AN URBAN RAINFALL STORM FLOOD SEVERITY INDEX

Prepared by:

Erik Jobin

Submitted under supervision of:

Dr. Ioan Nistor

In partial fulfillment of the requirements for the
Master of Applied Science in Civil Engineering degree

Department of Civil Engineering

University of Ottawa

May, 2013

Acknowledgements

I would like to thank my thesis supervisor Nistor for his help, trust, guidance and, suggestions throughout my Masters program.

I would like to thank my parents for their patience, encouragement and considerable help throughout this study and my Masters program.

I would like to thank my brother for his help in formatting media files.

I would like to thank Benoît Pagé for his input, company and for reviewing my thesis work.

I would like to thank Vincent Gabriel-Vallée for his help in producing the polygon storm dataset and providing programming guidance.

I would like to thank Kije Sipi Ltd and RadHyPS Inc for putting their resources at my disposal during the course of this study.

I would like to thank Sid Lodewyk for his input during the data gathering phase.

I would like to thank Hongliang Wang for providing data, maps, information as well as answering questions during the data gathering phase.

I would like to thank The City of Edmonton for allowing their Drainage Services branch to provide information crucial to this study.

Abstract

Extreme rainfall statistics are important for the design and management of the water resource infrastructure. The standard approach for extreme rainfall event severity assessment is the Intensity-Duration-Frequency (IDF) method. However, this approach does not consider the spatial context of rainfall and consequently does not properly describe rainfall storm severity, nor rarity.

This study provides a critical account of the current standard practice and presents an approach that takes into consideration both the spatial context of rainfall storms, and indirectly incorporates runoff to produce a representative approach to assessing urban rainfall storm severity in terms of flood potential. A stepwise regression analysis was performed on a dataset of individual rainfall storm characteristics to best represent documented basement floodings in the City of Edmonton. Finally, the urban rainfall storm flood severity index was shown to be most representative of the documented basement floodings' severity when compared to that of the IDF method.

Table of Contents

List of Tables

List of Figures

List of Symbols

a	=	Coefficient of the Z-R relation (non-dimensional)
b	=	Coefficient of the Z-R relation (non-dimensional)
c	=	Speed of light constant (299,792,458 m/s)
C	=	Weather radar constant (non-dimensional)
C_c	=	Retention correction (mm per rain measurement period)
d	=	Coefficient of the fitted IDF relation (non-dimensional)
D	=	Duration of the rainfall (hr)
D_a	=	Average rainfall duration (hr)
e	=	Coefficient of the fitted IDF relation (non-dimensional)
E_c	=	Evaporation in container/receiver (mm/hr)
f	=	Signal frequency (Hz)
f_d	=	Doppler frequency shift (Hz)
F_c	=	Funnel wetting correction (mm per rain measurement period)
F_e	=	Expected number of basement floodings (flooding(s))
F_r	=	Reported number of basement floodings (flooding(s))
g	=	Coefficient of the fitted IDF relation (non-dimensional)
G	=	Gain from the antenna system (dB)
h	=	Coefficient of the fitted IDF relation (non-dimensional)
H_0	=	Null hypothesis (non-numeric)
H_1	=	Alternative hypothesis (non-numeric)
I	=	Rainfall intensity (mm/hr)
I_a	=	Average rainfall intensity (mm/hr)
k	=	Number of variables (variable(s))
N	=	Number of scatterers per unit volume (non-dimensional)
p	=	Significance level (non-dimensional)
P_a	=	Average total rainfall (mm)
P_r	=	Signal strength (dBm)
P_t	=	Peak transmit power (watt)
r	=	Range to the target (m)
R	=	Rainfall intensity estimate (mm/hr)
R_a	=	Adjusted rainfall (mm)
R_m	=	Measured rainfall (mm)
R^2	=	Coefficient of determination (non-dimensional)
$\overline{R}^2$	=	Adjusted coefficient of determination (non-dimensional)
S_i	=	Rainfall storm flood severity index (non-dimensional)
t	=	Hours since last measurement (hr)
T	=	Return period of the rainfall (yr)
v_r	=	Radial velocity (m/s)
W_c	=	Wind correction factor (non-dimensional)
Z	=	Radar reflectivity (dBZ = mm^6/m^3)
Z_e	=	Effective radar reflectivity (dBZ = mm^6/m^3)
α	=	Significance level (0.05) (non-dimensional)

θ	=	Azimuth of the beam-width (degrees)
K	=	Kruskal-Wallis chi-squared test statistic (non-dimensional)
λ	=	Wavelength (m), $\lambda = c/f$
σ_i	=	Backscatter cross section (m^2)
τ	=	Radar pulse width (μs)
$\chi^2_{1-\alpha,(k-1)}$	=	Quantile of a chi-square distribution (non-dimensional)
Φ	=	Elevation of the beam-width (degrees)

Note that all units in this document are in the SI system unless otherwise stated.

List of Abbreviations

AMS	Annual Maximum Series
CAPPI	Constant Altitude Plan Position Indicator
EC	Environment Canada
ENSO	El Niño-Southern Oscillation
ERSPD	Edmonton Rainfall Storm Polygon Dataset
IDF	Intensity-Duration-Frequency
MJO	Madden-Julian Oscillation
MMS	Monthly Maxima Series
NAO	North Atlantic Oscillation
PE	'Pineapple Express'
PDO	Pacific Decadal Oscillation
POT	Peak-Over-Threshold
PPI	Plan Position Indicator
RERSD	Regional Edmonton Rainfall Storm Database
RP	Return Period
UTM	Universal Transverse Mercator

1 Introduction

Extreme rainfall statistics are important components for the design and management of the water resource infrastructure. The standard approach used to assess the severity of extreme rainfall events by the Canadian water resources community is the Intensity-Duration-Frequency (IDF) method. However, this approach most notably does not at all consider the spatial context of rainfall. Consequently, the IDF method does not properly describe individual rainfall storm's severity, rarity, and characteristics. Furthermore, it is the rainfall runoff that directly and indirectly impacts the operation of water resource infrastructures.

1.1 Study Objectives

The objectives of this research is to offer a critical account of the current standard practice and develop a more representative approach for assessing urban rainfall storm severity in terms of flooding potential. This will be achieved by using a better spatial description of recent regional rainfall tendencies in a form that is more convenient and intuitively understood than the currently used IDF statistical method.

1.2 Significance of Study

The significance of the study can be assessed in terms of three different perspectives: 1) scientific, 2) engineering and, 3) economical/social.

1.2.1 Scientific Perspective

The goal of scientific inquiry has always been to further understand natural phenomena in order to best describe, analyze, model and predict them. Unfortunately, extreme rainfall events are commonly being observed more frequently on a yearly basis, while the current IDF statistics suggest otherwise (Ramos *et al.*, 2006). These continued discrepancies in observations have rightfully taken credibility away from the IDF statistical method. Unfortunately, the Canadian

water resources community uses this standard approach when assessing the severity of extreme rainfall events without realizing the limitations and possible consequences of ignoring significant flaws, notably, that the IDF method does not at all consider the spatial context of rainfall. As a result, practioners overestimate the individual rainfall storm's severity, rarity and characteristics while underestimating the rainfall input design level. An approach that takes into consideration the spatial context of rainfall storms should yield a more representative assessment of its severity.

While rainfall data is readily available for water resources infrastructure design and analyses, it is the rainfall runoff that directly and indirectly impacts the operation of water resource infrastructure (CSA, 2010). However, since runoff statistics are obviously not available, rainfall data is used for design purposes. Nevertheless, this study attempts to indirectly incorporate runoff into the development process by using basement floodings as a measure of rainfall storm severity. These issues will be discussed in the following literature review (Section 2).

1.2.2 Practical Engineering Perspective

Extreme rainfall statistics are key components for the design and management of the water resource infrastructure (CSA, 2010). However, as previously mentioned, it is the rainfall runoff that directly and indirectly impacts the operation of water resource infrastructures (CSA, 2010). Therefore a more representative rainfall storm severity approach could be used to reliably communicate to the municipal water resources operators the severity of impending rainfall storms. When faced with extreme events, the performance of infrastructure often has a direct impact on deaths, injuries and, the magnitude of incurred damages (Mileti, 1999; CANHP, 2010). Using a more accurate assessment of severe rainfall storms relative to their flood potential during forecasts, planning and real-time storm tracking could help emergency response preparedness, improve security, help inform the public and allow residents to take necessary measures to minimize damages.

1.2.3 Economic and Social Perspectives

A correctly designed and operated water resource infrastructure effectively and safely directs runoff and storm sewage while providing peace of mind by offering a certain measure of protection from floods and thus allowing a region to prosper economically. Again, the proper description of actual local rainfall storms and their related occurrence is important for use in design and operational guidelines; otherwise, the risk of inadequate infrastructure capacity may either lead to: 1) increased flood risks, 2) decreased design life due to frequent operation under extreme and demanding conditions or, 3) the perceived misuse of public funds (CSA, 2010). Unfortunately, current design and operational practices fail on many fronts to properly assess the true rainfall event's magnitude and frequency of occurrence which results in more frequent floodings and operations at a higher hydraulic regime than designed.

Additionally, Gourley *et al.* (2010) stressed that representative rainfall descriptions are required while inaccuracies impact the well-being of the economy, not only in terms of flood prevention, but also in terms of normal operations in agriculture, hydroelectric power generation and drinking water supply.

The inaccurate description of rainfall event occurrences has also led to more frequent observations of events currently deemed as rare. This has had the effect of increasing the skepticism regarding water resource practices where too frequently claims of ‘Acts of God’ are being made in litigation. The criticism is geared towards what is felt as an exaggerated discourse of extreme weather that now frequently includes qualifiers such as ‘Storm-of-the-century’ and ‘Once-in-10,000-year events’ (Ungar, 1999). This only creates public confusion that stems from the inherently complex yet, flawed IDF return period (RP) method for characterizing significant rainfall events. Furthermore, the limited comprehension of risk and risk management associated with natural hazards come from the inability to grasp their complexity, statistics, resulting impacts and, effective protection and mitigation measures (Mileti, 1999). Consequently, the IDF method, also known as the

'extreme rainfall statistics' method is rightfully losing its credibility. This comes at a time when there's an increasing public demand for accurate information regarding extreme weather events (Ungar, 1999).

The majority of developed countries, including Canada, use hazard warning systems that combined with the application of building and construction standards contributed to a decrease in deaths (CANHP, 2010). In urbanized regions, where often limited flood protection options are feasible, flood watch systems could provide the most effective measure in minimizing losses due to floods (Vieux & Bedient, 2004). However, there is still a requirement of effectively communicating warnings to the public, community managers and decision makers, as well as a need to standardize the hazard warning systems (Mileti, 1999; Doswell, 2001). For example, Doswell (2001) contrasted that while tornados and hurricanes can easily be identified as threatening to one's well-being, severe rainfall conditions causing flash floods may not instill enough concern to generate the undertaking of appropriate measures. As a solution, public education campaigns can successfully impact the understanding and expectations of hazards (Sandink, 2007). Public relations and outreach regarding hazard risk and education that targets those potentially most affected should be a long-term endeavour. Furthermore, these programs offer access to a wealth of information from various reputable sources spanning different media formats while being respectful of the public's capacity to understand (Sandink, 2007). A key component that Sandink (2007) suggests for developing a 'Hazards Education Program' is that all information that is conveyed to the public should be concise, easy to understand and, include properly explained concepts such as RP. The proposed development of an alternative rainfall storm severity classification method would provide improved content to public education programs by providing more rigorous descriptions of extreme events, consistent with the recurrence rates.

The performance of the drainage infrastructure when subject to extreme events has a direct impact on deaths, injuries, the magnitude of damages incurred as well as

the ability of a community to cope with the disaster (Mileti, 1999). It is well known that significant rainfall runoff can overload the urban drainage infrastructure thereby causing surface floodings (CSA, 2010). Direct flood impacts can include road closures; infrastructure collapse, property and environmental damage including erosion and pollution. Furthermore, living through an extreme event and coping with the associated losses can also result in a stressful experience to those affected (Sandink, 2007) and can at the very least create a large scale inconvenience for the population at large.

The use of hazard warning systems as well as the beneficial changes to established construction and building standards has certainly contributed to a decrease in deaths in Canada; however, the on-going adverse economic impact from hazards has increased (CANHP, 2010). The insurance industry bears a significant portion of these costs; however, some hazards are typically not covered by insurance in Canada; for example, residential flood damages from surface water flow (IBC, 2006 in Sandink, 2007; IBC, 2009; CANHP, 2010). Furthermore, residential sewer backup loss protection from the insurance industry is only offered in specific communities (CANHP, 2010). In those cases, this protection is often only available as an optional rider to the standard policy (IBC, 2006 in Sandink, 2007; IBC, 2009; CANHP, 2010; IBC, 2011). Facing large payouts, insurance companies have also substantially increased the cost for such hazards (Sandink, 2007) or, altogether ceases offering the coverage in flood-prone areas, leaving residents to solely carry the entire risk (Sandink, 2007; CANHP, 2010).

The anticipated changes in climate and the north American population demographics including income gaps and urbanization trends are expected to increase the overall vulnerability of the population to hazard losses (Mileti, 1999; Rozalis et al., 2010). To clarify, with severe rainfall events as the root cause of many of sewer backup damages, it's probable that more frequent and severe sewer backups will occur as the climate changes (Sandink, 2007). Also, the state of the critical infrastructure has a direct impact on loss of life, injuries and damages

inflicted by extreme events (CANHP, 2010). Unfortunately, increased flood risks are expected as urbanization continues and is characterized by an increase in urban population and asset density (CANHP, 2010), as well as a deteriorating drainage infrastructure due to lack of proper funding for maintenance, upgrades and monitoring (Sandink, 2007). It may very well be that Canadian municipalities will eventually be unable to fiscally address the continued deterioration of the public drainage infrastructure with purely engineering approaches (Sandink, 2007). In light of this situation, many Canadian municipalities now turn to social approaches in addition to engineering means for basement flood risk reductions. These approaches include studies, education and activities that improves awareness as well as mitigative behavior alongside engineering strategies (Sandink, 2007). Considering the increasingly cost prohibitive nature of sewer network modifications, the seemingly erratic nature of extreme rainfall events and the anticipated worsening of rainfall event severity due to climate change, Sandink (2007) suggested to the cities of Edmonton and Toronto to get their residents more involved in sewer backup prevention. He notably suggested the use of education programs that provides detailed information on flood hazards stemming from several independent sources through various channels. This could include the use of a more intuitive index for characterizing the magnitude of rainfall storms that could be part of a modified standard of practice.

1.3 Study Innovation

Several innovations are used in this study, including: 1) use of spatio-temporal rainfall storm attribute data derived from adjusted weather radar data, 2) use of rainfall storms that are identified using hydro-meteorologic criteria rather than strictly meteorologic attributes, and 3) development of a rainfall storm severity classification method based on correlating spatial storm characteristics and observed flood occurrences.

1.4 Scope of Study

The work undertaken for this study mostly comprised of mapping and programming. Digitizing the study area's sub-region boundaries was required. Which radar rainfall storm data contributed to each respective sub-region was determined for storm characteristics computation. Relevant rainfall events to the study area from the Regional Edmonton Rainfall Storm Database (RERSD) needed to be identified. Identification of relevant events was done visually but first required programming of code for automated mapping of radar rainfall storm data. Once the relevant storms were identified, storm characteristics were computed for each respective sub-region.

In preparation for the stepwise regression analysis, the study-relevant storm set was separated in two: a calibration set and a validation set. The goal of programming of the stepwise regression analysis was to produce the model to present a measure of the potential urban rainfall storm flood severity. The regression analysis was performed on the data of sub-region that was most representative of the overall study storm dataset. The model that resulted from the analysis was tested against the overall study storm dataset for all sub-regions to assess the degree of success. Finally, the model was transformed into a potential flood severity index by curve fitting. The resulting storm severity index was compared to the IDF method with respect to predicting reported storm flood severity.

1.5 Outline/Organization

This work has 6 sections, which present the following content:

Section 1 introduces the study, presents its significance, describes its innovations, outlines its scope and presents its organization.

Section 2 presents an extensive literature review on the knowledge base of severe rainfall statistics.

Section 3 elaborates on the sources of data used in the study.

Section 4 presents the study's methodology, statistical analysis and formulation of the flood severity index.

Section 5 presents a comparison of the predictive ability of storm flood severity for the proposed index versus the IDF method. Also, a sensitivity analysis of the proposed storm flood severity index and discussion of its limitations and issues are presented.

Section 6 presents a brief conclusion of the work and mentions recommendations for future work.

2 Literature Review

A previous literature review (Jobin, 2010) performed as part of a directed study provided an opportunity to identify the advantages and disadvantages of existing classification methods of natural phenomena and, as a result, established target objectives for an alternative rainfall storm classification method. An expanded and comprehensive review of scientific articles, technical specifications, reports and other publications on subject matter of interest was eventually completed. Only pertinent and key findings from over five hundred (500+) documents are summarized and organized in the following sub-sections by topics.

This study provided further insight on issues regarding the currently preferred methods of identifying the rainfall storm severity. Particularly, advantages and limitations of current cutting-edge instruments and data analysis were researched in order to improve upon these techniques and methods. The most promising alternative classification methods were isolated and desirable aspects were considered for subsequent integration in the new rainfall storm classification method.

2.1 Introduction

Our country has expended a considerable amount of resources towards obtaining and maintaining its current water resource infrastructure (CSA, 2010). It is expected that it performs as intended in safeguarding lives and protecting property and the environment against damages from floods (CSA, 2010).

2.2 Meteorology

Water resources engineering is particularly interested in the atmospheric component of the hydrological cycle, particularly aspects that lead to, or influence precipitation. From a global perspective, precipitation and its distribution is closely associated with atmospheric circulations (Michaelides et al., 2009). Precipitation can replenish soil

water content and aquifers as well as supply water to natural surface water drainage which carves the landscape from its runoff (Michaelides et al., 2009).

2.2.1 Storms Types of Interest

A cloud is an observable agglomeration of water and/or ice particles located above ground (WMO, 2008a). Generally, clouds are produced by the upward movement of moist air through convection, orographic effects or, as a result of low-pressure areas and weather fronts (WMO, 2008a).

Recently, significant effort in hydrology has been directed towards producing better temporal and spatial descriptions of precipitation events (Bara et al., 2009). A particularly interesting research has identified precipitation storm cell types and scales through time and space (Bara et al., 2009). At the atmospheric level, multiple scales of meteorological events can occur (WMO, 2010). The WMO diagram below (Figure 2.1) illustrates what each scale represents, in terms of time and space with examples of expected meteorological events (WMO, 2010). Five scales are depicted: 1) microscale, which represents meteorological events smaller than 100 m; 2) toposcale are events between 100 m and 3 km in size; 3) mesoscale events of 3 km to 100 km in size; 4) large scale, 100 km to 3,000 km and; 5) planetary scale are events greater than 3,000 km in size (WMO, 2008a; WMO, 2010).

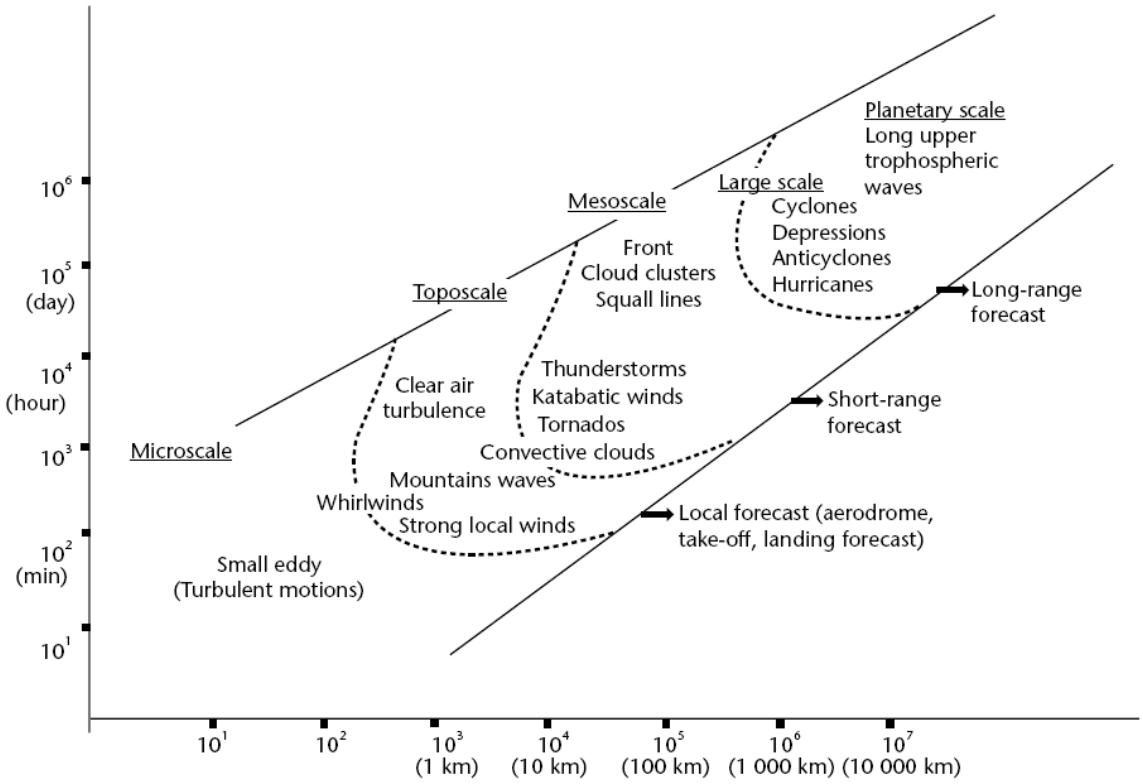


Figure 2.1 - Spatial and temporal scales of meteorolocial phenomena (WMO, 2010)

Rainfall clouds, aggregates of suspended water particles, are dynamic spatially and temporally heterogeneous entities that are influenced by the local atmospheric wind, temperature and pressure conditions but also induce natural occurring meteorological responses to its precipitation phenomenon, which in turn conspire to make it an ever changing entity. The very action of the cloud heterogeneously shedding water droplets or crystals accompanies localized physical responses such as a drop in atmospheric pressure and changes in temperature as well as impacting local wind patterns as water droplets precipitate towards the ground.

Severe short-lived convective rain storms often are the most spatially variable of all rainfall events (CSA, 2010). Their significant spatial variability make it hard to record this type of storm using the typically sparse networks of weather stations. A particular problem is the fact that the maximum storm rainfall over an area of interest is seldom, if ever, recorded by rain gauges (CSA, 2010).

Also, the type of patterns that multiple convective storm cells cluster together to form a mesoscale convective system have significant impacts on that region's flood response, particularly when it comes to relatively small urban catchments (Zhang & Smith, 2003). Furthermore, the dynamic and microphysical processes that occur in convective systems largely dictate the spatial and temporal variability of rainfall, and by extension is responsible for severe flash floods in small urban catchments (Zhang & Smith, 2003).

2.2.1.1 Flood Event Types

Floods in Canada are typically the result of either: 1) snowmelt, 2) prolonged rainfall on saturated soils, 3) intense localized rainfall or, 4) snowmelt accompanied with rainfall (CANHP, 2010). The flood severity is dictated by the rainfall intensity, rate of snowmelt and area topography (CANHP, 2010). Flood trends in Canada exhibit seasonality (CANHP, 2010). At the lower and mid-latitudes in Canada, snowmelt due to rainfall in January to March may produce floodings throughout warm winter periods (CANHP, 2010). However, approximately 40% of large-scale floods happen between April and May during southern Canada's spring snowmelt (CANHP, 2010). Nevertheless, severe rainfalls are the leading cause of floods from May to September (CANHP, 2010) and toward the end of summer and fall there are the occasional hurricanes and tropical storms that produce significant flood conditions in eastern Canada. The months of November and December typically have the least amounts of floods (CANHP, 2010).

The total rainfall depth, intensity and duration of a rainfall event influences the soil's capacity to cope with draining the rain water (CANHP, 2010). As the severity of rainfall events increase in terms of intensity and/or duration, the infiltration capacity of the soil decreases and therefore the runoff to rainfall ratio increases proportionately contributing to the rise in water level of connected water courses (CANHP, 2010). In relatively small catchments, roughly in the order of no more than a couple hundred square kilometres (Smith et al., 2007; Creutin et al., 2009; Borga et al., 2007 in Rusjan et al., 2009), severe rainfall can quickly produce significant

floodings referred to as flash floods (CANHP, 2010). Flash floods are generally the result of convective storm cells that produced a severe rainfall event with intense rainfall (Rozalis et al., 2010). A particular characteristic of flash floods is that flooding tends to occur in the immediate vicinity of the rainfall storm (Rozalis et al., 2010). Typically, such floods are deemed to have occurred if the worse flooding conditions are observed within the first few hours (first 6 hours according to (CANHP, 2010)) of the beginning of the rainfall (Smith et al., 2007; Creutin et al., 2009; Borga et al., 2007 in Rusjan et al., 2009). Areas with considerable topographical gradients tend to experience flash floods more frequently due to the high runoff flow rate that generate quick accumulations of water in channels (Creutin et al., 2009; CANHP, 2010). Another important causative factor of floods is the level of antecedent soil moisture content if the soil's saturated, it will produce more runoff that could overwhelm the drainage system and effectively creating a flash flood condition (Creutin et al., 2009). Urbanization also affects negatively the overall surface permeability and the interception attributes; consequently, relatively high water retention capabilities of what was once a natural environment, is effectively reduced and; consequently, results in more rainfall runoff that is ultimately conveyed more rapidly throughout the watershed (Creutin et al., 2009; Rozalis et al., 2010).

In an urbanized environment, the majority of floods arise from either surface water induced floodings or sewer backup (Sandink, 2007). Surface water induced floodings can take place when the municipal drainage infrastructure's conveyance capacity is exceeded (Kulkarni, 1999 in Sandink, 2007). Property damage can occur when the municipal drainage infrastructure is overwhelmed and the surplus water spills outside purpose-built channels and onto private property (Sandink, 2007). In urban settings, both stormwater and sanitary sewers can backup, and the outcome worsened if private property owners have direct stormwater drains connected to the infrastructure (Sandink, 2007). Sanitary sewers backup occurs when inflow and infiltration of storm sewer and ground water reduce the design capacity via breaches in the pipe network (UMA, 2005 in Sandink, 2007). It is noteworthy to retain that the majority of Canadian municipalities still have combined

storm and sanitary sewers (Pleau et al., 2005 in Sandink, 2007) even though they are slowly being replaced by separate systems. Combined sewer networks are particularly prone to flooding because of their direct connection to households via their sanitary service connections. As a result, when a particularly severe storm occurs, the system often exceeds its capacity and backs up into the connected buildings (Sandink, 2007).

Finally, Smith et al. (2007) summarized that rainfall which results in urban floods may present significant spatial and temporal variability.

2.2.2 Climate and Weather Patterns

The difference between weather and climate patterns can be defined as the day-to-day observation in the variability in meteorological parameters for the former (e.g. precipitation, temperature, winds, atmospheric pressure, relative humidity) versus the long-term variability of the same parameters over a broader area for the latter (CANHP, 2010).

There are well known natural climate and weather patterns that influence the Canadian rainfall distribution trends including extremes. These include: El Niño-Southern Oscillation (ENSO), North Atlantic Oscillation (NAO), Arctic Oscillation, Pacific Decadal Oscillation (PDO), Pacific North American Oscillation, 'Pineapple Express' (PE) and the Madden-Julian Oscillation (MJO) (CSA, 2010).

Many of the climate oscillations stem from the capacity of oceans to retain their temperature longer than can land and masses of air that allows significant temperature and air pressure differentials to be created (CSA, 2010). An understanding of these climate oscillations and associated weather patterns is relatively new (CSA, 2010). Atmospheric and oceanic oscillations play an important role in determining predominant hurricane paths as well as influence the degree to which a specific area will experience wet, dry, warm, or cold weather patterns for a given time period (CSA, 2010). Specifically, the ENSO cycles through phases

where El Niño occurs each 3 to 7 years for periods of 12 to 18 months then switches to a La Niña phase (CSA, 2010). Long-term trends of ENSO form the southern part of the PDO whose cycles last anywhere from 50 to 70 years and each of its positive and negative phase ((warm and dry, cool and wet weather respectively) lasts between 20 and 30 years (CSA, 2010).

The severity of each oscillation or weather pattern at any given time changes from one cycle to another. For example, El Niño events are most strongly felt during Canadian winters as well as early springs and notably more often in the western parts of Canada (Shabbar et al., 1997 in CSA, 2010). However, nearly half the time, no El Niño or La Niña effects are felt at all (CSA, 2010).

The ENSO and PDO are significantly connected with the atmospheric air mass circulations over the North American continent as well as over the North Pacific sea (CSA, 2010). Warm phases of the PDO were linked with a higher frequency of severely intense precipitation events over the Greater Vancouver Region (CSA, 2010).

The PE refers to a one to two month winter weather anomaly that brings about warm and moist air that induces significant rainfall and floods to the North American Pacific coast and even reaches inland as far as the British Columbia interior (CSA, 2010). Favourable conditions for the formation of PE events are created by a 30 to 60-day atmospheric phenomenon referred to as the MJO, which mostly influences the tropical and sub-tropical regions of the globe but also circulations over the North Pacific (CSA, 2010).

At times, the NAO and the Arctic Oscillation are considered different descriptions of the same oscillation (CSA, 2010). The NAO has an important impact on weather and climate patterns, particularly significant rainfall trends in eastern Canada (CSA, 2010). Both the Arctic and NAO cycles change anywhere from a weekly to decadal time scale (CSA, 2010). These cycles result in either severe cold temperatures in

central North America, or in contrast, warm winter extremes with more wet air masses near the Gulf of Mexico with significantly increased probability of convective rainfall events (CSA, 2010).

Other naturally occurring phenomena may impact the hydrological cycle to a significant degree. Such phenomena includes large volcanic discharges of ash and gas in the atmosphere (WMO, 2009).

2.3 Precipitation Measurements

A measurement is an attempt to quantify a characteristic with an instrument. However, constraints are always present when performing observations (e.g. the instrument's physical limitations with regards to accuracy). As a result, precipitation measurements always remain an interpretation of the true precipitation values (WMO, 2008a; WMO, 2008b). Precipitation measurements are not instantaneous observations but averages over set finite timeframes and over a given area and elevation.

Note that precipitation is a meteorological phenomenon that includes water meteors in all three physical phases including rain, drizzle, freezing rain, freezing drizzle, hail, pellets and snow (EC, 2010a).

Rainfall measurements are vital for water resources planning and design such as reservoir and storm sewer design, flood control, forecasting, and weather radar observation adjustments (Devine & Mekis, 2008). These observations are typically recorded as total daily amounts or precipitation intensities. A WMO (2008a) reference document has established expected operational measurement uncertainty and instrument performance as shown in Table 2.1 along with a listing of typical characteristics.

Table 2.1 - Expected operational measurement uncertainty, instrument performance and typical characteristics (WMO, 2008a)

	(1) Variable	(2) Range	(3) Reported resolution	(4) Mode of measurement/ observation	(5) Required measurement uncertainty	(6) Sensor time constant	(7) Output averaging time	(8) Achievable measurement uncertainty	(9) Remarks
6.	**Precipitation**								
6.1	Amount (daily)	0 – 500 mm	0.1 mm	T	0.1 mm for ≤ 5 mm 2% for > 5 mm	n/a	n/a	The larger of 5% or 0.1 mm	Quantity based on daily amounts Measurement uncertainty depends on aerodynamic collection efficiency of gauges and evaporation losses in heated gauges
6.4	Precipitation intensity	0.02 mm h^{-1} – 2 000 mm h^{-1}	0.1 mm h^{-1}	I	(trace): n/a for 0.02 – 0.2 mm h^{-1} 0.1 mm h^{-1} for 0.2 – 2 mm h^{-1} 5% for > 2 mm h^{-1}	< 30 s	1 min		Uncertainty values for liquid precipitation only Uncertainty is seriously affected by wind Sensors may show significant non-linear behaviour For < 0.2 mm h^{-1}: detection only (yes/no) sensor time constant is significantly affected during solid precipitation using catchment type of gauges

Notes:
1. Column 1 gives the basic variable.
2. Column 2 gives the common range for most variables; limits depend on local climatological conditions.
3. Column 3 gives the most stringent resolution as determined by the *Manual on Codes* (WMO-No. 306).
4. In column 4:
   I = Instantaneous: In order to exclude the natural small-scale variability and the noise, an average value over a period of 1 min is considered as a minimum and most suitable; averages over periods of up to 10 min are acceptable.
   T: = Totals: Totals over a fixed period, as specified by coding requirements.
5. Column 5 gives the recommended measurement uncertainty requirements for general operational use, i.e. of Level II data according to FM 12, 13, 14, 15 and its BUFR equivalents. They have been adopted by all eight technical commissions and are applicable for synoptic, aeronautical, agricultural and marine meteorology, hydrology, climatology, etc. These requirements are applicable for both manned and automatic weather stations as defined in the *Manual on the Global Observing System* (WMO-No. 544). Individual applications may have less stringent requirements. The stated value of required measurement uncertainty represents the uncertainty of the reported value with respect to the true value and indicates the interval in which the true value lies with a stated probability. The recommended probability level is 95 per cent (k = 2), which corresponds to the 2 σ level for a normal (Gaussian) distribution of the variable. The assumption that all known corrections are taken into account implies that the errors in reported values will have a mean value (or bias) close to zero. Any residual bias should be small compared with the stated measurement uncertainty requirement. The true value is the value which, under operational conditions, perfectly characterizes the variable to be measured/observed over the representative time interval, area and/or volume required, taking into account siting and exposure.
6. Columns 2 to 5 refer to the requirements established by the CBS Expert Team on Requirements for Data from Automatic Weather Stations in 2004.
7. Columns 6 to 8 refer to the typical operational performance established by the CIMO Expert Team on Surface Technology and Measurement Techniques in 2004.
8. Achievable measurement uncertainty (column 8) is based on sensor performance under nominal and recommended exposure that can be achieved in operational practice. It should be regarded as a practical aid to users in defining achievable and affordable requirements.
9. n/a = not applicable.
10. The term *uncertainty* has preference over *accuracy* (i.e. uncertainty is in accordance with ISO standards on the uncertainty of measurements (ISO, 1995)).
11. Dewpoint temperature, relative humidity and air temperature are linked, and thus their uncertainties are linked. When averaging, preference is given to absolute humidity as the principal variable.

Typically, precipitation is expressed as a depth in millimetres and actually represents a volume over a standard area (WMO, 2008a). When precipitation depths smaller than 0.1 mm (or 0.2 mm) are detected, these observations are usually labelled as a trace amounts (WMO, 2008a).

Rain gauge instruments measure precipitation directly and typically near the surface while weather radars and satellites indirectly estimate precipitation at a distance well above the ground. Although both types of instrument measure precipitation, they have significantly different approaches to observing and quantizing rainfall. Since rainfall measurement is a key component of the study, further assessments of these technologies will be carried out and presented in subsequent sections.

2.3.1 Measurement Error

Measurement uncertainty can be described probabilistically, as illustrated in Figure 2.2, where the breadth of the confidence interval can also be referred to as the error band (WMO, 2008b). Although hydrological measurements usually deviate from the underlying assumption of independent random variables from set statistical distributions (e.g. stream stage data correlate with previous values) (WMO, 2008b). WMO (2008b) does not consider it a significant issue,

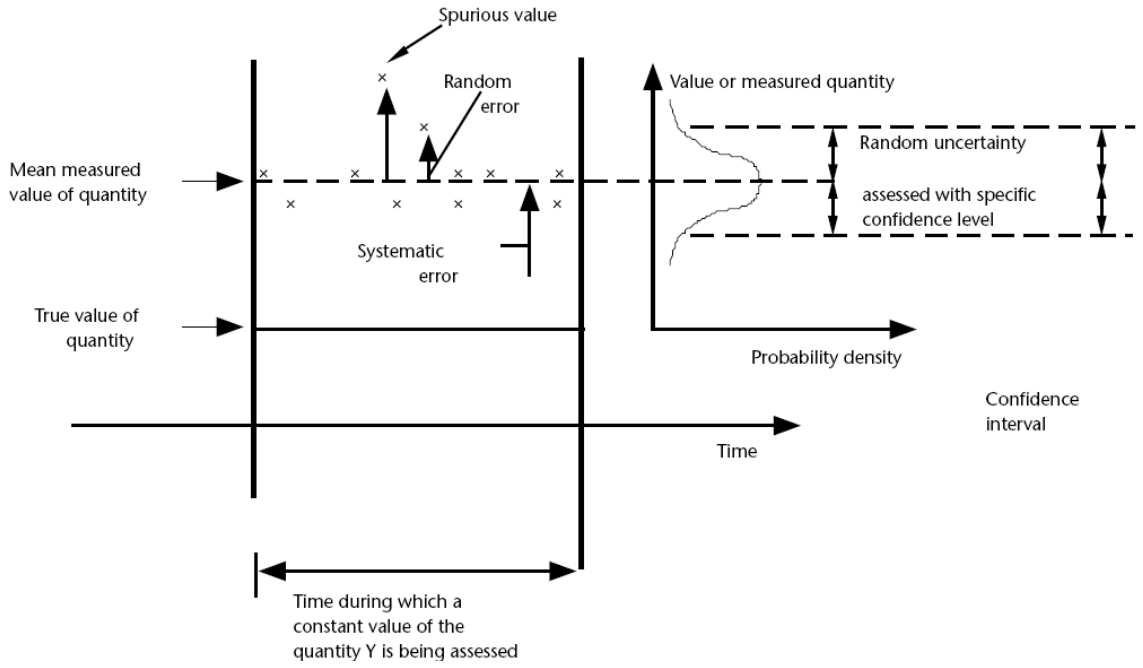


Figure 2.2 - Probabilistic uncertainty of measured values (WMO, 2008b)

Observations of climatic characteristics vary in uncertainty and unpredictability (only with respect to our current understanding and representation), temperature measurements for instance tend to vary less than rainfall, which in turn varies less than runoff (Koutsoyiannis, 2003). An increase in uncertainty and unpredictability is exhibited when comparing climate models of coarse spatial scale with hydrological models of finer spatial scale (Koutsoyiannis, 2003). Using sources of systematic and random measurement error listed by WMO (2008a) for air temperature as an example, relevant items were retained and modified for rain gauges and are listed as follows: 1) errors when comparing field instruments to standards; 2) non-linearity,

3) drift, 4) repeatability and reproducibility of the instrument. Transducer or recording device errors also need to be considered and include: 1) effectiveness in terms of lag-time of the controlled measurement component location in the instrument to capture the actual desired field meteorological parameter outside the instrument, 2) the degree to which the protective measures minimize effects from other undesired actors; and, 3) the degree to which the protective measures allows adequate exposure to the desired characteristic to be measured. WMO (2008a) also stated that significant sources of error might lie outside the instrument such as topographical or environmental features. Diligent instrument operation, monitoring and siting may help minimize the magnitude of such errors that, if left unchecked could produce considerable deviations from true values (WMO, 2008a). An occasional mistake when instrumentation errors are discussed is the assumption that the laboratory calibration describes completely the total error (WMO, 2008a).

Observational systematic errors represent an instrument's average deviation from the true value (e.g. originating from environmental or siting sources) (WMO, 2008a). These errors can only be minimized, rendering measurements only approximations of true values (WMO, 2008a; WMO, 2008b); however, applying a correction factor can sometimes account for that deviation (WMO, 2008a). Generally, systematic sensor errors can be eliminated with proper calibrations and adjustments (WMO, 2008a). Operationally, measured values, after the removal of systematic errors are considered as reasonable estimates of true values (WMO, 2008a).

Random errors originate from unpredictable variations (WMO, 2008a) and the magnitude of this type of error can be estimated by the standard deviation after several observations (WMO, 2008a). Essentially, random errors represent the degree of uncertainty with which systematic errors are known (WMO, 2008a).

Standard meteorological observations of precipitation are completed with instruments that are generally at ground level; however, they generally do not include attributes describing the structure and character of the observations such as

drop size distribution and state (WMO, 2008a). Obtaining representative precipitation information is a major issue in meteorology (WMO, 2008a). Precipitation measurements for instance can strongly be influenced by exposure, wind and topography (WMO, 2008a); hence the importance of choosing a suitable site and documenting potential external influences.

The amount of instruments and their spatial distribution within the area of interest greatly influences the precipitation's representativeness when used for applications of different meteorological scales such as synoptic, meso or microscales (WMO, 2008a). For example a location's particular wind fields due to its nearby surroundings can induce local maxima or minima in precipitation (WMO, 2008a). WMO (2008a) provides guidance on how to choose a suitable instrument location such as significant sized objects should be kept away from the instrument at least twice their height above the instrument's opening (WMO, 2008a). Also, slopes and rooftops should not be considered as suitable locations (WMO, 2008a). Instead, areas that adequately shield the instrument from wind in all directions should be sought, such as clearings in forests or shrub forests (WMO, 2008a). Wind effects can be minimized either by placing rain gauges at ground elevation or by inducing a horizontal air-flow over the instrument opening either with vegetation or fencing kept at the same height as the instrument's opening or, simply using of a windshield (WMO, 2008a).

2.3.2 Precipitation Gauges

Precipitation gauges are the most popular of precipitation measuring devices and are usually made of an open and vertical cylindrical container that includes an internal funnel (WMO, 2008a). Precipitation gauges describe incremental accumulations of water through time at near ground-level with direct measurements of its weight or volume (Michaelides et al., 2009; WMO, 2008a).

2.3.2.1 Types of Instruments

WMO (2008a; 2008b) and Michaelides et al. (2009) describe various approaches to measuring precipitation and types of instrument, each having their own particular

strengths and weaknesses. The following is a summary of these sources of information. Generally, precipitation gauges can be categorized into two groups or types: non-recording and recording.

2.3.2.1.1 Non-Recording Precipitation Gauges

Non-recording precipitation gauges can be further categorized into two types: standard gauges and storage gauges. Generally, these precipitation gauges can be described as unobstructed containers, typically of an upright cylindrical shape, whose specifications and instrument height from ground varies by country. The precipitation collected by these containers is determined via direct measurement of its content's depth, weight or volume.

2.3.2.1.1.1 Standard Precipitation Gauges

Ordinary precipitation gauges are made of a collector, which captures the precipitation, located over a funnel that guides rain or melted solid precipitation particles in the container to be kept until measured. Figure 2.3 illustrates several common shapes of precipitation gauge (Sevruk & Nespor, 1994 in WMO, 2008a).

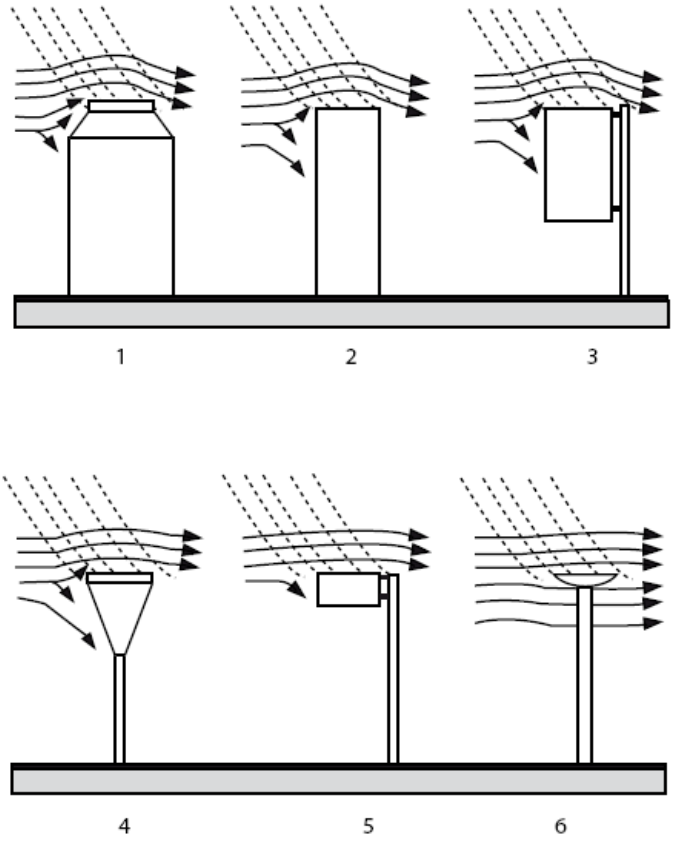


Figure 2.3 - Various shapes of standard rain gauges with wind field deformation (Sevruk & Nespor, 1994 in WMO, 2008a)

2.3.2.1.1.2 Storage Precipitation Gauges

Storage gauges allow the capture, and therefore description of total seasonal precipitation for remote and sparsely populated regions. Similar to the standard gauge, the storage gauge is composed of a collector, a funnel and a container that enables the storage of collected precipitation for extended periods between measurements.

2.3.2.1.2 Recording Precipitation Gauges

Although there are several types of recording precipitation gauges, three of the most commonly used will be discussed, namely: 1) weighing–recording gauges, 2) float gauges and, 3) tipping-bucket gauges. Of these three, only the weighing gauge can adequately describe all forms of precipitation while the float and tipping-bucket gauges are mostly suited for rainfall measurement only. Other types of recording

precipitation gauges exist, such as rainfall-rate recorders; disdrometers, optical gauges and acoustic type gauges, however these instruments are not as widely used because of their complexity, high costs or in some cases represent unproven new technologies.

Recording precipitation gauges have the significant advantage of automatically making measurements in the absence of human presence; hence, allow systematic readings at smaller time resolutions as compared to manual readings. This can minimize evaporation and wetting losses from precipitation measurement; however, the potential wind effects still remain.

2.3.2.1.2.1 Weighing–Recording Gauge

One aspect of the weighing-recording type of gauge that sets it apart from other precipitation gauges is its weighing mechanism. It can continuously record the weight of the container and its content using either a spring mechanism, a system of balanced weights or, a pressure transducer. Measurements of precipitation and its accumulation is made instantaneously irrespective of the type of precipitation. Contrary to tipping bucket rain gauges, the weighing-recording gauge can characterize rainfall intensity, including intense rainfall rates. However, these types of gauges are typically costlier and require more maintenance as compared to tipping-bucket rain gauges.

2.3.2.1.2.2 Float Gauge

Float gauges measure the changes in the amount of precipitation collected in its container with float devices, that transfers the level of the float to a paper chart via a pen or, to a digital data recorder via a transducer.

2.3.2.1.2.3 Tipping-Bucket Gauge

Tipping-bucket gauges are perhaps the most frequently used rain-measuring devices of automatic weather stations. The main reason is their internal workings that are simple and easily lend themselves to digital data recording. The instrument is basically composed of a collector surface that funnels the captured rainfall onto

the tipping-bucket mechanism. A double-bucket component, similar to a teeter-totter and made of light metal, forms the essential part of this tipping-bucket mechanism. The cantilevered double-buckets are kept on a fulcrum that alternatively tips downward when the upper bucket is filled with a set quantity of water. When in the lower position, the collected rain is emptied from the bucket and, either discarded or captured for subsequent measurement. Considering that each tip represents a known quantity of rainfall and given the timing of each tip is recorded then the rainfall totals and intensity can be determined. Observations can either be summed at set time intervals or time-stamped one event (tip) (Habib et al., 2001).

2.3.2.1.3 Reference Gauges

There are limitations to testing instrument in laboratory settings; hence, determination of an instrument's accuracy in measuring rainfall is generally done in field comparison tests alongside pit gauges (Devine & Mekis, 2008). Pit gauges produce the best rainfall measurements excluding adjusted gauge data observations (Devine & Mekis, 2008).

The purpose of reference gauges is to provide estimates of observational error magnitudes for specific types of gauges. This is achieved by co-locating a given test gauge next to a reference gauge and comparing their recorded rainfall amounts. Note that there is no specific type of gauge that serves as a reference gauge. A common feature of reference gauges is the significant effort put into minimizing wind effects that are responsible for the greatest amount of measurement error as shown in Figure 2.4 (WMO, 2008b). If precipitation in the liquid form is the target variable to be measured, the reference gauge is set at ground-level and in a gridded pit far from its edges in order to minimize in-splashing. This setup typically captures more precipitation than above-ground gauge installations. However, pit gauges are quite labour intensive to operate and maintain (Devine & Mekis, 2008). Nonetheless, all rain gauges, even reference gauges, are not exempt from errors (Devine & Mekis, 2008).

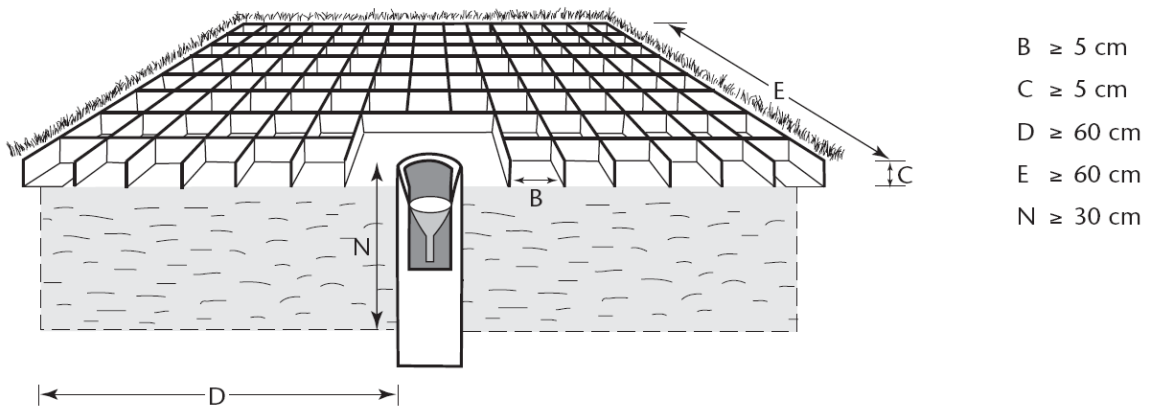


Figure 2.4 - Standard reference rain gauge (WMO, 2008b)

2.3.2.2 Sources of Measurement Errors

Irrespective of the instrument's accuracy, the quality of the precipitation observation is subject to measurement errors from several sources as will be discussed in the following sub-sections (Devine & Mekis, 2008).

2.3.2.2.1 Common Errors

The extent of common errors found in all gauges, as listed by WMO (2008a; 2008b), vary depending on the gauge design and operation and include: wind effects, wetting loss, evaporation loss, splash and random measurement errors. For greater detail about the meteorological and instrumental factors, the reader is invited to consult the above-referenced WMO documents. Post-measurement adjustments to rainfall data can be completed in an attempt to account for each of the systematic errors as elaborated in the many key reference documents (Mekis and Hogg, 1999; WMO, 2008a; Devine & Mekis, 2008; Mekis & Vincent, 2011 among others). Several of the recommended adjustment approaches were produced using reference gauge comparison; however, adjustments can now also be made by numerical modelling (e.g. wind field effects on rainfall collectors) (Habib et al., 2008). Nevertheless, WMO (2008a) cautions that data adjustments for errors are not perfect and there is even the possibility of exacerbating the aggregate errors. Complimentary weather information is often needed in order to apply data adjustments such as wind speed and temperature corrections (WMO, 2008a; WMO,

2008b). Aside from wetting loss, wind speed is a common factor that tends to exacerbate rainfall catch errors (WMO, 2008b).

2.3.2.2.1.1 Wind Effects

Wind pushes rain particles which in turn fall at an angle to the vertical (Devine & Mekis, 2008). Wind effects over the gauge's opening negatively impacts the instrument's catch; hence, undermine accurate measurements (Goodrich et al., 1995; Groisman and Legates, 1994 via Allen & DeGaetano, 2005; WMO, 2008a; WMO, 2008b; Mekis & Vincent, 2011). As Figure 2.3 illustrates (WMO, 2008a), the on-site presence of a gauge deforms the local wind field and thereby influences the rainfall catch (Goodrich et al., 1995; WMO, 2008a; WMO, 2008b) by producing a wind speed increase of about 35% above a cylindrical gauge opening (Sevruk & Klemm, 1989 in Devine & Mekis, 2008). This increase in the wind speed above the opening reduces the rain particle catch, especially at the lower end of the wind spectrum (Devine & Mekis, 2008). Also, since the vertical wind speed profile increases logarithmically as a function of distance above the ground, a proportionally increasing impact on rainfall measurement errors is observed with higher instrument opening installations (Devine & Mekis, 2008). Furthermore, wind induced turbulence eddies can also form above the gauge opening and, under extreme conditions can even fling out rain particles before they even come in contact with the instrument (Krajewski et al., 1998 in Devine & Mekis, 2008) All these issues can inevitably modify the captured drop size distribution from that of the unimpeded path that precipitation would have taken had the instrument not been there (Devine & Mekis, 2008).

Wind induced errors usually account for a 2 to 10 % under estimate of the true rainfall values; however, larger errors can occur (Sieck et al., 2007; WMO, 2008a; WMO, 2008b) due to high winds. As previously mentioned, wind is typically responsible for most measurement errors (Goodrich et al., 1995; WMO, 2008a; Devine & Mekis, 2008).

WMO (2008a, equation Annex 6.B) suggests a specific correction function in regards to adjusting data for wind effects on precipitation catch. However, random components of wind patterns, such as speed and direction as well as rainfall drop size distribution have been observed and; unfortunately, these cannot be accounted for, nor corrected in post-recording data adjustments (Goodrich et al., 1995).

Local topography plays a role in defining the wind regime (Goodrich et al., 1995; EC, 2010a) thereby creating local precipitation maxima and minima (WMO, 2008a; WMO, 2008b; EC, 2010a). Alternatively, the installation of wind shields on rain gauges are known to help reduce wind effects (WMO, 2008a; WMO, 2008b). Also, proper site selection and maintenance of the surrounding environment can play an important role in minimizing adverse wind field effects on gauge catch (WMO, 2008a; WMO, 2008b). Ideally, the instrument would be placed at a minimum distance of twice the tallest nearby object height, but preferably at four times the distance (WMO, 2008a; WMO, 2008b). As previously mentioned, an alternative may be to install an actual windshield apparatus (WMO, 2008a; WMO, 2008b). Certain sites such as hill-sides or rooftops are best avoided (WMO, 2008a) due to the potential for modifying local wind fields, hence, rainfall catch.

2.3.2.2.1.2 Wetting Loss

Wetting loss occurs due to the remaining water on the inner walls of the instrument and is consequently not measured (WMO, 2008a; Mekis & Vincent, 2011). This unavoidable loss is caused by the minimum amount of water needed to moisten the gauge's internal collecting surface in order to subsequently let rainfall flow downward into the measuring mechanism (Devine & Mekis, 2008). The potential wetting loss is in part a function of the gauge surface area leading to the measuring apparatus (Devine & Mekis, 2008). Intermittent rainfall with periods of no rain of several hours further increases such losses since it repeatedly allows the gauge surfaces to dry (WMO, 2008b; Devine & Mekis, 2008). The condition and type of material used for the collection surface of the gauge influence its roughness and consequently the wetting loss (Devine & Mekis, 2008). Total wetting loss errors are estimated to represent a 2 to 15% reduction from true values during the summer and

1 to 8% during winter months (WMO, 2008a). However, the loss is generally in the range of 2 to 10% (WMO, 2008b).

2.3.2.2.1.3 Evaporation

Evaporation loss occurs when captured water is evaporated prior to measurement (Devine & Mekis, 2008). During the day, the temperature within the gauges rise, increasing the potential evaporation losses (Devine & Mekis, 2008). Also, the presence of wind can further increase the evaporation loss (Devine & Mekis, 2008). Errors due to evaporation loss typically range from 0 to 4% (WMO, 2008a; WMO, 2008b); however, it can be a more significant issue in warm and dry climates (WMO, 2008a; WMO, 2008b) or, in installations with infrequent measurements (WMO, 2008b).

2.3.2.2.1.4 Splash

Splash errors refer to water either splashing into the instrument (e.g. after hitting the ground) or splashing out of the instrument before it is recorded (WMO, 2008a; Devine & Mekis, 2008). A splash occurs when a larger rain drop breaks up into multiple smaller drops upon impacting a moist surface (Devine & Mekis, 2008). Greater wind speeds and larger fall angles result in greater splashing (Devine & Mekis, 2008). Such errors contribute only a 1 or 2% difference to the true rainfall (WMO, 2008a; WMO, 2008b). Although in-splashing is less of an issue when gauge openings are high off the ground (Goodrich et al., 1995), splash has been shown to reach heights greater than 1 m (Kurtyka, 1953 in Devine & Mekis, 2008). Note that in-splashing is the only type of error that can produce an overestimation of the true rainfall - all other errors produce underestimations (WMO, 2008a; Devine & Mekis, 2008). Also, this error is hard to quantify and measurements are typically left unadjusted for this loss/gain; hence, splash is best avoided (WMO, 2008a). Generally, it is simply assumed that the losses and gains cancel each other out (WMO, 2008b).

Measures to minimize in- and out-splashing include having instruments surrounded by grass as well as using precipitation collectors designed with the wall heights

proportional to the funnel angle as shown in Figure 2.5 (WMO, 2008a; WMO, 2008b).

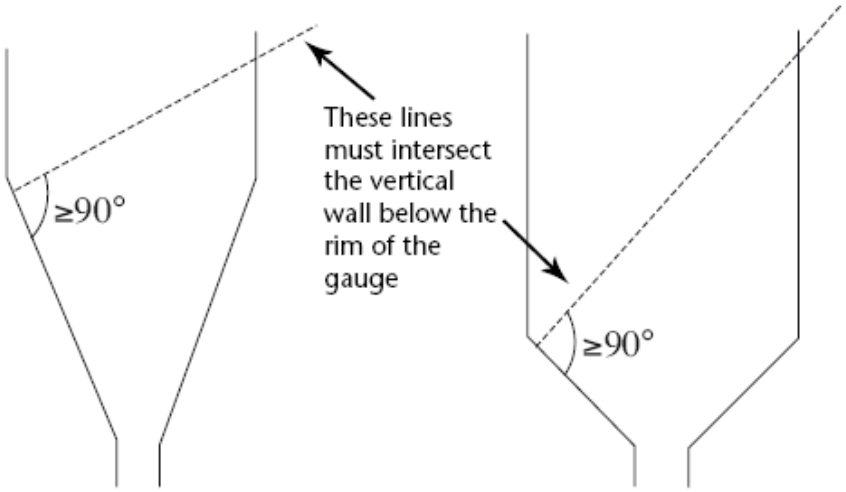


Figure 2.5 - Splash reduction rain gauge collector design (WMO, 2008a; 2008b)

2.3.2.2.1.5 Random Measurement Errors

Random measurement and instrument errors, including issues such as inaccurate hardware clocks, can also affect the accuracy of measurements (WMO, 2008a). Data logger clocks can drift by as much as 'one minute per month' (Michaelides et al., 2009); thereby, potentially affecting the actual length of time series and observation time stamps.

Random errors due to a lack in precision of measurements are generally small and tend to cancel each other (Goodrich et al., 1995). Instrumental errors can also occur as a result of the equipment and sensors being out of calibration or even out-of-level (Sieck et al., 2007). However, these types of errors are generally considered as negligible for well-designed and maintained instruments since systematic errors are a magnitude greater (WMO, 2008b).

2.3.2.2.2 Errors Relating to the Type of Instrument

Considering that the susceptibility of each type of error varies for each type of instrument, collocating different type of rain gauges is an ideal approach to

compensate each gauge's shortcoming (Ciach, 2003; Michaelides et al., 2009). Nevertheless, measured rainfall quantities can still vary within the collocating surface area due to the discrete measurement approach of each instrument as well as the slight difference in wind regime over each instrument (Michaelides et al., 2009). In absence of collocated rain gauges, data from isolated instruments are assumed to be error-free, unless obvious anomalies are identified (Ciach, 2003). In such a case, data quality and credibility is in question and can undermine comparative analysis with remotely-sensing data such as weather radar (Ciach, 2003). Nevertheless, for reasons that will be discussed in subsequent sections, the true rainfall at collocated rain gauges may differ but still be spatially accurate (Michaelides et al., 2009).

2.3.2.2.2.1 Non-Recording Rain Gauges

Wetting loss is a particularly concern with non-recording gauges that require transferring the collected precipitation into a separate container for measurement (WMO, 2008a). In this case, the accrued error is proportional to the frequency of transfers into the measuring container (WMO, 2008a; WMO, 2008b).

Although evaporation loss can be significant, it can be minimized by simply reducing the delay of measurement after each rainfall event (Michaelides et al., 2009).

Splash out is potentially also an issue unless efforts are made to limit the degree to which the accumulation gauge is filled since the loss of water is more likely in nearly full gauges (Michaelides et al., 2009).

2.3.2.2.2.1.1 Standard Rain Gauges

The use of rain gauges that are designed with a narrow opening will keep evaporation losses to a minimum by simply reducing the surface that is exposed to solar radiation (WMO, 2008a; WMO, 2008b).

2.3.2.2.2.1.2 Storage Rain Gauges

These types of gauges can be left unattended or go without being measured for long periods of time. Measures guarding against evaporation losses should be of the highest priority when operating this type of equipment as they could experience large errors. Error mitigation measures may include the addition of oil atop the water in the collection tank while preferring the use of precipitation gauges that are specifically designed to minimize evaporation losses (i.e. minimized exposed water surface and ventilation while keeping the gauge temperature low) (WMO, 2008a; WMO, 2008b). Also, the addition of antifreeze to the precipitation container in cold climates can help guard against instrument damage and maintain precipitation in liquid state (WMO, 2008a; WMO, 2008b).

A significant issue for this type of gauge is the potential accumulation of debris or obstruction considering they are left unattended for long periods of time.

2.3.2.2.2.2 Recording Rain Gauges

Instruments fitted with digital recorders are usually limited in their sensitivity (Morin et al., 2003) to about 0.1 mm (WMO, 2008a). As a result, precipitation amounts that are less than this minimum trigger resolution are not recorded or worst, erroneously added to the next rainfall event. This highlights that the true duration of a partitioned rainfall is in reality unknown, and under-estimated as water remains unaccounted for in the apparatus (Morin et al., 2003). Furthermore, this issue also adds to the uncertainty of accurately identifying the beginning and end of events (Habib et al., 2008). This also has implications for the determination of low rainfall intensity events since the duration is likely overestimated; thereby underestimating the actual intensity.

A common practice of flagging trace precipitation (amounts less than the instruments trigger value) as zero values in recorded data series can lead to underestimate the actual seasonal precipitation values (WMO, 2008a).

The data recorder clock time accuracy can be of significant issue with these types of gauges since data estimates, especially those pertaining to precipitation intensity, tend to increasingly differ from actual conditions with time. Furthermore, precipitation estimates may be incorrectly attributed to the wrong data and time (WMO, 2008a).

2.3.2.2.2.2.1 Weighing Recording Gauges

Evaporation losses can also be an issue for this type of gauge; however, it can be minimized by adding a layer of oil on top of the collected water (WMO, 2008a).

'Wind pumping' is a specific issue with this type of rain gauge and consists of a wind-induced oscillation of the water surface within the collecting area during high winds (WMO, 2008a) that is observed as fluctuating measurements in the recorded data. A hydraulic damping mechanism can minimize such an error while there is growing evidence that digital compensation may be possible by simply reducing the time resolution (WMO, 2008a).

As with non-recording accumulation gauges, the use of antifreeze during cold weather can effectively be used to melt any collected solid precipitation into liquid (WMO, 2008a) thereby creating more space for collection.

A particular advantage of this type of gauge is that it is not subject to wetting loss errors since the entire gauge container and content, including the film of water on the container walls, is weighed and consequently measured (WMO, 2008a). However, precipitation in the form of freezing rain and wet snow may adhere to the opening until it is melted; thereby, being recorded at a later time than it actually arrived (WMO, 2008a).

2.3.2.2.2.2.2 Float Gauges

Maintaining the water level in the container within the operational range of the float is crucial for this type of gauge. This can be achieved by choosing a float chamber that is large enough for the period of observation or; alternatively, equipping the

gauge with a mechanism that will automatically empty the float chamber in a timely manner when the water reaches a preset level (WMO, 2008a). Unfortunately, if precipitation were to fall during the emptying process, a most likely scenario, it would obviously not be recorded. However, some of these instruments are designed to minimize such errors by installing siphons that accelerate the emptying process. Alternatively, a second chamber is sometimes added to temporarily hold the collected precipitation during the emptying process (WMO, 2008a).

These instruments require a heating system when installed in cold climate regions in order to keep the collected precipitation in a liquid form for ongoing measurements but also to guard against equipment damage (WMO, 2008a). Unfortunately, heating the collected precipitation can increase evaporation losses (WMO, 2008a).

2.3.2.2.2.2.3 Tipping-Bucket Gauges

Significant errors can occur with tipping-bucket rain gauges if proper dynamic calibration is not completed (Habib et al., 2001; Habib et al., 2008). Manufacturers and meteorological agencies usually calibrate their instruments at low rainfall intensities; however, this approach fails to verify the performance at high rainfall intensities (Habib et al., 2008). The consequence is a potential loss of precipitation between tips (Habib et al., 2008; WMO, 2008a; WMO, 2008b; Michaelides et al., 2009) with the resulting under-estimation of the actual rainfall intensities and rainfall totals (Habib et al., 2008). The error is particularly important at greater intensities and is generally nonlinear in trend (Habib et al., 2008).

According to Habib et al. (2008), errors associated with un-calibrated (dynamic) tipping-bucket rain gauge measurements are in the range of 2 to 5% as compared to wind effects errors that are in the range of 3 to 9%.

As previously mentioned, there are two possible methods of processing tipping-bucket rain gauge data: 1) counting the number of tips within a fixed interval or, 2) collecting the date and time of every tip (Habib et al., 2001). When rainfall intensities are derived from fixed timeframes, the minimum recordable rainfall

intensity would be equivalent to one tip per set time interval; hence, all higher intensities would be expressed in terms of multiple of this aggregate value (Habib et al., 2001). However, when rainfall intensities are computed by considering that the total time that has passed since the last tip represents the total time it took to fill the bucket with rainfall, the rainfall intensity is consequently assumed to be uniform throughout that time period (Habib et al., 2001). This assumption is likely not always true when considering low rainfall intensities at coarse timeframes (Habib et al., 2001). Another related issue is the fact that tip data cannot be relied upon to properly characterize periods without rainfall since it only considers the time it takes to produce a tip, as an indication of precipitation (Habib et al., 2001). Consequently, periods without rainfall end up being described by prolonged periods of non-zero, low rainfall intensities (Habib et al., 2001). The figure below (Figure 2.6) illustrates the typical errors produced by a simulated tipping-bucket rain gauge when using the fixed timeframe data processing approach; hence, averaging rainfall intensities over a pre-set interval (Habib et al., 2001). Notice how the tipping-bucket instrument fails to record the finer details as it smoothes out the peaks and troughs (Habib et al., 2001). Also, notice that the tipping-bucket instrument fails to adequately describe both the start and the end of rainfall (Habib et al., 2001).

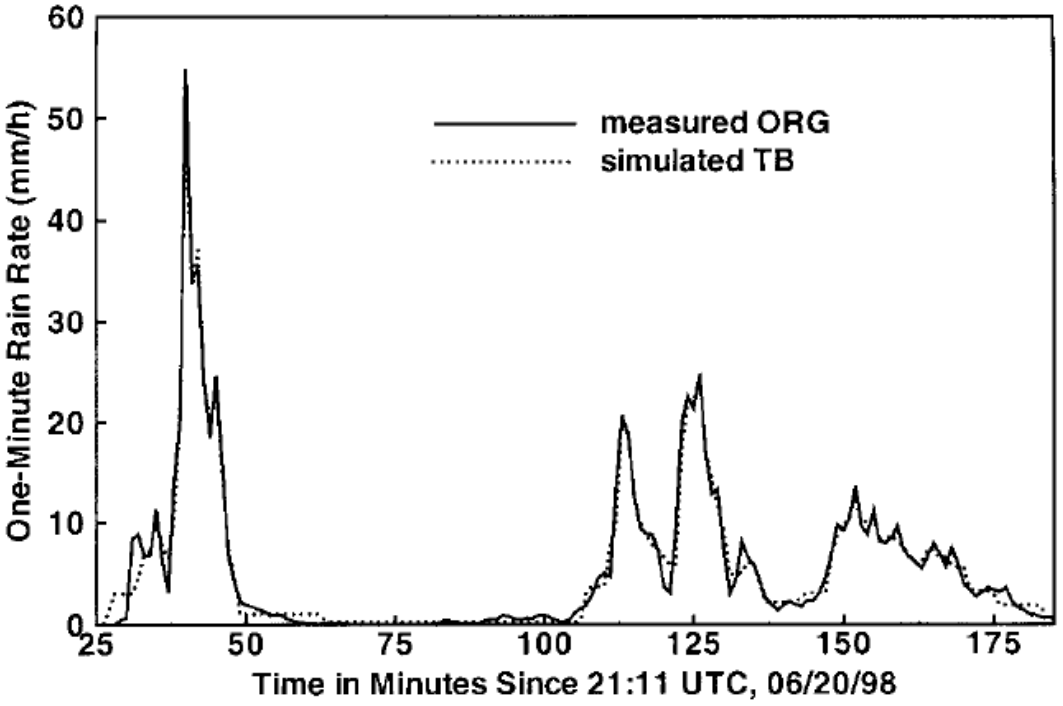


Figure 2.6 - Simulated tipping bucket rain gauge rainfall temporal averaging error (Habib et al., 2001)

Habib et al. (2001) also found that by increasing the tipping-bucket's fixed timeframe for counting tips to 5 or 15 minutes, as opposed to 1 minute, it minimizes error as compared to optical rain gauge data (see Figure 2.7).

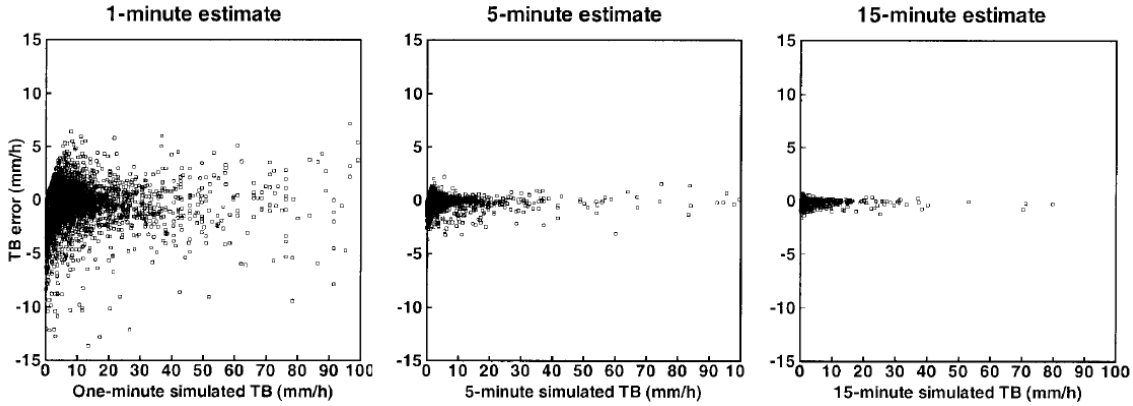


Figure 2.7 - Simulated tipping bucket rain gauge timeframe sampling comparison (Habib et al., 2001)

Research from Ciach (2003) also revealed that the selection of a larger timeframe helps reduce errors at any given rainfall intensity (see Figure 2.8).

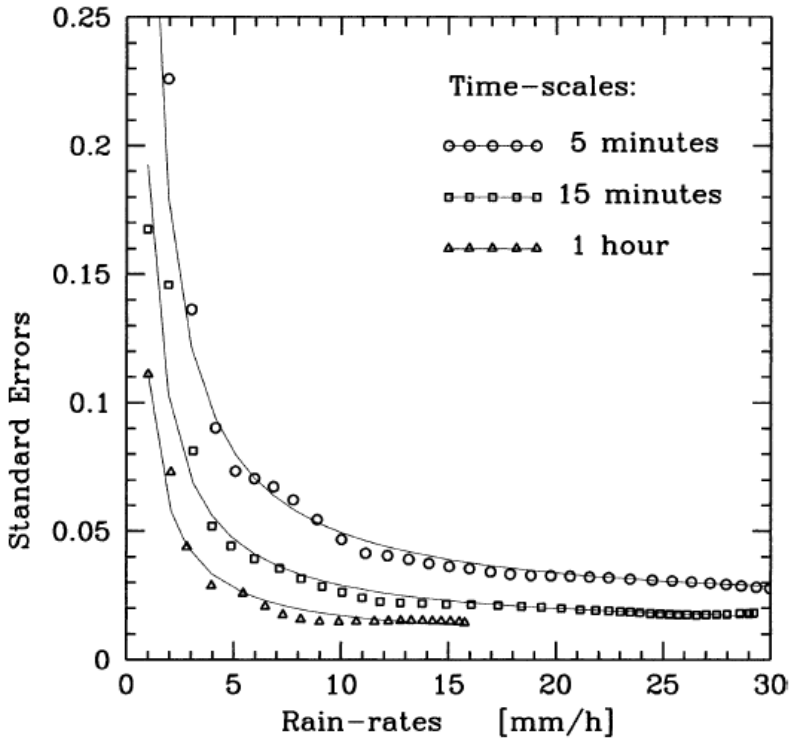


Figure 2.8 - Rain rate measurement uncertainty with time-scale (Ciach, 2003)

Also, Habib et al. (2001) found that by using a smaller bucket size (equivalent to 0.1 mm of rainfall rather than 0.254 mm and 0.5 mm) measurements improved when compared to the optical rain gauge data (see Figure 2.9) Based on these and more recent findings (Habib et al., 2008) the instrument's sampling time resolution should be as high as possible in order to minimize errors.

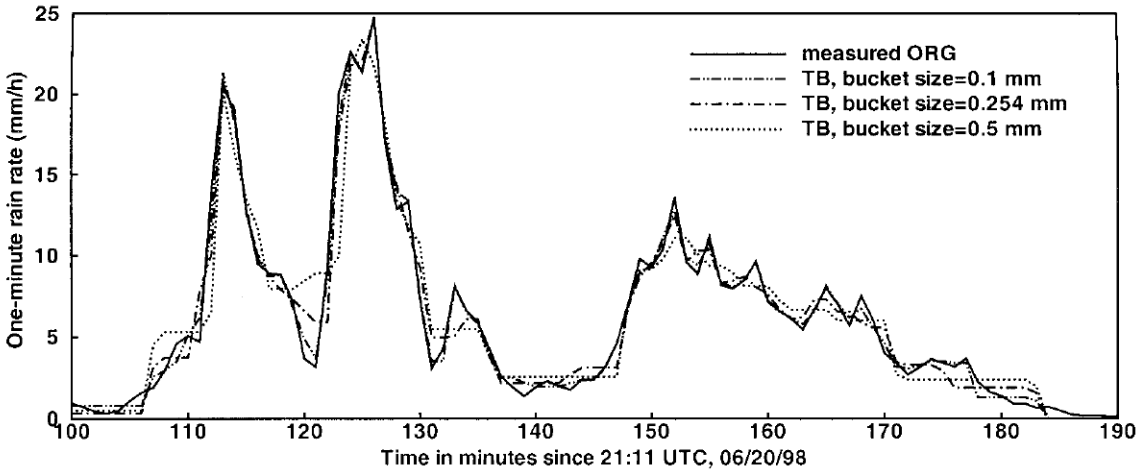


Figure 2.9 - Simulated impact of varying the tipping bucket rain gauge bucket size (Habib et al., 2001)

Unfortunately, tipping-bucket rain gauges are particularly susceptible to clogging or jamming due in part to their somewhat delicate measuring mechanism (WMO, 2008a; Michaelides et al., 2009). Furthermore, some elements of the gauge must be heated for use in sub-zero temperatures, which unfortunately increases evaporation losses (WMO, 2008a).

The typical rainfall measurement accuracy of this type of gauge lies between 5 to 10% of true rainfall, although better results may be achieved if proper wind shielding is provided (WMO, 2008a). However, evaporation losses can be significant in periods of light rain and warm weather due to the proportionately small volume of the bucket (WMO, 2008a; WMO, 2008b).

The volume of water that is required to tip the bucket can also change with time due to: 1) buckets that do not completely empty its full volume of water upon tipping, especially during periods of rapid tipping due to the presence of very high rainfall intensities or, 2) changing bucket properties such as increased surface oxidation and contamination by impurities that increased water retention (WMO, 2008a).

Also, if the rate at which the water is funnelled onto the bucket is large enough, the increased kinetic energy may prematurely induce a tip (WMO, 2008a).

Asymmetry with regards to the dual-bucket tipping movement as well as wear of the fulcrum will also produce errors (Michaelides et al., 2009).

2.3.2.3 Canadian Instruments Measuring Precipitation

Canadian rainfall information is gathered by all levels of government, including the Meteorological Service of Canada, a division of the federal ministry of Environment Canada (EC), as well as similar provincial and territorial ministries, municipalities, energy utilities and other public agencies (CSA, 2010). The data that is accumulated by these agencies collected by a wide selection of instruments and types of instruments that are operated under various standards (CSA, 2010).

The purpose of EC's meteorological systems is to provide a historical perspective of the country's climate and weather (CSA, 2010). Its networks and instruments are designed, located, and operated in a manner mostly adhering to guidelines produced by the World Meteorological Organization (WMO) (CSA, 2010).

Considering that the focus of this study is extreme rainfall statistics, only instruments whose data is used for that purpose will be further discussed. However, for a more comprehensive historical account of Canadian rain gauges, the reader is referred to the works by Metcalfe et al. (1997) as well as Devine and Mekis (2008).

The following two instruments are the most widely used gauges to derive EC's extreme rainfall statistics throughout its data gathering network: the MSC Type B standard rain gauge and, the MSC tipping-bucket rain gauge (CSA, 2010). Since 2002 the MSC tipping-bucket rain gauges are gradually being replaced with the newer TB-3 tipping-bucket rain gauge (Devine & Mekis, 2008; CSA, 2010). EC also uses the F&P/Belfort weighing gauge for year-round liquid and solid precipitation measurements (Devine & Mekis, 2008; CSA, 2010). However, while in use since approximately 1965, the FP/Belfort rain gauge is also being replaced and expected

to be phased out by the end of 2012 (EC, 2010a) - being replaced by the Geonor and Pluvio weighing gauge (Devine & Mekis, 2008; CSA, 2010). Both replacement instruments showed improved performance over their predecessors (Devine & Mekis, 2008). Figure 2.10 presents sketches of these common precipitation gauges (CSA, 2010). Also, Figure 2.11 shows a photograph and a sketch of the two models of weighing gauges used in Canada (Devine & Mekis, 2008).

The MSC Type B standard rain gauge was adopted in 1970 to replace the standard MSC copper gauge within that decade (Devine & Mekis, 2008; CSA, 2010; Mekis & Vincent, 2011). Manually operated rainfall stations like the MSC Type B standard rain gauge are usually read on a 6-hour basis, totalled daily (EC, 2010a).

The currently used MSC and TB-3 tipping-bucket rain gauges are fitted with buckets whose tip volumes represent 0.2 mm of rainfall depth (Devine & Mekis, 2008; CSA, 2010). However, prior to its metric conversion in 1978, the MSC tipping-bucket rain gauge instruments recorded rainfall at a resolution of 0.254 mm (Metcalfe et al., 1997; Devine & Mekis, 2008).

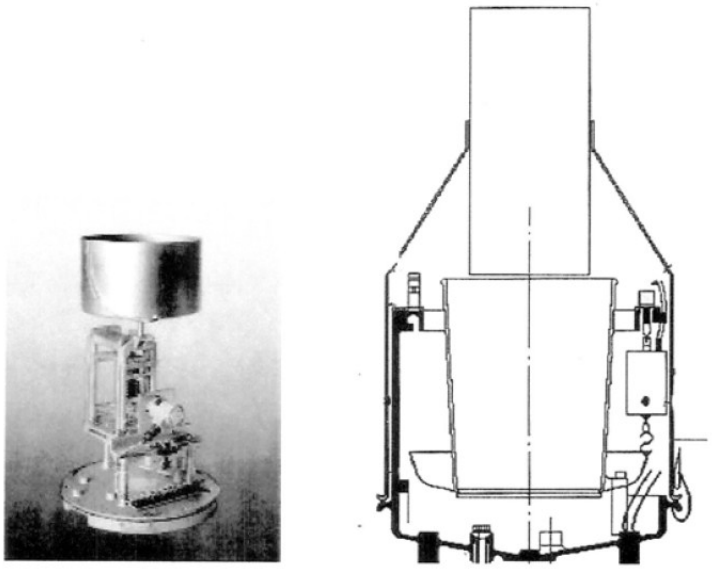
MSC Type B standard rain gauge — non-recording
(Instruction Manual 43 – First Edition, Atmospheric Environment Service, 1971)

MSC Tipping bucket rain gauge — recording
(TM 04-01-03, Environment Canada, 1981)

TB-3 Tipping bucket rain gauge 2002 — recording
(Instruction Manual Tipping Bucket Rain Gauge Model T.B.3, issue 2, Hydrological Services PTY Ltd, 1990)

F&P/Belfort weighing gauge 1965 — recording
(Belfort Model 3000 Specificati ns, Belfort Instruments, 1998)

Figure 2.10 - Most common Canadian IDF rain gauges (CSA, 2010)



a) b)

Figure 2.11 - Weighing gauges used in Canada (Devine & Mekis, 2008)

Among all EC's instruments, only tipping-bucket rain gauges have sufficient recording resolution for generating severe rainfall statistics (Hogg & Hogg, 2010). Data from other rainfall observing instruments are archived as daily precipitation totals only (Hogg & Hogg, 2010). Before the advent of digital data loggers, tips were mechanically recorded on paper charts (CSA, 2010).

Figure 2.12 shows the distribution across the country of 557 EC tipping-bucket rain gauge stations that were in operation in 2010 and are used to produce extreme rainfall statistics (CSA, 2010).

Figure 2.12 - EC's 2010 IDF gauge network (CSA, 2010)

The MSC tipping-bucket rain gauge, introduced after the second World War (Devine & Mekis, 2008), has shown itself to be both simple to operate and reliable, despite its known measurement errors, notably pronounced during high rainfall intensities (Hogg et al., 1989; CSA, 2010); errors that are typical to tipping-bucket rain gauges.

EC seasonally operates its tipping-bucket rain gauges when the ambient temperature is above freezing. The exact timing varies across the country; however, the operating period is generally mid-April to mid-October (CSA, 2010). These instruments are maintained regularly during site inspections and, on at least a yearly basis, they are calibration-tested and re-calibrated if necessary (CSA, 2010).

The weighing gauges, like the F&P/Belfort, are used in remote regions and often left unattended for months at a time (Hogg et al., 1989; Devine & Mekis, 2008). These recording instruments are notably limited in their observational sensitivity and temporal resolution (Hogg et al., 1989). The Canadian weighing gauge recording resolution is only 0.6 mm (Devine & Mekis, 2008). Furthermore, the Geonor and Pluvio weighing gauges record precipitation totals on only an hourly basis (EC, 2010a).

Table 2.2 conveniently lists the characteristics of each type of rain gauge thereby facilitating their comparison (Devine & Mekis, 2008). Notice that tipping-bucket rain gauges have been in use for 110 years.

Table 2.2 - Canadian rain gauge characteristics (Devine & Mekis, 2008)

Rain Gauge		Orifice diameter (cm)	Resolution (mm)	Height (m)	Time Period
Type	Model				
Manual	Kingston skirted	9.1	0.254	0.3	1869-1914
	Cast-iron base	9.1	0.254	0.3	1914-35
	MSC copper/plastic	9.1	0.254	0.3	1935-85
	Type B plastic	11.3	0.254 then 0.1	0.4	1970-present
Recording	Friez tipping bucket	30.5	0.254	0.91	1902-50
	MSC tipping bucket	25.4	0.254 then 0.2	0.75	1937-present
	Hydrological Services TB3	20.0	0.2	0.75	2002-present
All-weather	F&P/Belfort	20.3	0.6	2.13	1967-present
	Geonor T-200B	16.0	0.6	2.13	2002-present

2.3.2.3.1 Rainfall Estimate Error and Adjustments

Both automated and manual data quality control logs are also archived in the national climatological database along with the actual station precipitation data.

These quality control logs mainly flag outliers and validate internal consistency. (CSA, 2010).

A Provision for signalling the presence of precipitation below the instrument's measurement resolution is also available and, in such cases, a trace precipitation amount flag is set. However, this trace amount of precipitation is excluded from any rainfall accumulation statistics (Mekis & Vincent, 2011). Unfortunately, this method of accounting for trace amounts of precipitation may become an issue, especially for large regions in Canada where its frequent occurrence become significant, notably in the Arctic (Mekis & Vincent, 2011).

Devine and Mekis, (2008) present a comprehensive discussion on the sources of error that affects rain gauges and also provide a great historical account of the use of these instruments in Canada.

CSA (2010) conveniently summarized the experimental findings of Devine and Mekis (2008) regarding EC rain gauge errors (see Table 2.3). As can be observed, the TB3 and Geonor gauges show improved measurement certainty as compared to the instruments they are respectively replacing; namely, the MSC tipping-bucket and the Belfort F&P gauges.

Table 2.3 - Observed field accuracy of EC rain gauges compared to the WMO reference gauge (Devine and Mekis, 2008 in CSA, 2010)

EC Rain Gauges	Percentage Bias
MSC Type B	-0.6%
TB3	-3.5%
Belfort/F&P	-7.1%
Geonor	-4.7%

During low wind speed conditions, Equation 2.1 can be used to adjust measured rainfall data (Devine & Mekis, 2008) as described below:

$$R_a = (R_m + F_c + tE_c + C_c) \times (1 + W_c) \quad \textbf{(2.1)}$$

where the parameters are as follows:

R_a = adjusted rainfall (mm),
R_m = measured rainfall (mm),
F_c = funnel wetting correction (mm per rain measurement period),
t = hours since last measurement (hr),
E_c = evaporation in container/receiver (mm/hr),
C_c = container/receiver retention correction (mm per rain measurement period),
W_c = wind correction factor.

Suggested values for each parameter in the equation are listed in Table 2.4 as a function of the type of instrument (Devine & Mekis, 2008).

Table 2.4 - "Rain gauge corrections for manual and recording gauges (estimates are in brackets)" (Devine & Mekis, 2008, p. 225)

			Manual Gauges			Recording Gauges		
Type of correction	Unit	Add or Multiply	MSC, copper receiver	MSC, plastic receiver	Type B Gauge	TB3 Tipping Bucket	MSC Tipping Bucket	F&P c/w Alter shield
Wind at orifice level	-	x	0.04	0.04	0.02	-	-	0.02
Wetting at funnel area	mm/event	+	0.13	0.13	0.08	(0.05)	(0.03)	-
Evaporation	mm/h	+	0.005	0.008	0.002	(0.06)	-	slight
Evaporation (previous work)	mm	+	0.02	0.03	0.01	-	-	-
Wetting of receiver of container	mm/event	+	0.06	0.03	0.04	-	-	0

2.3.2.3.1.1 Manual Gauges

Adjustments to each observation from manual gauges required corrections for systematic errors as listed in Table 2.4 (Devine & Mekis, 2008). **F**urthermore, according to research findings by Mekis and Vincent (2011), the magnitude of required adjustments that Canadian manual gauges data needed is illustrated in the Figure 2.13.

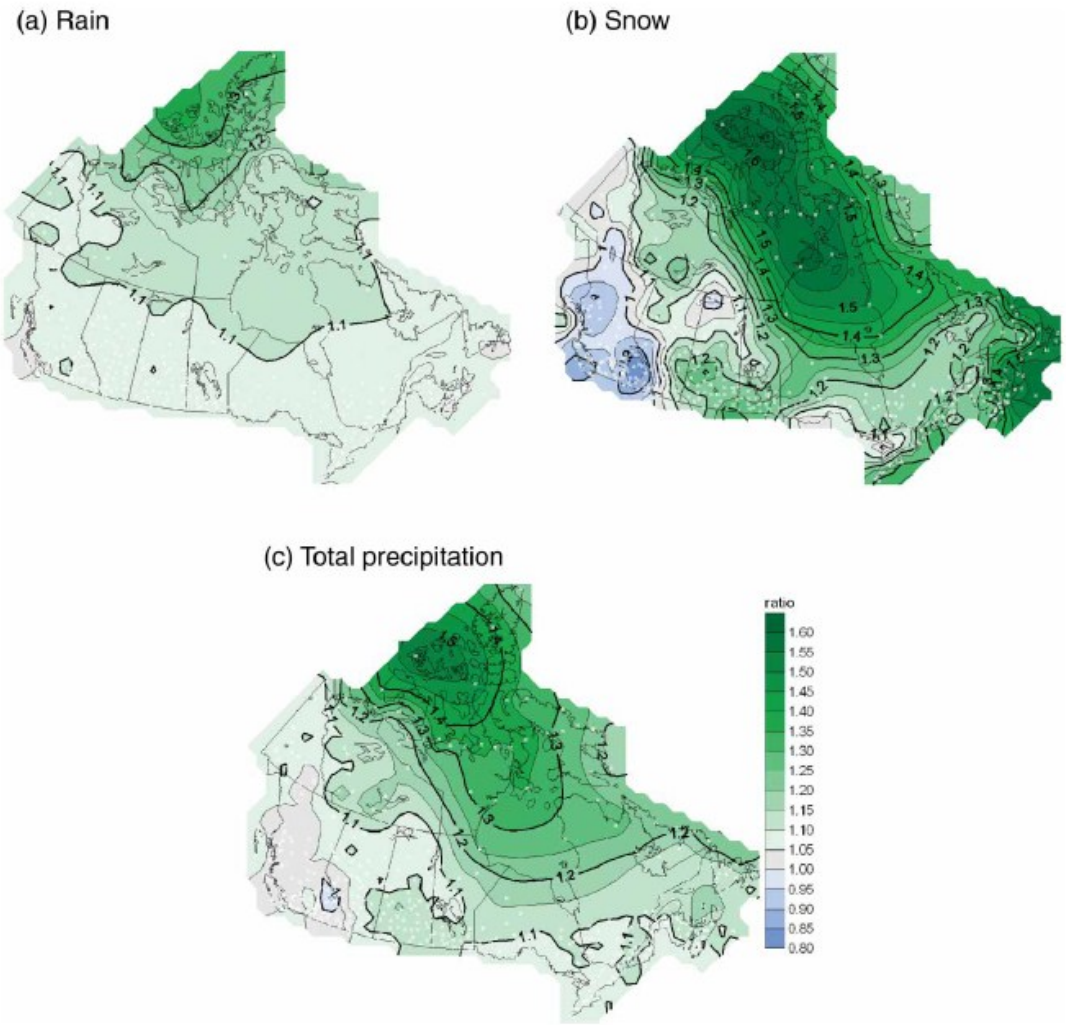


Figure 2.13 - "Magnitude of adjustments for all known rain and snow measurement issues using the average ratio of the annual rain (snow, total) with all adjustments to the original rain (snow, total) measurements from 1950 to 2009" (Mekis and Vincent, 2011, p. 170)

As previously mentioned, manual gauges are typically subject to a larger wetting loss due to the transfer of the captured water into a graduated cylinder for measurement (Devine & Mekis, 2008). This is not the case for Type B rain gauges that have a built-in graduated cylinder (Devine & Mekis, 2008).

2.3.2.3.1.2 Tipping-Bucket Rain Gauges

EC does not shield its tipping-bucket rain gauges against wind effects (Devine & Mekis, 2008).

Table 2.4 provides limited rainfall data correction opportunities for use with tipping-bucket rain gauges if adjustments with the equation are desired. Furthermore, the MSC tipping-buckets require data adjustments proportional to the rainfall intensity (Devine & Mekis, 2008).

The standard operation for EC's tipping-bucket rain gauges is to compare and adjust the reading using the daily totals from collocated standard gauges (Hogg et al., 1989; Metcalfe et al., 1997; Hogg & Hogg, 2010; CSA, 2010) considering that this type of gauge produces a better volumetric measurement (Devine & Mekis, 2008). However, this procedure may cease to be the norm as half of EC's tipping-bucket rain gauges are now without collocated standard gauges (CSA, 2010). As the MSC tipping-bucket rain gauges are being replaced by TB3 instruments, the standard gauges are also removed; consequently, data recorded by these new gauges are archived without adjustments (CSA, 2010).

The TB3 type of gauge is designed to facilitate drainage off the instrument's surfaces and the measurement mechanism thereby minimizing its retention losses (Devine & Mekis, 2008).

The MSC tipping-bucket rain gauges are calibrated for ideal measurement at 50 mm/hr while catch error at intensities of 100 mm/hr is approximately -3% and have also been found to considerably under-estimate the actual rainfall in periods of very low rainfall intensities (Devine & Mekis, 2008; CSA, 2010). In contrast, the TB3 tipping-bucket rain gauge has a siphon below the funnel that acts as a flow control device that discharges water at a constant rate (Devine & Mekis, 2008). This advantage, coupled with the adjustable tipping-bucket mechanism, allows for accurate rainfall intensity measurements up to 300 mm/h (Devine & Mekis, 2008).

Contrary to the cited CSA (2010) work, Devine and Mekis (2008) lists historical and experimental results that clearly shows that MSC tipping-bucket rain gauge measurement errors range between -5.3 to -22.2%, and generally staying above -10%.

A common practise of indiscriminately applying a constant data adjustment factor to MSC tipping-bucket rain gauge measurements across all rainfall intensities in order to match daily precipitation totals with data obtained from a collocated standard rain gauge, typically results in an underestimation of short duration rainfall intensity values at the higher end of the precipitation intensity scale (CSA, 2010).

2.3.2.3.1.3 Weighing Gauges

EC weighing gauges are fitted with wind shields to reduce the risk that precipitation measurements are negatively impacted by wind effects (Devine & Mekis, 2008; CSA, 2010). This is particularly important since, as previously mentioned, the wind can induce an oscillation of the water surface in the collection container thereby creating an artificial head of water that can appear as a false rainfall reading of up to 10 mm in magnitude (Devine & Mekis, 2008).

The weighing gauges require maintenance service two or more times per year in order to empty the collected precipitation, add antifreeze and oil (Devine & Mekis, 2008).

2.3.2.4 Spatial Point Measurements

As the most prevalent source of precipitation information, gauge precipitation data are the most widely used for spatial analysis (WMO, 2008a). However, these instruments measure over such a small area as compared to the actual surface areas of interest (i.e. watersheds) that it's essentially considered as point-source instruments. In fact, the most widely acknowledged short-coming of rain gauge data is its limited spatial representation due to its 'near-point sampling size and sparse spatial coverage' (Habib et al., 2008).

2.3.2.4.1 A Significant Source of Representativeness Error

Possibly the most important factor that contributes to the current poor representativeness of true extreme rainfall statistics is the ongoing accumulation of point source data rather than storm cell maxima over the region of interest. Also, short-lived high intensity rainfall events occur over small surface areas (Hogg & Hogg, 2010) while the most severe rainfall of such events does not appear over the exact same geographic location for every storm (Hogg & Hogg, 2010). Consequently, extreme rainfall data derived from point-source rainfall measurements of even the most densely populated instrument networks rarely, if ever, capture the most severe portion of the rainfall storm cells that fall onto the region for which the instrument is meant to represent (Durrans et al., 2002; Hogg & Hogg, 2010). Concerning extreme rainfall measurements, rain gauges effectively record sub-maxima precipitation quantities that are taken at varying distances away from each storm cell's center (Durrans et al., 2002; Hogg & Hogg, 2010).

2.3.2.4.2 Assumption of Point Source Data Representation of Regional Climate Patterns

Local meteorological conditions established by spatial extrapolation of point source data are often assumed as representative of a much greater region than the immediate vicinity of the instrument, this may not be valid. The high spatial variability of rainfall as well as site specific climate characteristics due not support a homogeneous rainfall assumption; hence, make this an imprudent approach for analysis (Goodrich et al., 1995). An indication of this premise is easily found by locating neighboring rain gauges with vastly different extreme rainfall statistics, even if they are located in the same climatically homogeneous region.

This leads to the current dilemma regarding the areal extent for which point-source statistics can confidently be used to describe the actual extreme rainfall occurrences (CSA, 2010). In this regard, the validity of a point source rainfall measurement is limited in its spatial extent by the region's period of record, its physiographic characteristics and types of precipitation events (WMO, 2008a).

Work by Ohara et al. (2011) points at the disadvantage of spatially interpolating point source rainfall measurements over mountainous terrain as well as other data from instruments only located in valleys. However, the description of the spatial variability of rainfall was determined as especially important in summers for a small mountainous British Columbia watershed (Hrachowitz & Weiler, 2011). It is easy to demonstrate the inadequacy of the homogeneous rainfall assumption over mountainous regions where orographic effects are quite pronounced.

Furthermore, Hrachowitz and Weiler (2011) has indicated that research clearly demonstrate that significant spatial variability of precipitation may occur at small-scales. Precipitation measurements over a large area (e.g. city) can exhibit strikingly different values depending on the location or timing of the observation (EC, 2010a). This is due, in part, to the high variability of rainfall intensities associated with the turbulent nature of the phenomenon as well as the weather system's constant state of movement (EC, 2010a).

The highly variable spatio-temporal nature of rainfall has made accurate measurements difficult (Gourley et al., 2010). Rain gauge observations are carried out over tiny areas, considered as a point measurement, whose measurements can be considered as reliable when its quantifiable errors are accounted for (Zawadzki, 1975 in Gourley et al., 2010). However, high rainfall intensities are usually associated with significant spatial variability that typically makes point source data a poor estimate of the nearby rainfall regimes (Zawadzki, 1975 in Gourley et al., 2010).

A study of 9 rain gauges within a 0.25 km^2 area found that recorded amounts varied up to 90% between instruments (Jensen and Pedersen, 2005 in Villarini et al., 2008).

2.3.2.4.3 What Influences Spatial Variability

It has been demonstrated that the most important elements affecting the spatial variability of precipitation are: topography (Hogg & Hogg, 2010), sampling time,

areal extent of the study location, storm event cloud type and whether liquid or solid precipitation is falling (Hrachowitz & Weiler, 2011).

2.3.2.4.4 Instrument Network Density

As outlined in the very first figure of this report (Figure 2.1), there are various scales pertaining to meteorological phenomena (WMO, 2010). Thunderstorms belong to the mesoscale category of meteorological phenomena are typically 3 to 100 km in extent (WMO, 2010). Meteorological phenomena that are smaller than the spacing of instruments within a network can easily escape detection (WMO, 2010); therefore, instrument spacing needs to be appropriately designed (WMO, 2010). Pedersen et al. (2007) also showed that networks with a low rain gauge density are hard-pressed to accurately describe the spatio-temporal characteristics of small events such as convective storms.

Furthermore, Morin et al. (2003) showed the importance of instrument density by plotting the correlation between rain gauge intensity measurements as a function of gauge spacing for significant storm events (see Figure 2.14). Also, a low instrument density across a network has been shown to increase uncertainty in their precipitation measurements' representation (Hrachowitz & Weiler, 2011).

Additionally, the network's layout of instruments has been shown to significantly influence the uncertainty of the precipitation measurements' representation (Hrachowitz & Weiler, 2011). Therefore, an ideal network instrument density should be implemented for best precipitation representation (Hrachowitz & Weiler, 2011). Work by Hrachowitz and Weiler's (2011) shows that a gauge density of at least 4 instruments per 5 to 10 km^2 in the mountainous regions on Vancouver Island, BC is required for best representation of precipitation over the area. Nevertheless, field observations showed a noticeable seasonal difference in recorded precipitation in a dense mountainous network of 6 gauges with an instrument density of 1 gauge per 2.5 km^2 (Hrachowitz & Weiler, 2011). Summer precipitation trends demonstrated a predominant storm population of shorter, smaller and spatially limited cells compared to those of winter (Hrachowitz & Weiler, 2011).

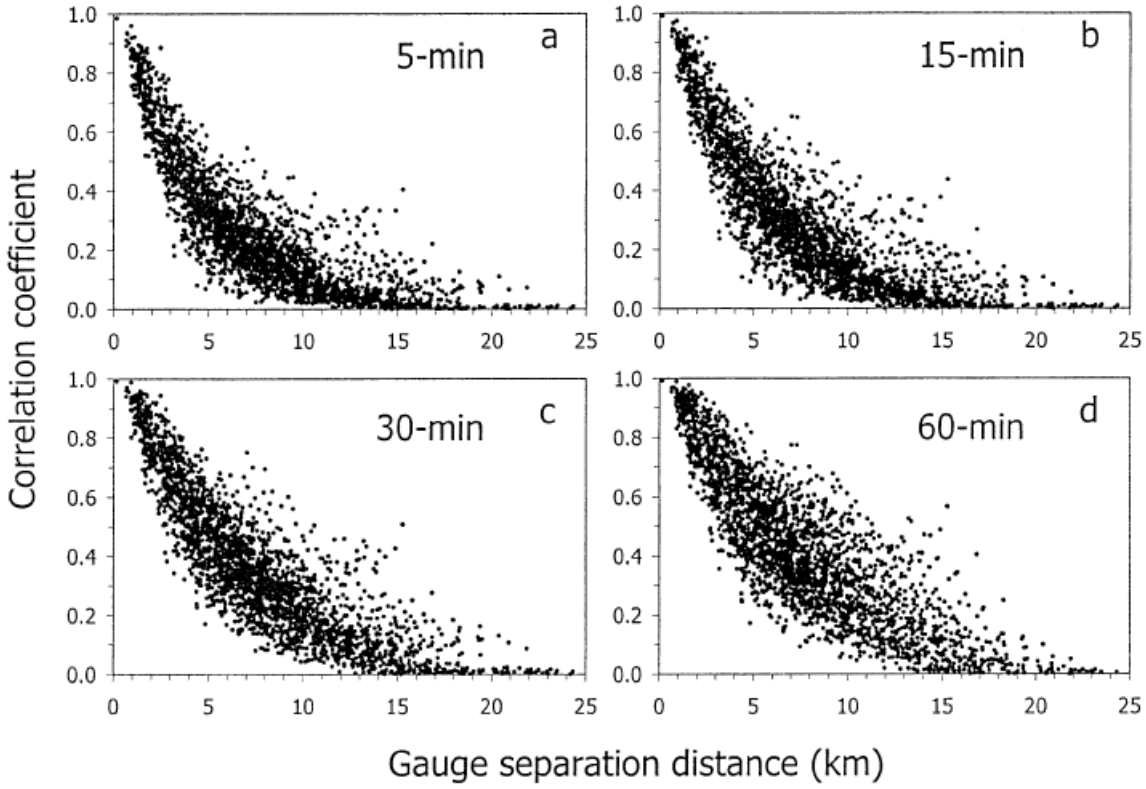


Figure 2.14 - "Correlation of gauge rain intensities as a function of gauge-separation distance. The analysis is based on the 15 storms used in this study for durations of (a) 5, (b) 15, (c) 30, and (d) 60 min" (Morin et al., 2003, p. 786)

High intensity short-lived convective rainfall storms typically have the most spatially variable rainfall distributions (CSA, 2010). This is of particularly interest because areas subject to the most severe rainfall would most likely occur between instruments. This is especially the case for observation networks with low instrument density and are subject to storm sizes that are smaller than the station spacing (CSA, 2010).

The location of instruments within a network is also important issue that may have significant implications in the interpretation of the precipitation climate (Kutiel and Kay, 1996 in Michaelides et al., 2009 ; WMO, 2010). WMO (2010) notes that a precipitation gauge spacing of at least 10 km may be needed in some regions.

Generally, the spatial variability of precipitation is much greater than other meteorological elements such as temperature and atmospheric pressure (CSA, 2010). Consequently, an adequate description of the precipitation regime requires a denser instrument network than otherwise would be needed for other weather variables (Konrad, 2001). Considering that storms are usually 15 to 30 km in extent (Gandin and Kagan, 1976 in CSA, 2010) and, that the high intensity short-lived convective rainfall storms are important events to observe, a dense rainfall observation network is needed to adequately characterize the spatial and temporal variability of these extreme events as well as their recurrence (Goodrich et al., 1995 and Wheater et al., 2000 in CSA, 2010). However, as noted by Wheater et al. (2000 in CSA, 2010, p. 18), a rain gauge spacing of 10 km might still be insufficient to adequately describe the actual "spatial heterogeneity of mesoscale or fine scale organized convective storms". WMO (2008b) recommends a minimum network density for recording precipitation gauges within an urban region of 1 gauge per 10 to 20 km^2.

For a region in Minnesota that included the Cities of Minneapolis and St-Paul, Blumenfeld et al. (2004) noted that the area of highest average annual 24-hr precipitation correlates well with the region's highest instrument density, as can be observed in Figure 2.15. According to Blumenfeld et al. (2004), the instrument density correlates well with the variance of the average annual 24-hr precipitation for up to 9 instruments per 100km^2; however, the correlation breaks down with a higher gauge density.

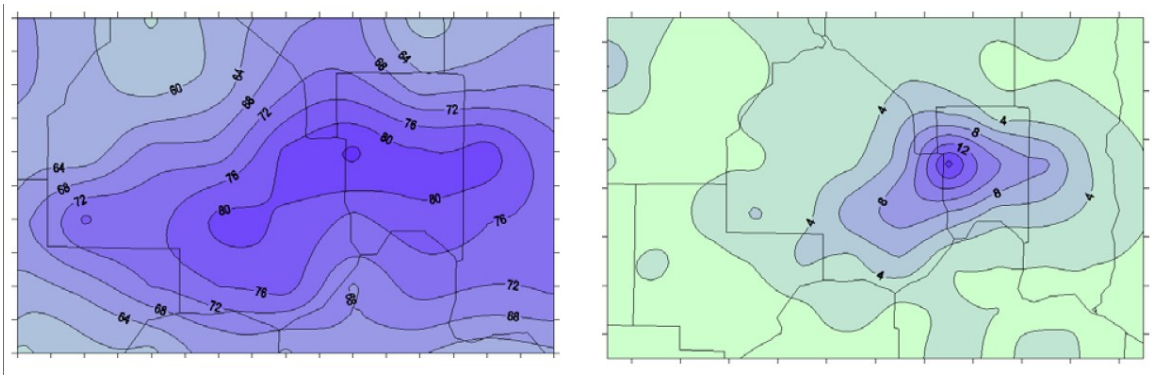


Figure 2.15 - (Left) "Mean annual daily maximum precipitation (mm)", (Right) "average observer density per 100 km^2"
(Blumenfeld et al., 2004, p. 2)

The authors surmised that their results may indicate a gauge density of up to 9 instruments per 100 km^2 is ideal for measuring event variability at the mesoscale level since it increases the probability that the event's maximum rain area is actually measured. However, the corrollary is that a greater distance between instruments leads to a lesser chance of observing severe rainfall events (Blumenfeld et al., 2004). Their results point to the fact that an instrument density greater than 9 gauges per 100 km^2 does not significantly increase the level of useful information (Blumenfeld et al., 2004). Blumenfeld et al. (2004, p. 3) also concluded that "indices such as return-period statistics", can "underestimate the true potential for heavy precipitation" when they are derived from data of a relatively coarse density. Furthermore, according to García-Pintado et al. (2009) a spacing of rain gauges in the order of several tens of kilometers is considered large and can produce a systematic under-representation of high precipitation amounts and over-representation of low amounts.

Furthermore, work by Groisman et al. (2005) as discussed in the CSA guidelines (2010) indicates that a sparse instrument network is ill-equipped to adequately describe the actual extreme events' magnitudes since the network tends to miss the portion of each storm with the most severe rainfall.

Rain gauge spacing of the order of 10 km is not adequate in describing small-sized (less than 1 km^2) intense convective storm cells that often cause significant flooding problems in urban sewersheds. In fact, Smith et al. (2007) indicated that a gauge density of no less than 1 rain gauge per km^2 is critical in describing rainfall variability. CSA guidelines (2010, p. 66) also point to network density deficiencies: *"In many regions of southern Canada, convective rainstorm events such as thunderstorms that can occur at relatively small spatial scales often result in the most intense rainfall at short time durations (e.g., 15 minutes, 30 minutes, 1 hour). Thus the spatial density of the Environment Canada recording network is insufficient to capture all of these intense storms."*

A comprehensive study by Villarini et al. (2008) studied the uncertainties associated with the use of point-source measurements for describing the spatial characteristics of rainfall. They used stations with at least six years of data form a network of 50 rain gauges over a 135 km^2 region of England. It was determined that a dense instrument network is required for generating acceptable representation of the actual rainfall. They also discovered that if a single rainfall gauge is used to describe an area's spatial regime, a complete description of the underlying statistical distribution is required (Villarini et al., 2008).

Goodrich et al. (1995), in summarizing the findings of a literature review, revealed that networks with a relatively high density of 4 rain gauges placed within a 2 m by 2 m area exhibited significant differences in measured rainfall amounts.

Michaelides et al. (2009) also points to the uneven distribution of rain gauges within existing public managed networks, the majority of which are located on land and concentrated in densely populated areas. Furthermore, the cost to build and maintain precipitation stations increases with gauge density and recent economic pressures push public agencies to close stations rather than adding new ones.

2.3.2.4.5 3D Precipitation Variability

The definition and use of the word 'spatial' in this study refers to the two-dimensional, planimetric description of a unit surface area. Although the spatial variability of rainfall has been discussed, precipitation measurements at one altitude may not be representative of other elevations due to orographic influences on rainfall for instance; thereby, pointing to a spatial variability of rainfall in the vertical dimension. This topic will be discussed further in association with weather radar data sampling of rainfall in Sections 2.3.3.5.6 through 2.3.3.5.8.

2.3.2.4.6 A Growing Consensus

A resounding theme throughout CSA's (2010) IDF Technical Guide is the insufficient spatial description in the current historical rainfall data that is used to derive long-term local and regional standards.

2.3.2.4.7 Interpolation

A variety of computational techniques are used for spatially interpolating point data sets while only a few can be used for extrapolation beyond the maximum extent of the point data-space. Kriging and co-Kriging techniques are well known techniques that can be used when more than 50 data points are available for analysis (Hrachowitz & Weiler, 2011). Interpolations using splines also need a dense set of data points (Hrachowitz & Weiler, 2011). However, in an operational context, multiple linear regressions, Thiessen polygons (Thiessen, 1911), percentage weighted polygons and, inverse distance weighted interpolations are the most commonly used approaches (Hrachowitz & Weiler, 2011). Multiple linear regressions can produce more accurate interpolated estimates for mountainous regions when used with elevation information (Hrachowitz & Weiler, 2011).

A particular challenge of interpolating point-source rainfall data is the obvious total lack of information between data points. For example, spatially propagating spikes in data values cannot readily be explained and justified; hence, spatial interpolation would be tenuous at best. In fact, the spatial interpolation of point-source data to

obtain a contiguous spatial estimate of rainfall is not a reasonable alternative to using areal averages obtained by weather radar or satellite data.

2.3.2.5 Temporally Near Continuous Measurements

Tipping-bucket rain gauge measurements are deemed as near continuous observations since the time resolution can be theoretically be recorded at very small intervals, in the order of fractions of a second. However, the normal operational resolution hinders the ability to ideally capture and, therefore, represent the true rainfall patterns. The resolution issue arises since these measurements are assumed to represent the true rainfall accumulations during the sampling period although in reality, it is only an approximation. The potentially highly variable true rates of rainfall accumulation that occur within the sampling interval is as a single constant rain rate.

Villarini et al. (2008) described the effect of temporal resolution sampling and how lower temporal resolution tends to smooth the spatial variability of measurements between instruments as a function of gauge spacing. Although a low data resolution tends to reduce uncertainty; it also results in a loss of most extreme measurements since the values of these events are averaged out over a longer period of time.

Figure 2.16 from Villarini et al. (2008) shows that the distance at which instruments decorrelate in terms of their rainfall measurements is dependent of the timescale.

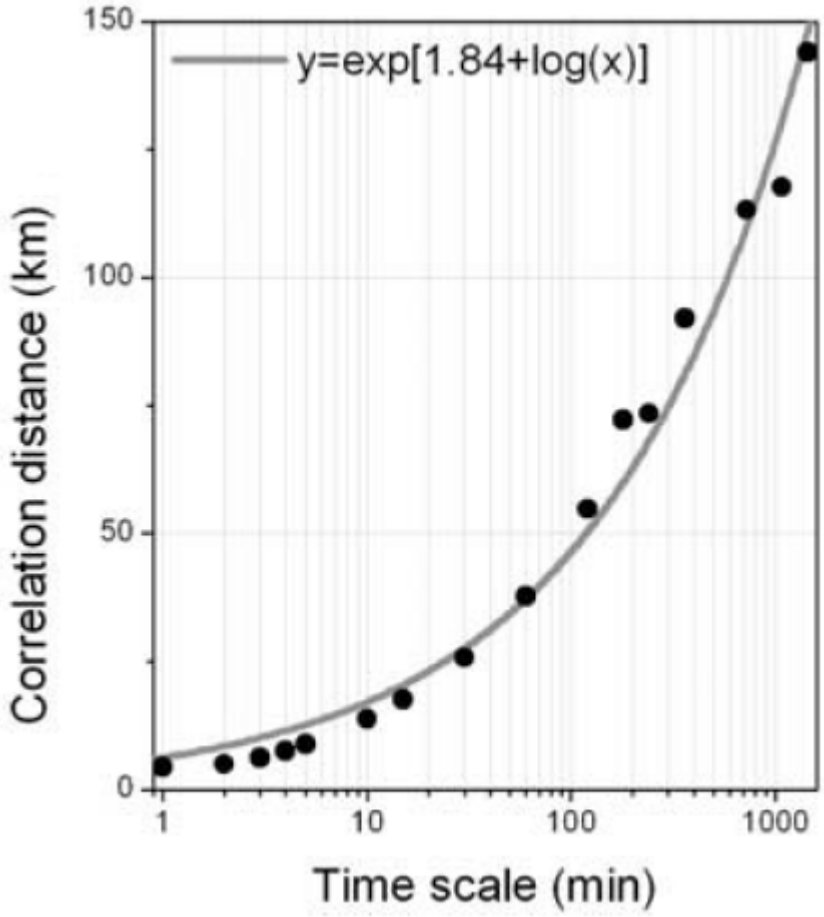


Figure 2.16 - Timescale dependence of inter-instrument decorrelation spacing distance (Villarini et al., 2008)

The work of Habib et al. (2001) used a simulated rain gauge to also present its temporal sampling limitations. The loss of information by sampling resolution due to different recording thresholds is presented in Figure 2.17 (Habib et al., 2001). However, their results show that high data uncertainty arises at sampling timescales lower than 15 minutes.

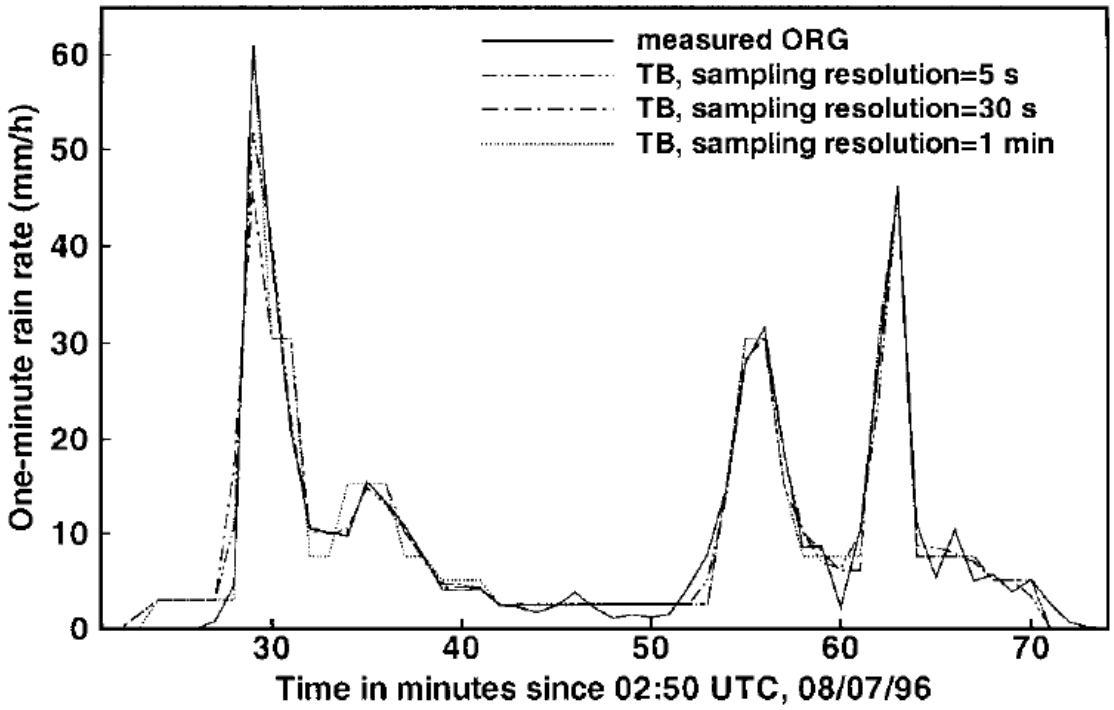


Figure 2.17 - Rain gauge temporal resolution effects (Habib et al., 2001)

2.3.3 Weather Radar

Remote sensing types of instruments enable indirect measurements of various environmental variables such as precipitation (Gourley et al., 2010). The radar technology was initially developed during World War II as a means of detecting military bombers (Skolnik, 2008). Improvements in computing capability over the years enabled the transition from analog to the digital data processing (Skolnik, 2008). This huge advancement had a significant impact on the evolution of radar technology and its application, especially since the 1970s (Skolnik, 2008). As a result, there are today many different types of radars in use covering a wide range of applications as presented by Skolnik (2008). This study, and more specifically this section, will concentrate on radars that are specifically designed for weather monitoring while focusing on the rainfall detection capabilities.

The majority of operational weather radar instruments are pulse-type radars that use linear horizontal polarization and are often equipped with Doppler sensors (Skolnik, 2008; WMO, 2008a) for wind movement detection. These instruments enable the

real-time characterization of temporal and spatial patterns of precipitation intensity, wind speed and direction that would otherwise be impossible to obtain using only point-source ground instruments (Skolnik, 2008; WMO, 2008a; Overeem et al., 2009b; Gourley et al., 2010). The relatively far range of radars also enables characterization over previously unobservable areas such as sparsely populated regions and oceans (Skolnik, 2008; WMO, 2008a; Overeem et al., 2009b; Gourley et al., 2010). This feature is particularly beneficial for obtaining data in hard to reach places and for obtaining measurements over a large region in a cost-effective manner (WMO 2008a).

Real-time radar weather information is considered most helpful in weather surveillance of synoptic and mesoscale scale meteorology that include the issuance of severe weather warnings (i.e. cyclones, tornadoes, thunderstorms, flash floods, etc) (Skolnik, 2008; WMO, 2008a; WMO, 2010). The current state of technology now allows a high degree of spatial insight on the dynamics of hydro-meteorological events, as well as significant increase in the level of spatio-temporal characterization for monitoring precipitation and wind (Skolnik, 2008; WMO, 2008a; WMO, 2010). In fact, weather radars are the sole ground instruments able to monitor severe weather over large regions in a practical manner (WMO, 2008a).

Real-time weather radar information is an operational requirement at large airports and, although this particular application was the initial justification to build the weather radar networks, radar-derived data is currently being used by several other applications such as hydrologic modeling and forensic analysis of extreme weather events (WMO, 2008a). In fact, weather radar information has now become familiar to the general public and is commonly used by weather forecasters (Skolnik, 2008).

The work by Skolnik (2008) is a highly recommended reference for radar-related information as it presents an in-depth depiction of the instruments and dedicates an entire chapter on weather related applications.

2.3.3.1 *Types of Instruments, General Specifications and Measurement Approaches*

This section reviews the types of weather radars and how they basically operate in producing rainfall estimates.

2.3.3.1.1 Types of Instruments

As was previously stated, the focus of this section is on weather radars; however, there exist other meteorologically relevant radar instruments that will briefly be mentioned here. These include Terminal Doppler Weather Radars and wind profilers. At large airports, Terminal Doppler Weather Radars are used to rapidly and continually scan the areas immediately above the runways for the presence of dangerous wind shear (i.e. microbursts, strong gust fronts and sudden wind shifts) that would jeopardize the safe departure and landing of aircrafts (Skolnik, 2008). Wind profilers are also used at airports and basically consist of vertically oriented radars that are used to characterize wind speed and direction as a function of altitude (Skolnik, 2008).

2.3.3.1.2 How the Technology Works

Most weather radars emit their electromagnetic energy in the form of pulses that travel at the speed of light using parabolic antennas (Skolnik, 2008; WMO, 2008a). Generally, when a fraction of the emitted energy encounters precipitation or differences in the atmosphere's refractivity, it will scatter and a portion of it will reflect back towards the antenna (Skolnik, 2008; WMO, 2008a; Overeem et al., 2009b). Typically, the pulse is in the order of one microsecond in duration with one millisecond pauses between each pulse to listen for returning echoes (Skolnik, 2008; WMO, 2008a). A typical weather radar station systematic scans the surrounding sky to provide estimates of precipitation above ground. Although radar antennas can be configured to complete rotations at a single vertical angle, most installations, including those within national weather networks, scan the skies at multiple vertical angles thereby providing a more comprehensive volumetric description of precipitation (He et al., 2011; WMO 2008a). Common weather radars perform a full volumetric scan cycle every 5 to 10 minutes (WMO, 2008a; WMO, 2010) that result in millions of data points.

The weather radar's beam width is nominally 1^{o} for long-range instruments (i.e. a couple hundred kilometers in range) (Skolnik, 2008; WMO, 2008a; WMO, 2008b). This empirically-derived preferred setting yields a 1 km wide beam at a 60 km distance from the instrument (Skolnik, 2008) yet forms a beam width of several kilometers in width beyond a range of 160 km (WMO, 2008b). The typical weather radar installation yields a useful measurement range of 40 to 200 km in distance (WMO, 2008a; WMO, 2008b). The spatial resolution of these installations can typically provide a reasonable description of the important spatial features embedded within thunderstorms including the areas of intense precipitation (Skolnik, 2008).

2.3.3.1.3 General Components

Figure 2.18 presents a diagram taken from Skolnik (2008) that simplistically illustrates the interaction among the main components of a weather radar system.

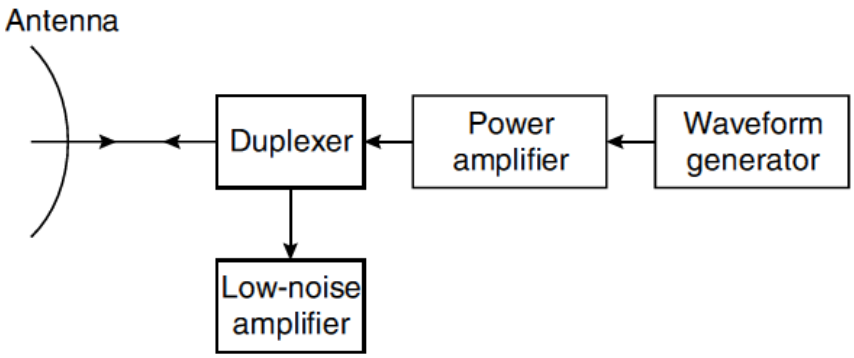


Figure 2.18 - Single antenna radar component interaction (Skolnik, 2008)

A transmitter is used to create and amplify a series of short pulses that is the preferred waveform for the use in a single antenna system in order to accommodate both transmission and reception functions (Skolnik, 2008; WMO, 2008a). The duplexer sub-system directs emitted energy to the antenna and away from the sensitive receiver component; however, it also serves to direct the reflectivity echoes to the receiver (Skolnik, 2008). The antenna focuses and emits the electromagnetic energy pulse in addition to capturing any returning echo signals (Skolnik, 2008;

WMO, 2008a). The receiver, referred to in the figure as the low-noise amplifier, increases the return echoes' signal to permit a better quantification (Skolnik, 2008; WMO, 2008a).

Operating a weather radar is a computationally intensive endeavor due to the particularly huge number of data points per scan (Skolnik, 2008). The system typically gathers over 2 million data points during each interval, each data point being the resulting average of 10 samples; hence, 20 million sample points every 5 to 6 minute (Jobin, 2012). This prodigious amount of data also requires high capacity storage and transmission capabilities in order to complete downstream analyses and archiving (Skolnik, 2008).

The design of theses systems is typically a compromise in terms of performance as it often serves many applications (WMO, 2008a). Usually, a balance is struck between the desired operational requirements in the given region's climate and other criteria such as cost (WMO, 2008a).

2.3.3.1.4 Frequency Bands of Operation

The most commonly used bands within the electromagnetic spectrum for weather radars are: S, C, X, and K that respectively corresponds to frequencies 2 to 4 GHz, 4 to 8 GHz, 8 to 12 GHz and 18 to 27 GHz (Skolnik, 2008; WMO, 2008a). The radar's sensing performance will vary according to the frequency band within which it operates (Skolnik, 2008; WMO, 2008a).

Radars that operate at low frequencies are more expensive; however, can measure at longer distances due to their 'high-power transmitters and physically large antennas' (Skolnik, 2008). Conversely, radars that operate at high frequencies more accurately resolve range and location data due to their larger bandwidths and smaller antennas (Skolnik, 2008).

Figure 2.19 presents a sketch that illustrates, for a range of radar frequencies, that the longer the emitting signal's wavelength, the greater the sensing capability (Einfalt et al., 2004).

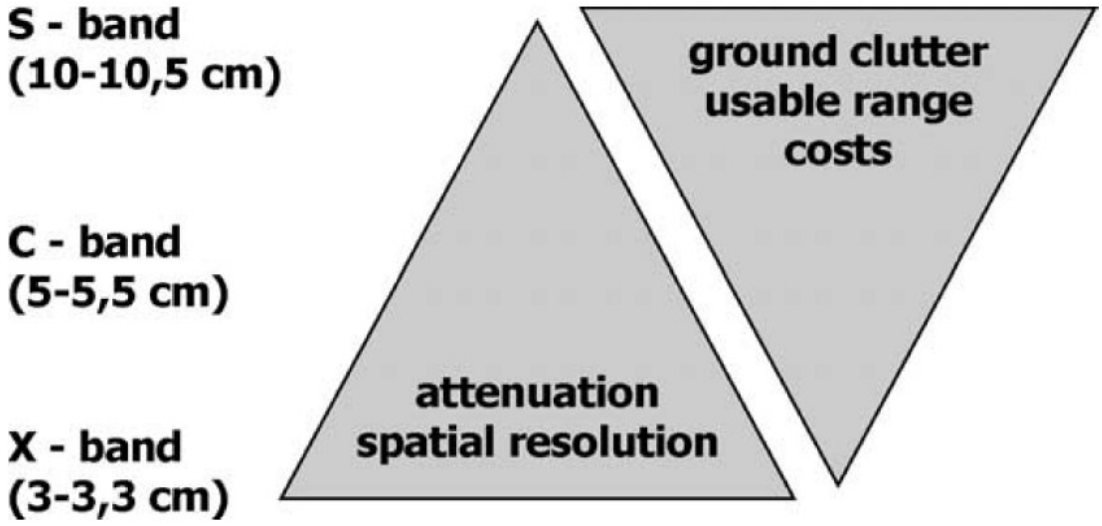


Figure 2.19 - Weather radar frequency band concerns (Einfalt et al., 2004)

The United States' meteorological monitoring radar network operates in the S-band frequency (Skolnik, 2008). Lower frequencies tend to discriminate smaller rainfall echoes while higher frequencies are known to attenuate under heavy rainfall conditions (Skolnik, 2008; WMO, 2008b). The X-band is known to have significant signal attenuation due to rainfall (Skolnik, 2008). The C-band is a compromise frequency range that lies between the S and X-band frequencies and, as expected, its performance is somewhat of an average of the two (Skolnik, 2008). Not surprisingly, the C-band tends to be the preferred frequency range for weather radar installations (WMO, 2008a). Radars operating in the K-band are range-limited and suffer from signal attenuation due to rainfall, even more so than the X-band frequency range; however, it offers better measurement resolutions (Skolnik, 2008).

The majority of weather monitoring radars currently in use for severe storm weather research and warnings are S and C-band frequency systems (Skolnik, 2008). Outside of the United States, the C-band is preferred since these systems are

typically half the cost of comparable S-band station (WMO, 2008b) and are also found to be superior at sensing light rain and snow (WMO, 2008b).

The wavelengths corresponding to the most common weather radar equipment are shown in Table 2.5 (WMO, 2008a).

Table 2.5 - **Weather radar frequencies (WMO, 2008a)**

Radar band	*Frequency*	*Wavelength*	*Nominal*
UHF	300–1 000 MHz	1–0.3 m	70 cm
L	1 000–2 000 MHz	0.3–0.15 m	20 cm
S[a]	2 000–4 000 MHz	15–7.5 cm	10 cm
C[a]	4 000–8 000 MHz	7.5–3.75 cm	5 cm
X[a]	8 000–12 500 MHz	3.75–2.4 cm	3 cm
K_u	12.5–18 GHz	2.4–1.66 cm	1.50 cm
K	18–26.5 GHz	1.66–1.13 cm	1.25 cm
K_a	26.5–40 GHz	1.13–0.75 cm	0.86 cm
W	94 GHz	0.30 cm	0.30 cm

[a] Most common weather radar bands.

A radar's detection sensitivity is proportional to its transmitting wavelength, antenna size, gain and beam-width (WMO, 2008a). The beam-width is a function of the wavelength (i.e. longer wavelengths yield wider beams) (Skolnik, 2008; WMO, 2008a; WMO, 2008b; Einfalt et al., 2004). WMO (2008a) explains that shorter wavelengths produce better sensitivity, that enables smaller beam-width and in turn, better resolution and gain. This is of particular interest when considering

measurements at greater distances from the instrument (WMO, 2008a). Nevertheless, while small wavelengths are prone to significant signal attenuation by rainfall, the cost of the radar increases proportionately to wavelength (WMO, 2008a).

2.3.3.1.5 Strength of Signal

The radar's ability to detect precipitation is proportional to the strength of the signal emitted from the radar transmitter (WMO, 2008b). There are generally two types of transmitter: klystron and magnetron (Skolnik, 2008; WMO, 2008a; WMO, 2008b) and each has distinct advantages and disadvantages that are documented in the WMO (2008b) reference. Furthermore, radars operating in lower frequencies require a greater output power to produce comparable results to that of those operating at higher frequency (WMO, 2008b).

2.3.3.1.6 Radar Pulse Duration and Rate of Repetition

The duration of the emitted electromagnetic pulse establishes the radar's ability to detect weather targets at a distance (WMO, 2008a) while a short pulse length allows a greater number of independent measurements which in turn results in greater measurement accuracies (WMO, 2008a). Ideally, the radar pulse repetition rate should be set as high as practically possible in order to gather as many readings per unit time (WMO, 2008a). In reality, a balance must be reached between maximizing pulse repetition rates and minimizing second trip echoes an intrinsic problem to all active remote sensors (WMO, 2008a).

2.3.3.1.7 Antenna, Beam-Width, Rotation Speed and Gain

Radar antennas have typically parabolic shapes and project an ever expanding cone-shaped signal (WMO, 2008a). The antenna's dimension is directly proportional to the radar's wavelength and beam-width (WMO, 2008a). A small beam-width is a desirable characteristic since it produces higher spatial resolution data - a significant benefit when considering measurements at great distances from the radar (WMO, 2008a). Also, the smaller the beam-width, the greater the antenna gain; however, at a higher instrument cost (WMO, 2008a).

A minimum amount of repeated target hits are required to produce a reasonably accurate reflectivity estimate (WMO, 2008a). This requirement places an upper limit on the antenna's rate of rotation as it must stay on target for a set amount of time (WMO, 2008a). By extension, the instrument's rate of rotation sets the amount of time that is required to perform a full scan cycle.

2.3.3.1.8 Single and Dual-Polarisation

The majority of weather radars use horizontally polarized beams that are preferred for single-polarisation systems as they minimize sea and ground clutter effects (Michaelides et al., 2009; Skolnik, 2008; WMO, 2008a). They also provide stronger echo signal from rainfall because of the droplet's shape in the horizontal plane (WMO, 2008a).

Polarimetric or, dual-polarization instruments can emit and capture electromagnetic signals on two axes: vertical and horizontal (Michaelides et al., 2009; Skolnik, 2008). These type of radar can characterize the oblate spheroid shape of rain particles and the size-velocity relation to produce better estimates of rainfall (Skolnik, 2008; Horstmeyer, 2008; Han, 2010). Also, radars using dual-polarization technologies can differentiate between various types of precipitation (i.e. hail, snow, rain, etc.) as well as produce better estimates of precipitation intensities. Furthermore, these radars can better filter erroneous signals from ground or sea clutter as well as echoes from birds and insects (Skolnik, 2008).

Unfortunately, polarimetric radars are mostly used for research although a few are now being used operationally (Skolnik, 2008). Villarini and Krajewski (2010) provide several references that describe the methods used by radars with dual-polarization to successfully overcome several known limitations of the conventional radars and; consequently, provide more accurate rainfall estimates.

The Canadian and USA government are now at the early stages of upgrading their network with dual-polarization capabilities (NOAA, 2011; EC, 2012).

2.3.3.1.9 Doppler Radar

The majority of weather radars currently in service are equipped with instruments that are capable of assessing the radial velocity (along the beam) of its target by exploiting the detected Doppler frequency shift in the echo signal (Villarini & Krajewski, 2010; Skolnik, 2008; WMO, 2008a). The Doppler frequency shift can be described by the following function (Equation 2.2):

$$f_d = -2v_r/\lambda \qquad \textbf{(2.2)}$$

where v_r is the radial velocity, λ is the carrier wavelength, which is $\lambda = c/f$, f is the signal frequency and c is the speed of light constant (Skolnik, 2008).

Weather radars with Doppler capabilities typically require transmitters that are more stable (Skolnik, 2008). Also, the processing of its signal is far more complex than analysing conventional reflectivity scans (Skolnik, 2008). However, Doppler weather radar data can be used directly to compute wind and wind shear, gust fronts and downbursts (Villarini & Krajewski, 2010; Skolnik, 2008; WMO, 2008a). Furthermore, the velocity estimates can yield invaluable insight on the kinematic structure of rainfall storms (Villarini & Krajewski, 2010).

The Doppler radar data is particularly useful when applied to signal processing of readings with a high signal noise (Skolnik, 2008). The result is a far better discrimination of stationary clutter with significant improvements in measurements (Skolnik, 2008).

In addition to measuring wind velocity, the Doppler technology can also be used to assess the target's range and reflectivity (Skolnik, 2008). Unfortunately, the signal frequencies needed for Doppler implementations impose a more restrictive limit on the useful range of the radar (WMO, 2008a).

Vertical wind profiling radars are another type of weather monitoring instrument that use Doppler radar technology (Skolnik, 2008; WMO, 2008a). As previously discussed, these instruments are used to describe the vertical wind velocity profile while some variants can be used to assess air pollution (Skolnik, 2008; WMO, 2008a).

2.3.3.2 Procedures for Estimating Rainfall

The raw data that weather radar stations actually output are measures of the reflected energy that require subsequent processing in order to generate precipitation estimates. This section will review computational techniques that accomplish this transformation.

2.3.3.2.1 Reflectivity

Rainfall estimates are produced from the strength signal of the return echo (Overeem et al., 2009b; Skolnik, 2008; WMO, 2008a). The strength of the backscatter signal, when considering point targets, is proportional to the target's distance to the radar as well as the volume of the radar cross section (Skolnik, 2008). However, for spatially distributed targets, as is the case for radar weather observations, a more elaborate description is needed and includes the precipitation particle density, the size and phase (Skolnik, 2008; WMO, 2008a). Figure 2.20 illustrates the elements of the weather radar scanning and detection processes that will be discussed in the following sections.

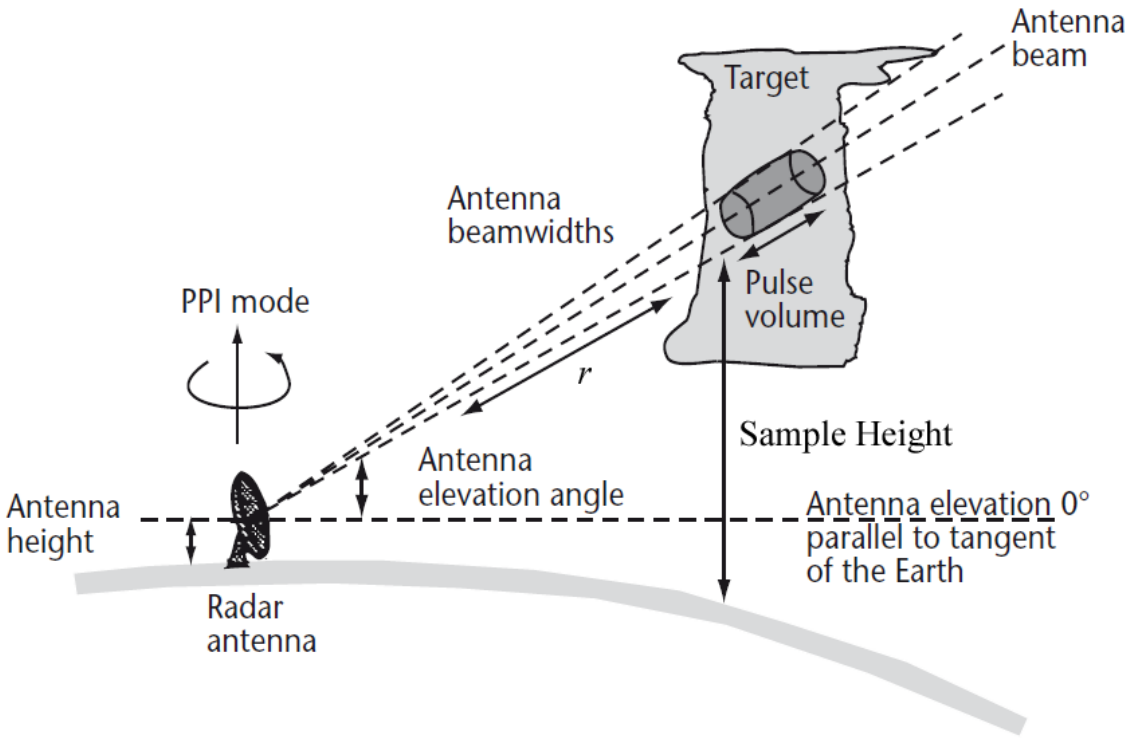


Figure 2.20 - Illustration of a radar weather target scan (WMO, 2008a)

The radar pulse echo signal shape can be described as a Gaussian distribution due to the larger number of individual targets - greater than 1,000,000 water droplets (Skolnik, 2008). The echo signal strength, also known as reflectivity, is represented by the surface under the curve as given by the probability density function whose mean value distribution is exponential (Skolnik, 2008).

The signal strength P_r can be expressed by Equation 2.3 (Probert-Jones, 1962 in Skinner et al., 2009; Skolnik, 2008 and WMO, 2008a):

$$P_r = \frac{P_t G^2 \lambda^2 \theta \phi c \tau}{512(2\ln 2)\pi^2 r^2} \sum_{i=1}^{N} \sigma_i \qquad \textbf{(2.3)}$$

where P_t is the peak transmit power, G is the gain from the antenna system, λ is the wavelength, θ is the azimuth of the beam-width, Φ is the elevation of the beam-width, c is the speed of light, τ is the radar pulse width, N represents the number of

scatterers per unit volume, σ_i is the backscatter cross section, 2ln2 accounts for the signal's Gaussian shape, and r is the range to the target.

From the signal strength, the weather-related target reflectivity can be estimated by Equation 2.4 (Skolnik, 2008; WMO, 2008b):

$$Z = C + P_r + 20\log r \qquad \textbf{(2.4)}$$

where Z is reflectivity in dBZ (also known as decibels, described as mm^6/m^3), C is the Weather Radar Constant and P_r as well as r are identical to the previously defined variables of Equation 2.3 (Skolnik, 2008). The unit of measure for the radar reflectivity actually represent a volume density for rain drops in terms of a diameter within a unit volume and expressed as the mm^6/m^3 (Skolnik, 2008; WMO, 2008a; Wilson & Brandes, 1979). Refer to Skolnik (2008) for more information on the derivation of the reflectivity equation. Several hypotheses are incorporated in Equation 2.4, including beam filling, the validity of the Rayleigh approximation for spheres and, water particles are either water liquid or solid. Although these might not be satisfied at all times, Z_e, the effective radar reflectivity, is used when the assumptions are assumed to be satisfied (Skolnik, 2008; WMO, 2008a; WMO, 2008b). Also, the effective reflectivity is assumed to represent the average reflectivity value over the radar cross section in non-ideal conditions and is also identified as the effective reflectivity factor (Skolnik, 2008).

Typical weather targets of interest have reflectivity signal strengths in the range of -20 to 70 dBZ while non-precipitation clouds return signals as low as -40 dBZ (Skolnik, 2008). Rainfall rates of extreme rainfall events that typically cause severe flooding have echo signals in the range of 55 to 60 dBZ, while those during hailstorms can return reflectivity values above 70 dBZ (Skolnik, 2008).

2.3.3.2.2 Conversion of Reflectivity Into Rainfall Estimates

The conventional approach in deriving radar rainfall intensity estimates from the detected radar reflectivity data is to apply one of several empirically-derived conversion functions (WMO, 2008a). Many of these empirical relations link radar signal reflectivity to rainfall intensity in the following form (Equation 2.5):

$$Z = aR^b \quad \textbf{(2.5)}$$

with a and b as constants and R the rainfall intensity (mm/hr) (WMO, 2008a; WMO, 2008b; Einfalt et al., 2004; Morin et al., 2003; Wilson & Brandes, 1979; Atlas, 1964). The equation's constants are related to the rainfall drop size distribution and drop density as it can be described in terms of the fourth power rain drop diameter (Martner & Dubovskiy, 2005 in Parzybok et al., 2010; WMO, 2008a; Skolnik, 2008; Wilson & Brandes, 1979). A plethora of variants of this equation can be found in literature and are based on site specific research across the world representing different regions, weather conditions, type of storms and seasons (Villarini & Krajewski, 2010; Skolnik, 2008; WMO, 2008a; Wilson & Brandes, 1979). Skolnik (2008) pointed to the fact that the absence of a universal, "all-weather" equation, results in numerous conversion functions. This is not entirely unexpected considering the high variability in rainfall particle size distributions. Rainfall storm conditions are temporally and spatially dynamic, differing between storm cells but also throughout each storm cell but also changing in time depending on the type of storm (Lee et al., 2009; Einfalt et al., 2004; Morin et al., 2003).

Battan (1973 in Skolnik, 2008) presented the following set of four relations that are commonly used for specific rainfall conditions:

1) Equation 2.6 is best for stratiform rainfall, which is a large scale and mostly homogeneous type of rainfall. Known as the Marshall-Palmer (1948) and Marshall et al. (1955) equation, it is among the most frequently implemented equation (Villarini & Krajewski, 2010; Skolnik, 2008; Wilson & Brandes, 1979).

$$Z = 200R^{1.6} \quad \textbf{(2.6)}$$

2) Blanchard (1953) proposed the following relation (Equation 2.7) that is most representative of orographic rainfall, a type of rainfall caused by geographic lifting.

$$Z = 31R^{1.71} \quad \textbf{(2.7)}$$

3) Equation 2.8 is recommended for use during convective storm cells which is typically associated with thunderstorm rainfall (Jones, 1955).

$$Z = 486R^{1.37} \quad \textbf{(2.8)}$$

4) Snowfall intensity can be estimated with the use of Equation 2.9 proposed by Gunn and Marshall (1958). Note that the precipitation intensity for this equation is expressed as water-equivalent.

$$Z = 2000R^{2} \quad \textbf{(2.9)}$$

Operational rainfall estimates within the continental network of the U.S.A. NEXRAD weather radar system are obtained by using Equation 2.10, a default relation for convective rain storms (Krajewski et al., 2010; Parzybok et al., 2010; Skolnik, 2008; OFCM, 2006; Einfalt et al., 2004; Morin et al., 2003; Fulton et al., 1998). However, other relations are also used depending on the particular weather conditions (Parzybok et al., 2010; Morin et al., 2003; Fulton et al., 1998).

$$Z = 300R^{1.4} \quad \textbf{(2.10)}$$

2.3.3.2.3 CAPPI and Other Products

As previously mentioned, weather radars continuously complete volume scans of the sky surrounding the radar, as shown in Figure 2.20 (WMO, 2008a). It does this by rotating about the vertical axis and incrementally changing the instrument's vertical angle after each complete rotation (WMO, 2008a).

As previously discussed, the amount of data produced by weather radars is computationally taxing and requires efficient strategies for analysis and visualization. A simple approach for quick visualization and weather monitoring of precipitation is to use Plan Position Indicators (PPI) which effectively presents the radar data of only a complete sweep at a single vertical angle (WMO, 2008a). Data presented in such a manner have observations closest to the instrument originating from lower altitudes than those farther away. This approach has become increasingly unpopular due to ground clutter near the instrument during scans at low vertical angles (Crozier et al., 1991).

A second approach is to produce a horizontal cross-section at a fixed altitude across the entire volume scan. This technique produces a Constant Altitude Plan Position Indicator or, a CAPPI arrangement (WMO, 2008a; Atlas, 1964). This is achieved by using radar data points on PPIs that are closest to the desired altitude as was first elaborated by Langleben (1956) and later refined by Marshall (1957 according to Marshall & Ballantyne, 1975 and Atlas, 1964). Figure 2.21 shows the CAPPI data arrangement for different altitudes including the various vertical angles used in building a given CAPPI. The bold, offset parallelograms along the radar signal angles represent the beam width. Note that the dotted elevation lines curved to account for the Earth's surface curvature. Weather radar precipitation data presented in a CAPPI format is commonly used for extreme storms monitoring and identification (WMO, 2008a).

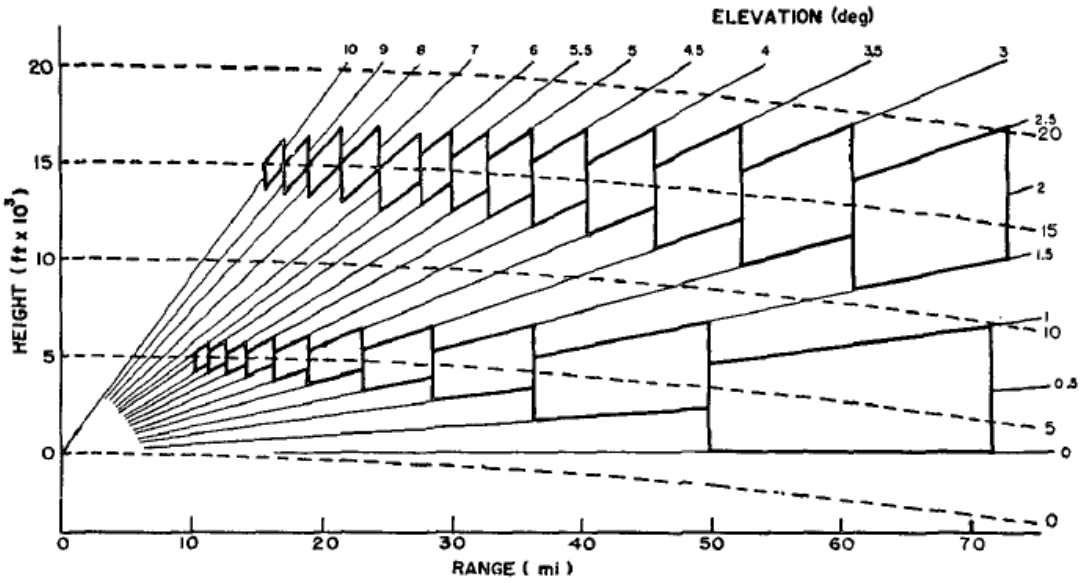


Figure 2.21 - Weather radar CAPPI data presentation (Langleben, 1956)

In order to obtain the most representative surface precipitation of, ground catch, it is desirable to use the lowest possible CAPPI altitude that is clutter-free in the area of interest (Fulton et al., 1998; Crozier et al., 1991). The lowest usable altitude is governed by the lowest available vertical angle a given radar station can use without encountering ground related impacts along the length of its beam. These impacts include partial of full beam blockage due to the intersection with the ground of the forest canopy. Increased reflectivity is also an issue at lower beam angles due to reflections off tall buildings in urban centers. Figure 2.21 shows that this issue becomes a bigger concern at greater distances to the radar since the sampling volume size increase along with the observing altitude due in part to the Earth's curvature (WMO, 2008a; Fulton et al., 1998; Crozier et al., 1991). In fact, the regional topography can significantly constrain the use of particularly low vertical angles, to a point where CAPPIs may be impractical (Fulton et al., 1998).

Note that the typical CAPPI images are not produced for elevations lower than 1 km in altitude while 1.5 and 2.5 km are commonly used (Crozier et al., 1991).

Radars with a limited number of useful tilt angles are restricted to using a pseudo-CAPPI approach and consequently result in a mix of data taken from different altitude (e.g. Bouilloud et al. (2010)).

2.3.3.3 Canadian Weather Radars

This research uses data from EC's network of radars; hence, it was deemed appropriate to review in more detail its specific characteristics.

2.3.3.3.1 Canadian Weather Radar Network

Additions, modifications and enhancements to EC's weather radar network, completed in 2004, enable the Canadian air space to be monitored by a network of 31 weather radars (EC, 2012; EC, 2011b; EC, 2011c; Zhang et al., 2008; OAGC, 2008; Joe & Lapczak, 2002) as shown in Figure 2.22. The large circles show the extent of each radar's conventional reflectivity scan while the points are the actual radar sites. Twenty-eight of the instruments are owned and operated by EC, while the National Defence has two and the McGill University has one (EC, 2012).

Figure 2.22 - The Canadian weather radar network (OAGC, 2008)

The layout of the network is such that it covers the most populated areas in Canada and also happens to lie along the Canada/U.S.A border. This provides the weather radar network with an effective coverage of over 95% of the population (EC, 2011b; EC, 2011c). Furthermore the network scans reach into the U.S.A while the U.S.A. network scans into Canada, thus resulting in a seamless cross-border weather monitoring and coverage redundancy (Zhang et al., 2008; Joe et al., 2002; Joe & Lapczak, 2002).

2.3.3.3.2 Canadian Weather Radar Instrument Recent Historic Capabilities

Several radars within the current network have been operational for 30 to 40 years (EC, 2012). Prior to the relatively recent network additions and upgrades of 1997 to 2004 (OAGC, 2008; Joe et al., 2002; Joe & Lapczak, 2002), the Canadian territory had a heterogeneous mix of instruments with varying capabilities and operational settings scattered over the country and operating as independent stations (Joe & Lapczak, 2002). The major network enhancements of 1997 to 2004 included upgrades to the then-twenty radar network (Joe et al., 2002) but most notably, the addition of 10 new Doppler radars as well as the retrofit with Doppler capabilities to sixteen existing radars (Joe et al., 2002; Joe & Lapczak, 2002). The network changes also included: 1) the strategic relocation of many existing radars (Joe & Lapczak, 2002), 2) software development and modifications (Joe et al., 2002; Joe & Lapczak, 2002) and, 3) implementation of a new logistic platform for the system (Joe & Lapczak, 2002). One of the goals of the upgrade was to group the weather radars into a homogenous network of similar instrument capabilities, all operating under a unified system (Joe & Lapczak, 2002). More detailed information on the network upgrades can be obtained in the following two references: Joe et al. (2002) and, Joe and Lapczak (2002).

The current reach of Canadian weather radar's conventional reflectivity scans extends to 256 km from the radar with a nominal resolution of 1^{o} x 1km (Zhang et al., 2008). The Doppler range is limited to 128km from the radar (EC, 2011b) and has a nominal resolution of either 0.5^{o} x 1 km or 1^{o} x 1 km depending on the scan cycle (Joe & Lapczak, 2002). All Canadian weather radars operate in the C-band,

with the exception of the McGill University radar, which operates in the S-band (Lee et al., 2009; Zhang et al., 2008). The McGill and King City radars are fitted with dual-polarization capabilities and considered as research instruments while still operating within the Canadian weather radar network (EC, 2012; Lee et al., 2009; Joe et al., 2002; Joe & Lapczak, 2002; Crozier et al., 1991). All other radar instruments operate with a horizontally polarized beam (Joe & Lapczak, 2002). However, the Canadian federal government have recently announced plans for upgrading the existing weather radar network with dual-polarization capabilities within the next 5 years (EC, 2012).

2.3.3.3.3 Canadian Weather Radar Scan Strategy and CAPPI Product

The radar scanning strategy, as elaborated by Marshall and Ballantyne (1978), has been in operational use for over 30 years (Joe & Lapczak, 2002). EC uses a different weather radar scanning strategy in the summer than in the winter for both its conventional reflectivity scan and the Doppler scans (Zhang et al., 2008). Canadian weather radar instruments can perform conventional reflectivity scans at up to 24 vertical angles as shown in Figure 2.23 (Zhang et al., 2008; Joe & Lapczak, 2002). The figure also shows the angle-range data combinations used in deriving CAPPIs for two standard altitudes. The bold lines that zig-zag across the beams illustrate the data locations that form the 1.5 and 4 km CAPPIs. Note that the radar beams in this figure are show as curves with respect to the Earth's surface (EC, 2011b). The lowest CAPPI that is typically used to provide the best representation of precipitation at ground-level is from a 1.5 km altitude in the summer and 1 km during the winter (EC, 2011b; Crozier et al., 1991). The standard Canadian weather radar CAPPI product from conventional reflectivity data is produced in a radial grid pattern (i.e. nominally at 1^{o} x 1km) (e.g. Islam & Rasmussen, 2008). However, this is typically transformed into a 1 km^2 Cartesian grid that is aligned to the Universal Transverse Mercator (UTM) projection for ease of interpretation and analysis (e.g. Islam & Rasmussen, 2008).

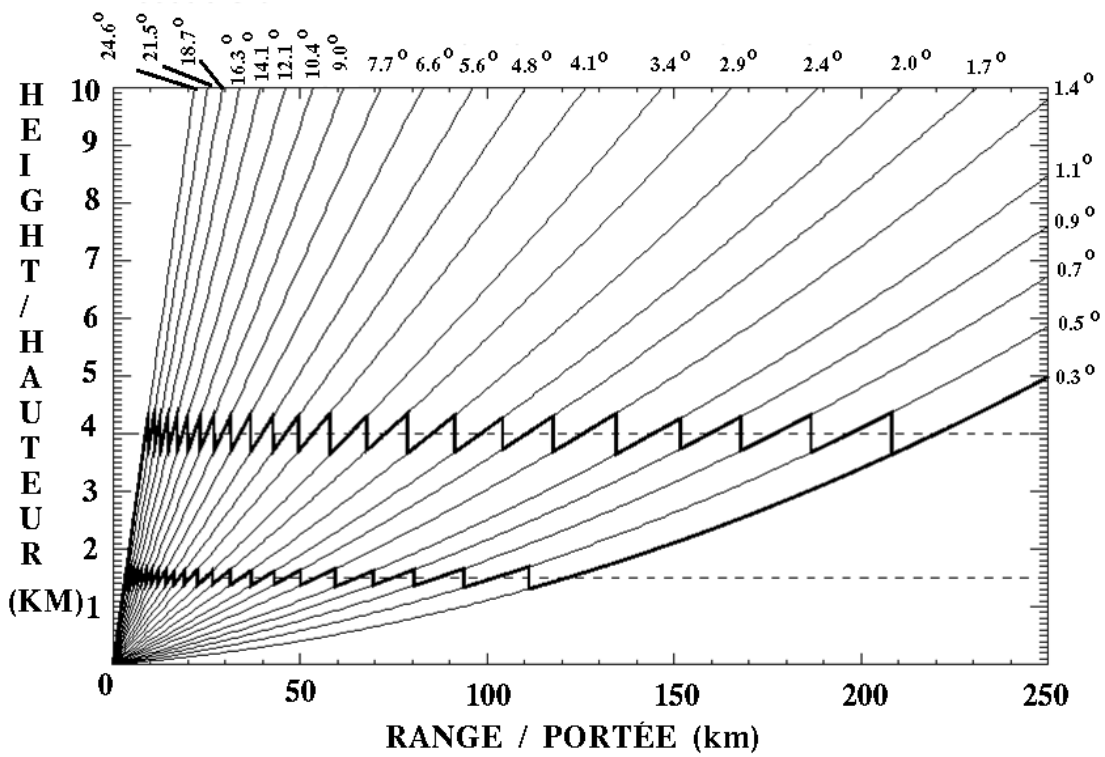


Figure 2.23 - EC weather radar CAPPI (Wikipedia, 2009)

The radars generally alternate between a 5-minute conventional reflectivity scan and a 5-minute Doppler scan (Joe et al., 2002). However, before the 1997 to 2004 network upgrade, instruments without Doppler capabilities would simply perform conventional reflectivity scans every 5 minutes (Joe et al., 2002). The McGill radar was the exception in the older system and was able to obtain both reflectivity and Doppler scans at the same time, every 5 minutes (Lee et al., 2009). Further information on Canadian weather radar scan strategies, instrument capabilities and specifications can be found in Joe and Lapczak (2002).

2.3.3.4 *Spatially Contiguous Measurements*

A far better spatial description of the actual rainfall regime is now possible with the use of meteorological radar data as compared to point source rain gauge data. This allows shifting analyses towards more accurate representations of extreme rainfall occurrences; from point-source measurements of storms to complete spatio-temporal descriptions of entire storm cells, including information on the maximum

total rainfall occurring anywhere within very large surface areas (Parzybok et al., 2010). Nevertheless, this new approach of analyzing precipitation information is still constrained by the technology's resolution limits that on average can provide rainfall estimates within a unit surface area of approximately 1^{o} by 1km.

The increased spatio-temporal accuracy plus the availability of contiguous measurements over large areas has eliminated the need to assume rainfall as uniform over a watershed; thereby, greatly providing a significant improvement in the precision and reliability of hydrologic analyses (Parzybok et al., 2010). Instead of only analyzing rainfall statistics gathered over only a tiny portion of the storm's total extent, as is the case with rain gauge data, weather radar observations enable the analysis of the full extent of rainfall storms (Parzybok et al., 2010).

The spatial density of weather radars derived precipitation estimates far surpasses that of any rain gauge network (Parzybok et al., 2010). As a result, events that might otherwise escape detection by rain gauge networks are likely never missed by weather radars thereby increasing the total number of storm characterizations (Parzybok et al., 2010).

Weather radars were found to offer the highest likelihood of detecting extreme rainfall while rain gauge provided the lowest likelihood of detection when comparing rain gauge, satellite and radar data (Gourley et al., 2010). This is due to the fact that the area with the largest total rainfall within any given storm seldom occurs, if ever, at the exact spot where the rain gauges are installed (Gourley et al., 2010). Figure 2.24 presents the results of this analysis (rain gauge in black and radar in red) whereas the green and blue lines represent satellite-based rain detection data products.

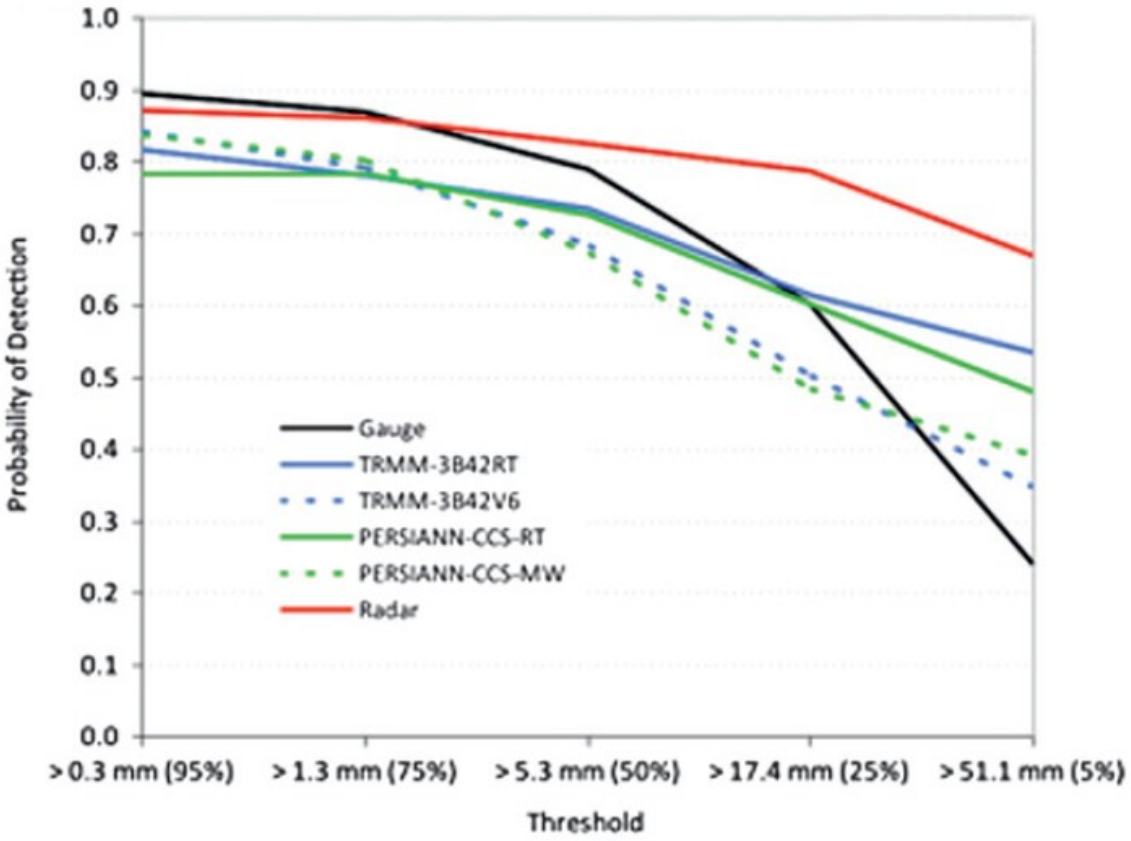


Figure 2.24 - Probability of detection for satellite, rain gauge and radar (Gourley et al., 2010)

The weather radar's ability to spatially characterize the entire rainfall event with time gives it the capacity to represent and follow in great detail the genesis of different types of storms (e.g. convective and tropical storms), something that a rain gauge network simply cannot do (Skinner et al., 2009).

Note that as is the case for rain gauge measurements, radar rainfall estimates also represent the mean value of rainfall intensity over the observed region.

Konrad (2001) argues that the spatial extent of a significant precipitation event is important since it provides an indication of its flooding potential. Also, in most cases, larger rainfall events produce increasingly greater spatial variations (Konrad, 2001).

2.3.3.5 Radar Measurement Errors and Limitations

Although weather radar technology offers significant benefits for rainfall characterization, it unfortunately has several inherent limitations that can result in large measurement uncertainties (Villarini & Krajewski, 2010). The following sub-sections present a summary of the most significant limitations, excluding those associated with instrumentation issues (hardware). Villarini and Krajewski (2010) prepared an extensive review of known causes of uncertainty in rainfall estimates from conventional weather radars. The level of measurement uncertainty for weather radars is not as well-known as that for rain gauges.

2.3.3.5.1 Signal Attenuation

As previously discussed, signal attenuation is a major concern for weather radars operating in the higher frequency bands (i.e. C-, X- and K-band) (Villarini & Krajewski, 2010; Skolnik, 2008; WMO, 2008a; WMO, 2008b; Einfalt et al., 2004). Weather radar signal attenuation can occur due to a wet radome, significant intervening precipitation, atmospheric gasses, and clouds (Overeem et al., 2009b; Battan, 1973 in Skinner et al., 2009; WMO, 2008a; Ryzhkov & Zrnic, 1995 in Fulton et al. 1998; Wilson and Brandes, 1979).

2.3.3.5.2 ***Ground Clutter and Other Sources of Error***

Ground clutter occurs as a result of the radar beam reflecting off ground objects in its path (e.g. buildings, topography, trees) (Villarini & Krajewski, 2010). In the lower atmosphere, airplanes, dust, birds and insects can be detected by the weather radar, the latter two objects often produce reflectivity ranging from -20 to 20 dBZ (Skolnik, 2008; WMO, 2008a). Also, atmospheric temperature inversions and large humidity gradients can influence the signal's reflectivity (WMO, 2008a). Sometimes, signal echoes from highly reflective targets beyond the scanning range of the radar is erroneously considered within the scanning range (WMO, 2008a). Several approaches have been proposed to minimize the impacts from these issues and Villarini and Krajewski (2010) provides a list of several available techniques. The use of a CAPPI radar data format for example helps minimize ground clutter effects

(Krajewski et al., 2011). However, the effects of clutter and other similar errors cannot be removed entirely (Villarini & Krajewski, 2010). Also, atmospheric beam bending due to atmospheric conditions can further accentuate ground clutter issues (Villarini & Krajewski, 2010).

2.3.3.5.3 Beam Blockage

Obstacles such as mountains and tall buildings near the weather radar can occult the radar signal, in part or completely, such that the radar detects only a fraction or none of the rainfall targets beyond the obstacle and; consequently, under-estimates the actual precipitation intensity (Villarini & Krajewski, 2010; WMO, 2008a).

When radar signals are blocked, it often becomes the most important source of error (WMO, 2008a). Blocked beams dictate the lowest usable radar angles (Fulton et al., 1998). This source of error is a particular concern in mountainous areas (Villarini & Krajewski, 2010).

2.3.3.5.4 Variability in the Reflectivity to Rainfall Conversion Relation

As presented in Section 2.3.3.2.2, there are many functions to convert radar reflectivity data into rainfall intensity estimates. This is in part due to the significant variability in the distribution of rainfall particles and how reflectivity values, by themselves, are insufficient in fully describing that variability (Villarini & Krajewski, 2010; Skolnik, 2008). Remember that the actual drop size, shape, distribution, density and phase, along with other characteristics of the sampled precipitation field are all inferred by reflectivity values. The conversion into precipitation rates through the use of any given power relation, significantly limits the ability to characterize the precipitation (Michaelides et al., 2009; WMO, 2008a).

Furthermore, hail in the scanning field is a common problem as its reflectivity signature is not discernible from extreme rainfall intensities and leads to erroneous water equivalent estimates (Villarini & Krajewski, 2010). Its presence is typically revealed by abnormally high reflectivity values and consequently high precipitation rates, if left uncorrected. Hail can typically have reflectivity values over 70 dBZ

(Skolnik, 2008). Usually, an upper limit on rainfall intensity is imposed on the converted weather radar data to mitigate the effects of hail contamination within the reflectivity data (e.g. Overeem et al., 2009). The threshold value is usually chosen as representative of a low probability, extreme value rainfall intensity. Overeem et al. (2009b) chose to use a threshold of 55 dBZ, which roughly represents a rainfall intensity of 100 mm/h, an intensity that has a 6.6 year recurrence at any radar pixel over the Netherlands.

2.3.3.5.5 Range Effects/Issues

Range related issues are intrinsically tied to weather radar scanning geometries as illustrated in Figures 2.20, 2.21 and 2.23. The sampling volume along any given beam increases with range while the lowest scanning altitude is limited by the Earth's curved surface (Villarini & Krajewski, 2010; WMO, 2008a).

At farther ranges (i.e. approximately 250 km), the scanning altitude of the lowest available radar beam is too high to be considered representative of precipitation at ground level. It can in fact overshoot low-lying clouds altogether (Villarini & Krajewski, 2010; WMO, 2008a; Einfalt et al., 2004).

Furthermore, as the beam widens with range, it reduces the sampling resolution (WMO, 2008a). It also makes the assumption of a homogeneous scanning surface increasingly less likely with range (WMO, 2008a), thus underestimating precipitation (Villarini & Krajewski, 2010). The reader is referred to Section 2.3.3.5.8 for more information on sub-sample variability.

2.3.3.5.6 Vertical Variability in Atmospheric Precipitation

The variation in precipitation rates with altitude is potentially a significant issue when the objective is to calculate the most representative ground-level values (WMO, 2008a). This vertical variability issue also becomes a concern at extented ranges of the beam because of the increasing sampling volume that potentially augments the samples' heterogeneity (Villarini & Krajewski, 2010).

Vertical variability of atmospheric precipitation is a result of particle collision, merger as well as mid-air splitting of rainfall particles (Villarini & Krajewski, 2010). Also, detected atmospheric precipitation is known to evaporate on its way to the ground (Villarini & Krajewski, 2010; Doviak and Zrnic 1993 in Skinner et al., 2009).

It is a well known fact that air temperatures varies with altitude and also exhibits seasonality. Figure 2.25 compares the vertical profile of rain (solid line) versus that of snow, shown as a dashed line (Koistinen et al., 2003 in WMO, 2008b) for two particular dates. The sudden spike of reflectivity (~7dBZ) for the atmospheric rainfall just above 1 km in altitude is due to the freezing level at that time. For an example of seasonal variability, consider the atmospheric freezing elevation over Sydney, Australia as shown in Figure 2.26 (Chumchean et al., 2003). However, note that winter in Australia is from June to August; hence, the period when the atmospheric freezing levels reach their lowest altitudes.

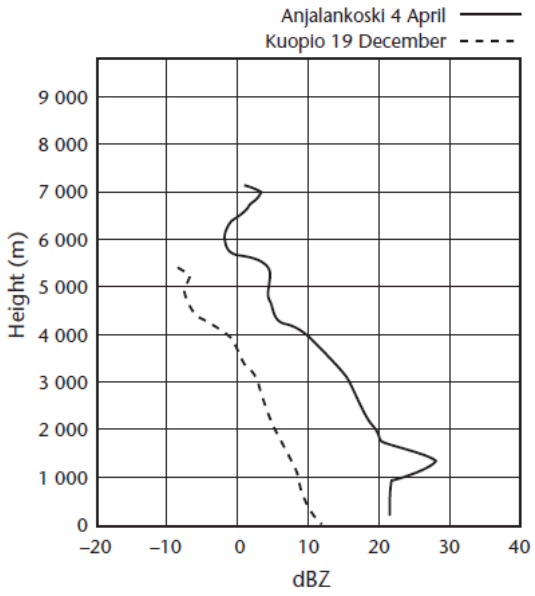


Figure 2.25 - Comparative vertical profile of rain and snow (WMO, 2008b)

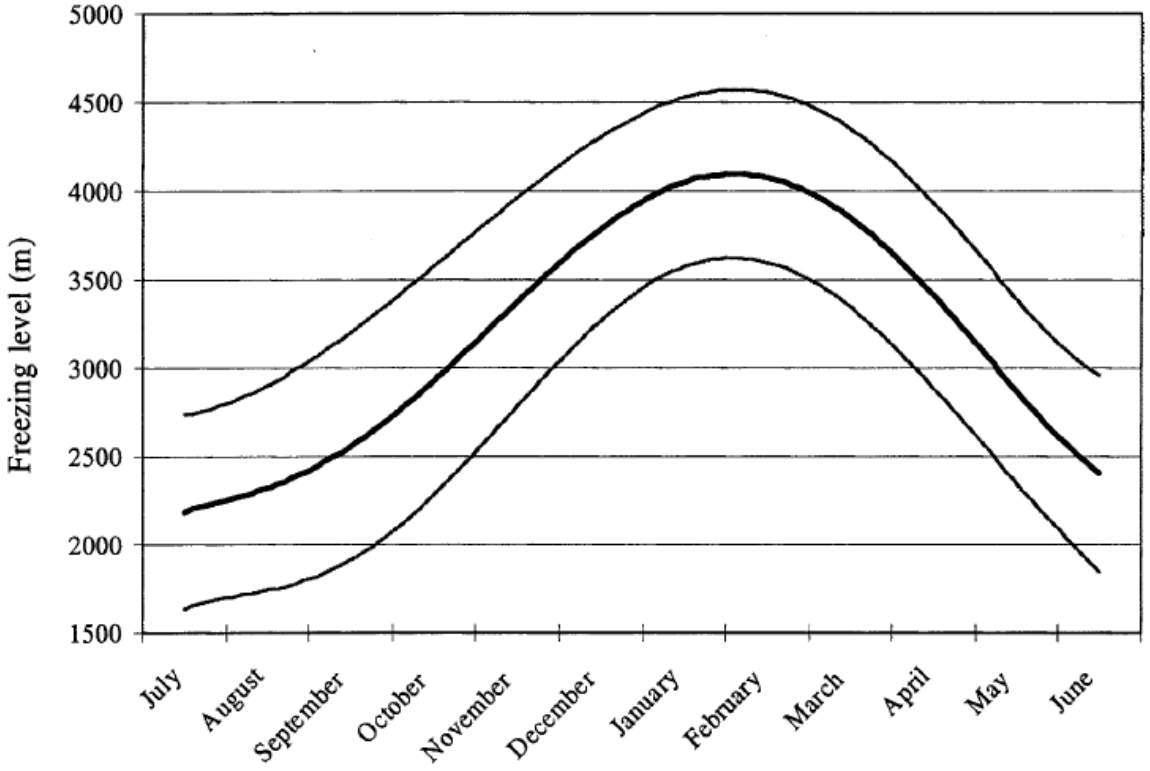


Figure 2.26 - **Atmospheric freezing elevation over Sydney, Australia (Chumchean et al., 2003)**

The freezing level is often referred to as 'bright band' in the weather radar literature due to its noticeable increase in local reflectivity. These bright bands are typically found in stratiform storms (Villarini & Krajewski, 2010). The overestimation in precipitation rate of bright bands is usually found to be in the orders 2 to 5 times the true water-equivalent rate and can reach at high as 10 times (Villarini & Krajewski, 2010). Villarini and Krajewski (2010) lists approaches that have been used to mitigate bright bands effects. Note that bright bands appear as ring patterns when radar precipitation data is presented in a CAPPI format. This occurs when the CAPPI altitude is near the bright band elevation and the scans of every vertical position repeatedly traverse this zone. A resulting ring pattern is created in the CAPPI products due to the data mapping strategy of this particular format. Reflectivity data points are successively selected below-within-above the bright band elevation, as presented in Figures 2.21 and 2.23.

2.3.3.5.7 Wind Drift and Vertical Air Movement

Wind drift and vertical air movement can be an issue when precipitation estimates derived from weather radar reflectivity data are used for ground-level applications. For instance, ground level precipitation is of particular interest for design and operation of water resource infrastructure (CSA, 2010). One particular issue is that atmospheric precipitation does not always reach the ground (CSA, 2010; Villarini & Krajewski, 2010). Additionally, atmospheric precipitation estimates from general reflectivity radar data does not reveal the vertical movement of precipitation (Michaelides et al., 2009); when, important upward air movements do occur (Michaelides et al., 2009), notably in convective storms.

Furthermore, as air masses move along the surface and through the atmosphere, the precipitation that is generated at altitude can continue to carry some of this lateral momentum on the way to the ground. This typically results in a lateral ground-level drift of the precipitation estimates observed by radars.

2.3.3.5.8 Sub-Pixel Variability

As previously discussed, weather radars provide a spatial average precipitation intensity value over its scanning volume of typically 1^{o} by 1 km. In terms of the spatial rain drop distribution, the averaging effect within the unit scanning volume, or a CAPPI pixel, becomes of particular interest as the spatial sampling area increases. Notice that in the example below (Figure 2.27), each of the four rainfall intensity value distributions has an identical average intensity; hence, would be described by the same numeric pixel value. Although they depict significantly different storm cells the radar reflectivities would be the same.

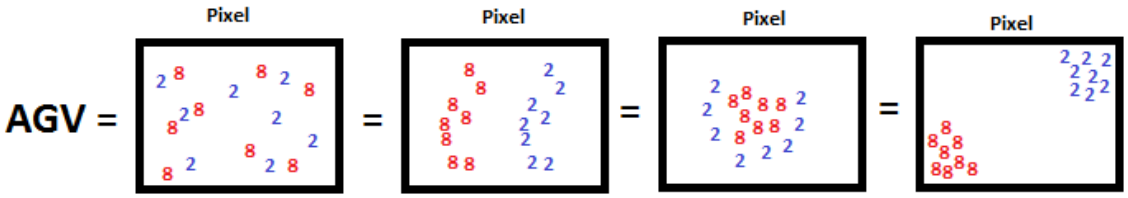


Figure 2.27 - Example of possible hidden value distribution within a single spatial sample due to averaging

Sub-pixel variability is essentially lost since the radar rainfall estimates represent an average rainfall intensity over the sampled area. For generally homogeneous rainfall intensities, little information is typically lost (e.g. stratiform rainfall events). However, storm cells with large spatial variability of rainfall intensities (e.g. convective storms), can have portions of their pixels experience no rainfall while other portions extremely high intensities. The resulting sampling area will average out the differences (Villarini et al., 2008); hence, radar-derived precipitation estimates typically underestimates the actual maximum storm intensities. The difference in spatial sampling scales between rain gauge and weather radar may help in quantifying this phenomenon.

Villarini et al. (2008) analyzed this sub-radar pixel variability issue by comparing 6 years of dense rain gauge network data with weather radar data. Figure 2.28 summarizes their findings with regards to sub-radar pixel rainfall variability. Specifically, it illustrates the probability that at least one rain gauge over the area of study reports no rainfall while the weather radar detected the presence of rainfall during the time period of observation. The basin in their study was 135 km^2 in size while the 2 km values represents a 2 km by 2 km area.

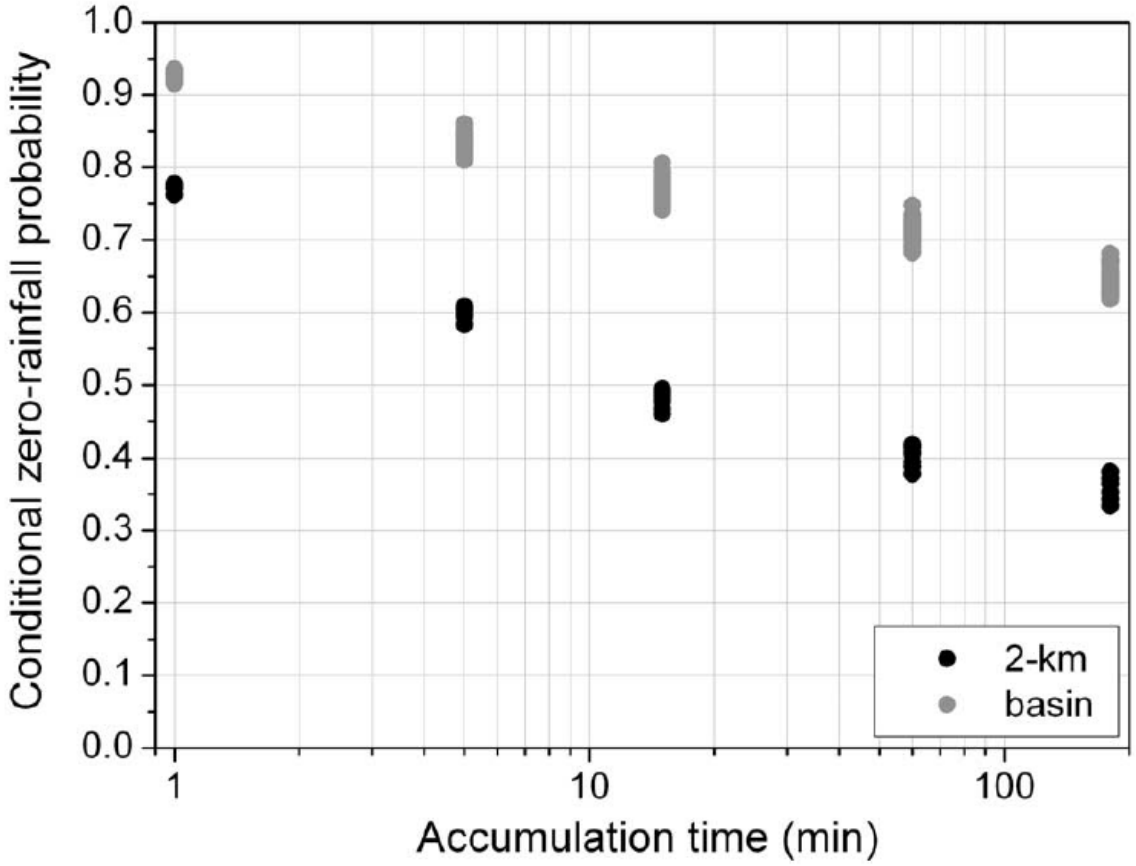


Figure 2.28 - Description of rainfall sub-radar pixel variability (Villarini et al., 2008)

Furthermore, the work of Zhang and Smith (2003) demonstrates that coarser sampling resolutions produce: 1) more areas with rainfall than that which actually occurred and, 2) effectively reduced the reported magnitude of the actual peak rainfall intensity by averaging with lower intensity rainfall values within the sampling volume.

2.3.3.5.9 Temporal Point Measurements

Weather radar precipitation estimates are effectively near-instantaneous samplings of atmospheric precipitation rates and, whose duration is in the order of a few micro-seconds (WMO, 2008a; Joe & Lapczak, 2002). However, each radar reflectivity volume scan takes approximately 5 minutes to complete and is followed by a 5-minute Doppler mode volume scan. This repeated pattern introduces large temporal

gaps in precipitation data sampling; hence, weather radar data are considered temporal point measurements. It is typically assumed that the instantaneous weather radar precipitation intensity estimates, sampled once every 5 to 10 minutes, are representative of a 5 to 10-minute rainfall depth over the sampling area.

Nystuen (1998) demonstrated the importance of sampling frequency gaps with rain gauge instruments operating at recording intervals of 1 minute every 10 minute. It was found that a 10% standard deviation about the average rainfall total were produced by the 90 % sampling frequency gaps.

Pastoriza et al. (2009) suggested that the temporal limitation of weather radar data sampling rates could be improved by use of rain gauge network data.

2.3.3.6 Radar Data Adjustment

Radar rainfall estimates represent detected rainfall at given altitudes over the Earth's surface. Unfortunately, It is far more useful for water resource managers to use near ground-level rainfall estimates in order to calculate runoff (CSA, 2010; WMO, 2008a).

In an effort to best represent ground-level precipitation and, despite the known spatial limitations of rain gauge data, weather radar precipitation estimates are often compared to observed ground point precipitation observations produced by rain gauges (He et al., 2011). The sampling discrepancies, both in time and space, between rain gauge and radar data has not yet been resolved. The impacts on the uncertainty of estimates for comparing precipitation data from vastly different sampling approaches (i.e. radar, rain gauge), whether for validation or merger is still not well understood (Michaelides et al., 2009). It is simply assumed that each data source is complementary to each other.

Considering the sampling differences between the two technologies (Table 2.6), it is possible that each instrument (i.e. radar, rain gauge) may produce largely different rain depth estimates in the same region and timeframe, while still producing equally

representative estimates of the true rainfall occurring in their respective sample area.

Table 2.6 - Sampling differences between rain gauges and weather radars

	Spatial	**Temporal**
Rain Gauge	Point	Continuous
Weather Radar	Continuous	Point

The mean-field bias radar data adjustment technique is commonly used for real-time precipitation data (Anagnostou et al., 2012). The technique consists of applying the overall gauge-over-radar ratio to the radar data uniformly over all of the observed area. One of the disadvantages of the mean-field bias technique is that it does not address the possibility of spatial precipitation variability such as that present in convective events (Anagnostou et al., 2012).

Generally, an agreement between rain gauge and radar data within roughly 10% is considered a good fit as it represents typical rain gauge data uncertainty. See Section 2.3.2.

2.4 Statistical Description of Severe Precipitation Event

A common approach used to describe severe rainfall event occurrences is the IDF method (CSA, 2010; Huard et al., 2010). The following sub-sections present a critical review of the IDF method. Other approaches to assess the level of rainfall severity include: Probable Maximum Precipitation, Areal Reduction Factors and, Intensity-Duration-Area-Frequency curves. Several of the listed alternative methods have been analyzed in a study by Jobin (2010). However, they generally all suffer from the same point-to-areal rainfall description.

2.4.1 Return Period

The Return Period (RP) states the expected average storm recurrence interval for a given severity (EC, 2011a; CSA, 2010). A storm of RP T years is expected to have a 1/T likelihood of occurring or have an event of greater magnitude occur every year (EC, 2011a; CSA, 2010). This method has historically been used to compute the expected probability of occurrence of rainfall events using rain gauge precipitation time series (CSA, 2010). This means that its validity is confined to the location of observation and the spatial extent of the sample size (i.e. rain gauge opening) (CSA, 2010).

Each individual storm is assumed to be independent (CSA, 2010) even though that may not be true. This means that the occurrence of a severe event one year is considered to not have any statistical impact on the expectation of a similar magnitude storm occurring during the following year (CSA, 2010). Furthermore, due to the statistical description of event occurrence, it is possible for a storm of a given importance to occur more than once within its prescribed RP. Table 2.7 lists the probability of event occurrence for standard RP.

Table 2.7 - RP probability of occurrence (CSA, 2010)

T Return Period	*N* Number of years														
	2	5	10	15	20	25	30	50	75	100	150	200	300	400	500
2	0.750	0.969	0.999	*	*	*	*	*	*	*	*	*	*	*	*
5	0.360	0.672	0.893	0.965	0.988	0.996	0.999	*	*	*	*	*	*	*	*
10	0.190	0.410	0.651	0.794	0.878	0.928	0.958	0.995	*	*	*	*	*	*	*
15	0.129	0.292	0.498	0.645	0.748	0.822	0.874	0.968	0.994	0.999	*	*	*	*	*
20	0.098	0.226	0.401	0.537	0.642	0.723	0.785	0.923	0.979	0.994	*	*	*	*	*
25	0.078	0.185	0.335	0.458	0.558	0.640	0.706	0.870	0.953	0.983	0.998	*	*	*	*
30	0.066	0.156	0.288	0.399	0.492	0.572	0.638	0.816	0.921	0.966	0.994	0.999	*	*	*
50	0.040	0.096	0.183	0.261	0.332	0.397	0.455	0.636	0.780	0.867	0.952	0.982	0.998	*	*
75	0.026	0.065	0.126	0.182	0.235	0.285	0.331	0.489	0.635	0.739	0.866	0.932	0.982	0.995	0.999
100	0.020	0.049	0.096	0.140	0.182	0.222	0.260	0.395	0.529	0.634	0.779	0.866	0.951	0.982	0.993
150	0.013	0.033	0.065	0.095	0.125	0.154	0.182	0.284	0.394	0.488	0.633	0.738	0.866	0.931	0.965
200	0.010	0.025	0.049	0.072	0.095	0.118	0.140	0.222	0.313	0.394	0.529	0.633	0.778	0.865	0.918
300	0.007	0.017	0.033	0.049	0.065	0.080	0.095	0.154	0.222	0.284	0.394	0.487	0.633	0.737	0.812
400	0.005	0.012	0.025	0.037	0.049	0.061	0.072	0.118	0.171	0.221	0.313	0.394	0.528	0.633	0.714
500	0.004	0.010	0.020	0.030	0.039	0.049	0.058	0.095	0.139	0.181	0.259	0.330	0.452	0.551	0.632

2.4.2 IDF

The IDF approach in an attempt at describing expected rainfall recurrence by relating rainfall depths with given durations of observation based on the historical rainfall statistics as gathered from rain gauge stations (Palynchuk & Guo, 2011; CSA, 2010). As a result, the statistics are inherently biased by the limitations of the observing instrument. The most notable limitations being the spatial point measurement characteristics as presented in Section 2.3.2.4.

RP values are often used to describe a rainfall's severity as the event's total rainfall is associated with the underlying rainfall statistics used to derive the IDF relation. A RP does not offer enough information by itself, it must be accompanied by a duration when making an IDF statement. This often causes confusion to those outside the domain of meteorology and hydrology since they are not familiar with the fundamental principles that relate to a rainfall storm's severity (Glazner et al., 1999). Ambiguity arises when three-dimensional IDF relations are discussed in a two-dimensional manner. This can easily lead to the conclusion that the relations are inconsistent. Table 2.8 is used to illustrate this point. Notice that the total rainfall depth for a 5-year and 6-hour storm is greater than a 1-hour, 100-year event. However, a 100-year event would intuitively be expected to produce a more severe rainfall event.

Table 2.8 – Edmonton International Airport IDF rainfall depths for 1-hour and 6-hour RPs (EC, 1990)

RP (years)	1-hour depth (mm)	6-hour depth (mm)
2	14.8	27.1
5	20.1	36.2
10	23.6	42.3
25	28.1	49.9
50	31.4	55.5
100	34.7	61.1

2.4.2.1 Approaches of Describing Event Occurrences

The following sub-sections present important components of the statistical approaches used in producing IDF relations.

2.4.2.1.1 AMS, MMS and POT

IDF relations are typically derived using Annual Maximum Series (AMS) statistics (Palynchuk & Guo, 2011; CSA, 2010). This approach consists of first producing daily rainfall totals for fixed timeframes (intervals) and keeping only the highest value for each given year in the period of record (Mailhot & Talbot, 2011; CSA, 2010; Svensson et al., 2007).

While the use of AMS is more prevalent for rainfall recurrence estimates, theoretically, the peak-over-threshold (POT) should be the preferred method as it is more representative of extreme rainfall (Svensson et al., 2007; Fowler and Kilsby, 2003). POT considers all events over the threshold value while the AMS is restricted to only using for analysis the greatest yearly value. As a result, AMS analyses may sometimes use annual values that might not be considered extreme and, may also exclude lesser extreme values that occurred within the same year (Svensson et al., 2007; Fowler & Kilsby, 2003). However, Fowler and Kilsby (2003) criticized the POT approach for being laborious when dealing with missing data and quite demanding when deciding what values to consider within single rainfall events. The selection of a correct threshold level is critical (CSA, 2010).

Svensson et al. (2007) provides a literature review and discussion on extreme rainfall AMS compared to monthly maxima series (MMS) and the POT approach. They noted that the MMS method produces larger RP than the AMS method even though the method is more likely to use a greater proportion of the dataset's extreme values. This is probably due to their small dataset size which forced the use of values that are not necessarily extremes. The same could be said of AMS method with respect to the actual extreme rainfall occurrence trends because the method is forced to consider values that generally would not be considered extreme (i.e. on unusually dry years) while it disregards multiple extreme events per year even if

such events surpass other years' severity values. Furthermore, seasonality may impact the AMS criteria of identically distributed values for a dataset's representation by the generalized extreme value distribution. When seasonal trends are significant in the rainfall data, which is often the case away from the tropics, only a few months in the year end up contributing to the annual maxima series. MMS is better to represent strong seasonal trends in extreme rainfall but when such trends are low, the POT approach is best. Svensson et al. (2007) found that MMS best deals with datasets that have missing data than AMS and POT (when approximately 3.5 or less values are retained for analysis per year).

Svensson et al. (2007) found that when compared with actual RP in the dataset, the MMS and POT approach faired similarly well in terms of biases. However, only the MMS yielded the best results in terms of its standard deviation while consistently performing better than the AMS with respect to both criteria. Thus, they concluded that MMS best estimates rainfall RPs.

Also, for extreme rainfall, the effect of exhibiting correlation between maxima values across several durations for any given rainfall event has been shown to be negligible (Svensson et al., 2007).

2.4.2.1.2 Data Distributions

Theoretically, extreme values for independent, identically distributed data (i.e. AMS) of considerable series length are expected to resemble one of the 3 types of extreme value distributions: Type I, Gumbel; Type II, Fréchet; and Type III, Weibull (CSA, 2010). The equation governing these 3 types can be written as the Generalized Extreme Value Distribution (CSA, 2010). Maxima datasets (e.g. AMS) with normal, exponential and log-normal distributions are well represented by the Gumbel extreme value distribution (CSA, 2010). However, the POT approach yields a Generalized Pareto Distribution for its extreme value distribution (CSA, 2010). The Gumbel distribution is often used for rainfall statistical analyses because extreme rainfall values are typically exponentially distributed (CSA, 2010). Note that the independent, identically distributed data assumptions are seldom met in

observational data for a number to reasons including seasonal and atmospheric circulation patterns (CSA, 2010).

2.4.2.1.3 Criticism with Approaches

The following sub-section presents the most notable issues with the common IDF relation and computation approach such as: the use of fixed point storm data as opposed to storm cell maximum statistics, the use of point-source data for spatial rainfall description and, the use of fixed temporal windows.

2.4.2.1.3.1 Storm Cell Maximum Statistics

In order to better capture the recurrence of extreme precipitation events in the context of climate change, Hogg and Hogg (2010) recommend that weather radar storm cell maximum statistics be explored. As previously explained, current extreme precipitation statistics are obtained via fixed location point-source instruments, which "may not sample the most intense rainfall in specific events" (Hogg & Hogg, 2010, p.14). This dated approach invariably results in underestimating the most severe portions of observed events in the region.

Ramos et al. (2006) showed that for the same simulated spatial rainfall when interpreted either as point source, mean areal or epicentre (centered on the maximum rainfall anywhere over the study area), they yielded significantly different severe rainfall statistic trends. The following figure (Figure 2.29) shows the considerable difference in rainfall depth statistics corresponding to similar RPs but computed by means of either a fixed point such as a rain gauge (left) or, point rainfall at any possible point over the entire area (right).

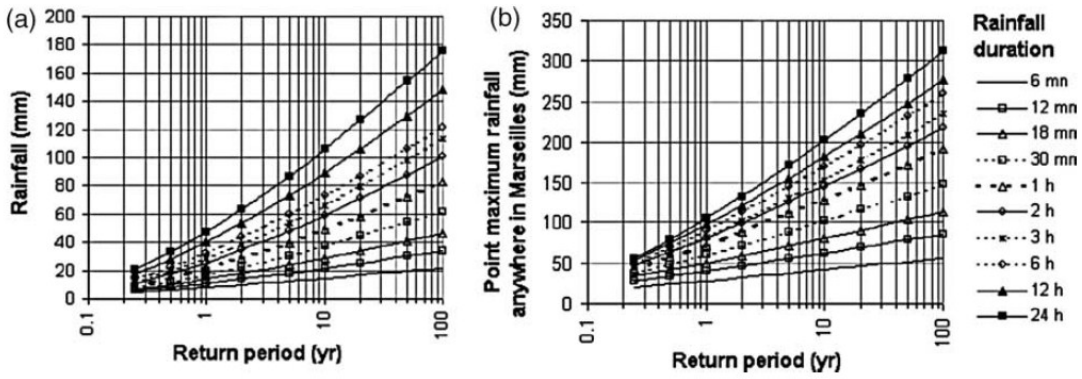


Figure 2.29 – Extreme rainfall recurrence derived from (a) fixed point compared to (b) any point over the region (Ramos et al., 2006)

Australia, Belgium, Denmark, Germany, Netherlands, and Canada are said to be exploring the possibility of producing IDF statistics from weather radar data in order to overcome the spatial shortcomings of rain gauge statistics in some region (CSA, 2010).

2.4.2.1.3.2 Use of Point Data for Spatial Description

Some of the implications of using point source data for the spatial description of phenomena has been presented in Section 2.3.2.4. As discussed, the use of point data represents a significant source of severe rainfall pattern error that has implications when used to represent regional climate patterns.

However, the current standard practice is to represent long-term rainfall statistics over a region using nearby point source-derived IDF relations. Doing so implicitly assumes that the instruments have repeatedly captured the worst rainfall totals and intensities in the region throughout the observation period. Unfortunately, rain gauges seldom, if ever, capture the most severe regions of rainfall storms (CSA, 2010).

This makes the direct use of point-source IDF statistics to larger areas questionable. The CSA (2010) Technical Guide on IDF information cautions that the probability of occurrence for an extreme event over a larger area than that of

the rain gauge opening will be greater than that of the stated IDF relation as derived for that instrument. EC states the following:

"Return period estimates correspond to probabilities of occurrence of rainfall rates or amounts at a single point: the location of the observing station. They do not reflect rainfall amounts or rates over wider areas such as river basins."(EC, 2011a, p. 3)

2.4.2.1.3.3 Fixed Window Temporal Issue

The concern regarding fixed temporal windows (e.g. mid-night to mid-night) come from the partitioning approach typically associated with the computation of AMS statistics which is based on each year's greatest 24 hr rainfall.

Longer duration storms (e.g. 12 hr) are more likely to straddle two calendar dates than shorter duration storms. This results in an underestimation of the storm's 24 hr severity as the portion of the storm in each calendar day is considered independent from the other (Mailhot & Talbot, 2011). Mailhot and Talbot (2011) found that deriving IDF relations from moving temporal windows helped mitigate the underestimation.

Also diurnal effects in rainfall patterns further skews the probability of adverse effects from use of fixed window temporal processing of daily rainfall data. Lopez and Holle (1986) noted predominant thunderstorm activity in the afternoon through the night and spilling into the early morning.

2.4.2.2 Post Event Severity Assessment

There is currently a discrepancy in extreme event description pertaining to post event severity assessment (CSA, 2010). The issue arises from the current standard practice which assesses specific event severity via its most severe attributes (irrespective of where it occurs within the event) while comparing it to descriptions of the local extreme precipitation history derived from a spatially fixed and limited sampling area size (i.e. rain gauges) (Rusjan et al., 2009; Ramos et al., 2006;

Ramos et al., 2005). The inconsistency between the two methods has implications that result in the overstatement of individual event severity while understating their true recurrence.

2.4.2.3 Stationarity Assumption

Historically and, to simplify severe rainfall statistical computations, the underlying data population was assumed stationary throughout the period of record, an approach that was considered conservative. However, this assumption is being reconsidered in light of increased understanding of climate change, naturally occurring climate and weather patterns, seasonal variability, impact of urbanization and the worsening infrastructure deficit.

2.4.2.3.1 Climate Change

Since the 1980s, climate change has become an issue of significant interest (Ungar, 1999). Global warming induced climate change is occurring (IPCC, 2007). It is highly probable that it's already having a negative impact on precipitation (IPCC, 2007) and, it's anticipated that it will further exacerbate weather-related hazards (IPCC, 2007; Mileti, 1999).

Climate change is projected to have long-term effects on rainfall patterns worldwide (IPCC, 2007) including Canada (CSA, 2010). However, the Canadian data collected to date fails to show the existence of such a trend over the territory (CSA, 2010). Nevertheless, comparison with data from south of the border show otherwise (CSA, 2010). This disparity could simply be due to the sparseness of the point-source rainfall instruments throughout Canada (CSA, 2010). Sparse instrument networks tend to miss each storm's most severe rainfall measurements, making it hard to adequately describe the actual climate, let alone any trends regarding the most extreme events (Groisman et al., 2005 in CSA, 2010).

2.4.2.3.2 Effect of Unstable Climate and Weather Patterns

Some yearly weather and climate variability is to be expected as global oscillations go through their cycles. As elaborated in Section 2.2.2, the cycles relevant to

Canada can take from months to decades to complete. The figure taken from Koutsoyiannis et al. (2009) (Figure 2.30) illustrates that the climate characteristics do not act as randomly distributed variables oscillating over an equilibrium value by comparing a simulated roulette outcome with Northern Hemisphere temperature trends over nearly the last 2000 years.

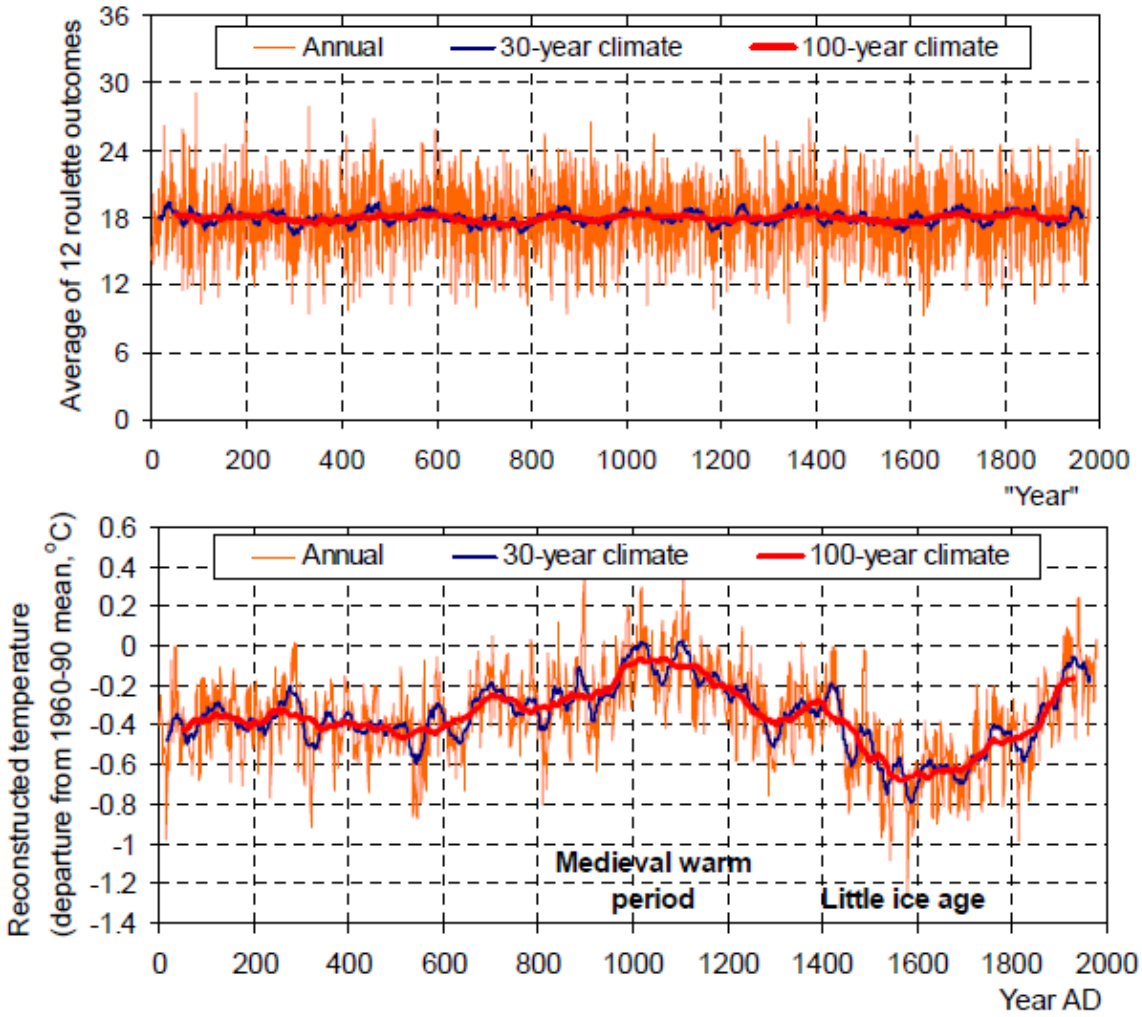


Figure 2.30 - Roulette versus Northern Hemisphere temperature trends (Koutsoyiannis et al., 2009)

2.4.2.3.3 Seasonality

Extreme rainfall trends in Canada exhibit seasonality (CSA, 2010). From January to March, snowmelt with rainfall may produce floodings throughout warm winter periods (CANHP, 2010). Approximately 40% of large-scale floods happen between

April and May during Southern Canada's spring snowmelt (CANHP, 2010). Severe rainfall is the leading cause of floods from May to September (CANHP, 2010) when convective storms occur (CSA, 2010). In end of summer and fall there are the occasional hurricane of tropical storm that produce floods in Eastern Canada while November and December typically have the least amounts of floods (CANHP, 2010).

Konrad (2001) presented the seasonal variability in extreme precipitation event occurrence over different regions of the U.S.A. Two region in the study may be of interest to Canadian extreme rainfall storm trends: Northeast may represent Eastern Canada and Midwest may represent the Canadian Prairies. Figure 2.31 presents the results where the different lines represent storm scales: solid line, 500 000 km^2; dashed line, 100 000 km^2; and dotted line, 2 500 km^2.

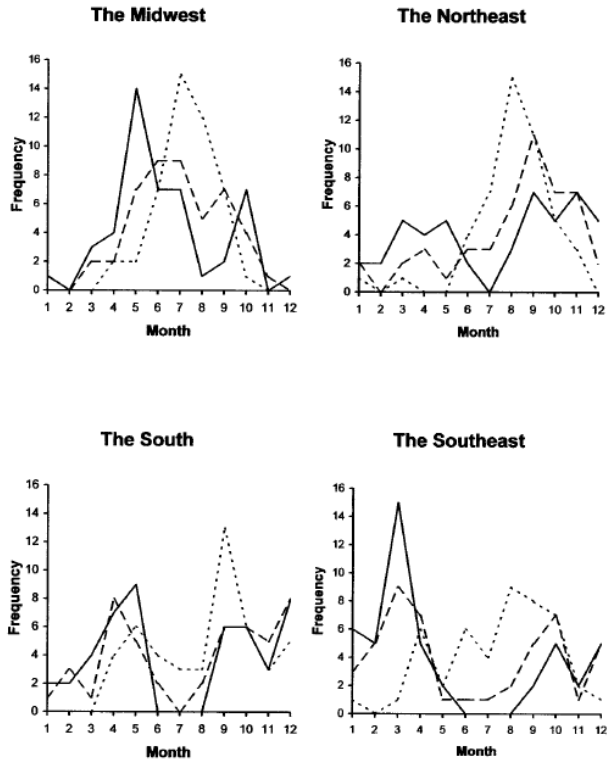


Figure 2.31 - Seasonality in extreme precipitation storm occurrence in the U.S.A. (Konrad, 2001)

2.4.2.3.4 Urbanization

Urbanization can increase the risk of flooding and related losses in two ways: 1) Increases in population and asset density can increase the probability of flooding (CANHP, 2010; Rozalis et al., 2010; Mileti, 1999) and, 2) the development of local ecosystems that once could have helped mitigate hazard's impacts no longer can and so contributes to a community's flooding vulnerability (CSA, 2010; Rozalis et al., 2010; Mileti, 1999). Flood mitigation measures often results in a negative impact on the environment (Mileti, 1999).

Furthermore, current and expected changes in the North American population demographics are and will worsen its overall vulnerability to hazards due to increased settlement in hazard-prone areas for example (Mileti, 1999). Finally, the growing income gap also decreases a community's ability to cope with and recover from hazards (Mileti, 1999).

2.4.2.3.5 Infrastructure Deficit Trend

A significant portion of the aging north American infrastructure nears the end of its design life (Mileti, 1999). Numerous aging water resources infrastructure networks can no longer offer the level of service they were design to deliver (CSA, 2010). Surplus capacity provided by conservative design practices can be lost through time for a number of reasons such as: increased urbanization, climate change and the growing amount of infrastructure components that are used past their design life (CSA, 2010).

The aging of certain types of infrastructure is of particular concern as much as they near the end of their design life (Mileti, 1999). Among those of concern are those that offer flood hazard protection (i.e. dams), which tend to exacerbate losses when they fail because they had allowed development in previously hazardous areas that are now dependent on that protection (Mileti, 1999).

2.4.2.4 EC IDF Curves and Statistics

EC is the authoritative entity that produces IDF statistics in Canada (CSA, 2010). Information regarding the EC IDF instruments is presented in Section 2.3.2.3 while Figure 2.12 presents EC's IDF instrument network. The figure is somewhat misleading with respect to the size of dots identifying the locations of instruments. The disproportionately large symbol may give the impression of a dense instrument network, which is not the case. EC IDF data sets have historically been plagued with many instrument site closures and relocations. Frequent operational changes in instrumentation, methodologies, network layout and site decommissioning introduce increasing complexity to climate data analyses (Mekis & Vincent, 2011).

Canadian IDF rain gauge data is generally processed in 5-minute fixed increments but AMS are derived from fixed 24-hour temporal windows of calendar days (Mailhot & Talbot, 2011; EC, 2011a). The use of a temporally fixed window in the computation of extreme rainfall statistics introduces a bias as described earlier in Section 2.4.2.1.3.3. Also the method's name 'AMS', as derived by EC, is misleading in itself due to seasonal operation restrictions of tipping bucket rain gauges in Canada actually limit AMS calculations at some stations to using data from only the non-freezing periods of each year (Mailhot & Talbot, 2011; CSA, 2010).

EC considers a minimum of 10 years of rainfall data adequate for producing rainfall depth RP estimates up to 100 years (EC, 2011a; CSA, 2010). However, 10 years of observational data may not be enough to fully describe the extreme rainfall trends due to some climate patterns having much longer cycles, as discussed in Section 2.4.2.3.2.

While EC uses the method of moments to estimate IDF parameters, other fitting methods have been developed (e.g. maximum likelihood, probability weighted moments and L-moments), some of which produce better fitting RP estimates (CSA, 2010).

The approach used by EC in computing IDF relations assumes the data population can be represented by the Gumbel GEV statistical distribution; however, this has not been tested (Mailhot & Talbot, 2011). As Mailhot and Talbot (2011) have shown, several data series distributions are best represented by other GEV distributions.

Similarly, the EC approach also assumes that trends are stationary through time without testing it's validity. However, Mailhot and Talbot (2011) discovered that the data stationarity assumption for statistical trends was not valid for several stations that were tested.

Up until recently, EC IDF statistics updates were infrequent. More than a decade passed between IDF update. Unfortunately, water resources designs using IDFs and rainfall event severity assessments were completed using the last update that did not reflect recent extreme rainfall trends (Mailhot & Talbot, 2011). The CSA (2010) recently recommended that IDF statistics be updated at a 5-year interval to minimize the effects of outliers and capture changes in the climate.

Furthermore, EC also recognizes that the probable future changes in climate are not represented in their IDF statistics as they are derived from historical observations (EC, 2011a).

Many IDF rain gauge datasets have missing data periods, some of which have several years missing from their records. These gaps in observational data have the potential to considerably bias the IDF statistics since several extreme rainfall events that may have been missed are excluded from the overall statistics; hence, skewing the IDF relations.

However, perhaps the most significant issue, is that EC recognizes the spatially limited applicability of the IDF relations due to its point-source nature and, since 2003, includes a warning with its data stating that their historical IDF statistics are

only valid in the immediate vicinity of each instrument and should not be directly applied to larger areas (EC, 2011a).

2.5 Extreme Precipitation Storms

The following sub-sections describe extreme precipitation storms from a national perspective, and from a regional perspective near the City of Edmonton (the area of interest for this study).

2.5.1 Recent Historical Extreme Events

Sewer backup related damages in Canada has a significant yearly cost to governments, homeowners and insurance companies (Sandink, 2007; IBC, 2011). Allouche and Freure (2002), as cited by Sandink (2007), found that of 26 Canadian cities, 42% experienced flooded basements more than once a year and, 92% once every few years. Many communities in Canada experience distress and suffering among other negative social impacts that accompany the direct physical damages of flooding (CANHP, 2010). In the past two decades, the City of Toronto (specifically 2005), Hamilton (2004, 2005 and 2006), Ottawa, Sarnia, Thunder Bay, Peterborough (2004), Port Alberni, Kenora, St. Johns, Winnipeg, Stratford, Edmonton (2004), Calgary and Moncton have all experienced significant urban flood damages (Sandink, 2007). The federal government program for Disaster Recovery Financial Assistance Arrangement distributed approximately $720 million dollars in compensation and support payments from 1975 to 1999 for 63 separate floodings (CANHP, 2010). Additionally, business-related flood insurance claims totalled over $750 million dollars from 1984 to 1998 (CANHP, 2010). Extensive sewer backup losses were incurred in Winnipeg in 1993 as a result of a severe rainfall storm that generated $213 million in insurance payments when adjusted to 2003 dollars (CANHP, 2010). Furthermore, the 1997 flooding of the Red River in Winnipeg resulted in the insurance industry paying over $200 million dollars in damages, $2 million of which is attributed to sewer backups (CANHP, 2010). It is estimated that Canadian insurance companies compensate $1.32 billion yearly for water-related damages (IBC, 2011). In 2004, a storm system caused $143 million in sewer

backup insurance claims in Edmonton, then $87 million in flood insurance claims and $25 million in government relief payments were made in the City of Peterborough (Sandink, 2007). In 2005, an extreme rainfall event was responsible for $500 million in insurance claims for flood damages throughout southern Ontario, $247 million of which were made due to sewer backup (Sandink, 2007).

2.5.2 Spatial Extreme Events as Described Near Edmonton

Analysis of the RERSD as described in Section 3.3 show an increased rainfall storm occurrence trend in the May to September period (shown in Figure 2.32) (Kije Sipi, 2007). Also, of interest, is Konrad's (2001) study regarding 1950 to 1996 extreme precipitation storms in the Midwest as they could be considered as representative of storm trends for the Canadian prairies. The upper left plot in Figure 2.31 illustrates the observed seasonal storm frequency in the Midwest.

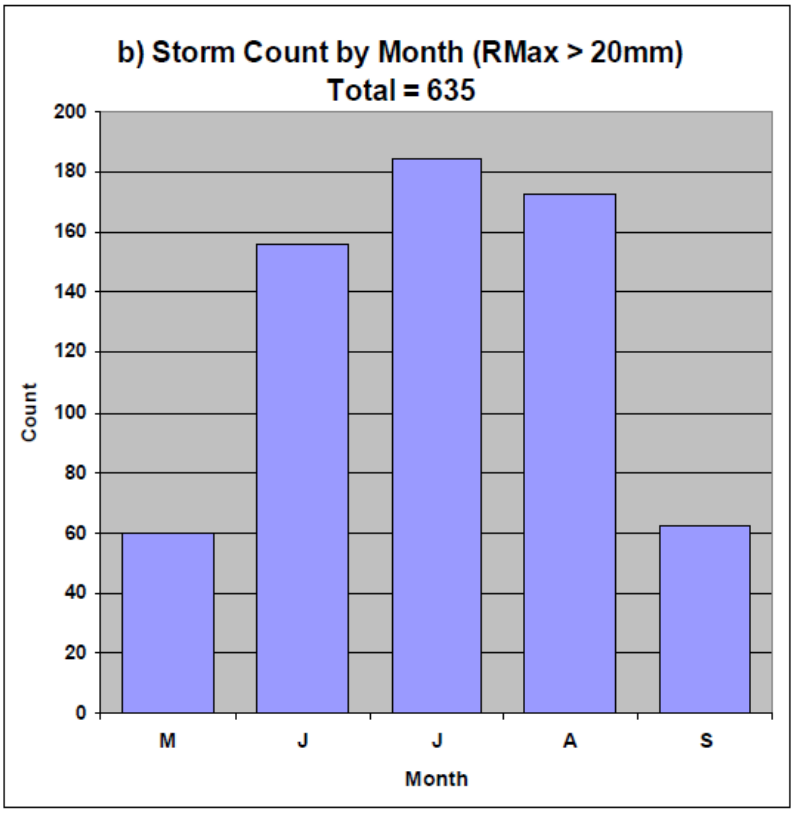


Figure 2.32 - Monthly distribution of significant rainfall storms near Edmonton (Kije Sipi, 2007)

The following figure (Figure 2.33) presents, on a semi-log plot, the number of significant storms observed in the City of Edmonton region between 1998 and 2005, excluding 2001 (see Section 3.2.2 for details), via their maximum observed rainfall total from the RERSD (Kije Sipi, 2007). There is a drastic reduction in the amount of observed storms the higher the total rainfall.

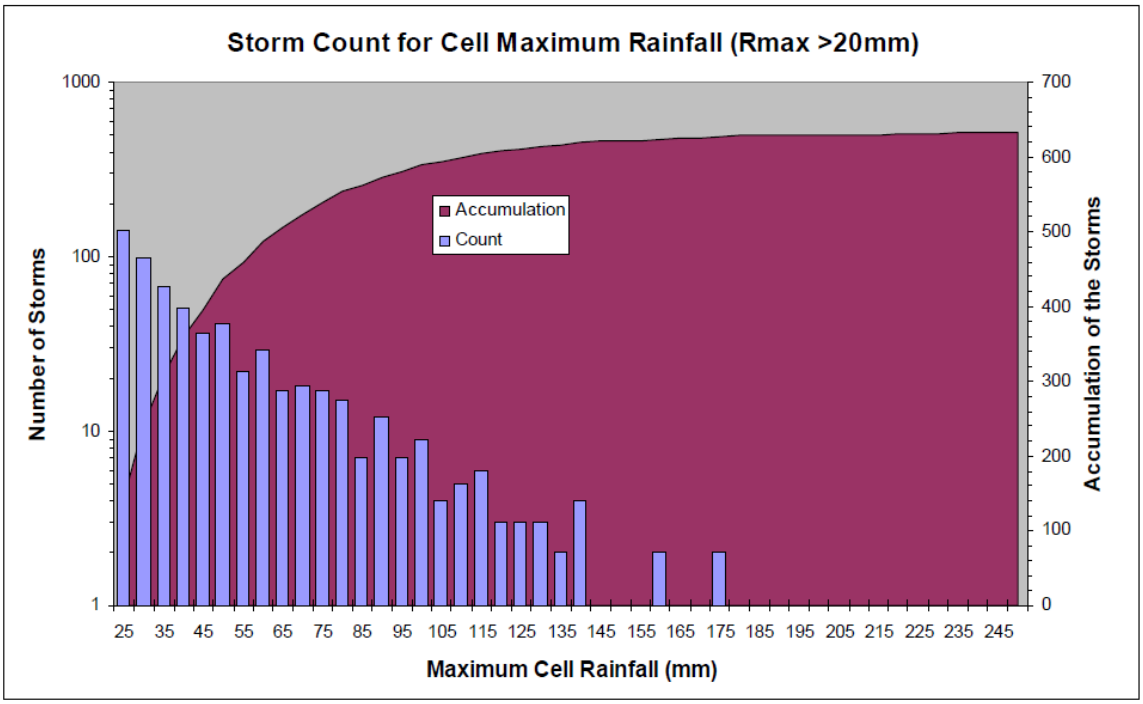


Figure 2.33 - Number of significant rainfall storms observed in the Edmonton region (Kije Sipi, 2007)

The trend in spatial storm extent of significant rainfall storms in City of Edmonton region between 1998 and 2005, excluding 2001 (see Section 3.2.2 for details), derived from the RERSD is shown in Figure 2.34 (Kije Sipi, 2007). As can be expected, the majority significant storms are of smaller size; however, results show that there is a considerable amount of large events as well.

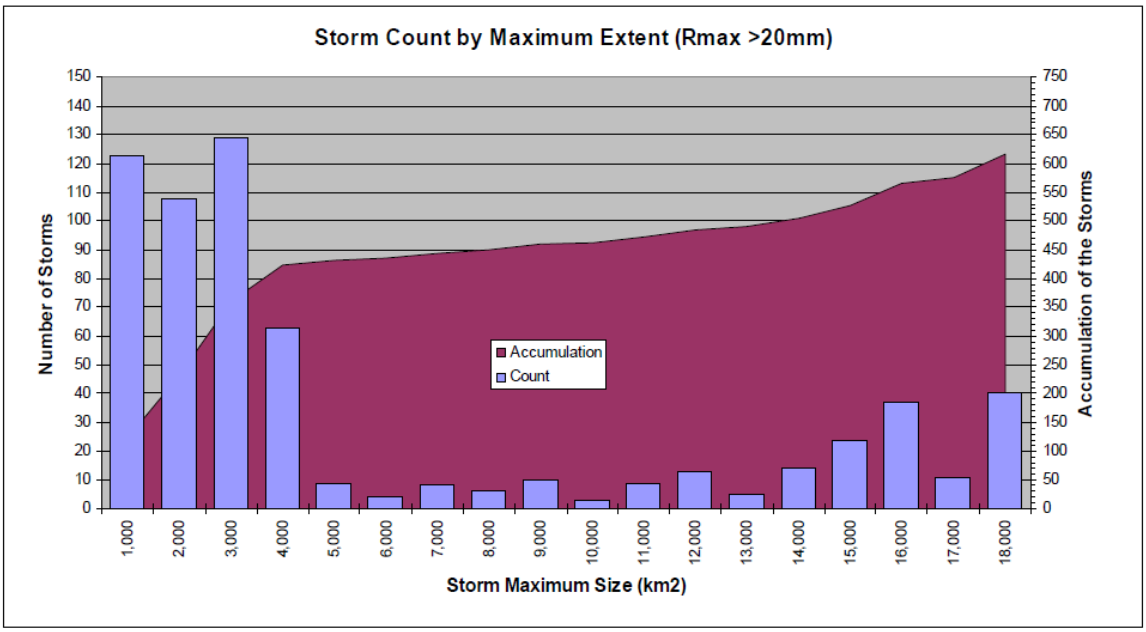


Figure 2.34 - Trend in spatial storm extent of significant rainfall storms in the Edmonton region (Kije Sipi, 2007)

2.6 Existing Storm Classification Methods

A comprehensive literature review had been performed previously in an effort to summarize the state of rainfall storm classifications for drainage design considerations (Jobin, 2010). Various storm importance classification methods from diverse storm types were consulted including: rainfall storm, winter storm, wind, tornado and hurricane classification methods. In all, 14 existing storm classification methods were assessed for adequacy and/or studied for inspiration as well as insight into producing a storm severity index. The Jobin (2010) study findings are summarized in Section 2.6.6.

2.6.1 Rainfall Storm Classification Methods

Five rainfall storm severity classification methods were examined: the Haestad Severity Index, the Rainfall Rate Weighted Index, the Alluvial Strength Index, the Intensity-Duration-Area-Frequency method as well as the use of Areal Reduction Factors, and the conceptual Edmonton Storm Classification Method. Below is a brief description of each method.

Glazner *et al.* (1999) developed the Heastad severity index, which further classifies rainfall RPs as determined by the existing IDF rainfall severity method into 10 subjectively chosen levels (and 5 categories) according to a rainfall's duration.

Casas *et al.* (2004) proposed an extreme rainfall importance index for the western Mediterranean region. The Rainfall Rate Weighted Index was produce by analyzing Barcelona, Spain rain gauge data from 1927 to 1992. A correlation and cluster analysis yielded four classes of extreme rainfall storms. The index considers the input rainfall in terms of 4 distinct durations, which represent the 4 extreme rainfall storm classes, and assigns each storm class an equal weighing in the severity index computation.

Mazzarella and Diodato (2002) developed a rainfall severity classification method in response to landslides near the Town of Sarno, in southern Italy. The Alluvial Strength Index value is computed by multiplying: 1) the sum of rainfall experience in the last 30 days prior to the date of interest (this represents a measure of the antecedent soil moisture conditions), and 2) the total rainfall for the last two days prior to the date of interest. Rainfall severity is assessed according to 6 levels, varying from very weak to catastrophic.

De Michele *et al.* (2002) produced an IDF relation that incorporated a measure of spatial rainfall by performing a frequency analysis on the mean area rainfall for a variety of surface sizes and durations. Similarly, Areal Reduction Factors can be used with the IDF method to adjust the rainfall severity assessment for spatial considerations. The works of Veneziano and Langousis (2005), De Michele *et al.* (2001), Lombardo *et al.* (2006) and Olivera *et al.* (2008) can be consulted for more information on Areal Reduction Factors.

After recently having experienced several significant flood-producing rainfall storms, the City of Edmonton sought to obtain an alternative approach to assessing extreme

rainfall severity and frequency in a more representative manner than the currently used IDF method. A storm classification method was developed by the City, though it is still in its conceptual phase and further work is required (Lodewyk, n. d.). This method incorporates a antecedent precipitation index (a measure of the soil moisture content), total rainfall, rainfall intensity and an adjustment for the location of interest within city location. A rainfall storm's severity is assessed according to 7 possible levels of importance for which potential impacts are presented.

2.6.2 Winter Storm Classification Methods

Three winter storm severity classification methods were examined: the Classification Scale for Winter Storms/Nor'Easters, the Northeast Snowfall Impact Scale, and the Storm Severity Index. Below is a brief description of each method.

Gregory Zielinski (2002) produced a storm severity classification approach for nor'easters (large Atlantic winter storms) in the Eastern and central U.S.A. Storms are classified on a 1 to 5 scale with respect an importance index and a duration factor. Importance is assessed according to barometric pressures while the storm duration is quantified by the storm's velocity.

Kocin and Uccellini (2004) produced winter storm severity approach. The Northeast Snowfall Impact Scale assesses a storm's severity with regards to its spatial extent, average snowfall and the affected region's population. The scale values range from 1 to 5 with increasing severity.

Nixon and Qiu (2005) produced the Storm Severity Index (SSI) to determine a winter storm's importance for managing winter road maintenance and clearing operations. A storm's severity is assessed with regards to storm type, pre and post storm temperatures, and wind speeds during and after the storm by a set of predefined levels with associated values.

2.6.3 Wind Classification Methods

With regards to wind severity classification, only the Beaufort Scale was examined. The Beaufort Scale is a longstanding approach of assessing wind speeds. It was developed by Francis Beaufort in nineteenth century to assist in the operation of sailboats (Peterson & Hasse, 1987). Eventually, the scale was modified for land use and associated with measured wind velocities (Sparks, 2003; Peterson & Hasse, 1987). The scale has 12 levels but can be extended if needed (Doswell, 2003).

2.6.4 Tornado Classification Methods

Two tornado importance classification methods were studied: the T-Scale and the Fujita Scale.

Terence Meaden developed a tornado severity scale, the T-Scale, based on the maxium observed wind speeds by extending the Beaufort Scale (Meaden *et al.,* 2007). Tornado importance is represented by a numerical scale ranging from 0 to 10.

The Fujita Scale presents a tornado severity scale that spans between the Beaufort wind severity scale value of 12 and the Mach 1 (the speed of sound) (Meaden *et al.,* 2007; Fujita, 1981). Tornado severity is presented on a scale of 0 to 5.

2.6.5 Hurricane Classification Methods

The three Hurricane importance approaches that were studied include: the Saffir-Simpson Scale, the Hurricane Classification System, and the Hurricane Severity Index. Below is a brief description of each method.

The Saffir-Simpson Scale was produced to fill the need to convey a hurricane storm's severity to the public (Senkbeil & Sheridan, 2006). A hurricane's severity is assessed from 1 to 5 in terms of maximum wind speed and storm surge (Senkbeil & Sheridan, 2006).

Senkbeil and Sheridan (2006) produced the Hurricane Classification System to describe post-landfall hurricane storm severity, a noted weakness in the Saffir-Simpson Scale. The approach uses the storm's central pressure, sustained wind speed, wind gust speed, storm duration, surge height and precipitation depth to assess the hurricane's severity, which is presented from 0 to 100.

Christopher Hebert and Robert Weinzapfel proposed a hurricane importance method, the Hurricane Severity Index, that in addition of wind speed and storm surge, also accounted for the storm size (Hebert *et al.,* 2008). Hurricane severity is presented from 0 to 50.

2.6.6 Notes on Developing a Storm Classification Method

The most noteworthy findings, as found by Jobin (2010), include:

- None were found to be satisfactory in producing a rainfall storm severity index.
- All studied rainfall storm classification methods, only considered rainfall in a point spatial context.
- The Main purpose of most methods, was to communicate storm severity to the public.
- Complex presentations of the information is to be avoided. The storm severity should be immediately obvious.
- The spatial distribution of rainfall should be considered, including storm cell size.
- The antecedent soil moisture conditions should be considered.
- The use of multiple rainfall storm descriptors is advised.

2.7 Discussion

The characteristics and limitations of rain gauge and radar data statistics have been presented. The most notable of which is that rain gauges provide a limited spatial description of rainfall. However, rain gauge's sample rainfall in a near temporal

continuous manner. In contrast, radars can provide contiguous spatial rainfall estimates in a point like manner temporally.

Despite the inherent instrumental limitations of both weather radars and rain gauges, the combination of both is thought to help overcome their limitations to produce a more representative description of actual precipitation.

The limited spatial description of rainfall by rain gauge data alone is of particular concern when attempting to describe extreme rainfall trends for an area greater than that of a rain gauge collector opening. Furthermore, issues with the current approach of describing extreme rainfall trends were presented. The most notable were: 1) the use of fixed-point storm data as opposed to storm cell maximum statistics, 2) the use of point-source data for spatial rainfall description and; 3) the use of fixed temporal windows.

The significance of using climate and weather patterns into consideration dictates that at least 10 years of data is required in order to ensure that the climate and weather cycles are considered in the analysis.

The impact of recent extreme precipitation storms in Canada was presented. Rainfall storm trends in the region of the City of Edmonton were used to introduce the study area.

Finally, a brief summary of a previous literature review's the findings was presented.

3 Study Data

The study data includes information on basement flooding in the City of Edmonton as well as a regional rainfall storm database that had previously been created from rain gauge and weather radar data.

The information contained in the following sub-sections that pertains to the Edmonton regional rainfall storm database, is taken from the spatial rainfall study report produced by Kije Sipi (2007) for the City of Edmonton. The study included the creation of a regional rainfall storm database, which has since been updated using the method outlined in said document.

3.1 Regional Edmonton Rain Gauge Data

Rain gauge data was used for calibrating the weather radar data in the production of the corporate storm database.

3.1.1 Source of Data

Although rain gauge data was extracted from the Kije Sipi Ltd corporate archives (Kije Sipi, 2007), the original data was obtained from EC and the City of Edmonton (Kije Sipi, 2007). The City of Edmonton only deploys their tipping-bucket rain gauge instruments during warm summer months (May through September).

3.1.2 Data Format

The data format of the retrieved rain gauge information came in various file formats, particularly with the earlier data, which required additional processing (Kije Sipi, 2007).

3.1.3 Content

The rain gauge rainfall information inclusively spans the months of May to September every year from 1998 to 2009 (Kije Sipi, 2007). The total number of rain

gauges whose data was available for analysis varied throughout the years from 19 to 23 (Kije Sipi, 2007). The EC stations were scattered throughout the region while the City of Edmonton's instruments were only located within city limits (Kije Sipi, 2007).

3.2 Edmonton CWHK Radar Rainfall Data

Data from the weather radar located near the City of Edmonton was the principal source of information in the production of the corporate storm database.

3.2.1 Introduction

Carvel, designated as CWHK and actually located near the Town of Carvel, is the actual name of the EC weather radar operating near the Edmonton area. The instrument is located approximately 30 km West of the City of Edmonton. It is this weather radar's data that had been used for the production of the regional rainfall storm database and made available for the analysis.

The instrument was part of the 1997 to 2004 radar network upgrades. Prior to being upgraded, the Carvel radar only returned conventional reflectivity estimates within a maximum range of approximately 120 km. In mid-2000, the Carvel radar station was upgraded with Doppler capability and a conventional reflectivity scanning range that doubled the original capability to 256 km.

3.2.2 Source of Data

Although the weather radar reflectivity data was originally provided by EC, the radar data set used for analysis was retrieved from the Kije Sipi Ltd corporate archives (Kije Sipi, 2007). Note that all of the weather radar data for 2001 was inexplicably lost to Kije Sipi Ltd. and therefore is not available.

3.2.3 Source Data Format

EC provided the radar reflectivity data in a 1.5 km altitude radial CAPPI format, as presented in Section 2.3.3.3.3. However, prior to 2000 the radar reflectivity data

was provided by EC as 1 km^2 Cartesian grid files. They effectively converted the radial data into radar-centered grid files in producing the deliverables. In order to have a single and consistent format for the production of the rainfall storm database, the company transformed the radial data into a grid format prior to the analyses. Kije Sipi Ltd emulated and used the EC data conversion approach in transforming radial data into the Cartesian format.

3.2.4 Content

The radar reflectivity data for the months of May through September and, from 1998 to 2009 was used in the analyses (Kije Sipi, 2007). However, radar data for the year of 2001 was unavailable for analysis (see Section 3.2.2 for details) (Kije Sipi, 2007).

Note that EC's summer scan strategy is usually in effect between mid-April to the 1st of December (Zhang et al., 2008). The use of data from different scan strategies would likely introduce unwanted data heterogeneity, which is not the case here.

The radar data used for deriving the storm database was restricted to a 120 km radius from the instrument (Kije Sipi, 2007). This avoided range-related effects and use of information at a 1 km^2 resolution (Kije Sipi, 2007).

3.2.5 Radar Data Processing

Radar reflectivity data as originally provided required the following processing in order to best represent rainfall at ground elevation.

3.2.5.1 Radar Reflectivity to Rainfall Estimate Conversion

Radar reflectivity data, as provided by EC, were converted to rainfall estimates using Equation 2.6, the Marshall-Palmer (1948) and Marshall et al. (1955) relation. See Section 2.3.3.2.2 for more information regarding the conversion of radar reflectivity into rainfall estimates.

3.2.5.2 Radial Data Spatial Arrangement to 1km^2 Grids

As previously mentioned, the radial to UTM Cartesian radar data conversion was carried out by Kije Sipi Ltd in a manner that emulates the standard EC approach for radar data from mid-2000 to 2009. The older data was already delivered by EC in a Cartesian grid format. The radial-to-grid conversion algorithm uses a closest neighbor technique to populate the 1 km^2 grids.

As previously discussed in Section 2.3.3.3.3, the standard Canadian weather radar conventional reflectivity product is given in 1.5 km CAPPI with a radial grid patter (e.g. Islam & Rasmussen, 2008). It is common practice to transform the data into a 1 km^2 Cartesian grid to simplify interpretation and analysis (e.g. Islam & Rasmussen, 2008). However, radial data conversion to Cartesian grid format invariably introduces known spatial distortions.

3.2.5.3 Projection of Edmonton Radar Data Grid

As is the case for almost all Canadian Provinces, Alberta's territory spans across more than one UTM geographic zone according to the North American Datum of 1983 (NAD83). Alberta has its western portion in UTM zone 11 while its eastern portion in UTM zone 12. Unlike the City of Calgary, which also happens to have its territory split between the two UTM zones, the City of Edmonton is entirely in UTM zone 12. However, Edmonton's closest EC weather radar (i.e. Carvel) is located in UTM zone 11. To avoid data distortion issues, the radar data's grids and all mapping features are projected to a single UTM zone (Z-one 11).

3.2.5.4 Rainfall Estimates During Doppler Scan

Intermediate conventional radar scan reflectivity values were computed using a linear average rainfall rate approach to estimate reflectivities during the Doppler scan cycles. However, a noted downside of using the average is that the method will always yield a lower value than the highest measurement being averaged.

3.2.5.5 Radar Area Blanking

Anomalous radar pixels were identified for each year of record by considering the total annual rainfall (Kije Sipi, 2007). Pixels exhibiting anomalous radar rainfall data

were excluded from further processing and analysis (Kije Sipi, 2007). This process yields a filtering spatial mask, whereas anomalous radar pixels were assigned a 'blanking' attribute 'zero'. Figure 3.1 presents the 2004 radar rainfall data mask, whereas areas shown in gray are masked, hence, eliminated from further considerations.

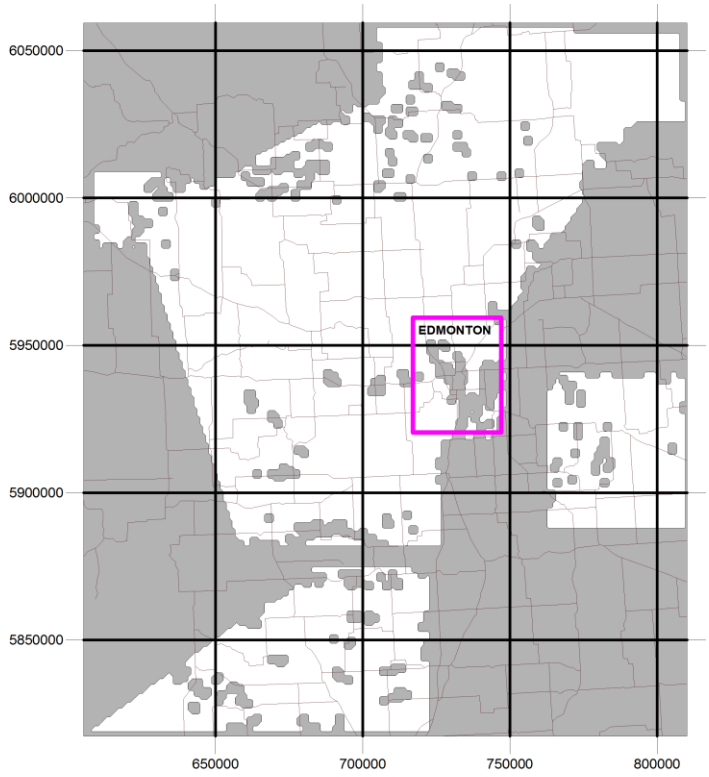


Figure 3.1 - Radar rainfall data mask of the storm database for 2004 (Kije Sipi, 2007)

3.2.5.6 Mitigation of Effects from Hail Detection

As previously discussed in Section 2.3.3.5.4, the presence of hail can skew radar rainfall estimates. A 250 mm/hr ceiling value for 5-minute rainfall rates was imposed onto radar rainfall estimates to mitigate hail contamination effects (Kije Sipi, 2007).

3.2.5.7 Calibration Approach

The spatially homogeneous but temporally-varying radar data adjustment technique known as 'Mean-Field Bias' was used. The mean-field bias is briefly discussed in Section 2.3.3.6. The specific approach used is similar to that described in Fulton

(1999). For a given temporal window (in this case 24 hours), the precipitation accumulation of a 3 by 3 km^2 set of radar pixels around each respective rain gauge are compared to that of the rain gauge. The radar pixel accumulation that best matches that of the given rain gauge is retained for computation of the radar data adjustment factor. The sum of all rain gauge totals over the sum of all radar pixels paired to rain gauges yields the calibration factor for the given interval (Kije Sipi, 2007).

Radar data calibration was performed using the available rain gauge data on a 24-hour fixed window basis from midnight to midnight (local time) (Kije Sipi, 2007). Daily radar calibration factors were restricted within the range of 0.1 to 1.3 to avoid unreasonably high adjustments that can result from periods with low rainfall (Kije Sipi, 2007).

3.3 Regional Edmonton Rainfall Storm Database

The RERSD was created as part of a study for the City of Edmonton to spatially characterize the rainfall storms occurring near the city.

3.3.1 Introduction

Calibrated radar rainfall data used for the creation of the RERSD spans the months of May through September from 1998 to 2005, excluding 2001 (see Section 3.2.2 for details) (Kije Sipi, 2007), and was later updated to include data up to 2009. The database is currently being expanded with 3 more years of data plus the missing 2001 dataset. At that point, the RERSD period of record will have reached 15 years (1998 to 2012).

Many extreme radar rainfall studies, including the RERSD limit the scope to only warm weather months that is typically the peak season for convective storms (e.g. Gourley et al., 2010; Mandapaka et al., 2009; Kaltenbock et al., 2009). This minimizes errors associated with the presence of the 'bright band'.

Note that the RERSD's temporal rainfall information was reduced in resolution from the original 5-minute estimates to 15-minute data intervals in order to reduce the significant processing times (Kije Sipi, 2007).

3.3.2 Methodology

The spatial progression of rainfall through time was tracked and aggregated into storm entities (Kije Sipi, 2007). The radar pixels that met the following requirements were grouped with the nearest storm:

- radar pixels with rainfall rates greater than 0.2 mm/hr
- located within 4 radar pixels from the edge of the closest storm, and
- must have occurred no more than 6 hours after the storm's last recorded storm rainfall (Kije Sipi, 2007).

Radar pixels that meet the first requirement but fail to meet the last 2 are considered to be part of a new storm (Kije Sipi, 2007).

3.3.3 Storm Parameters

Once individual storms were identified, hydrologically relevant parameters were computed. The following is a list the storm parameters which are of potential interest to this study. For a complete list, refer to (Kije Sipi, 2007).

- Start day, month, year, time
- End day, month, year, time
- Storm duration
- Geographic spatial storm extent
- Peak total rainfall location
- Peak total rainfall value
- Peak total rainfall duration
- Peak rainfall rate location
- Peak rainfall rate
- Total volumetric amount of rainfall
- Maximum storm size

- Average rainfall during storm
- Average North-South storm displacement
- Average East-West storm displacement
- Average storm speed
- Total storm distance
- Storm type
- Spatial decay ratio
- Areal reduction factor

3.4 Edmonton Flooded Basement Database

The following sub-sections present the information provided by the City of Edmonton regarding occurrences of basement floodings.

3.4.1 Introduction

The following clarification about the City of Edmonton's basement floodings database was given by Lodewyk and Wang (2010). The City of Edmonton's flood damage database does not relate damages in terms of dollar value; it does however, list the number of basement floodings as well as identifies some of the street floodings. It contains data from 1930s to present and notes the location of floodings per street address. Few commercial building floodings were noted. This is explained in part by the fact that most structures are set on grade, only residential buildings and structures in the downtown area are built below grade.

3.4.2 Source of Data

The following two documents were perhaps the most pertinent among those received from the City of Edmonton's Drainage Services Branch. Figure 3.2 presents a map, as provided, partitioning the City of Edmonton's territory into sub-areas according to Thiessen polygons (in pink); each centered on City rain gauge instruments according to its 2009 network layout. The City of Edmonton is approximately 30 km by 40 km in size; hence has a total surface of 1,200 km^2.

Notable geographic features that can be seen on the map are: the locations of City rain gauges previous and current to 2009, the Canadian Forces Base Edmonton (easily identifiable with its runways just north of City limits), Edmonton City Centre (Blatchford Field) Airport (within City limits) and the North Saskatchewan River running through the City along a Southwest to Northeast axis. The city also provided a spreadsheet containing the total basement floodings associated with different storms, broken down into the respective rain gauge Thiessen polygon. See the following table (Table 3.1) for a summary of said spreadsheet.

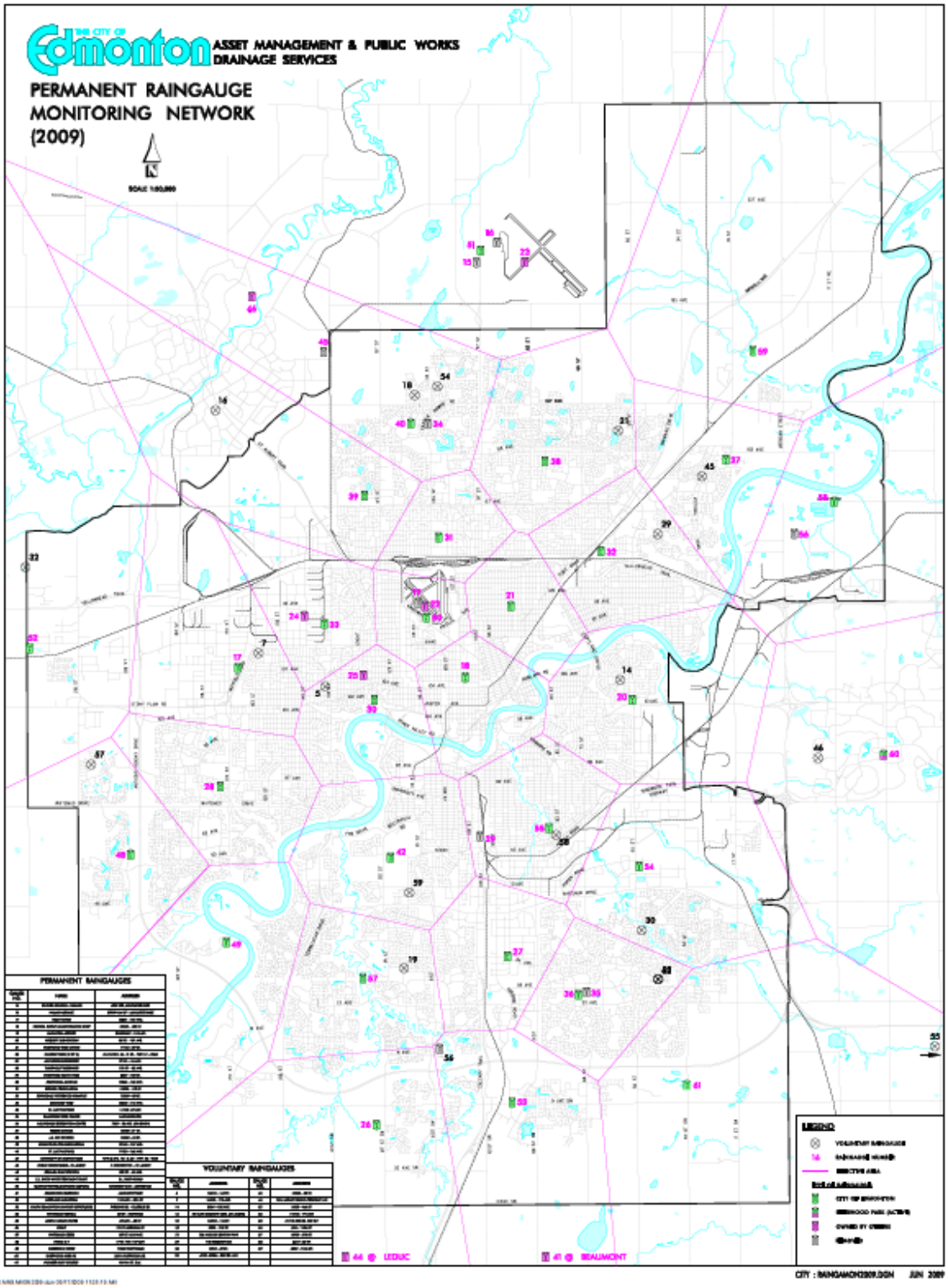


Figure 3.2 - City of Edmonton's rain gauge network Thiessen polygon map (City of Edmonton, 2009)

Table 3.1 - Summary of floodings per event per polygon

	Date									
Rain Gauge	**1998 June 29-July 5**	**1998 July 31**	**1999 May 21**	**1999 June 18**	**1999 July 9 & 13**	**2002 July 19**	**2003 August 6**	**2004 July 2 & 6**	**2004 July 11**	**2006 June 15**
17	1	2	2	1				2	18	3
18	24	14					17	9	8	32
19	2									
20	1	30	1				4	43	26	34
21	32	7					19	16	31	3
26								6	19	
27	2	3						34	438	2
28	4	25						10	1450	79
29	4	39				2	4	50	312	
30	5	16		1		1	12	5	80	36
31	6	1					54	1	6	1
32	4	5	2	7		1	14	25	24	3
33	2	9					12	2	12	3
34	1									
36	9		3				13	22	212	
37							8	1	1	
38					3	4	3	10	13	1
39							5		3	1
40				1			12	3	3	
42				1		1	1	70	410	13
48								2	6	1
49								15	18	1
50							45	1	11	
55			2	1						12
57										6
Total	97	151	10	12	3	9	223	327	3101	231

Note that only information related to floods that occurred during the RERSD's period of record were retained for subsequent analysis.

3.4.3 Content

The Thiessen polygons as depicted by the raster map in Figure 3.2 were digitized for ease of mapping and analysis.

Figure 3.3 presents a map of relevant regional features, including those of Figure 3.2. The regional base map layers were obtained from the Natural Resources Canada's online database of the 1:50,000 scale topographic map grid 83H (NRCan, 2007). Specifically, the following four maps taken from the National Topographic Data Base were assembled into a single map: 083H05, 083H05, 083H06, 083H11, and 083H12. The final map was produced using Golden Software's MapViewer for it's ability to manipulate map projections.

Figure 3.4 depicts the City's rain gauge station polygons as an overlay on the region's base map as shown in Figure 3.2. Notice that Figure 3.4 is a digital reproduction of Figure 3.2.

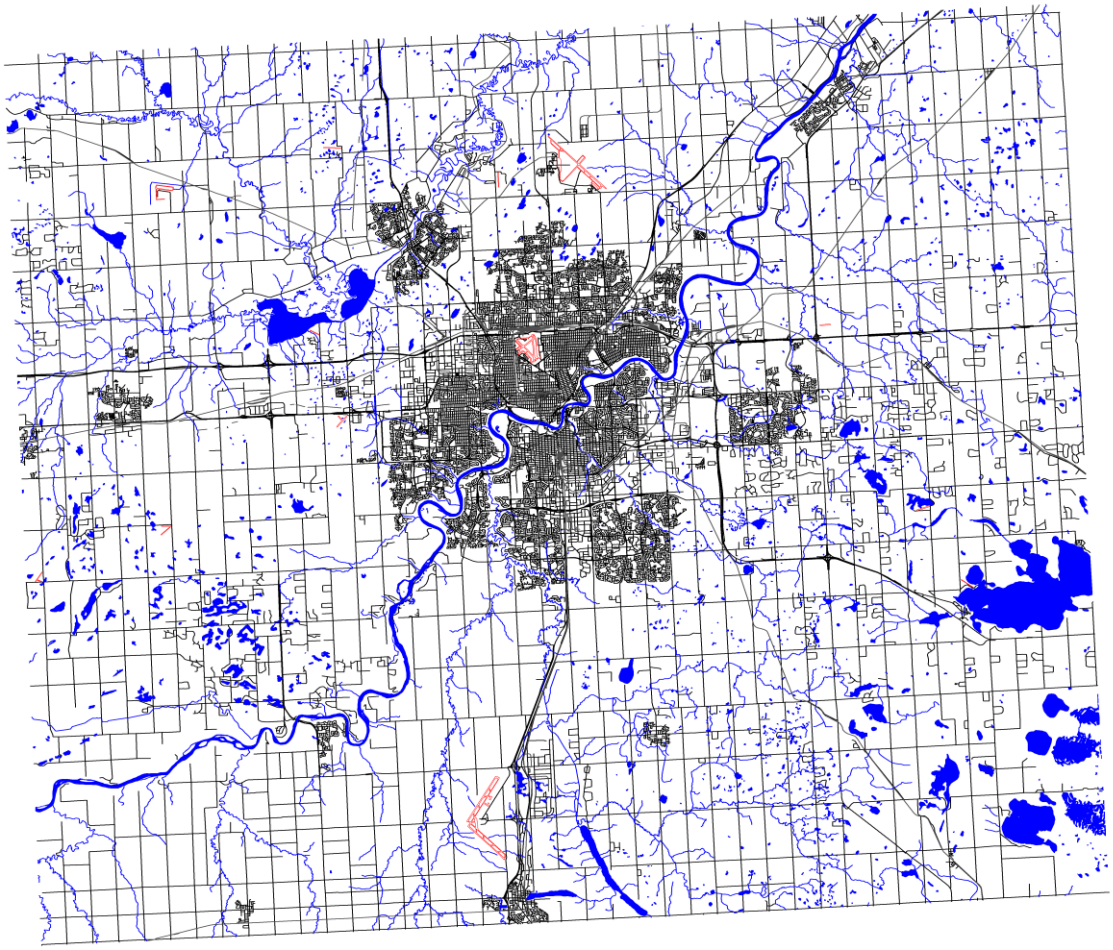


Figure 3.3 - The City of Edmonton regional map

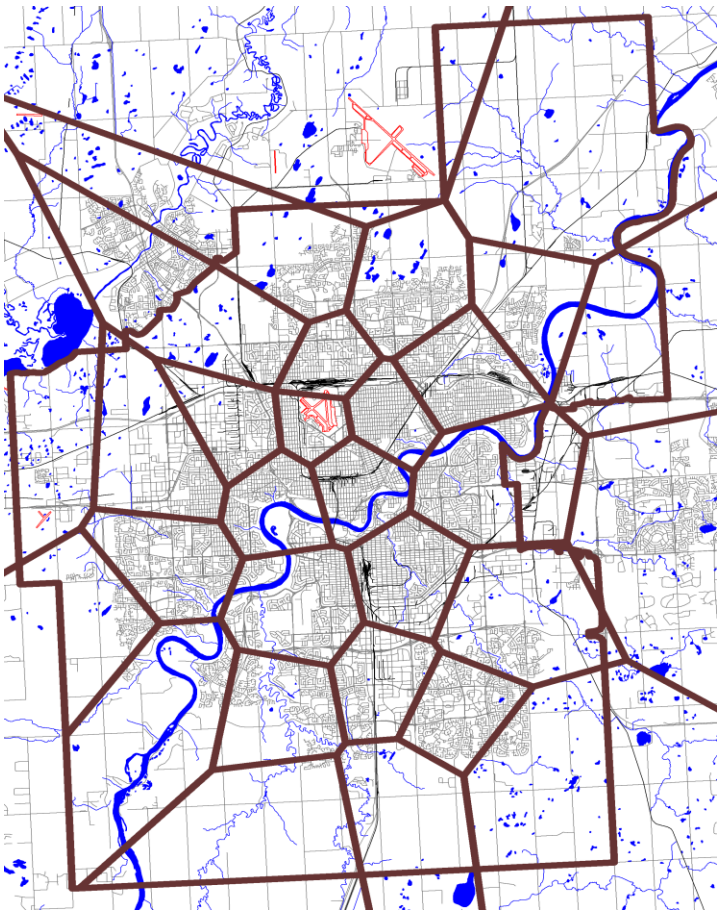


Figure 3.4 - Digitized City of Edmonton's rain gauge network Thiessen polygon map

Radar pixels of 1 km^2 in size that lie inside each red rectangular area on Figure 3.5 were assumed to contribute to each respective Thiessen polygon and used in computing storm characteristics. The rain gauge station reference number associated with each polygon is also shown.

Within the dataset's period of record, certain City rain gauges were either replaced and relocated nearby; hence, the reason why some polygons are associated with more than one rain gauge number.

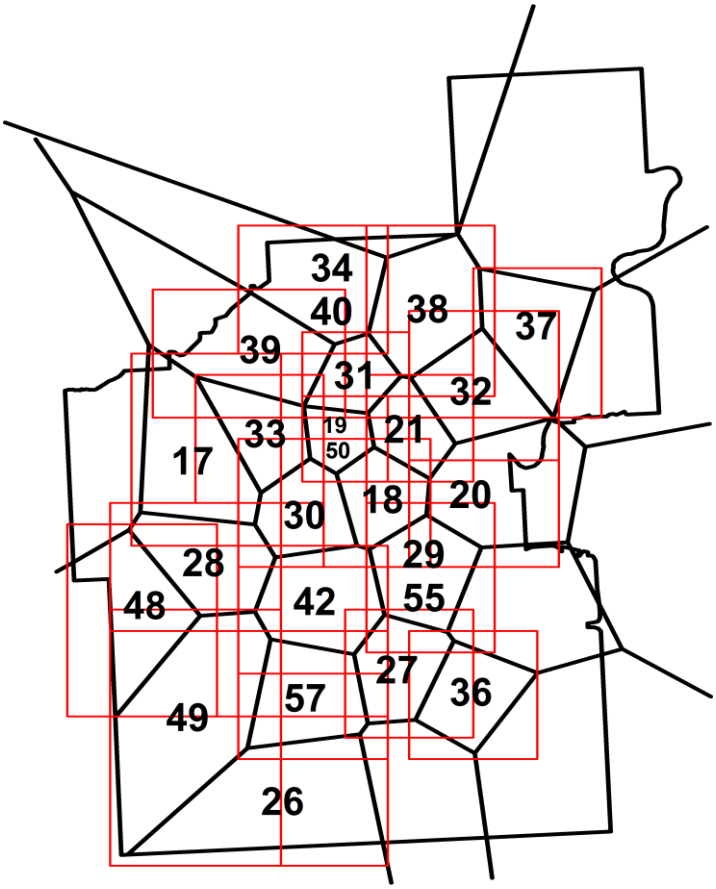


Figure 3.5 - Outline of assumed radar pixels which represent rainfall over each respective Thiessen polygon

4 Model Framework and Results

The following sub-sections were written in the same chronological order as was used to complete the analysis.

4.1 Determination of Analysis Dataset

The following sub-sections describes the steps taken to: 1) identify relevant events from the RERSD, 2) produce new sub-region specific storm characteristics and, 3) partition the relevant storms into two groups for calibration and validation of the regression analysis.

4.1.1 General Selection of Storms for Model Analysis

A summary list of all available RERSD events and their corresponding aggregated information was obtained and presented in an Excel spreadsheet. The list includes 3,414 storms with no less than 5 mm in maximum total rainfall. Compressed (zip format) data files for each storm was retrieved from the corporate archives and uncompressed for use.

The identification of relevant storms was accomplished by completing the following tasks:

1. Storms whose areal extent, as described by the smallest rectangle which encompasses the entire event, did not include at least a portion of the City of Edmonton's territory were not retained for further processing. This step was completed within Excel using intrinsic logic functions and resulted in 832 storms that were retained for further analysis.

2. Additional selection tasks, as elaborated below, were completed to isolate only storms that occurred during floodings. A total of only 25 storms were subsequently identified as potential candidates for further processing. Thematic maps of each storm's total precipitation were created from the

appropriate grid files using the default Kije Sipi Ltd/RadHyPS Inc. color scale with Golden Software's Surfer contouring software. Processing of the storms' total rainfall information required re-projection of the data to the UTM Zone 11 (i.e. the UTM Zone of the Carvel radar instrument) to ensure a common geographic frame of reference considering the City of Edmonton lies in UTM Zone 12 (see Section 3.2.5.3 for clarification). The following information was added to the maps: 1) the City's rain gauge Thiessen polygons as a separate overlay, 2) the radar's location and, 3) labels identifying the amount of corresponding basement floodings per storm for each polygon. These maps helped further identify which storms best corresponded with each flooding event. A visual inspection of the thematic maps revealed that several storms did not show the presence of rainfall over the City limits. As a result, these were discarded from further analysis. A total of 7 storms were eventually removed from the list due to the absence of any rainfall within the City limits. However, in a few marginal cases, several storms were retained for potentially contribution to flooding based on the previous assessment. Specifically, three periods of flooding were found to exhibit this issue: June 29 to July 5, 1998 [storms 1237456405, 1382032526]; July 9 and 13, 1999 [storms 1496568720, 1558109325] and July 2 and 6, 2004 [storms 1387611681, 1474785606]. In these cases, the storms with the highest overall total rainfall over the flooded areas within the City were subjectively assessed and retained as the most likely candidates for contributing to the floodings. The following storms were deemed as being linked to basement floodings: June 29 to July 5, 1998 [storm 1382032526]; July 9 and 13, 1999 [storm 1496568720] and July 2 and 6, 2004 [storm 1387611681]. Total rainfall maps of the 6 storms discussed are presented in Appendix A with each respective date's associated basement floodings overlaid. The storms that exhibited rainfall over the City but were not selected, were nevertheless retained for further analysis; however, were not associated to any floodings. At the conclusion of this last step in the selection process, the list of storms corresponding to the flood events was reduced to 10.

Considering that a large number of storms still remained to be assessed, the mapping and data handling processes were automated. Software script files were created that used the total precipitation grid files in order to produce contour maps for visualization of the remaining events of the RERSD. The scripts were written in Golden Software's Surfer Scripter (Surfer's native coding environment). One script file, applied on a yearly basis, systematically rescaled and transformed the individual storm information from its relative geographic framework to that of the UTM Zone 11. The second script produced the color contour maps on a per year basis. Once completed, an overlay of the rain gauge Thiessen polygons were added for use as a geo-reference to identify which City sectors were affected by basement floodings. This mapping approach allowed for visual inspection of whether or not any rain fell over the City of Edmonton during any given storm. Storms that did not show rainfall within the City limits were discarded from the list, leaving only an overall total of 583 storms for subsequent analysis. For further information regarding the Golden Software's Surfer Scripter code used, the reader is referred to Appendix B. Figure 4.1 shows a typical example of a color-coded thematic map for total storm rainfall.

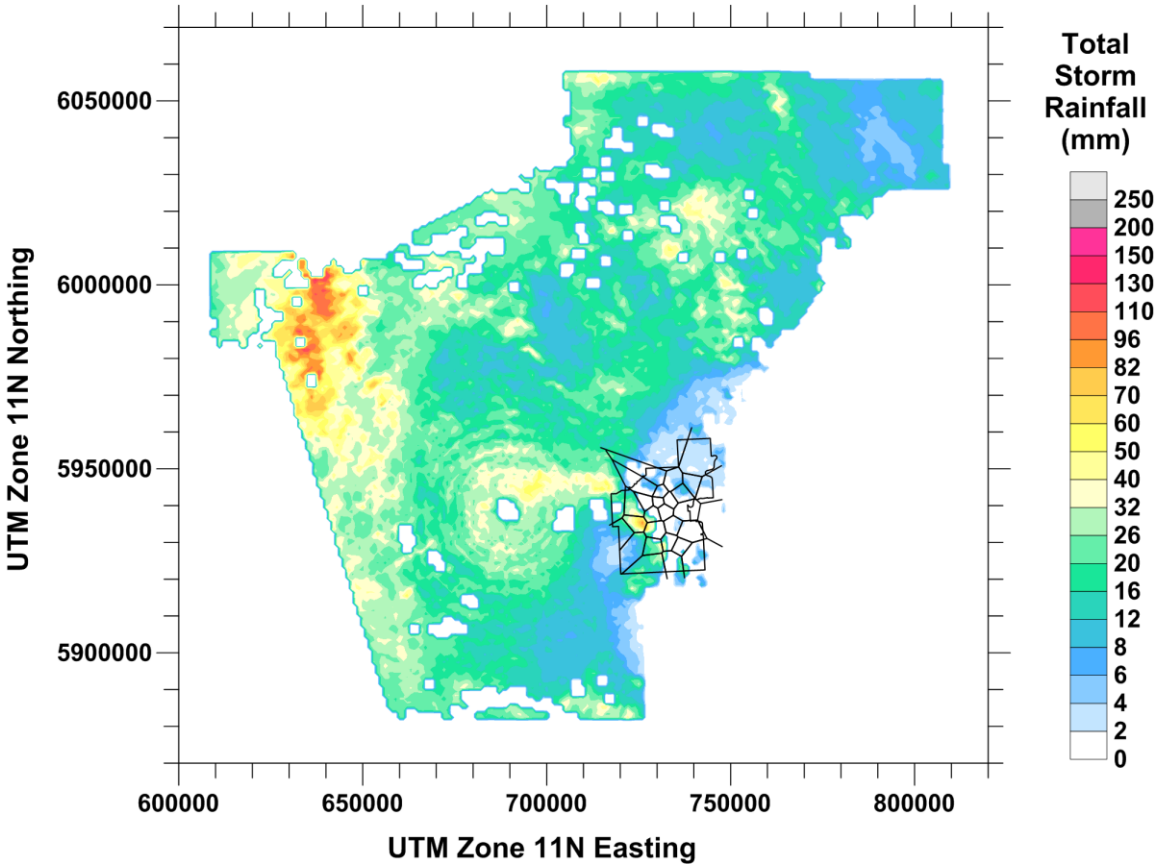


Figure 4.1 - Sample Edmonton database storm mapped in Surfer

4.1.2 List of Storm Classification Elements

A new set of City rain gauge polygon-specific storm characteristics was generated for each storm of interest, as per the final list of storms elaborated in Section 4.1.1. These new storm characteristics will henceforth be referred to as the Edmonton Rainfall Storm Polygon Dataset (ERSPD). Scripts were developed to compute the following ERSPD event characteristics using the RERSD files. These include:

- Average total rainfall (mm)
- Peak total rainfall (mm)
- Average rainfall intensity (mm/hr)
- Peak rainfall intensity (mm/hr)
- Average rainfall duration (hr)

The average total rainfall, measured in millimetres of rainfall depth, refers to a spatial mean of the total amounts of rainfall measured for each radar pixel deemed to have contributed to the rain gauge polygon for the duration of the storm. Similarly, the average rainfall intensity (measured in millimetres per hour according to rainfall depths obtained from 5-minute interval data). The average rainfall duration (measured in hours) also refers to a spatial mean, albeit with respect to the maximum 5-minute rainfall intensity of each radar pixel over the duration of the storm and, with each individual radar pixel's rainfall storm duration, respectively. In contrast, the peak total rainfall (measured in millimetres of rainfall depth) and the peak rainfall intensity (measured in millimetres per hour according to rainfall depths obtained on a 5-minute basis) both refer to the maximum value among the radar pixels deemed to have contributed to the rain gauge polygon during the course of the storm. Notice that each variable can plausibly be linked as a contributing factor to flooding.

4.1.3 Dataset Partitioning for Calibration and Validation

The purpose of segregating the entire storm dataset into two is to be able to test the validity of the fitted model with a dataset that is different from that used fitting. The list of storms, per respective polygon, was divided into two subsequent subsets, one used for calibration and the other used for validation. A subjective 80/20 split was selected whereas 80% of the storms on each polygon list of storms were placed into the calibration set while the other 20% were added to the validation set. The partitioning was done in two steps:

1. The storms that corresponded with basement floodings were considered. Note that the number of individual flood-positive storms recorded for each polygon varied. Generally, all but one or two flood-positive storms were retained for the calibration set. Storms identified as the 33rd percentile most and least numerous total basement flooding events (or storms closest to the 33rd percentile when the amount of flood-positive storms allowed). When a polygon only had two flood-positive storms, the storm with the largest amount

of floodings was retained for the calibration set while the other was added to the validation set. If only one flood-positive storm was associated with a polygon, that storm was retained for the calibration set.

2. The remaining storms (i.e. flood-negative storms) were sorted by descending order of magnitude of peak total rainfall. The storm with the third most greatest peak total rainfall value and every fifth storm thereafter was retained for the validation set, the remaining storms were used for calibration. This effectively represented 20% of storms on the final list which formed the validation set.

However, in order to have the characteristics of the flood-positive storms within the calibration and validation sets better represent each other, some slight modifications were made to the above-mentioned partitioning approach. For further discussion on this is, refer to Section 4.3.1.1 regarding the use of a peak total rainfall floor threshold value and refer to Section 5.3.1.5 for discussion of its impacts.

4.2 Multivariate Analyses

This section will discuss the spatial description offered by the radar data for each sub-region and the characteristics of the ERSPD for the entire dataset, as well as those for only Polygon 32. The following sub-sections present: 1) the characteristics of the various datasets used in the stepwise regression analysis, 2) the methodology used to produce the stepwise regression analysis and, 3) the stepwise regression analysis results.

4.2.1 Introduction

Table 4.1 serves to illustrate the degree of available spatial information for each respective rain gauge polygon. It lists the total number of radar pixels associated with each rain gauge polygon, the amount of which are masked due to the RERSD quality control process (see Section 3.2.5.5), the percentage of masked pixels with respect to overall associated number of pixels and, the actual number of pixels

available for analysis per polygon. The fourth column presents information on the masked data as follows:

- polygons associated with rain gauges 18, 19&50, 31, 34&40, 36, and 39 have the entire set of radar pixels associated with that polygon masked and; therefore, there is currently no available storm information for these areas;
- polygons associated with rain gauges 17, 20, 21, 27, 29&55, 33, and 48 have between 75 and 99.99% of the radar pixels associated with that polygon masked;
- polygons associated with rain gauges 28, 30, 38, 42, and 57 have less than 75% of the radar pixels associated with that polygon masked;
- polygons associated with rain gauges 26, 32, 37, and 49 have less than 40% of the radar pixels associated with that polygon masked.

The fifth column in Table 4.1 lists the respective total number of pixels remaining per polygon after masking. It also represents the information, as follows:

- polygons associated with rain gauges 18, 19&50, 31, 34&40, 36, and 39 have no radar pixels associated with their rain gauge polygons due to masking;
- polygons associated with rain gauges 21, and 29&55 have only one radar pixels associated with that rain gauge polygon due to masking of all other radar pixels; finally,
- polygons associated with rain gauges 17, 20, 26, 27, 28, 30, 32, 33, 37, 38, 42, 48, 49, and 57 all have more than one radar pixels left for analysis due to masking of certain radar pixels associated with that rain gauge polygon.

The sixth column lists the number of storms which produced floodings within the available storm event dataset. A cursory examination of the table reveals that rain gauge polygon 32 is the best candidate for further analysis. Its dataset has one of the highest number of storms with flooding data (7) and also has one of the largest amount of available pixels, as well as the lowest percentage of masked pixels among other rain gauge polygon datasets. Note that polygon 21 and 29&55 have

only one remaining pixel contributing to the storm characteristics computation, their peak and average values are the same.

Table 4.1 - Proportion of polygon masking

Rain Gauge	Number of Pixels	Number of Masked Pixels	Percent of Masked Pixels	Actual Number of Pixels Available for Analysis	Number of Storms with Floodings
17	63	52	82.54%	11	7
18	30	30	100.00%	0	-
19 & 50	16	16	100.00%	0	-
20	49	39	79.59%	10	6
21	25	24	96.00%	1	4
26	91	26	28.57%	65	2
27	36	30	83.33%	6	5
28	48	29	60.42%	19	5
29 & 55	42	41	97.62%	1	7
30	36	25	69.44%	11	7
31	20	20	100.00%	0	-
32	49	19	38.78%	30	7
33	36	29	80.56%	7	6
34 & 40	42	42	100.00%	0	-
36	36	36	100.00%	0	-
37	42	7	16.67%	35	3
38	48	26	54.17%	22	5
39	54	54	100.00%	0	-
42	42	19	45.24%	23	-
48	63	51	80.95%	12	3
49	96	38	39.58%	58	3
57	42	29	69.05%	13	1

Figure 4.2 presents the 16 rain gauge polygons for whish storm data information was available for analysis. The study covered the most of the City of Edmonton's territory with various degrees of data availability throughout the area as listed in Table 4.1.

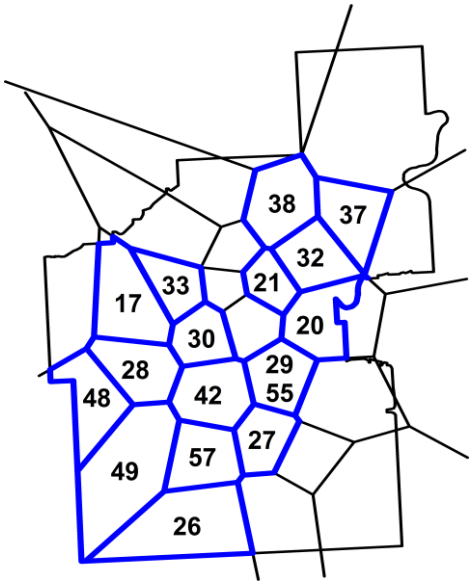


Figure 4.2 - Polygons for which storm data was available for analysis

4.2.1.1 *Characteristics of the Entire Dataset*

Figure 4.3, 4.4, 4.5, 4.6, and 4.7 respectively show scatterplots of the reported basement floodings for all storm-polygons against average total rainfall (mm), peak total rainfall (mm), average rainfall intensity (mm/hr), peak rainfall intensity (mm/hr) and, average rainfall duration (hr). The same three most severe storm-polygon floods are identified each of the figures. Note that the values identified in red were eventually considered as having no floodings as their peak total rainfall value did not exceed the subjectively chosen, yet common practice in urban hydrology of using a 5 mm threshold for initial abstraction and interception. Refer to Section 5.3.1.5 for a discussion regarding the use of the 5 mm urban interception threshold in this study. A common theme among the storm characteristics, as illustrated in the 5 figures, is that the zeroed storms (in red) tend to be located in the lower region of variable values, though associated flooding numbers range from 1 to 312. The majority of storm characteristic values can be found to be associated with less than 100 basement floodings. Specifically, the distribution of values in Figures 4.3, 4.4, 4.5 and 4.6 show a fairly uniform pattern across the entire range of values with a slight tendency to increase in floodings with increased storm characteristic. In contrast, the value distribution in Figure 4.7 seems to exhibit a clustering of values in two regions. This clustering is possibly indicative of the influence of a different rainfall

regimes or an underlying hydrologic characteristic (e.g. Time of Concentration). This will be further explored in Section 5.3. Note that the 3 storms, which produced floodings in excess of 100 basements appear to be outliers in Figure 4.3. A total of 70 flood-positive storms remain out of the original 86 that were identified prior to the threshold discrimination step.

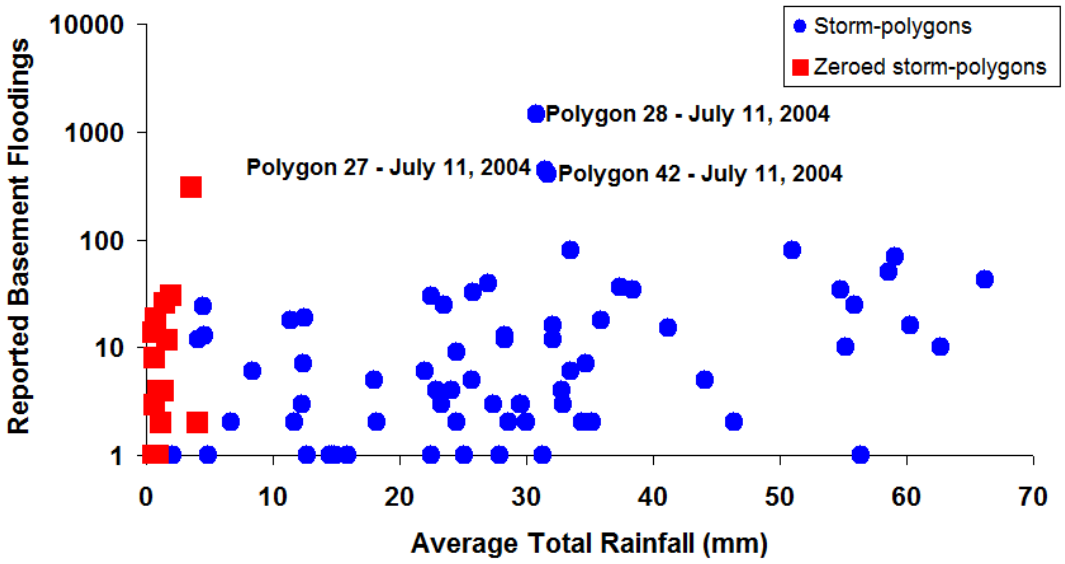


Figure 4.3 - Reported floodings associated with average total rainfall

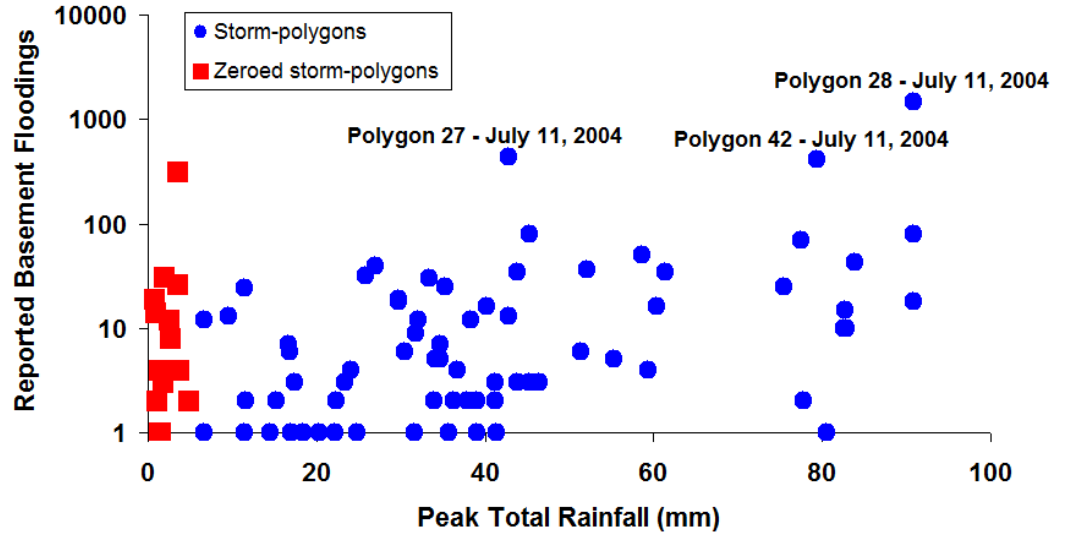


Figure 4.4 - Reported floodings associated with peak total rainfall

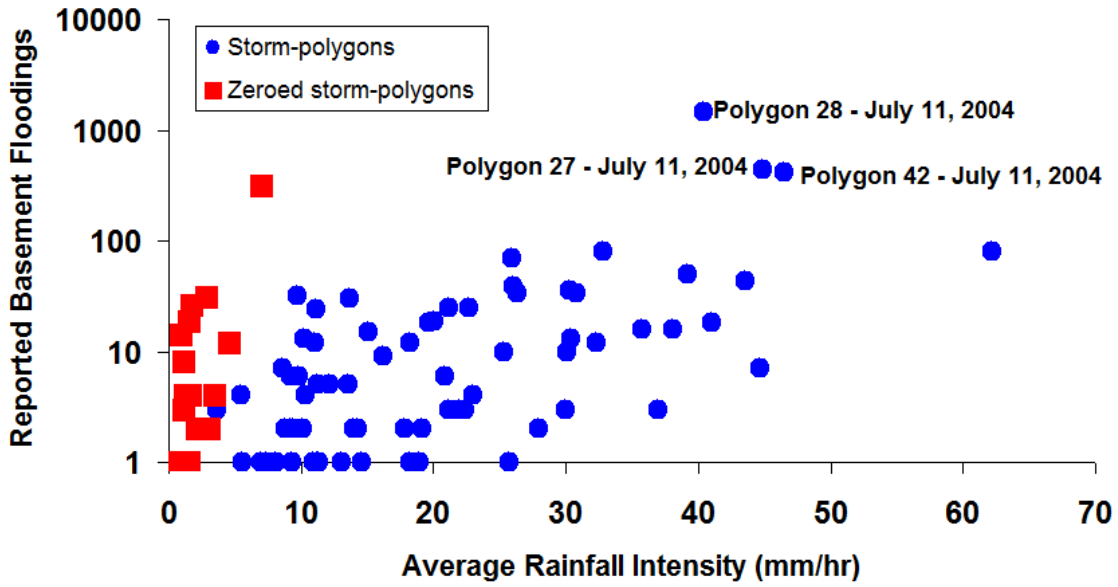


Figure 4.5 - Reported floodings associated with average rainfall intensity

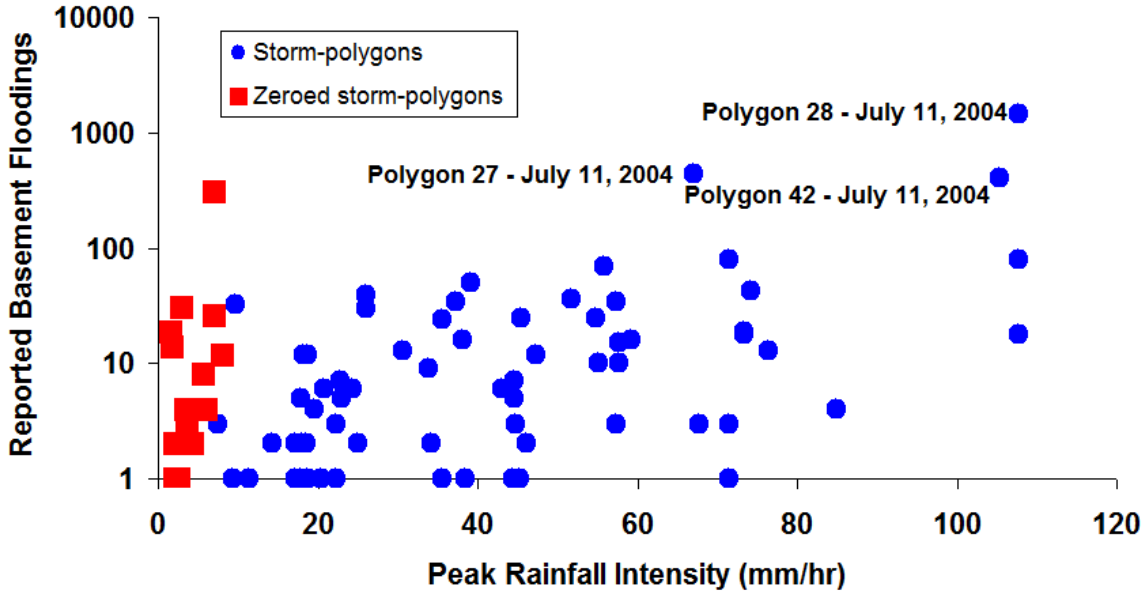


Figure 4.6 - Reported floodings associated with peak rainfall intensity

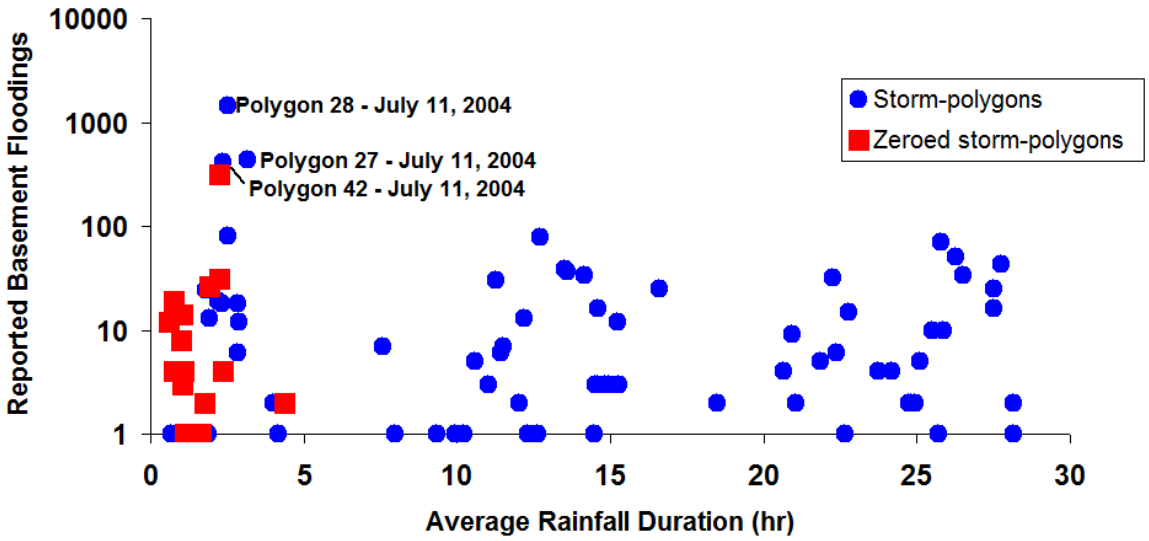


Figure 4.7 - Reported floodings associated with average rainfall duration

An indication of the spread of all five storm characteristic values for all 7,678 storms in the database can be seen in Figure 4.8. The box-plot is presented on a vertical log scale. The values shown for each box-whisker plot represent the following storm characteristic statistics (from top to bottom), the maximum, the 75th quartile, the median, the 25th quartile and the minimum values. The distribution of values for the three average storm characteristics are similar as those for the two peak storm characteristics. The peak value distributions are, as expected, higher than that of average characteristics. The disparity between average and peak values illustrate the presence of spatial variability in rainfall storm characteristics. Also, the figure effectively shows that the majority (those below the lower 75th quartile) of recorded storms for the 11 years of observed data exhibit storm characteristics lower than 10 (mm, mm/hr, hr) in terms of average total rainfall, peak total rainfall, average rainfall intensity, peak rainfall intensity, and average rainfall duration. A short discussion about the dataset's average rainfall duration is presented in Section 5.3.1.5. Also, refer to Section 3.2.5.6 with regards to ceiling radar rainfall intensity values due to hail.

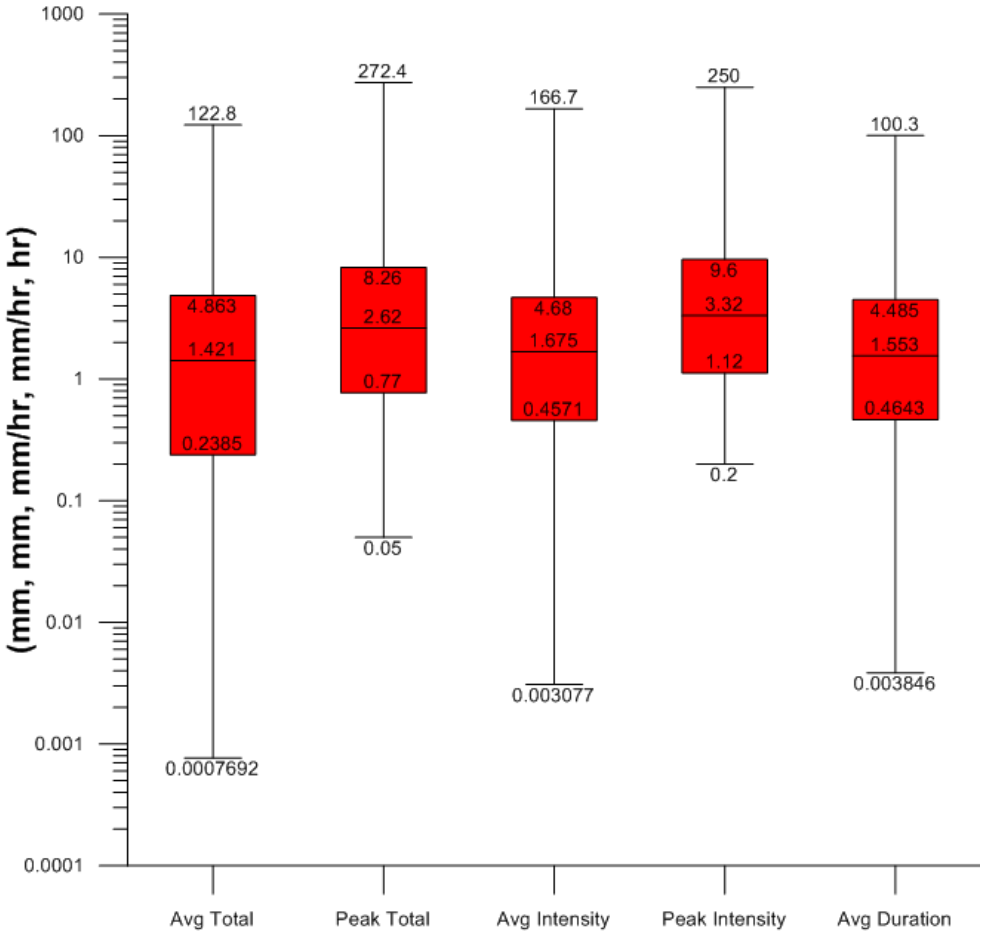


Figure 4.8 - Box-plot of entire polygon dataset storm characteristics

The full storm dataset for floodings could not be plotted on Figure 4.8 since zero values cannot be plotted on a logarithmic scale. Figure 4.9 shows the distribution of flood-positive storm characteristics values and the number of basement floodings. It is important to point out that 99.1% of the entire storm-polygon dataset (7,608 out of 7,678 storms) are not associated with any floodings, while the storms and the distribution of their characteristics, as shown in Figure 4.9**,** only represents 0.9% of storms. A comparison of the storm characteristics shown in Figure 4.9 with those of Figure 4.8, yields the following observations:

- all storm characteristics exhibit higher minimum, 25th quartile, median, and 75th quartile values; however,

- they also have lower maximum values, and;
- the distribution of the number of basement floodings per storm-polygon is such that the majority (up to 75th percentile) of storms causing floodings is below 20 basements; whereas, the entire range of flood-positive storms ranges from 1 to 1,450 flooding(s) per polygon, of which only 3 were above 80 flooded basements.

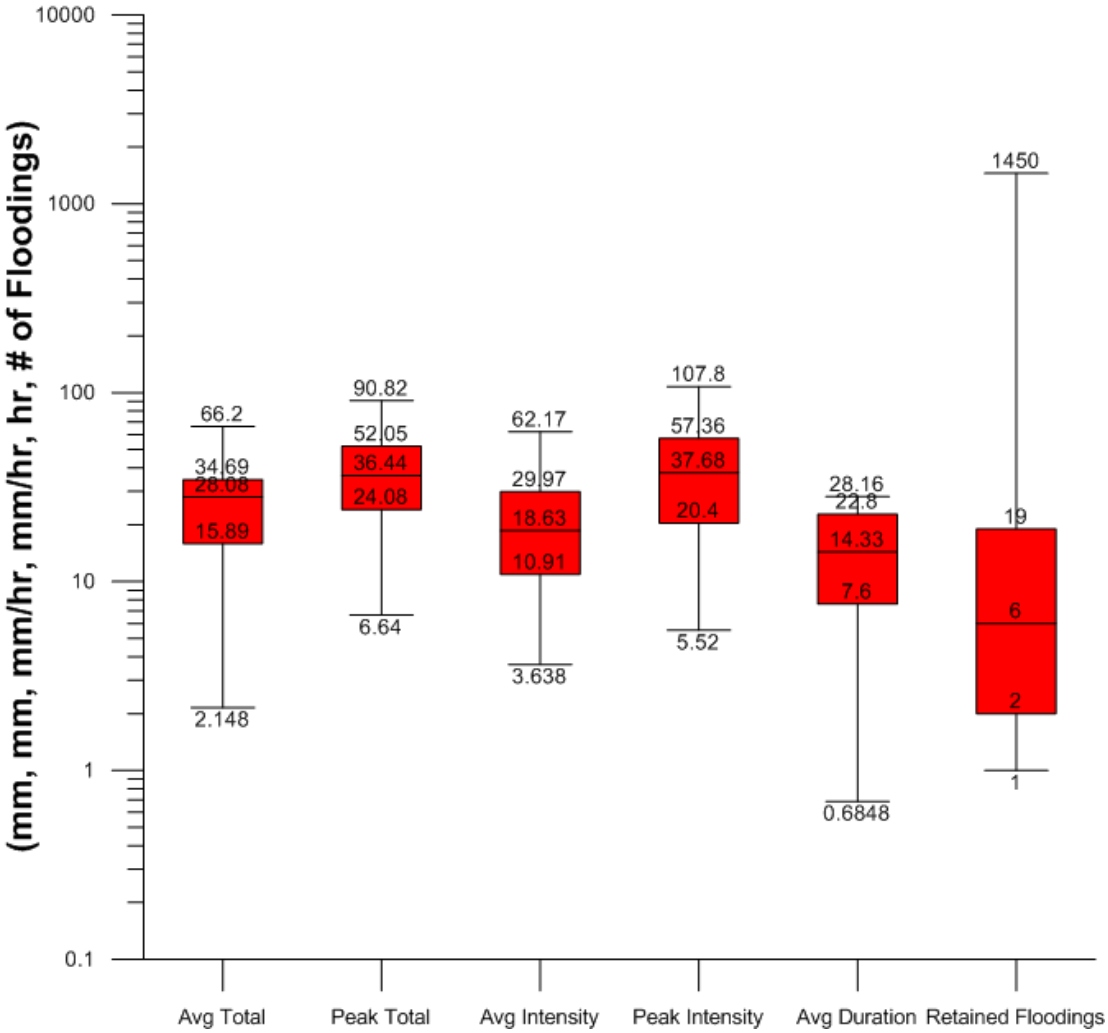


Figure 4.9 - Box-plot of non-zero storm floodings for all polygons

4.2.1.2 Characteristics of the Polygon 32 Dataset

As revealed in Table 4.1, the high level of masked pixels leads to a limited coverage of reasonably confident radar rainfall estimates for the City of Edmonton (see Section 3.2.4.6 for more information regarding pixel masking). Consequently, in an

effort to reduce computational requirements while using the most representative data of the rainfall regime over that area, therefore the desired storm rainfall severity relation was derived using a single polygon. In order to best describe the actual rainfall regime over the area to be analysed and, to offer a meaningful extreme data trend, the candidate polygon would need to: 1) minimize overall polygon pixel masking over its territory, 2) maximize the number of pixels that were used to produce the storm characteristics over its polygon and, 3) maximize the number of flood-positive storms associated with the polygon. Of the available polygons, polygon 32 is the best candidate in terms of the aforementioned criteria.

While the entire ERSPD contains 7,678 storms, only 511 of those occurred over polygon 32. Nevertheless, the distribution of polygon 32 storm characteristics are quite similar; hence, representative of the overall rainfall storms when comparing those of the polygon 32 dataset, as illustrated in Figure 4.10, to that of the entire storm-polygon dataset (as shown in Figure 4.8). While the minimum, 25th quartile, median, and 75th quartile values are generally the same, differences in the maximum values for peak total rainfall, average rainfall intensity, and peak rainfall intensity are notably of the same magnitude. Appendix C contains a more detailed assessment with regards to the similarity of these two above-mentioned datasets. The test results concluded that polygon 32's dataset has a similar data distribution as that of the entire ERSPD dataset, except for peak total rainfall and peak rainfall intensity which differ significantly.

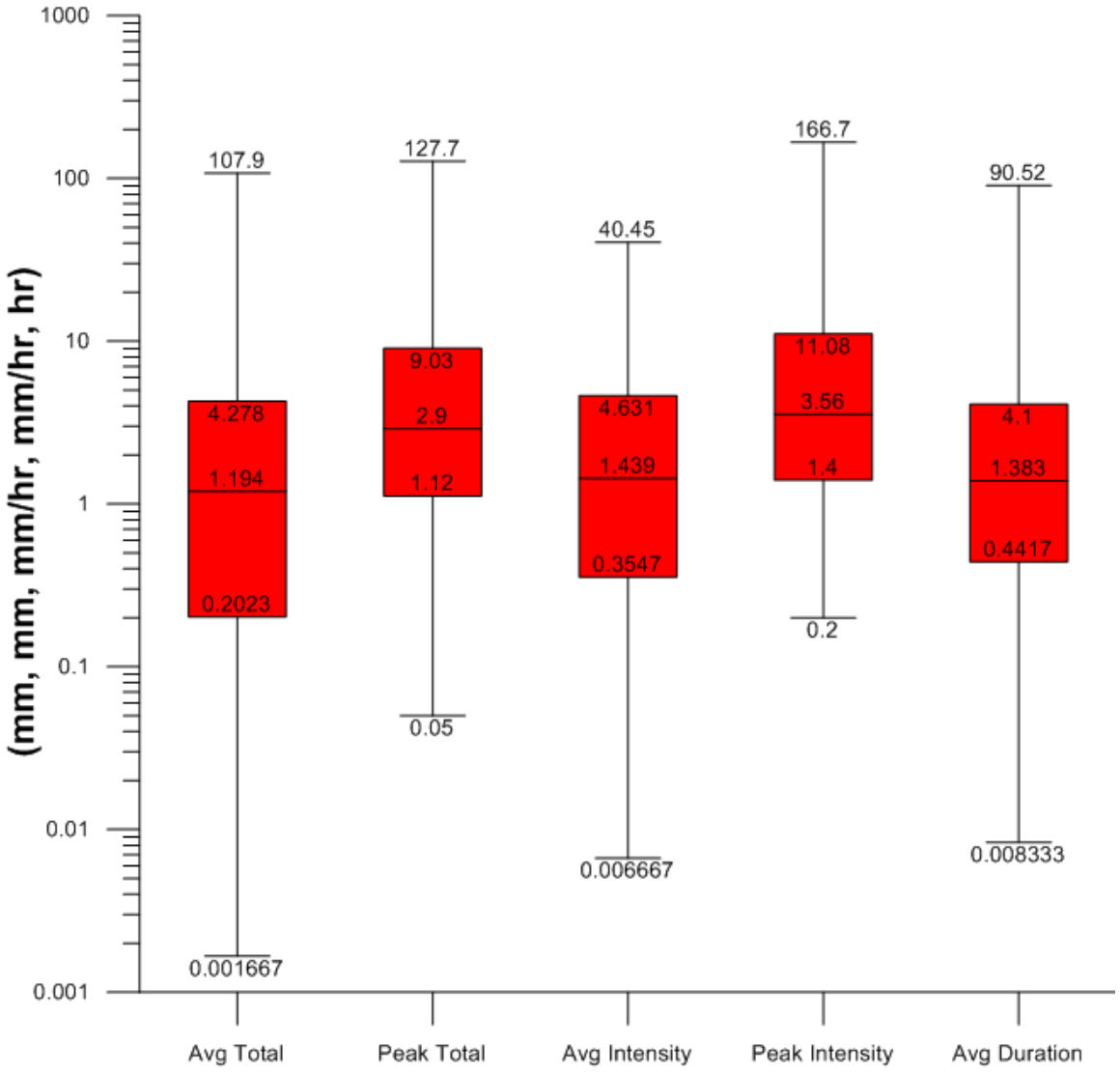


Figure 4.10 - Box-plot of polygon 32 dataset storm characteristics

The dataset for polygon 32 consists of 511 storms, of which 7 are known to have produced basement floodings. As previously mentioned, these 7 flood-positive storms represent only 1.4% of all storms, a similar proportion as that of the entire storm-polygon dataset. The distribution of the flood-positive storm characteristics can be seen in Figure 4.11. While the average and peak rainfall intensity values for polygon 32 flood-positive storms are restricted to values which generally range between the 25th and 75th quartile of the overall storm-polygon dataset, as seen in Figure 4.9. The remaining three storm characteristics generally represent well the

overall storm-polygon dataset's characteristics distribution with the exception of the maximum number of floodings. It is nowhere near the number of floodings of the overall storm-polygon dataset (i.e. 1,450 basement floodings) yet still represents a value above the 75th quartile.

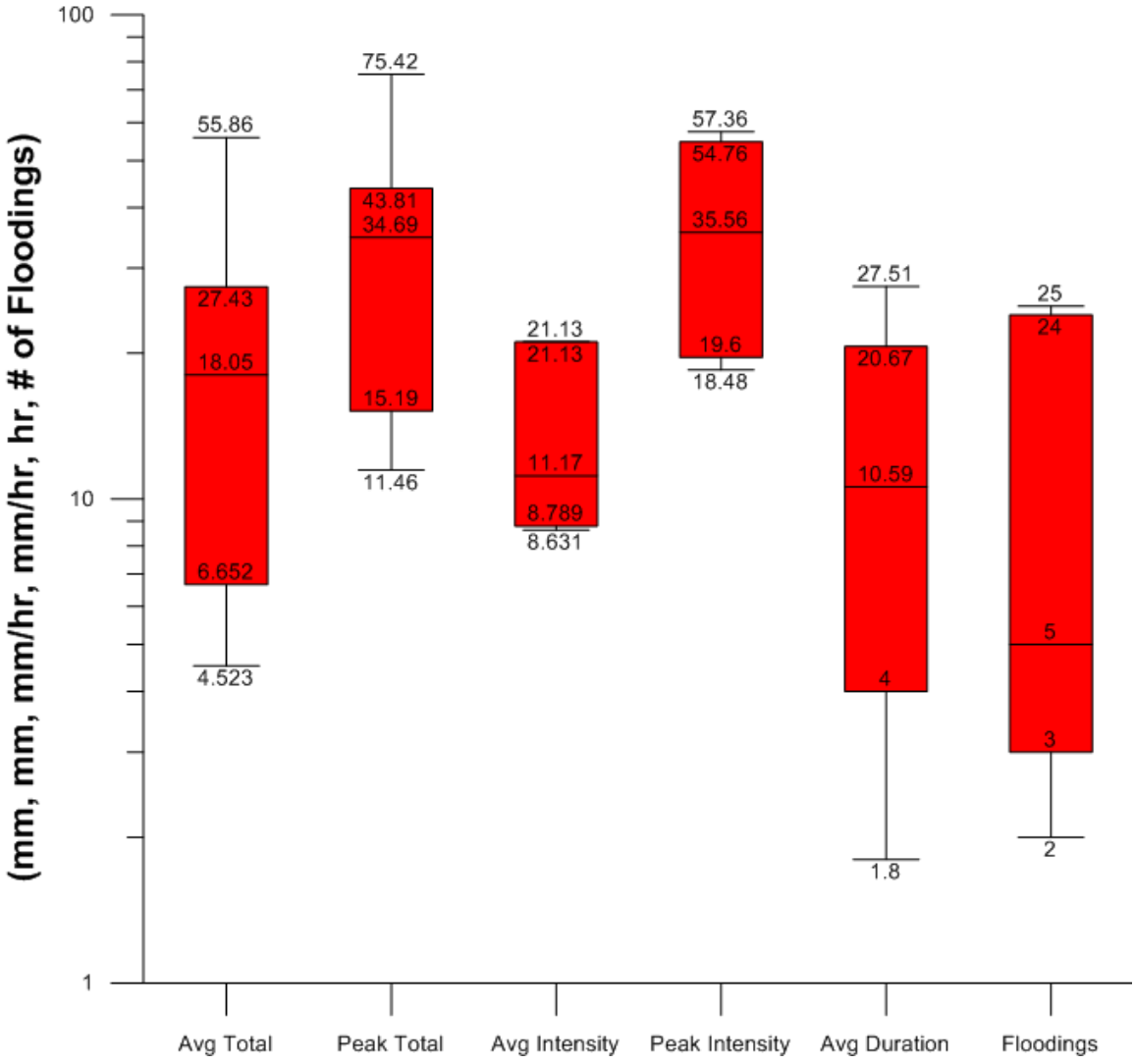


Figure 4.11 - Box-plot of polygon 32 non-zero floodings dataset storm

4.2.2 Calibration and Validation

As previously discussed in Section 4.2.3, the calibration and validation datasets are distinct storm data sets that were tested separately in order to produce the best

storm classification model. Each polygon's respective storm list was accordingly divided in two.

4.2.2.1 Introduction

This section mostly elaborates on the representativeness of polygon 32's calibration and validation datasets, in terms of its storm characteristics' distributions. The storm characteristics' distribution of polygon 32's calibration dataset, as shown in Figure 4.12, is statistically similar to that of the entire dataset of polygon 32, as seen illustrated in Figure 4.10. Since the storm characteristics' distribution is similar to that of the overall polygon 32 dataset, it is also similar to the overall storm-polygon dataset, as shown in Figure 4.8. They have a minimum, a 25^{th} quartile, a median, and a 75^{th} quartile value that are generally similar, while exhibiting differences in only the maximum values of approximately 100 mm or mm/hr in magnitude for peak total rainfall, average rainfall intensity, and peak rainfall intensity. Note that the dataset consists of 408 individual storm events. Appendix C has a more in-depth assessment of the similarity between the polygon 32's entire dataset compared to polygon 32's calibration dataset. The test results concluded that polygon 32's calibration dataset has a similar data distribution as that of the entire polygon 32's dataset.

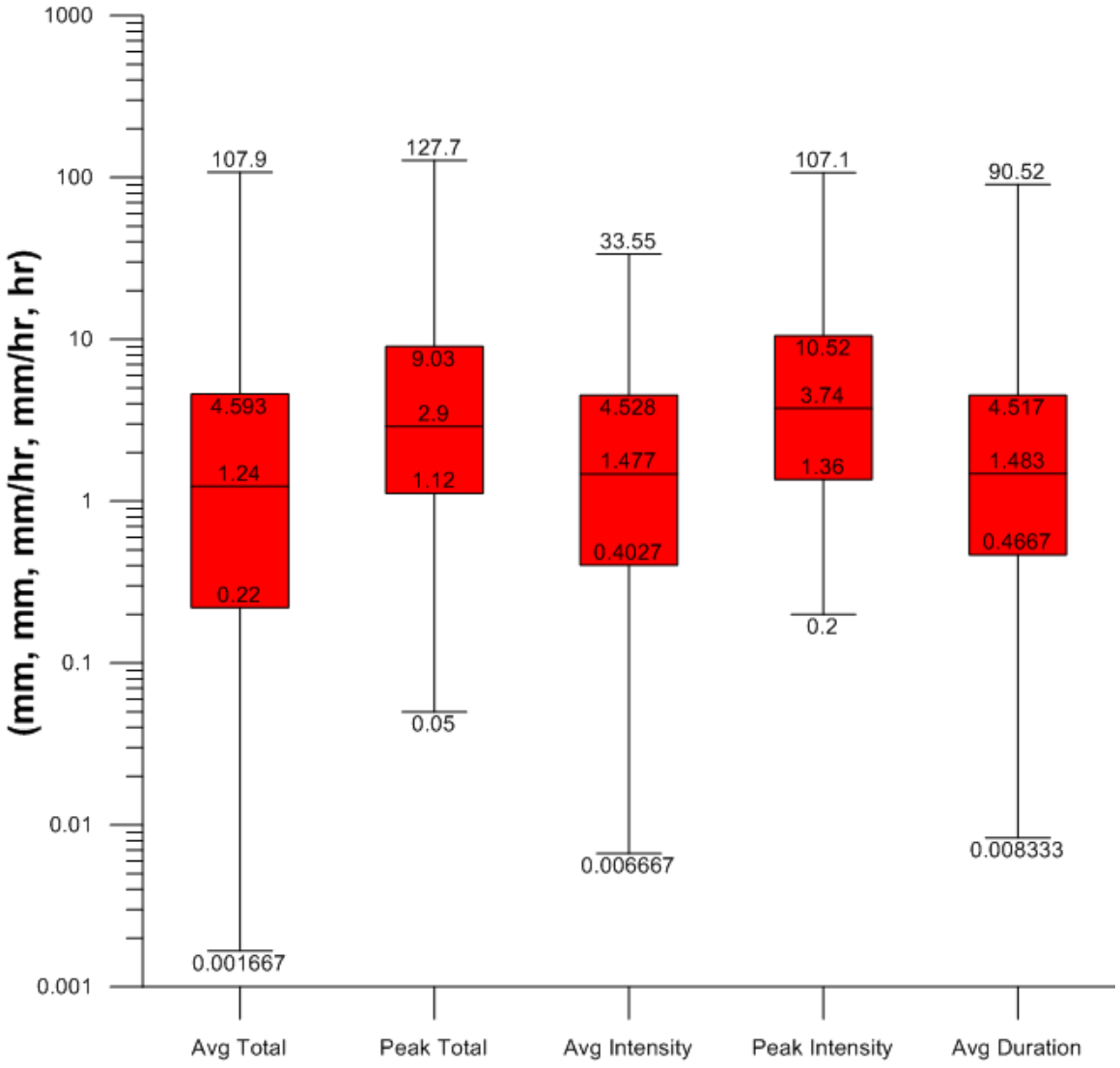


Figure 4.12 - Box-plot of polygon 32 calibration dataset storm characteristics

Table 4.2 lists the storm characteristics for the five flood-positive storms in the calibration dataset. When comparing values found in Table 4.2 to those of Figure 4.11, which represent the distribution of storm characteristics for the polygon 32 dataset as a whole, one realizes that the choice of flood-positive calibration storms represent a good spread of values spanning the entire range as presented in Figure 4.11. Also, as was the case with the overall flood-positive dataset of polygon 32, the average and peak rainfall Intensity values for polygon 32's calibration flood-positive storms set are restricted to values which generally range between the 25th and 75th

quartiles of the overall storm-polygon dataset. However, the average total rainfall, peak total rainfall and average rainfall duration generally represent the overall storm-polygon dataset's characteristics distribution aside for the maximum number of floodings which is above the 75th quartile but not close to the value of the overall storm-polygon dataset (i.e. 1,450 basement floodings).

Table 4.2 - Polygon 32 calibration set of storms with non-zero floodings

Average Total Rainfall (mm)	Peak Total Rainfall (mm)	Average Rainfall Intensity (mm/hr)	Peak Rainfall Intensity (mm/hr)	Average Rainfall Duration (hr)	Number of Floodings
55.86467	75.41998	21.12667	54.76	27.50833	25
4.523334	11.46	11.16933	35.56	1.8	24
18.048	34.69	12.192	44.64	10.59167	5
22.90433	36.66	10.37867	19.6	20.66667	4
6.652	15.19	8.789333	18.48	4	2

Polygon 32's validation dataset is made up of 103 individual storms whose characteristics distribution are presented in Figure 4.13. Compared to the storm characteristics of the calibration set (Figure 4.12), its storm characteristics are similar in value distribution except for having smaller maximum values for average total rainfall, peak total rainfall and average rainfall duration but larger maximum values for average and peak rainfall intensities. When compared against the entire polygon 32 dataset (Figure 4.10), polygon 32's validation data set fairs similarly to the previous comparison, for all but the average total rainfall, peak total rainfall and average rainfall duration's maximum values that are close to the values of the entire polygon 32 dataset while those three storm characteristics' maximum values still represent values in excess of the entire polygon 32 dataset's 75th quartile values. Furthermore, when comparing storm characteristics values as shown in Figure 4.13 (polygon 32's validation dataset) and Figure 4.9 (overall storm-polygon dataset), it is observed that all but each characteristics' maximum value are of generally comparable magnitudes, whereas the maximum values are significantly smaller in the validation data set yet, they still represent values greater than their 75th quartile. Appendix C presents a computed assessment, per storm characteristics and

reported basement floodings, of the similarity between the polygon 32's validation dataset against that of polygon 32's calibration dataset. The test results concluded that polygon 32's validation dataset has a similar data distribution as that of the polygon 32's calibration dataset

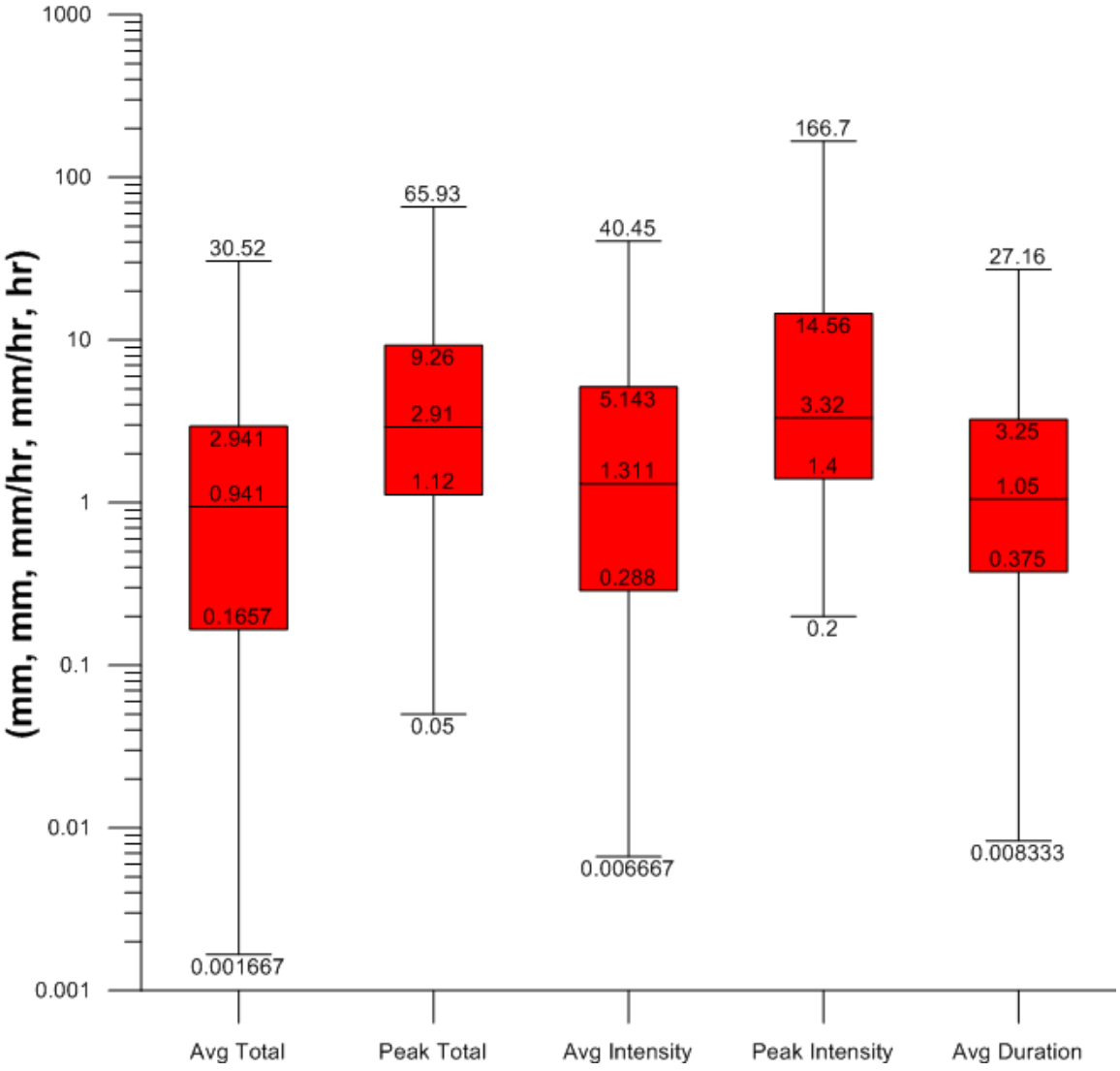


Figure 4.13 - Polygon 32 storm database validation characteristics box-plot

Table 4.3 lists the only two flood-positive storms and their characteristics in the validation storm data set. Comparing Table 4.3 values to those of Table 4.2 reveals that the flood-positive validation storms are within the range of values from those of

the calibration set for average total rainfall, peak total rainfall, average rainfall duration and number of reported basement floodings. However, they show values smaller than the smallest and equal to the greatest values for average rainfall intensity and, a storm characteristic greater than the greatest peak rainfall intensity. The values of the two flood producing storm characteristics, when compared the distribution of those in Figure 4.11 (entire polygon 32 storm dataset), represent values within the 25th to 75th quartile for all characteristics except for average and peak rainfall intensity where its values are equal to the maximum values. Also, when examining where along the entire storm dataset for all polygons (as seen in Figure 4.9), the two validation data set of flood-positive storms characteristics lie, it is observed that, while the storm which produced three basement floodings lie within the 25th and 75th quartile range and while storm characteristics for the storm which produced 7 basement floodings have values which lay below the 25th quartile for average total rainfall, peak total rainfall and average rainfall intensity, the remaining three storm characteristics lie at the 25th quartile or between the 25th and 75th quartile. Consequently, it is concluded that the calibration and validation storm datasets are generally representative of each other as well as being representative of the overall storm dataset which includes storm characteristics for all polygons.

Table 4.3 - Polygon 32 validation set of storms with non-zero floodings

Average Total Rainfall (mm)	Peak Total Rainfall (mm)	Average Rainfall Intensity (mm/hr)	Peak Rainfall Intensity (mm/hr)	Average Rainfall Duration (hr)	Number of Floodings
12.35233	16.75	8.630667	22.88	7.6	7
27.427	43.80999	21.12667	57.36	14.94167	3

4.2.2.2 *Methodology*

Initially, an artificial neural network approach was considered but was rejected for its black box approach and its tendency not to yield simple models. A logistic multivariate regression analysis was also considered as a binary approach (flood versus no flood) but that method was rejected due to the loss of flood severity description. Finally, a modified multivariate linear regression analysis was chosen

as the analysis approach due to its ability for continuous modeling of storm severity through the flood-positive range.

The R programming language environment was used to analyse the storm datasets and determine a best fit model. A program was written in R using available program functions (The R Core Team, 2012), see Appendix D for sample code, which was used to systematically cycle through a host of possible variable combinations. The program was used to determine the best fit model through an iterative stepwise linear regressive process. The stepwise regression function used in the program uses an iterative test ratio, the Akaike's Information Criteria, when comparing several equations to discriminate the worst terms fitted as well as justify when to eliminate candidate relations. The ratio iteratively compares the model fit while penalizing the use of too many variables (Helsel & Hirsch, 2002).

The desire to test nested models forced the adoption of a iterative workaround approach; whereas, the nested model with the coefficients in front of each variable within the nested model were systematically changed with a static array of values. The distribution of these values can be described as approaching that of a 2-term Pareto distribution from 0 to 500 and mirrored at 0 for a full range spanning -500 to 500.

To limit the amount of virtual memory being used during the analysis, only the top 100 equations were retained. The best equations were identified by assessing the average percent of false positives in the calibration and validation datasets.

False positives are defined as any flood-negative storm that ranks higher in severity than the lowest ranking flood-positive storm, according to the model being tested. Therefore, the percent false positive value represents the number of false positive storms divided by the total amount of storms in the dataset and subsequently multiplied 100 to give the percentage.

4.2.2.3 Results

As previously mentioned, the approach used to derive the model was a stepwise multiple linear regression. The final result of using this approach produced the best fit model to the floodings term using both visual inspection of a scatterplot and the rank position of flooded events as discussed in this section. The best fit model function is presented as Equation 4.1 and written as:

$$F_r = -0.1060 - 9.177 \times 10^{-11}\left(2D_a - 3P_a\right)^5 + 0.06957 I_a \qquad \textbf{(4.1)}$$

where F_r is the number of reported basement floodings, D_a is the average rainfall duration (hr), P_a is the average total rainfall (mm), and, I_a stands for the average rainfall intensity (mm/hr).

Figure 4.14 and Figure 4.15 illustrate the model fit against the actual recorded number of basement floodings. Notice that the model does not fully capture the actual number of recorded basement floodings but mostly succeeds in ordering the flood-positive storms among the highest values the model generates. As previously mentioned, the success of the model fit by the stepwise regression was measured in terms of percent false positives, which represents a modified hit-rate approach. The hit-rate approach is a fit assessment method typically used for logistic linear regressions where the percent correct category association is reported.

Calibration results show this equation to identify 5 flooding events among the top 41 or, out of a total of 408, which produces 8.8% false-positives. However, in the validation set, the model identifies 2 flooding events among the top 18 out of a total of 103, accounting for a rate of 15.5% false-positives according to the best fit model.

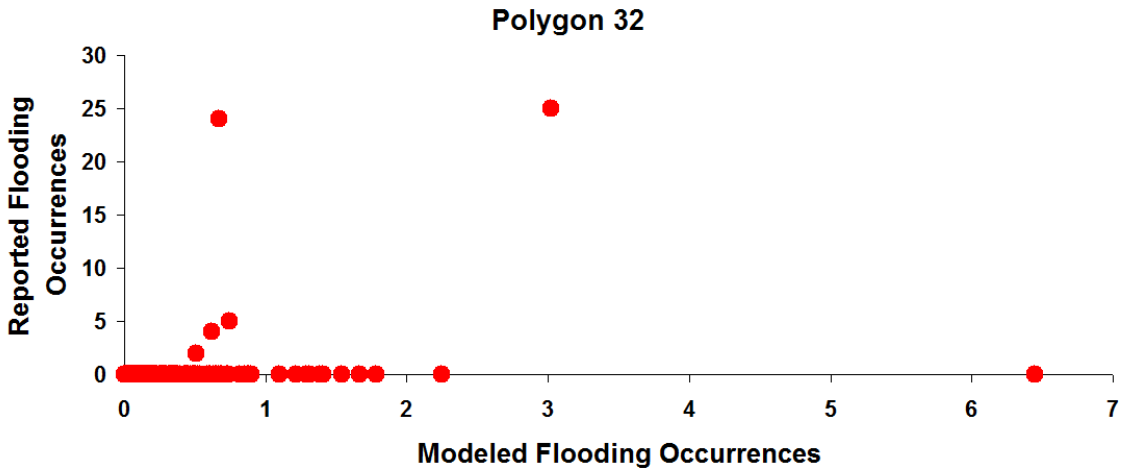


Figure 4.14 - Polygon 32 calibration results

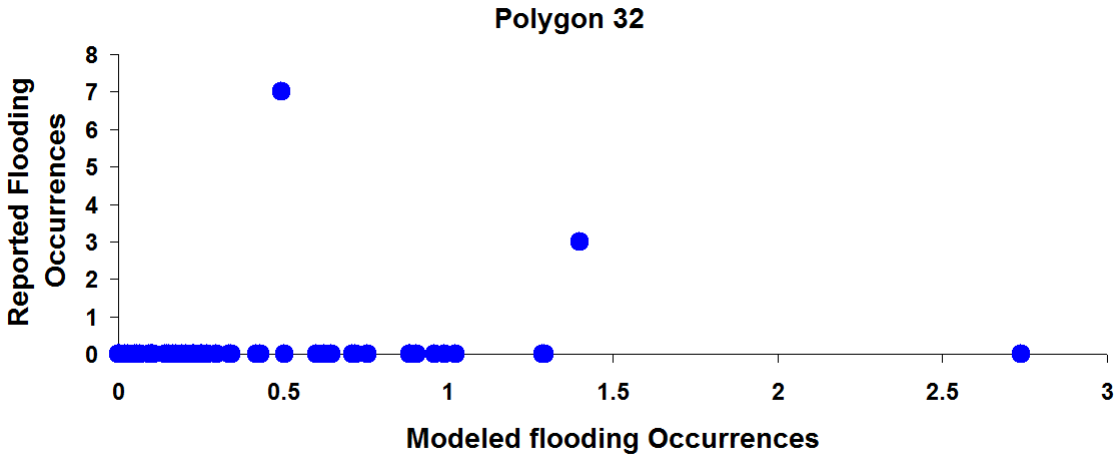


Figure 4.15 - Polygon 32 validation results

- In order to test the validity of the model equation and to assess the practicality of using this relation across the City of Edmonton, a comparison with other Thiessen polygons was completed. Model results from storm datasets associated with the remaining polygons yielded mostly positive results as can be seen in Table 4.4.

No false positive percentage could be calculated for polygon 57's validation set as it did not contain any flood-positive storm.

Note that the three storm datasets with the highest percent false positive, polygon 29 & 55's calibration set, polygon 38's validation set and polygon 42's calibration set, also happened to have among their datasets storms with the second, first, and third most smallest storm average rainfall intensity of all flood-positive events, respectively. This suggests that the model may be overemphasizing the importance of the average rainfall intensity storm characteristic in determining its severity.

However, with respect to the four storm datasets that have the lowest percent false positive, polygon 21's validation set, polygon 27's validation set, polygon 28's validation set, and polygon 37's calibration set, their success cannot as easily be attributed to the contribution of a single storm characteristic. The flood-positive storms for those datasets can be sporadically found among the top 30 flood-positive storms (out of 70) in terms of average total rainfall and among the top 35 flood-positive storms (out of 70) in terms of average rainfall intensity.

The disparity in number of storms available for analysis per individual polygon is directly linked to the nature of precipitation events; whereas, the large spatial variability of rainfall renders the information area-specific. When no rainfall was recorded over a polygon while it did elsewhere, no storm statistics were kept for further analysis pertaining to that specific polygon with respect to that particular storm. Fortunately, the number of available storms in polygon 32's datasets was among the most numerous.

Table 4.4 - Model false positive results according to each polygon

	Calibration			Validation			
Rain Gauge	**Number of Storms**	**Number of Floodings**	**% False Positives**	**Number of Storms**	**Number of Floodings**	**% False Positives**	**Average % False Positives**
17	378	4	9.5%	96	2	13.5%	12%
20	397	3	6.3%	100	2	2.0%	4%
21	372	2	9.1%	94	2	0%	5%
26	410	1	1.2%	103	1	7.8%	4%
27	391	3	9.2%	99	2	0%	5%
28	369	3	2.2%	93	2	0%	1%
29 & 55	346	4	19.4%	87	2	6.9%	13%
30	378	5	14.0%	95	2	5.3%	10%
32	408	5	8.8%	103	2	15.5%	12%
33	373	3	4.3%	94	2	11.7%	8%
37	410	1	0.2%	103	1	4.9%	3%
38	409	2	14.7%	104	2	29.8%	22%
42	365	3	20.3%	92	2	16.3%	18%
48	343	2	10.8%	86	1	3.5%	7%
49	396	2	4.5%	99	1	2.0%	3%
57	388	1	2.6%	97	0	-	-

The false positive approach was preferred over conventional measures of model fit because it was of particular interest to assess the model fit for flood-positive storms while the conventional methods consider the entire range of data indiscriminately.

However, conventional means of measuring model fit is listed below for comparison. The model fit to polygon 32's calibration dataset had a:

- R^2 for multiple variables = 0.0728
- $\overline{R}^2$ = 0.06823 and,
- p = 2.25e-07.

The coefficient of determination, also known as R^2, is a measure of the distribution of values away from the line of best fit. Here it essentially assesses the representiveness of the distribution of values of each method produced in terms of their linear prediction of the number of reported floodings. While the range of

possible values for R^2 spans from 0 to 1, an R^2 value of 1 denotes a perfect fit against the desired variable. Note that the adjusted R^2, also written as $\overline{R}^2$, is a more accurate computation than R^2 as it takes into account the effect that multiple variables have on the R^2 computation (Helsel & Hirsch, 2002). Also, α and p are as defined in Appendix B.

Interpretation of the model fit statistics unsurprisingly presents an unfavorable assessment. The $\overline{R}^2$ obtained is far from the ideal value of 1 and similarly the p value is well below α = 0.05, which would lead to the conclusion that the two datasets (the stepwise regression model and polygon 32's calibration dataset) are not similar.

However, for additional comparison purposes, the binary (flood/no-flood) hit-rate produced by Equation 4.1 was computed for all polygon calibration and validation datasets. The hit-rate computation results are presented in Table 4.5. The hit-rate for the entire storm-polygon dataset was found to be equal to 95.3%. In all, the hit-rate assessment of the stepwise regression fitted model highlights its near ideal (close to 100%) identification of flood-producing storm-polygons from those that did not.

Table 4.5 - Model hit-rate results according to each polygon

	Calibration			Validation			
Rain Gauge	**Number of Storms**	**Number of Floodings**	**% Hit-Rate**	**Number of Storms**	**Number of Floodings**	**% Hit-Rate**	**Average % Hit-Rate**
17	378	4	94.2%	96	2	96.9%	96%
20	397	3	95.0%	100	2	96.0%	95%
21	372	2	94.9%	94	2	90%	93%
26	410	1	96.3%	103	1	96.1%	96%
27	391	3	96.9%	99	2	95%	96%
28	369	3	95.9%	93	2	96%	96%
29 & 55	346	4	95.1%	87	2	92.0%	94%
30	378	5	94.7%	95	2	93.7%	94%
32	408	5	96.1%	103	2	93.2%	95%
33	373	3	96.0%	94	2	92.6%	94%
37	410	1	96.6%	103	1	97.1%	97%
38	409	2	94.9%	104	2	95.2%	95%
42	365	3	96.2%	92	2	91.3%	94%
48	343	2	95.3%	86	1	93.0%	94%
49	396	2	95.5%	99	1	97.0%	96%
57	388	1	94.8%	97	0	94.8%	95%

4.3 Formulation of Preferred Storm Classification Method

Adjustments to Equation 4.1, as described below, were made as a compromise to attempt to fit the most number of storms with floodings into the smallest flood-positive range for storm flood index values. Ultimately, three flood-positive storms were rejected (representing the lower 4.3% of recorded storms with floodings) in terms of devising the flood-positive range of the index. This was done due to the increasingly high ratio of false positives that would have been considered. As a result, this effectively translates into the flood-positive range of the index value representing only the top 8.3% of the storm dataset with a large 88.9% false positive within that range.

Using Equation 4.1 and the entire storm dataset for each polygon, a transformation was added to the model equation by dividing the model by 4 in order to produce a index of values which would fit within a 1 to 10 severity index for flood-positive

values. The 1 to 10 severity index value range was chosen for ease of application. The model was then offset by 0.855 in order to capture the largest amount of positive recorded floodings while keeping the false positives to a minimum. These modifications to the model as well as the truncation of coefficients to 5 significant figures produced the following index (Equation 4.2):

$$S_i = 0.82851 - 2.2942 \times 10^{-11}\left(2D_a - 3P_a\right)^5 + 0.017393 I_a \qquad \textbf{(4.2)}$$

where S_i is the rainfall storm flood severity index and D_a, P_a as well as I_a are identical to the previously defined variables of Equation 4.1.

In order to produce the scaled values, it is important to round the index values to 1 decimal place. Once rounded, if the index value falls below 1.0, it is considered to be negative in terms of expected flooding.

The maximum storm index value produced by a storm within the dataset was of 9.0; however, it was a false positive. The highest storm index value a flood-positive storm produced was 3.5 with 80 flooded basements.

Expected floodings corresponding to index values within the flood-positive range were derived by fitting a power function to only the index values within the range that were associated with floodings (see Figure 4.16 and 4.17). Notice that the derived trend was extrapolated to an index value of 10.

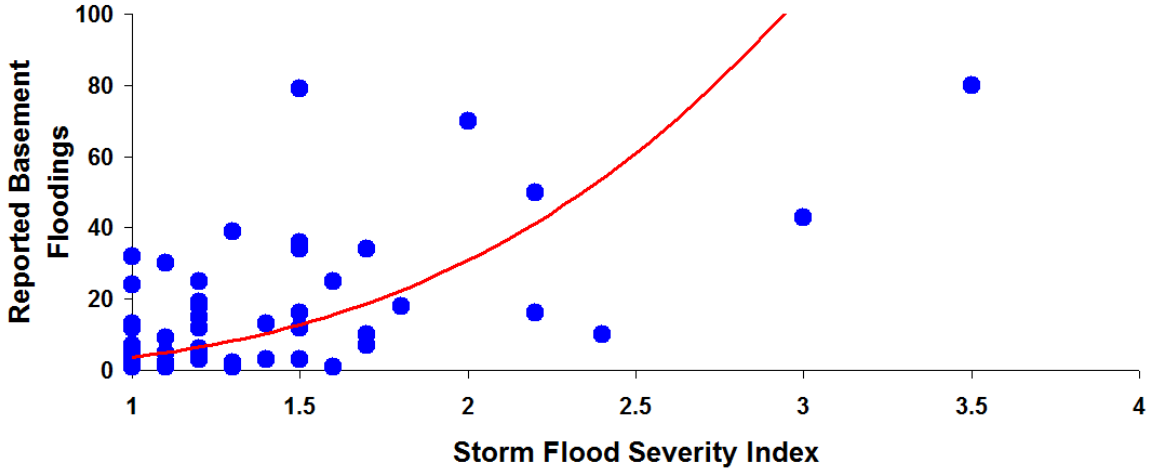


Figure 4.16 - Expected floodings associated with index values, 1 to 200 vertical axis

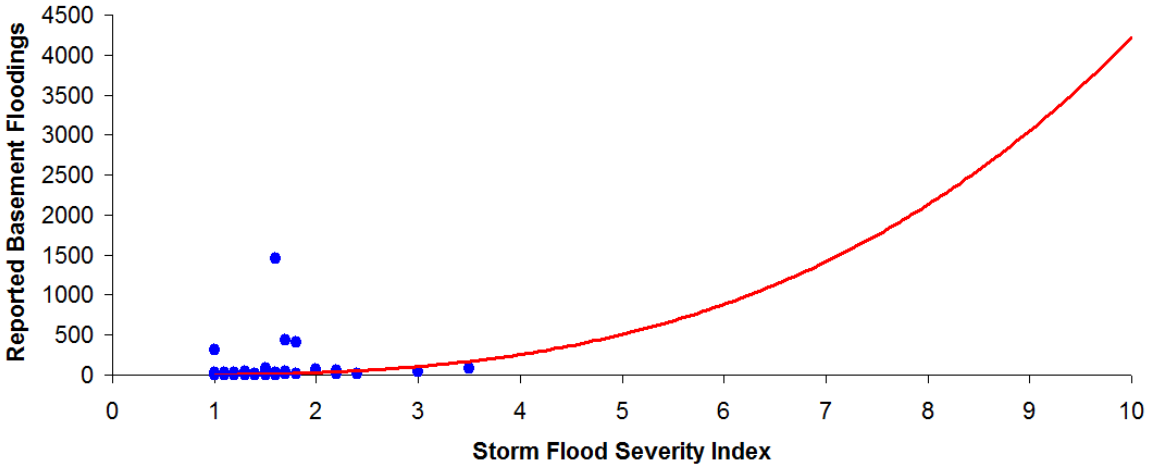


Figure 4.17 - Expected floodings associated with index values

A power function (see Equation 4.3) was found to reasonably fit the current dataset distribution and offers a reasonable extrapolation of the model in higher index values.

$$F_e = 3.7122 S_i^{3.0549} \quad \textbf{(4.3)}$$

where F_e is the expected number of basement floodings and S_i is identical to the previously defined variable of Equation 4.2.

The expected ranges of floodings corresponding to each index level were computed from Equation 4.3 and presented in Table 4.6.

Table 4.6 - Expected floodings associated with index values

Index number	Expected Floodings
1	4
2	31
3	106
4	256
5	507
6	885
7	1417
8	2130
9	3053
10	4212

5 Analysis and Discussion

The following sub-sections contain two separate analyses and a discussion on the issues and limitations pertaining to the study. The first analysis consisted of a comparison in the predictive performance of the proposed storm flood severity index versus the IDF method in assessing reported storm flood severity. The second, is a sensitivity analysis performed on the computed urban rainfall storm flood severity index.

5.1 Comparison of the Proposed Method Against the IDF Method

In this section, the predictive ability of both the proposed Storm Severity Index and the IDF method will be compared. To compare the performance of the proposed Storm Severity Index against that of the IDF method in predicting basement floodings, the closest up-to-date IDF station data was used. EC offers four stations with IDF data in or near the City of Edmonton (EC, 2012), as can be seen in Figure 5.1 taken from EC (1990). These stations are listed below with each station's years of operation in parentheses (EC, 2012):

- Edmonton Namao Airport (1965-1994)
- Edmonton Municipal Airport (1914-1993)
- Ellerslie (1965-1986)
- Edmonton Int'l A (1961-2006)

Figure 5.1 - EC IDF stations in or near the City of Edmonton (EC, 1990)

Notice that the Edmonton International Airport station is the only station with data beyond 1994. Therefore, it was chosen as most representative of the four stations for the City of Edmonton as its statistics include the most up-to-date observations of recent local climate trends. This station is geographically the farthest of the four, located approximately 10 km outside the study limits and approximately 28 km away from the centre of the study area. Also, as shown above, Edmonton International Airport station is the station with the second longest number of years of operation (45 years as opposed to Edmonton Municipal Airport with 79 years).

The mm/hr statistics of the Edmonton International Airport station (EC, 2012) were fitted to Equation 5.1 (Kije Sipi, 2007) via the Simplex method by using a program developed internally at Kije Sipi Ltd. The fitted equation allows for an expression of all IDF variables simultaneously. The IDF data was fitted to the following equation:

$$I = \frac{d \times T^{e}}{(g + D)^{h}} \qquad \textbf{(5.1)}$$

where I is for rainfall intensity (mm/hr); D is the duration of the event (hours); T is the RP of the event (years) and; d, e, g and h are the fitted coefficients. For the Edmonton International Airport station's 2012 data, the Simplex method yielded the following coefficients:

d = 14.9306
e = 0.222089
g = 0.0601719
h = 0.771013

When solving for the RP, Equation 5.1 takes the form of Equation 5.2.

$$T = \sqrt[e]{\frac{I \times (g + D)^h}{d}} \tag{5.2}$$

Two different storm-polygon characteristics combinations were used in computing each storm-polygon's IDF RP. As is standard practice for storm severity assessment, the most highest RP description was retained. The following table (Table 5.1) shows the three different storm-polygon characteristics combinations:

Table 5.1 - Storm-polygon characteristics combinations for IDF RP computation

Combination	Intensity (mm/hr)	Duration (hr)
1	Peak total rainfall/Average rainfall duration	Average rainfall duration
2	Peak intensity	0.25

The performance in terms of assessing rainfall storm severity in terms of the number of flooded basements for both the proposed Storm Severity Index and the IDF method were compared in two ways: 1) using only flood-positive storm-polygons and, 2) all storm-polygons.

5.1.1 Comparison of Both Methods for Flood-Positive Storm-Polygons

The first comparison comprises of assessing the predictive ability of the proposed Storm Severity Index against the IDF method of the number of reported basement floodings per storm-polygons for only the flood-positive events.

The predictive ability of each approach will be assess in terms of the R^2 as previously presented.

When considering only the flood-positive storm-polygons, the R^2 values for each severity assessment method are as follows: the proposed Storm Severity Index has an R^2 of 0.03307, whereas the IDF method as computed has an R^2 of 0.3902. The IDF method's R^2 when compared to that of the proposed Storm Severity Index's, though not great (not close to 1), indicates that the IDF method seems to better represent reported basement flooding severity when only considering flood-positive storms.

Figure 5.2 overlays the storm severity assessment for both the proposed Storm Severity Index method (in blue circles) as well as the IDF method (in red squares) for flood-positive storm-polygons. In order to overlay the values on the same plot, different vertical axes were used for each set: the IDF RP values are plotted according to the left-hand axis, while the proposed Storm Severity Index are plotted against the right-hand axis. Note that both the left-hand (IDF RP) and bottom axes (floodings) are logarithmic. Also visible in the figure are each set's best fit line. The mostly dotted line is the IDF RP's best fit line, whereas the full line is the proposed Storm Severity Index's best fit line. The figure shows that four storm-polygons have computed IDF values of over 1000 years. The figure illustrates what the R^2 analysis concluded: that the Storm Severity Index is best represented by its linear relation as compared to that of the IDF method.

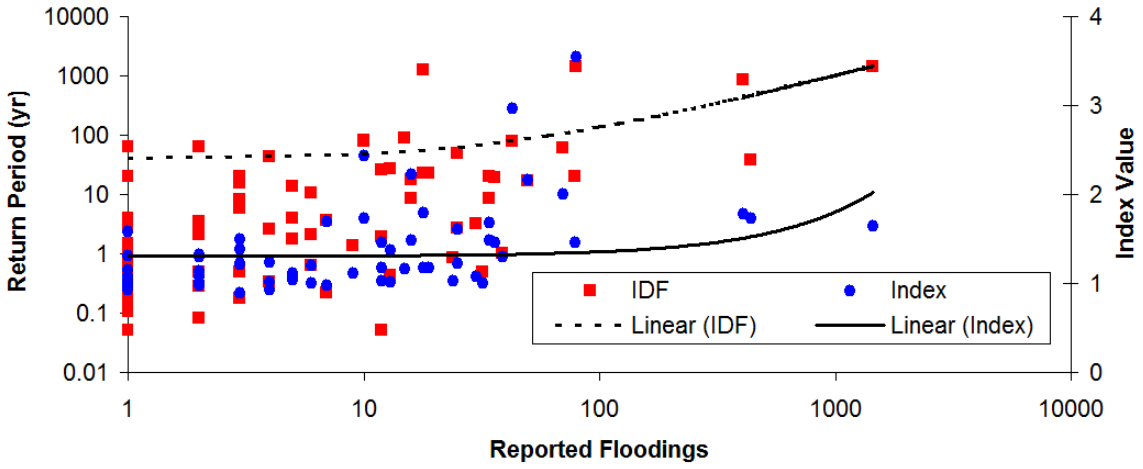


Figure 5.2 - Storm severity as assessed by both methods for flood-positive storm-polygons

5.1.2 Comparison of Both Methods for All Storm-Polygons

The second and most important comparison comprises of assessing the predictive ability of the proposed Storm Severity Index against the IDF method of the number of reported basement floodings per storm-polygons for all 7678 storm-polygon events.

As previously explained, the coefficient of determination, also known as R^2 will serve to assesse the representiveness of the distribution of values for each method produced in terms of their linear prediction of the number of reported floodings.

When considering all storm-polygons, the R^2 values for each severity assessment method are as follows: the proposed Storm Severity Index has an R^2 of 0.007976, whereas the IDF method as computed has an R^2 of 2.893×10^{-6}. While neither value is good (near the value of 1), the Storm Severity Index method has a much better R^2 than the IDF method. Therefore, it can be said that the Storm Severity Index does a much better job at describing the overall storm-polygon dataset than the IDF method. It does so in the sense that it best describes flooding severity for both flood-positive and flood-negative storms.

Like Figure 5.2, Figure 5.3 presents the storm severity assessment for both the proposed Storm Severity Index method (in blue circles) and the IDF method (in red squares) but for all storm-polygons. Again, different vertical axes were used for each set: the IDF RP values are plotted according to the left-hand axis, while the proposed Storm Severity Index are plotted against the right-hand axis. Also note that, like for Figure 5.2, both the left-hand (IDF RP) and bottom axes (floodings) are logarithmic. Visible in the figure are each set's best fit line. The mostly dotted line is the IDF RP's best fit line, whereas the full line is the proposed Storm Severity Index's best fit line. Figure 5.3 also illustrates, as Figure 5.2 did, what the R^2 analysis concluded with respect to all storm-polygons: that the Storm Severity Index is best represented by its linear relation as compared to that of the IDF method.

Note that in order to fit the entire storm-polygon dataset onto the plot, the storm-polygons with no floodings are giving a value of 0.1 to allow for logarithmic plotting. The Edmonton floodings dataset does not contain any partial flood values. Of interest is the fact that the greater majority of storm-polygons have no reported floodings (plotted on Figure 5.3 as having 0.1 floodings).

While the majority of storm-polygon events were reported to not have caused basement floodings, there is an alarming amount storm-polygons that were computed to have an IDF recurrence of over 1000-years.

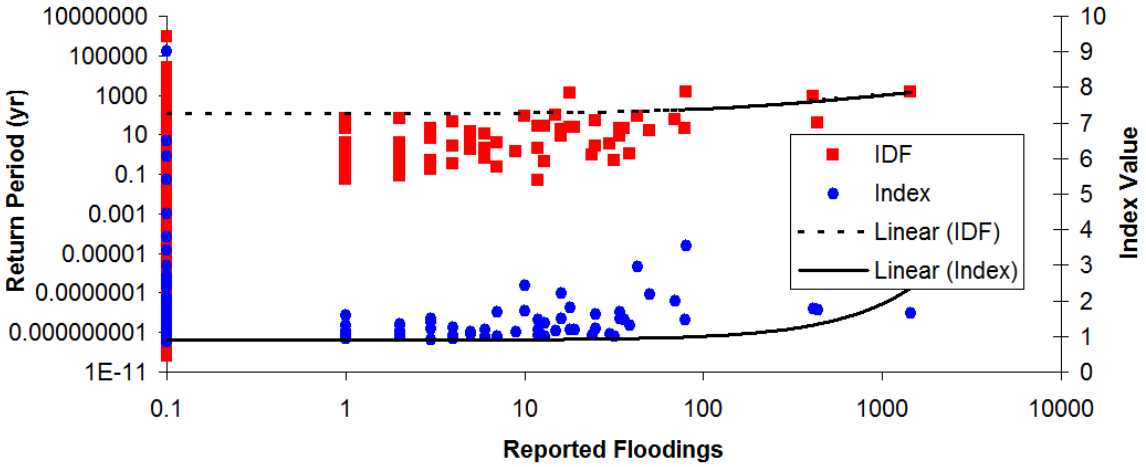


Figure 5.3 - Storm severity as assessed by both methods for all storm-polygons

5.2 Sensitivity Analysis of Proposed Method

A sensitivity analysis was performed on the model to illustrate the magnitude of its response with respect to each of its variables. The largest value in the overall storm dataset, rounded up to nearest 10 from dataset statistics (as shown in Figure 4.8), was used to determine the range of sensitivity analysis. The following values ranges were used to produce the sensitivity analysis:

- Average total rainfall**:** 0 to 130 (mm)
- Average rainfall intensity**:** 0 to 170 (mm/hr)
- Average rainfall duration**:** 0 to 100 (hr)

The color scale values used in Figures 5.4 to 5.8 were restricted to meaningful ranges corresponding to those of the storm flood severity index as elaborated in the previous sections. These figures illustrate the sensitivity of the storm flood severity index plotting the expected basement floodings against the average total rainfall versus the average rainfall intensity for fixed average rainfall durations from 0.1 to 100 hrs.

Prevailing patterns among the figures include:

- smaller average rainfall durations lead to more floodings; however,
- greater average total rainfall and average rainfall intensity yield more floodings.

Specifically, the five figures exhibit a gradual increase in expected basement floodings as the average rainfall intensity increases while presenting a sharp gradient in expected floodings as the average total rainfall increases. The lower maximum expected basement floodings that can be attained with low average total rainfall and high average rainfall intensity as compared to that of the opposite suggests that the weights of the average rainfall intensity is not the main contributor in determining the storm importance. In contrast, the sharp gradient in expected basement floodings along the average total rainfall figure axes and the significant impact that the average rainfall duration has on the distribution of the expected basement floodings demonstrate that these two storm characteristics are the main contributors in determining the storm importance.

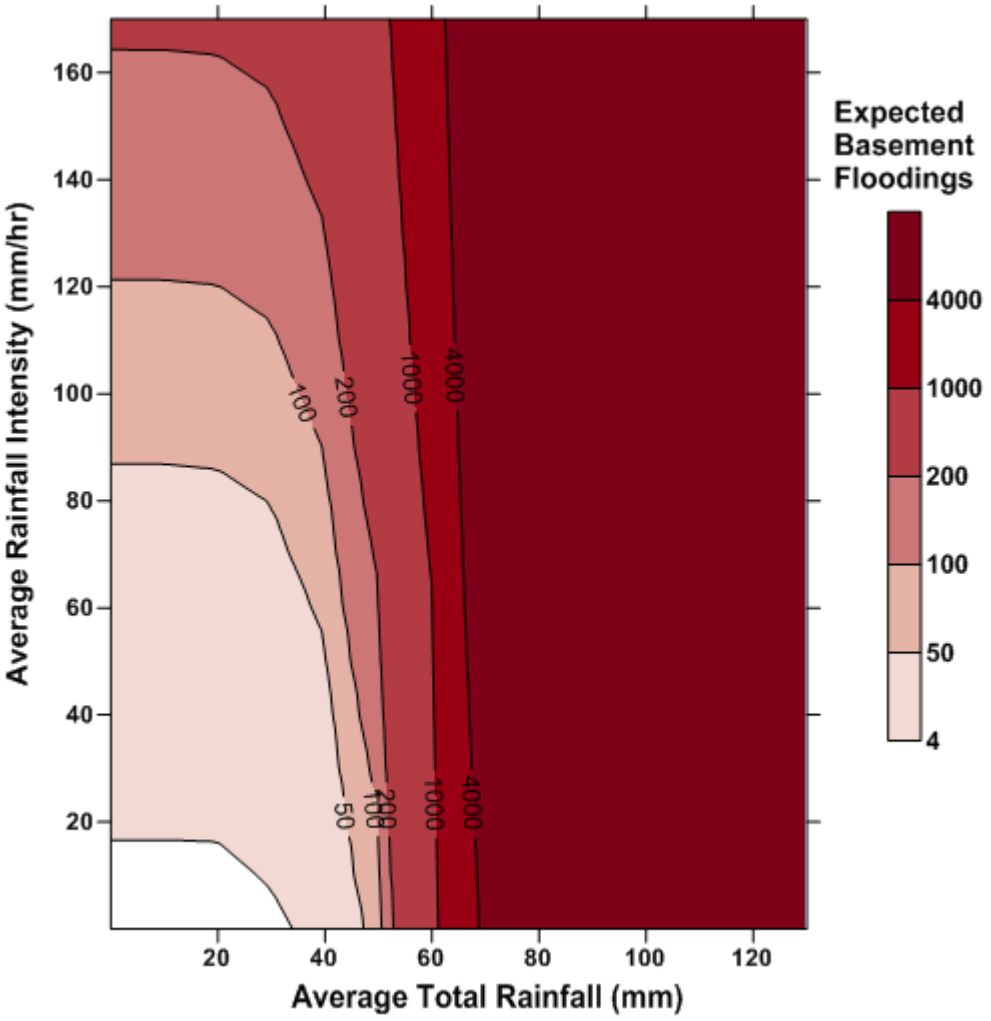


Figure 5.4 - Expected basement floodings from changes in average total rainfall versus average rainfall intensity when the average rainfall duration is equal to 0.1 hr

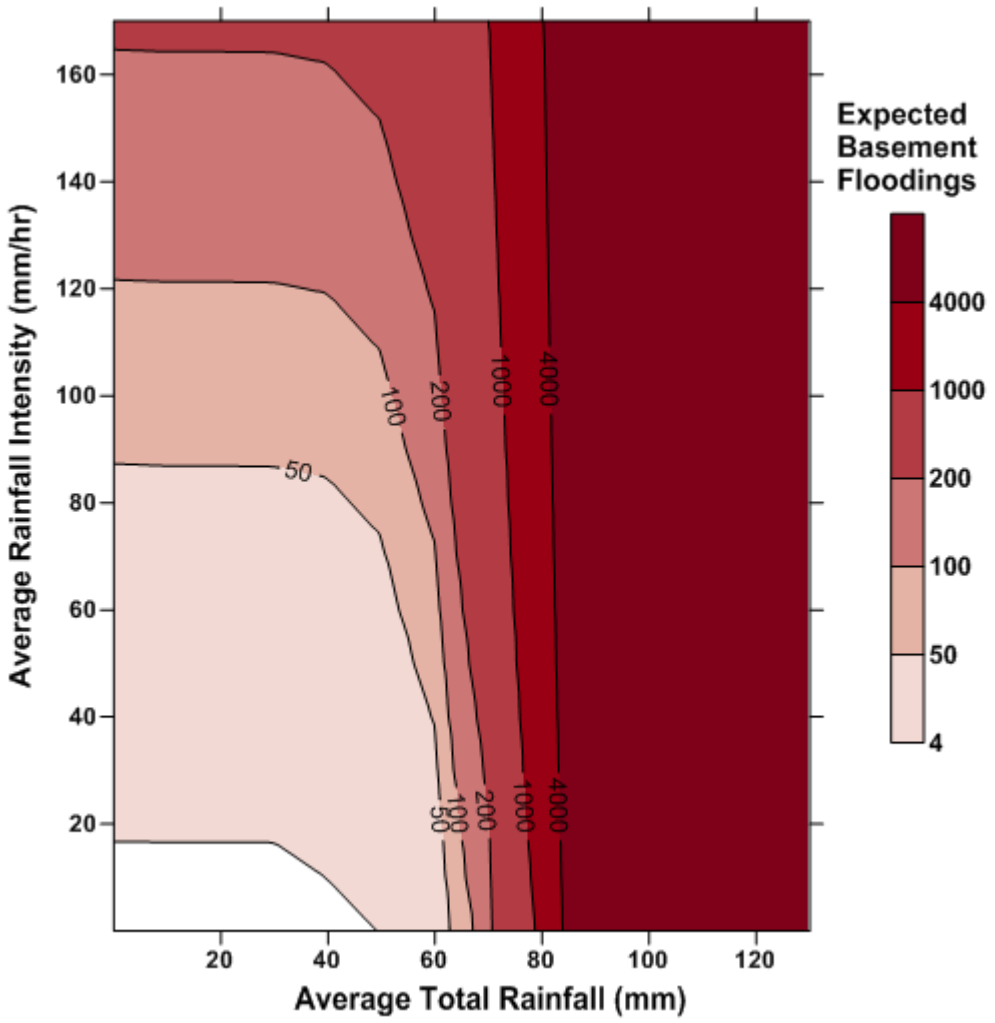


Figure 5.5 - Expected basement floodings from changes in average total rainfall versus average rainfall intensity when the average rainfall duration is equal to 25 hrs

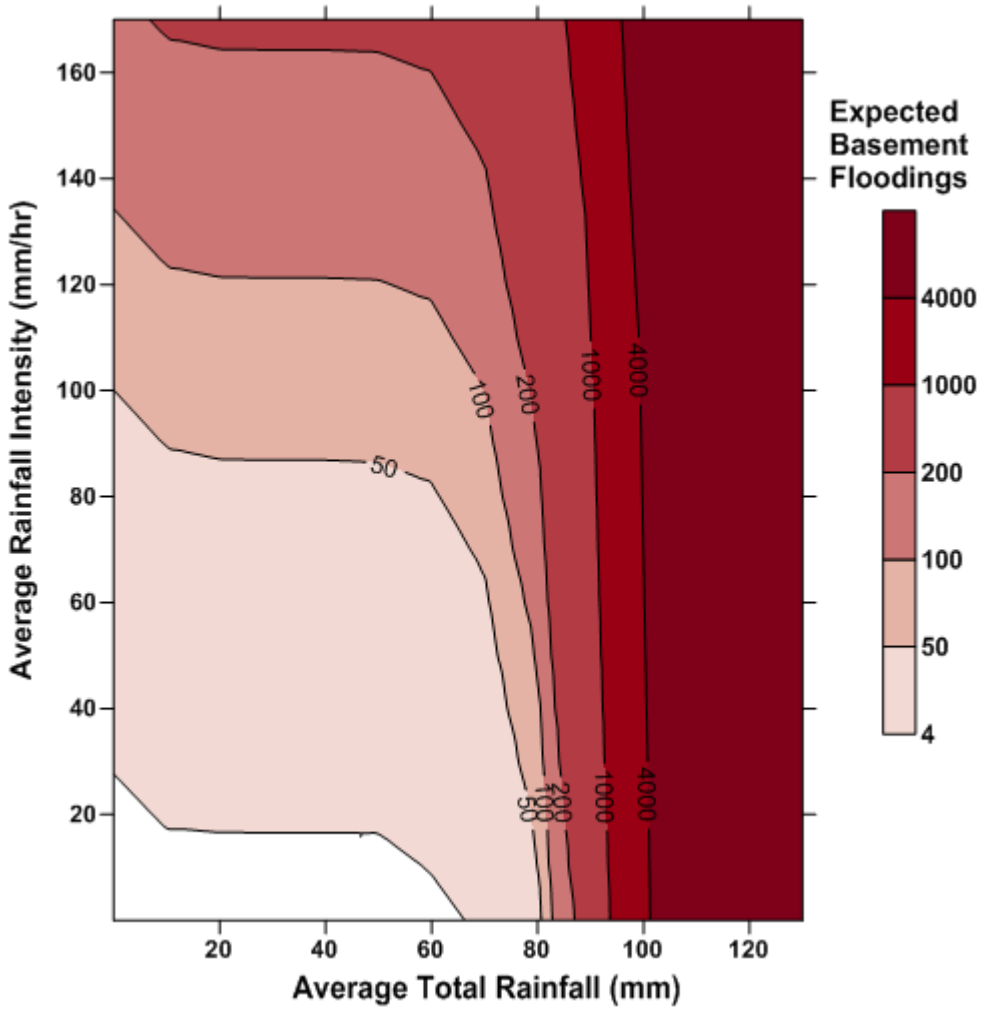


Figure 5.6 - Expected basement floodings from changes in average total rainfall versus average rainfall intensity when the average rainfall duration is equal to 50 hrs

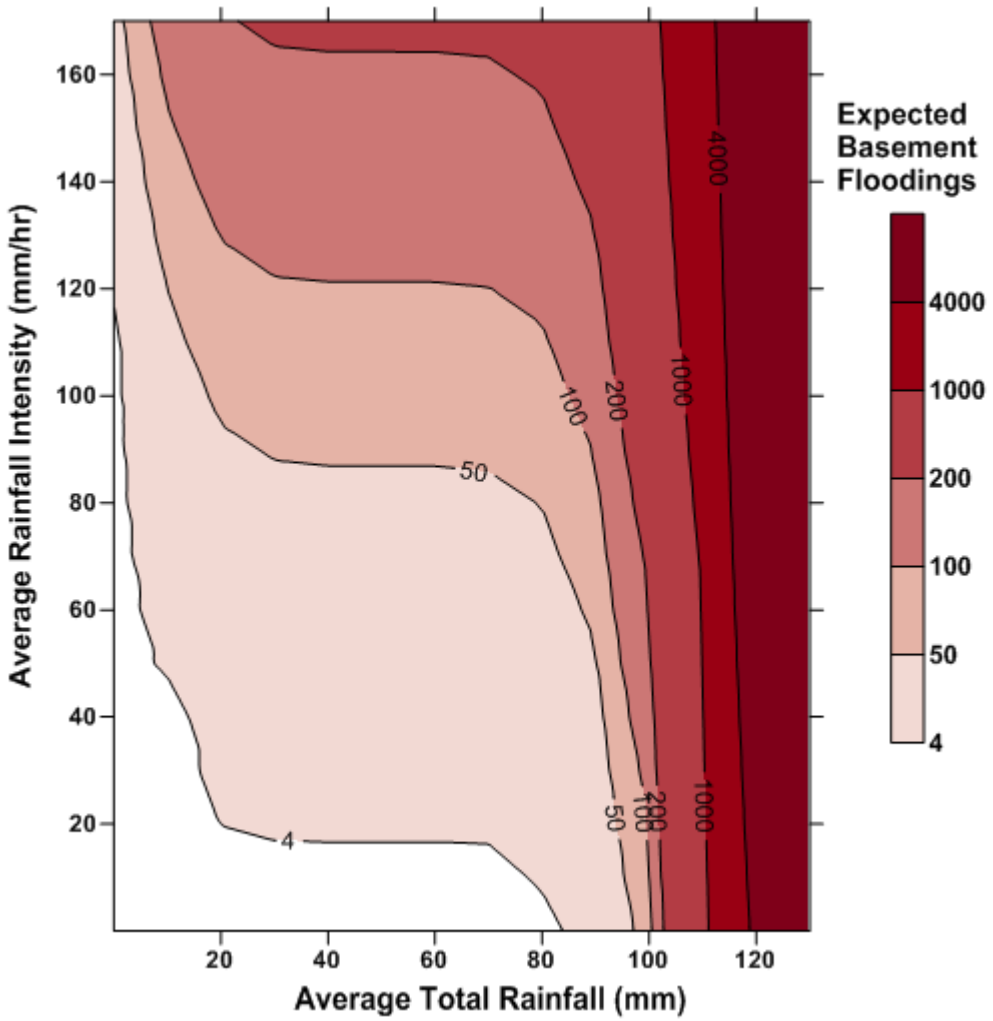


Figure 5.7 - Expected basement floodings from changes in average total rainfall versus average rainfall intensity when the average rainfall duration is equal to 75 hrs

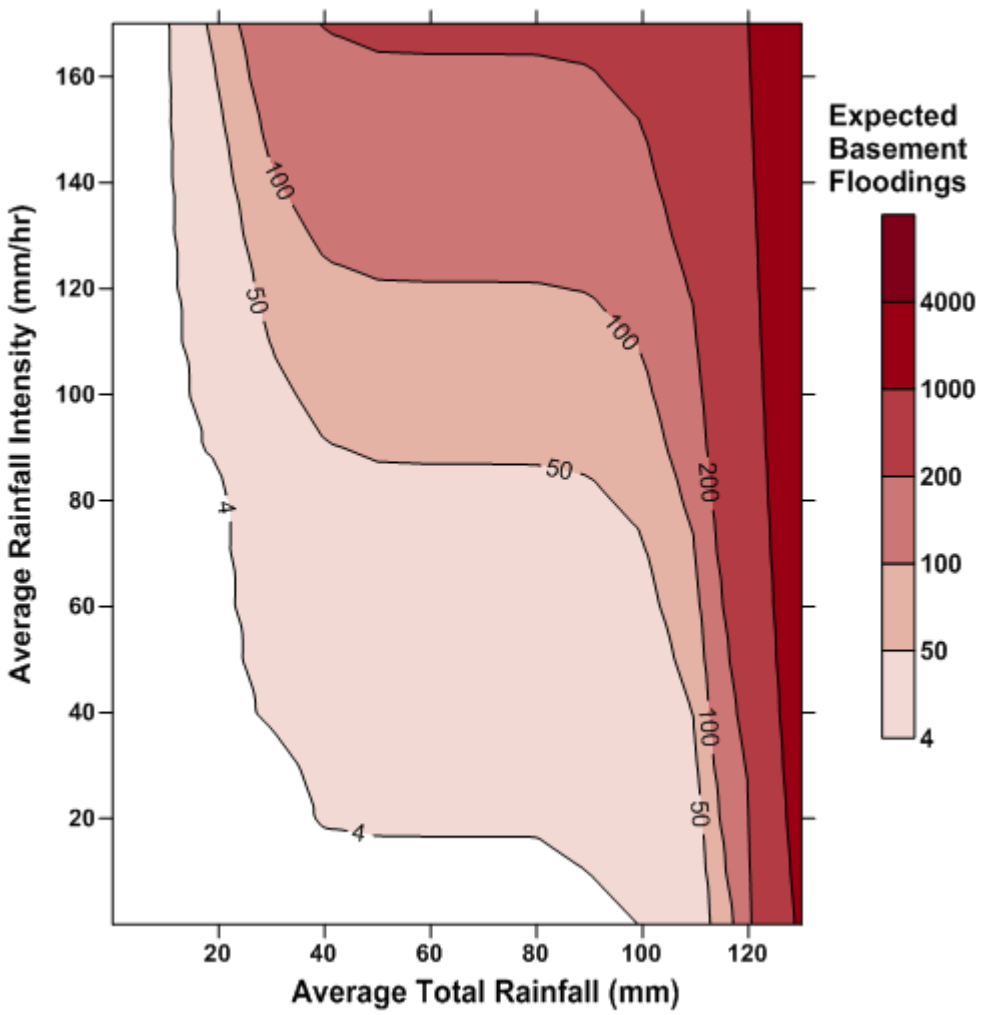


Figure 5.8 - Expected basement floodings from changes in average total rainfall versus average rainfall intensity when the average rainfall duration is equal to 100 hrs

Overall, the figures show an increasing trend of floods severity, which is in line with expected urban watershed hydrological responses - a shorter and more intense rainfall storm with more greater rainfall depths tend to produce flash floods (see Section 2.2.1.1).

5.3 Discussion

The following sub-sections discuss the limitations and potential issues associated with the work presented in this study.

5.3.1 **Limitations and Issues Regarding the Model**

This section elaborates on model limitations in terms of representativeness and particular issues, some of which were raised during analysis. In particular, the following sections discuss 2 limitations imposed by the original corporate storm characteristics database.

5.3.1.1 Hydrologic Characteristics

The model, as derived with the available database storm characteristics, could not take into account the temporal distribution of the rainfall. The importance of examining the temporal distribution of rainfall within events could help refine the analysis in determining periods of particularly intense or voluminous rainfall; hence, improve the description of severity. It also highlights concerns regarding the RERSD's working definition of a storm (see Section 3.3). Although storm characteristics are derived at the scale approaching that of the weather radar's full reach, it only examines a fraction of that area, the City of Edmonton. This issue expresses itself in the form of tracking storms, and continuing to acquire storm characteristics, before and after it has actually crossed City limits. Consequently, storm characteristics in the database could be hydrologically irrelevant to the study. This issue could also impact storm characteristics by grouping several hydrologically distinct rainfall storm events into one. Further storm dataset processing may be in order to derive only locally relevant storm characteristics.

Also, since no rainfall time series are directly available for storm cells within storms, the antecedent hydrologic conditions, that normally could be estimated via antecedent rainfall, would likely be unreliable. Antecedent rainfall is typically used to provide an indication of soil moisture, which influences rainfall-runoff processes. For example, dry soils would typically absorb the early rainfall, an effective process that delays the peak runoff discharge. This is particularly important when considering storm events that tend to produce flash floods.

Additionally, the model development did not consider the particular hydrologic/hydraulic characteristics of each test areas' (polygon), including:

- soil type;
- ground cover;
- sub-bassin shape;
- spatial distribution of the rainfall over the area;
- sewer network, and;
- building types, distributions and vulnerability to basement floodings.

This omission can potentially impact the degree of vulnerability; hence, susceptibility to flooding for buildings in the area of consideration, especially if significant hydrological/hydraulic variations exist in the study areas.

A perhaps more meaningful (hydrologically) approach in delineating areas of interest for analysis would be to use watershed delineations. This would likely lead to better results, as compared to using rain gauge-centered Thiessen polygons scattered across the City.

Regarding grouping basement floodings by polygons, knowledge of the exact location of each basement flooding, even if available, would likely not improve the analysis since the precision would still be limited by the 1 km^2 radar data grids.

5.3.1.2 Impact of the Number of Available Radar Pixels on Storm Characteristics

Another issue relates to the number of available radar pixel for storm characteristic computation and how this impacts the representativeness of the rainfall over the polygon areas.

Cases with a smaller number of available radar pixels tend to skew the generated storm characteristics distribution, as can be seen in its most extreme situation when only one radar pixel is used. A comparison of the storm characteristics distributions of polygons 21 and 29&55 that have only one available radar pixel each, as seen in

Figures 5.9 and 5.10, to that of the entire polygon-storm dataset, as seen in Figure 4.8, has revealed that lower numbers of available radar pixels lead to:

- lower peak values' (peak total rainfall and peak rainfall intensities) distributions as they approach the average values (average total rainfall and average rainfall intensities) and become the same when only one radar pixel is used, and;
- generally, to slightly higher average values (average total rainfall, average rainfall intensities and average rainfall duration) distribution, when compared to those of the overall storm-polygon dataset (except for the maximum values).

The rationale behind these observations is that a smaller number of available pixels lead to a greater weight on each pixel's value when computing the average. Furthermore, with extreme values being less likely, a smaller number of radar pixels provide a reduced sample size of recorded values. This impacts the spatially averaged values as extremes and usually averages down values; hence, reduces the probability of finding extreme values. This also has the effect of reducing the likelihood that each storm's most extreme rainfall value, which occurred over the study area, will be described and considered for analysis.

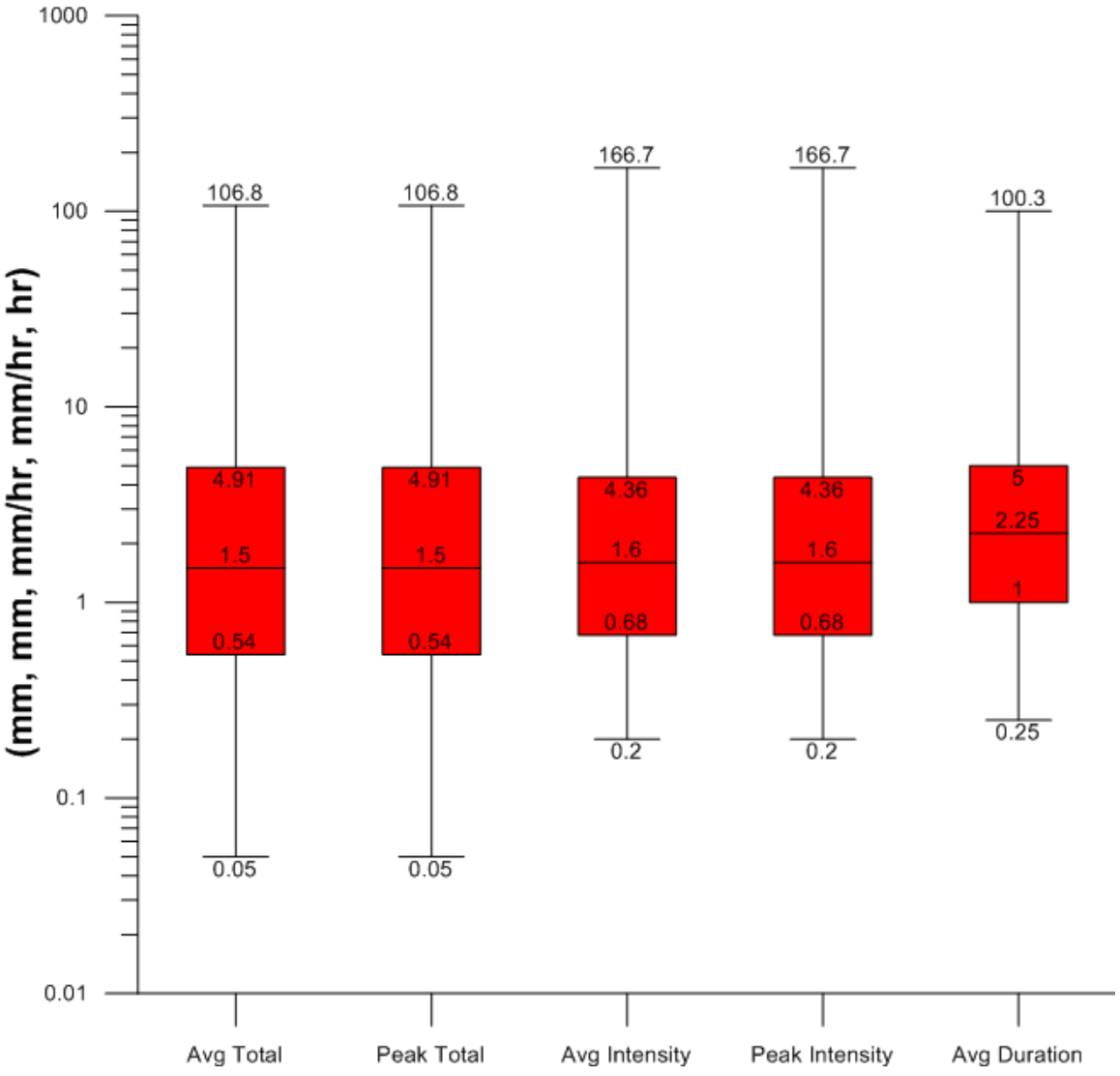


Figure 5.9 - Box-plot of polygon 21 dataset storm characteristics

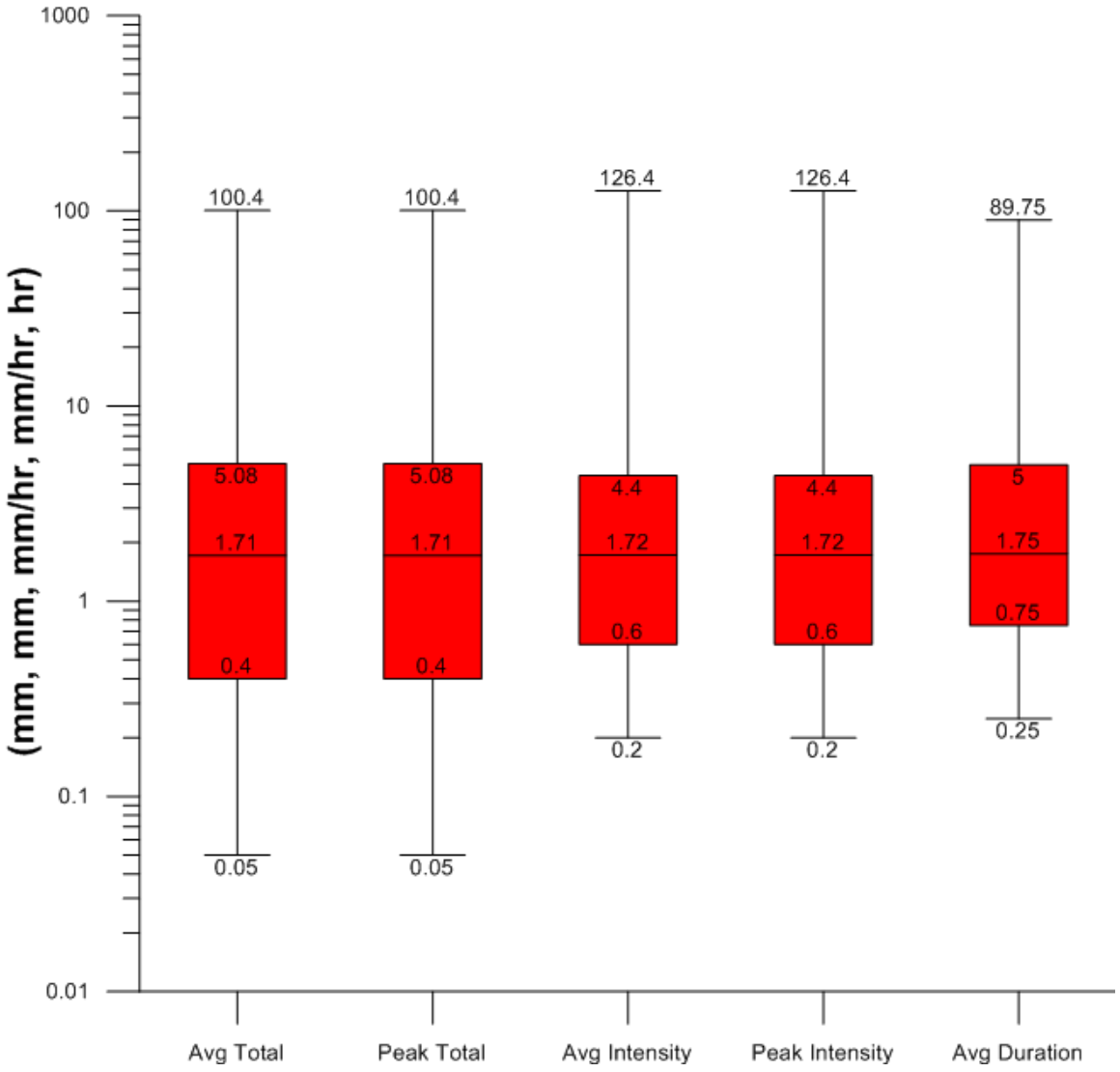


Figure 5.10 - Box-plot of polygon 29 & 55 dataset storm characteristics

In contrast to the previously mentioned effects of a limited number of radar pixels on the storm characteristics value distributions, consider the storm characteristics value distributions of polygon 26 and 49. They have the two most numerous number of available pixels in the study, 65 and 58 respectively. When comparing the storm characteristics value distributions of polygon 26 and 49, as seen in Figures 5.11 and 5.12 respectively, to that of the entire storm-polygon dataset, as seen in Figure 4.8, an opposite effect as presented earlier can be observed. This effectively indicates that a larger amount of available radar pixels lead to:

- higher peak values' (peak total rainfall and peak rainfall intensities) distributions as they increasingly diverge from that of the average values (average total rainfall and average rainfall intensities), and;
- generally, slightly lower the Average values (average total rainfall, average rainfall intensities and average rainfall duration) distribution when compared to those of the overall storm-polygon dataset.

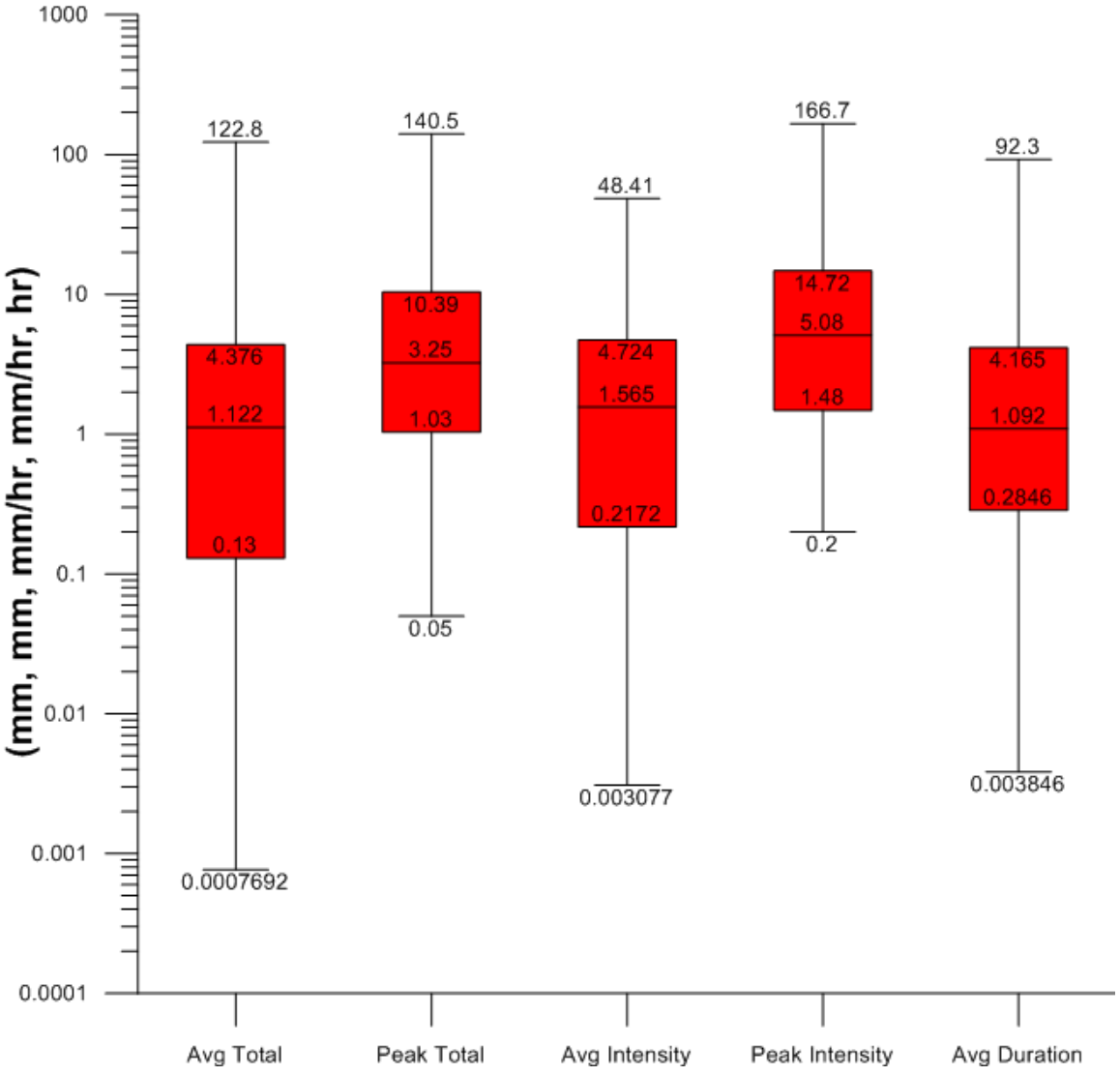


Figure 5.11 - Box-plot of polygon 26 dataset storm characteristics

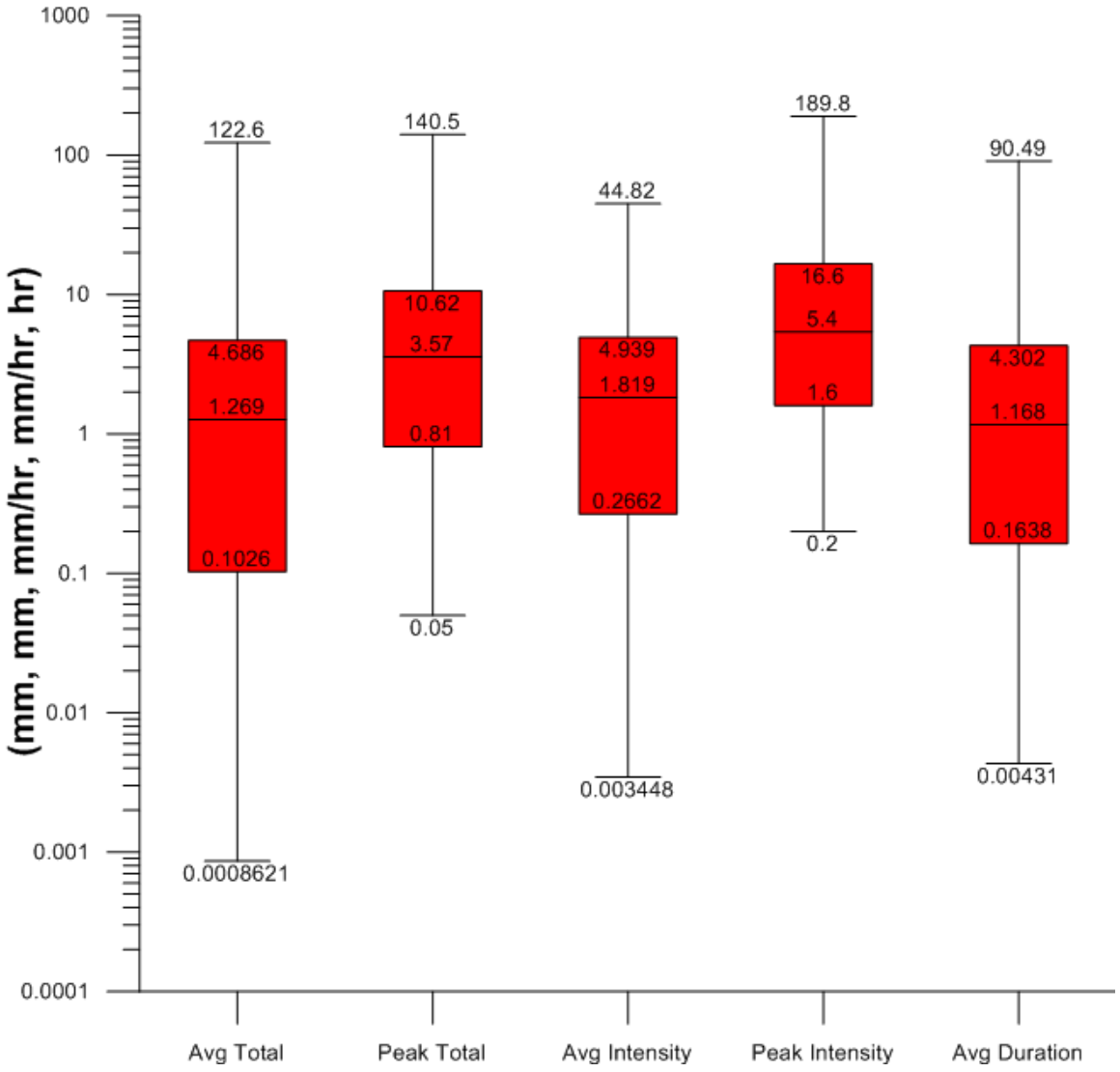


Figure 5.12 - Box-plot of polygon 49 dataset storm characteristics

Both extremes in the number of available radar pixel for storm characteristic computation, as elaborated above, impact their distribution as shown in the various figures cited. A concern is that this biasing of storm characteristics distribution for the overall storm-polygon dataset has potentially impacted the study results. While the model was derived from a dataset produced from 30 available radar pixels and the average number of radar pixels used per polygon considered in the final storm flood severity index development is 20, it is possible that values derived from polygons with the lowest number of pixels skewed the overall storm-polygon dataset from what it truly should be. However, it is fortunate that the model faired well for

polygons with multiple pixels used for storm characteristics computation, as seen in Table 4.4.

5.3.1.3 Masking

A second look at Table 4.1 reveals that overall, a staggering 67.8% of radar pixels assigned to polygons over the study area are masked. This is a fairly high amount of data which is not being considered in the analysis. This brings into question the representativeness of the study findings over the entire area which it is meant to represent.

The distribution of the number of floodings as found in Table 3.1 are illustrated in Figure 5.13. Note that despite the removal of 45 flood-positive polygon-storms from the final dataset, it appears that the selection of which has been random since little impact in the distribution of number of floodings can be observed when comparing Figure 5.13 to Figure 4.9.

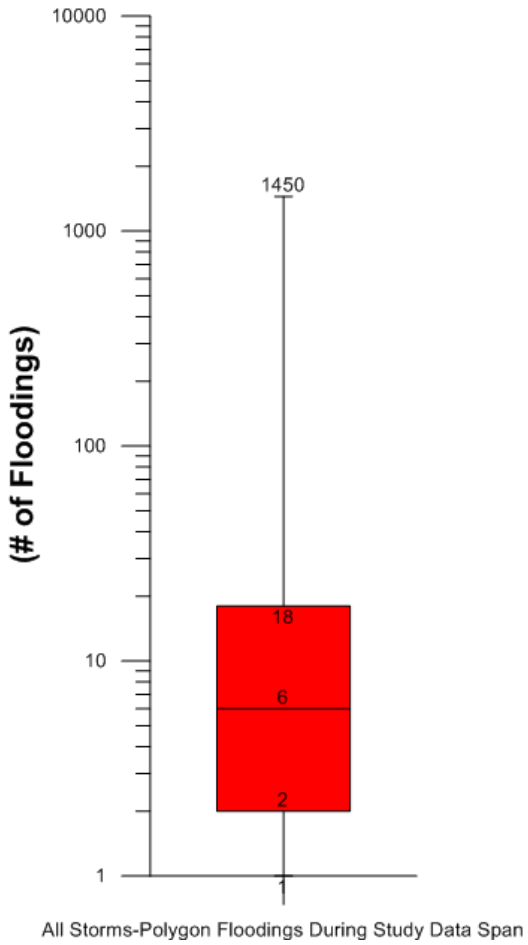


Figure 5.13 - All storm-polygon floodings during study data span

5.3.1.4 Assigning Flood Fault

Decisions regarding the processing of storm-flood fault assignments when multiple storms in the database were found to have produced significant rainfall over the City on dates associated with floodings could possibly have had impacts on the analysis results, especially if the flood responsibility was wrongfully attributed. The impact of such decisions is currently unknown though assumed to be negligible in size. Storm flood fault assignment was attributed to the worst storms, among those considered, as determined by having been subjectively selected in terms of the areal distribution and severity of the total rainfall over the Thiessen polygons representing flooded basements. Furthermore, this again brings up the impact of antecedent soil moisture conditions, which in turn could have been impacted by prior storms over

the area and consequently impacted the rainfall runoff response of the storm that were analyzed. This was not taken into account in this study.

5.3.1.5 Flash Floods

Figures 4.2 through 4.6 presented in Section 4.3.1.1 show that the average rainfall duration storm characteristics produced a curious clustering of reported floodings that does not tend to increase with an increase in average rainfall duration.

Figure 5.14 presents the same plot as previously seen in Figure 4.7 with two red circles superimposed to coarsely group the values into 2 groups. These two groups actually represent two distinct types of rainfall storms: local convective storm cells and regional scale stratiform storms. Storm and flood types were presented in Sections 2.2.1 and 2.2.1.1.

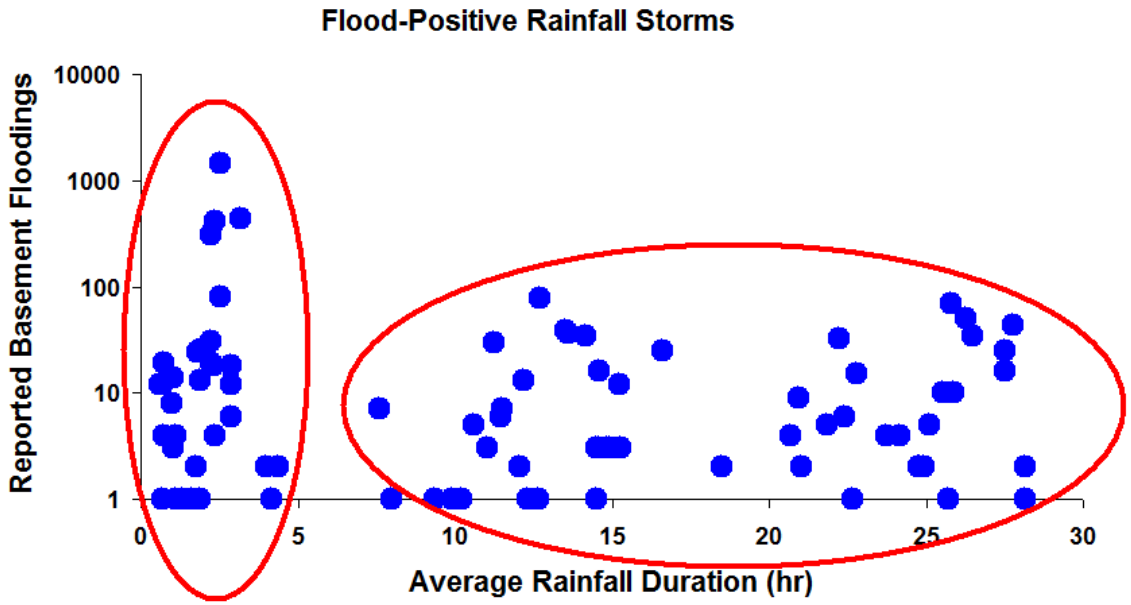


Figure 5.14 - Storm types on floodings

The figure suggests that the type of storms has a different impact on the number of the basement floodings, the most severe of which seem to have been produced by convective storm cells which likely resulted in flash floods. Figure 5.15 illustrates the storm flood severity index's estimated number of floodings. Comparison of Figure

5.14 to Figure 5.15 reveals that the storm flood severity index is quite successful at emulating the expected number of basement floodings, albeit with lower estimated values. Further comparison between the two figures reveal the presence of the same three data point clusters, that potentially represent the influence of different storm types which include flash floods. Notable in Figure 5.15 is the absence of all flood-positive storms below the 5 mm threshold, except one, which suggests that the storm flood severity index may not yield a good fit for the severity of flash floods. This is to be expected as the model which produced the index was derived from a dataset where storms' associated with this type of flooding were zeroed.

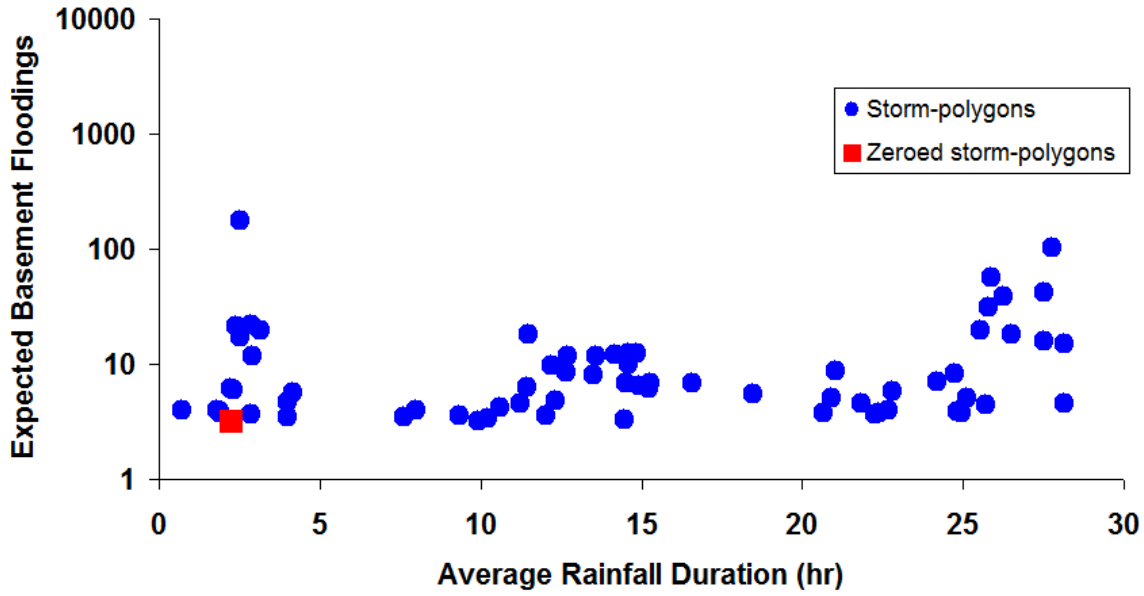


Figure 5.15 - Expected floodings as per storm Scale

Further concerns arise from the use of the 5 mm peak total rainfall threshold imposed as previously flood-positive storms which were zeroed. Section 4.3.1.1 discusses this issue that relates to storms that did not exceed the urban hydrology interception threshold. The 5 mm threshold may be too high a value for certain areas, such as those with combined sewer systems, and specific weather conditions, such as smaller, yet still extreme sub-radar pixel rainfall events. However, it is possible that the 5 mm peak total rainfall threshold is adequate and that the discrepancies noted are related to the amount of radar pixel masking and/or

severe sub-radar pixel rainfall storm cells. It is also possible that floodings occurring during such events are caused by factors other than direct rainfall runoff, including sewer backups.

5.3.1.6 Use of the Entire Storm-Polygon Dataset for the Production of the Final Storm Flood Severity Index

Caution is in order when using the model produced from polygon 32's dataset as compared to all storms from all polygons. This is due to the fact that the hydrologic properties of each represented polygon storm statistics are not the same and could vary drastically. Furthermore, as shown in Section 5.3.1.2, the amount of radar pixels available for storm characteristic computation effectively skews their values in a manner that will differ from their true values.

5.3.1.7 Discussion Regarding the Comparison in Storm Severity Assessment Between the IDF Method and the Storm Severity Index Method

There are two main areas that merit discussion due to limitations when the storm severity assessment of each method was compared. The first is with respect to the IDF RP computation while the second is with respect to the approach used to assess the level of fit.

In computing each storm-polygon's IDF RP, it is useful to remember that as mentioned in Section 3.3.1, the RERSD was used to produce the ERSPD. Therefore, data limitations present in the former is also present in the latter. Of particular interest, is that the RERSD used radar rainfall data that had a coarser temporal resolution than the original 5-minute estimates (15-minute data intervals). This undoubtedly effects of peak rainfall intensities by smoothing these values as extremes are averaged over a longer temporal period. Consequently, the storm-polygon's IDF RP estimates will be underestimated for events with particularly high 5-minute rainfall intensities.

Additionally, further IDF RP computation limitation arise from storm characteristics information originating from the original storm database (RERSD) in the form of fixed descriptions of storms, such as: peak total rainfall, average rainfall duration and

peak rainfall intensity. Though these storm characteristics can happen to describe the most severe aspect of certain rainfall storms, that is not always the case. It is not uncommon for the most severe rainfall depth over time description of a rainfall event to lay in between the total storm depth/duration and the shortest temporal intensity measurement. Again, this implies that the computed IDF RP for each storm-polygon may potentially be underestimates of their true IDF RP, as would typically be computed as part of forensic analysis.

Regarding the coefficient of determination, also known as R^2, note that it is a measure of the distribution of values away from the straight line of best fit. Assessment via the coefficient of determination is most informative when the data distribution tend is linear. However, this approach is not as useful at assessing data fit when there are other data distribution trends (e.g. logarithmic).

5.3.1.8 Notes Stemming From the Sensitivity Analysis

Plotting the recorded flood-positive storm values on one of the sensitivity analysis figures, as seen in Figures 5.16 and 5.17, illustrates that recorded storm characteristic values only represent a small region of the plausible value space. Dots in red with the number of associated basement floodings represent flood-positive storms while orange dots represent the rest of the entire storm-polygon dataset values, for all average rainfall durations.

Note that recorded storm value distribution (red and orange) was visibly absent from certain quadrants of plausible storm characteristic values, such that:

- high average rainfall intensity and high average rainfall duration;
- low average total rainfall and high average rainfall duration;
- high average total rainfall and low average rainfall duration, as well as;
- high average rainfall intensity with either high or low average total rainfall.

A plausible explanation for the lack of observed storm characteristics in quadrants with such value ranges could be that they are highly unlikely due to their high or low extreme nature.

Also, while Figure 5.17 did reveal three recorded storms with characteristics in the high average rainfall intensity quadrant, they only had mid-range values. A further interesting observation is the distribution of the flood-positive storms, shown in red. Their distribution is strictly limited to the lower quadrants with values in all two figures.

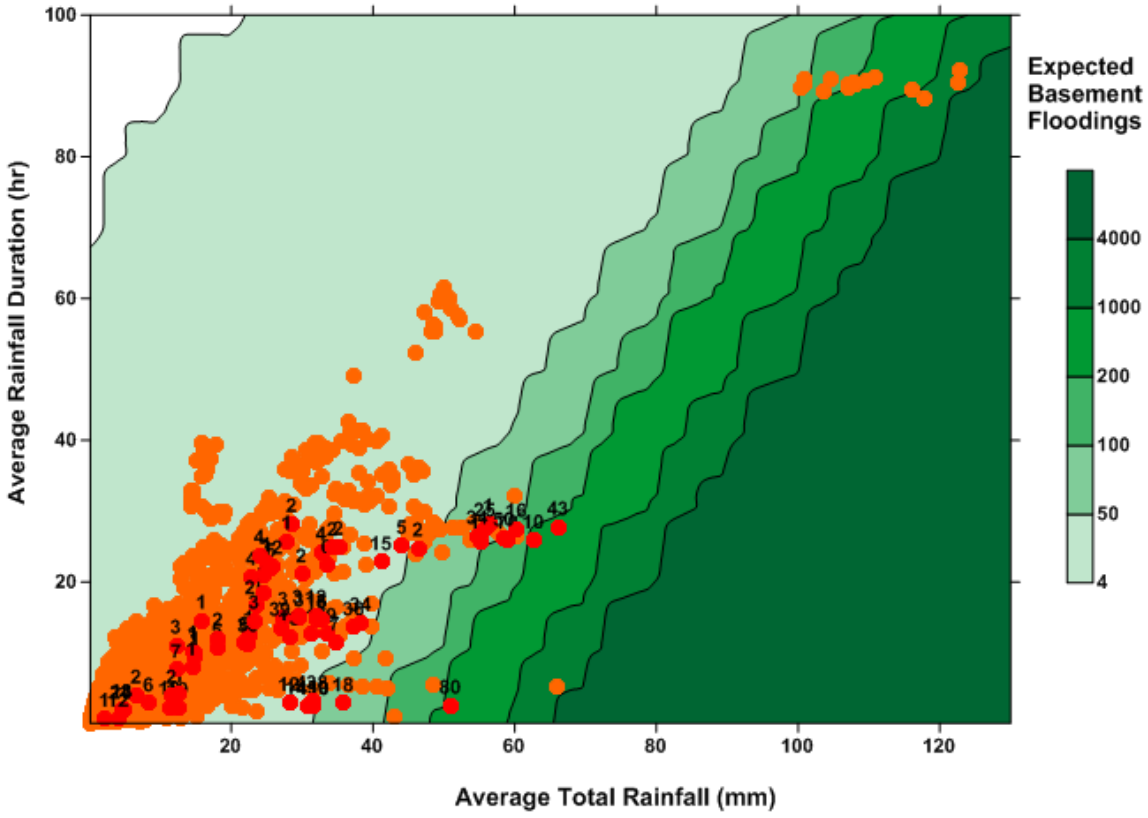


Figure 5.16 - Entire dataset storms with respect to expected basement floodings (average total rainfall vs. average rainfall duration)

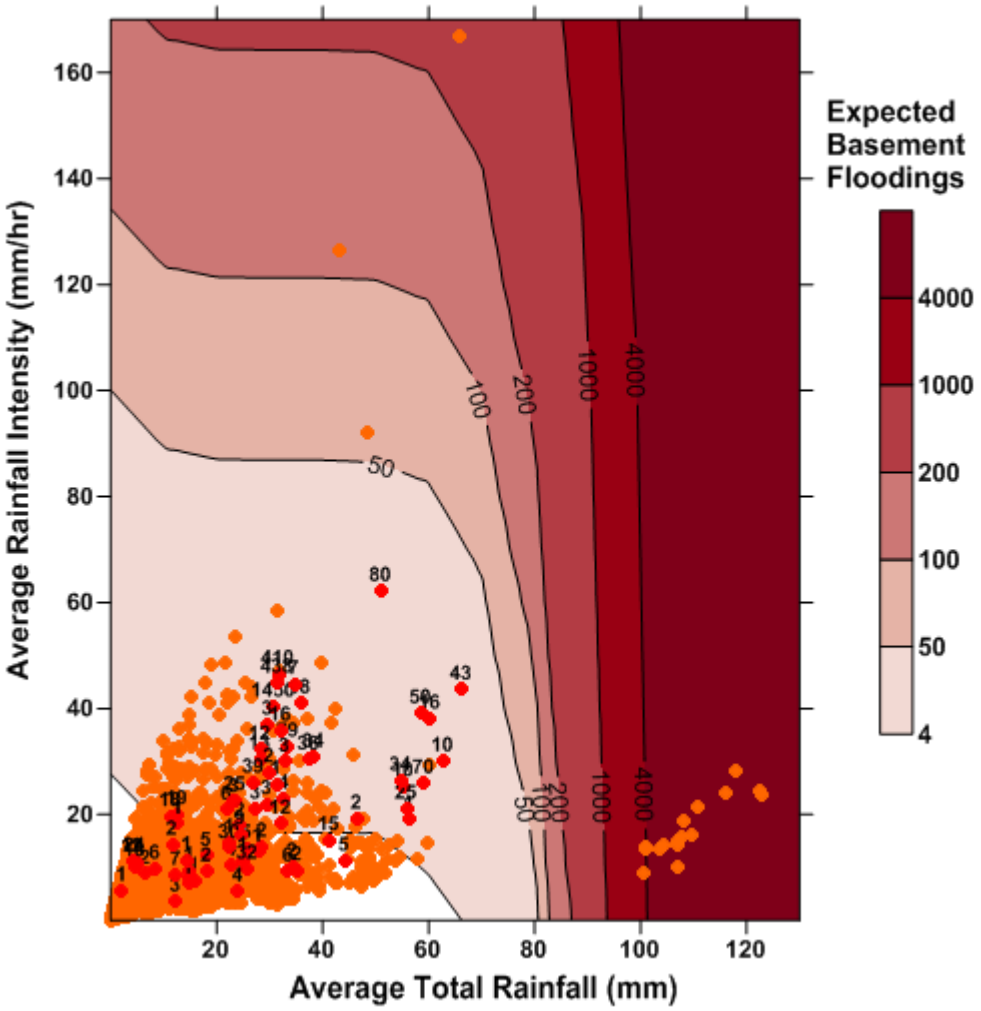


Figure 5.17 - Entire dataset storms with respect to expected basement floodings (average total rainfall vs. average rainfall intensity)

A closer inspection of Figure 5.16 helps discern a pattern that is also present in the other two figures. Expected floodings, derived the storm flood severity index, did not show a progressive severity in the expected number of floodings within the range of values encompassing flood-positive storms when considering average total rainfall and/or average rainfall duration as illustrated in Figure 5.16. This suggests the possibility that these two storm characteristics' are less important in the model that was produced in the study. In other words, the average total rainfall and average rainfall duration, may have a negligible role with respect to determining storm

severity. However, this is contrary to results found when seeking the best fitting model the polygon 32 storm dataset.

Furthermore, it is not yet clear whether the storm index is actually representative of expected floodings in specific ranges of values considering certain ranges of possible storm characteristics combinations have not yet been observed as part of the available storm dataset. Index values in such ranges remain extrapolations, and as such caution should be used when considering index outputs.

The sporadic distribution of red points amidst the orange clusters in the previous three figures help illustrate the challenge it's been to best differentiate the flood-positive events from the flood-negative storms whilst modeling their severity. This reveals that perhaps other factors may be responsible in determining a rainfall storm's severity with respect to producing urban floodings.

Also, the abrupt increase in expected floodings distribution along the average total rainfall axes as shown in Figures 5.4 through 5.8, and 5.17 hints at the possibly of a threshold value. Some hazard protection measures, such as dams, can simultaneously lessen risk from hazards within design parameters while exacerbating the magnitude of losses. This typically occurs when their design levels are exceeded when land development was allowed in a previously hazardous areas that are now dependent on that protection (Mileti, 1999; CANHP, 2010). Numerous examples can be found in Canada and the United States where flood prevention infrastructure has allowed urban sprawl to extend in areas previously considered at risk of floodings (CANHP, 2010).

5.3.1.9 The City's Basement Floodings Dataset

Limitations with regards to the use of the City's flooding data include both the binary nature of the data (i.e. flood versus no flood) and the reliability of the information. Flood data on private property is considered to have a low degree of certainty due to self-reporting and possible censorship. Flooding occurrence reporting to the City and to insurance companies are typically lower than actual for a number of reasons

including concerns about impacts on property values and increase to insurance premiums.

5.3.1.10 Damage-Based Severity Index

A major underlying assumption in the study is that the number of flooded basements is directly related to the severity of storm, which may not be true. Also, damage importance classification methods are not suitable for long-term use because building standards are dynamic through time as technology and our understanding of the phenomenon in question increases coupled with infrastructure upgrades. This has been the case with regards to the historic development of earthquake severity scales (Desrochers, 2007).

5.3.2 General Limitations and Concerns

This section briefly elaborates on this study's limitations and concerns with it's representativeness with respect to more general points.

The chosen approach of using nested static coefficients in the stepwise regression analysis proved quite computationally intensive. Also, during computation of the stepwise regression, notable instability in the R compiling environment was encountered and caused frequent errors and session halts. This resulted in staggered progress of the stepwise regression analysis. Originally, several types of functions were to be analyzed, namely: power functions from 2 to 10, square root, inverse square root, exponential, inverse exponential, logarithm and inverse logarithm. However, in light of the setbacks due to the programming environment instability and due to time constraints the first set (power 5) were analyzed. The stepwise regression analysis was halted when the false-positive storm percentage reached a minima in the 3-term function. Admittedly, the chosen approach proved time consuming.

It is important to stress that while storm dataset was composed of a particularly long series for radar rainfall data (11 years), it represents a relatively short data series in terms of historical rainfall observation. Eleven years of observational data is likely

not be enough to fully describe the extreme rainfall trends due to certain climate patterns having much longer cycles, as discussed in Section 2.4.2.3.2.

6 Conclusions

This study provides a critical account of the current standard of practice for assessing rainfall storm severity as well as presents an entirely new approach to characterize the magnitude of these events. The proposed rainfall storm severity index considers both the spatial context of rainfall storms and indirectly incorporates runoff potential to produce a more representative approach to assessing urban rainfall storm severity. The result is a new rainfall storm classification method that expresses, in a convenient and intuitively fashion, the event's magnitude in terms of flood potential.

The severity level of urban rainfall storm floods was established by successfully applying a stepwise regression function to individual rainfall storm characteristics that best represented the documented basement flooding within 16 sectors in the City of Edmonton. The rainfall storm characteristics that were considered include: spatial storm average total rainfall, peak storm total rainfall, spatial storm average rainfall intensity, peak storm rainfall intensity, and spatial storm average rainfall duration.

The final model used to derive the urban rainfall storm flood severity index was shown to provide a better fit to the documented basement flooding than the IDF method for the entire study area and period of record.

6.1 Recommendations for Future Research

Considering the research and findings, the following set of recommendations are put forward for future research:

1. Future work could also include an attempt at producing a realistic estimate of radar rainfall for pixels that were masked as a function of neighbouring un-masked pixels. This would likely provide for a better handling of erroneous

radar pixel data or, of questionable quality since they are currently masked and eliminated for further analysis.

2. The stepwise regression analysis could be expanded in order to include other model functions and improving the predictive capability. Should such an effort be considered, it is highly recommended that a computationally more efficient approach be used with a stable programming environment in order to obtain and assess the new model functions in a timely manner.

3. The applicability of the storm flood severity index method could be tested for other large Canadian municipalities in order to assess whether it is universally applicable in Canada and add to this new approach's credibility.

Further development on an alternative to the current IDF approach for input to drainage system designs should not use damage-related severity relations due to difficulties in establishing cause and effect; hence, the high likelihood of imparting unsubstantiated biases in the predictive model. However, adding a spatial component to an IDF-like approach and addressing other related and discussed issues could yield promising results.

References

Allouche, D., & Freure, P. (2002). Management and maintenance practices of storm and sanitary sewers in Canadian Municipalities. Institute for Catastrophic Loss Reduction Paper Series, No. 18. Institute for Catastrophic Loss Reduction: Toronto. In Sandink, D. [2007]. *Sewer backup: Homeowner perception and mitigative behaviour in Edmonton and Toronto*. Institute for Catastrophic Loss Reduction research paper series number 44, 83 pp. Retrieved from http://www.iclr.org/images/ICLR_Report_sewer_backup.pdf

Anagnostou, E. N., Pathak, C. S., & Morales, C. A. (2012). Use of storms life cycle information and lightning data in radar-rainfall estimation. *Journal of Hydrologic Engineering*. [In Press] doi:10.1061/(ASCE)HE.1943-5584.0000557

Atlas, D. (1964). Advances in radar meteorology. *Advances in Geophysics*, *10*, 317-478. doi: 10.1016/S0065-2687(08)60009-6

Bara, M., Kohnová, S., Gaál, L., Szolgay, J., & Hlavčová, K. (2009). Estimation of IDF curves of extreme rainfall by simple scaling in Slovakia. *Geophysics and Geodesy, 39*(3), 187–206.

Battan, L. J. (1973). *Radar observation of the atmosphere*, Chicago: University of Chicago Press. In Skinner, C., Bloetscher, F., & Pathak C. S. (2009). Comparison of NEXRAD and rain gauge precipitation measurements in South Florida. *Journal of Hydrologic Engineering, 14*(3), pp. 248-260. doi: 10.1061/(ASCE)1084-0699(2009)14:3(248)

Battan, L. J. (1973). *Radar observation of the atmosphere*, Chicago: University of Chicago Press. In Skolnik, M. (2008). *Radar handbook* (3rd edition). New York, NY: McGraw-Hill.

Blanchard, D. C. (1963). Raindrop size distribution in Hawaiian rains. *Journal of Meteorology. 10*, 457-473. In Battan, L. J. (1973). *Radar observation of the atmosphere*, Chicago: University of Chicago Press. In Skolnik, M. (2008). *Radar handbook* (3rd edition). New York, NY: McGraw-Hill.

Blumenfeld, K.A., Skaggs, R., & Zandlo J. (2004). Using a dense precipitation gage network to estimate annual maximum daily precipitation. In *Proceedings of 14th Conference on Applied Climatology*, Seattle, Washington.

Borga, M., Boscolo, P., Zanon, F., & Sangati, M. (2007). Hydrometeorological analysis of the August 29, 2003 flash flood in the eastern Italian Alps. *J. Hydrometeorol., 8*(5), 1049–1067. In Rusjan, S., Kobold, M., & Mikos, M. (2009). Characteristics of the extreme rainfall event and consequent flash floods in W Slovenia in September 2007. *Natural Hazards and Earth System*

Sciences, 9, 947–956. Retrieved from www.nat-hazards-earth-syst-sci.net/9/947/2009/

Bouilloud, L., Delrieu, G., Boudevillain, B., & Kirstetter, P.-E. (2010). Radar rainfall estimation in the context of post-event analysis of flash-flood events. *Journal of Hydrology, 394*(1-2), 17-27. doi:10.1016/j.jhydrol.2010.02.035

Canadian Assessment of Natural Hazards Project [CANHP] (2010). *Canadians at risk: Our exposure to natural hazards*. D. Etkin (Editor). Institute for Catastrophic Loss Reduction research paper series number 48, 223 pp. Retrieved from http://www.preventionweb.net/english/professional/publications/v.php?id=13008

Canadian Standards Association [CSA] (2010). *Technical guide: Development, interpretation and use of rainfall intensity-duration-frequency (IDF) information: Guideline for Canadian water resources practitioners*. (Lead Authors) Auld, H., Burton, B., Morris, R., & Klaassen, J.; (Contributing Author) Manson, J., Bishop, R., Haley, D., & Mailhot, A. CSA special publication, Mississauga, Ontario, Canada.

Casas, M. C., Codina, B., Redaño, A., & Lorente, J. (2004). A methodology to classify extreme rainfall events in the western Mediterranean area. *Theoretical and Applied Climatology, 77*(3-4), 139-150.

Chumchean, S., Sharma, A., & Seed, A. (2003). Radar rainfall error variance and its impact on radar rainfall calibration. *Physics and Chemistry of the Earth 28*, 27–39. doi:10.1016/S1474-7065(03)00005-6

Ciach, G. J. (2003). Local random errors in tipping-bucket rain gauge measurements. *Journal of Atmospheric and Oceanic Technology, 20*, 752–759. doi: 10.1175/1520-0426(2003)20<752:LREITB>2.0.CO;2

Creutin, J. D., Borga, M., Lutoff, C., Scolobig, A., Ruind, I., & Creton-Cazanave, L. (2009). Catchment dynamics and social response during flash floods: the potential of radar rainfall monitoring for warning procedures. *Meteorological Applications, 16*, 115–125. doi: 10.1002/met.128

Crozier, C. L., Joe, P. I., Scott, J. W., Herscovitch, H. N., & Nichols, T. R. (1991). The king city operational doppler radar: Development, all-season applications and forecasting, *Atmosphere-Ocean, 29*(3), 479-516. doi: http://dx.doi.org/10.1080/07055900.1991.9649414

Curtis, D. C., & Humphrey, J. (2010). *Fountain Creek watershed: Rainfall characterization study*. Prepared for the City of Colorado Springs. Carlton Engineering Inc., Shingle Springs, California, U.S.A.

De Michele, C., Kottegoda, N. T., & Rosso, R. (2001). The derivation of areal reduction factor of storm rainfall from its scaling properties. *Water Resources Research, 37*(12), 3247-3252.

De Michele, C., Kottegoda, N. T., & Rosso, R. (2002). IDAF (intensity-durationarea-frequency) curves of extreme storm rainfall: a scaling approach. *Water Science and Technology, 45*(2), 83-90.

Devine, K. A., & Mekis, É. (2008). Field accuracy of Canadian rain measurements. *Atmosphere-Ocean, 46*(2), 213–227. doi:10.3137/ao.460202

Doswell, C. A. (2001). Severe Convective Storms—An Overview. *Meteorological Monographs, 28*, 1–26. doi: http://dx.doi.org/10.1175/0065-9401-28.50.1

Doswell, C. A. (2003). *A Guide to F-Scale Damage Assessment*. U.S. Department of Commerce, NOAA, NWS, Silver Spring, Maryland, 94pp.

Doviak, R. J., & Zrnic, D. S. (1993). *Doppler radar and weather observations*, 2nd Ed., San Diego: Academic. In Skinner, C., Bloetscher, F., & Pathak C. S. (2009). Comparison of NEXRAD and rain gauge precipitation measurements in South Florida. *Journal of Hydrologic Engineering, 14*(3), pp. 248-260. doi: 10.1061/(ASCE)1084-0699(2009)14:3(248)

Durrans, S. R., Julian, L. T., & Yekta, M. (2002). Estimation of depth-area relationships using radar-rainfall data. *Journal of Hydrologic Engineering, 7*(5), 356-367. doi: 10.1061/(ASCE)1084-0699(2002)7:5(356)

Environment Canada [EC] (1990). *Edmonton International Airport IDF*, IDF files, National Climate Data and Information Archive, Weather Office. Retrieved from ftp://arcdm20.tor.ec.gc.ca/pub/dist/IDF/

Environment Canada [EC] (2011a). *Documentation on Environment Canada rainfall intensity-duration-frequency (IDF) tables and graphs*. What's new in EC IDF version 2, Notes on EC IDF. Retrieved from ftp://arcdm20.tor.ec.gc.ca/pub/dist/IDF/IDF_v_2.100_2011_05_17/

Environment Canada [EC] (2012). *Edmonton Int'l A IDF*, IDF v2.20, IDF files, National Climate Data and Information Archive, Weather Office. Retrieved from ftp://ftp.tor.ec.gc.ca/Pub/Engineering_Climate_Dataset/IDF/

Einfalt, T., Karsten, A. N., Golz, C., Jensen, N. E., Quirmbach, M., Vaes, G., & Vieux, B. E. (2004). Towards a roadmap for use of radar rainfall data in urban drainage. *Journal of Hydrology, 299*(3), 186–202.

Fowler, H. J., & Kilsby, C. G. (2003). A regional frequency analysis of United Kingdome extreme rainfall from 1961 to 2000. *International Journal of Climatology, 23*, 1313–1334. doi: 10.1002/joc.943

Fujita, T. T. (1981). Tornadoes and downbursts in the context of generalized planetary scales. *Journal of the Atmospheric Sciences, 38*(8), 1511-1534.

Fulton, R. A., Breidenbach, J. P., Seo, D.-J., Miller, D. A., & O'Bannon, T. (1998). The WSR-88D rainfall algorithm. *Weather and Forecasting, 13*, 377–395. doi: http://dx.doi.org/10.1175/1520-0434(1998)013<0377:TWRA>2.0.CO;2

Gandin L. S., & Kagan, R. L. (1976). *Statistics in methods of meteorological data interpretation*. Gidrometeoizdat, 359 pp. In Canadian Standards Association [CSA] (2010). *Technical guide: Development, interpretation and use of rainfall intensity-duration-frequency (IDF) information: Guideline for Canadian water resources practitioners*. (Lead Authors) Auld, H., Burton, B., Morris, R., & Klaassen, J.; (Contributing Author) Manson, J., Bishop, R., Haley, D., & Mailhot, A. CSA special publication, Mississauga, Ontario, Canada.

García-Pintado, J., Barberá, G. G., Erena, M., & Castillo, V. M. (2009). Rainfall estimation by rain gauge-radar combination: A concurrent multiplicative-additive approach, *Water Resources Research, 45*, W01415. doi:10.1029/2008WR007011

Glazner, M.K., Tomic, S., & White, B. (1999). Redefining rainfall classification. *American Society of Civil Engineers*, *5*, 1-10.

Goodrich, D. C., Faures, J. M., Woolhiser, D. A., Lane, L. J., & Sorooshian, S. (1995). Measurement and analysis of small-scale convective storm rainfall variability. *Journal of Hydrology*, *173*, 283–308. doi:10.1016/0022-1694(95)02703-R

Gourley, J. J., Hong, Y., Flamig, Z. L., Li, L., & Wang, J. (2010). Intercomparison of rainfall estimates from radar, satellite, gauge, and combinations for a season of record rainfall. *Journal of Applied Meteorology and Climatology 49*(3), 437-452.

Groisman, P. Y., & Legates, D. R., (1994). The accuracy of United States precipitation data. *Bulletin of the American Meteorological Society, 75*, 215–227. In Allen, R. J., & DeGaetano, A. T. (2005). Considerations for the use of radar-derived precipitation estimates in determining return intervals for extreme

areal precipitation amounts. *Journal of Hydrology, 315*, 203–219. doi:10.1016/j.jhydrol.2005.03.028

Groisman, P., & Easterling, D. (1994). Variability and trends of total precipitation and snowfall over the United States and Canada. *Journal of Climate, 7*(1), 184-2005. In Canadian Standards Association [CSA] (2010). *Technical guide: Development, interpretation and use of rainfall intensity-duration-frequency (IDF) information: Guideline for Canadian water resources practitioners*. (Lead Authors) Auld, H., Burton, B., Morris, R., & Klaassen, J.; (Contributing Author) Manson, J., Bishop, R., Haley, D., & Mailhot, A. CSA special publication, Mississauga, Ontario, Canada.

Gunn, K. L. S., & Marshall, J. S. (1958). The distribution with size of aggregate snowflakes, *Journal of Meteorology, 15*, 452–466. In Battan, L. J. (1973). *Radar observation of the atmosphere*, Chicago: University of Chicago Press. In Skolnik, M. (2008). *Radar handbook* (3rd edition). New York, NY: McGraw-Hill.

Habib, E. H., Krajewski, W. F., & Kruger, A. (2001). Sampling errors of tipping-bucket rain gauge measurements. *Journal of Hydrologic Engineering, 6*(2), 159-166.

Habib, E. H., Meselhe, E. A., & Aduvala, A. V. (2008). Effect of local errors of tipping-bucket rain gauges on rainfall-runoff simulations. *Journal of Hydrologic Engineering, 13*(6), 488-496. doi: 10.1061/(ASCE)1084-0699(2008)13:6(488)

Han, D. (2010). *Concise hydrology*. Ventus Publishing ApS. Book Boon. Bookboon.com

He, X., Refsgaard, J. C., Sonnenborg, T. O., Vejen, F., & Jensen, K. H. (2011). Statistical analysis of the impact of radar rainfall uncertainties on water resources modeling. *Water Resources Research, 47*, W09526. doi:10.1029/2011WR010670

Hebert, C., Weinzapfel, B., & Chambers, M. (2008). Hurricane Severity Index: A new way of estimating a tropical cyclone's destructive potential. *19th Conference on Probability and Statistics*, 21 Januar 2008, New Orleans, Louisiana, U.S.A. The American Meteorological Society.

Hogg, W. D., Carr, D. A., & Routledge, B. (1989). *Rainfall intensity-duration frequency values for Canadian locations*. Atmospheric Environment Service, Downsview, Ontario: Environment Canada. Retrieved from ftp://arcdm20.tor.ec.gc.ca/pub/dist/IDF/IDF_v_2.100_2011_05_17

Hogg, W. D., & Hogg, A.R. (2010). *Historical trends in short duration rainfall in the greater Toronto area*. Report prepared for the Toronto and Region Conservation Authority.

Hrachowitz, M., & Weiler, M. (2011). Uncertainty of precipitation estimates caused by sparse gauging networks in a small, mountainous watershed. *Journal of Hydrologic Engineering, 16*(5), 460-471. doi:10.1061/(ASCE)HE.1943-5584.0000331

Insurance Bureau of Canada [IBC] (2011). *Water damage is on the rise: are you protected?* Information brochure. Retrieved from http://www.ibc.ca/en/Home_Insurance/documents/brochures/Water_Damage_on_Rise_en_web.pdf

Independent Community Panel [ICP] (2006). Independent community panel report to the City of Hamilton. City of Hamilton: Canada. In Sandink, D. [2007]. *Sewer backup: Homeowner perception and mitigative behaviour in Edmonton and Toronto*. Institute for Catastrophic Loss Reduction research paper series number 44, 83 pp. Retrieved from http://www.iclr.org/images/ICLR_Report_sewer_backup.pdf

Intergovernmental Panel on Climate Change [IPCC] (2007). *Summary for policymakers*. In: *Climate change 2007: Synthesis report*. Contribution of working groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Core writing team, Pachauri, R. K., & Reisinger, A. (eds.)]. IPCC, Geneva, Switzerland, 104 pp.

Islam, R., & Rasmussen, P. F. (2008). Improved high-resolution radar-based rainfall estimation. *Journal of Hydrologic Engineering, 13*(9), 910-913. doi:10.1061/(ASCE)1084-0699(2008)13:9(910)

Jensen, N. E., & Pedersen, L. (2005), Spatial variability of rainfall: Variations within a single radar pixel, *Atmospheric Research, 77*, 269– 277. In Villarini, G., Mandapaka, P. V., Krajewski, W. F., & Moore, R. J. (2008), Rainfall and sampling uncertainties: A rain gauge perspective, *Journal of Geophysical Research, 113*, D11102. doi:10.1029/2007JD009214

Jobin, E. (2010). *Classification of rainfall storm events for drainage design considerations, phase 1: Current knowledge base*, Prepared for the City of Edmonton. Ottawa, Ontario, Canada.

Jobin, D. (2012). *Conversation notes regarding Environment Canada radar data processing*. December 16, 2012. Ottawa, Ontario, Canada.

Joe, P., Falla, M., Rijn, P.V., Stamadianos, L., Falla, T., Magosse, D., Ing, L., & Dobson, J. (2003). Radar data processing for severe weather in the National Radar Project of Canada. Extended poster abstract. *21st Conference on Severe Local Storms*, San Antonio, Texas, American Meteorological Society. 221-224.

Joe, P., & Lapczak, S. (2002). Evolution of the Canadian operational radar network, *2nd European Conference on Radar Meteorology (ERAD)*, Delft, Netherlands. 370-382.

Jones, D. M. A. (1955). 3 cm and 10 cm wavelength radiation backscatter from rain, *5th Weather Radar Conference*, AMS, Boston, pp. 281–285. In Battan, L. J. (1973). *Radar observation of the atmosphere*, Chicago: University of Chicago Press. In Skolnik, M. (2008). *Radar handbook* (3rd edition). New York, NY: McGraw-Hill.

Kije Sipi Ltd (2007). *Spatial analysis of rainfall over & near Edmonton*, Study report, prepared for the City of Edmonton, September 2007, Ottawa, Ontario, Canada.

Kocin, P.J. & Uccellini, L.W. (2004). A snowfall impact scale derived from Northeast storm snowfall distributions. *American Meteorological Society, 85*(2), 177-194.

Koistinen, J., Michelson, D. B., Hohti, H., & Peura, M. (2003). *Operational measurement of precipitation in cold climates*. In: *Weather radar: Principles and advanced applications*, Meischner P. (ed.), pp. 78–114. Springer, Berlin. In World Meteorological Organization (2008). *Guide to Hydrological Practices: Volume I Hydrology – From Measurement to Hydrological Information*, WMO-No. 168, Sixth Edition, Geneva. Retrieved from ftp://ftp.wmo.int/Documents/MediaPublic/Publications/Guide_to_Hydrological_Practices_WMO_no_168/WMOENG_v6_voll.pdf, on Nov 3rd 2010.

Konrad, C. E. (2001) The most extreme precipitation events over the Eastern United States from 1950 to 1996: Considerations of scale. *Journal of Hydrometeorology, 2*, 309–325.

Koutsoyiannis, D. (2003). Climate change, the Hurst phenomenon, and hydrological statistics. *Hydrological Sciences Journal, 48*(1), 3-24.

Koutsoyiannis, D., Makropoulos, C., Langousis, A., Baki, S., Efstratiadis, A., Christofides, A., Karavokiros, G., & Mamassis, N. (2009). HESS Opinions: Climate, hydrology, energy, water: recognizing uncertainty and seeking sustainability. *Hydrology and Earth System Sciences, 13*, 247–257. Retrieved from www.hydrol-earth-syst-sci.net/13/247/2009/

Krajewski, W. F., Kruger, A. & Nespor, V. (1998). Experimental and numerical studies of small-scale rainfall measurements and variability. *Water Science and*

Technology, 37(11), 131–138. In Devine, K. A., & Mekis, É. (2008). Field accuracy of Canadian rain measurements. *Atmosphere-Ocean, 46*(2), 213–227. doi:10.3137/ao.460202

Krajewski, W. F., Villarini, G., & Smith, J. A. (2010). Radar-rainfall uncertainties. *Bulletin of the American Meteorological Society, 91*, 87–94. doi: http://dx.doi.org/10.1175/2009BAMS2747.1

Kulkarni, T. (1999). *Urban infrastructure floods in Southern Ontario: A methodology to determine causality*. Unpublished masters research paper, Institute for Environmental Studies, University of Toronto, Toronto, Ontario. In Sandink, D. [2007]. *Sewer backup: Homeowner perception and mitigative behaviour in Edmonton and Toronto*. Institute for Catastrophic Loss Reduction research paper series number 44, 83 pp. Retrieved from http://www.iclr.org/images/ICLR_Report_sewer_backup.pdf

Kurtyka, J. C. (1953). *Precipitation measurements study for use with the automatic weather station*. Report of Investigation No. 20, State Water Survey Division, State of Illinois, USA. In Devine, K. A., & Mekis, É. (2008). Field accuracy of Canadian rain measurements. *Atmosphere-Ocean 46*(2), 213–227. doi:10.3137/ao.460202

Kutiel, H., Kay, P.A., (1996). Effects of network design on climatic maps of precipitation. *Climate Research, 7*, 1–10. In Michaelides, S., Levizzani, V., Anagnostou, E., Bauer, P., Kasparis, T., & Lane, J. E. (2009). Precipitation - measurement, remote sensing, climatology and modeling. *Atmospheric Research, 94*(4), 512–533. doi:10.1016/j.atmosres.2009.08.017

Langleben, M. P., (1956). The plan pattern of snow echoes at the generating level. *Journal of Meteorology, 13*, 554-560.

Lee, C. H., Lee, G., Zawadzki, I., & Kim, K.-E. (2009), A preliminary analysis of spatial variability of raindrop size distributions during stratiform rain events, *Journal of Applied Meteorology and Climatology, 48*(2), 270–283. doi:10.1175/2008JAMC1877.1

Lodewyk, S. (n. d.). Conceptual Edmonton storm classification method, The City of Edmonton. Edmonton, Alberta, Canada.

Lombardo, F., Napolitano, F., & Russo, F. (2006). On the use of radar reflectivity for estimation of the areal reduction factor. *Natural Hazards and Earth System Sciences, 6*(3), 377-386.

López, R. E., & Holle, R. L. (1986). Diurnal and spatial variability of lightning activity in northeastern Colorado and central Florida during the summer. *Monthly Weather Review, 114*, 1288-1312.

Mailhot, A., & Talbot, G. (2011). *Mise à jour des estimateurs intensité-durée-fréquence (IDF) pour le sud-Québec, Tome I - Données et méthodes*. Projet Mise à jour des normes et procédures de conception des ouvrages hydro-agricoles dans un contexte de changements climatiques - Volet Météorologique. Rapport de recherche R-1259. Institut national de la recherche scientifique, Centre Eau Terre Environnement.

Marshall, M. P. (1957). *The constant-altitude presentation of radar weather patterns*. Preprints Sixth Radar Weather Conference, Cambridge, Mass., Amer. Meteor. Soc., 321-324; reprinted in Selected Meteorological Papers, The Meteorological Society of Japan, 11, December 1972, 52-55.

Marshall, J. S., & Ballantyne, E. H. (1975). Weather surveillance radar, *Journal of Applied Meteorology, 14*, 1317–1338.

Marshall, J. S., & Palmer, W. McK. (1948). The distribution of raindrops with size. *Journal of Meteorology, 6*, 165-166.

Marshall, J. S., Hitschfeld, W., & Gunn, K. L. S. (1955), Advances in radar weather, in *Advances in Geophysics*, 2, 1 –56, New York: Academic Press.

Martner, B. E., & Dubovskiy, V. (2005). Z-R relations from raindrop disdrometers: Sensitivity to regression methods and DSD data refinements, *32nd Radar Meteorology Conference*, Albuquerque, NM, October, 2005. In Parzybok, T. W., Hultstrand, D. M., Clarke, B., Tomlinson, E. M. & Kappel, W. D. (2010). Improving hydrologic analysis and applications through the use of quality controlled radar data and the storm precipitation analysis system, *Canadian Dam Association 2010 Annual Conference*, Niagara Falls, Ontario, Canada. 2010 Oct 2-7.

Mazzarella, A., & Diodato, N. (2002). The alluvial events in the last two centuries at Sarno, southern Italy: their classification and power-law time-occurrence. *Theoretical and Applied Climatology, 72*(1-2). 75-84.

Meaden, G.T., Kochev, S., Kolendowicz, L., Kosa-Kiss, A., Marcinoniene, I., Sioutas, M., Tooming, H., & Tyrrell, J. (2007). Comparing the theoretical versions of the Beaufort scale, the T-Scale and the Fujita scale. *Atmospheric Research, 83*, 446-449.

Mekis, É., & Hogg., W. (1999). Rehabilitation and analysis of Canadian daily precipitation time series. *Atmosphere – Ocean, 37*(1), 53-85.

Mekis, É., & Vincent, L. A. (2011). An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada. *Atmosphere-Ocean, 49*(2), 163-177.

Metcalfe, J. R., Routledge, B., & Devine, K. (1997). Rainfall measurement in Canada: Changing observational methods and archive adjustment procedures. *Journal of Climate, 10*, 92–101. doi: http://dx.doi.org/10.1175/1520-0442(1997)010<0092:RMICCO>2.0.CO;2

Michaelides, S., Levizzani, V., Anagnostou, E., Bauer, P., Kasparis, T., & Lane, J. E. (2009). Precipitation - measurement, remote sensing, climatology and modeling. *Atmospheric Research 94*(4), 512–533. doi:10.1016/j.atmosres.2009.08.017

Mileti, D. (1999). *Disasters by design: A reassessment of natural hazards in the United States*. Washington DC: Joseph Henry Press.

Morin, E., Krajewski, W. F., Goodrich, D. C., Gao, X., & Sorooshian, S. (2003). Estimating rainfall intensities from weather radar data: The scale-dependency problem. *Journal of Hydrometeorology, 4*, 782–797. doi: http://dx.doi.org/10.1175/1525-7541(2003)004<0782:ERIFWR>2.0.CO;2

Nixon, W. A., & Qiu, L. (2005). Developing a storm severity index. Transportation Research Record. *Journal of the Transportation Research Board, 1911*, 143-148.

Natural Resources Canada [NRCan] (2007). *National Topographic Data Base (NTDB)* 083H05, 083H05, 083H06, 083H11, 083H12. Government of Canada, Centre for Topographic Information. Sherbrooke, Quebec, Canada. Retrieved from http://www.GeoGratis.gc.ca

Nystuen, J. A. (1998). Temporal sampling requirements for automatic rain gauges. *Journal of Atmospheric and Oceanic Technology, 15*, 1253–1260. doi: http://dx.doi.org/10.1175/1520-0426(1998)015<1253:TSRFAR>2.0.CO;2

Office of the Auditor General of Canada [OAGC] (2008). *Report of the Commissioner of the Environment and Sustainable Development to the House of Commons. Chapter 2: Managing Severe Weather Warnings — Environment Canada.* December 2008. Minister of Public Works and Government Services Canada 2008. Retrieved from http://www.oag-bvg.gc.ca/internet/docs/parl_cesd_200812_02_e.pdf

Office of the Federal Coordinator's Interdepartmental Committee for Meteorological Services and Supporting Research [OFCM] (2006). *Doppler radar meteorological observations, Part C: WSR-88D products and algorithms*. Federal Meteorological Handbook #11, FCM-H 11C-2006, US Department of Commerce, Washington, DC. Retrieved from http://www.ofcm.gov/fmh11/fmh11C.htm

Ohara, N., Kavvas, M. L., Kure, S., Chen, Z. Q., Jang, S., & Tan, E. (2011). Physically based estimation of maximum precipitation over American River watershed, California. *Journal of Hydrologic Engineering, 16*(4), 351-361. doi:10.1061/(ASCE)HE.1943-5584.0000324

Olivera, F., Choi, J., Kim, D., & Li, M. (2008). Estimation of average rainfall areal reduction factors in Texas using NEXRAD data. *Journal of Hydrologic Engineering, 13*(6), 438-448.

Overeem, A., Holleman, I., & Buishand, A. (2009). Derivation of a 10-Year radar-based climatology of rainfall. *Journal of Applied Meteorology and Climatology, 48*, 1448–1463. doi: http://dx.doi.org/10.1175/2009JAMC1954.1

Parzybok, T. W., Hultstrand, D. M., Clarke, B., Tomlinson, E. M., & Kappel, W. D. (2010). Improving hydrologic analysis and applications through the use of quality controlled radar data and the storm precipitation analysis system, *Canadian Dam Association 2010 Annual Conference*, Niagara Falls, Ontario, Canada. 2010 Oct 2-7.

Pastoriza, V., Núñez, A., Machado, F., Mariño, P., Fontán F. P., & Fiebig, U.-C. (2009). Combining meteorological radar and network of rain gauges data for space–time model development. *International Journal of Satellite Communications and Networking, 29*, 61-78. DOI: 10.1002/sat.954

Pedersen, L., Jensen, N. E., & Madsen, H. (2007). *Extreme rainfall statistics based on rain gauges and radar measurement*. Department of Informatics and Mathematical Modelling, Technical University of Denmark.

Peterson, E. W., & Hasse, L. (1987). Did the Beaufort scale or the wind climate change? *American Meteorological Society, 17*(7), 1071-1074.

Pleau, M., Colas, H., Lavallee, P., Genevieve, P., & Bonin, R. (2005). Global optimal real-time control of the Quebec urban drainage system. *Environmental Modeling & Software, 20*, 401-413. In Sandink, D. [2007]. *Sewer backup: Homeowner perception and mitigative behaviour in Edmonton and Toronto*. Institute for Catastrophic Loss Reduction research paper series number 44, 83 pp. Retrieved from http://www.iclr.org/images/ICLR_Report_sewer_backup.pdf

R Core Team, The (2012). R: A language and environment for statistical computing. Reference index. R Foundation for Statistical Computing, version 2.15.1 (2012-06-22). Retrieved from http://cran.r-project.org/doc/manuals/fullrefman.pdf

Ramos, M. H., Creutin, J. D., & Leblois, E. (2005). Visualization of storm severity, *Journal of Hydrology, 315*, 295–307.

Ramos, M. H., Leblois, E., & Creutin, J.-D. (2006). From point to areal rainfall: linking the different approaches for the frequency characterization of rainfalls in urban areas. *Water Science & Technology*, *54*(6-7), 33-40.

Rozalis, S., Morin, E., Yair, Y., & Price, C. (2010). Flash flood prediction using an uncalibrated hydrological model and radar rainfall data in a Mediterranean watershed under changing hydrological conditions. *Journal of Hydrology*, [Article in press]. doi:10.1016/j.jhydrol.2010.03.021

Rusjan, S., Kobold, M., & Mikos, M. (2009). Characteristics of the extreme rainfall event and consequent flash floods in W Slovenia in September 2007. *Natural Hazards and Earth System Sciences, 9*, 947–956. Retrieved from www.nat-hazards-earth-syst-sci.net/9/947/2009/

Ryzhkov, A., & Zrnic´, D. (1995). Precipitation and attenuation measurements at a 10-cm wavelength. *Journal of Applied Meteorology, 34*, 2121–2134. In Fulton, R. A., Breidenbach, J. P., Seo, D.-J., Miller, D. A., & O'Bannon, T. (1998). The WSR-88D rainfall algorithm. *Weather and Forecasting, 13*, 377–395. doi: http://dx.doi.org/10.1175/1520-0434(1998)013<0377:TWRA>2.0.CO;2

Sandink, D. [2007]. *Sewer backup: Homeowner perception and mitigative behaviour in Edmonton and Toronto*. Institute for Catastrophic Loss Reduction research paper series number 44, 83 pp. Retrieved from http://www.iclr.org/images/ICLR_Report_sewer_backup.pdf

Schilling, W. (1991). Rainfall data for urban hydrology: what do we need? *Atmospheric Research, 27*, 5-21.

Senkbeil, J. C., & Sheridan, S. C. (2006). A postlandfall hurricane classification system for the United States. *Journal of Coastal Research, 22*(5), 1025-1034.

Sevruk, B., & Klemm. S. (1989). *Catalogue of national standard precipitation gauges*. Instrument and Observing Methods Report No.39, WMO/TD-No. 313. In Devine, K. A., & Mekis, É. (2008). Field accuracy of Canadian rain measurements. *Atmosphere-Ocean 46*(2), 213–227. doi:10.3137/ao.460202

Sevruk, B., & Nespor, V. (1994). The effect of dimensions and shape of precipitation gauges on the wind-induced error. In: Desbois, M. & Desalmand, F. (eds.): *Global Precipitation and Climate Change*, NATO ASI Series, I26, Springer Verlag, Berlin, pp. 231–246. In: World Meteorological Organization [WMO] (2008). *Guide to Meteorological Instruments and Methods of Observation*, WMO-No. 8, Seventh edition, Geneva. Retrieved from ftp://ftp.wmo.int/Documents/MediaPublic/Publications/CIMOguide_WMO_no_08/Part%20I.pdf

Shabbar A., Bonsal, B., & Khandekar, M. (1997). Canadian precipitation patterns associated with the Southern oscillation. *Journal of Climate*, *10*, 3016-3027. In Canadian Standards Association [CSA] (2010). *Technical guide: Development, interpretation and use of rainfall intensity-duration-frequency (IDF) information: Guideline for Canadian water resources practitioners*. (Lead Authors) Auld, H., Burton, B., Morris, R., & Klaassen, J.; (Contributing Author) Manson, J., Bishop, R., Haley, D., & Mailhot, A. CSA special publication, Mississauga, Ontario, Canada.

Sieck, L. C., Burges, S. J., & Steiner, M. (2007), Challenges in obtaining reliable measurements of point rainfall, *Water Resources Research, 43*, W01420, doi:10.1029/2005WR004519.

Skinner, C., Bloetscher, F., & Pathak, C. S. (2009). Comparison of NEXRAD and rain gauge precipitation measurements in South Florida. *Journal of Hydrologic Engineering, 14*(3), pp. 248-260 doi: 10.1061/(ASCE)1084-0699(2009)14:3(248)

Skolnik, M. (2008). *Radar Handbook* (3rd edition). New York, NY: McGraw-Hill.

Smith, J. A., Baeck, M. L., Meierdiercks, K. L., Miller, A. J., & Krajewski, W. F. (2007). Radar rainfall estimation for flash flood forecasting in small urban watersheds, *Advances in Water Resources, 30*, 2087–2097. doi:10.1016/j.advwatres.2006.09.007

Sparks, P. R. (2003). Wind speeds in tropical cyclones and associated insurance losses. *Journal of Wind Engineering and Industrial Aerodynamics, 91*(12), 1731-1751.

Svensson, C., Clarke, R. T., & Jones, D. A. (2007). An experimental comparison of methods for estimating rainfall intensity-duration-frequency relations. *Journal of Hydrology, 341*, 79–89. doi:10.1016/j.jhydrol.2007.05.002

UMA Engineering Limited (2005). Master plan for the City of Peterborough. Unpublished report commissioned by the City of Peterborough. Mississauga: UMA Engineering Limited. In Sandink, D. [2007]. *Sewer backup: Homeowner perception and mitigative behaviour in Edmonton and Toronto*. Institute for Catastrophic Loss Reduction research paper series number 44, 83 pp. Retrieved from http://www.iclr.org/images/ICLR_Report_sewer_backup.pdf

Ungar, S. (1999). Is strange weather in the air? A study of U.S. national network news coverage of extreme weather events. *Climatic Change, 41*(2), 133–150.

Helsel, D. R., & Hirsch, R. M. (2002). *Statistical Methods in Water Resources Techniques of Water Resources Investigations*, Book 4, chapter A3. 522 pp.

Veneziano, D., & Langousis, A. (2005). The areal reduction factor: A multifractal analysis. *Water Resources Research, 41*(7), W07008.

Vieux, B. E., & Bedient, P. B. (2004). Assessing urban hydrologic prediction accuracy through event reconstruction. *Journal of Hydrology, 299*, 217–236. doi:10.1016/j.jhydrol.2004.08.005

Villarini, G., Mandapaka, P. V., Krajewski, W. F., & Moore, R. J. (2008). Rainfall and sampling uncertainties: A rain gauge perspective, *Journal of Geophysical Research, 113*, D11102, doi:10.1029/2007JD009214.

Villarini, G., & Krajewski, W. F. (2010). Review of the different sources of uncertainty in single-polarization radar-based estimates of rainfall. *Surveys in Geophysics, 31*(1). 107-129. doi: 10.1007/s10712-009-9079-x

World Meteorological Organization [WMO] (2008a). *Guide to meteorological instruments and methods of observation*, WMO-No. 8, seventh edition, Geneva. Retrieved from ftp://ftp.wmo.int/Documents/MediaPublic/Publications/CIMOguide_WMO_no_08/Cov_Pref_etc.pdf, ftp://ftp.wmo.int/Documents/MediaPublic/Publications/CIMOguide_WMO_no_08/Part%20I.pdf, ftp://ftp.wmo.int/Documents/MediaPublic/Publications/CIMOguide_WMO_no_08/Part%20II.pdf, ftp://ftp.wmo.int/Documents/MediaPublic/Publications/CIMOguide_WMO_no_08/Part%20III.pdf,

World Meteorological Organization [WMO] (2008b). *Guide to hydrological practices: Volume I Hydrology – From measurement to hydrological information*, WMO-No. 168, sixth edition, Geneva. Retrieved from ftp://ftp.wmo.int/Documents/MediaPublic/Publications/Guide_to_Hydrological_Practices_WMO_no_168/WMOENG_v6_volI.pdf,

World Meteorological Organization [WMO] (2009). *Guide to hydrological practices: Volume II Hydrology – Management of water resources and application of hydrological practices*, WMO-No. 168, sixth edition, Geneva. Retrieved from ftp://ftp.wmo.int/Documents/MediaPublic/Publications/Guide_to_Hydrological_Practices_WMO_no_168/WMOENG_v6_volII.pdf

World Meteorological Organization [WMO] (2010). *Guide to the global observing system*, WMO-No. 488. Retrieved from ftp://ftp.wmo.int/Documents/MediaPublic/Publications/488_en.pdf

Wheater, H. S., Isham, V.S., Cox, D. R., Chandler, R. E., Kakou, A., Northrop, P. J., Oh, L., Onaf, C., & Rodriguez-Iturbe, I. (2000). Spatial-temporal rainfall fields: modelling and statistical aspects. *Hydrology and Earth System Sciences, 4*, 581-601. In Canadian Standards Association [CSA] (2010). *Technical guide: Development, interpretation and use of rainfall intensity-duration-frequency (IDF) information: Guideline for Canadian water resources practitioners*. (Lead Authors) Auld, H., Burton, B., Morris, R., & Klaassen, J.; (Contributing Author) Manson, J., Bishop, R., Haley, D., & Mailhot, A. CSA special publication, Mississauga, Ontario, Canada.

Wilson, J. W., & Brandes, E. A. (1979). Radar measurement of rainfall — A summary. *Bulletin of The American Meteorological Society, 60*(9), 1048-1060. doi: 10.1175/1520-0477(1979)060<1048:RMORS>2.0.CO;2

Zawadzki, I. I. (1975). On radar-raingage comparison. *Journal of Applied Meteorology*, *14*, 1430–1436.

Zhang, Y., & Smith, J. A. (2003). Space–time variability of rainfall and extreme flood response in the Menomonee River basin, Wisconsin. *Journal of Hydrometeorology, 4*, 506–517. doi: http://dx.doi.org/10.1175/1525-7541(2003)004<0506:SVORAE>2.0.CO;2

Zhang, J., Langston, C., & Howard, K. (2008). Three-dimensional radar mosaic integrating WSR-88Ds and Canadian radar network. Preprints, *The 13th Conference on Aviation, Range, and Aerospace Meteorology*, New Orleans, LA, USA, American Meteorological Society, CD-ROM, p4.4.

Zielinski, G. (2002). A classification scheme for winter storms in the Eastern and Central United States with an emphasis on nor'easters. *American Meteorological Society, 83*(1), 37-51.

Programs

Golden Software Inc. (2010). MapViewer (version 7.6.3305) [Computer software] Golden, Colorado. Released on August 13th, 2010. Available from www.goldensoftware.com

Golden Software Inc. (2012). Surfer (version 10.7.972 64-bit) [Computer software] Golden, Colorado. Released on March 5th, 2012. Available from www.goldensoftware.com

Golden Software Inc. (2012). Grapher (version 9.6.1001 64-bit) [Computer software] Golden, Colorado. Released on September 26th, 2012. Available from www.goldensoftware.com

R Foundation for Statistical Computing, The (2012). (version 2.15.1) R [Computer software] Vienna, Austria. Released on June 22nd, 2012. Available from http://www.r-project.org/foundation/

Online Resources

Environment Canada [EC] (2010a). *Weather tools - Frequently asked questions: Current weather*. Accessed on December 13, 2011: http://www.ec.gc.ca/meteo-weather/default.asp?lang=En&n=108C6C74-1.

Environment Canada [EC] (2010b). *Atmospheric hazards - Prairie and Northern region*. Accessed on February 2, 2012: http://pnr.hazards.ca/lightning.html

Environment Canada [EC] (2011b). *Weather and meteorology - Weather tools - About radar*. Accessed on June 21, 2012, last modified on November 15th, 2011. http://www.ec.gc.ca/meteo-weather/default.asp?lang=En&n=2B931828-1.

Environment Canada [EC] (2011c). *Weather and meteorology - Weather tools - radar*. Accessed on June 21, 2012, last modified on July 11th, 2011. http://www.ec.gc.ca/meteo-weather/default.asp?lang=En&n=B761FDEA-1.

Environment Canada [EC] (2012). *Media room - Backgrounders - Weather and meteorology*. Last Modified January 25th, 2012. Accessed on September 21st, 2012. http://www.ec.gc.ca/default.asp?lang=En&n=592AB94B-1&news=06F87D0A-4EC0-41F2-99EE-729855FCEA65

Horstmeyer, S. L. (2008). *Deformation of water in the air*. Retrieved on February 18th, 2012 from http://www.shorstmeyer.com/wxfaqs/float/dropdeform.html

National Oceanic and Atmospheric Administration [NOAA] (2011). *WSR-88D dual polarization background and status*. NOAA's National Weather Service - Radar Operations Center - NEXRAD WSR-88D. United States Department of Commerce. Accessed on September 21st, 2012. Last update on September 28th, 2011. http://www.roc.noaa.gov/WSR88D/dualpol/DualPolOverview.aspx

Wikipedia (2009). *Trajectoiries at differents angles of the radar beam*. This diagram is showing angles used by the Canadian Meteorological Service of Environment Canada. Accessed on January 4, 2013. Last update on February 13th, 2009. http://en.wikipedia.org/wiki/File:Radar-angles.png

Appendices

Appendix A - Total Storm Rainfall Maps of the 6 Marginal Cases for 3 Basement Flooding Events

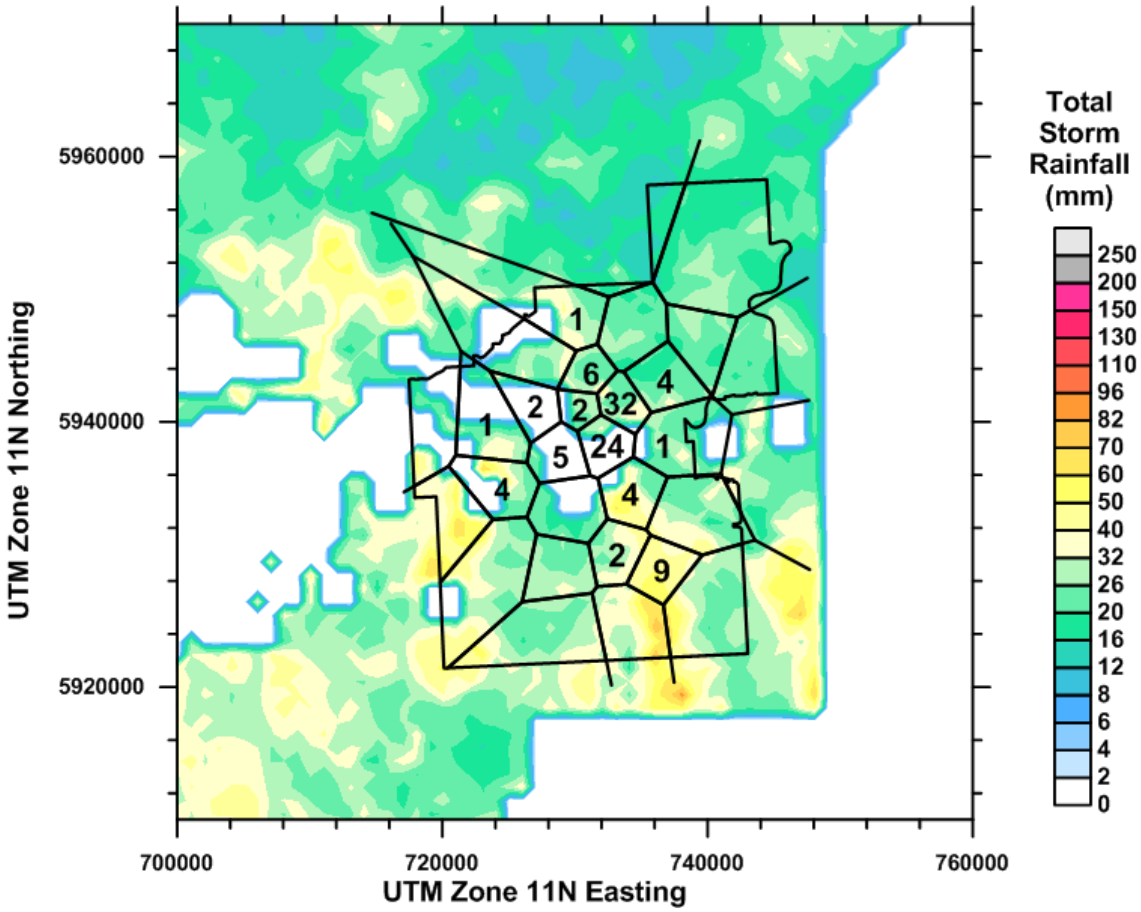


Figure A.1 - RERSD storm 1382032526 retained as likely causing basement floodings on June 29 to July 5, 1998

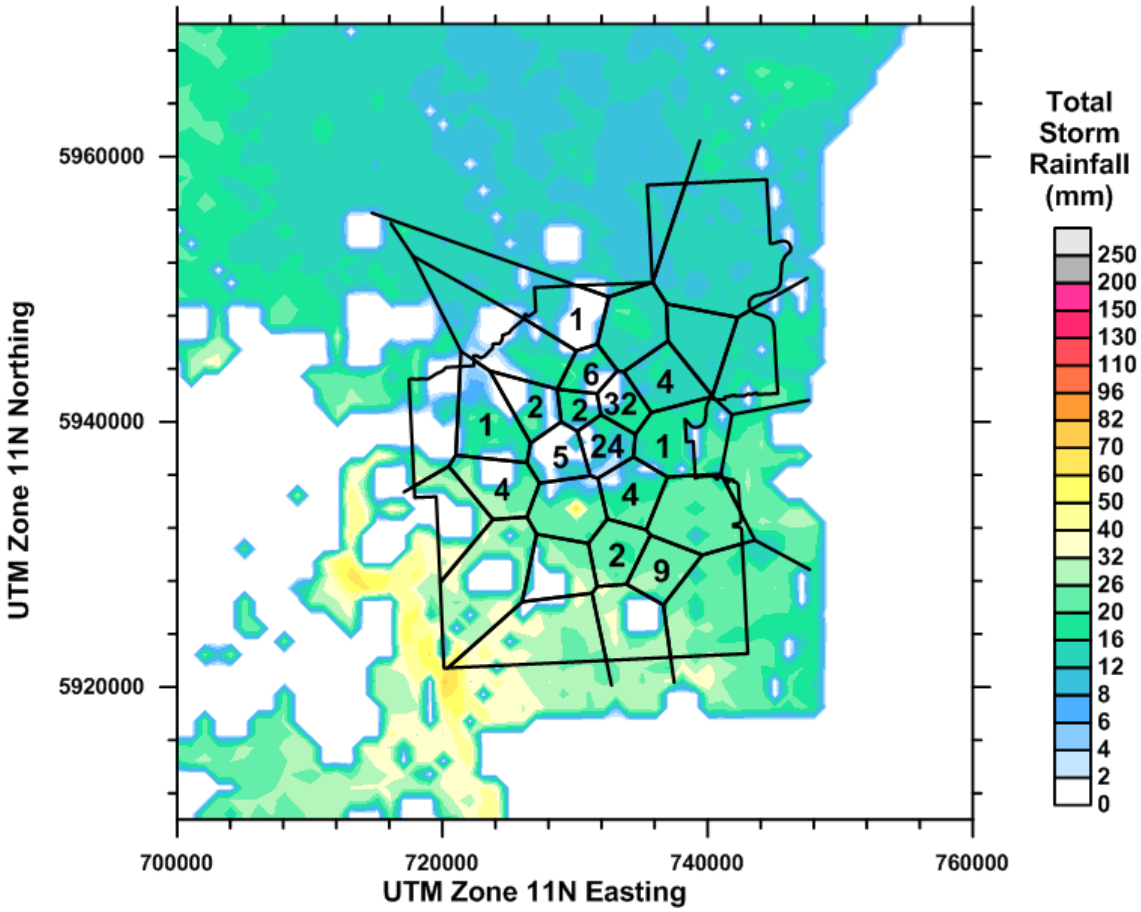


Figure A.2 - RERSD storm 1237456405 deemed to not have caused the basement floodings of June 29 to July 5, 1998

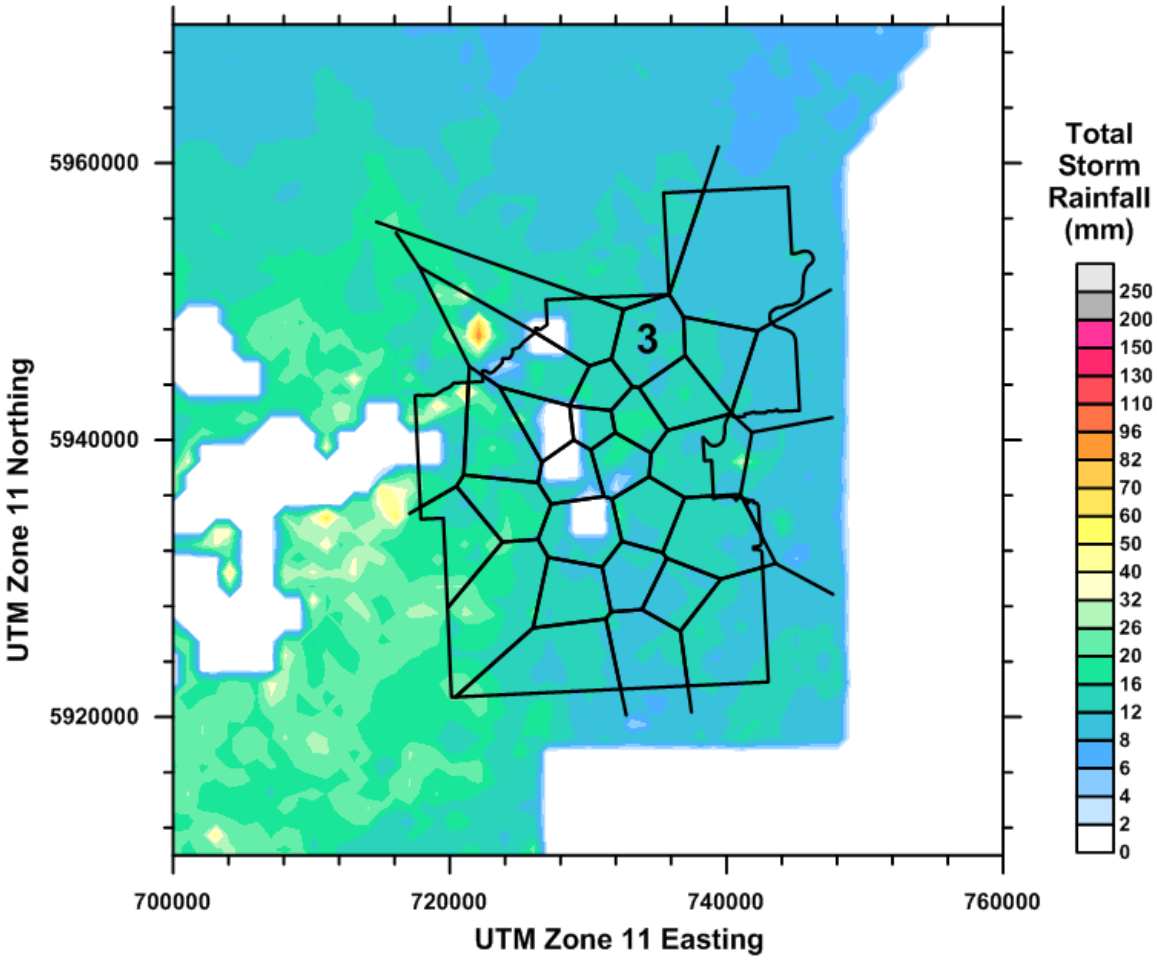


Figure A.3 - RERSD storm 1496568720 retained as likely causing basement floodings on July 9 and 13, 1999

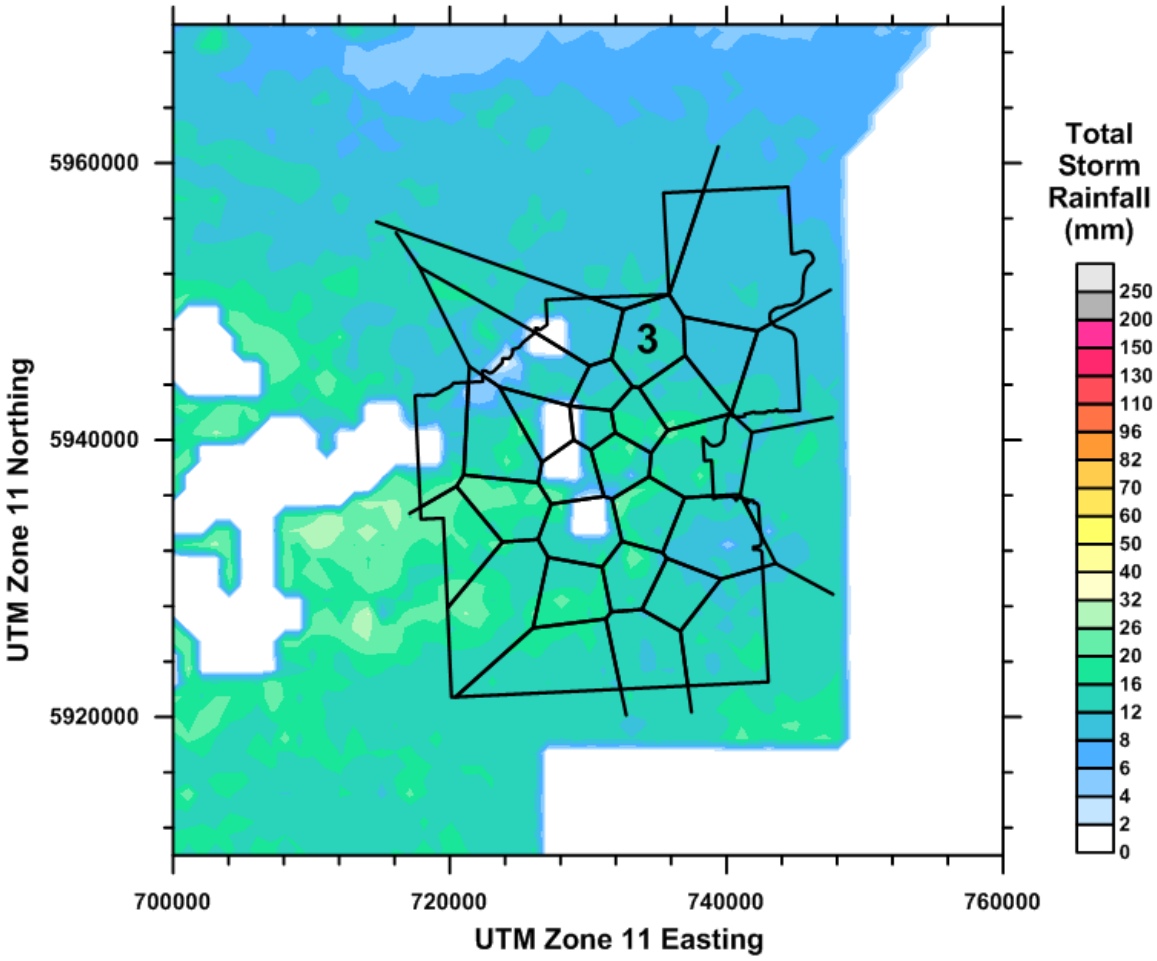


Figure A.4 - RERSD storm 1558109325 deemed to not have caused the basement floodings of July 9 and 13, 1999

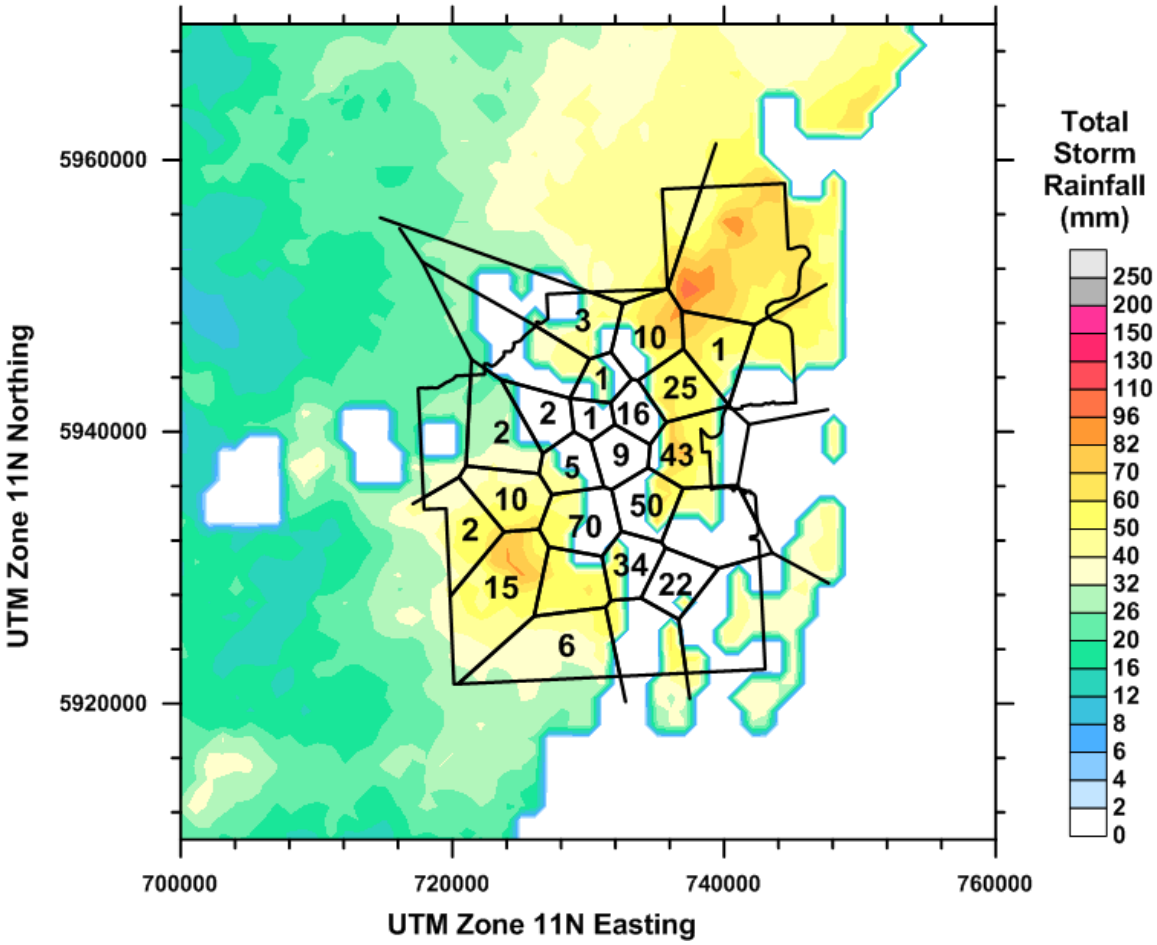


Figure A.5 - RERSD storm 1387611681 retained as likely causing basement floodings on July 2 and 6, 2004

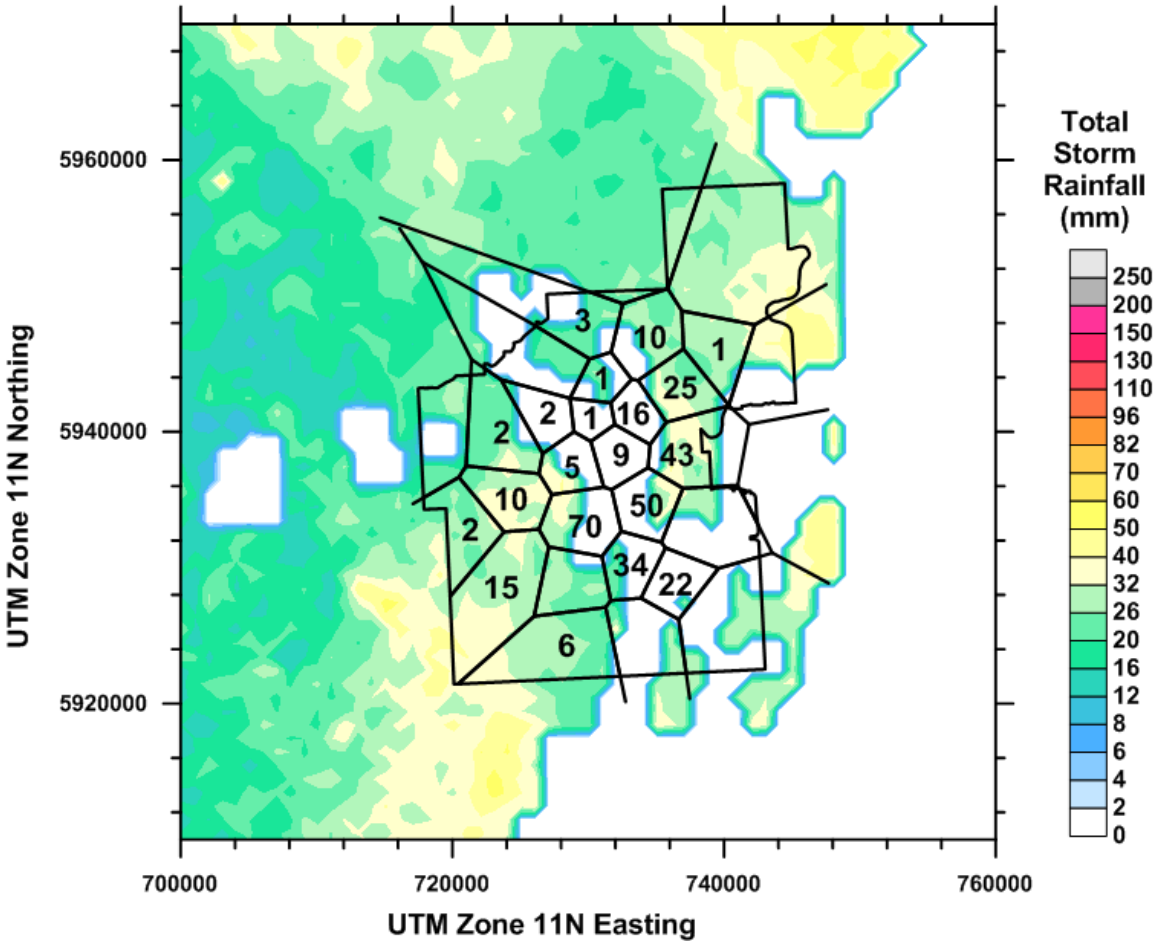


Figure A.6 - RERSD storm 1474785606 deemed to not have caused the basement floodings of July 2 and 6, 2004

Appendix B - Scripter Code

```
Const template As String = "D:\Documents\Masters\Thesis\New Radar Data\Modified Unzipped Database Cells Over
        Edmonton\Cells Over Edmonton - 1998.srf"

Sub Main
        Set SurferApp = CreateObject("Surfer.Application")

Dim counter As Integer
        counter = 0

Dim cell (65) As String

cell (0) = "48695278"
cell (1) = "321198612"
cell (2) = "339296732"
cell (3) = "366154785"
cell (4) = "410781408"
cell (5) = "411459657"
cell (6) = "412151361"
cell (7) = "412176881"
cell (8) = "469992617"
cell (9) = "544734669"
cell (10) = "621792805"
cell (11) = "634936514"
cell (12) = "672630178"
cell (13) = "715439216"
cell (14) = "800164929"
cell (15) = "811017229"
cell (16) = "822556924"
cell (17) = "822624219"
cell (18) = "893170131"
cell (19) = "998618319"
cell (20) = "1025969875"
cell (21) = "1031546574"
cell (22) = "1039156512"
cell (23) = "1191391981"
cell (24) = "1212419980"
cell (25) = "1237456405"
cell (26) = "1320254390"
cell (27) = "1342951737"
cell (28) = "1354964257"
cell (29) = "1356357275"
cell (30) = "1363029205"
cell (31) = "1366533795"
cell (32) = "1382032526"
cell (33) = "1385716095"
cell (34) = "1646250637"
cell (35) = "1680042821"
cell (36) = "1780217805"
cell (37) = "1802187473"
cell (38) = "1810288491"
cell (39) = "1810988788"
cell (40) = "1915088319"
cell (41) = "2020362391"
cell (42) = "2075647567"
cell (43) = "2197575764"
cell (44) = "2220950450"
cell (45) = "2220951398"
cell (46) = "2240856582"
cell (47) = "2260279850"
cell (48) = "2354394251"
cell (49) = "2356707861"
cell (50) = "2407886685"
cell (51) = "2521937645"
cell (52) = "2620748002"
cell (53) = "2639502663"
cell (54) = "2641335285"
cell (55) = "2684364075"
cell (56) = "2756779142"
cell (57) = "2805364929"
cell (58) = "2969398936"
```

```
cell (59) = "2969866919"
cell (60) = "3107987074"
cell (61) = "3169545872"
cell (62) = "3232883522"
cell (63) = "3274516949"
cell (64) = "3321539512"


While (counter<65)
	'Scale Grid from km to m
	SurferApp.GridTransform( _
	   InGrid:="D:\Documents\Masters\Thesis\New Radar Data\Unzipped Database Cells Over
	Edmonton\1998\"+cell(counter)+"\"+cell(counter)+"_total.grd", _
	   Operation:=srfGridTransScale, _
	   xScale:=1000, _
	   yScale:=1000, _
	   OutGrid:="D:\Documents\Masters\Thesis\New Radar Data\Modified Unzipped Database Cells Over
	Edmonton\1998\"+cell(counter)+"_total - scaled to m.grd", _
	   OutFmt:=srfGridFmtAscii)



	'Offset Grid from relative reference frame to UTM
	SurferApp.GridTransform( _
	   InGrid:="D:\Documents\Masters\Thesis\New Radar Data\Modified Unzipped Database Cells Over
	Edmonton\1998\"+cell(counter)+"_total - scaled to m.grd", _
	   Operation:=srfGridTransOffset, _
	   xOffset:=449089.648, _
	   yOffset:=5698422.567, _
	   OutGrid:="D:\Documents\Masters\Thesis\New Radar Data\Modified Unzipped Database Cells Over
	Edmonton\1998\"+cell(counter)+"_total - scaled to m - offset to UTM.grd", _
	   OutFmt:=srfGridFmtAscii)

	   counter = counter + 1

Wend


	'cleanup
	Set SurferApp = Nothing
	'cell (65) = 0
	counter = 0

End Sub


Sub Main

	'Declare object and string variables used in the script
	Dim SurferApp, Plot, ContourMapFrame, ContourMap As Object

	'Create the Surfer Application object and assign it to the "SurferApp" variable
	Set SurferApp = CreateObject("Surfer.Application")

	SurferApp.Visible = True 'Make Surfer visible

	'Create a plot document in Surfer and assign it to the variable named "Plot"
	Set Plot = SurferApp.Documents.Add(srfDocPlot)


Dim counter As Integer
	counter = 0

Dim cell (65) As String

cell (0) = "48695278"
cell (1) = "321198612"
```

```
cell (2) = "339296732"
cell (3) = "366154785"
cell (4) = "410781408"
cell (5) = "411459657"
cell (6) = "412151361"
cell (7) = "412176881"
cell (8) = "469992617"
cell (9) = "544734669"
cell (10) = "621792805"
cell (11) = "634936514"
cell (12) = "672630178"
cell (13) = "715439216"
cell (14) = "800164929"
cell (15) = "811017229"
cell (16) = "822556924"
cell (17) = "822624219"
cell (18) = "893170131"
cell (19) = "998618319"
cell (20) = "1025969875"
cell (21) = "1031546574"
cell (22) = "1039156512"
cell (23) = "1191391981"
cell (24) = "1212419980"
cell (25) = "1237456405"
cell (26) = "1320254390"
cell (27) = "1342951737"
cell (28) = "1354964257"
cell (29) = "1356357275"
cell (30) = "1363029205"
cell (31) = "1366533795"
cell (32) = "1382032526"
cell (33) = "1385716095"
cell (34) = "1646250637"
cell (35) = "1680042821"
cell (36) = "1780217805"
cell (37) = "1802187473"
cell (38) = "1810288491"
cell (39) = "1810988788"
cell (40) = "1915088319"
cell (41) = "2020362391"
cell (42) = "2075647567"
cell (43) = "2197575764"
cell (44) = "2220950450"
cell (45) = "2220951398"
cell (46) = "2240856582"
cell (47) = "2260279850"
cell (48) = "2354394251"
cell (49) = "2356707861"
cell (50) = "2407886685"
cell (51) = "2521937645"
cell (52) = "2620748002"
cell (53) = "2639502663"
cell (54) = "2641335285"
cell (55) = "2684364075"
cell (56) = "2756779142"
cell (57) = "2805364929"
cell (58) = "2969398936"
cell (59) = "2969866919"
cell (60) = "3107987074"
cell (61) = "3169545872"
cell (62) = "3232883522"
cell (63) = "3274516949"
cell (64) = "3321539512"

'-----------------
While (counter<65)

        'Create a contour map. Assign the map frame to the " ContourMapFrame" variable
```

```
        Set ContourMapFrame = Plot.Shapes.AddContourMap("D:\Documents\Masters\Thesis\New Radar Data\Modified
        Unzipped Database Cells Over Edmonton\1998\"+cell(counter)+"_total - scaled to m - offset to UTM.grd")

        'Assign the contour map properties to the variable named "ContourMap"
        Set ContourMap = ContourMapFrame.Overlays(1)

        'Fill the contour map levels
        ContourMap.FillContours = True

        'Shows contour map's color scale
        ContourMap.ShowColorScale = True


        Dim Levels As Object 'Declares Levels as an object
        Set Levels = ContourMap.Levels 'Assigns the Levels collection to the variable named "Levels"


        'ColorMap.LoadFile(path/file)
        ContourMap.Levels.LoadFile("D:\Documents\Masters\Thesis\New Radar Data\0TO250-LOG.lvl")

        'Assign coutour map name
        ContourMap.Name = cell(counter)

        counter = counter + 1

'---------------------------
Wend


        'cleanup
        Set SurferApp = Nothing
        Set Plot = Nothing
        Set ContourMapFrame = Nothing
        Set ContourMap = Nothing
        Set Levels = Nothing
        counter = 0

End Sub
```

Appendix C - Assessing Dataset Distribution Similarity

Two-sided Kruskal-Wallis rank sum tests (using the R programing environment and its native functions) were performed to assess whether datasets had similar distributions. This nonparametric approach can be used to compare several groups of data with equal or unequal dataset sizes.

Thus the null hypothesis can be written as:

H_0: both datasets have identical distributions; alternatively,

H_1: the dataset distribution differ.

H_0 is rejected if K, the Kruskal-Wallis chi-squared test statistic, is greater or equal to the $\chi^2_{1-\alpha,(k-1)}$ (Helsel & Hirsch, 2002). $\chi^2_{1-\alpha,(k-1)}$ is the 1-α quantile of a chi-square distribution with (k-1) degrees of freedom where k is the number of datasets being tested, α is the significance level (Helsel & Hirsch, 2002).

Storm characteristics and reported basement floodings of the ERSPD were respectively tested against their counterpart dataset as follows:

- the entire ERSPD dataset compared to polygon 32's dataset,
- polygon 32's entire dataset compared against polygon 32's calibration dataset and,
- polygon 32's calibration dataset versus polygon 32's validation dataset.

Note that a typical α value of 0.05 was used. Also, for comparisons between only two datasets at a time, the $\chi^2_{0.95,(1)}$ is equal to 3.841.

Of further interest is the Helsel and Hirsch's (2002, p. 108) definition of the p-value, also known as p or the 'significance level': *"[It] is the probability of obtaining the computed test statistic, or one even less likely, when the null hypothesis is true. (...) The smaller the p-value, the less likely is the observed test statistic when H_0 is true, and the stronger the evidence for rejection of the null hypothesis."* Essentially, the null hypothesis is rejected when the p-value is smaller than the α value.

Tables C.1, C.2 and C.3 summarize the dataset distribution comparison test results. The notable results presented in tables reveal the followings:

1. Both methods are in complete agreement with regards to each dataset distribution similarity assessment.
2. Polygon 32's dataset has a similar data distribution as that of the entire ERSPD dataset, except for peak total rainfall and peak rainfall intensity which differ significantly.
3. Polygon 32's calibration dataset has a similar data distribution as that of the entire polygon 32's dataset.
4. Polygon 32's validation dataset has a similar data distribution as that of the polygon 32's calibration dataset.

Table C.1 - Comparison of the distributions of entire ERSPD dataset against polygon 32's dataset

	K	$K \geq X^2$?	Test Conclusion	p	$p < \alpha$?	p Conclusion
Average Total Rainfall	1.2468	No	H_0 not rejected	0.2642	No	H_0 not rejected
Peak Total Rainfall	5.05	Yes	Reject H_0	0.0246	Yes	Reject H_0
Average Rainfall Intensity	1.3879	No	H_0 not rejected	0.2388	No	H_0 not rejected
Peak Rainfall Intensity	4.9661	Yes	Reject H_0	0.0259	Yes	Reject H_0
Average Rainfall Duration	0.877	No	H_0 not rejected	0.349	No	H_0 not rejected
Reported Basement Floodings	1.081	No	H_0 not rejected	0.2985	No	H_0 not rejected

Table C.2 - Comparison of the distributions of polygon 32's dataset against polygon 32's calibration dataset

	K	$K \geq X^2$?	Test Conclusion	p	$p < \alpha$?	p Conclusion
Average total rainfall	0.1174	No	H_0 not rejected	0.7318	No	H_0 not rejected
Peak total rainfall	4.00E-04	No	H_0 not rejected	0.9834	No	H_0 not rejected
Average rainfall intensity	0.0092	No	H_0 not rejected	0.9237	No	H_0 not rejected
Peak rainfall intensity	0.0161	No	H_0 not rejected	0.8992	No	H_0 not rejected
Average rainfall duration	0.2856	No	H_0 not rejected	0.5931	No	H_0 not rejected
Reported basement floodings	0.0362	No	H_0 not rejected	0.8492	No	H_0 not rejected

Table C.3 - Comparison of the distributions of polygon 32's calibration dataset against polygon 32's validation dataset

	K	$K \geq X^2$?	Test Conclusion	p	$p < \alpha$?	p Conclusion
Average total rainfall	1.0468	No	H_0 not rejected	0.3062	No	H_0 not rejected
Peak total rainfall	0.0038	No	H_0 not rejected	0.9506	No	H_0 not rejected
Average rainfall intensity	0.0818	No	H_0 not rejected	0.7749	No	H_0 not rejected
Peak rainfall intensity	0.1431	No	H_0 not rejected	0.7052	No	H_0 not rejected
Average rainfall duration	2.5458	No	H_0 not rejected	0.1106	No	H_0 not rejected
Reported basement floodings	0.3075	No	H_0 not rejected	0.5792	No	H_0 not rejected

The R code used to compute the Kruskal-Wallis chi-squared test statistics results follow.

```
#1-alpha quantile of a chi-square distribution with (k-1) degrees of freedom where alpha = 0.05 and k = 2
qchisq(0.95,df=1)

#Assigning the entire ERSPD dataset table to a temporary matrix
tempmatrixAll <- read.csv("C:/R/Thesis/Final Datasets Distribution Tests/Entire ERSPD dataset.csv", header = TRUE, sep =
        ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
AllAvgTotal <- tempmatrixAll[1:length(tempmatrixAll[,5]),5]
AllPeakTotal <- tempmatrixAll[1:length(tempmatrixAll[,6]),6]
AllAvgIntensity <- tempmatrixAll[1:length(tempmatrixAll[,7]),7]
AllPeakIntensity <- tempmatrixAll[1:length(tempmatrixAll[,8]),8]
AllAvgDuration <- tempmatrixAll[1:length(tempmatrixAll[,9]),9]
AllFloodings <- tempmatrixAll[1:length(tempmatrixAll[,10]),10]


#Assigning the entire polygon 32 dataset table to a temporary matrix
tempmatrix32 <- read.csv("C:/R/Thesis/Final Datasets Distribution Tests/Polygon 32.csv", header = TRUE, sep = ",",
        quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
P32AvgTotal <- tempmatrix32[1:length(tempmatrix32[,5]),5]
P32PeakTotal <- tempmatrix32[1:length(tempmatrix32[,6]),6]
P32AvgIntensity <- tempmatrix32[1:length(tempmatrix32[,7]),7]
P32PeakIntensity <- tempmatrix32[1:length(tempmatrix32[,8]),8]
P32AvgDuration <- tempmatrix32[1:length(tempmatrix32[,9]),9]
P32Floodings <- tempmatrix32[1:length(tempmatrix32[,10]),10]


#Assigning Polygon 32 Calibration set table to a temporary matrix
tempmatrixC32 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Calibration Set/Calibration 32.csv",
        header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
CAvgTotal <- tempmatrixC32[1:length(tempmatrixC32[,5]),5]
CPeakTotal <- tempmatrixC32[1:length(tempmatrixC32[,6]),6]
CAvgIntensity <- tempmatrixC32[1:length(tempmatrixC32[,7]),7]
CPeakIntensity <- tempmatrixC32[1:length(tempmatrixC32[,8]),8]
CAvgDuration <- tempmatrixC32[1:length(tempmatrixC32[,9]),9]
CFloodings <- tempmatrixC32[1:length(tempmatrixC32[,10]),10]

#Assigning Polygon 32 Validation set table to a temporary matrix
tempmatrixV32 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Validation Set/Validation 32.csv",
        header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Validation set Database variable to its corresponding array
VAvgTotal <- tempmatrixV32[1:length(tempmatrixV32[,5]),5]
VPeakTotal <- tempmatrixV32[1:length(tempmatrixV32[,6]),6]
VAvgIntensity <- tempmatrixV32[1:length(tempmatrixV32[,7]),7]
VPeakIntensity <- tempmatrixV32[1:length(tempmatrixV32[,8]),8]
VAvgDuration <- tempmatrixV32[1:length(tempmatrixV32[,9]),9]
VFloodings <- tempmatrixV32[1:length(tempmatrixV32[,10]),10]


#Kruskal-Wallis rank sum test them all

#Testing entire ERSPD dataset compared to polygon 32's dataset
kruskal.test(list(AllAvgTotal,P32AvgTotal))
kruskal.test(list(AllPeakTotal,P32PeakTotal))
kruskal.test(list(AllAvgIntensity,P32AvgIntensity))
kruskal.test(list(AllPeakIntensity,P32PeakIntensity))
kruskal.test(list(AllAvgDuration,P32AvgDuration))
kruskal.test(list(AllFloodings,P32Floodings))


#Testing polygon 32's entire dataset compared against polygon 32's calibration dataset
kruskal.test(list(P32AvgTotal,CAvgTotal))
kruskal.test(list(P32PeakTotal,CPeakTotal))
kruskal.test(list(P32AvgIntensity,CAvgIntensity))
```

```
kruskal.test(list(P32PeakIntensity,CPeakIntensity))
kruskal.test(list(P32AvgDuration,CAvgDuration))
kruskal.test(list(P32Floodings,CFloodings))


#Testing polygon 32's calibration dataset versus polygon 32's validation dataset
kruskal.test(list(VAvgTotal,CAvgTotal))
kruskal.test(list(VPeakTotal,CPeakTotal))
kruskal.test(list(VAvgIntensity,CAvgIntensity))
kruskal.test(list(VPeakIntensity,CPeakIntensity))
kruskal.test(list(VAvgDuration,CAvgDuration))
kruskal.test(list(VFloodings,CFloodings))
```

Appendix D - Example of R Code

```
#Assigning Polygon 17 Calibration set table to a temporary matrix
tempmatrixC17 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Calibration Set/Calibration 17.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixC17 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Calibration Set/Calibration 17.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
C17AvgTotal <- tempmatrixC17[1:length(tempmatrixC17[,5]),5]
C17PeakTotal <- tempmatrixC17[1:length(tempmatrixC17[,6]),6]
C17AvgIntensity <- tempmatrixC17[1:length(tempmatrixC17[,7]),7]
C17PeakIntensity <- tempmatrixC17[1:length(tempmatrixC17[,8]),8]
C17AvgDuration <- tempmatrixC17[1:length(tempmatrixC17[,9]),9]
C17Floodings <- tempmatrixC17[1:length(tempmatrixC17[,10]),10]

#Assigning Polygon 17 Validation set table to a temporary matrix
tempmatrixV17 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Validation Set/Validation 17.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixV17 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Validation Set/Validation 17.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Validation set Database variable to its corresponding array
V17AvgTotal <- tempmatrixV17[1:length(tempmatrixV17[,5]),5]
V17PeakTotal <- tempmatrixV17[1:length(tempmatrixV17[,6]),6]
V17AvgIntensity <- tempmatrixV17[1:length(tempmatrixV17[,7]),7]
V17PeakIntensity <- tempmatrixV17[1:length(tempmatrixV17[,8]),8]
V17AvgDuration <- tempmatrixV17[1:length(tempmatrixV17[,9]),9]
V17Floodings <- tempmatrixV17[1:length(tempmatrixV17[,10]),10]


#Assigning Polygon 20 Calibration set table to a temporary matrix
tempmatrixC20 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Calibration Set/Calibration 20.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixC20 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Calibration Set/Calibration 20.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
C20AvgTotal <- tempmatrixC20[1:length(tempmatrixC20[,5]),5]
C20PeakTotal <- tempmatrixC20[1:length(tempmatrixC20[,6]),6]
C20AvgIntensity <- tempmatrixC20[1:length(tempmatrixC20[,7]),7]
C20PeakIntensity <- tempmatrixC20[1:length(tempmatrixC20[,8]),8]
C20AvgDuration <- tempmatrixC20[1:length(tempmatrixC20[,9]),9]
C20Floodings <- tempmatrixC20[1:length(tempmatrixC20[,10]),10]

#Assigning Polygon 20 Validation set table to a temporary matrix
tempmatrixV20 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Validation Set/Validation 20.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixV20 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Validation Set/Validation 20.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Validation set Database variable to its corresponding array
V20AvgTotal <- tempmatrixV20[1:length(tempmatrixV20[,5]),5]
V20PeakTotal <- tempmatrixV20[1:length(tempmatrixV20[,6]),6]
V20AvgIntensity <- tempmatrixV20[1:length(tempmatrixV20[,7]),7]
V20PeakIntensity <- tempmatrixV20[1:length(tempmatrixV20[,8]),8]
V20AvgDuration <- tempmatrixV20[1:length(tempmatrixV20[,9]),9]
V20Floodings <- tempmatrixV20[1:length(tempmatrixV20[,10]),10]


#Assigning Polygon 21 Calibration set table to a temporary matrix
tempmatrixC21 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Calibration Set/Calibration 21.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixC21 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Calibration Set/Calibration 21.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
C21AvgTotal <- tempmatrixC21[1:length(tempmatrixC21[,5]),5]
C21PeakTotal <- tempmatrixC21[1:length(tempmatrixC21[,6]),6]
C21AvgIntensity <- tempmatrixC21[1:length(tempmatrixC21[,7]),7]
C21PeakIntensity <- tempmatrixC21[1:length(tempmatrixC21[,8]),8]
C21AvgDuration <- tempmatrixC21[1:length(tempmatrixC21[,9]),9]
```

```
C21Floodings <- tempmatrixC21[1:length(tempmatrixC21[,10]),10]

#Assigning Polygon 21 Validation set table to a temporary matrix
tempmatrixV21 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Validation Set/Validation 21.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixV21 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Validation Set/Validation 21.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Validation set Database variable to its corresponding array
V21AvgTotal <- tempmatrixV21[1:length(tempmatrixV21[,5]),5]
V21PeakTotal <- tempmatrixV21[1:length(tempmatrixV21[,6]),6]
V21AvgIntensity <- tempmatrixV21[1:length(tempmatrixV21[,7]),7]
V21PeakIntensity <- tempmatrixV21[1:length(tempmatrixV21[,8]),8]
V21AvgDuration <- tempmatrixV21[1:length(tempmatrixV21[,9]),9]
V21Floodings <- tempmatrixV21[1:length(tempmatrixV21[,10]),10]


#Assigning Polygon 26 Calibration set table to a temporary matrix
tempmatrixC26 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Calibration Set/Calibration 26.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixC26 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Calibration Set/Calibration 26.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
C26AvgTotal <- tempmatrixC26[1:length(tempmatrixC26[,5]),5]
C26PeakTotal <- tempmatrixC26[1:length(tempmatrixC26[,6]),6]
C26AvgIntensity <- tempmatrixC26[1:length(tempmatrixC26[,7]),7]
C26PeakIntensity <- tempmatrixC26[1:length(tempmatrixC26[,8]),8]
C26AvgDuration <- tempmatrixC26[1:length(tempmatrixC26[,9]),9]
C26Floodings <- tempmatrixC26[1:length(tempmatrixC26[,10]),10]

#Assigning Polygon 26 Validation set table to a temporary matrix
tempmatrixV26 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Validation Set/Validation 26.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixV26 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Validation Set/Validation 26.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Validation set Database variable to its corresponding array
V26AvgTotal <- tempmatrixV26[1:length(tempmatrixV26[,5]),5]
V26PeakTotal <- tempmatrixV26[1:length(tempmatrixV26[,6]),6]
V26AvgIntensity <- tempmatrixV26[1:length(tempmatrixV26[,7]),7]
V26PeakIntensity <- tempmatrixV26[1:length(tempmatrixV26[,8]),8]
V26AvgDuration <- tempmatrixV26[1:length(tempmatrixV26[,9]),9]
V26Floodings <- tempmatrixV26[1:length(tempmatrixV26[,10]),10]


#Assigning Polygon 27 Calibration set table to a temporary matrix
tempmatrixC27 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Calibration Set/Calibration 27.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixC27 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Calibration Set/Calibration 27.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
C27AvgTotal <- tempmatrixC27[1:length(tempmatrixC27[,5]),5]
C27PeakTotal <- tempmatrixC27[1:length(tempmatrixC27[,6]),6]
C27AvgIntensity <- tempmatrixC27[1:length(tempmatrixC27[,7]),7]
C27PeakIntensity <- tempmatrixC27[1:length(tempmatrixC27[,8]),8]
C27AvgDuration <- tempmatrixC27[1:length(tempmatrixC27[,9]),9]
C27Floodings <- tempmatrixC27[1:length(tempmatrixC27[,10]),10]

#Assigning Polygon 27 Validation set table to a temporary matrix
tempmatrixV27 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Validation Set/Validation 27.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixV27 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Validation Set/Validation 27.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Validation set Database variable to its corresponding array
V27AvgTotal <- tempmatrixV27[1:length(tempmatrixV27[,5]),5]
V27PeakTotal <- tempmatrixV27[1:length(tempmatrixV27[,6]),6]
V27AvgIntensity <- tempmatrixV27[1:length(tempmatrixV27[,7]),7]
```

```
V27PeakIntensity <- tempmatrixV27[1:length(tempmatrixV27[,8]),8]
V27AvgDuration <- tempmatrixV27[1:length(tempmatrixV27[,9]),9]
V27Floodings <- tempmatrixV27[1:length(tempmatrixV27[,10]),10]


#Assigning Polygon 28 Calibration set table to a temporary matrix
tempmatrixC28 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Calibration Set/Calibration 28.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixC28 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Calibration Set/Calibration 28.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
C28AvgTotal <- tempmatrixC28[1:length(tempmatrixC28[,5]),5]
C28PeakTotal <- tempmatrixC28[1:length(tempmatrixC28[,6]),6]
C28AvgIntensity <- tempmatrixC28[1:length(tempmatrixC28[,7]),7]
C28PeakIntensity <- tempmatrixC28[1:length(tempmatrixC28[,8]),8]
C28AvgDuration <- tempmatrixC28[1:length(tempmatrixC28[,9]),9]
C28Floodings <- tempmatrixC28[1:length(tempmatrixC28[,10]),10]

#Assigning Polygon 28 Validation set table to a temporary matrix
tempmatrixV28 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Validation Set/Validation 28.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixV28 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Validation Set/Validation 28.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Validation set Database variable to its corresponding array
V28AvgTotal <- tempmatrixV28[1:length(tempmatrixV28[,5]),5]
V28PeakTotal <- tempmatrixV28[1:length(tempmatrixV28[,6]),6]
V28AvgIntensity <- tempmatrixV28[1:length(tempmatrixV28[,7]),7]
V28PeakIntensity <- tempmatrixV28[1:length(tempmatrixV28[,8]),8]
V28AvgDuration <- tempmatrixV28[1:length(tempmatrixV28[,9]),9]
V28Floodings <- tempmatrixV28[1:length(tempmatrixV28[,10]),10]


#Assigning Polygon 29&55 Calibration set table to a temporary matrix
tempmatrixC2955 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Calibration Set/Calibration 29&55.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixC2955 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Calibration Set/Calibration 29&55.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
C2955AvgTotal <- tempmatrixC2955[1:length(tempmatrixC2955[,5]),5]
C2955PeakTotal <- tempmatrixC2955[1:length(tempmatrixC2955[,6]),6]
C2955AvgIntensity <- tempmatrixC2955[1:length(tempmatrixC2955[,7]),7]
C2955PeakIntensity <- tempmatrixC2955[1:length(tempmatrixC2955[,8]),8]
C2955AvgDuration <- tempmatrixC2955[1:length(tempmatrixC2955[,9]),9]
C2955Floodings <- tempmatrixC2955[1:length(tempmatrixC2955[,10]),10]

#Assigning Polygon 29&55 Validation set table to a temporary matrix
tempmatrixV2955 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Validation Set/Validation 29&55.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixV2955 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Validation Set/Validation 29&55.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Validation set Database variable to its corresponding array
V2955AvgTotal <- tempmatrixV2955[1:length(tempmatrixV2955[,5]),5]
V2955PeakTotal <- tempmatrixV2955[1:length(tempmatrixV2955[,6]),6]
V2955AvgIntensity <- tempmatrixV2955[1:length(tempmatrixV2955[,7]),7]
V2955PeakIntensity <- tempmatrixV2955[1:length(tempmatrixV2955[,8]),8]
V2955AvgDuration <- tempmatrixV2955[1:length(tempmatrixV2955[,9]),9]
V2955Floodings <- tempmatrixV2955[1:length(tempmatrixV2955[,10]),10]


#Assigning Polygon 30 Calibration set table to a temporary matrix
tempmatrixC30 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Calibration Set/Calibration 30.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixC30 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Calibration Set/Calibration 30.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
```

```
C30AvgTotal <- tempmatrixC30[1:length(tempmatrixC30[,5]),5]
C30PeakTotal <- tempmatrixC30[1:length(tempmatrixC30[,6]),6]
C30AvgIntensity <- tempmatrixC30[1:length(tempmatrixC30[,7]),7]
C30PeakIntensity <- tempmatrixC30[1:length(tempmatrixC30[,8]),8]
C30AvgDuration <- tempmatrixC30[1:length(tempmatrixC30[,9]),9]
C30Floodings <- tempmatrixC30[1:length(tempmatrixC30[,10]),10]

#Assigning Polygon 30 Validation set table to a temporary matrix
tempmatrixV30 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Validation Set/Validation 30.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixV30 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Validation Set/Validation 30.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Validation set Database variable to its corresponding array
V30AvgTotal <- tempmatrixV30[1:length(tempmatrixV30[,5]),5]
V30PeakTotal <- tempmatrixV30[1:length(tempmatrixV30[,6]),6]
V30AvgIntensity <- tempmatrixV30[1:length(tempmatrixV30[,7]),7]
V30PeakIntensity <- tempmatrixV30[1:length(tempmatrixV30[,8]),8]
V30AvgDuration <- tempmatrixV30[1:length(tempmatrixV30[,9]),9]
V30Floodings <- tempmatrixV30[1:length(tempmatrixV30[,10]),10]


#Assigning Polygon 32 Calibration set table to a temporary matrix
tempmatrixC32 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Calibration Set/Calibration 32.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixC32 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Calibration Set/Calibration 32.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
CAvgTotal <- tempmatrixC32[1:length(tempmatrixC32[,5]),5]
CPeakTotal <- tempmatrixC32[1:length(tempmatrixC32[,6]),6]
CAvgIntensity <- tempmatrixC32[1:length(tempmatrixC32[,7]),7]
CPeakIntensity <- tempmatrixC32[1:length(tempmatrixC32[,8]),8]
CAvgDuration <- tempmatrixC32[1:length(tempmatrixC32[,9]),9]
CFloodings <- tempmatrixC32[1:length(tempmatrixC32[,10]),10]

#Assigning Polygon 32 Validation set table to a temporary matrix
tempmatrixV32 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Validation Set/Validation 32.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixV32 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Validation Set/Validation 32.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Validation set Database variable to its corresponding array
VAvgTotal <- tempmatrixV32[1:length(tempmatrixV32[,5]),5]
VPeakTotal <- tempmatrixV32[1:length(tempmatrixV32[,6]),6]
VAvgIntensity <- tempmatrixV32[1:length(tempmatrixV32[,7]),7]
VPeakIntensity <- tempmatrixV32[1:length(tempmatrixV32[,8]),8]
VAvgDuration <- tempmatrixV32[1:length(tempmatrixV32[,9]),9]
VFloodings <- tempmatrixV32[1:length(tempmatrixV32[,10]),10]


#Assigning Polygon 33 Calibration set table to a temporary matrix
tempmatrixC33 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Calibration Set/Calibration 33.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixC33 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Calibration Set/Calibration 33.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
C33AvgTotal <- tempmatrixC33[1:length(tempmatrixC33[,5]),5]
C33PeakTotal <- tempmatrixC33[1:length(tempmatrixC33[,6]),6]
C33AvgIntensity <- tempmatrixC33[1:length(tempmatrixC33[,7]),7]
C33PeakIntensity <- tempmatrixC33[1:length(tempmatrixC33[,8]),8]
C33AvgDuration <- tempmatrixC33[1:length(tempmatrixC33[,9]),9]
C33Floodings <- tempmatrixC33[1:length(tempmatrixC33[,10]),10]

#Assigning Polygon 33 Validation set table to a temporary matrix
tempmatrixV33 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Validation Set/Validation 33.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixV33 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Validation Set/Validation 33.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
```

```
#Assigning each Validation set Database variable to its corresponding array
V33AvgTotal <- tempmatrixV33[1:length(tempmatrixV33[,5]),5]
V33PeakTotal <- tempmatrixV33[1:length(tempmatrixV33[,6]),6]
V33AvgIntensity <- tempmatrixV33[1:length(tempmatrixV33[,7]),7]
V33PeakIntensity <- tempmatrixV33[1:length(tempmatrixV33[,8]),8]
V33AvgDuration <- tempmatrixV33[1:length(tempmatrixV33[,9]),9]
V33Floodings <- tempmatrixV33[1:length(tempmatrixV33[,10]),10]


#Assigning Polygon 37 Calibration set table to a temporary matrix
tempmatrixC37 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Calibration Set/Calibration 37.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixC37 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Calibration Set/Calibration 37.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
C37AvgTotal <- tempmatrixC37[1:length(tempmatrixC37[,5]),5]
C37PeakTotal <- tempmatrixC37[1:length(tempmatrixC37[,6]),6]
C37AvgIntensity <- tempmatrixC37[1:length(tempmatrixC37[,7]),7]
C37PeakIntensity <- tempmatrixC37[1:length(tempmatrixC37[,8]),8]
C37AvgDuration <- tempmatrixC37[1:length(tempmatrixC37[,9]),9]
C37Floodings <- tempmatrixC37[1:length(tempmatrixC37[,10]),10]

#Assigning Polygon 37 Validation set table to a temporary matrix
tempmatrixV37 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Validation Set/Validation 37.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixV37 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Validation Set/Validation 37.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Validation set Database variable to its corresponding array
V37AvgTotal <- tempmatrixV37[1:length(tempmatrixV37[,5]),5]
V37PeakTotal <- tempmatrixV37[1:length(tempmatrixV37[,6]),6]
V37AvgIntensity <- tempmatrixV37[1:length(tempmatrixV37[,7]),7]
V37PeakIntensity <- tempmatrixV37[1:length(tempmatrixV37[,8]),8]
V37AvgDuration <- tempmatrixV37[1:length(tempmatrixV37[,9]),9]
V37Floodings <- tempmatrixV37[1:length(tempmatrixV37[,10]),10]


#Assigning Polygon 38 Calibration set table to a temporary matrix
tempmatrixC38 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Calibration Set/Calibration 38.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixC38 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Calibration Set/Calibration 38.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
C38AvgTotal <- tempmatrixC38[1:length(tempmatrixC38[,5]),5]
C38PeakTotal <- tempmatrixC38[1:length(tempmatrixC38[,6]),6]
C38AvgIntensity <- tempmatrixC38[1:length(tempmatrixC38[,7]),7]
C38PeakIntensity <- tempmatrixC38[1:length(tempmatrixC38[,8]),8]
C38AvgDuration <- tempmatrixC38[1:length(tempmatrixC38[,9]),9]
C38Floodings <- tempmatrixC38[1:length(tempmatrixC38[,10]),10]

#Assigning Polygon 38 Validation set table to a temporary matrix
tempmatrixV38 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Validation Set/Validation 38.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixV38 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Validation Set/Validation 38.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Validation set Database variable to its corresponding array
V38AvgTotal <- tempmatrixV38[1:length(tempmatrixV38[,5]),5]
V38PeakTotal <- tempmatrixV38[1:length(tempmatrixV38[,6]),6]
V38AvgIntensity <- tempmatrixV38[1:length(tempmatrixV38[,7]),7]
V38PeakIntensity <- tempmatrixV38[1:length(tempmatrixV38[,8]),8]
V38AvgDuration <- tempmatrixV38[1:length(tempmatrixV38[,9]),9]
V38Floodings <- tempmatrixV38[1:length(tempmatrixV38[,10]),10]


#Assigning Polygon 42 Calibration set table to a temporary matrix
```

```
tempmatrixC42 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Calibration Set/Calibration 42.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixC42 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Calibration Set/Calibration 42.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
C42AvgTotal <- tempmatrixC42[1:length(tempmatrixC42[,5]),5]
C42PeakTotal <- tempmatrixC42[1:length(tempmatrixC42[,6]),6]
C42AvgIntensity <- tempmatrixC42[1:length(tempmatrixC42[,7]),7]
C42PeakIntensity <- tempmatrixC42[1:length(tempmatrixC42[,8]),8]
C42AvgDuration <- tempmatrixC42[1:length(tempmatrixC42[,9]),9]
C42Floodings <- tempmatrixC42[1:length(tempmatrixC42[,10]),10]

#Assigning Polygon 42 Validation set table to a temporary matrix
tempmatrixV42 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Validation Set/Validation 42.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixV42 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Validation Set/Validation 42.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Validation set Database variable to its corresponding array
V42AvgTotal <- tempmatrixV42[1:length(tempmatrixV42[,5]),5]
V42PeakTotal <- tempmatrixV42[1:length(tempmatrixV42[,6]),6]
V42AvgIntensity <- tempmatrixV42[1:length(tempmatrixV42[,7]),7]
V42PeakIntensity <- tempmatrixV42[1:length(tempmatrixV42[,8]),8]
V42AvgDuration <- tempmatrixV42[1:length(tempmatrixV42[,9]),9]
V42Floodings <- tempmatrixV42[1:length(tempmatrixV42[,10]),10]


#Assigning Polygon 48 Calibration set table to a temporary matrix
tempmatrixC48 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Calibration Set/Calibration 48.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixC48 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Calibration Set/Calibration 48.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
C48AvgTotal <- tempmatrixC48[1:length(tempmatrixC48[,5]),5]
C48PeakTotal <- tempmatrixC48[1:length(tempmatrixC48[,6]),6]
C48AvgIntensity <- tempmatrixC48[1:length(tempmatrixC48[,7]),7]
C48PeakIntensity <- tempmatrixC48[1:length(tempmatrixC48[,8]),8]
C48AvgDuration <- tempmatrixC48[1:length(tempmatrixC48[,9]),9]
C48Floodings <- tempmatrixC48[1:length(tempmatrixC48[,10]),10]

#Assigning Polygon 48 Validation set table to a temporary matrix
tempmatrixV48 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Validation Set/Validation 48.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixV48 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Validation Set/Validation 48.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Validation set Database variable to its corresponding array
V48AvgTotal <- tempmatrixV48[1:length(tempmatrixV48[,5]),5]
V48PeakTotal <- tempmatrixV48[1:length(tempmatrixV48[,6]),6]
V48AvgIntensity <- tempmatrixV48[1:length(tempmatrixV48[,7]),7]
V48PeakIntensity <- tempmatrixV48[1:length(tempmatrixV48[,8]),8]
V48AvgDuration <- tempmatrixV48[1:length(tempmatrixV48[,9]),9]
V48Floodings <- tempmatrixV48[1:length(tempmatrixV48[,10]),10]


#Assigning Polygon 49 Calibration set table to a temporary matrix
tempmatrixC49 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Calibration Set/Calibration 49.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixC49 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Calibration Set/Calibration 49.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
C49AvgTotal <- tempmatrixC49[1:length(tempmatrixC49[,5]),5]
C49PeakTotal <- tempmatrixC49[1:length(tempmatrixC49[,6]),6]
C49AvgIntensity <- tempmatrixC49[1:length(tempmatrixC49[,7]),7]
C49PeakIntensity <- tempmatrixC49[1:length(tempmatrixC49[,8]),8]
C49AvgDuration <- tempmatrixC49[1:length(tempmatrixC49[,9]),9]
C49Floodings <- tempmatrixC49[1:length(tempmatrixC49[,10]),10]
```

```
#Assigning Polygon 49 Validation set table to a temporary matrix
tempmatrixV49 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Validation Set/Validation 49.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixV49 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Validation Set/Validation 49.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Validation set Database variable to its corresponding array
V49AvgTotal <- tempmatrixV49[1:length(tempmatrixV49[,5]),5]
V49PeakTotal <- tempmatrixV49[1:length(tempmatrixV49[,6]),6]
V49AvgIntensity <- tempmatrixV49[1:length(tempmatrixV49[,7]),7]
V49PeakIntensity <- tempmatrixV49[1:length(tempmatrixV49[,8]),8]
V49AvgDuration <- tempmatrixV49[1:length(tempmatrixV49[,9]),9]
V49Floodings <- tempmatrixV49[1:length(tempmatrixV49[,10]),10]


#Assigning Polygon 57 Calibration set table to a temporary matrix
tempmatrixC57 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Calibration Set/Calibration 57.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixC57 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Calibration Set/Calibration 57.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Calibration set Database variable to its corresponding array
C57AvgTotal <- tempmatrixC57[1:length(tempmatrixC57[,5]),5]
C57PeakTotal <- tempmatrixC57[1:length(tempmatrixC57[,6]),6]
C57AvgIntensity <- tempmatrixC57[1:length(tempmatrixC57[,7]),7]
C57PeakIntensity <- tempmatrixC57[1:length(tempmatrixC57[,8]),8]
C57AvgDuration <- tempmatrixC57[1:length(tempmatrixC57[,9]),9]
C57Floodings <- tempmatrixC57[1:length(tempmatrixC57[,10]),10]

#Assigning Polygon 57 Validation set table to a temporary matrix
tempmatrixV57 <- read.csv("D:/Documents/Masters/Thesis/New Radar Data - Polygons/Validation Set/Validation 57.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")
#tempmatrixV57 <- read.csv("C:/Erik/Edmonton R&D/Thesis - 2012Nov1/Validation Set/Validation 57.csv", header = TRUE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#Assigning each Validation set Database variable to its corresponding array
V57AvgTotal <- tempmatrixV57[1:length(tempmatrixV57[,5]),5]
V57PeakTotal <- tempmatrixV57[1:length(tempmatrixV57[,6]),6]
V57AvgIntensity <- tempmatrixV57[1:length(tempmatrixV57[,7]),7]
V57PeakIntensity <- tempmatrixV57[1:length(tempmatrixV57[,8]),8]
V57AvgDuration <- tempmatrixV57[1:length(tempmatrixV57[,9]),9]
V57Floodings <- tempmatrixV57[1:length(tempmatrixV57[,10]),10]

#=====================================================================

Variables <- c("CAvgTotal","CPeakTotal","CAvgIntensity","CPeakIntensity","CAvgDuration")

#Set empty table for Top 100 best equations
Top100RankTable <- matrix(1,100,2)

#Set table's header
#Use a weighed rank in ascending order as the measure of each particular equation's success
colnames(Top100RankTable) <- c("Equations","% False Positives")

#equation number for which the computation is made
equnumber <- 1

#All possible combinations of 5 terms without repetitions = 120
Combinations <- read.csv("C:/R/Thesis/Combinations.csv", header = FALSE, sep = ",", quote="\"", dec=".", fill = TRUE, comment.char="")

#coefficients used to vary weight of variables within I()
coeffs <- c(-500,-250,-100,-50,-25,-10,-5,-4,-3,-2,-1,0,1,2,3,4,5,10,25,50,100,250,500)

#Start of bigger overall loop
#Start of computation of new equation
```

```
#length(Variables) = 5
for (i10 in 0:length(Variables))
{
	#This level is ^5

	#i10 refers to how many terms are in the parenthese

	#length(Combinations[,1]) = 120
	for(j10 in 1:length(Combinations[,1]))
	{

	   #Creates Formulas according to the number of parentheses
	   switch(i10+1,
		Formulas <-
paste(Variables[Combinations[j10,1]],Variables[Combinations[j10,2]],Variables[Combinations[j10,3]],Variables[Combin
ations[j10,4]],Variables[Combinations[j10,5]], sep = "+"),
		Formulas <-
paste(paste("I((",Variables[Combinations[j10,1]],")^5)"),Variables[Combinations[j10,2]],Variables[Combinations[j10,3]],V
ariables[Combinations[j10,4]],Variables[Combinations[j10,5]], sep = "+"),
		Formulas <- paste(paste("I((",paste(Variables[Combinations[j10,1]],Variables[Combinations[j10,2]], sep =
"+"),")^5)"),Variables[Combinations[j10,3]],Variables[Combinations[j10,4]],Variables[Combinations[j10,5]], sep = "+"),
		Formulas <-
paste(paste("I((",paste(Variables[Combinations[j10,1]],Variables[Combinations[j10,2]],Variables[Combinations[j10,3]],
sep = "+"),")^5)"),Variables[Combinations[j10,4]],Variables[Combinations[j10,5]], sep = "+"),
		Formulas <-
paste(paste("I((",paste(Variables[Combinations[j10,1]],Variables[Combinations[j10,2]],Variables[Combinations[j10,3]],V
ariables[Combinations[j10,4]], sep = "+"),")^5)"),Variables[Combinations[j10,5]], sep = "+"),
		Formulas <-
paste("I((",paste(Variables[Combinations[j10,1]],Variables[Combinations[j10,2]],Variables[Combinations[j10,3]],Variable
s[Combinations[j10,4]],Variables[Combinations[j10,5]], sep = "+"),")^5)"),
		)

	   #Here coefficients are computed for variables within I()
	   if(i10>1)
	   {
		#Produce equations & sMLR
		for(k10 in 1:length(coeffs))
		{
			for(l10 in 1:length(coeffs))
			{
				#If there are only 2 variables within the parentheses...
				if(i10==2)
				{
					#partly inspired by post at https://stat.ethz.ch/pipermail/r-help/2008-
October/175554.html
					#Character Locations of opening and closing parenthese in string
					OpenParLoc <- which(strsplit(Formulas, '')[[1]]=='(')
					CloseParLoc <- which(strsplit(Formulas, '')[[1]]==')')
					PlusLoc <- which(strsplit(Formulas, '')[[1]]=='+')

					#Cuts appart the Formulas equation and inserts coeffs as per k10 and l10
for loops
					TempFormula <-
paste(substr(Formulas,1,OpenParLoc[length(OpenParLoc)]),coeffs[k10],"*",substr(Formulas,OpenParLoc[length(Open
ParLoc)]+1,PlusLoc[1]),coeffs[l10],"*",substr(Formulas,PlusLoc[1]+1,nchar(Formulas)))


					#Perform stepwise regression with formulas
					sMLR <- step(lm(as.formula(paste("CFloodings~",TempFormula))))

					#extract the stepwise regression equation
					sMLRtempEqu <-
substr(toString(formula(sMLR)),tail(which(strsplit(toString(formula(sMLR)),
')[[1]]==','),1)+2,nchar(toString(formula(sMLR))))

					#extract the stepwise regression coefficients
					sMLRcoeffs <- coefficients(sMLR)
```

```
#Finding the spaces in the equation which should be before and after each
operation and place it into an array
SpaceLoc <- unlist(gregexpr(" ",sMLRtempEqu),1)

switch(length(coefficients(sMLR)),
        sMLRtempEquNCoef <- sMLRcoeffs[1],
        sMLRtempEquNCoef <-
paste(sMLRcoeffs[1],"+",sMLRcoeffs[2],"*",sMLRtempEqu),
        sMLRtempEquNCoef <-
paste(sMLRcoeffs[1],"+",sMLRcoeffs[2],"*",substr(sMLRtempEqu,1,tail(SpaceLoc,1)-
1),sMLRcoeffs[3],"*",substr(sMLRtempEqu,tail(SpaceLoc,1),nchar(sMLRtempEqu))),
        sMLRtempEquNCoef <-
paste(sMLRcoeffs[1],"+",sMLRcoeffs[2],"*",substr(sMLRtempEqu,1,SpaceLoc[length(SpaceLoc)-2]-
1),sMLRcoeffs[3],"*",substr(sMLRtempEqu,SpaceLoc[length(SpaceLoc)-2]+1,tail(SpaceLoc,1)-
1),sMLRcoeffs[4],"*",substr(sMLRtempEqu,tail(SpaceLoc,1),nchar(sMLRtempEqu))),
        sMLRtempEquNCoef <-
paste(sMLRcoeffs[1],"+",sMLRcoeffs[2],"*",substr(sMLRtempEqu,1,SpaceLoc[length(SpaceLoc)-4]-
1),sMLRcoeffs[3],"*",substr(sMLRtempEqu,SpaceLoc[length(SpaceLoc)-4]+1,SpaceLoc[length(SpaceLoc)-2]-
1),sMLRcoeffs[4],"*",substr(sMLRtempEqu,SpaceLoc[length(SpaceLoc)-2]+1,tail(SpaceLoc,1)-
1),sMLRcoeffs[5],"*",substr(sMLRtempEqu,tail(SpaceLoc,1),nchar(sMLRtempEqu))),
        )


#Assessing the Calibration and Validation fit for all polygons
#Compute Test Rank for Polygon 17 Calibration Set
C17Model <- eval(parse(text = gsub("C","C17",sMLRtempEquNCoef)))
RankofC17Model <- rank(C17Model)

#Compute Test Rank for Polygon 17 Validation Set
V17Model <- eval(parse(text = gsub("C","V17",sMLRtempEquNCoef)))
RankofV17Model <- rank(V17Model)

#Compute Test Rank for Polygon 20 Calibration Set
C20Model <- eval(parse(text = gsub("C","C20",sMLRtempEquNCoef)))
RankofC20Model <- rank(C20Model)

#Compute Test Rank for Polygon 20 Validation Set
V20Model <- eval(parse(text = gsub("C","V20",sMLRtempEquNCoef)))
RankofV20Model <- rank(V20Model)

#Compute Test Rank for Polygon 21 Calibration Set
C21Model <- eval(parse(text = gsub("C","C21",sMLRtempEquNCoef)))
RankofC21Model <- rank(C21Model)

#Compute Test Rank for Polygon 21 Validation Set
V21Model <- eval(parse(text = gsub("C","V21",sMLRtempEquNCoef)))
RankofV21Model <- rank(V21Model)

#Compute Test Rank for Polygon 26 Calibration Set
C26Model <- eval(parse(text = gsub("C","C26",sMLRtempEquNCoef)))
RankofC26Model <- rank(C26Model)

#Compute Test Rank for Polygon 26 Validation Set
V26Model <- eval(parse(text = gsub("C","V26",sMLRtempEquNCoef)))
RankofV26Model <- rank(V26Model)


#Compute Test Rank for Polygon 27 Calibration Set
C27Model <- eval(parse(text = gsub("C","C27",sMLRtempEquNCoef)))
RankofC27Model <- rank(C27Model)

#Compute Test Rank for Polygon 27 Validation Set
V27Model <- eval(parse(text = gsub("C","V27",sMLRtempEquNCoef)))
RankofV27Model <- rank(V27Model)

#Compute Test Rank for Polygon 28 Calibration Set
C28Model <- eval(parse(text = gsub("C","C28",sMLRtempEquNCoef)))
RankofC28Model <- rank(C28Model)

#Compute Test Rank for Polygon 28 Validation Set
```

```
V28Model <- eval(parse(text = gsub("C","V28",sMLRtempEquNCoef)))
RankofV28Model <- rank(V28Model)

#Compute Test Rank for Polygon 2955 Calibration Set
C2955Model <- eval(parse(text = gsub("C","C2955",sMLRtempEquNCoef)))
RankofC2955Model <- rank(C2955Model)

#Compute Test Rank for Polygon 2955 Validation Set
V2955Model <- eval(parse(text = gsub("C","V2955",sMLRtempEquNCoef)))
RankofV2955Model <- rank(V2955Model)

#Compute Test Rank for Polygon 30 Calibration Set
C30Model <- eval(parse(text = gsub("C","C30",sMLRtempEquNCoef)))
RankofC30Model <- rank(C30Model)

#Compute Test Rank for Polygon 30 Validation Set
V30Model <- eval(parse(text = gsub("C","V30",sMLRtempEquNCoef)))
RankofV30Model <- rank(V30Model)

#Compute Test Rank for Polygon 32 Calibration Set
C32Model <- eval(parse(text = sMLRtempEquNCoef))
RankofC32Model <- rank(C32Model)

#Compute Test Rank for Polygon 32 Validation Set
V32Model <- eval(parse(text = gsub("C","V",sMLRtempEquNCoef)))
RankofV32Model <- rank(V32Model)

#Compute Test Rank for Polygon 33 Calibration Set
C33Model <- eval(parse(text = gsub("C","C33",sMLRtempEquNCoef)))
RankofC33Model <- rank(C33Model)

#Compute Test Rank for Polygon 33 Validation Set
V33Model <- eval(parse(text = gsub("C","V33",sMLRtempEquNCoef)))
RankofV33Model <- rank(V33Model)

#Compute Test Rank for Polygon 37 Calibration Set
C37Model <- eval(parse(text = gsub("C","C37",sMLRtempEquNCoef)))
RankofC37Model <- rank(C37Model)

#Compute Test Rank for Polygon 37 Validation Set
V37Model <- eval(parse(text = gsub("C","V37",sMLRtempEquNCoef)))
RankofV37Model <- rank(V37Model)

#Compute Test Rank for Polygon 38 Calibration Set
C38Model <- eval(parse(text = gsub("C","C38",sMLRtempEquNCoef)))
RankofC38Model <- rank(C38Model)

#Compute Test Rank for Polygon 38 Validation Set
V38Model <- eval(parse(text = gsub("C","V38",sMLRtempEquNCoef)))
RankofV38Model <- rank(V38Model)

#Compute Test Rank for Polygon 42 Calibration Set
C42Model <- eval(parse(text = gsub("C","C42",sMLRtempEquNCoef)))
RankofC42Model <- rank(C42Model)

#Compute Test Rank for Polygon 42 Validation Set
V42Model <- eval(parse(text = gsub("C","V42",sMLRtempEquNCoef)))
RankofV42Model <- rank(V42Model)

#Compute Test Rank for Polygon 48 Calibration Set
C48Model <- eval(parse(text = gsub("C","C48",sMLRtempEquNCoef)))
RankofC48Model <- rank(C48Model)

#Compute Test Rank for Polygon 48 Validation Set
V48Model <- eval(parse(text = gsub("C","V48",sMLRtempEquNCoef)))
RankofV48Model <- rank(V48Model)

#Compute Test Rank for Polygon 49 Calibration Set
C49Model <- eval(parse(text = gsub("C","C49",sMLRtempEquNCoef)))
RankofC49Model <- rank(C49Model)
```

```
#Compute Test Rank for Polygon 49 Validation Set
V49Model <- eval(parse(text = gsub("C","V49",sMLRtempEquNCoef)))
RankofV49Model <- rank(V49Model)

#Compute Test Rank for Polygon 57 Calibration Set
C57Model <- eval(parse(text = gsub("C","C57",sMLRtempEquNCoef)))
RankofC57Model <- rank(C57Model)



#Measure of model fit
PercentC17 <- ((length(C17AvgTotal) - min(RankofC17Model[1],RankofC17Model[2],RankofC17Model[3],RankofC17Model[4])) -3)/length(C17AvgTotal)
PercentV17 <- ((length(V17AvgTotal) - min(RankofV17Model[1],RankofV17Model[2])) -1)/length(V17AvgTotal)
PercentC20 <- ((length(C20AvgTotal) - min(RankofC20Model[1],RankofC20Model[2],RankofC20Model[3])) -2)/length(C20AvgTotal)
PercentV20 <- ((length(V20AvgTotal) - min(RankofV20Model[1],RankofV20Model[2])) -1)/length(V20AvgTotal)
PercentC21 <- ((length(C21AvgTotal) - min(RankofC21Model[1],RankofC21Model[2])) -1)/length(C21AvgTotal)
PercentV21 <- ((length(V21AvgTotal) - min(RankofV21Model[1],RankofV21Model[2])) -1)/length(V21AvgTotal)
PercentC26 <- (length(C26AvgTotal) - RankofC26Model[1])/length(C26AvgTotal)
PercentV26 <- (length(V26AvgTotal) - RankofV26Model[1])/length(V26AvgTotal)
PercentC27 <- ((length(C27AvgTotal) - min(RankofC27Model[1],RankofC27Model[2],RankofC27Model[3])) -2)/length(C27AvgTotal)
PercentV27 <- ((length(V27AvgTotal) - min(RankofV27Model[1],RankofV27Model[2])) -1)/length(V27AvgTotal)
PercentC28 <- ((length(C28AvgTotal) - min(RankofC28Model[1],RankofC28Model[2],RankofC28Model[3])) -2)/length(C28AvgTotal)
PercentV28 <- ((length(V28AvgTotal) - min(RankofV28Model[1],RankofV28Model[2])) -1)/length(V28AvgTotal)
PercentC2955 <- ((length(C2955AvgTotal) - min(RankofC2955Model[1],RankofC2955Model[2],RankofC2955Model[3],RankofC2955Model[4])) - 3)/length(C2955AvgTotal)
PercentV2955 <- ((length(V2955AvgTotal) - min(RankofV2955Model[1],RankofV2955Model[2])) -1)/length(V2955AvgTotal)
PercentC30 <- ((length(C30AvgTotal) - min(RankofC30Model[1],RankofC30Model[2],RankofC30Model[3],RankofC30Model[4],RankofC30Model[5])) - 4)/length(C30AvgTotal)
PercentV30 <- ((length(V30AvgTotal) - min(RankofV30Model[1],RankofV30Model[2])) -1)/length(V30AvgTotal)
PercentC32 <- ((length(CAvgTotal) - min(RankofC32Model[1],RankofC32Model[2],RankofC32Model[3],RankofC32Model[4],RankofC30Model[5])) - 4)/length(CAvgTotal)
PercentV32 <- ((length(VAvgTotal) - min(RankofV32Model[1],RankofV32Model[2])) -1)/length(VAvgTotal)
PercentC33 <- ((length(C33AvgTotal) - min(RankofC33Model[1],RankofC33Model[2],RankofC33Model[3])) -2)/length(C33AvgTotal)
PercentV33 <- ((length(V33AvgTotal) - min(RankofV33Model[1],RankofV33Model[2])) -1)/length(V33AvgTotal)
PercentC37 <- (length(C37AvgTotal) - RankofC37Model[1])/length(C37AvgTotal)
PercentV37 <- (length(V37AvgTotal) - RankofV37Model[1])/length(V37AvgTotal)
PercentC38 <- ((length(C38AvgTotal) - min(RankofC38Model[1],RankofC38Model[2])) -1)/length(C38AvgTotal)
PercentV38 <- ((length(V38AvgTotal) - min(RankofV38Model[1],RankofV38Model[2])) -1)/length(V38AvgTotal)
PercentC42 <- ((length(C42AvgTotal) - min(RankofC42Model[1],RankofC42Model[2],RankofC42Model[3])) -2)/length(C42AvgTotal)
PercentV42 <- ((length(V42AvgTotal) - min(RankofV42Model[1],RankofV42Model[2])) -1)/length(V42AvgTotal)
PercentC48 <- ((length(C48AvgTotal) - min(RankofC48Model[1],RankofC48Model[2])) -1)/length(C48AvgTotal)
```

```
PercentV48 <- (length(V48AvgTotal) -
RankofV48Model[1])/length(V48AvgTotal)
PercentC49 <- ((length(C49AvgTotal) -
min(RankofC49Model[1],RankofC49Model[2])) -1)/length(C49AvgTotal)
PercentV49 <- (length(V49AvgTotal) -
RankofV49Model[1])/length(V49AvgTotal)
PercentC57 <- (length(C57AvgTotal) -
RankofC57Model[1])/length(C57AvgTotal)

PercentAll <- (length(C17AvgTotal)*PercentC17+
        length(V17AvgTotal)*PercentV17+
        length(C20AvgTotal)*PercentC20+
        length(V20AvgTotal)*PercentV20+
        length(C21AvgTotal)*PercentC21+
        length(V21AvgTotal)*PercentV21+
        length(C26AvgTotal)*PercentC26+
        length(V26AvgTotal)*PercentV26+
        length(C27AvgTotal)*PercentC27+
        length(V27AvgTotal)*PercentV27+
        length(C28AvgTotal)*PercentC28+
        length(V28AvgTotal)*PercentV28+
        length(C2955AvgTotal)*PercentC2955+
        length(V2955AvgTotal)*PercentV2955+
        length(C30AvgTotal)*PercentC30+
        length(V30AvgTotal)*PercentV30+
        length(CAvgTotal)*PercentC32+
        length(VAvgTotal)*PercentV32+
        length(C33AvgTotal)*PercentC33+
        length(V33AvgTotal)*PercentV33+
        length(C37AvgTotal)*PercentC37+
        length(V37AvgTotal)*PercentV37+
        length(C38AvgTotal)*PercentC38+
        length(V38AvgTotal)*PercentV38+
        length(C42AvgTotal)*PercentC42+
        length(V42AvgTotal)*PercentV42+
        length(C48AvgTotal)*PercentC48+
        length(V48AvgTotal)*PercentV48+
        length(C49AvgTotal)*PercentC49+
        length(V49AvgTotal)*PercentV49+
        length(C57AvgTotal)*PercentC57)/(
        length(C17AvgTotal)+
        length(V17AvgTotal)+
        length(C20AvgTotal)+
        length(V20AvgTotal)+
        length(C21AvgTotal)+
        length(V21AvgTotal)+
        length(C26AvgTotal)+
        length(V26AvgTotal)+
        length(C27AvgTotal)+
        length(V27AvgTotal)+
        length(C28AvgTotal)+
        length(V28AvgTotal)+
        length(C2955AvgTotal)+
        length(V2955AvgTotal)+
        length(C30AvgTotal)+
        length(V30AvgTotal)+
        length(CAvgTotal)+
        length(VAvgTotal)+
        length(C33AvgTotal)+
        length(V33AvgTotal)+
        length(C37AvgTotal)+
        length(V37AvgTotal)+
        length(C38AvgTotal)+
        length(V38AvgTotal)+
        length(C42AvgTotal)+
        length(V42AvgTotal)+
        length(C48AvgTotal)+
        length(V48AvgTotal)+
        length(C49AvgTotal)+
        length(V49AvgTotal)+
```

```
                                        length(C57AvgTotal))

                                #Test if this Formula is worthy of Top100, if rank is better than 12345
                                        #If worthy, record Equ & rank data in Top100 table
                                if(!is.na(PercentAll))
                                        if(PercentAll<max(Top100RankTable[,2]))
                                        {

  #match(min(Top100RankTable[,2]),Top100RankTable[,2])    this is the line that needs to be overwritten


  Top100RankTable[match(max(Top100RankTable[,2]),Top100RankTable[,2]),1] <- gsub("C","",sMLRtempEquNCoef)

  Top100RankTable[match(max(Top100RankTable[,2]),Top100RankTable[,2]),2] <- PercentAll
                                        }

                                equnumber <- equnumber+1
                                print(equnumber)
                        }
                        else
                        {
                                for(m10 in 1:length(coeffs))
                                {
                                        #If there are only 3 variables within the parentheses...
                                        if(i10==3)
                                        {
                                                #partly inspired by post at
https://stat.ethz.ch/pipermail/r-help/2008-October/175554.html
                                                #Character Locations of opening and closing
parenthese in string
                                                OpenParLoc <- which(strsplit(Formulas, '')[[1]]=='(')
                                                CloseParLoc <- which(strsplit(Formulas, '')[[1]]==')')
                                                PlusLoc <- which(strsplit(Formulas, '')[[1]]=='+')

                                                #Cuts appart the Formulas equation and inserts coeffs
as per k10, l10 and m10 for loops
                                                TempFormula <-
paste(substr(Formulas,1,OpenParLoc[length(OpenParLoc)]),coeffs[k10],"*",substr(Formulas,OpenParLoc[length(Open
ParLoc)]+1,PlusLoc[1]),coeffs[l10],"*",substr(Formulas,PlusLoc[1]+1,PlusLoc[2]),coeffs[m10],"*",substr(Formulas,PlusL
oc[2]+1,nchar(Formulas)))

                                                #Perform stepwise regression with formulas
                                                sMLR <-
step(lm(as.formula(paste("CFloodings~",TempFormula))))

                                                #extract the stepwise regression equation
                                                sMLRtempEqu <-
substr(toString(formula(sMLR)),tail(which(strsplit(toString(formula(sMLR)),
')[[1]]==','),1)+2,nchar(toString(formula(sMLR))))

                                                #extract the stepwise regression coefficients
                                                sMLRcoeffs <- coefficients(sMLR)

                                                #Finding the spaces in the equation which should be
before and after each operation and place it into an array
                                                SpaceLoc <- unlist(gregexpr(" ",sMLRtempEqu),1)


                                                switch(length(coefficients(sMLR)),
                                                        sMLRtempEquNCoef <- sMLRcoeffs[1],
                                                        sMLRtempEquNCoef <-
paste(sMLRcoeffs[1],"+",sMLRcoeffs[2],"*",sMLRtempEqu),
                                                        sMLRtempEquNCoef <-
paste(sMLRcoeffs[1],"+",sMLRcoeffs[2],"*",substr(sMLRtempEqu,1,tail(SpaceLoc,1)-
1),sMLRcoeffs[3],"*",substr(sMLRtempEqu,tail(SpaceLoc,1),nchar(sMLRtempEqu))),
                                                        sMLRtempEquNCoef <-
paste(sMLRcoeffs[1],"+",sMLRcoeffs[2],"*",substr(sMLRtempEqu,1,SpaceLoc[length(SpaceLoc)-2]-
1),sMLRcoeffs[3],"*",substr(sMLRtempEqu,SpaceLoc[length(SpaceLoc)-2]+1,tail(SpaceLoc,1)-
1),sMLRcoeffs[4],"*",substr(sMLRtempEqu,tail(SpaceLoc,1),nchar(sMLRtempEqu))),
                                                        )
```

```
#Assessing the Calibration and Validation fit for all
polygons

#Compute Test Rank for Polygon 17 Calibration Set
C17Model <- eval(parse(text =
gsub("C","C17",sMLRtempEquNCoef)))

RankofC17Model <- rank(C17Model)

#Compute Test Rank for Polygon 17 Validation Set
V17Model <- eval(parse(text =
gsub("C","V17",sMLRtempEquNCoef)))

RankofV17Model <- rank(V17Model)

#Compute Test Rank for Polygon 20 Calibration Set
C20Model <- eval(parse(text =
gsub("C","C20",sMLRtempEquNCoef)))

RankofC20Model <- rank(C20Model)

#Compute Test Rank for Polygon 20 Validation Set
V20Model <- eval(parse(text =
gsub("C","V20",sMLRtempEquNCoef)))

RankofV20Model <- rank(V20Model)

#Compute Test Rank for Polygon 21 Calibration Set
C21Model <- eval(parse(text =
gsub("C","C21",sMLRtempEquNCoef)))

RankofC21Model <- rank(C21Model)

#Compute Test Rank for Polygon 21 Validation Set
V21Model <- eval(parse(text =
gsub("C","V21",sMLRtempEquNCoef)))

RankofV21Model <- rank(V21Model)

#Compute Test Rank for Polygon 26 Calibration Set
C26Model <- eval(parse(text =
gsub("C","C26",sMLRtempEquNCoef)))

RankofC26Model <- rank(C26Model)

#Compute Test Rank for Polygon 26 Validation Set
V26Model <- eval(parse(text =
gsub("C","V26",sMLRtempEquNCoef)))

RankofV26Model <- rank(V26Model)


#Compute Test Rank for Polygon 27 Calibration Set
C27Model <- eval(parse(text =
gsub("C","C27",sMLRtempEquNCoef)))

RankofC27Model <- rank(C27Model)

#Compute Test Rank for Polygon 27 Validation Set
V27Model <- eval(parse(text =
gsub("C","V27",sMLRtempEquNCoef)))

RankofV27Model <- rank(V27Model)

#Compute Test Rank for Polygon 28 Calibration Set
C28Model <- eval(parse(text =
gsub("C","C28",sMLRtempEquNCoef)))

RankofC28Model <- rank(C28Model)

#Compute Test Rank for Polygon 28 Validation Set
V28Model <- eval(parse(text =
gsub("C","V28",sMLRtempEquNCoef)))

RankofV28Model <- rank(V28Model)

#Compute Test Rank for Polygon 2955 Calibration Set
C2955Model <- eval(parse(text =
gsub("C","C2955",sMLRtempEquNCoef)))

RankofC2955Model <- rank(C2955Model)

#Compute Test Rank for Polygon 2955 Validation Set
```

```
V2955Model <- eval(parse(text = gsub("C","V2955",sMLRtempEquNCoef)))
RankofV2955Model <- rank(V2955Model)

#Compute Test Rank for Polygon 30 Calibration Set
C30Model <- eval(parse(text = gsub("C","C30",sMLRtempEquNCoef)))
RankofC30Model <- rank(C30Model)

#Compute Test Rank for Polygon 30 Validation Set
V30Model <- eval(parse(text = gsub("C","V30",sMLRtempEquNCoef)))
RankofV30Model <- rank(V30Model)

#Compute Test Rank for Polygon 32 Calibration Set
C32Model <- eval(parse(text = sMLRtempEquNCoef))
RankofC32Model <- rank(C32Model)

#Compute Test Rank for Polygon 32 Validation Set
V32Model <- eval(parse(text = gsub("C","V",sMLRtempEquNCoef)))
RankofV32Model <- rank(V32Model)

#Compute Test Rank for Polygon 33 Calibration Set
C33Model <- eval(parse(text = gsub("C","C33",sMLRtempEquNCoef)))
RankofC33Model <- rank(C33Model)

#Compute Test Rank for Polygon 33 Validation Set
V33Model <- eval(parse(text = gsub("C","V33",sMLRtempEquNCoef)))
RankofV33Model <- rank(V33Model)

#Compute Test Rank for Polygon 37 Calibration Set
C37Model <- eval(parse(text = gsub("C","C37",sMLRtempEquNCoef)))
RankofC37Model <- rank(C37Model)

#Compute Test Rank for Polygon 37 Validation Set
V37Model <- eval(parse(text = gsub("C","V37",sMLRtempEquNCoef)))
RankofV37Model <- rank(V37Model)

#Compute Test Rank for Polygon 38 Calibration Set
C38Model <- eval(parse(text = gsub("C","C38",sMLRtempEquNCoef)))
RankofC38Model <- rank(C38Model)

#Compute Test Rank for Polygon 38 Validation Set
V38Model <- eval(parse(text = gsub("C","V38",sMLRtempEquNCoef)))
RankofV38Model <- rank(V38Model)

#Compute Test Rank for Polygon 42 Calibration Set
C42Model <- eval(parse(text = gsub("C","C42",sMLRtempEquNCoef)))
RankofC42Model <- rank(C42Model)

#Compute Test Rank for Polygon 42 Validation Set
V42Model <- eval(parse(text = gsub("C","V42",sMLRtempEquNCoef)))
RankofV42Model <- rank(V42Model)

#Compute Test Rank for Polygon 48 Calibration Set
C48Model <- eval(parse(text = gsub("C","C48",sMLRtempEquNCoef)))
RankofC48Model <- rank(C48Model)

#Compute Test Rank for Polygon 48 Validation Set
```

```
V48Model <- eval(parse(text = gsub("C","V48",sMLRtempEquNCoef)))
RankofV48Model <- rank(V48Model)

#Compute Test Rank for Polygon 49 Calibration Set
C49Model <- eval(parse(text = gsub("C","C49",sMLRtempEquNCoef)))
RankofC49Model <- rank(C49Model)

#Compute Test Rank for Polygon 49 Validation Set
V49Model <- eval(parse(text = gsub("C","V49",sMLRtempEquNCoef)))
RankofV49Model <- rank(V49Model)

#Compute Test Rank for Polygon 57 Calibration Set
C57Model <- eval(parse(text = gsub("C","C57",sMLRtempEquNCoef)))
RankofC57Model <- rank(C57Model)


#Measure of model fit
PercentC17 <- ((length(C17AvgTotal) - min(RankofC17Model[1],RankofC17Model[2],RankofC17Model[3],RankofC17Model[4])) -3)/length(C17AvgTotal)
PercentV17 <- ((length(V17AvgTotal) - min(RankofV17Model[1],RankofV17Model[2])) -1)/length(V17AvgTotal)
PercentC20 <- ((length(C20AvgTotal) - min(RankofC20Model[1],RankofC20Model[2],RankofC20Model[3])) -2)/length(C20AvgTotal)
PercentV20 <- ((length(V20AvgTotal) - min(RankofV20Model[1],RankofV20Model[2])) -1)/length(V20AvgTotal)
PercentC21 <- ((length(C21AvgTotal) - min(RankofC21Model[1],RankofC21Model[2])) -1)/length(C21AvgTotal)
PercentV21 <- ((length(V21AvgTotal) - min(RankofV21Model[1],RankofV21Model[2])) -1)/length(V21AvgTotal)
PercentC26 <- (length(C26AvgTotal) - RankofC26Model[1])/length(C26AvgTotal)
PercentV26 <- (length(V26AvgTotal) - RankofV26Model[1])/length(V26AvgTotal)
PercentC27 <- ((length(C27AvgTotal) - min(RankofC27Model[1],RankofC27Model[2],RankofC27Model[3])) -2)/length(C27AvgTotal)
PercentV27 <- ((length(V27AvgTotal) - min(RankofV27Model[1],RankofV27Model[2])) -1)/length(V27AvgTotal)
PercentC28 <- ((length(C28AvgTotal) - min(RankofC28Model[1],RankofC28Model[2],RankofC28Model[3])) -2)/length(C28AvgTotal)
PercentV28 <- ((length(V28AvgTotal) - min(RankofV28Model[1],RankofV28Model[2])) -1)/length(V28AvgTotal)
PercentC2955 <- ((length(C2955AvgTotal) - min(RankofC2955Model[1],RankofC2955Model[2],RankofC2955Model[3],RankofC2955Model[4])) -3)/length(C2955AvgTotal)
PercentV2955 <- ((length(V2955AvgTotal) - min(RankofV2955Model[1],RankofV2955Model[2])) -1)/length(V2955AvgTotal)
PercentC30 <- ((length(C30AvgTotal) - min(RankofC30Model[1],RankofC30Model[2],RankofC30Model[3],RankofC30Model[4],RankofC30Model[5])) -4)/length(C30AvgTotal)
PercentV30 <- ((length(V30AvgTotal) - min(RankofV30Model[1],RankofV30Model[2])) -1)/length(V30AvgTotal)
PercentC32 <- ((length(CAvgTotal) - min(RankofC32Model[1],RankofC32Model[2],RankofC32Model[3],RankofC32Model[4],RankofC30Model[5])) -4)/length(CAvgTotal)
PercentV32 <- ((length(VAvgTotal) - min(RankofV32Model[1],RankofV32Model[2])) -1)/length(VAvgTotal)
PercentC33 <- ((length(C33AvgTotal) - min(RankofC33Model[1],RankofC33Model[2],RankofC33Model[3])) -2)/length(C33AvgTotal)
PercentV33 <- ((length(V33AvgTotal) - min(RankofV33Model[1],RankofV33Model[2])) -1)/length(V33AvgTotal)
PercentC37 <- (length(C37AvgTotal) - RankofC37Model[1])/length(C37AvgTotal)
PercentV37 <- (length(V37AvgTotal) - RankofV37Model[1])/length(V37AvgTotal)
```

```
                                                    PercentC38      <-      ((length(C38AvgTotal)       -
min(RankofC38Model[1],RankofC38Model[2])) -1)/length(C38AvgTotal)
                                                    PercentV38      <-      ((length(V38AvgTotal)       -
min(RankofV38Model[1],RankofV38Model[2])) -1)/length(V38AvgTotal)
                                                    PercentC42      <-      ((length(C42AvgTotal)       -
min(RankofC42Model[1],RankofC42Model[2],RankofC42Model[3])) -2)/length(C42AvgTotal)
                                                    PercentV42      <-      ((length(V42AvgTotal)       -
min(RankofV42Model[1],RankofV42Model[2])) -1)/length(V42AvgTotal)
                                                    PercentC48      <-      ((length(C48AvgTotal)       -
min(RankofC48Model[1],RankofC48Model[2])) -1)/length(C48AvgTotal)
                                                    PercentV48      <-      (length(V48AvgTotal)        -
RankofV48Model[1])/length(V48AvgTotal)
                                                    PercentC49      <-      ((length(C49AvgTotal)       -
min(RankofC49Model[1],RankofC49Model[2])) -1)/length(C49AvgTotal)
                                                    PercentV49      <-      (length(V49AvgTotal)        -
RankofV49Model[1])/length(V49AvgTotal)
                                                    PercentC57      <-      (length(C57AvgTotal)        -
RankofC57Model[1])/length(C57AvgTotal)

                                                    PercentAll <- (length(C17AvgTotal)*PercentC17+
                                                            length(V17AvgTotal)*PercentV17+
                                                            length(C20AvgTotal)*PercentC20+
                                                            length(V20AvgTotal)*PercentV20+
                                                            length(C21AvgTotal)*PercentC21+
                                                            length(V21AvgTotal)*PercentV21+
                                                            length(C26AvgTotal)*PercentC26+
                                                            length(V26AvgTotal)*PercentV26+
                                                            length(C27AvgTotal)*PercentC27+
                                                            length(V27AvgTotal)*PercentV27+
                                                            length(C28AvgTotal)*PercentC28+
                                                            length(V28AvgTotal)*PercentV28+
                                                            length(C2955AvgTotal)*PercentC2955+
                                                            length(V2955AvgTotal)*PercentV2955+
                                                            length(C30AvgTotal)*PercentC30+
                                                            length(V30AvgTotal)*PercentV30+
                                                            length(CAvgTotal)*PercentC32+
                                                            length(VAvgTotal)*PercentV32+
                                                            length(C33AvgTotal)*PercentC33+
                                                            length(V33AvgTotal)*PercentV33+
                                                            length(C37AvgTotal)*PercentC37+
                                                            length(V37AvgTotal)*PercentV37+
                                                            length(C38AvgTotal)*PercentC38+
                                                            length(V38AvgTotal)*PercentV38+
                                                            length(C42AvgTotal)*PercentC42+
                                                            length(V42AvgTotal)*PercentV42+
                                                            length(C48AvgTotal)*PercentC48+
                                                            length(V48AvgTotal)*PercentV48+
                                                            length(C49AvgTotal)*PercentC49+
                                                            length(V49AvgTotal)*PercentV49+
                                                            length(C57AvgTotal)*PercentC57)/(
                                                            length(C17AvgTotal)+
                                                            length(V17AvgTotal)+
                                                            length(C20AvgTotal)+
                                                            length(V20AvgTotal)+
                                                            length(C21AvgTotal)+
                                                            length(V21AvgTotal)+
                                                            length(C26AvgTotal)+
                                                            length(V26AvgTotal)+
                                                            length(C27AvgTotal)+
                                                            length(V27AvgTotal)+
                                                            length(C28AvgTotal)+
                                                            length(V28AvgTotal)+
                                                            length(C2955AvgTotal)+
                                                            length(V2955AvgTotal)+
                                                            length(C30AvgTotal)+
                                                            length(V30AvgTotal)+
                                                            length(CAvgTotal)+
                                                            length(VAvgTotal)+
                                                            length(C33AvgTotal)+
                                                            length(V33AvgTotal)+
```

```
                                                length(C37AvgTotal)+
                                                length(V37AvgTotal)+
                                                length(C38AvgTotal)+
                                                length(V38AvgTotal)+
                                                length(C42AvgTotal)+
                                                length(V42AvgTotal)+
                                                length(C48AvgTotal)+
                                                length(V48AvgTotal)+
                                                length(C49AvgTotal)+
                                                length(V49AvgTotal)+
                                                length(C57AvgTotal))

                                        #Test if this Formula is worthy of Top100, if rank is
better than 12345
                                                #If worthy, record Equ & rank data in Top100
table
                                        if(!is.na(PercentAll))
                                                if(PercentAll<max(Top100RankTable[,2]))
                                                {

   #match(min(Top100RankTable[,2]),Top100RankTable[,2])    this is the line that needs to be overwritten


   Top100RankTable[match(max(Top100RankTable[,2]),Top100RankTable[,2]),1] <- gsub("C","",sMLRtempEquNCoef)

   Top100RankTable[match(max(Top100RankTable[,2]),Top100RankTable[,2]),2] <- PercentAll
                                                }


                                        equnumber <- equnumber+1
                                        print(equnumber)
                                }
                                else
                                {
                                        for(n10 in 1:length(coeffs))
                                        {
                                                #If there are only 4 variables within the
parentheses...
                                                if(i10==4)
                                                {
                                                        #partly inspired by post at
https://stat.ethz.ch/pipermail/r-help/2008-October/175554.html
                                                        #Character Locations of opening
and closing parenthese in string
                                                        OpenParLoc <-
which(strsplit(Formulas, '')[[1]]=='(')
                                                        CloseParLoc <-
which(strsplit(Formulas, '')[[1]]==')')
                                                        PlusLoc <- which(strsplit(Formulas,
'')[[1]]=='+')

                                                        #Cuts appart the Formulas
equation and inserts coeffs as per k10, l10, m10 and n10 for loops
                                                        TempFormula <-
paste(substr(Formulas,1,OpenParLoc[length(OpenParLoc)]),coeffs[k10],"*",substr(Formulas,OpenParLoc[length(Open
ParLoc)]+1,PlusLoc[1]),coeffs[l10],"*",substr(Formulas,PlusLoc[1]+1,PlusLoc[2]),coeffs[m10],"*",substr(Formulas,PlusL
oc[2]+1,PlusLoc[3]),coeffs[n10],"*",substr(Formulas,PlusLoc[3]+1,nchar(Formulas)))

                                                        #Perform stepwise regression with
formulas
                                                        sMLR <-
step(lm(as.formula(paste("CFloodings~",TempFormula))))


                                                        #extract the stepwise regression
equation
                                                        sMLRtempEqu <-
substr(toString(formula(sMLR)),tail(which(strsplit(toString(formula(sMLR)),
'')[[1]]==','),1)+2,nchar(toString(formula(sMLR))))
```

```
#extract the stepwise regression coefficients
sMLRcoeffs <- coefficients(sMLR)

#Finding the spaces in the equation which should be before and after each operation and place it into an array
SpaceLoc <- unlist(gregexpr(" ",sMLRtempEqu),1)

switch(length(coefficients(sMLR)),
       sMLRtempEquNCoef <- sMLRcoeffs[1],
       sMLRtempEquNCoef <- paste(sMLRcoeffs[1],"+",sMLRcoeffs[2],"*",sMLRtempEqu),
       sMLRtempEquNCoef <- paste(sMLRcoeffs[1],"+",sMLRcoeffs[2],"*",substr(sMLRtempEqu,1,tail(SpaceLoc,1)-1),sMLRcoeffs[3],"*",substr(sMLRtempEqu,tail(SpaceLoc,1),nchar(sMLRtempEqu))),
       )

#Assessing the Calibration and Validation fit for all polygons
#Compute Test Rank for Polygon 17 Calibration Set
C17Model <- eval(parse(text = gsub("C","C17",sMLRtempEquNCoef)))
RankofC17Model <- rank(C17Model)

#Compute Test Rank for Polygon 17 Validation Set
V17Model <- eval(parse(text = gsub("C","V17",sMLRtempEquNCoef)))
RankofV17Model <- rank(V17Model)

#Compute Test Rank for Polygon 20 Calibration Set
C20Model <- eval(parse(text = gsub("C","C20",sMLRtempEquNCoef)))
RankofC20Model <- rank(C20Model)

#Compute Test Rank for Polygon 20 Validation Set
V20Model <- eval(parse(text = gsub("C","V20",sMLRtempEquNCoef)))
RankofV20Model <- rank(V20Model)

#Compute Test Rank for Polygon 21 Calibration Set
C21Model <- eval(parse(text = gsub("C","C21",sMLRtempEquNCoef)))
RankofC21Model <- rank(C21Model)

#Compute Test Rank for Polygon 21 Validation Set
V21Model <- eval(parse(text = gsub("C","V21",sMLRtempEquNCoef)))
RankofV21Model <- rank(V21Model)

#Compute Test Rank for Polygon 26 Calibration Set
C26Model <- eval(parse(text = gsub("C","C26",sMLRtempEquNCoef)))
RankofC26Model <- rank(C26Model)
```

```
26 Validation Set
gsub("C","V26",sMLRtempEquNCoef)))
rank(V26Model)

27 Calibration Set
gsub("C","C27",sMLRtempEquNCoef)))
rank(C27Model)

27 Validation Set
gsub("C","V27",sMLRtempEquNCoef)))
rank(V27Model)

28 Calibration Set
gsub("C","C28",sMLRtempEquNCoef)))
rank(C28Model)

28 Validation Set
gsub("C","V28",sMLRtempEquNCoef)))
rank(V28Model)

2955 Calibration Set
gsub("C","C2955",sMLRtempEquNCoef)))
rank(C2955Model)

2955 Validation Set
gsub("C","V2955",sMLRtempEquNCoef)))
rank(V2955Model)

30 Calibration Set
gsub("C","C30",sMLRtempEquNCoef)))
rank(C30Model)

30 Validation Set
gsub("C","V30",sMLRtempEquNCoef)))
rank(V30Model)

32 Calibration Set
sMLRtempEquNCoef))
```

```
#Compute Test Rank for Polygon
V26Model <- eval(parse(text =
RankofV26Model <-

#Compute Test Rank for Polygon
C27Model <- eval(parse(text =
RankofC27Model <-

#Compute Test Rank for Polygon
V27Model <- eval(parse(text =
RankofV27Model <-

#Compute Test Rank for Polygon
C28Model <- eval(parse(text =
RankofC28Model <-

#Compute Test Rank for Polygon
V28Model <- eval(parse(text =
RankofV28Model <-

#Compute Test Rank for Polygon
C2955Model <- eval(parse(text =
RankofC2955Model <-

#Compute Test Rank for Polygon
V2955Model <- eval(parse(text =
RankofV2955Model <-

#Compute Test Rank for Polygon
C30Model <- eval(parse(text =
RankofC30Model <-

#Compute Test Rank for Polygon
V30Model <- eval(parse(text =
RankofV30Model <-

#Compute Test Rank for Polygon
C32Model <- eval(parse(text =
```

```
rank(C32Model)

32 Validation Set
gsub("C","V",sMLRtempEquNCoef)))
rank(V32Model)

33 Calibration Set
gsub("C","C33",sMLRtempEquNCoef)))
rank(C33Model)

33 Validation Set
gsub("C","V33",sMLRtempEquNCoef)))
rank(V33Model)

37 Calibration Set
gsub("C","C37",sMLRtempEquNCoef)))
rank(C37Model)

37 Validation Set
gsub("C","V37",sMLRtempEquNCoef)))
rank(V37Model)

38 Calibration Set
gsub("C","C38",sMLRtempEquNCoef)))
rank(C38Model)

38 Validation Set
gsub("C","V38",sMLRtempEquNCoef)))
rank(V38Model)

42 Calibration Set
gsub("C","C42",sMLRtempEquNCoef)))
rank(C42Model)

42 Validation Set
gsub("C","V42",sMLRtempEquNCoef)))
rank(V42Model)

48 Calibration Set
gsub("C","C48",sMLRtempEquNCoef)))
RankofC32Model <-

#Compute Test Rank for Polygon
V32Model <- eval(parse(text =
RankofV32Model <-

#Compute Test Rank for Polygon
C33Model <- eval(parse(text =
RankofC33Model <-

#Compute Test Rank for Polygon
V33Model <- eval(parse(text =
RankofV33Model <-

#Compute Test Rank for Polygon
C37Model <- eval(parse(text =
RankofC37Model <-

#Compute Test Rank for Polygon
V37Model <- eval(parse(text =
RankofV37Model <-

#Compute Test Rank for Polygon
C38Model <- eval(parse(text =
RankofC38Model <-

#Compute Test Rank for Polygon
V38Model <- eval(parse(text =
RankofV38Model <-

#Compute Test Rank for Polygon
C42Model <- eval(parse(text =
RankofC42Model <-

#Compute Test Rank for Polygon
V42Model <- eval(parse(text =
RankofV42Model <-

#Compute Test Rank for Polygon
C48Model <- eval(parse(text =
```

```
                                                                RankofC48Model <-
rank(C48Model)

                                                                #Compute Test Rank for Polygon
48 Validation Set
                                                                V48Model <- eval(parse(text =
gsub("C","V48",sMLRtempEquNCoef)))
                                                                RankofV48Model <-
rank(V48Model)

                                                                #Compute Test Rank for Polygon
49 Calibration Set
                                                                C49Model <- eval(parse(text =
gsub("C","C49",sMLRtempEquNCoef)))
                                                                RankofC49Model <-
rank(C49Model)

                                                                #Compute Test Rank for Polygon
49 Validation Set
                                                                V49Model <- eval(parse(text =
gsub("C","V49",sMLRtempEquNCoef)))
                                                                RankofV49Model <-
rank(V49Model)

                                                                #Compute Test Rank for Polygon
57 Calibration Set
                                                                C57Model <- eval(parse(text =
gsub("C","C57",sMLRtempEquNCoef)))
                                                                RankofC57Model <-
rank(C57Model)



                                                                #Measure of model fit
                                                                PercentC17 <-
((length(C17AvgTotal) - min(RankofC17Model[1],RankofC17Model[2],RankofC17Model[3],RankofC17Model[4])) -
3)/length(C17AvgTotal)
                                                                PercentV17 <-
((length(V17AvgTotal) - min(RankofV17Model[1],RankofV17Model[2])) -1)/length(V17AvgTotal)
                                                                PercentC20 <-
((length(C20AvgTotal) - min(RankofC20Model[1],RankofC20Model[2],RankofC20Model[3])) -2)/length(C20AvgTotal)
                                                                PercentV20 <-
((length(V20AvgTotal) - min(RankofV20Model[1],RankofV20Model[2])) -1)/length(V20AvgTotal)
                                                                PercentC21 <-
((length(C21AvgTotal) - min(RankofC21Model[1],RankofC21Model[2])) -1)/length(C21AvgTotal)
                                                                PercentV21 <-
((length(V21AvgTotal) - min(RankofV21Model[1],RankofV21Model[2])) -1)/length(V21AvgTotal)
                                                                PercentC26 <-
(length(C26AvgTotal) - RankofC26Model[1])/length(C26AvgTotal)
                                                                PercentV26 <-
(length(V26AvgTotal) - RankofV26Model[1])/length(V26AvgTotal)
                                                                PercentC27 <-
((length(C27AvgTotal) - min(RankofC27Model[1],RankofC27Model[2],RankofC27Model[3])) -2)/length(C27AvgTotal)
                                                                PercentV27 <-
((length(V27AvgTotal) - min(RankofV27Model[1],RankofV27Model[2])) -1)/length(V27AvgTotal)
                                                                PercentC28 <-
((length(C28AvgTotal) - min(RankofC28Model[1],RankofC28Model[2],RankofC28Model[3])) -2)/length(C28AvgTotal)
                                                                PercentV28 <-
((length(V28AvgTotal) - min(RankofV28Model[1],RankofV28Model[2])) -1)/length(V28AvgTotal)
                                                                PercentC2955 <-
((length(C2955AvgTotal)                                                                -
min(RankofC2955Model[1],RankofC2955Model[2],RankofC2955Model[3],RankofC2955Model[4]))                -
3)/length(C2955AvgTotal)
                                                                PercentV2955 <-
((length(V2955AvgTotal) - min(RankofV2955Model[1],RankofV2955Model[2])) -1)/length(V2955AvgTotal)
                                                                PercentC30 <-
((length(C30AvgTotal)                                                                -
min(RankofC30Model[1],RankofC30Model[2],RankofC30Model[3],RankofC30Model[4],RankofC30Model[5]))              -
4)/length(C30AvgTotal)
```

```
                                                                    PercentV30                  <-
((length(V30AvgTotal) - min(RankofV30Model[1],RankofV30Model[2])) -1)/length(V30AvgTotal)
                                                                    PercentC32 <- ((length(CAvgTotal)
-       min(RankofC32Model[1],RankofC32Model[2],RankofC32Model[3],RankofC32Model[4],RankofC30Model[5]))     -
4)/length(CAvgTotal)
                                                                    PercentV32 <- ((length(VAvgTotal)
- min(RankofV32Model[1],RankofV32Model[2])) -1)/length(VAvgTotal)
                                                                    PercentC33                  <-
((length(C33AvgTotal) - min(RankofC33Model[1],RankofC33Model[2],RankofC33Model[3])) -2)/length(C33AvgTotal)
                                                                    PercentV33                  <-
((length(V33AvgTotal) - min(RankofV33Model[1],RankofV33Model[2])) -1)/length(V33AvgTotal)
                                                                    PercentC37                  <-
(length(C37AvgTotal) - RankofC37Model[1])/length(C37AvgTotal)
                                                                    PercentV37                  <-
(length(V37AvgTotal) - RankofV37Model[1])/length(V37AvgTotal)
                                                                    PercentC38                  <-
((length(C38AvgTotal) - min(RankofC38Model[1],RankofC38Model[2])) -1)/length(C38AvgTotal)
                                                                    PercentV38                  <-
((length(V38AvgTotal) - min(RankofV38Model[1],RankofV38Model[2])) -1)/length(V38AvgTotal)
                                                                    PercentC42                  <-
((length(C42AvgTotal) - min(RankofC42Model[1],RankofC42Model[2],RankofC42Model[3])) -2)/length(C42AvgTotal)
                                                                    PercentV42                  <-
((length(V42AvgTotal) - min(RankofV42Model[1],RankofV42Model[2])) -1)/length(V42AvgTotal)
                                                                    PercentC48                  <-
((length(C48AvgTotal) - min(RankofC48Model[1],RankofC48Model[2])) -1)/length(C48AvgTotal)
                                                                    PercentV48                  <-
(length(V48AvgTotal) - RankofV48Model[1])/length(V48AvgTotal)
                                                                    PercentC49                  <-
((length(C49AvgTotal) - min(RankofC49Model[1],RankofC49Model[2])) -1)/length(C49AvgTotal)
                                                                    PercentV49                  <-
(length(V49AvgTotal) - RankofV49Model[1])/length(V49AvgTotal)
                                                                    PercentC57                  <-
(length(C57AvgTotal) - RankofC57Model[1])/length(C57AvgTotal)

                                                                    PercentAll                  <-
(length(C17AvgTotal)*PercentC17+

   length(V17AvgTotal)*PercentV17+

   length(C20AvgTotal)*PercentC20+

   length(V20AvgTotal)*PercentV20+

   length(C21AvgTotal)*PercentC21+

   length(V21AvgTotal)*PercentV21+

   length(C26AvgTotal)*PercentC26+

   length(V26AvgTotal)*PercentV26+

   length(C27AvgTotal)*PercentC27+

   length(V27AvgTotal)*PercentV27+

   length(C28AvgTotal)*PercentC28+

   length(V28AvgTotal)*PercentV28+

   length(C2955AvgTotal)*PercentC2955+

   length(V2955AvgTotal)*PercentV2955+

   length(C30AvgTotal)*PercentC30+

   length(V30AvgTotal)*PercentV30+

   length(CAvgTotal)*PercentC32+

   length(VAvgTotal)*PercentV32+
```

```
    length(C33AvgTotal)*PercentC33+
    length(V33AvgTotal)*PercentV33+
    length(C37AvgTotal)*PercentC37+
    length(V37AvgTotal)*PercentV37+
    length(C38AvgTotal)*PercentC38+
    length(V38AvgTotal)*PercentV38+
    length(C42AvgTotal)*PercentC42+
    length(V42AvgTotal)*PercentV42+
    length(C48AvgTotal)*PercentC48+
    length(V48AvgTotal)*PercentV48+
    length(C49AvgTotal)*PercentC49+
    length(V49AvgTotal)*PercentV49+
    length(C57AvgTotal)*PercentC57)/(
                                                    length(C17AvgTotal)+
                                                    length(V17AvgTotal)+
                                                    length(C20AvgTotal)+
                                                    length(V20AvgTotal)+
                                                    length(C21AvgTotal)+
                                                    length(V21AvgTotal)+
                                                    length(C26AvgTotal)+
                                                    length(V26AvgTotal)+
                                                    length(C27AvgTotal)+
                                                    length(V27AvgTotal)+
                                                    length(C28AvgTotal)+
                                                    length(V28AvgTotal)+
                                                    length(C2955AvgTotal)+
                                                    length(V2955AvgTotal)+
                                                    length(C30AvgTotal)+
                                                    length(V30AvgTotal)+
                                                    length(CAvgTotal)+
                                                    length(VAvgTotal)+
                                                    length(C33AvgTotal)+
                                                    length(V33AvgTotal)+
                                                    length(C37AvgTotal)+
                                                    length(V37AvgTotal)+
                                                    length(C38AvgTotal)+
                                                    length(V38AvgTotal)+
                                                    length(C42AvgTotal)+
                                                    length(V42AvgTotal)+
                                                    length(C48AvgTotal)+
                                                    length(V48AvgTotal)+
                                                    length(C49AvgTotal)+
                                                    length(V49AvgTotal)+
                                                    length(C57AvgTotal))

                                                #Test if this Formula is worthy of
Top100, if rank is better than 12345
                                                    #If worthy, record Equ &
rank data in Top100 table
                                                if(!is.na(PercentAll))

    if(PercentAll<max(Top100RankTable[,2]))
                                                    {
    #match(min(Top100RankTable[,2]),Top100RankTable[,2])    this is the line that needs to be overwritten
```

```
Top100RankTable[match(max(Top100RankTable[,2]),Top100RankTable[,2]),1] <- gsub("C","",sMLRtempEquNCoef)

Top100RankTable[match(max(Top100RankTable[,2]),Top100RankTable[,2]),2] <- PercentAll
                                                                                }


                                                                        equnumber <- equnumber+1
                                                                        print(equnumber)
                                                                }
                                                                else
                                                                {
                                                                        #If there are all of 5 variables
within the parentheses...
                                                                        for(o10 in 1:length(coeffs))
                                                                        {
                                                                                #partly inspired by post
at https://stat.ethz.ch/pipermail/r-help/2008-October/175554.html
                                                                                #Character Locations of
opening and closing parenthese in string
                                                                                OpenParLoc <-
which(strsplit(Formulas, '')[[1]]=='(')
                                                                                CloseParLoc <-
which(strsplit(Formulas, '')[[1]]==')')
                                                                                PlusLoc <-
which(strsplit(Formulas, '')[[1]]=='+')

                                                                                #Cuts appart the
Formulas equation and inserts coeffs as per k10, l10, m10, n10 and o10 for loops
                                                                                TempFormula <-
paste(substr(Formulas,1,OpenParLoc[length(OpenParLoc)]),coeffs[k10],"*",substr(Formulas,OpenParLoc[length(Open
ParLoc)]+1,PlusLoc[1]),coeffs[l10],"*",substr(Formulas,PlusLoc[1]+1,PlusLoc[2]),coeffs[m10],"*",substr(Formulas,PlusL
oc[2]+1,PlusLoc[3]),coeffs[n10],"*",substr(Formulas,PlusLoc[3]+1,PlusLoc[4]),coeffs[o10],"*",substr(Formulas,PlusLoc[4
]+1,nchar(Formulas)))

                                                                                #Perform stepwise
regression with formulas
                                                                                sMLR <-
step(lm(as.formula(paste("CFloodings~",TempFormula))))



                                                                                #extract the stepwise
regression equation
                                                                                sMLRtempEqu <-
substr(toString(formula(sMLR)),tail(which(strsplit(toString(formula(sMLR)),
")[[1]]==','),1)+2,nchar(toString(formula(sMLR))))

                                                                                #extract the stepwise
regression coefficients
                                                                                sMLRcoeffs <-
coefficients(sMLR)

                                                                                #Finding the spaces in
the equation which should be before and after each operation and place it into an array
                                                                                SpaceLoc <-
unlist(gregexpr(" ",sMLRtempEqu),1)


   switch(length(coefficients(sMLR)),

   sMLRtempEquNCoef <- sMLRcoeffs[1],

   sMLRtempEquNCoef <- paste(sMLRcoeffs[1],"+",sMLRcoeffs[2],"*",sMLRtempEqu),
                                                                                )
```

Calibration and Validation fit for all polygons

Polygon 17 Calibration Set

```
eval(parse(text = gsub("C","C17",sMLRtempEquNCoef)))
rank(C17Model)
```

Polygon 17 Validation Set

```
eval(parse(text = gsub("C","V17",sMLRtempEquNCoef)))
rank(V17Model)
```

Polygon 20 Calibration Set

```
eval(parse(text = gsub("C","C20",sMLRtempEquNCoef)))
rank(C20Model)
```

Polygon 20 Validation Set

```
eval(parse(text = gsub("C","V20",sMLRtempEquNCoef)))
rank(V20Model)
```

Polygon 21 Calibration Set

```
eval(parse(text = gsub("C","C21",sMLRtempEquNCoef)))
rank(C21Model)
```

Polygon 21 Validation Set

```
eval(parse(text = gsub("C","V21",sMLRtempEquNCoef)))
rank(V21Model)
```

Polygon 26 Calibration Set

```
eval(parse(text = gsub("C","C26",sMLRtempEquNCoef)))
rank(C26Model)
```

Polygon 26 Validation Set

```
eval(parse(text = gsub("C","V26",sMLRtempEquNCoef)))
rank(V26Model)
```

Polygon 27 Calibration Set

```
eval(parse(text = gsub("C","C27",sMLRtempEquNCoef)))
rank(C27Model)
```

Polygon 27 Validation Set

```
#Assessing the
#Compute Test Rank for
C17Model <-
RankofC17Model <-

#Compute Test Rank for
V17Model <-
RankofV17Model <-

#Compute Test Rank for
C20Model <-
RankofC20Model <-

#Compute Test Rank for
V20Model <-
RankofV20Model <-

#Compute Test Rank for
C21Model <-
RankofC21Model <-

#Compute Test Rank for
V21Model <-
RankofV21Model <-

#Compute Test Rank for
C26Model <-
RankofC26Model <-

#Compute Test Rank for
V26Model <-
RankofV26Model <-

#Compute Test Rank for
C27Model <-
RankofC27Model <-

#Compute Test Rank for
```

```
eval(parse(text = gsub("C","V27",sMLRtempEquNCoef)))
rank(V27Model)

Polygon 28 Calibration Set
eval(parse(text = gsub("C","C28",sMLRtempEquNCoef)))
rank(C28Model)

Polygon 28 Validation Set
eval(parse(text = gsub("C","V28",sMLRtempEquNCoef)))
rank(V28Model)

Polygon 2955 Calibration Set
eval(parse(text = gsub("C","C2955",sMLRtempEquNCoef)))
rank(C2955Model)

Polygon 2955 Validation Set
eval(parse(text = gsub("C","V2955",sMLRtempEquNCoef)))
rank(V2955Model)

Polygon 30 Calibration Set
eval(parse(text = gsub("C","C30",sMLRtempEquNCoef)))
rank(C30Model)

Polygon 30 Validation Set
eval(parse(text = gsub("C","V30",sMLRtempEquNCoef)))
rank(V30Model)

Polygon 32 Calibration Set
eval(parse(text = sMLRtempEquNCoef))
rank(C32Model)

Polygon 32 Validation Set
eval(parse(text = gsub("C","V",sMLRtempEquNCoef)))
rank(V32Model)

Polygon 33 Calibration Set
eval(parse(text = gsub("C","C33",sMLRtempEquNCoef)))
rank(C33Model)

Polygon 33 Validation Set

V27Model <-
RankofV27Model <-

#Compute Test Rank for
C28Model <-
RankofC28Model <-

#Compute Test Rank for
V28Model <-
RankofV28Model <-

#Compute Test Rank for
C2955Model <-
RankofC2955Model <-

#Compute Test Rank for
V2955Model <-
RankofV2955Model <-

#Compute Test Rank for
C30Model <-
RankofC30Model <-

#Compute Test Rank for
V30Model <-
RankofV30Model <-

#Compute Test Rank for
C32Model <-
RankofC32Model <-

#Compute Test Rank for
V32Model <-
RankofV32Model <-

#Compute Test Rank for
C33Model <-
RankofC33Model <-

#Compute Test Rank for
```

```
eval(parse(text = gsub("C","V33",sMLRtempEquNCoef)))
rank(V33Model)

Polygon 37 Calibration Set
eval(parse(text = gsub("C","C37",sMLRtempEquNCoef)))
rank(C37Model)

Polygon 37 Validation Set
eval(parse(text = gsub("C","V37",sMLRtempEquNCoef)))
rank(V37Model)

Polygon 38 Calibration Set
eval(parse(text = gsub("C","C38",sMLRtempEquNCoef)))
rank(C38Model)

Polygon 38 Validation Set
eval(parse(text = gsub("C","V38",sMLRtempEquNCoef)))
rank(V38Model)

Polygon 42 Calibration Set
eval(parse(text = gsub("C","C42",sMLRtempEquNCoef)))
rank(C42Model)

Polygon 42 Validation Set
eval(parse(text = gsub("C","V42",sMLRtempEquNCoef)))
rank(V42Model)

Polygon 48 Calibration Set
eval(parse(text = gsub("C","C48",sMLRtempEquNCoef)))
rank(C48Model)

Polygon 48 Validation Set
eval(parse(text = gsub("C","V48",sMLRtempEquNCoef)))
rank(V48Model)

Polygon 49 Calibration Set
eval(parse(text = gsub("C","C49",sMLRtempEquNCoef)))
rank(C49Model)

Polygon 49 Validation Set

V33Model <-
RankofV33Model <-

#Compute Test Rank for
C37Model <-
RankofC37Model <-

#Compute Test Rank for
V37Model <-
RankofV37Model <-

#Compute Test Rank for
C38Model <-
RankofC38Model <-

#Compute Test Rank for
V38Model <-
RankofV38Model <-

#Compute Test Rank for
C42Model <-
RankofC42Model <-

#Compute Test Rank for
V42Model <-
RankofV42Model <-

#Compute Test Rank for
C48Model <-
RankofC48Model <-

#Compute Test Rank for
V48Model <-
RankofV48Model <-

#Compute Test Rank for
C49Model <-
RankofC49Model <-

#Compute Test Rank for
```

```
                                                                    V49Model        <-
eval(parse(text = gsub("C","V49",sMLRtempEquNCoef)))
                                                                    RankofV49Model  <-
rank(V49Model)

                                                                    #Compute Test Rank for
Polygon 57 Calibration Set
                                                                    C57Model        <-
eval(parse(text = gsub("C","C57",sMLRtempEquNCoef)))
                                                                    RankofC57Model  <-
rank(C57Model)



                                                                    #Measure of model fit
                                                                    PercentC17      <-
((length(C17AvgTotal) - min(RankofC17Model[1],RankofC17Model[2],RankofC17Model[3],RankofC17Model[4])) -
3)/length(C17AvgTotal)
                                                                    PercentV17      <-
((length(V17AvgTotal) - min(RankofV17Model[1],RankofV17Model[2])) -1)/length(V17AvgTotal)
                                                                    PercentC20      <-
((length(C20AvgTotal) - min(RankofC20Model[1],RankofC20Model[2],RankofC20Model[3])) -2)/length(C20AvgTotal)
                                                                    PercentV20      <-
((length(V20AvgTotal) - min(RankofV20Model[1],RankofV20Model[2])) -1)/length(V20AvgTotal)
                                                                    PercentC21      <-
((length(C21AvgTotal) - min(RankofC21Model[1],RankofC21Model[2])) -1)/length(C21AvgTotal)
                                                                    PercentV21      <-
((length(V21AvgTotal) - min(RankofV21Model[1],RankofV21Model[2])) -1)/length(V21AvgTotal)
                                                                    PercentC26      <-
(length(C26AvgTotal) - RankofC26Model[1])/length(C26AvgTotal)
                                                                    PercentV26      <-
(length(V26AvgTotal) - RankofV26Model[1])/length(V26AvgTotal)
                                                                    PercentC27      <-
((length(C27AvgTotal) - min(RankofC27Model[1],RankofC27Model[2],RankofC27Model[3])) -2)/length(C27AvgTotal)
                                                                    PercentV27      <-
((length(V27AvgTotal) - min(RankofV27Model[1],RankofV27Model[2])) -1)/length(V27AvgTotal)
                                                                    PercentC28      <-
((length(C28AvgTotal) - min(RankofC28Model[1],RankofC28Model[2],RankofC28Model[3])) -2)/length(C28AvgTotal)
                                                                    PercentV28      <-
((length(V28AvgTotal) - min(RankofV28Model[1],RankofV28Model[2])) -1)/length(V28AvgTotal)
                                                                    PercentC2955    <-
((length(C2955AvgTotal)                                                             -
min(RankofC2955Model[1],RankofC2955Model[2],RankofC2955Model[3],RankofC2955Model[4]))                -
3)/length(C2955AvgTotal)
                                                                    PercentV2955    <-
((length(V2955AvgTotal) - min(RankofV2955Model[1],RankofV2955Model[2])) -1)/length(V2955AvgTotal)
                                                                    PercentC30      <-
((length(C30AvgTotal)                                                               -
min(RankofC30Model[1],RankofC30Model[2],RankofC30Model[3],RankofC30Model[4],RankofC30Model[5]))              -
4)/length(C30AvgTotal)
                                                                    PercentV30      <-
((length(V30AvgTotal) - min(RankofV30Model[1],RankofV30Model[2])) -1)/length(V30AvgTotal)
                                                                    PercentC32      <-
((length(CAvgTotal)                                                                 -
min(RankofC32Model[1],RankofC32Model[2],RankofC32Model[3],RankofC32Model[4],RankofC30Model[5]))              -
4)/length(CAvgTotal)
                                                                    PercentV32      <-
((length(VAvgTotal) - min(RankofV32Model[1],RankofV32Model[2])) -1)/length(VAvgTotal)
                                                                    PercentC33      <-
((length(C33AvgTotal) - min(RankofC33Model[1],RankofC33Model[2],RankofC33Model[3])) -2)/length(C33AvgTotal)
                                                                    PercentV33      <-
((length(V33AvgTotal) - min(RankofV33Model[1],RankofV33Model[2])) -1)/length(V33AvgTotal)
                                                                    PercentC37      <-
(length(C37AvgTotal) - RankofC37Model[1])/length(C37AvgTotal)
                                                                    PercentV37      <-
(length(V37AvgTotal) - RankofV37Model[1])/length(V37AvgTotal)
                                                                    PercentC38      <-
((length(C38AvgTotal) - min(RankofC38Model[1],RankofC38Model[2])) -1)/length(C38AvgTotal)
```

```
                                                                    PercentV38          <-
((length(V38AvgTotal) - min(RankofV38Model[1],RankofV38Model[2])) -1)/length(V38AvgTotal)
                                                                    PercentC42          <-
((length(C42AvgTotal) - min(RankofC42Model[1],RankofC42Model[2],RankofC42Model[3])) -2)/length(C42AvgTotal)
                                                                    PercentV42          <-
((length(V42AvgTotal) - min(RankofV42Model[1],RankofV42Model[2])) -1)/length(V42AvgTotal)
                                                                    PercentC48          <-
((length(C48AvgTotal) - min(RankofC48Model[1],RankofC48Model[2])) -1)/length(C48AvgTotal)
                                                                    PercentV48          <-
(length(V48AvgTotal) - RankofV48Model[1])/length(V48AvgTotal)
                                                                    PercentC49          <-
((length(C49AvgTotal) - min(RankofC49Model[1],RankofC49Model[2])) -1)/length(C49AvgTotal)
                                                                    PercentV49          <-
(length(V49AvgTotal) - RankofV49Model[1])/length(V49AvgTotal)
                                                                    PercentC57          <-
(length(C57AvgTotal) - RankofC57Model[1])/length(C57AvgTotal)

                                                                    PercentAll          <-
(length(C17AvgTotal)*PercentC17+
   length(V17AvgTotal)*PercentV17+
   length(C20AvgTotal)*PercentC20+
   length(V20AvgTotal)*PercentV20+
   length(C21AvgTotal)*PercentC21+
   length(V21AvgTotal)*PercentV21+
   length(C26AvgTotal)*PercentC26+
   length(V26AvgTotal)*PercentV26+
   length(C27AvgTotal)*PercentC27+
   length(V27AvgTotal)*PercentV27+
   length(C28AvgTotal)*PercentC28+
   length(V28AvgTotal)*PercentV28+
   length(C2955AvgTotal)*PercentC2955+
   length(V2955AvgTotal)*PercentV2955+
   length(C30AvgTotal)*PercentC30+
   length(V30AvgTotal)*PercentV30+
   length(CAvgTotal)*PercentC32+
   length(VAvgTotal)*PercentV32+
   length(C33AvgTotal)*PercentC33+
   length(V33AvgTotal)*PercentV33+
   length(C37AvgTotal)*PercentC37+
   length(V37AvgTotal)*PercentV37+
   length(C38AvgTotal)*PercentC38+
   length(V38AvgTotal)*PercentV38+
   length(C42AvgTotal)*PercentC42+
   length(V42AvgTotal)*PercentV42+
```

length(C48AvgTotal)*PercentC48+

length(V48AvgTotal)*PercentV48+

length(C49AvgTotal)*PercentC49+

length(V49AvgTotal)*PercentV49+

length(C57AvgTotal)*PercentC57)/(

length(C17AvgTotal)+

length(V17AvgTotal)+

length(C20AvgTotal)+

length(V20AvgTotal)+

length(C21AvgTotal)+

length(V21AvgTotal)+

length(C26AvgTotal)+

length(V26AvgTotal)+

length(C27AvgTotal)+

length(V27AvgTotal)+

length(C28AvgTotal)+

length(V28AvgTotal)+

length(C2955AvgTotal)+

length(V2955AvgTotal)+

length(C30AvgTotal)+

length(V30AvgTotal)+

length(CAvgTotal)+

length(VAvgTotal)+

length(C33AvgTotal)+

length(V33AvgTotal)+

length(C37AvgTotal)+

length(V37AvgTotal)+

length(C38AvgTotal)+

length(V38AvgTotal)+

length(C42AvgTotal)+

length(V42AvgTotal)+

length(C48AvgTotal)+

length(V48AvgTotal)+

length(C49AvgTotal)+

length(V49AvgTotal)+

```
    length(C57AvgTotal))

                                                                                #Test if this Formula is
worthy of Top100, if rank is better than 12345
                                                                                    #If worthy,
record Equ & rank data in Top100 table
                                                                                if(!is.na(PercentAll))

    if(PercentAll<max(Top100RankTable[,2]))
                                                                                    {

    #match(min(Top100RankTable[,2]),Top100RankTable[,2])    this is the line that needs to be overwritten


    Top100RankTable[match(max(Top100RankTable[,2]),Top100RankTable[,2]),1] <- gsub("C","",sMLRtempEquNCoef)

    Top100RankTable[match(max(Top100RankTable[,2]),Top100RankTable[,2]),2] <- PercentAll
                                                                                    }



                                                                                    equnumber               <-
equnumber+1
                                                                                    print(equnumber)
                                                                                }
                                                                                #End for o10
                                                                            }
                                                                            #End else
                                                                        }
                                                                        #End for n10
                                                                    }
                                                                    #End else
                                                                }
                                                                #End for m10
                                                            }
                                                            #End else
                                                        }
                                                        #End for l10
                                                    }
                                                    #End for k10
    }
    else
    {
                                                    #If there are no variables within parentheses...


                                                    #Perform stepwise regression with formulas
                                                    sMLR <- step(lm(as.formula(paste("CFloodings~",Formulas))))



                                                    #extract the stepwise regression equation
                                                    sMLRtempEqu             <-             substr(toString(formula(sMLR)),tail(which(strsplit(toString(formula(sMLR)),
")[[1]]==','),1)+2,nchar(toString(formula(sMLR))))

                                                    #extract the stepwise regression coefficients
                                                    sMLRcoeffs <- coefficients(sMLR)

                                                    #Finding the spaces in the equation which should be before and after each operation and place it into an
array
                                                    SpaceLoc <- unlist(gregexpr(" ",sMLRtempEqu),1)

                                                    switch(length(coefficients(sMLR)),
                                                            sMLRtempEquNCoef <- sMLRcoeffs[1],
                                                            sMLRtempEquNCoef <- paste(sMLRcoeffs[1],"+",sMLRcoeffs[2],"*",sMLRtempEqu),
                                                            sMLRtempEquNCoef                                                                    <-
paste(sMLRcoeffs[1],"+",sMLRcoeffs[2],"*",substr(sMLRtempEqu,1,tail(SpaceLoc,1)-
1),sMLRcoeffs[3],"*",substr(sMLRtempEqu,tail(SpaceLoc,1),nchar(sMLRtempEqu))),
```

```
                    sMLRtempEquNCoef                                                                    <-
paste(sMLRcoeffs[1],"+",sMLRcoeffs[2],"*",substr(sMLRtempEqu,1,SpaceLoc[length(SpaceLoc)-2]-
1),sMLRcoeffs[3],"*",substr(sMLRtempEqu,SpaceLoc[length(SpaceLoc)-2]+1,tail(SpaceLoc,1)-
1),sMLRcoeffs[4],"*",substr(sMLRtempEqu,tail(SpaceLoc,1),nchar(sMLRtempEqu))),
                    sMLRtempEquNCoef                                                                    <-
paste(sMLRcoeffs[1],"+",sMLRcoeffs[2],"*",substr(sMLRtempEqu,1,SpaceLoc[length(SpaceLoc)-4]-
1),sMLRcoeffs[3],"*",substr(sMLRtempEqu,SpaceLoc[length(SpaceLoc)-4]+1,SpaceLoc[length(SpaceLoc)-2]-
1),sMLRcoeffs[4],"*",substr(sMLRtempEqu,SpaceLoc[length(SpaceLoc)-2]+1,tail(SpaceLoc,1)-
1),sMLRcoeffs[5],"*",substr(sMLRtempEqu,tail(SpaceLoc,1),nchar(sMLRtempEqu))),
                    sMLRtempEquNCoef                                                                    <-
paste(sMLRcoeffs[1],"+",sMLRcoeffs[2],"*",substr(sMLRtempEqu,1,SpaceLoc[length(SpaceLoc)-6]-
1),sMLRcoeffs[3],"*",substr(sMLRtempEqu,SpaceLoc[length(SpaceLoc)-6]+1,SpaceLoc[length(SpaceLoc)-4]-
1),sMLRcoeffs[4],"*",substr(sMLRtempEqu,SpaceLoc[length(SpaceLoc)-4]+1,SpaceLoc[length(SpaceLoc)-2]-
1),sMLRcoeffs[5],"*",substr(sMLRtempEqu,SpaceLoc[length(SpaceLoc)-2]+1,tail(SpaceLoc,1)-
1),sMLRcoeffs[6],"*",substr(sMLRtempEqu,tail(SpaceLoc,1),nchar(sMLRtempEqu))),
                    )


        #Assessing the Calibration and Validation fit for all polygons
        #Compute Test Rank for Polygon 17 Calibration Set
        C17Model <- eval(parse(text = gsub("C","C17",sMLRtempEquNCoef)))
        RankofC17Model <- rank(C17Model)

        #Compute Test Rank for Polygon 17 Validation Set
        V17Model <- eval(parse(text = gsub("C","V17",sMLRtempEquNCoef)))
        RankofV17Model <- rank(V17Model)

        #Compute Test Rank for Polygon 20 Calibration Set
        C20Model <- eval(parse(text = gsub("C","C20",sMLRtempEquNCoef)))
        RankofC20Model <- rank(C20Model)

        #Compute Test Rank for Polygon 20 Validation Set
        V20Model <- eval(parse(text = gsub("C","V20",sMLRtempEquNCoef)))
        RankofV20Model <- rank(V20Model)

        #Compute Test Rank for Polygon 21 Calibration Set
        C21Model <- eval(parse(text = gsub("C","C21",sMLRtempEquNCoef)))
        RankofC21Model <- rank(C21Model)

        #Compute Test Rank for Polygon 21 Validation Set
        V21Model <- eval(parse(text = gsub("C","V21",sMLRtempEquNCoef)))
        RankofV21Model <- rank(V21Model)

        #Compute Test Rank for Polygon 26 Calibration Set
        C26Model <- eval(parse(text = gsub("C","C26",sMLRtempEquNCoef)))
        RankofC26Model <- rank(C26Model)

        #Compute Test Rank for Polygon 26 Validation Set
        V26Model <- eval(parse(text = gsub("C","V26",sMLRtempEquNCoef)))
        RankofV26Model <- rank(V26Model)


        #Compute Test Rank for Polygon 27 Calibration Set
        C27Model <- eval(parse(text = gsub("C","C27",sMLRtempEquNCoef)))
        RankofC27Model <- rank(C27Model)

        #Compute Test Rank for Polygon 27 Validation Set
        V27Model <- eval(parse(text = gsub("C","V27",sMLRtempEquNCoef)))
        RankofV27Model <- rank(V27Model)

        #Compute Test Rank for Polygon 28 Calibration Set
        C28Model <- eval(parse(text = gsub("C","C28",sMLRtempEquNCoef)))
        RankofC28Model <- rank(C28Model)

        #Compute Test Rank for Polygon 28 Validation Set
        V28Model <- eval(parse(text = gsub("C","V28",sMLRtempEquNCoef)))
        RankofV28Model <- rank(V28Model)

        #Compute Test Rank for Polygon 2955 Calibration Set
```

```
C2955Model <- eval(parse(text = gsub("C","C2955",sMLRtempEquNCoef)))
RankofC2955Model <- rank(C2955Model)

#Compute Test Rank for Polygon 2955 Validation Set
V2955Model <- eval(parse(text = gsub("C","V2955",sMLRtempEquNCoef)))
RankofV2955Model <- rank(V2955Model)

#Compute Test Rank for Polygon 30 Calibration Set
C30Model <- eval(parse(text = gsub("C","C30",sMLRtempEquNCoef)))
RankofC30Model <- rank(C30Model)

#Compute Test Rank for Polygon 30 Validation Set
V30Model <- eval(parse(text = gsub("C","V30",sMLRtempEquNCoef)))
RankofV30Model <- rank(V30Model)

#Compute Test Rank for Polygon 32 Calibration Set
C32Model <- eval(parse(text = sMLRtempEquNCoef))
RankofC32Model <- rank(C32Model)

#Compute Test Rank for Polygon 32 Validation Set
V32Model <- eval(parse(text = gsub("C","V",sMLRtempEquNCoef)))
RankofV32Model <- rank(V32Model)

#Compute Test Rank for Polygon 33 Calibration Set
C33Model <- eval(parse(text = gsub("C","C33",sMLRtempEquNCoef)))
RankofC33Model <- rank(C33Model)

#Compute Test Rank for Polygon 33 Validation Set
V33Model <- eval(parse(text = gsub("C","V33",sMLRtempEquNCoef)))
RankofV33Model <- rank(V33Model)

#Compute Test Rank for Polygon 37 Calibration Set
C37Model <- eval(parse(text = gsub("C","C37",sMLRtempEquNCoef)))
RankofC37Model <- rank(C37Model)

#Compute Test Rank for Polygon 37 Validation Set
V37Model <- eval(parse(text = gsub("C","V37",sMLRtempEquNCoef)))
RankofV37Model <- rank(V37Model)

#Compute Test Rank for Polygon 38 Calibration Set
C38Model <- eval(parse(text = gsub("C","C38",sMLRtempEquNCoef)))
RankofC38Model <- rank(C38Model)

#Compute Test Rank for Polygon 38 Validation Set
V38Model <- eval(parse(text = gsub("C","V38",sMLRtempEquNCoef)))
RankofV38Model <- rank(V38Model)

#Compute Test Rank for Polygon 42 Calibration Set
C42Model <- eval(parse(text = gsub("C","C42",sMLRtempEquNCoef)))
RankofC42Model <- rank(C42Model)

#Compute Test Rank for Polygon 42 Validation Set
V42Model <- eval(parse(text = gsub("C","V42",sMLRtempEquNCoef)))
RankofV42Model <- rank(V42Model)

#Compute Test Rank for Polygon 48 Calibration Set
C48Model <- eval(parse(text = gsub("C","C48",sMLRtempEquNCoef)))
RankofC48Model <- rank(C48Model)

#Compute Test Rank for Polygon 48 Validation Set
V48Model <- eval(parse(text = gsub("C","V48",sMLRtempEquNCoef)))
RankofV48Model <- rank(V48Model)

#Compute Test Rank for Polygon 49 Calibration Set
C49Model <- eval(parse(text = gsub("C","C49",sMLRtempEquNCoef)))
RankofC49Model <- rank(C49Model)

#Compute Test Rank for Polygon 49 Validation Set
V49Model <- eval(parse(text = gsub("C","V49",sMLRtempEquNCoef)))
RankofV49Model <- rank(V49Model)
```

```
#Compute Test Rank for Polygon 57 Calibration Set
C57Model <- eval(parse(text = gsub("C","C57",sMLRtempEquNCoef)))
RankofC57Model <- rank(C57Model)



#Measure of model fit
PercentC17 <- ((length(C17AvgTotal) - min(RankofC17Model[1],RankofC17Model[2],RankofC17Model[3],RankofC17Model[4])) -3)/length(C17AvgTotal)
PercentV17 <- ((length(V17AvgTotal) - min(RankofV17Model[1],RankofV17Model[2])) - 1)/length(V17AvgTotal)
PercentC20 <- ((length(C20AvgTotal) - min(RankofC20Model[1],RankofC20Model[2],RankofC20Model[3])) -2)/length(C20AvgTotal)
PercentV20 <- ((length(V20AvgTotal) - min(RankofV20Model[1],RankofV20Model[2])) - 1)/length(V20AvgTotal)
PercentC21 <- ((length(C21AvgTotal) - min(RankofC21Model[1],RankofC21Model[2])) - 1)/length(C21AvgTotal)
PercentV21 <- ((length(V21AvgTotal) - min(RankofV21Model[1],RankofV21Model[2])) - 1)/length(V21AvgTotal)
PercentC26 <- (length(C26AvgTotal) - RankofC26Model[1])/length(C26AvgTotal)
PercentV26 <- (length(V26AvgTotal) - RankofV26Model[1])/length(V26AvgTotal)
PercentC27 <- ((length(C27AvgTotal) - min(RankofC27Model[1],RankofC27Model[2],RankofC27Model[3])) -2)/length(C27AvgTotal)
PercentV27 <- ((length(V27AvgTotal) - min(RankofV27Model[1],RankofV27Model[2])) - 1)/length(V27AvgTotal)
PercentC28 <- ((length(C28AvgTotal) - min(RankofC28Model[1],RankofC28Model[2],RankofC28Model[3])) -2)/length(C28AvgTotal)
PercentV28 <- ((length(V28AvgTotal) - min(RankofV28Model[1],RankofV28Model[2])) - 1)/length(V28AvgTotal)
PercentC2955 <- ((length(C2955AvgTotal) - min(RankofC2955Model[1],RankofC2955Model[2],RankofC2955Model[3],RankofC2955Model[4])) - 3)/length(C2955AvgTotal)
PercentV2955 <- ((length(V2955AvgTotal) - min(RankofV2955Model[1],RankofV2955Model[2])) - 1)/length(V2955AvgTotal)
PercentC30 <- ((length(C30AvgTotal) - min(RankofC30Model[1],RankofC30Model[2],RankofC30Model[3],RankofC30Model[4],RankofC30Model[5])) - 4)/length(C30AvgTotal)
PercentV30 <- ((length(V30AvgTotal) - min(RankofV30Model[1],RankofV30Model[2])) - 1)/length(V30AvgTotal)
PercentC32 <- ((length(CAvgTotal) - min(RankofC32Model[1],RankofC32Model[2],RankofC32Model[3],RankofC32Model[4],RankofC30Model[5])) - 4)/length(CAvgTotal)
PercentV32 <- ((length(VAvgTotal) - min(RankofV32Model[1],RankofV32Model[2])) -1)/length(VAvgTotal)
PercentC33 <- ((length(C33AvgTotal) - min(RankofC33Model[1],RankofC33Model[2],RankofC33Model[3])) -2)/length(C33AvgTotal)
PercentV33 <- ((length(V33AvgTotal) - min(RankofV33Model[1],RankofV33Model[2])) - 1)/length(V33AvgTotal)
PercentC37 <- (length(C37AvgTotal) - RankofC37Model[1])/length(C37AvgTotal)
PercentV37 <- (length(V37AvgTotal) - RankofV37Model[1])/length(V37AvgTotal)
PercentC38 <- ((length(C38AvgTotal) - min(RankofC38Model[1],RankofC38Model[2])) - 1)/length(C38AvgTotal)
PercentV38 <- ((length(V38AvgTotal) - min(RankofV38Model[1],RankofV38Model[2])) - 1)/length(V38AvgTotal)
PercentC42 <- ((length(C42AvgTotal) - min(RankofC42Model[1],RankofC42Model[2],RankofC42Model[3])) -2)/length(C42AvgTotal)
PercentV42 <- ((length(V42AvgTotal) - min(RankofV42Model[1],RankofV42Model[2])) - 1)/length(V42AvgTotal)
PercentC48 <- ((length(C48AvgTotal) - min(RankofC48Model[1],RankofC48Model[2])) - 1)/length(C48AvgTotal)
PercentV48 <- (length(V48AvgTotal) - RankofV48Model[1])/length(V48AvgTotal)
PercentC49 <- ((length(C49AvgTotal) - min(RankofC49Model[1],RankofC49Model[2])) - 1)/length(C49AvgTotal)
PercentV49 <- (length(V49AvgTotal) - RankofV49Model[1])/length(V49AvgTotal)
PercentC57 <- (length(C57AvgTotal) - RankofC57Model[1])/length(C57AvgTotal)

PercentAll <- (length(C17AvgTotal)*PercentC17+
        length(V17AvgTotal)*PercentV17+
        length(C20AvgTotal)*PercentC20+
        length(V20AvgTotal)*PercentV20+
```

```
                length(C21AvgTotal)*PercentC21+
                length(V21AvgTotal)*PercentV21+
                length(C26AvgTotal)*PercentC26+
                length(V26AvgTotal)*PercentV26+
                length(C27AvgTotal)*PercentC27+
                length(V27AvgTotal)*PercentV27+
                length(C28AvgTotal)*PercentC28+
                length(V28AvgTotal)*PercentV28+
                length(C2955AvgTotal)*PercentC2955+
                length(V2955AvgTotal)*PercentV2955+
                length(C30AvgTotal)*PercentC30+
                length(V30AvgTotal)*PercentV30+
                length(CAvgTotal)*PercentC32+
                length(VAvgTotal)*PercentV32+
                length(C33AvgTotal)*PercentC33+
                length(V33AvgTotal)*PercentV33+
                length(C37AvgTotal)*PercentC37+
                length(V37AvgTotal)*PercentV37+
                length(C38AvgTotal)*PercentC38+
                length(V38AvgTotal)*PercentV38+
                length(C42AvgTotal)*PercentC42+
                length(V42AvgTotal)*PercentV42+
                length(C48AvgTotal)*PercentC48+
                length(V48AvgTotal)*PercentV48+
                length(C49AvgTotal)*PercentC49+
                length(V49AvgTotal)*PercentV49+
                length(C57AvgTotal)*PercentC57)/(
                length(C17AvgTotal)+
                length(V17AvgTotal)+
                length(C20AvgTotal)+
                length(V20AvgTotal)+
                length(C21AvgTotal)+
                length(V21AvgTotal)+
                length(C26AvgTotal)+
                length(V26AvgTotal)+
                length(C27AvgTotal)+
                length(V27AvgTotal)+
                length(C28AvgTotal)+
                length(V28AvgTotal)+
                length(C2955AvgTotal)+
                length(V2955AvgTotal)+
                length(C30AvgTotal)+
                length(V30AvgTotal)+
                length(CAvgTotal)+
                length(VAvgTotal)+
                length(C33AvgTotal)+
                length(V33AvgTotal)+
                length(C37AvgTotal)+
                length(V37AvgTotal)+
                length(C38AvgTotal)+
                length(V38AvgTotal)+
                length(C42AvgTotal)+
                length(V42AvgTotal)+
                length(C48AvgTotal)+
                length(V48AvgTotal)+
                length(C49AvgTotal)+
                length(V49AvgTotal)+
                length(C57AvgTotal))

        #Test if this Formula is worthy of Top100, if rank is better than 12345
                #If worthy, record Equ & rank data in Top100 table
        if(!is.na(PercentAll))
                if(PercentAll<max(Top100RankTable[,2]))
                {
                        #match(min(Top100RankTable[,2]),Top100RankTable[,2])    this is the line that needs
to be overwritten
                        Top100RankTable[match(max(Top100RankTable[,2]),Top100RankTable[,2]),1]      <-
gsub("C","",sMLRtempEquNCoef)
                        Top100RankTable[match(max(Top100RankTable[,2]),Top100RankTable[,2]),2]      <-
PercentAll
```

```
                                }


                        equnumber <- equnumber+1
                        print(equnumber)
                    }
                    #End Else

                #Export Top100 table to file
                write.csv(Top100RankTable, file = "C:/R/Thesis/PerPolygon/Top100RankTable - Overall Percentage -
                2013Jan24.csv",row.names = TRUE, col.names = TRUE)
                }
                #End for j10


#Export Top100 table to file
write.csv(Top100RankTable, file = "C:/R/Thesis/PerPolygon/Top100RankTable - Overall Percentage -
        2013Jan24.csv",row.names = TRUE, col.names = TRUE)
}
#End for i10
```

Combinations.csv

1	2	3	4	5
1	2	3	5	4
1	2	4	3	5
1	2	4	5	3
1	2	5	3	4
1	2	5	4	3
1	3	2	4	5
1	3	2	5	4
1	3	4	2	5
1	3	4	5	2
1	3	5	2	4
1	3	5	4	2
1	4	2	3	5
1	4	2	5	3
1	4	3	2	5
1	4	3	5	2
1	4	5	2	3
1	4	5	3	2
1	5	2	4	3
1	5	2	3	4
1	5	4	2	3
1	5	4	3	2
1	5	3	2	4
1	5	3	4	2
2	1	3	4	5
2	1	3	5	4
2	1	4	3	5
2	1	4	5	3
2	1	5	3	4
2	1	5	4	3
2	3	1	4	5
2	3	1	5	4
2	3	4	1	5
2	3	4	5	1
2	3	5	1	4
2	3	5	4	1
2	4	1	3	5
2	4	1	5	3
2	4	3	1	5
2	4	3	5	1
2	4	5	1	3
2	4	5	3	1
2	5	1	4	3
2	5	1	3	4

2	5	4	1	3
2	5	4	3	1
2	5	3	1	4
2	5	3	4	1
3	1	2	4	5
3	1	2	5	4
3	1	4	2	5
3	1	4	5	2
3	1	5	2	4
3	1	5	4	2
3	2	1	4	5
3	2	1	5	4
3	2	4	1	5
3	2	4	5	1
3	2	5	1	4
3	2	5	4	1
3	4	1	2	5
3	4	1	5	2
3	4	2	1	5
3	4	2	5	1
3	4	5	1	2
3	4	5	2	1
3	5	1	4	2
3	5	1	2	4
3	5	4	1	2
3	5	4	2	1
3	5	2	1	4
3	5	2	4	1
4	1	2	3	5
4	1	2	5	3
4	1	3	2	5
4	1	3	5	2
4	1	5	2	3
4	1	5	3	2
4	2	1	3	5
4	2	1	5	3
4	2	3	1	5
4	2	3	5	1
4	2	5	1	3
4	2	5	3	1
4	3	1	2	5
4	3	1	5	2
4	3	2	1	5
4	3	2	5	1
4	3	5	1	2
4	3	5	2	1
4	5	1	3	2

4	5	1	2	3
4	5	3	1	2
4	5	3	2	1
4	5	2	1	3
4	5	2	3	1
5	1	2	3	4
5	1	2	4	3
5	1	3	2	4
5	1	3	4	2
5	1	4	2	3
5	1	4	3	2
5	2	1	3	4
5	2	1	4	3
5	2	3	1	4
5	2	3	4	1
5	2	4	1	3
5	2	4	3	1
5	3	1	2	4
5	3	1	4	2
5	3	2	1	4
5	3	2	4	1
5	3	4	1	2
5	3	4	2	1
5	4	1	3	2
5	4	1	2	3
5	4	3	1	2
5	4	3	2	1
5	4	2	1	3
5	4	2	3	1