal hydrograph shape produces better results than the triangular shape. Thus, the trapezoidal shape based on what is described in Ponce (1989) is used in this study. According to Ponce (1989), the shape of a runoff hydrograph depends on the relative magnitudes of the catchment's time of concentration and the storm duration. For cases with approximately uniform rainfall excesses, the approximate shape of the event hydrographs may be divided into three cases. First, if $t \leq t_c$, runoff rate increases approximately linearly and reaches a maximum (peak discharge) right at the end of the rainfall event, this maximum lasts $(t_c - t)$ long until the whole catchment has contributed runoff to the outlet then it will decrease linearly to zero over a period equaling t. Second, if $t = t_c$, the hydrograph is approximately triangular where the runoff rate increases linearly over t_c to reach an instant maximum then deceases linearly to zero over a period of time equaling t_c as well. Third, if $t > t_c$, the runoff increases linearly until it reaches a maximum at t_c when the whole catchment is contributing, this maximum lasts until the end of the rainfall duration, then the flow rate decreases linearly to zero over a period of time equaling t_c. All in all, the hydrograph shape is assumed to be isosceles trapezoidal (a triangle may be viewed as a special trapezoid) with lower and upper bases equaling $(t + t_c)$ and $|t - t_c|$, respectively [more details can be found in Ponce (1989) and Hassini and Guo (2017)].

The area of the above-described trapezoids represents the runoff event volume from

an urban catchment (i.e. v_r). Therefore the resulting peak discharge rate (Q_p) can be geometrically determined as

$$Q_p = \begin{cases} \frac{v_r}{t_c}, & t \le t_c \\ \frac{v_r}{t}, & t > t_c \end{cases} \tag{6}$$

The analytical expression of Q_p as a function of the rainfall event volume v and duration t can be obtained by substituting the expression for runoff event volume (v_r) [Eq. (5)] into Eq. (6). Since v_r depends on t_s and Q_p depends on v_r and t_c, the relative magnitude of t_s and t_c, which are both catchment characteristics, needs to be known in order to express Q_p analytically.

If a catchment's time of concentration is less than or equal to its time of saturation (i.e., $t_c \le t_s$), this type of catchment is referred to as Type I. The analytical expression of Q_p for Type I catchments is obtained by substituting Eq. (5) into Eq. (6) whereas the possible t value is divided further into the following three ranges: $t \le t_c$, $t_c < t \le t_s$, and $t > t_s$. The results are as follows:

$$Q_p = \begin{cases} 0, & v \le S_{di} \\ \frac{h(v - S_{di})}{t_c}, & S_{di} < v \le S_{il} + f_c t \text{ and } t \le t_c \\ \frac{v - S_d - f_c(1-h)t}{t_c}, & v > S_{il} + f_c t \text{ and } t \le t_c \\ \frac{h(v - S_{di})}{t}, & S_{di} < v \le S_{il} + f_c t \text{ and } t_c < t \le \\ \frac{v - S_d - f_c(1-h)t}{t}, & v > S_{il} + f_c t \text{ and } t_c < t \le t_s \\ \frac{h(v - S_{di})}{t}, & S_{di} < v \le S_{il} + S_m \text{ and } t > t_s \\ \frac{v - S_d - (1-h)S_m}{t}, & v > S_{il} + S_m \text{ and } t > t_s \end{cases} \tag{7}$$

If the catchment time of concentration is greater than its time of saturation (i.e., $t_c > t_s$), this type of catchment is referred to as Type II. The analytical expression of Q_p for Type II catchments is also obtained by substituting Eq. (5) into Eq. (6) whereas the possible t value is divided into the following three ranges: $t \leq t_s$, $t_s < t \leq t_c$, and $t > t_c$. Note that these three ranges are different from the three ranges for Type I catchments. The substitution results are:

$$Q_p = \begin{cases} 0, & v \leq S_{di} \\ \dfrac{h(v - S_{di})}{t_c}, & S_{di} < v \leq S_{il} + f_c t \text{ and } t \leq t_s \\ \dfrac{v - S_d - f_c(1-h)t}{t_c}, & v > S_{il} + f_c t \text{ and } t \leq t_s \\ \dfrac{h(v - S_{di})}{t_c}, & S_{di} < v \leq S_{il} + S_m \text{ and } t_s < t \leq t_c \\ \dfrac{v - S_d - f_c(1-h)t}{t_c}, & v > S_{il} + S_m \text{ and } t_s < t \leq t_c \\ \dfrac{h(v - S_{di})}{t}, & S_{di} < v \leq S_{il} + S_m \text{ and } t > t_c \\ \dfrac{v - S_d - (1-h)S_m}{t}, & v > S_{il} + S_m \text{ and } t > t_c \end{cases} \quad (8)$$

4.4. Derivation of the Peak Discharge Rate Exceedence Probabilities

As mentioned earlier, the derived probability distribution is the basis for the development of the analytical expressions of the peak discharge exceedence probabilities. In this case the mathematical relationships are the peak discharge

expressions shown in Eqs. (7) and (8), where Q_p is expressed as functions of the two random variables: rainfall event volume (v) and duration (t). Many researchers (e.g. Eagleson 1972, 1978; Howard 1976; Adams et al., 1986; Guo and Adams, 1998a; Adams and Papa, 2000; Guo, 2001; Guo and Baetz, 2007; Zhang and Guo, 2013a, b; and Hassini and Guo, 2016) found that these two random variables follow exponential distributions at many different geographical locations especially in North America. Hassini and Guo (2016) proposed a procedure for testing the exponentiality of rainfall event volume, duration and interevent time. In this paper v and t are assumed to have the following exponential distributions:

$$f(v) = \zeta e^{-\zeta v}, \qquad v \geq 0 \tag{9}$$

$$f(t) = \lambda e^{-\lambda t}, \qquad t \geq 0 \tag{10}$$

where ζ and λ are the exponential distribution parameters for rainfall event volume and duration, respectively.

Since the peak discharge rate depends on the two random variables v and t, the exceedence probability of Q_p per rainfall event can be determined using the joint probability density function of v and t. To relatively ease the derivations, random variables v and t are assumed to be statistically independent as it was done by many researchers (e.g. Eagleson, 1972; Guo and Adams, 1998a; Adams and Papa, 2000; Guo, 2001; Guo and Baetz, 2007; Guo et al., 2012; Zhang and Guo, 2013a, b; and Hassini and Guo, 2017). Thus the joint probability density function of v and t is the product of their marginal probability density functions [Eqs. (9) and (10)]. According

to Eqs. (7) and (8), a zero peak discharge value (i.e. $Q_p = 0$) occurs when $v \leq S_{di}$ for any values of t, independent of the relative magnitude of t_s and t_c. Making use of the derived probability distribution theory and the independence assumption between v and t, the probability per rainfall event with a null peak discharge rate is

$$P[Q_p = 0] = \int_0^{S_{di}} \int_0^{\infty} \zeta e^{-\zeta v} \lambda e^{-\lambda t} dt\, dv = 1 - e^{-\zeta S_{di}} \tag{11}$$

In the above integration, the region of integration as required in using the derived probability distribution theory is the region on the v-t plane covered by $t > 0$ and $0 < v \leq S_{di}$.

For a given non-zero peak discharge rate q_p (i.e. $q_p > 0$) and based on Eqs. (7) and (8), the exceedence probability $P[Q_p > q_p]$ depends on the relative magnitude of the catchment's time of concentration (t_c) and the catchment soil's time of saturation (t_s). For either cases, the peak discharge rate expressions [i.e., Eqs. (7) and (8)] are piece-wise functions defined on seven subdomains of v and t. As a result, the exceedence probability expression is determined by integrating the joint probability of v and t over the union of the sub-regions delineated based on each of the subdomains described in Eqs. (7) and (8). For each type of catchments there are six mutually exclusive sub-regions of integration as given by each applicable v and t subdomains (those subdomains other than $v \leq S_{di}$). This is because $v \leq S_{di}$ can only result in a zero peak discharge; while each of the other remaining six set of v and t conditions may result in a Q_p being greater than the given non-zero q_p. Detailed delineation of the sub-regions of integration and derivation of the exceedence probabilities are

illustrated in Appendix A.

In order to simplify the derivation results for zero Q_p [Eq. (11)] and non-zero Q_p values [Eq. (A.1) through Eq. (A.9)], the following short-hand notation is introduced:

$$g_1(q_p) = \exp\left(-\frac{\zeta q_p t_c}{h} - \zeta S_{di}\right);$$

$$g_2(q_p) = \frac{\zeta q_p}{\lambda h+\zeta q_p}\exp\left(-\lambda t_c - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right);$$

$$g_3(q_p) = \frac{(1-h)\lambda\zeta(q_p-hf_c)}{(\lambda h+\zeta q_p)(\lambda+\zeta q_p+(1-h)\zeta f_c)}\exp\left(\frac{-\zeta S_{il}q_p+h\zeta f_c S_{di}-h\lambda S_{dd}}{q_p-hf_c}\right);$$

$$g_4(q_p) = \frac{\lambda}{\lambda+(1-h)\zeta f_c}\exp\left(-\zeta q_p t_c - \zeta S_d\right);$$

$$g_5(q_p) = \frac{(1-h)\zeta f_c}{\lambda+(1-h)\zeta f_c}\exp\left(\frac{-f_c\zeta q_p t_c-h\zeta f_c S_{di}-\lambda q_p t_c+h\lambda S_{dd}}{hf_c}\right);$$

$$g_6(q_p) = \frac{\lambda\zeta q_p}{(\lambda+\zeta q_p+(1-h)\zeta f_c)(\lambda+(1-h)\zeta f_c)}\exp\left(-\lambda t_c - \zeta q_p t_c - (1-h)\zeta f_c t_c - \zeta S_d\right);$$

$$g_7(q_p) = \frac{\lambda\zeta q_p}{(\lambda h+\zeta q_p)(\lambda+\zeta q_p)}\exp\left(-\lambda\frac{h(S_{dd}+S_m)}{q_p} - \zeta(S_{il}+S_m)\right);$$

$$g_8(q_p) = \frac{\lambda\zeta(1-h)f_c}{(\lambda+\zeta q_p)(\lambda+\zeta q_p+(1-h)\zeta f_c)}\exp\left(-\lambda t_s - \zeta q_p t_s - (1-h)\zeta S_m - \zeta S_d\right);$$

$$g_9(q_p) = \frac{\zeta q_p}{(\lambda+\zeta q_p)}\exp\left(-\lambda t_c - \zeta q_p t_c - (1-h)\zeta S_m - \zeta S_d\right);$$

$$g_{10}(q_p) = \frac{(1-h)\zeta f_c}{\lambda+(1-h)\zeta f_c}\exp\left(-\lambda t_s - \zeta q_p t_c - (1-h)\zeta S_m - \zeta S_d\right);$$

$$g_{11}(q_p) = \frac{\zeta q_p}{(\lambda h+\zeta q_p)}\exp\left(-\lambda\frac{h(S_{dd}+S_m)}{q_p} - \zeta(S_{il}+S_m)\right);$$

$$g_{12}(q_p) = \frac{\lambda h}{\lambda h+\zeta q_p}\exp\left(-\lambda t_s - \frac{\zeta q_p t_s}{h} - \zeta S_{di}\right);$$

$$g_{13}(q_p) = \exp\left(-\lambda t_s - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right);$$

$$g_{14}(q_p) = \frac{\lambda\zeta(1-h)q_p}{(\lambda h+\zeta q_p)(\lambda+\zeta q_p)}\exp\left(-\lambda\frac{h(S_{dd}+S_m)}{q_p} - \zeta(S_{il}+S_m)\right);$$

$$E = \exp(-\lambda t_s - \zeta(S_{il}+S_m)).$$

Incorporating the above notation into the final derivation results, the peak discharge rate exceedence probability per rainfall event is summarized in Table 1.

4.5. Annual Exceedence Probability and Flood Frequency Distribution

In flood control design, annual exceedence probability or return period is often used instead of exceedence probability per rainfall event. By definition, return period (T_R, in years) is the inverse of the exceedence probability per year. The exceedence probabilities $P[Q_p > q_p]$ described in Table 1 [along with the expressions of $g_1(q_p)$ through $g_{14}(q_p)$ and E] are exceedence probabilities per rainfall event, which must be multiplied by the average number of events per year (denoted as θ) in order to obtain exceedence probabilities per year. The return period of a given peak discharge rate q_p can therefore be calculated as

$$T_R = \frac{1}{\theta P[Q_p > q_p]} \tag{12}$$

The expressions shown in Table 1 and Eq. (12) constitute a new derived flood frequency distribution model considering both infiltration- and saturation-excess runoff.

4.6. Comparison with SWMM Design Storm Results

4.6.1. Rainfall input data

Rainfall data prepared for APSWM in an earlier study (Hassini and Guo 2016) are used in this paper to check the performance of the new models. Hassini and Guo (2016) performed rainfall data analysis for seven stations in the Midwest region of the

United States. For each station, continuous long rainfall records (53 years) were separated into independent individual events. The pdfs of rainfall characteristics were found to fit well exponential distributions. Refer to Hassini and Guo (2016) for more details about the selected stations, their rainfall data analysis, and fitting of rainfall event characteristics' distributions. The rainfall input data required by APSWM (i.e. average rainfall event volume $\bar{v}$, duration $\bar{t}$, and annual average number θ) of the selected stations are summarized in Table 2. Using the design storm approach, the required rainfall input data are design storm duration, volume, and hyetograph. For better results, the design storm duration of 24 hours is selected and partial duration series (PDS) volumes for different return periods are used (Table 3). The PDS volumes of the selected stations are retrieved from the U.S. National Weather Service's website. The same period of rainfall data as used in APSWM, from 1949 to 2001, is used in PDS analysis.

4.6.2. Catchment's characteristics

In this study, the catchment of interest is an actual catchment located in the east part of Hamilton in Ontario, Canada. The catchment's characteristics that need to be determined depend on the selected rainfall-runoff transformation model. SWMM version 5.1, developed by the United States Environmental Protection Agency, is the single event model that is used in this study. The design storm's hyetograph is determined using the methodology developed by the United States Department of Agriculture - Natural Resources Conservation Service (USDA-NRCS). All the stations

used in this study are in the area where Type 2 storms apply. A time step of 30 minutes is used for hyetograph generation. Horton infiltration equation is used to calculate the infiltration losses. The catchment's characteristics required by SWMM are calibrated by Miao (2016) and are summarized in Table 4. Most of the catchment's characteristics required by APSWM are also available in Table 4. The initial soil wetting infiltration volume (S_{iw}) and the catchment's time of concentration (t_c) are the only parameters missing; however, they can be estimated using data from Table 4.

The initial soil wetting infiltration volume S_{iw} can be estimated using the formula proposed by Guo and Adams (1998a):

$$S_{iw} = \frac{f_0 - f_c}{k}[1 - \exp(-kt)]$$

The variability in the rainfall event duration t has negligible effect on the S_{iw} calculations, thus t can be replaced by the average rainfall event duration in the above equation.

Based on the available catchment's characteristics (Table 4), the kinematic wave equation is a good candidate for estimating the catchment's time of concentration; however this equation is recommended for catchments with flow sheets less than 91 m in length, which is not the case for the catchment under study (the flow sheet is 1487 m in length as shown in Table 4). There are other equations that can be used to estimate t_c such as the Kirpich, the US Federal Aviation Administration (FAA) and Kerby equations. The FAA equation is selected for this study because it is the

most widely used equation due to its connection with the rational formula, it is also recommended by the American Society of Civil Engineers (ASCE), and it is suitable for urban catchments (Ken 2015). The FAA equation is as follows:

$$t_c = \frac{3.26\,(1.1 - C)L^{0.5}}{(100S)^{\frac{1}{3}}}$$

where t_c is in minutes, L is the overland flow length (m) and S is the slope of the surface (m/m). The rational coefficient C in the above equation depends on the climate, soil characteristics and land cover where land cover is the most important factor. For urban catchments, C depends largely on the impervious fraction and can be estimated as follows

$$C = C_i h + C_p(1 - h)$$

where C_i and C_p are the rational coefficients for impervious and pervious areas of the catchment, respectively; h is the impervious fraction of the catchment. Assuming that C_i and C_p are 0.85 and 0.35, respectively, t_c is estimated to be 69.34 min (1.16 h).

4.6.3. Results and Discussion

The catchment's time of concentration (t_c) is the only parameter that is not directly used by the SWMM design storm modeling and may have an important effect on flood frequency estimations. Thus t_c is the only parameter that needs to be verified in this study in order to make sure that the $APSWM_{tis}$ and SWMM input data do represent the same catchment. Rainfall data from St. Cloud Airport, MN are used to calibrate while rainfall data from Des Moines Airport, IA are used to validate

$APSWM_{tis}$ against SWMM results. During the calibration and validation processes, t_c used in $APSWM_{tis}$ was modified to match the $APSWM_{tis}$ results closely to SWMM results. However, some differences between the results from the two models are expected due to their different ways of flood frequency estimations. The best value found of t_c is 1.25 hours. Estimated and calibrated t_c values (1.16 h and 1.25 h, respectively) are fairly close, which confirm that the formula used to estimate t_c is a suitable one for the catchment of this study.

Fig. 1 represents the peak discharge rates resulting from $APSWM_{tis}$ and SWMM using the rainfall data of MN and IA stations for different frequencies ($t_c = 1.25$ h). As shown in Fig. 1, $APSWM_{tis}$ and SWMM results are fairly close for both stations. Therefore 1.25 h is used as the t_c value when comparing $APSWM_{tis}$ and SWMM results for the rest of the stations. Table 5 summarizes the differences in results between the two models (i.e. $APSWM_{tis}$ and SWMM) for all the stations used in this Study. Other than the stations used for calibration and validation (MN and IA), minimal differences are shown for ND and SD stations followed by the stations of KS, MO, and then NE (Table 5). The best and worst results which correspond to the SD and NE stations, respectively, are plotted in Fig. 2. Overall, the differences in results are almost all less than 20%, this level of difference may also occur just between design storm modeling results if different design storms are used (Guo and Zhuge 2008; Hassini and Guo 2017). For the same location, design storms used may be different in duration or hyetograph shape. If this level of differences is acceptable

when using the design storm approach, it should also be acceptable for $APSWM_{tis}$. Overall $APSWM_{tis}$ results are slightly higher than those from SWMM design storm modeling, which may be in favor of $APSWM_{tis}$ when safety is a priority. Continuous simulation results or observed runoff data are required to judge which modeling approach is more accurate; however, this is not the focus of this study as it was already investigated in other studies (e.g. Guo and Adam (1998a,b); and Guo 2001).

4.7. Summary and Conclusions

This study brings the development of analytical probabilistic stormwater models for flood frequency estimation for small urban catchments to a new stage where saturation-excess runoff, infiltration-excess runoff, and various possible hydrograph shapes are all taken into consideration. Although quite lengthy, the derived probability distributions of peak discharge rate are still in closed analytical form, and are computationally efficient and easy to use if they are implemented into a spreadsheet or a small computer program.

The derived analytical expressions are collectively referred to as the $APSWM_{tis}$ model. $APSWM_{tis}$ results are compared to those from the use of the design storm approach using SWMM to build the catchment hydrologic model. Seven stations from the Midwest region of the United States where used together with a real catchment in Hamilton, Ontario, Canada. The catchment parameters that are required by SWMM were already calibrated. Matching input data are used for both $APSWM_{tis}$ and

SWMM. Comparable results were found with reasonable differences which are mainly due to the dissimilarities in modeling approaches. These reasonable results also imply that the incorporated assumptions are acceptable for this case study and for many locations.

The main purpose of this paper is the derivation and verification of $APSWM_{tis}$, verification of the derived model under different catchment conditions is not performed in this study because it was completed in other studies such as Guo and Adams (1998b) and Hassini and Guo (2017). The saturation parameter S_s (i.e. maximum possible infiltration volume) is the only additional catchment characteristic, however its sensitivity analysis was not conducted in this study due to our focus on the analytical derivation on one hand and the use of an actual catchment on the other hand. Guo et al. (2012) examined the analytical probabilistic model's sensitivity of runoff event volume on the changing magnitude of S_s, similar studies can be performed in the future examining the analytical probabilistic model's sensitivity of peak discharge rate on the changing magnitude of S_s, using the newly derived $APSWM_{tis}$ model.

Due to its incorporation of the parameter of maximum saturation, $APSWM_{tis}$ is expected to produce more accurate results especially when pervious areas are more predominant within the catchment of interest. The use of low impact development practices (LIDs) such as green roofs and infiltration trenches as urban stormwater management measures increase equivalently the pervious areas of urban catchments.

LIDs together with the impervious areas that they serve can be modeled as pervious areas which allow a limited amount of rainfall to be infiltrated. With the incorporation of the additional parameter S_s, APSWM$_{tis}$ can be used to effectively model the functionality of LIDs.

Appendix A: Detailed Derivations

4.A.1. Integration Sub-regions of Type I Catchments (i.e. catchments with $t_c \leq t_s$)

According to Eq. (7), for a given non-zero q_p value, the v and t values that would result in a Q_p being greater than q_p may come from either of the six subdomains of v and t values, i.e. those subdomains of the Q_p function as defined in Eq. (7) other than the first one with $v < S_{di}$. Each of the six subdomains would therefore result in a sub-region of integration for determining $P[Q_p > q_p]$. Those sub-regions of integrations are determined as follows.

The first sub-region of integration, referred to as R_{I1}, is determined as the sub-region of v and t given that $S_{di} < v \leq S_{il} + f_c t$ and $t < t_c$, whereas the resulting Q_p must be greater than q_p. Based on Eq. (7), R_{I1} is defined by

$$\begin{cases} \dfrac{h(v - S_{di})}{t_c} > q_p \\ S_{di} < v \leq S_{il} + f_c t \\ t \leq t_c \end{cases} \xrightarrow{yields} \begin{cases} \dfrac{q_p t_c}{h} + S_{di} < v \leq S_{il} + f_c t \\ t \leq t_c \end{cases}$$

Let L_1, L_2, and L_c be the lines defined by $v = f_c t + S_{il}$, $v = \frac{q_p t_c}{h} + S_{di}$, and $t = t_c$, respectively (all lines in this paper are plotted with v on the vertical axis and t on the horizontal axis). Then R_{I1} is the area above L_2 but below L_1, and on the left side of the vertical line L_c.

In a way similar to R_{I1}, the second sub-region of integration, referred to as R_{I2}, is determined as the sub-region of v and t given that $v > S_{il} + f_c t$ and $t \leq t_c$, whereas the resulting Q_p must be greater than q_p. Again based on Eq. (7), R_{I2} is defined by

$$\begin{cases} \dfrac{v - S_d - f_c(1-h)t}{t_c} > q_p \\ v > S_{il} + f_c t \\ t \leq t_c \end{cases} \xrightarrow{yields} \begin{cases} v > q_p t_c + S_d + (1-h) f_c t \\ v > S_{il} + f_c t \\ t \leq t_c \end{cases}$$

Let L_3 be the line defined by $v = f_c(1-h)t + t_c q_p + S_d$, R_{I2} is the area above L_1 and L_3, and also on the left side of L_c.

Similar to R_{I1} and R_{I2}, the third sub-region of integration R_{I3} is defined by

$$\begin{cases} \dfrac{h(v - S_{di})}{t} > q_p \\ S_{di} < v \leq S_{il} + f_c t \\ t_c < t \leq t_s \end{cases} \xrightarrow{yields} \begin{cases} \dfrac{q_p t}{h} + S_{di} < v \leq S_{il} + f_c t \\ t_c < t \leq t_s \end{cases}$$

Let L_4 be the line defined by $v = \frac{q_p}{h} t + S_{di}$, R_{I3} is the area above L_4 but below L_1,

and also on the right side of L_c and the left side of L_s.

The fourth sub-region of integration R_{I4} is defined by

$$\begin{cases} \dfrac{v - S_d - f_c(1-h)t}{t} > q_p \\ v > S_{il} + f_c t \\ t_c < t \leq t_s \end{cases} \xrightarrow{yields} \begin{cases} v > \left[q_p + (1-h)f_c\right]t + S_d \\ v > S_{il} + f_c t \\ t_c < t \leq t_s \end{cases}$$

Let L_5 be the line defined by $v = \left[f_c(1-h) + q_p\right]t + S_d$, R_{I4} is the area above both L_1 and L_5, and also on the right side of L_c and the left side of L_s.

The fifth sub-region of integration R_{I5} is defined by

$$\begin{cases} \dfrac{h(v - S_{di})}{t} > q_p \\ S_{di} < v \leq S_{il} + S_m \\ t > t_s \end{cases} \xrightarrow{yields} \begin{cases} \dfrac{q_p t}{h} + S_{di} < v \leq S_{il} + S_m \\ t > t_s \end{cases}$$

Let L_6 be the line defined by $v = S_{il} + S_m$, R_{I5} is the area above L_4 but below L_6, and also on the right side of L_s.

The sixth sub-region of integration R_{I6} is defined by

$$\begin{cases} \dfrac{v - S_d - (1-h)S_m}{t} > q_p \\ v > S_{il} + S_m \\ t > t_s \end{cases} \xrightarrow{yields} \begin{cases} v > q_p t + (1-h)S_m + S_d \\ v > S_{il} + S_m \\ t > t_s \end{cases}$$

Let L_7 be the line defined by $v = q_p t + (1-h)S_m + S_d$, R_{I6} is the area above both L_6 and L_7, and also on the right side of L_s.

The relative locations of the sub-regions depend on the relative positions of the lines that delineate them. According to the above definitions, sub-regions R_{I1} and R_{I2}

are on the left side of the vertical line L_c and delineated by the group of lines L_1, L_2, and L_3 (referred to as the G1 group). Sub-regions R_{I3} and R_{I4} are situated between the two vertical lines L_c and L_s and bounded by the group (referred to as the G2 group) of lines L_1, L_4, and L_5. Sub-regions R_{I5} and R_{I6} are located on the right side of L_s and bounded by the group (referred to as the G3 group) of lines L_4, L_6, and L_7. Then it makes sense, in order to determine the relative locations of the sub-regions, to investigate the relative positions of the lines of each group separately. The relative positions of lines depend on the relative magnitudes of their intercepts and slopes. Let I_1, I_2, and I_3 be the intercepts of lines L_1, L_2, and L_3 at the v-axis, respectively; I_{1c}, I_{2c}, and I_{3c} be the intercepts of lines L_1, L_2, and L_3 at L_c, respectively; I_{4c} and I_{5c} be the intercepts of lines L_4 and L_5 at L_c, respectively; I_{1s}, I_{4s}, and I_{5s} be the intercepts of lines L_1, L_4, and L_5 at L_s, respectively; and I_{4s}, I_{6s}, and I_{7s} be the intercepts of lines L_4, L_6, and L_7 at L_s, respectively. Let S_1, S_2, S_3, S_4, S_5, S_6, and S_7 be the slopes of lines L_1 through L_7, respectively. The required intercepts and slopes for the delineation of the different sub-regions are summarized in Table A.1. In order to determine the relative magnitudes of these intercepts and slopes, subtractions from each other are applied to the expressions of each couple of slopes and each couple of intercepts in each group of lines (i.e. G1, G2, and G3). The signs of these subtractions based on the values of q_p determine the relative magnitudes of the slopes or intercepts. The relative magnitudes of slopes of the lines in G1 are $S_1 > S_3 > S_2$ for all q_p; in G2 are $S_1 \leq S_4 \leq S_5$ if $q_p \leq hf_c$ and $S_5 \leq S_4 < S_1$ if $q_p > hf_c$; and in G3,

$S_6 \leq S_7 \leq S_4$ for all q_p.

For the three groups of lines, the relative magnitudes of intercepts depend on the catchment characteristics and q_p values. As summarized in Table A.2, the intercepts' relative magnitude differ depending on whether $\frac{S_{dd}}{t_c}$ is less or greater than $\frac{S_{dd}}{t_s} + f_c$. As a result, Type I catchments need to be divided into two sub-types: catchment Type I_a where $\frac{S_{dd}}{t_c} \leq \frac{S_{dd}}{t_s} + f_c$ and catchment Type I_b where $\frac{S_{dd}}{t_c} > \frac{S_{dd}}{t_s} + f_c$. For each sub-type of catchments there are four possible intervals of q_p where the relative magnitudes of intercepts of the three groups of lines may differ (Table A.2). For catchments of Type I_a, the four intervals are $0 < q_p \leq \frac{hS_{dd}}{t_c}$, $\frac{hS_{dd}}{t_c} < q_p \leq \frac{hS_{dd}}{t_s} + hf_c$, $\frac{hS_{dd}}{t_s} + hf_c < q_p \leq \frac{hS_{dd}}{t_c} + hf_c$, and $q_p > \frac{hS_{dd}}{t_c} + hf_c$, respectively. However, for the catchments of Type I_b, the four intervals are $0 < q_p \leq \frac{hS_{dd}}{t_s} + hf_c$, $\frac{hS_{dd}}{t_s} + hf_c < q_p \leq \frac{hS_{dd}}{t_c}$, $\frac{hS_{dd}}{t_c} < q_p \leq \frac{hS_{dd}}{t_c} + hf_c$, and $q_p > \frac{hS_{dd}}{t_c} + hf_c$, respectively. Thus, for each catchment types there are four possible configurations of the region of integration. The four possible intervals of q_p for Type I_a catchments are referred to as Int_{a1}, Int_{a2}, Int_{a3}, and Int_{a4} and those for Type I_b catchments are referred to as Int_{b1}, Int_{b2}, Int_{b3}, and Int_{b4}. As illustrated in Table A.2., the relative magnitudes of intercepts of the different groups of lines (i.e. G1, G2, and G3) when q_p values are in intervals Int_{a1}, Int_{a3}, and Int_{a4} for Type I_a catchments are the same as for Type I_b catchments when q_p values are in intervals Int_{b1}, Int_{b3}, and Int_{b4}; respectively. The relative magnitudes of intercepts in interval Int_{b2} are different from those in Int_{a2} (Table A.2). Thus in total there are five different configurations of the region of integration for Types I_a and I_b catchments.

4.A.2. Derivation of the Peak Discharge Exceedence Probability for Type I Catchments

Plotting the region of integration, which is the union of sub-regions R_{I1} through R_{I6}, facilitates the determination of the peak discharge exceedence probability $P[Q_p > q_p]$. As mentioned earlier the configuration of the region of integration depends on the catchment type and the interval that the q_p value of interest is located in. In the following, each configuration of the region of integration among the five possible configurations as discussed earlier is described together with the derivation of the corresponding exceedence probability.

The first region of integration configuration (Fig. A.1) is applicable for catchments of Types I_a and I_b with q_p taking on values from intervals Int_{a1} and Int_{b1}, respectively, as they have the same relative magnitudes of intercepts and slopes (Table A.2). In addition to the determination of the relative magnitudes of the intercepts and slopes of lines in each group, it is essential to figure out whether the lines of each group intersect. Given the conditions defining the two sub-types of catchments and the range of values of q_p for this configuration (i.e. Int_{a1} and Int_{b1}), the lines in G1 do not intersect, neither would lines in G2; however, the lines in G3 (i.e. L_4, L_6, and L_7) intersect at a common point referred to as I_{467} with coordinates $t_{467} = \frac{h(S_{dd}+S_m)}{q_p}$ and $v_{467} = S_{il} + S_m$. A shown in Fig. A.1, the area of integration is the area above L_2 for $t \leq t_c$, above L_4 for $t_c < t \leq t_{467}$, and above L_7 for $t > t_{467}$. Consequently, the exceedence probability of peak discharge can be determined as follows:

$$P[Q_p > q_p] = \int_0^{t_c} \int_{\frac{q_p t_c}{h}+S_{di}}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt + \int_{t_c}^{t_{467}} \int_{\frac{q_p}{h}t+S_{di}}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt + \int_{t_{467}}^{\infty} \int_{q_p t+(1-h)S_m+S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt$$

Carrying out the integration, the following expression is obtained:

$$P[Q_p > q_p] = \exp\left(-\frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) - \frac{\zeta q_p}{\lambda h + \zeta q_p} \exp\left(-\lambda t_c - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) + \frac{\lambda\zeta q_p (1-h)}{(\lambda h + \zeta q_p)(\lambda + \zeta q_p)} \exp\left(-\lambda \frac{h(S_{dd} + S_m)}{q_p} - \zeta S_m - \zeta S_{il}\right) \quad \text{(A.1)}$$

The second configuration (Fig. A.2) represents the region of integration for Type I_a catchments with q_p in Int_{a2}. For this configuration the lines of G1 meet at a common point referred to as I_{123} with coordinates $t_{123} = \frac{q_p t_c - h S_{dd}}{h f_c}$ and $v_{123} = \frac{q_p t_c}{h} + S_{di}$. The G2 lines do not intersect. The G3 lines intersect at a common point as already described for the first configuration. As shown in Fig. A.2, the region of integration is the area above L_3 for $t \leq t_{123}$, above L_2 for $t_{123} < t \leq t_c$, above L_4 for $t_c < t \leq t_{467}$, and above L_7 for $t > t_{467}$. Consequently, the exceedence probability of peak discharge can be determined as follows:

$$P[Q_p > q_p] = \int_0^{t_{123}} \int_{f_c(1-h)t+t_c q_p+S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt + \int_{t_{123}}^{t_c} \int_{\frac{q_p t_c}{h}+S_{di}}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt + \int_{t_c}^{t_{467}} \int_{\frac{q_p}{h}t+S_{di}}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt + \int_{t_{467}}^{\infty} \int_{q_p t+(1-h)S_m+S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt$$

Carrying out the integration results in the following expression:

$$P[Q_p > q_p] = \frac{\lambda}{\lambda + \zeta f_c(1-h)} \exp(-\zeta q_p t_c - \zeta S_d) - \frac{\zeta q_p}{\lambda h + \zeta q_p} \exp\left(-\lambda t_c - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) + \frac{\zeta f_c(1-h)}{\lambda + \zeta f_c(1-h)} \exp\left(-\lambda \frac{q_p t_c - hS_{dd}}{h f_c} - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) + \frac{\lambda\zeta q_p(1-h)}{(\lambda h + \zeta q_p)(\lambda + \zeta q_p)} \exp\left(-\lambda \frac{h(S_{dd}+S_m)}{q_p} - \zeta S_m - \zeta S_{il}\right) \quad \text{(A.2)}$$

The third configuration (Fig. A.3) represents the region of integration for Type I_a catchments with q_p in Int_{a3} and Type I_b catchments with q_p in Int_{b3} (Table A.2). For this configuration the lines of G1 intersect at a common point referred to as I_{123} as for the second configuration. The G2 lines intersect at a common point referred to as I_{145} with coordinates $t_{145} = \frac{hS_{dd}}{q_p - hf_c}$ and $v_{145} = \frac{q_p S_{dd}}{q_p - hf_c} + S_{di}$. The G3 lines do not intersect for the conditions defining this configuration (Table A.2). As shown in Fig. A.3, the

region of integration is the area above L_3 for $t \leq t_{123}$, above L_2 for $t_{123} < t \leq t_c$, above L_4 for $t_c < t \leq t_{145}$, above L_5 for $t_{145} < t \leq t_s$, and above L_7 for $t > t_s$. Consequently, the exceedence probability of peak discharge can be determined as follows:

$$
\begin{aligned}
P[Q_p > q_p] = & \int_0^{t_{123}} \int_{f_c(1-h)t+t_c q_p+S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt \\
& + \int_{t_{123}}^{t_c} \int_{\frac{q_p t_c}{h}+S_{di}}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt \\
& + \int_{t_c}^{t_{145}} \int_{\frac{q_p}{h}t+S_{di}}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt \\
& + \int_{t_{145}}^{t_s} \int_{[f_c(1-h)+q_p]t+S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt \\
& + \int_{t_s}^{\infty} \int_{q_p t+(1-h)S_m+S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt
\end{aligned}
$$

Carrying out the integration results in the following expression:

$$
\begin{aligned}
&P[Q_p > q_p] \\
&= \frac{\lambda}{\lambda + \zeta f_c(1-h)} \exp(-\zeta q_p t_c - \zeta S_d) \\
&- \frac{\zeta q_p}{\lambda h + \zeta q_p} \exp\left(-\lambda t_c - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) \\
&+ \frac{\zeta f_c(1-h)}{\lambda + \zeta f_c(1-h)} \exp\left(-\lambda \frac{q_p t_c - h S_{dd}}{h f_c} - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) \\
&+ \frac{(1-h)\lambda\zeta(q_p - h f_c)}{(\lambda h + \zeta q_p)(\lambda + \zeta q_p + (1-h)\zeta f_c)} \exp\left(-\frac{\lambda h S_{dd} - h\zeta f_c S_{di} + q_p \zeta S_{il}}{q_p - h f_c}\right) \\
&+ \frac{\lambda\zeta(1-h) f_c}{(\lambda + \zeta q_p)(\lambda + \zeta q_p + (1-h)\zeta f_c)} \exp(-\lambda t_s - \zeta q_p t_s - \zeta S_d \\
&- \zeta(1-h) S_m)
\end{aligned}
\tag{A.3}
$$

The fourth configuration (Fig. A.4) represents the region of integration for Type I_a catchments with q_p in Int_{a4} and Type I_b catchments with q_p in Int_{b4} (Table A.2). For this configuration none of the lines intersect. As shown in Fig. A.4, the region of integration is the area above L_3 for $t \le t_c$, above L_5 for $t_c < t \le t_s$, and above L_7 for $t > t_s$. Consequently, the exceedence probability of peak discharge can be determined as follows:

$$
\begin{aligned}
P[Q_p > q_p] = &\int_0^{t_c} \int_{f_c(1-h)t + t_c q_p + S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dv dt \\
&+ \int_{t_c}^{t_s} \int_{[f_c(1-h) + q_p]t + S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dv dt \\
&+ \int_{t_s}^{\infty} \int_{q_p t + (1-h)S_m + S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dv dt
\end{aligned}
$$

Carrying out the integration results in the following expression:

$$P[Q_p > q_p] = \frac{\lambda}{\lambda + \zeta f_c(1-h)} \exp(-\zeta q_p t_c - \zeta S_d) - \frac{\lambda \zeta q_p}{(\lambda h + \zeta(1-h)f_c)(\lambda + \zeta q_p + (1-h)\zeta f_c)} \exp(-\lambda t_c - \zeta[q_p t_c + S_d + f_c(1-h)t_c]) + \frac{\lambda \zeta (1-h) f_c}{(\lambda + \zeta q_p)(\lambda + \zeta q_p + (1-h)\zeta f_c)} \exp(-\lambda t_s - \zeta q_p t_s - \zeta S_d - \zeta(1-h)S_m) \quad \text{(A.4)}$$

The fifth configuration (Fig. A.5) represents the region of integration for Type I_b catchments with q_p in Int_{b2} (Table A.2). For this configuration, the lines of G1 do not intersect nor do the lines of G3. The G2 lines intersect at a common point referred to as I_{145} as described in the third configuration. As shown in Fig. A.5, the region of integration is the area above L_2 for $t \leq t_c$, above L_4 for $t_c < t \leq t_{145}$, above L_5 for $t_{145} < t \leq t_s$, and above L_7 for $t > t_s$. Consequently, the exceedence probability of peak discharge can be determined as follows:

$$P[Q_p > q_p] = \int_0^{t_c} \int_{\frac{q_p t_c}{h} + S_{di}}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dv dt + \int_{t_c}^{t_{145}} \int_{\frac{q_p}{h} t + S_{di}}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dv dt + \int_{t_{145}}^{t_s} \int_{[f_c(1-h)+q_p]t + S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dv dt + \int_{t_s}^{\infty} \int_{q_p t + (1-h)S_m + S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dv dt$$

Carrying out the integration results in the following expression:

$$P[Q_p > q_p]$$

$$= \exp\left(-\frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) - \frac{\zeta q_p}{\lambda h + \zeta q_p}\exp\left(-\lambda t_c - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right)$$

$$+ \frac{(1-h)\lambda\zeta(q_p - hf_c)}{(\lambda h + \zeta q_p)(\lambda + \zeta q_p + (1-h)\zeta f_c)}\exp\left(-\frac{\lambda h S_{dd} - h\zeta f_c S_{di} + q_p \zeta S_{il}}{q_p - hf_c}\right)$$

$$+ \frac{\lambda\zeta(1-h)f_c}{(\lambda + \zeta q_p)(\lambda + \zeta q_p + (1-h)\zeta f_c)}\exp\big(-\lambda t_s - \zeta q_p t_s - \zeta S_d$$

$$- \zeta(1-h)S_m\big) \qquad \text{(A.5)}$$

4.A.3. Integration Sub-regions of Type II Catchments (i.e. Catchments with $\mathbf{t_c > t_s}$)

Type II catchments have saturation time less than their respective time of concentration ($t_s < t_c$). Similar to Type I catchments, the region of integration is the union of the sub-regions of integration. The sub-regions of integrations for Type II catchments are defined in a similar way as they are defined for Type I catchments. Six sub-regions referred to as R_{II1}, R_{II2}, R_{II3}, R_{II4}, R_{II5}, and R_{II6} are defined in a similar way as R_{I1}, R_{I2}, R_{I3}, R_{I4}, R_{I5}, and R_{I6} are defined for Type I catchments but Eq. (8) instead of Eq. (7) is used as the basis. In the following, each of the sub-regions is defined and explained.

R_{II1} is defined by

$$\begin{cases} \dfrac{h(v - S_{di})}{t_c} > q_p \\ S_{di} < v \le S_{il} + f_c t \\ t \le t_s \end{cases} \xrightarrow{yields} \begin{cases} \dfrac{q_p t_c}{h} + S_{di} < v \le S_{il} + f_c t \\ t \le t_s \end{cases}$$

Then R_{II1} is the area above L_2 but below L_1, and on the left side of the vertical line L_s.

R_{II2} is defined by

$$\begin{cases} \dfrac{v - S_d - f_c(1-h)t}{t_c} > q_p \\ v > S_{il} + f_c t \\ t \le t_s \end{cases} \xrightarrow{yields} \begin{cases} v > q_p t_c + S_d + (1-h) f_c t \\ v > S_{il} + f_c t \\ t \le t_s \end{cases}$$

Then R_{II2} is the area above L_1 and L_3, and also on the left side of L_s.

R_{II3} is defined by

$$\begin{cases} \dfrac{h(v - S_{di})}{t} > q_p \\ S_{di} < v \le S_{il} + S_m \\ t_s < t \le t_c \end{cases} \xrightarrow{yields} \begin{cases} \dfrac{q_p t}{h} + S_{di} < v \le S_{il} + S_m \\ t_s < t \le t_c \end{cases}$$

Then R_{II3} is the area above L_4 but below L_7, and between L_s and L_c.

R_{II4} is defined by

$$\begin{cases} \dfrac{v - S_d - (1-h)S_m}{t_c} > q_p \\ v > S_{il} + S_m \\ t_s < t \le t_c \end{cases} \xrightarrow{yields} \begin{cases} v > q_p t_c + (1-h)\, S_m \ + S_d \\ v > S_{il} + S_m \\ t_s < t \le t_c \end{cases}$$

Let L_8 be the line defined by $v = q_p t_c + (1-h)\, S_m \ + S_d$, R_{II4} is the area above both L_1 and L_8, and also between L_s and L_c.

R_{II5} is defined by

$$\begin{cases} \dfrac{h(v - S_{di})}{t} > q_p \\ S_{di} < v \leq S_{il} + S_m \\ t > t_c \end{cases} \xrightarrow{yields} \begin{cases} \dfrac{q_p t}{h} + S_{di} < v \leq S_{il} + S_m \\ t > t_c \end{cases}$$

Then R_{II5} is the area above L_4 but below L_6, and also on the right side of L_c.

R_{II6} is defined by

$$\begin{cases} \dfrac{v - S_d - (1-h)S_m}{t} > q_p \\ v > S_{il} + S_m \\ t > t_c \end{cases} \xrightarrow{yields} \begin{cases} v > q_p t + (1-h)S_m + S_d \\ v > S_{il} + S_m \\ t > t_c \end{cases}$$

Then R_{II6} is the area above both L_6 and L_7, and also on the right side of L_c.

The sub-regions R_{II1} and R_{II2} are on the left side of the vertical line L_s and delineated by the lines L_1, L_2, and L_3 (G1 group of lines as defined for Type I catchments). The sub-regions R_{II3} and R_{II4} are situated between the two vertical lines L_s and L_c and defined by the lines L_4, L_6, and L_8; these three lines together form a group referred to as the G4 group. The sub-regions R_{II5} and R_{II6} are located on the right side of L_c and delineated by the lines L_4, L_6, and L_7 (G3 group of lines). The relative magnitudes of the intercepts at v-axis of the G1 lines are already evaluated. Let I_{1s}, I_{2s}, I_{3s}, I_{4s}, I_{6s}, and I_{8s} be, respectively, the intercepts at L_s of the lines L_1, L_2, L_3, L_4, L_6, and L_8; then $I_{1s} = S_{il} + S_m = I_{6s}$, $I_{2s} = \frac{q_p t_c}{h} + S_{di} = I_2$, $I_{3s} = q_p t_c + S_d + (1-h)S_m = I_{8s}$, and $I_{4s} = \frac{q_p t_s}{h} + S_{di}$. Let I_{4c}, I_{6c}, I_{7c}, and I_{8c} be, respectively, the intercepts at L_c of the lines L_4, L_6, L_7, and L_8; then $I_{4c} = \frac{q_p t_c}{h} + S_{di}$, $I_{6c} = S_{il} + S_m$, and $I_{7c} = q_p t_c + S_d + (1-h)S_m = I_{8c}$.

Similar to what was done for Type I catchments, the relative magnitudes of intercepts are determined by subtracting one from the other for each pair of intercepts within each group of lines. The relative magnitudes of intercepts of each group of lines within different possible intervals of q_p values are summarized in Table A.3. For G1 and G2 groups of lines, the relative magnitudes of slopes are the same as described for Type I catchments. For group G4, let S_8 be the slope of line L_8, which is newly introduced for Type II catchments, $S_8 = 0$ since L_8 is a horizontal line. Therefore $S_4 \geq S_6 = S_8$. As shown in Table A.3, there are four possible intervals of q_p (referred to as Int_1, Int_2, Int_3, and Int_4) values where the relative magnitudes of the intercepts of the G1, G4, and G3 three groups of lines may differ. Thus there are four possible configurations of the region of integration for Type II catchments, these configurations are referred to as the sixth though ninth configurations.

4.A.4. Derivation of the Peak Discharge Exceedence Probability for Type II Catchments

The sixth configuration (Fig. A.6) represents the region of integration for Type II catchments with q_p in Int_1 (Table A.3). For this configuration, the lines of G1 do not intersect nor do the lines of G4. The G3 lines intersect at I_{467} as described in the first configuration. As shown in Fig. A.6, the region of integration is the area above L_2 for $t \leq t_s$, above L_4 for $t_s < t \leq t_{467}$, and above L_7 for $t > t_{467}$. Consequently, the exceedence probability of peak discharge can be determined as follows:

$$P[Q_p > q_p] = \int_0^{t_s} \int_{\frac{q_p t_c}{h}+S_{di}}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt + \int_{t_s}^{t_{467}} \int_{\frac{q_p}{h}t+S_{di}}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt + \int_{t_{467}}^{\infty} \int_{q_p t+(1-h)S_m+S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt$$

Carrying out the integration results in the following expression:

$$P[Q_p > q_p] = \exp\left(-\frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) - \exp\left(-\lambda t_s - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) + \frac{\lambda h}{\lambda h + \zeta q_p} \exp\left(-\lambda t_s - \frac{\zeta q_p t_s}{h} - \zeta S_{di}\right) + \frac{\lambda\zeta(1-h)q_p}{(\lambda h + \zeta q_p)(\lambda + \zeta q_p)} \exp\left(-\frac{\lambda h(S_{dd}+S_m)}{q_p} - \zeta(S_{il}+S_m)\right) \quad \text{(A.6)}$$

The seventh configuration (Fig. A.7) represents the region of integration for Type II catchments with q_p in Int_2 (Table A.3). For this configuration, G1 lines intersect at I_{123} as mentioned in the second configuration, G4 lines do not intersect and G3 lines meet at I_{467} as described in the first configuration. As shown in Fig. A.7, the region of integration is the area above L_3 for $t \leq t_{123}$, above L_2 for $t_{123} < t \leq t_s$, above L_4 for $t_s < t \leq t_{467}$, and above L_7 for $t > t_{467}$. Consequently, the exceedence probability of peak discharge can be determined as follows:

$$P[Q_p > q_p] = \int_0^{t_{123}} \int_{f_c(1-h)t+t_c q_p+S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt + \int_{t_{123}}^{t_s} \int_{\frac{q_p t_c}{h}+S_{di}}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt + \int_{t_s}^{t_{467}} \int_{\frac{q_p}{h}t+S_{di}}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt + \int_{t_{467}}^{\infty} \int_{q_p t+(1-h)S_m+S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt$$

Carrying out the integration results in the following expression:

$$P[Q_p > q_p] = \frac{\lambda}{\lambda + \zeta f_c(1-h)} \exp(-\zeta q_p t_c - \zeta S_d) + \frac{\zeta f_c(1-h)}{\lambda + \zeta f_c(1-h)} \exp\left(-\lambda \frac{q_p t_c - hS_{dd}}{h f_c} - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) - \exp\left(-\lambda t_s - \frac{\zeta q_p t_c}{h} - \zeta S_{di}\right) + \frac{\lambda h}{\lambda h + \zeta q_p} \exp\left(-\lambda t_s - \frac{\zeta q_p t_s}{h} - \zeta S_{di}\right) + \frac{\lambda\zeta(1-h)q_p}{(\lambda h + \zeta q_p)(\lambda + \zeta q_p)} \exp\left(-\frac{\lambda h(S_{dd} + S_m)}{q_p} - \zeta(S_{il} + S_m)\right) \quad \text{(A.7)}$$

The eighth configuration (Fig. A.8) represents the region of integration for Type II catchments with q_p in Int_3 (Table A.3). For this configuration, G1 lines do not intersect neither do the G3 lines. According to the relative magnitudes of the intercepts of G1 and G4 lines at L_s and those of G3 and G4 lines at L_c (Table A.3), L_4 crosses L_6 at I_{46} and L_8 at I_{48}. The point I_{46} has the same coordinates as the point I_{467}

defined in the first configuration. The coordinates of point I_{48} are not important to be determined here because L_4 and L_8 are not part of the lines that define a sub-region of integration; neither for R_{II4} nor for R_{II5}. As shown in Fig. A.8, the region of integration is the area above L_3 for $t \leq t_s$, above L_8 for $t_s < t \leq t_c$, above L_4 for $t_s < t \leq t_{467}$, and above L_7 for $t > t_c$. Consequently, the exceedence probability of peak discharge can be determined as follows:

$$\begin{aligned} P[Q_p > q_p] = & \int_0^{t_s} \int_{f_c(1-h)t+t_c q_p+S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt \\ & + \int_{t_s}^{t_c} \int_{q_p t_c+(1-h)S_m+S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt \\ & + \int_{t_s}^{t_{467}} \int_{\frac{q_p}{h}t+S_{di}}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt \\ & + \int_{t_c}^{\infty} \int_{q_p t+(1-h)S_m+S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt \end{aligned}$$

Carrying out the integration results in the following expression:

$$\begin{aligned} P[Q_p > q_p] = & \frac{\lambda}{\lambda + \zeta f_c(1-h)} \exp\left(-\zeta q_p t_c - \zeta S_d\right) \\ & + \frac{\zeta f_c(1-h)}{\lambda + \zeta f_c(1-h)} \exp\left(-\lambda t_s - \zeta(1-h)S_m - \zeta q_p t_c \right. \\ & \left. - \zeta S_d\right) \\ & - \frac{\zeta q_p}{\lambda + \zeta q_p} \exp\left(-\lambda t_c - \zeta(1-h)S_m - \zeta q_p t_c - \zeta S_d\right) \\ & + \frac{\lambda q_p}{\lambda h + \zeta q_p} \exp\left(-\frac{\lambda h(S_{dd}+S_m)}{q_p} - \zeta(S_{il}+S_m)\right) \\ & - \exp\left(-\lambda t_s - \zeta(S_{il}+S_m)\right) \\ & + \frac{\lambda h}{\lambda h + \zeta q_p} \exp\left(-\lambda t_s - \frac{\zeta q_p t_s}{h} - \zeta S_{di}\right) \end{aligned} \quad \text{(A.8)}$$

The ninth configuration (Fig. A.9) represents the region of integration for Type II catchments with q_p in Int_4 (Table A.3). This configuration is the same as the eighth configuration except that the lines L_4 and L_6 do not intersect. As shown in Fig. A.9, the region of integration is the area above L_3 for $t \leq t_s$, above L_8 for $t_s < t \leq t_c$, and above L_7 for $t > t_c$. Consequently, the exceedence probability of peak discharge can be determined as follows:

$$P[Q_p > q_p] = \int_0^{t_s} \int_{f_c(1-h)t+t_c q_p+S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt + \int_{t_s}^{t_c} \int_{q_p t_c+(1-h)S_m+S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt + \int_{t_c}^{\infty} \int_{q_p t+(1-h)S_m+S_d}^{\infty} \zeta\lambda \exp(-\zeta v - \lambda t)\, dvdt$$

Carrying out the integration results in the following expression:

$$P[Q_p > q_p] = \frac{\lambda}{\lambda + \zeta f_c(1-h)} \exp\left(-\zeta q_p t_c - \zeta S_d\right) + \frac{\zeta f_c(1-h)}{\lambda + \zeta f_c(1-h)} \exp\left(-\lambda t_s - \zeta(1-h)S_m - \zeta q_p t_c - \zeta S_d\right) - \frac{\zeta q_p}{\lambda + \zeta q_p} \exp\left(-\lambda t_c - \zeta(1-h)S_m - \zeta q_p t_c - \zeta S_d\right) \quad \text{(A.9)}$$

Eq. (11) and Eqs. (A.1) through (A.9) are all the analytical expressions needed to determine the exceedence probability of peak discharge per rainfall event adopting trapezoidal runoff hydrograph simplifications and taking into consideration both infiltration- and saturation-excess runoff. They form the fundamentals of what we refer to as the $\text{APSWM}_{\text{tis}}$.

Acknowledgements: This work was supported by the Natural Sciences and Engineering Research Council of Canada. This work was also partially completed within the framework of the Panta Rhei Research Initiative of the International Association of Hydrological Sciences. The data used are listed on the references and tables.

References

Adams, B. J., H. G. Fraser, C. D. D. Howard, and M. S. Hanafy (1986). Meteorologic data analysis for drainage system design. *Journal of Enviromental Engineering*, *ASCE*, *Vol. 112*, *No. 5*, 827-848.

Adams, B. J., and F. Papa (2000). *Urban Stormwater Management Planning with Analytical Probabilistic Models*. John Wiley & Sons, Inc., New York, NY, USA.

Akan, A. O., and Houghtalen, R. J. (2003). *Urban Hydrology, Hydraulics, and Stormwater Quality: engineering applications and computer modeling*. John Wiley & Sons, Inc., New York, NY, USA.

Bacchi, B., M. Balistrocchi, and G. Grossi (2008). Proposal of a semi-probabilistic approach for storage facility design. *Urban Water J.*, 5(3), 195- 208, doi:10.1080/15730620801980723.

Balistrocchi, M., G. Grossi, and B. Bacchi (2009). An analytical probabilistic model

of the quality efficiency of a sewer tank. *Water Resour. Res*., 45, W12420, doi:10.1029/2009WR007822, 2009

Benjamin, J. R., and C. A. Cornell (1970). *Probability, Statistics and Decision for Civil Engineers*. McGraw-Hill, New York.

Chen, J., and Adams, B. J., (2005). Urban storm water control evaluation with analytical probabilistic models. *Journal of Water Resources Planning and Management*, 131(5), 362-374.

Chen, J., and Adams, B. J., (2007). Development of analytical models for estimation of urban stormwater runoff. *Journal of Hydrology*, 336(3), 458-469.

Eagleson, P. S. (1972). Dynamics of flood frequency. *Water Resources Research*, 8(4), 878–898.

Eagleson, P. S. (1978). Climate, soil, and vegetation, 2, the distribution of annual precipitation derived from observed storm sequences. *Water Resources Research*, 14(5), 713–721.

Guo, Y. (2001). Hydrologic design of urban flood control detention ponds. *Journal of Hydrologic Engineering*, 6(6), 472–479.

Guo, Y., and Adams, B. J. (1998a). Hydrologic analysis of urban catchments with event-based probabilistic models: 1. Runoff volume. *Water Resources Research*, 34(12), 3421-3431.

Guo, Y., and Adams, B. J. (1998b). Hydrologic analysis of urban catchments with event-based probabilistic models: 2. Peak discharge rate. *Water Resources*

Research, 34(12), 3433-3443.

Guo, Y., and Adams, B. J. (1999a). Analysis of detention ponds for storm water quality control. *Water Resources Research,* 35(8), 2447-2456.

Guo, Y., and Adams, B. J. (1999b). An analytical probabilistic approach to sizing flood control detention facilities. *Water Resources Research,* 35(8), 2457-2468.

Guo, Y., and Baetz, B. W. (2007). Sizing of rainwater storage units for green building applications. *Journal of Hydrologic Engineering*, 12(2), 197– 205.

Guo, Y., Hansen, D. and Li, C. (2009). Probabilistic approach to estimating the effects of channel reaches on flood frequencies. *Water Resources Research,* 45, W08404, Doi:10.1029/2008WR007387

Guo, Y. and Markus, M. (2011). An Analytical Probabilistic Approach for Estimating Design Floods of Small Watersheds. *Journal of Hydrologic Engineering*, Vol. 16, No. 11, 847-857. doi:10.1061/(ASCE)HE.1943-5584.0000380

Guo, Y., Liu, S., & Baetz, B. W. (2012). Probabilistic rainfall-runoff transformation considering both infiltration and saturation excess runoff generation processes. *Water Resources Research*, 48, W06513. DOI: 10.1029/2011WR011613

Guo, Y., and Zhuge, Z. (2008). Analytical probabilistic flood routing for urban stormwater management purposes. *Canadian Journal of Civil Engineering*, 35(5), 487-499.

Hassini, S. and Guo, Y. (2016). Exponentiality Test Procedures for Large Samples of Rainfall Event Characteristics. *Journal of Hydrologic Engineering,* 21*(4).*

https://doi.org/10.1061/(ASCE)HE.1943-5584.0001352

Hassini, S. and Guo, Y. (2017). Derived flood frequency distributions considering individual event hydrograph shapes. *Journal of Hydrology*, *547*, 296-308. https://doi.org/10.1016/j.jhydrol.2017.02.003

Hicks, W. I. (1944). A method of computing urban runoff. *Transactions of the American Society of Civil Engineers*, *109*(1), 1217-1253.

Howard, C. D. D. (1976). Theory of storage and treatment plant overflows. *Journal of the Environmental Engineering Division,* ASCE, 102(EE4), 709–722.

Ken, E. (2015). *Compute watershed time of concentration using FAA equation (rational method), Kirpich equation, or Kerby equation*, <http://www.LMNOeng.com> (November, 2016).

Levy, B., and McCuen, R. (1999). Assessment of storm duration for hydrologic design. *Journal of Hydrologic Engineering*, 4(3), 209-213.

Miao, W. (2016*). Modeling and Analysing the Impacts of Future Land Use and Climate Changes on the Stormwater Management in the Davis Creek Subwatershed*. Master of Engineering report, McMaster University, Hamilton, ON, Canada.

NOAA's National Weather Service (2011). *The Precipitation Frequency Data Server*, <http://hdsc.nws.noaa.gov/hdsc/pfds/> (September 2011)

Quader, A., and Guo, Y. (2006). Peak discharge estimation using analytical probabilistic and design storm approaches. *Journal of Hydrologic Engineering*,

11(1), 46-54.

Ponce, V. M. (1989). *Engineering Hydrology: Principles and Practices*. Prentice Hall, New Jersey, USA.

Viessman Jr, W., & Lewis, G. L. (2003). *Introduction to Hydrology, Fifth Edition*. Pearson Education, Inc., Upper Saddle River, NJ, 187-190.

Wurbs, R. A., and James, P. W. (2002). *Water Resources Engineering*. Prentice-Hall, Inc., NJ, USA.

Zhang, S., and Guo, Y. (2013a). Analytical Probabilistic Model for Evaluating the Hydrologic Performance of Green Roofs. *Journal of Hydrologic Engineering*, 18(1), 19-28.

Zhang, S., and Guo, Y. (2013b). Explicit Equation for Estimating Storm-Water Capture Efficiency of Rain Gardens. *Journal of Hydrologic Engineering*, 18(12), 1739-1748.

Table 1. Peak Discharge Exceedence Probability Expressions for all Catchment Types

$P[Q_p > q_p]$	**Interval**
Type I_a catchments (i.e. $t_c \le t_s$ and $\frac{S_{dd}}{t_c} \le \frac{S_{dd}}{t_s} + f_c$)	
$g_1(q_p) - g_2(q_p) + g_7(q_p)$,	$q_p \le \frac{hS_{dd}}{t_c}$
$g_4(q_p) - g_2(q_p) + g_3(q_p) + g_7(q_p)$,	$\frac{hS_{dd}}{t_c} < q_p \le \frac{hS_{dd}}{t_s} + hf_c$
$g_4(q_p) - g_2(q_p) + g_3(q_p) + g_5(q_p) + g_8(q_p)$,	$\frac{hS_{dd}}{t_s} + hf_c < q_p \le \frac{hS_{dd}}{t_c} + hf_c$
$g_4(q_p) - g_6(q_p) + g_8(q_p)$,	$q_p > \frac{hS_{dd}}{t_c} + hf_c$
Type I_b catchments (i.e. $t_c \le t_s$ and $\frac{S_{dd}}{t_c} > \frac{S_{dd}}{t_s} + f_c$)	
$g_1(q_p) - g_2(q_p) + g_7(q_p)$,	$q_p \le \frac{hS_{dd}}{t_s} + hf_c$
$g_1(q_p) - g_2(q_p) + g_3(q_p) + g_8(q_p)$,	$\frac{hS_{dd}}{t_s} + hf_c < q_p \le \frac{hS_{dd}}{t_c}$
$g_4(q_p) - g_2(q_p) + g_3(q_p) + g_5(q_p) + g_8(q_p)$,	$\frac{hS_{dd}}{t_c} < q_p \le \frac{hS_{dd}}{t_c} + hf_c$
$g_4(q_p) - g_6(q_p) + g_8(q_p)$,	$q_p > \frac{hS_{dd}}{t_c} + hf_c$
Type II catchments (i.e. $t_c > t_s$)	
$g_1(q_p) - g_{13}(q_p) + g_{12}(q_p) + g_{14}(q_p)$,	$q_p \le \frac{hS_{dd}}{t_c}$
$g_4(q_p) + g_5(q_p) - g_{13}(q_p) + g_{12}(q_p) + g_{14}(q_p)$,	$\frac{hS_{dd}}{t_c} < q_p \le \frac{h(S_{dd} + S_m)}{t_c}$
$g_4(q_p) - g_9(q_p) + g_{10}(q_p) + g_{11}(q_p) - E + g_{11}(q_p)$,	$\frac{h(S_{dd} + S_m)}{t_c} < q_p \le \frac{h(S_{dd} + S_m)}{t_s}$
$g_4(q_p) - g_9(q_p) + g_{10}(q_p)$,	$q_p > \frac{h(S_{dd} + S_m)}{t_s}$

Table 2. $APSWM_{tis}$ Rainfall Input Data (Hassini and Guo 2016)

Station	$\bar{v}$ (mm)	$\bar{t}$ (h)	θ
St. Cloud Airport, MN	16.5	11.5	31.8
Des Moines Airport, IA	16.5	10	36.2
North Platte Airport, NE	14.8	9.6	26.4
Fargo Airport, ND	16.8	12	22.4
Huron Airport, SD	13.5	10.2	29.1
Doge Airport, KS	16.4	9.1	24.6
Springfield Airport, MO	23.4	13.3	35.5

Table 3. 24 – Hour Partial Duration Series Rainfall Volumes for Different Return Periods

Station	2 - yr	5 - yr	10 - yr	25 - yr	50 - yr	100 - yr
St. Cloud Airport, MN	68.6	85.3	100.3	122.7	141.5	161.5
Des Moines Airport, IA	77.2	95.0	111.5	136.7	157.7	180.3
North Platte Airport, NE	58.7	71.9	83.6	100.6	114.6	129.0
Fargo Airport, ND	62.2	79.8	96.0	120.9	142.0	165.1
Huron Airport, SD	59.4	73.9	86.9	106.4	122.7	139.7
Doge Airport, KS	65.5	81.0	94.7	115.1	131.6	148.8
Springfield Airport, MO	91.9	111.3	128.3	153.4	174.2	196.1

Table 4. Test Catchment's Characteristics

Area (ha)	105.73623
Width (m)	1195.057
Length L(m)	1487
Slope S (%)	0.87
Imperviousness (%)	44.7
Manning's N-Impervious	0.012
Manning's N-pervious	0.25
Impervious depression storage (mm)	0.049
Pervious depression storage (mm)	0.098
Maximum Infiltration rate f_0(mm/h)	21.48
Minimum Infiltration rate f_c (mm/h)	0.36
Decay constant k (1/h)	4.14
Drying time (days)	10
Saturation volume S_s (mm)	100

Table 5. Percentage of Flood Frequency Differences between $APSWM_{tis}$ and SWMM under Different Climate Conditions

T_R (yrs)	MN	IA	NE	ND	SD	KS	MO
2	-0.1	-3.7	6.0	-4.6	-4.0	2.4	0.8
5	7.6	3.0	17.2	3.8	3.9	11.7	12.7
10	8.9	3.0	19.9	4.1	4.8	14.3	16.7
25	7.1	0.1	20.4	-0.4	3.0	13.7	18.0
50	4.6	-3.1	19.2	-2.3	0.5	12.1	17.1
100	1.6	-6.4	17.6	-6.1	-2.1	10.2	15.8

Table A.1. Intercepts and Slopes of Lines L_1 through L_7

	Intercept at v-axis	Intercept at L_c	Intercept at L_s	Slope
L_1	S_{il}	$f_c t_c + S_{il}$	$S_m + S_{il}$	f_c
L_2	$\frac{q_p}{h} t_c + S_{di}$	$\frac{q_p}{h} t_c + S_{di}$	NN	0
L_3	$t_c q_p + S_d$	$[f_c(1-h) + q_p]t_c + S_d$	NN	$f_c(1-h)$
L_4	NN*	$\frac{q_p}{h} t_c + S_{di}$	$\frac{q_p}{h} t_s + S_{di}$	$\frac{q_p}{h}$
L_5	NN	$[f_c(1-h) + q_p]t_c + S_d$	$[f_c(1-h) + q_p]t_s + S_d$	$f_c(1-h) + q_p$
L_6	NN	NN	$S_m + S_{il}$	0
L_7	NN	NN	$q_p t_s + (1-h)S_m + S_d$	q_p

* NN means not necessary.

Table A.2. Relative Magnitudes of the Intercepts of the Lines Defining the Sub-regions of Integration for Type I Catchments ($t_c \leq t_s$)

Type I_a Catchments with $t_c \leq t_s$ and $\frac{S_{dd}}{t_c} \leq \frac{S_{dd}}{t_s} + f_c$			
$0 < q_p \leq \frac{hS_{dd}}{t_c}$	$\frac{hS_{dd}}{t_c} < q_p \leq \frac{hS_{dd}}{t_s} + hf_c$	$\frac{hS_{dd}}{t_s} + hf_c < q_p \leq \frac{hS_{dd}}{t_c} + hf_c$	$q_p > \frac{hS_{dd}}{t_c} + hf_c$
(Int_{a1})	(Int_{a2})	(Int_{a3})	(Int_{a4})
$I_1 \geq I_3 \geq I_2$ $I_{1c} \geq I_{3c} \geq I_{2c}$ $I_{1s} \geq I_{5s} \geq I_{4s}$	$I_2 > I_3 > I_1$ $I_{1c} \geq I_{3c} \geq I_{2c}$ $I_{1s} \geq I_{5s} \geq I_{4s}$	$I_2 > I_3 > I_1$ $I_{1c} \geq I_{3c} \geq I_{2c}$ $I_{4s} > I_{5s} > I_{1s}$	$I_2 > I_3 > I_1$ $I_{2c} \geq I_{3c} \geq I_{1c}$ $I_{4s} > I_{5s} > I_{1s}$
Type I_b Catchments with $t_c \leq t_s$ and $\frac{S_{dd}}{t_c} > \frac{S_{dd}}{t_s} + f_c$			
$0 < q_p \leq \frac{hS_{dd}}{t_s} + hf_c$	$\frac{hS_{dd}}{t_s} + hf_c < q_p \leq \frac{hS_{dd}}{t_c}$	$\frac{hS_{dd}}{t_c} < q_p \leq \frac{hS_{dd}}{t_c} + hf_c$	$q_p > \frac{hS_{dd}}{t_c} + hf_c$
(Int_{b1})	(Int_{b2})	(Int_{b3})	(Int_{b4})
$I_1 \geq I_3 \geq I_2$ $I_{1c} \geq I_{3c} \geq I_{2c}$ $I_{1s} \geq I_{5s} \geq I_{4s}$	$I_2 > I_3 > I_1$ $I_{1c} \geq I_{3c} \geq I_{2c}$ $I_{4s} > I_{5s} > I_{1s}$	$I_2 > I_3 > I_1$ $I_{1c} \geq I_{3c} \geq I_{2c}$ $I_{4s} > I_{5s} > I_{1s}$	$I_2 > I_3 > I_1$ $I_{2c} \geq I_{3c} \geq I_{1c}$ $I_{4s} > I_{5s} > I_{1s}$

Table A.3. Relative Magnitudes of the Intercepts of the Lines Defining the Sub-regions of Integration for Type II Catchments ($t_c > t_s$)

$0 < q_p \leq \frac{hS_{dd}}{t_c}$	$\frac{hS_{dd}}{t_c} < q_p \leq \frac{h(S_{dd}+S_m)}{t_c}$	$\frac{h(S_{dd}+S_m)}{t_c} < q_p \leq \frac{h(S_{dd}+S_m)}{t_s}$	$q_p > \frac{h(S_{dd}+S_m)}{t_s}$
(Int_1)	(Int_2)	(Int_3)	(Int_4)
$I_1 \geq I_3 \geq I_2$	$I_2 > I_3 > I_1$	$I_2 > I_3 > I_1$	$I_2 > I_3 > I_1$
$I_{1s} \geq I_{3s} \geq I_{2s} \geq I_{4s}$	$I_{1s} \geq I_{3s} \geq I_{2s} \geq I_{4s}$	$I_{2s} > I_{3s} > I_{1s} \geq I_{4s}$	$I_{2s} > I_{3s} > I_{1s}$
		$I_{4c} > I_{7c} > I_{6c}$	$I_{2s} > I_{4s} > I_{1s}$
$I_{6c} \geq I_{7c} \geq I_{4c}$	$I_{6c} \geq I_{7c} \geq I_{4c}$		$I_{4c} > I_{7c} > I_{6c}$

Figure Captions

Fig. 1. Flood frequency results from $APSWM_{tis}$ (analytical) and SWMM (design storm) with the test catchment under the two climate conditions used for calibration and validation

Fig. 2. Best (SD) and worst (NE) comparison results between $APSWM_{tis}$ and SWMM

Fig. A.1. Region of integration for Type I_a catchments with $q_p \leq \frac{hs_{dd}}{t_c}$ and Type I_b catchments with $q_p \leq \frac{hs_{dd}}{t_c} + hf_c$

Fig. A.2. Region of integration for Type I_a catchments with $\frac{hs_{dd}}{t_c} < q_p \leq \frac{hs_{dd}}{t_S} + hf_c$

Fig. A.3. Region of integration for Type I_a catchments with $\frac{hs_{dd}}{t_S} + f_c h < q_p \leq \frac{hs_{dd}}{t_c} + hf_c$ and Type I_b catchments with $\frac{hs_{dd}}{t_c} < q_p \leq \frac{hs_{dd}}{t_c} + hf_c$

Fig. A.4. Region of integration for Type I_a catchments with $q_p > \frac{hs_{dd}}{t_c} + hf_c$ and Type I_b catchments with $q_p > \frac{hs_{dd}}{t_c} + hf_c$

Fig. A.5. Region of integration for Type I_b catchments with $\frac{hs_{dd}}{t_S} + hf_c < q_p \leq \frac{hs_{dd}}{t_c}$

Fig. A.6. Region of integration for Type II catchments with $0 < q_p \leq \frac{hs_{dd}}{t_c}$

Fig. A.7. Region of integration for Type II catchments with $\frac{hs_{dd}}{t_c} < q_p \leq \frac{h(s_{dd}+S_m)}{t_c}$

Fig. A.8. Region of integration for Type II catchments with $\frac{h(s_{dd}+S_m)}{t_c} < q_p \leq \frac{h(s_{dd}+S_m)}{t_s}$

Fig. A.9. Region of integration for Type II catchments with $q_p > \frac{h(s_{dd}+S_m)}{t_s}$

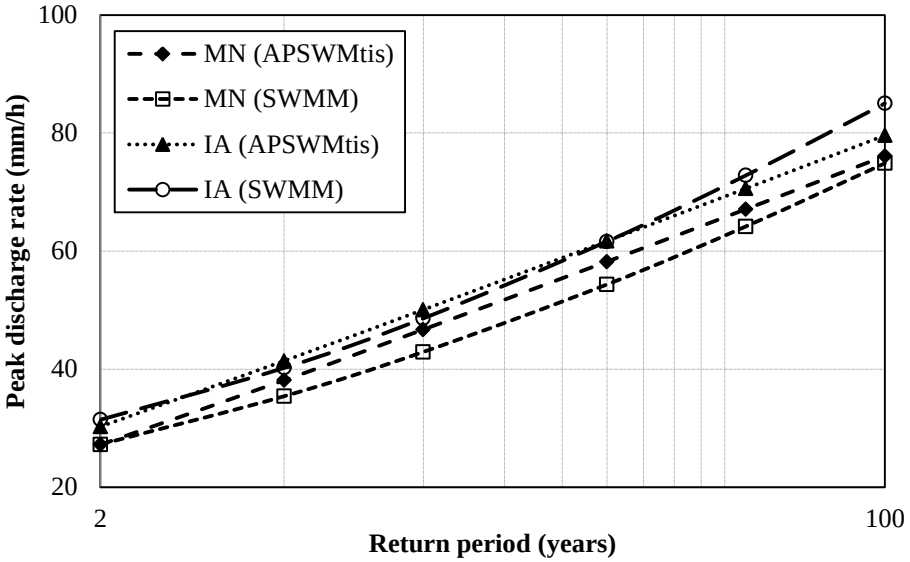

Fig. 1. Flood frequency results from $APSWM_{tis}$ (analytical) and SWMM (design storm) with the test catchment under the two climate conditions used for calibration and validation

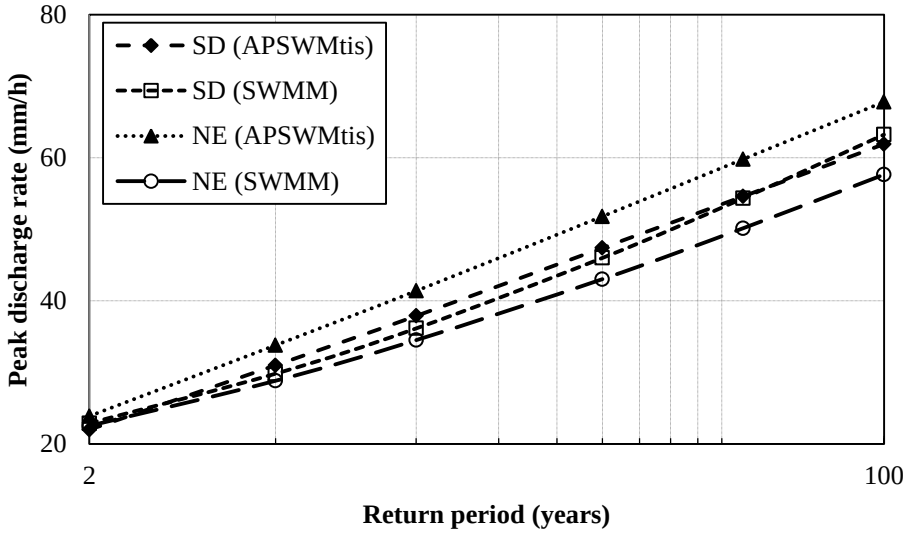

Fig. 2. Best (SD) and worst (NE) comparison results between $APSWM_{tis}$ and SWMM

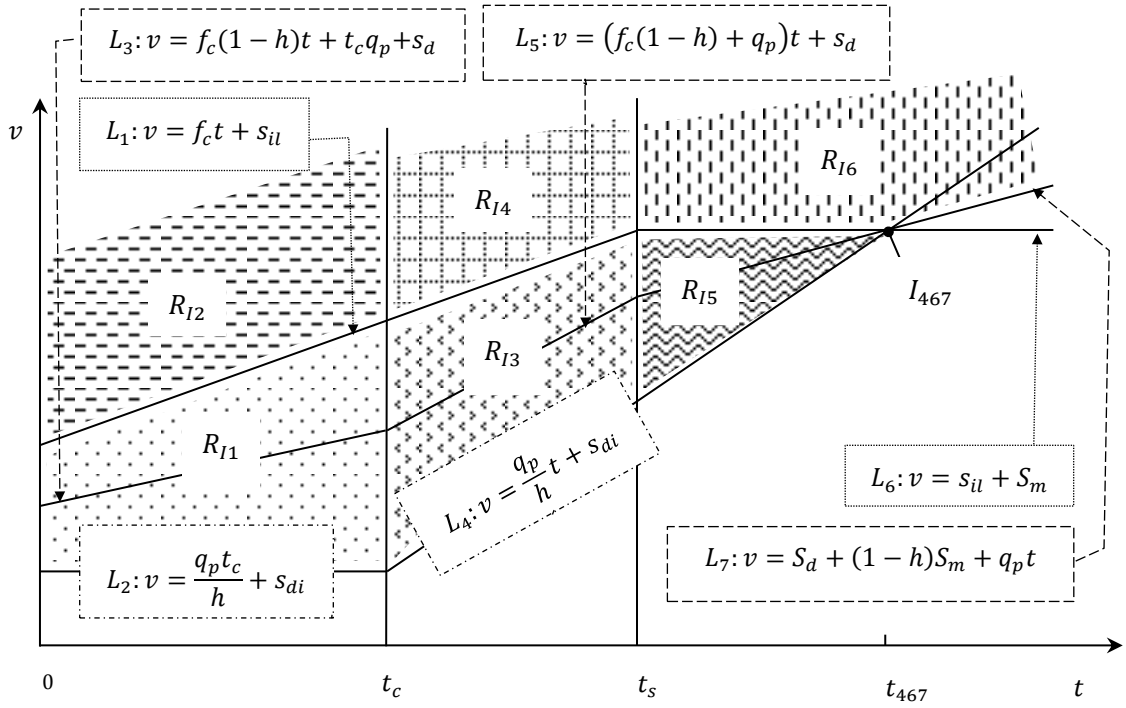

Fig. A.1. Region of integration for Type I_a catchments with $q_p \leq \frac{hs_{dd}}{t_c}$ and Type I_b catchments with $q_p \leq \frac{hs_{dd}}{t_c} + hf_c$

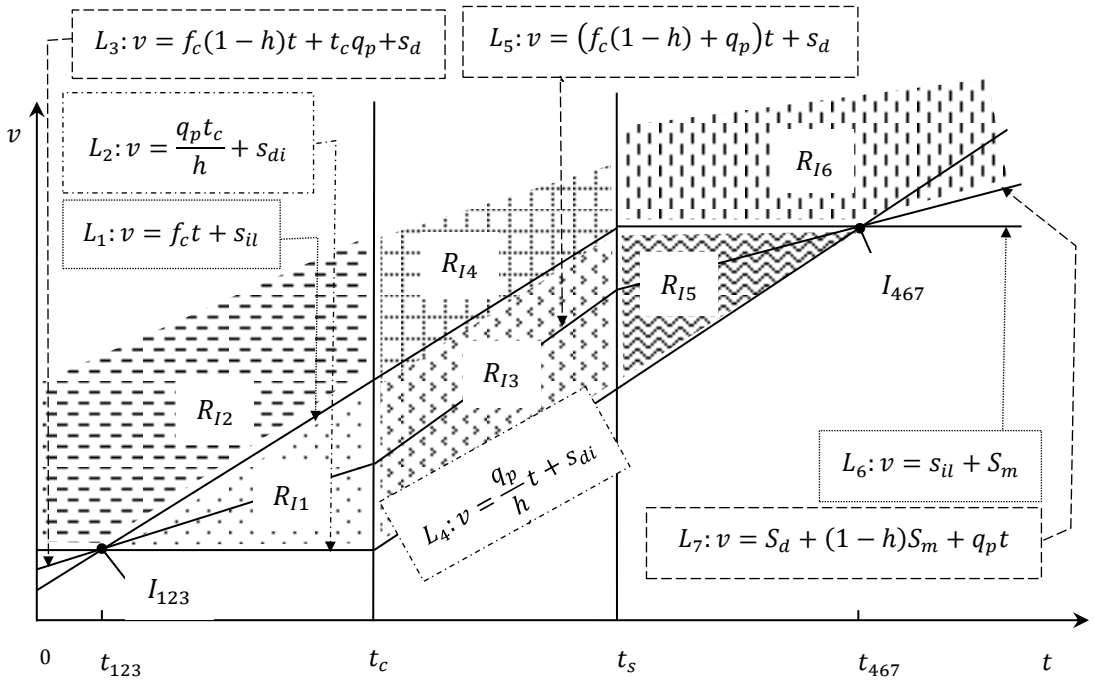

Fig. A.2. Region of integration for Type I_a catchments with

$$\frac{hs_{dd}}{t_c} < q_p \leq \frac{hs_{dd}}{t_S} + hf_c$$

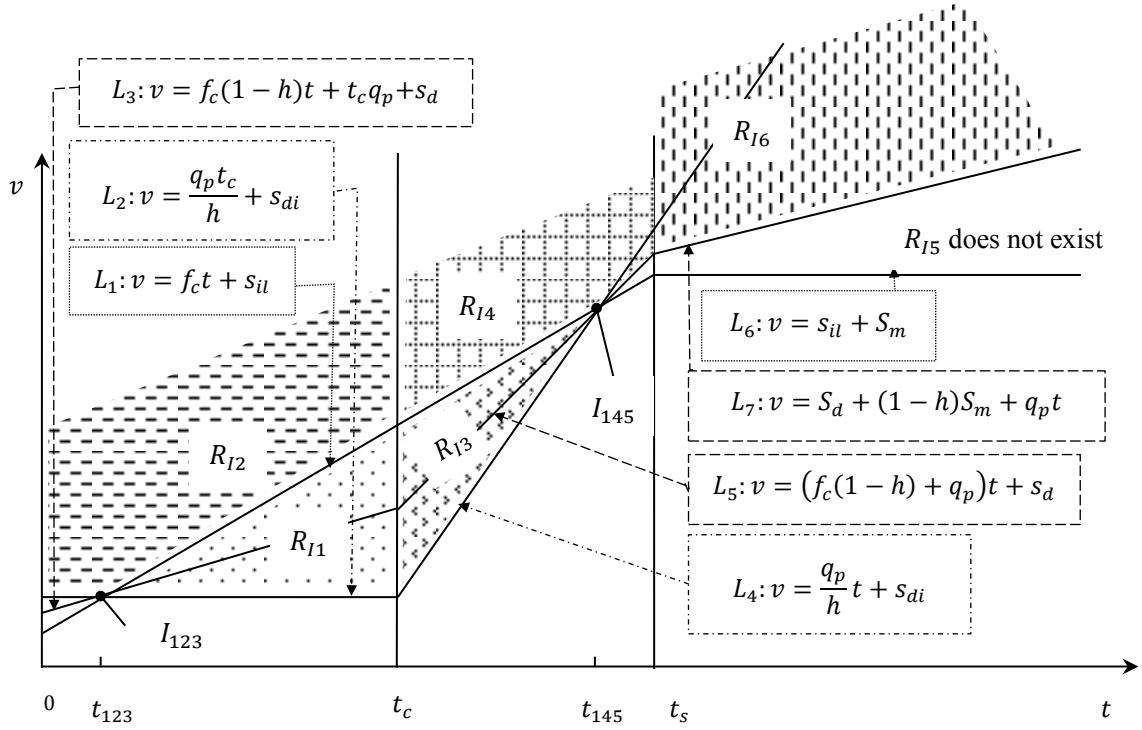

Fig. A.3. Region of integration for Type I_a catchments with

$$\frac{hs_{dd}}{t_S} + f_c h < q_p \leq \frac{hs_{dd}}{t_c} + hf_c$$

and Type I_b catchments with $\frac{hs_{dd}}{t_c} < q_p \leq \frac{hs_{dd}}{t_c} + hf_c$

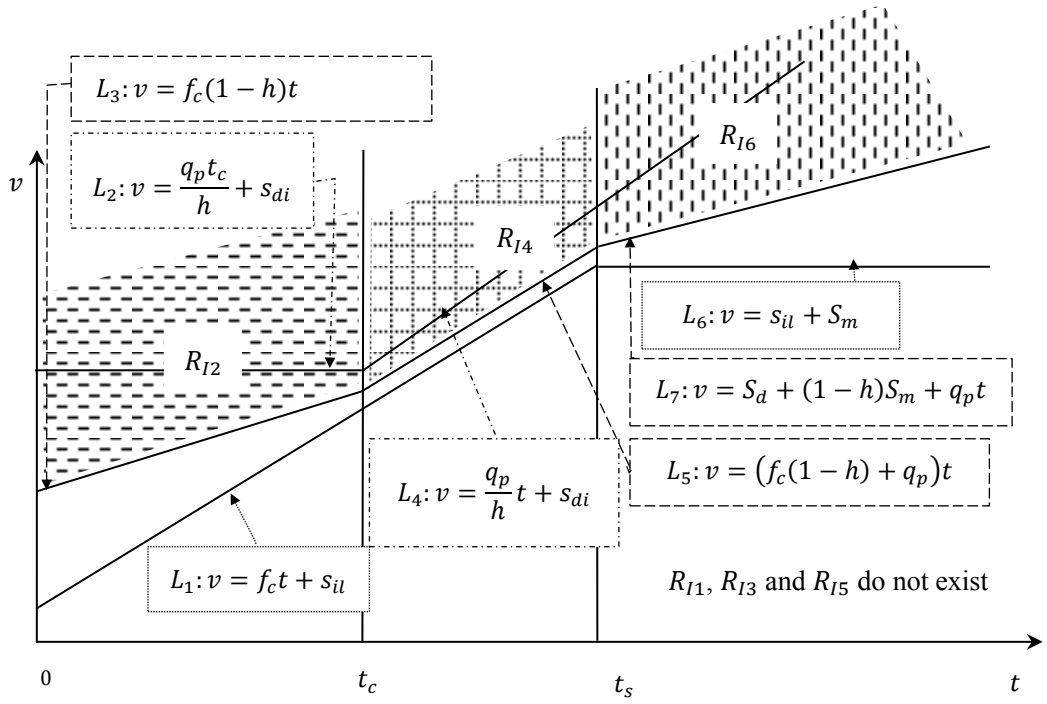

Fig. A.4. Region of integration for Type I_a catchments with $q_p > \frac{hs_{dd}}{t_c} + hf_c$ and Type I_b catchments with $q_p > \frac{hs_{dd}}{t_c} + hf_c$

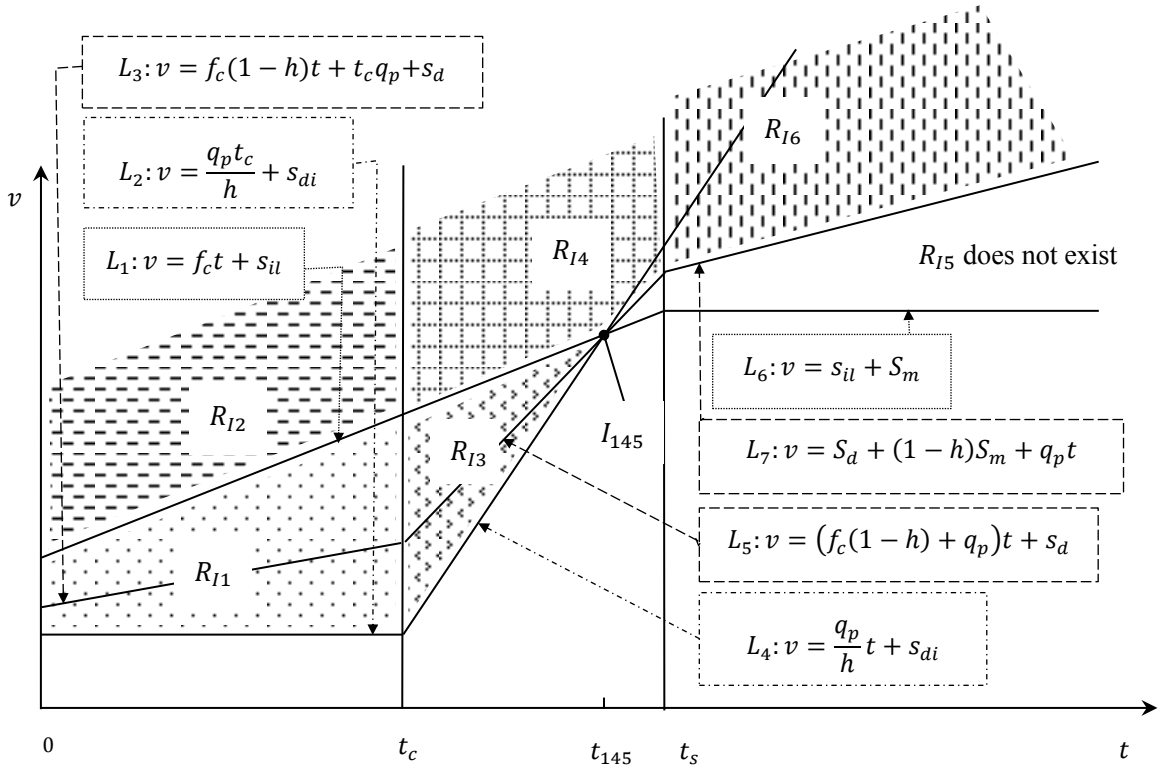

Fig. A.5. Region of integration for Type I_b catchments with

$$\frac{hs_{dd}}{t_S} + hf_c < q_p \leq \frac{hs_{dd}}{t_c}$$

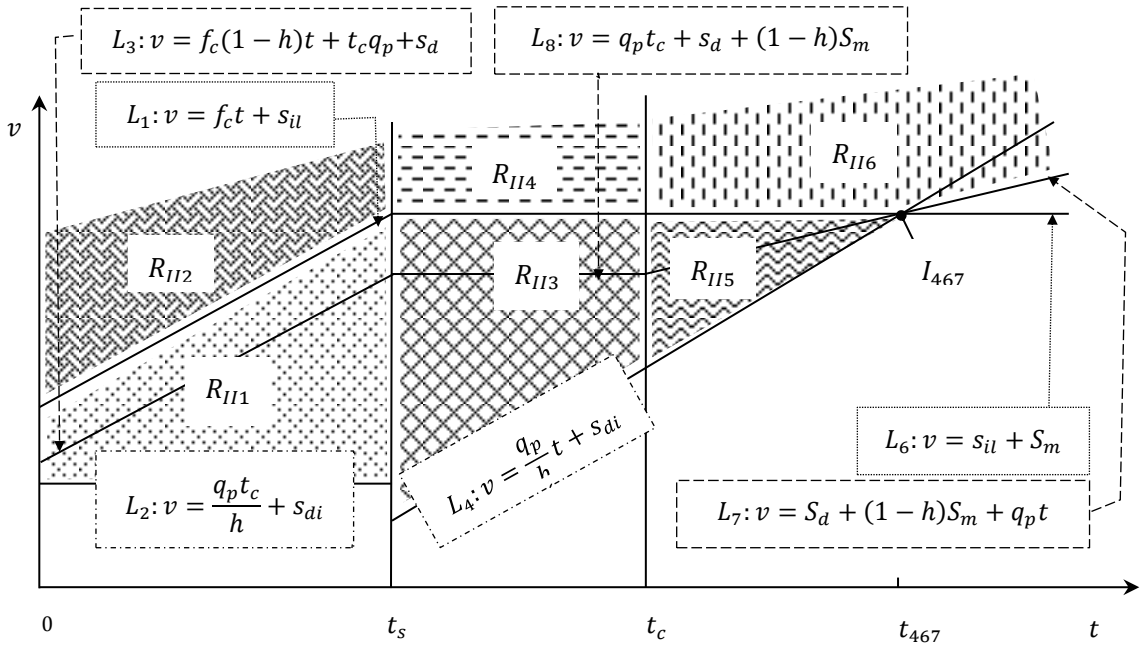

Fig. A.6. Region of integration for Type II catchments with $0 < q_p \leq \frac{hs_{dd}}{t_c}$

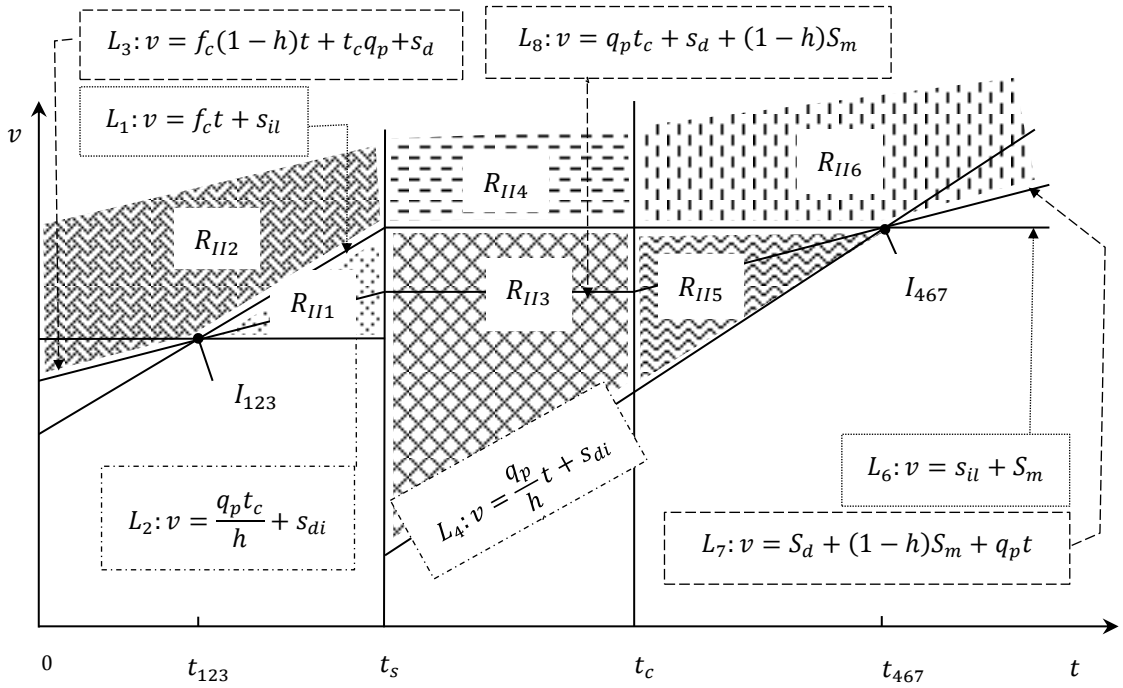

Fig. A.7. Region of integration for Type II catchments with $\frac{hs_{dd}}{t_c} < q_p \leq \frac{h(s_{dd}+S_m)}{t_c}$

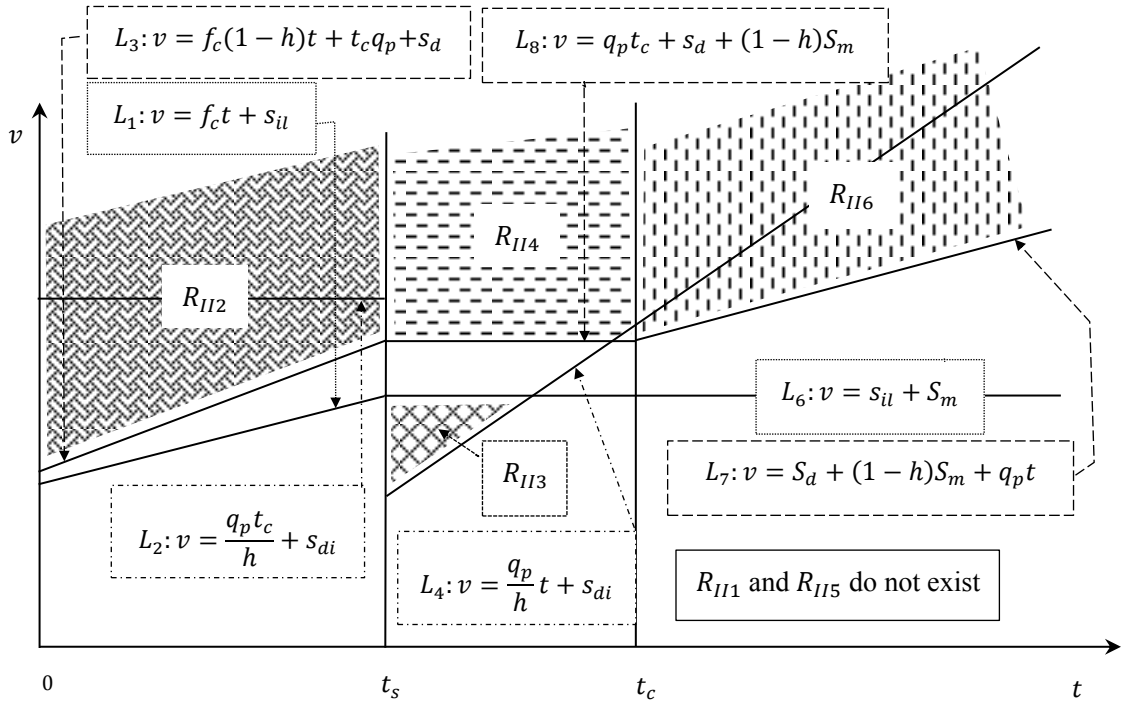

Fig. A.8. Region of integration for Type II catchments with $\frac{h(s_{dd}+S_m)}{t_c} < q_p \leq \frac{h(s_{dd}+S_m)}{t_s}$

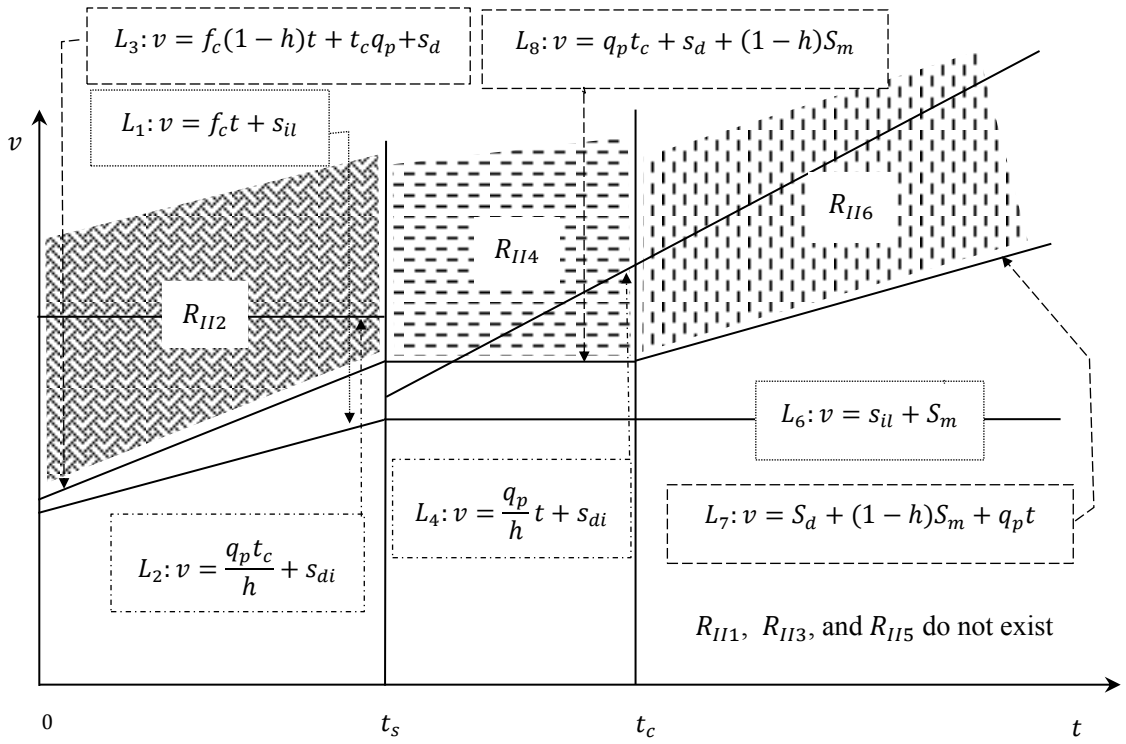

Fig. A.9. Region of integration for Type II catchments with $q_p > \frac{h(s_{dd}+S_m)}{t_s}$

CHAPTER 5

Overall Summary and Recommendations for Future Research

5.1. Overall Summary

The analytical probabilistic approach was first developed to investigate problems in the field of water resources in the late seventies, since then it evolved rapidly particularly for urban stormwater management purposes. Due to their use of closed-form mathematical equations, analytical probabilistic models are computationally more efficient than numerical models. These models are usually verified and found to produce results comparable to either the continuous simulation or design storm approach; nevertheless, the analytical probabilistic models are not yet widely used in practice. In order to promote this promising approach, further expansions and improvements are crucial.

In chapter 2, a procedure was proposed to test the exponentiality of a location's rainfall event characteristics. The event-based rainfall frequency analysis starts by selecting an inter-event time definition (IETD), which is the minimum dry time

between events, to be used as the criterion to separate long continuous rainfall records into individual events. This results in a large sample of rainfall events, each event is characterized by its volume v, duration t , and interevent b. Then a Poisson test can be applied to test the independence of these events and the exponentiality of inter-event time b. Chi-square goodness-of-fit tests can be employed to test the exponentiality of rainfall event volume v and duration t. Small rainfall events with volumes less than or equal to a volume threshold v_t that have negligible hydrological effects can be removed to improve the goodness-of-fit. An IETD of 6-12 h and volume threshold $v_t \leq 5$ mm for small urban catchments are reasonable. The most suitable values of IETD and v_t may depend on the specific location's climatic conditions. Goodness-of-fit tests of large samples are challenging and may be misleading if not applied with caution thus engineering common sense should also be jointly considered. The resulting distribution models for rainfall event characteristics can be used as input to the analytical probabilistic models developed for stormwater management planning and design purposes (Adams and Papa 2000); similar to the use of the entire set of rainfall data with continuous simulation models or individual design storms with the design storm approach.

As mentioned earlier, APSWM is an event-based approach where rainfall data is treated as individual events; each event has its own characteristics. If its volume exceeds the losses, a rainfall event produces a runoff event characterized by runoff duration, volume, and peak discharge rate. The runoff event duration is equivalent to

the corresponding rainfall event duration plus the time for a drop of rain fallen on the hydraulically farthest point of the catchment to reach the catchment's outlet; known as the catchment's time of concentration. The runoff event volume is a direct estimate of its corresponding rainfall event volume less all the losses. However the runoff peak discharge rate estimation depends on the hydrograph shape in addition to the runoff event volume and duration. Thus further assumption about the hydrograph shape is required. In the past, in order to facilitate the derivations of the peak discharge pdfs in APSWM, the hydrograph shape was assumed to be triangular. In chapter 3 three different hydrograph shapes are considered for the development of new peak discharge rate pdfs. The three possible shapes of runoff event hydrographs are two trapezoids with different upper bases and a triangle, depending on the magnitude of the rainfall event duration as compared to that of the catchment's time of concentration. The newly developed models are compared to their preceding models developed with triangular hydrographs only. The HEC-HMS design-storm with SCS and Clark's unit hydrographs were used as references. The new models produced better results than the models with triangular hydrographs as compared to design storm results.

In chapter 4, new pdfs of peak discharge rate were derived considering the different hydrograph shapes as discussed in chapter 3 and also the saturation excess runoff volume. The newly developed models were referred to as $APSWM_{tis}$ which is an acronym for analytical probabilistic storm water models which uses trapezoidal

hydrographs and includes both infiltration and saturation excess volumes into runoff volume calculations. $APSWM_{tis}$ was applied to the seven stations analyzed in chapter 2, results comparable to SWMM with design storm inputs were found. The differences in results between the two approaches are usually less than 20%.

5.2. Recommendations for Future Research

The analytical probabilistic approach has been investigated and occasionally applied in urban stormwater management for more than three decades, many expansions and improvements were made since its first appearance. However, there are still areas for the model to expand and improve. Starting from this thesis, first it would be great to have a detailed literature review of the analytical probabilistic models developed so far. This will facilitate an access to the approach's collective information for a better awareness. Second, a more accurate model for runoff volume frequencies is available in Appendix 1, which is ready to be tested and can be applied for the study of LIDs such as infiltration trenches; continuous simulations are required for comparison purposes because design storm results are not accurate for runoff volume estimates. This model also can be extended and used along with $APSWM_{tis}$, the model described in chapter 4, to test the functionality of LIDs.

$APSWM_{tis}$ can also be used to check the ability of the analytical probabilistic approach in detecting the effect of the regional climate differences on the pdfs of peak

discharge rate. In chapter 4, rainfall data of some stations in the Midwest region of the USA were used to test $APSWM_{tis}$ and compared against SWMM design-storm results. Some of those results are replotted here to get an idea about the capability of $APSWM_{tis}$ to detect/quantify the aforementioned regional differences (Fig. 1) as compared to the design storm approach (Fig. 2).

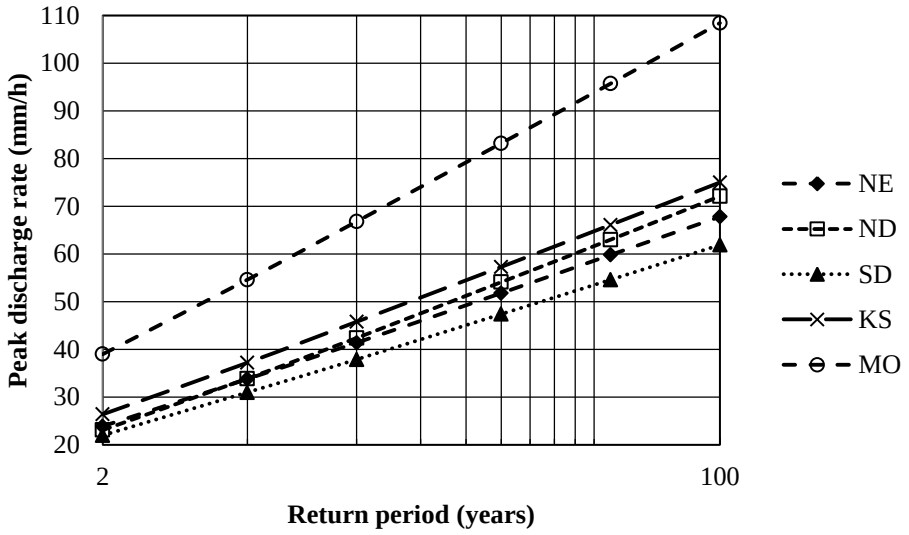

Fig. 1. Flood frequency results of the analytical modeling ($APSWM_{tis}$) with catchment under different climate conditions

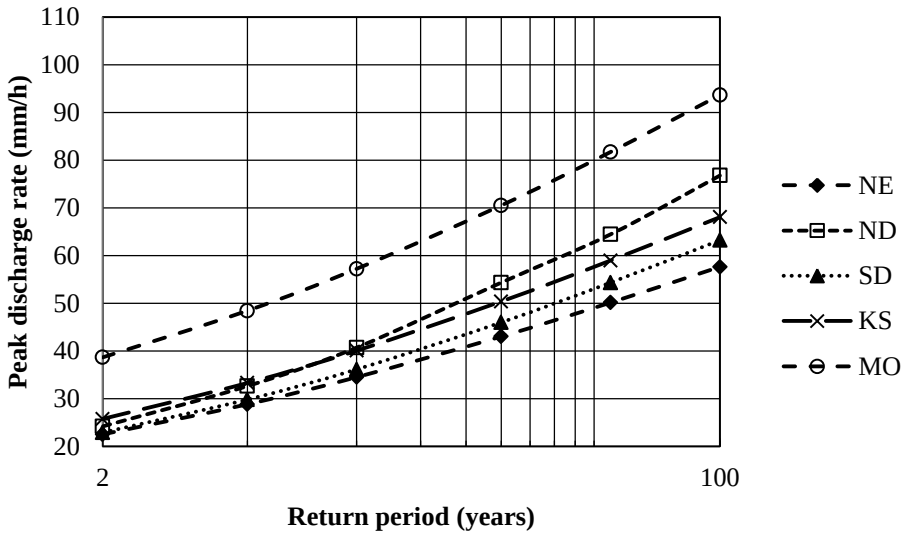

Fig. 2. Flood frequency results of SWMM design storm modeling with catchment under different climate conditions

Comparing the results summarized in Figs. 1 and 2, it can be seen that $APSWM_{tis}$ seems to be able to fairly detect/quantify the regional differences; however, more detailed further investigations are required. More stations covering a wider area may be included.

To simplify the development of APSWM, it was assumed that rainfall event characteristics (duration, volume and interevent time) are statistically independent. In fact, these characteristics are found to be statistically independent except for rainfall event volume and rainfall duration (Adams and Papa, 2000; and Rivera et al., 2005). The degree of dependency varies from one region to another (Rivera et al., 2005). It is expected that better results can be obtained if the dependence between rainfall event volume and duration is considered (Adams and Papa, 2000). Thus including the rainfall event volume-duration dependency for the development of alternative closed-

form analytical expressions would be very useful especially where rainfall event volume and duration are highly correlated.

For most of the developed analytical probabilistic models, the rainfall event characteristics are assumed to be exponentially distributed. This limits the use of the analytical approach to locations where rainfall event characteristics do follow exponential distributions. For example, for some rainfall stations in Italy, Bacchi et al. (2008) and Balistrocchi et al. (2009) found that Weibull distribution is more appropriate for rainfall event volume. In order to extend the use of the analytical probabilistic models, other distributions should also be investigated.

REFERENCES

Adams, B. J., and Howard C. D. D. (1986). "Design storm pathology." *Canadian water resources journal*, 11(3), 49-55.

Adams, B. J. and Papa, F. (2000). *Urban Stormwater Management Planning with Analytical Probabilistic Models*, John Wiley & Sons, Inc., New York, NY, USA.

Bacchi, B., Balistrocchi, M., and Grossi, G. (2008). "Proposal of a semi-probabilistic approach for storage facility design." *Urban Water Journal*, 5(3), 195-208.

Balistrocchi, M., Grossi, G., and Bacchi, B. (2009). "An analytical probabilistic model of the quality efficiency of a sewer tank." *Water Resources Research*, 45(12).

Behera, P. K., Adams, B. J., & Li, J. Y. (2006). "Runoff quality analysis of urban catchments with analytical probabilistic models." *Journal of Water Resources Planning and Management*, *132*(1), 4-14.

Benjamin, J. R., and Cornell C. A. (1970). *Probability, Statistics, and decision for civil engineers.* McGraw-Hill, New York, USA.

Chen, J., and Adams, B. J. (2005). "Urban storm water control evaluation with analytical probabilistic models." *Journal of water resources planning and management*, 131(5), 362-374.

Chen, J., and Adams, B. J. (2007). "Development of analytical models for estimation of urban stormwater runoff." *Journal of hydrology*, 336(3), 458-469.

Cheng, K., Wai, C., Cheng, Y., and Yeh, H. (2003). "Effect of spatial variation characteristics on contouring of design storm depth." *Hydrological processes*,

17(9), 1755-1769.

Eagleson, P. S. (1972). "Dynamics of flood frequency." *Water Resources Research*, 8(4), 878–898.

Guo, Y. (1998). *Development of analytical probabilistic urban stormwater models*. University of Toronto, Toronto, Canada.

Guo, Y. (2001). "Hydrologic design of urban flood control detention ponds." *Journal of Hydrologic Engineering*, 6(6), 472–479.

Guo, Y., and Adams, B. J. (1998a). "Hydrologic analysis of urban catchments with event-based probabilistic models: 1. Runoff volume." *Water Resources Research*, 34(12), 3421-3431.

Guo, Y., and Adams, B. J. (1998b). "Hydrologic analysis of urban catchments with event-based probabilistic models: 1. Peak discharge rate." *Water Resources Research*, 34(12), 3433-3443.

Guo, Y., and Adams, B. J. (1999a). "Analysis of detention ponds for storm water quality control." *Water Resources Research,* 35(8), 2447-2456.

Guo, Y., and Adams, B. J. (1999b). "An analytical probabilistic approach to sizing flood control detention facilities." *Water Resources Research,* 35(8), 2457-2468.

Guo, Y. and Zhuge, Z. (2008). "Analytical probabilistic flood routing for urban stormwater management purposes." *Canadian Journal of Civil Engineering*, 35(5): 487-499, https://doi.org/10.1139/L07-131.

Guo, Y., Zhang, S., and Liu, S. (2014). "Runoff reduction capabilities and irrigation requirements of green roofs." *Water resources management*, *28*(5), 1363-1378.

Hydrocom, Inc. (2007). Hydrologic Simulation models: An Overview, http://hydrocomp.com/simoverview.html.

Hromadka, T. V. (1997). "Balanced design storm UH, rational, and regression equation methods." *Journal of hydrologic engineering*, 2(3), 129-132.

Kottegoda, N.T. (1980). *Stochastic Water Resources Technology*, John Wiley & Sons, Inc. New York.

Levy, B., and McCuen, R. (1999). "Assessment of storm duration for hydrologic design." *Journal of hydrologic engineering*, 4(3), 209-213.

Linsley, R. K., Kohler, M. A., and Paulhus, J. L. H. (1982). *Hydrology for Engineers*, McGraw-Hill Book Company, New York, NY.

Marsalek, J., and Watt, W.E. (1984). "Design storms for urban drainage design." *Canadian Journal of Civil Engineering*, 11: 574-584.

McCuen, R. H. (2003). *Modeling hydrologic change: Statistical methods*, Lewis Publishers, Boca Raton.

Nnadi, F. N., Kline, F. X., Wary, H. L., and Wanielista, M. P. (1999). "Comparison of design storm concepts using continuous simulation with short duration storms." *Journal of the American Water Resources Association*, 35(1), 61-72.

Ontario Ministry of the Environment (2003). *Stormwater Management Planning and Design Manual*, Ontario, Canada.

Packman, J. C., and Kidd, C. H. R. (1980). "Logical approach to the design storm concept." *Water resources research*, 16(6), 994-1000.

Ponce, V. M. (1989). *Engineering Hydrology: Principles and Practices*, Prentice Hall, New Jersey.

Prodanovic, P., and Simonovic, S. P. (2004). CFCAS project, generation of synthetic design storms for the upper Thames River basin, Assessment of Water Resources Risk and Vulnerability to Changing Climatic Conditions.

Quader, A., and Guo, Y. (2006). "Peak discharge estimation using analytical probabilistic and design storm approaches." *Journal of hydrologic engineering*, 11(1), 46-54.

Rivera, P., Gironas J., Montt, J. P., and Fernandez, B. (2005). "An analytical model for hydrologic analysis in urban watersheds." *10 th International Conference on Urban Drainage, Copenhagen, Denmark, 21-26*.

SMHI-HBV-model (2007). http://www.smhi.se/foretag/m/hbv_demo/html/welcome.html

The U. S. Environmental Protection Agency (USEPA) (1986). *Methodology for analysis of detention basins for control of urban runoff quality*. Report EPA, 440, 5-87, Washington, D. C.

Urbonas, B. (1979). "Reliability of design storms in modeling." *Proceedings International Symposium On Urban Storm Runoff*, 23-27.

Vaes, G., Willems, P., and Berlamont, J. (2001). "Rainfall input requirements for hydrological calculations." *Urban Water*, 3(1-2), 107-112.

Watt, W. E., Chow, C. A., Hogg, W. D., and Lathem, K. W. (1986). "1-h urban design storm for Canada." *Canadian journal of civil engineering*, 13(3), 293-300.

Wurbs, R. A., and James, P. W. (2002). *Water resources engineering,* Prentice-Hall, Inc., USA, 2002.

Zhang, S., and Guo, Y. (2012a). "Analytical Probabilistic Model for Evaluating the Hydrologic Performance of Green Roofs." *Journal of Hydrologic Engineering*, 18(1), 19-28.

Zhang, S., and Guo, Y. (2012b). "Explicit Equation for Estimating Storm-Water Capture Efficiency of Rain Gardens." *Journal of Hydrologic Engineering*, 18(12), 1739-1748.

Zhang, S., & Guo, Y. (2014). "Stormwater capture efficiency of bioretention systems." *Water resources management*, *28*(1), 149-168.

APPENDIX 1: Derivation of runoff volume frequency distribution accounting the excess-saturation runoff

The estimation of runoff volume that results from a rainfall event considering saturation excess runoff is given in Chapter 4. The equation of runoff volume found in chapter 4 is rewritten here for an easier access

$$v_r = \begin{cases} 0, & v \le S_{di} \\ h(v - S_{di}), & S_{di} < v \le S_{il} + f_c t \ and \ t < t_s \\ v - S_d - f_c(1-h)t, & v > S_{il} + f_c t \ and \ t < t_s \\ h(v - S_{di}), & S_{di} < v < S_{il} + S_m \ and \ t \ge t_s \\ v - S_d - (1-h)S_m, & v \ge S_{il} + S_m \ and \ t \ge t_s \end{cases} \quad (1)$$

For any runoff volume v_0, the cumulative distribution function of an event runoff volume v_r valued at $v_0$referred to as $F_{v_r}(v_0)$ is the probability that v_r is less or equal to v_0 ($P[v_r \le v_0]$). Since v_r expression described in Eq. (1) is a piecewise equation, the corresponding $F_{v_r}(v_0)$ can be determined as the union of the sub-regions of integration defined by the specified intervals in Eq. (1). There are in total six sub-regions of integration; three sub-regions on the left of the vertical line defined by $t = t_s$ and referred to as L_s. The other three sub-regions are on the right of L_s.

The sub-region $R_1$is the area defined by

$$\begin{cases} 0 < v \leq S_{di} \\ t < t_s \end{cases}$$

Let L_1 be the line defined by $v = S_{di}$, R_1 is the region on the left of L_s, above the time axis, and below L_1.

The sub-region R_2 is defined by

$$\begin{cases} h(v - S_{di}) \leq v_0 \\ S_{di} < v \leq S_{il} + f_c t \\ t < t_s \end{cases} \xrightarrow{yields} \begin{cases} v \leq \frac{v_0}{h} + S_{di} \\ S_{di} < v \leq S_{il} + f_c t \\ t < t_s \end{cases}$$

Let L_2 and L_3 be the lines defined by $v = S_{il} + f_c t$ and $v = \frac{v_0}{h} + S_{di}$, respectively. R_2 is the area on the left of L_s, above L_1, and below L_2 and L_3.

The sub-region R_3 is defined by

$$\begin{cases} v - S_d - f_c(1-h)t \leq v_0 \\ v > S_{il} + f_c t \\ t < t_s \end{cases} \xrightarrow{yields} \begin{cases} v \leq S_d + v_0 + f_c(1-h)t \\ v > S_{il} + f_c t \\ t < t_s \end{cases}$$

Let L_4 be the line defined by $v = S_d + v_0 + f_c(1-h)t$. R_3 is the area on the left of L_s, above L_2, and below L_4.

The sub-region $R_4$is the area defined by

$$\begin{cases} 0 < v \leq S_{di} \\ t \geq t_s \end{cases}$$

R_4 is the region on the right of L_s, above the time axis, and below L_1.

The sub-region R_5 is defined by

$$\begin{cases} h(v - S_{di}) \le v_0 \\ S_{di} < v \le S_{il} + S_m \\ t \ge t_s \end{cases} \xrightarrow{yields} \begin{cases} v \le \dfrac{v_0}{h} + S_{di} \\ S_{di} < v \le S_{il} + S_m \\ t \ge t_s \end{cases}$$

Let L_5 be the line defined by $v = S_{il} + S_m$. R_5 is the area on the right of L_s, above L_1, and below L_3 and L_5.

The sub-region R_6 is defined by

$$\begin{cases} v - S_d - (1-h)S_m \le v_0 \\ v > S_{il} + S_m \\ t \ge t_s \end{cases} \xrightarrow{yields} \begin{cases} v \le S_d + v_0 + (1-h)S_m \\ v > S_{il} + S_m \\ t \ge t_s \end{cases}$$

Let L_6 be the line defined by $v = S_d + v_0 + (1-h)S_m$. R_6 is the area on the right of L_s, above L_5, and below L_6.

The areas of the six sub-regions of integrations R_1 through R_6 depend on the relative positions of the above defined lines L_1 through L_6 and the vertical line L_s. The lines positions are determined based on their corresponding intercepts and slopes. The sub-regions R_1 through R_3 are on the left of L_s and delineated by the segments of lines L_1 through L_4 situated between the volume axis and L_s. Thus the delineation of the sub-regions R_1 through R_3 can be achieved by determining the relative magnitudes of L_1 through L_4 slopes and intercepts at volume axis and L_s. For $j = 1$ to 4, let I_j and I_{js} be the intercepts of lines L_j at the volume axis and the vertical line L_s, respectively. The lines L_1 and L_3 are horizontal lines, consequently, $I_{1s} = I_1$ and $I_{3s} = I_3$. Let S_2 and S_4 be the slopes of the lines L_2 and L_4, respectively; $S_2 = f_c$ and $S_4 = (1-h)f_c$, thus S_2 is always greater or equal to $S_4 (S_2 \ge S_4)$.

The sub-regions R_4 through R_6 are on the right of L_s and delineated by the lines L_1, L_2, L_5, and L_6. The lines L_1, L_3, L_5, and L_6 are horizontal lines, where the intercepts of the lines L_5 and L_6 are the same as those of the lines L_2 and L_4, respectively. Thus the delineation of the sub-regions R_4 through R_6 depends on the relative magnitudes of the intercepts I_{1s}, I_{2s}, I_{3s}, and I_{4s}.

The relative magnitudes of the intercepts of lines L_1 through L_6 depend on the values of v_0, as shown in Table 1, there are three $v_0$intervals. The six sub-regions of integrations R_1 through R_6 corresponding to Intervals 1, 2, and 3 are plotted in Figs. 1, 2, and 3, respectively. The lines do not intersect except for the Interval 2 of v_0, where the lines L_2, L_3, and L_4 intersect at common point referred to as I_{234} with coordinates $(t_{234} = \frac{v_0 - hS_{dd}}{hf_c}\ , v_{234} = \frac{v_0 + hS_{di}}{h})$.

For Interval 1 ($v_0 \leq hS_{dd}$) the regions R_3 and R_6 do not exist and the union of the other sub-regions is the area between the time axis and L_3 (Fig. 1.). Thus

$$P[v_r \leq 0] = \int_0^{\frac{v_0}{h}+S_{di}} \int_0^{\infty} \zeta e^{-\zeta v} \lambda e^{-\lambda t} dt\, dv = 1 - e^{-\zeta(\frac{v_0}{h}+S_{di})}$$

For Interval 2 ($hS_{dd} < v_0 \leq h(S_{dd} + S_m)$) the region R_6 does not exist and the union of the other sub-regions is the area below L_3 and L_4 (Fig. 2.). Thus

$$P[v_r \leq 0] = \int_0^{t_{234}} \int_0^{S_d + v_0 + f_c(1-h)t} \lambda e^{-\lambda t} \zeta e^{-\zeta v} dv\, dt + \int_{t_{234}}^{\infty} \int_0^{\frac{v_0}{h}+S_{di}} \lambda e^{-\lambda t} \zeta e^{-\zeta v} dv\, dt$$

$$= 1 - \frac{\lambda}{\lambda + \zeta f_c(1-h)} \exp(-\zeta S_d - \zeta v_0) - \frac{\zeta f_c(1-h)}{\lambda + \zeta f_c(1-h)} \exp\left(-\zeta S_{di} - \frac{\zeta v_0}{h} - \frac{\lambda(v_0 - hS_{dd})}{hf_c}\right)$$

For Interval 3 ($v_0 > h(S_{dd} + S_m)$) the union of the sub-regions is the area below L_4 and L_6 (Fig. 3.). Thus

$$P[v_r \le 0] = \int_0^{t_s} \int_0^{S_d + v_0 + f_c(1-h)t} \lambda e^{-\lambda t} \zeta e^{-\zeta v} dv\, dt + \int_{t_s}^{\infty} \int_0^{S_d + v_0 + (1-h)S_m} \lambda e^{-\lambda t} \zeta e^{-\zeta v} dv\, dt$$

$$= 1 - \frac{\lambda}{\lambda + \zeta f_c(1-h)} \exp(-\zeta S_d - \zeta v_0) - \frac{\zeta f_c(1-h)}{\lambda + \zeta f_c(1-h)} \exp(-\zeta(1-h)S_m - \lambda t_s)$$

Let $C = \frac{\lambda}{\lambda + \zeta f_c(1-h)}$, in summary, the CDF of v_r is

$$P[v_r \le v_0] = \begin{cases} 1 - \exp(-\zeta S_{di} - \frac{\zeta v_0}{h}), & v_0 \le hS_{dd} \\ 1 - (1-C)\exp\left(-\zeta S_{di} - \frac{\zeta v_0}{h} - \frac{\lambda(v_0 - hS_{dd})}{hf_c}\right) - C\exp(-\zeta S_d - \zeta v_0), & S_{dd} < v_0 \le h(S_{dd} + S_m) \\ 1 - (1-C)\exp(-\zeta(1-h)S_m - \lambda t_s) - C\exp(-\zeta S_d - \zeta v_0), & v_0 > h(S_{dd} + S_m) \end{cases} \quad (2)$$

Consequently, the exceedance probability models of runoff volume, considering the saturation excess amount, are as follows:

$$P[v_r > v_0]$$

$$= \begin{cases} \exp(-\zeta S_{di} - \dfrac{\zeta v_0}{h}), & v_0 \le hS_{dd} \\ (1-C)\exp\left(-\zeta S_{di} - \dfrac{\zeta v_0}{h} - \dfrac{\lambda(v_0 - hS_{dd})}{hf_c}\right) + & \\ \quad C\exp(-\zeta S_d - \zeta v_0), & S_{dd} < v_0 \le h(S_{dd} + S_m) \\ (1-C)\exp(-\zeta(1-h)S_m - \lambda t_s) + & \\ \quad C\exp(-\zeta S_d - \zeta v_0), & v_0 > h(S_{dd} + S_m) \end{cases} \tag{3}$$

Table 1. The relative magnitudes of the intercepts of lines L_1 through L_6

Interval 1	Interval 2	Interval 3
$v_0 \le hS_{dd}$	$hS_{dd} < v_0 \le h(S_{dd} + S_m)$	$v_0 > h(S_{dd} + S_m)$
$I_2 \ge I_4 \ge I_3$	$I_3 > I_4 > I_2$	$I_3 > I_4 > I_2$
$I_{2s} \ge I_{4s} \ge I_{3s}$	$I_{2s} \ge I_{4s} \ge I_{3s}$	$I_{3s} > I_{4s} > I_{2s}$

Figure Captions

Fig. 1. Region of integration when $v_0 \leq hS_{dd}$

Fig. 2. Region of integration when $hS_{dd} < v_0 \leq h(S_{dd} + S_m)$

Fig. 3. Region of integration when $v_0 > h(S_{dd} + S_m)$

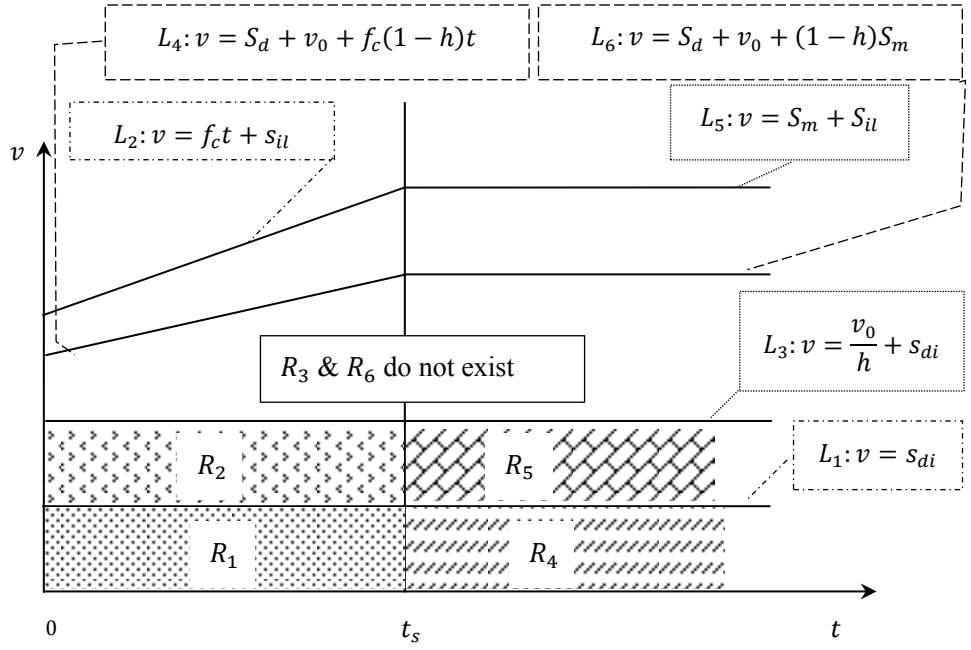

Fig.1. Region of integration when $v_0 \leq hS_{dd}$

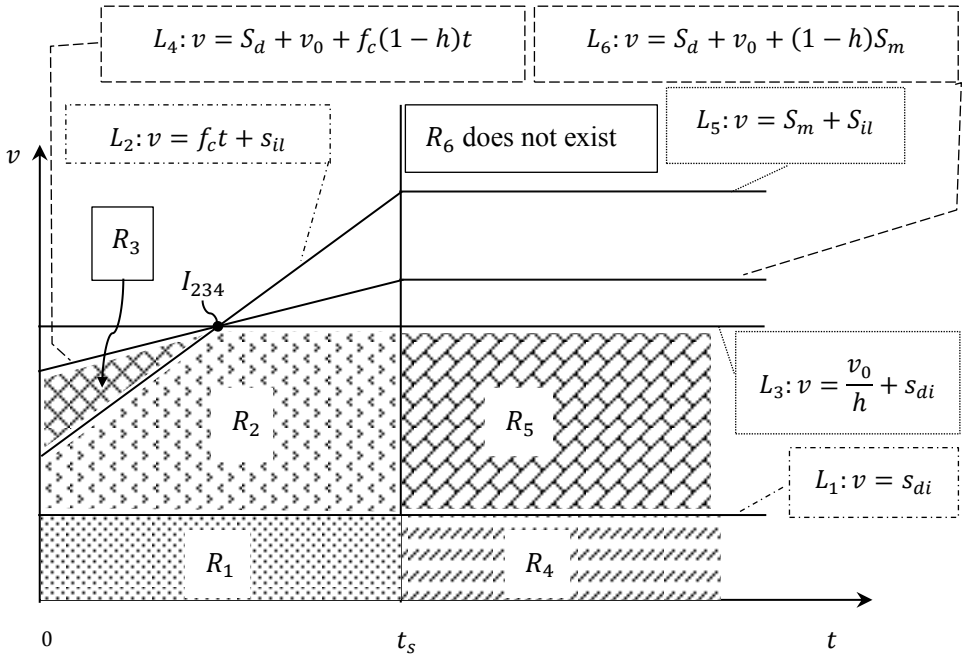

Fig.2. Region of integration when $hS_{dd} < v_0 \leq h(S_{dd} + S_m)$

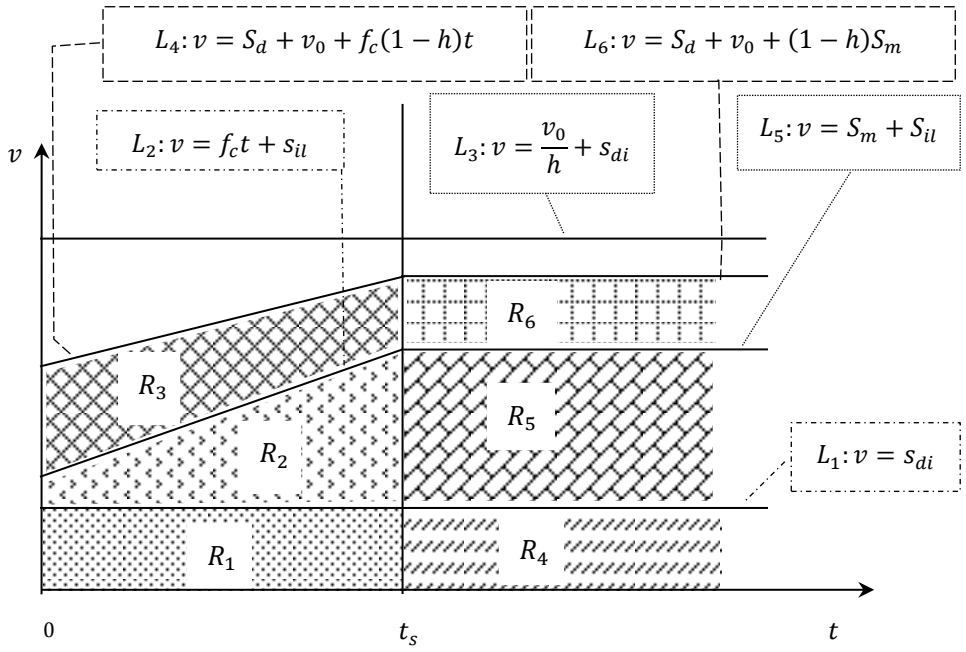

Fig.3. Region of integration when $v_0 > h(S_{dd} + S_m)$