Ryerson University
Digital Commons @ Ryerson

Theses and dissertations

1-1-2002

Uncertainty and sensitivity analysis of GIS based continuous hydrological modelling

Harry R. Manson
Ryerson University

Follow this and additional works at: http://digitalcommons.ryerson.ca/dissertations
Part of the Environmental Engineering Commons

Recommended Citation
Manson, Harry R., "Uncertainty and sensitivity analysis of GIS based continuous hydrological modelling" (2002). *Theses and dissertations.* Paper 5.

This Thesis is brought to you for free and open access by Digital Commons @ Ryerson. It has been accepted for inclusion in Theses and dissertations by an authorized administrator of Digital Commons @ Ryerson. For more information, please contact bcameron@ryerson.ca.

NOTE TO USERS

This reproduction is the best copy available.

UMI®

UNCERTAINTY AND SENSITIVITY ANALYSIS OF GIS BASED CONTINUOUS HYDROLOGICAL MODELLING

By

Harry R. Manson

Honors Bachelor of Science
April 20, 2000
Wilfrid Laurier University
Waterloo Ontario Canada

Thesis Submitted to the Faculty of Graduate Studies Ryerson University in Partial Fulfillment of the Requirements for the Degree of

Masters of Applied Science and Management
In
Environmental Applied Science and Management

Dr. James Li
Dr. Doug Banting

September 30, 2003
Toronto Ontario Canada

Copyright © 2003 Harry R. Manson

National Library of Canada
Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services
Acquisisitons et services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-85322-5
Our file Notre référence
ISBN: 0-612-85322-5

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

Canadä

I AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis.

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

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

II BORROWER'S PAGE

Ryerson University requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

III ABSTRACT

. The impact of uncertainty in spatial and a-spatial lumped model parameters for a continuous rainfall-runoff model is evaluated with respect to model prediction. The model uses a modified SCS-Curve Number approach that is loosely coupled with a geographic information system (GIS). The rainfall-runoff model uses daily average inputs and is calibrated using a daily average streamflow record for the study site. A Monte Carlo analysis is used to identify total model uncertainty while sensitivity analysis is applied using both a one-at-a–time (OAT) approach as well as through application of the extended Fourier Amplitude Sensitivity Technique (FAST). Conclusions suggest that the model is highly sensitive to uncertainties associated with the initial abstraction estimates followed by model inputs and finally the Curve Number. While the model does not indicate a high degree of sensitivity to the Curve Number at present conditions, uncertainties in Curve Number estimation can potentially be the cause of high predictive errors when future development scenarios are evaluated. The author of this research is Harry Manson. This research is presented to the department of graduate studies at Ryerson University, Toronto Ontario Canada on September 30, 2003. This work is submitted as partial fulfillment for the degree of Masters of Applied Science in Environmental Science and Management

IV ACKNOWLEDGEMENTS

I would like to take this opportunity to thank the following individuals and organizations for financial and in-kind contributions to my research.

The Toronto and Regional Conservation Authority

Clarifica Consultants.

Wilderness Preservation Committee of Ontario

The National Science and Engineering Research Council (NSERC)

The members of my graduate committee including:

Dr. James Li
Dr. Doug Banting

Special thanks to Dr. R. Pushchak for aiding my transition to study at Ryerson University.

The following individuals for their continued support and encouragement:

Mr. Steven McComb
Mr. Scott B McCormick
Mr. Cory Smyth
Mrs. Melissa Stamplecoski

V DEDICATIONS

I would like to dedicate this work to my mother and father for, without their respective patience and guile I would not have been presented with the genetics capable of withstanding graduate school.

VII List of Tables

VIII List of Figures

Introduction

Often in science, it is necessary to study and comprehend systems that for a variety of reasons are too complex or impractical to effectively study in reality. This notion is especially true in the case of hydrological models and as a result, various stakeholders often utilize a modelling assisted approach in order to simplify systems of interest. While environmental models range in complexity and application, Singh (1996) notes that all hydrological modelling requires accurate and representative data, the manipulation of which is cumbersome largely because of the spatial nature of hydrological data, the large volumes of data needed to support effective modelling, as well as challenges surrounding data organization and management.

Within the last 15 years, rapid advances in computing power, now commonly available to the average user, have motivated the evolution of computer based modelling of the environment and one of the most significant advances in computer-assisted modelling is the development and application of geographic information systems (GIS). Fundamentally, GIS is a system for the collection, storage, manipulation, analyzing and display of spatial data. While basic GIS systems have been in existence for the last 30-35 years, it has been only in the last 10-15 years that GIS has evolved from its original emphasis of storing and displaying spatial data to the development of complex and sophisticated data manipulation and analysis. The push towards expanding the capabilities of GIS beyond a tool for convenient cartography has been emphasized by a relatively recent shift in the focus of the discipline from simple geographic information systems (GISystems) towards what has been identified as *geographic information science* (GIScience), (Goodchild, 1992), a topic explored extensively in the book "Geographic Information Systems and Science" (Longley et al., 2001).

The abilities of GIS to manage the challenges associated with computer based hydrological modelling effectively have been discussed by a number of authors (Bobba et al., 2000; Sui et al., 2000; Correia et al, 1998; Singh and Fiorentino, 1996; Singh, 1995), largely because GIS provides a powerful means for the management, comprehension, visualization and analysis of hydrologic problems. While the characteristics of watershed attributes such as soil, land use, slope and antecedent moisture conditions vary spatially in complex ways, this variation can be managed effectively within a GIS framework. As a result, issues surrounding data integration, compatibility and structure often significantly challenging for even simplified hydrological models become more manageable through the application of GIS. Singh, (1996) has reinforced this utility:

> *...because many models have different data requirements, a collection program tailored to the demand of a particular model cannot be used for another with different data requirements. Consequently a separate collection program has to be developed and data problems can be resolved through application of a GIS. Its ability to extract, overlay and delineate watershed characteristics permits integration with watershed models. Design calibration and modification and comparison of these models can be significantly facilitated by use of the GIS.*

Further, superior means of data visualization not readily achievable with traditional modelling approaches can be made use of within the GIS environment.

This use of GIS is also particularly suitable for investigating issues of scale associated with traditional hydrological modelling. Such issues are tied to aspects of heterogeneity within the model system, the degree of which varies with the geographical focus of the model. Geographic modelling of spatially varied information such as, precipitation, evaporation, land use, soil properties and soil moisture contents, etc. opens the door to more physically realistic approaches to simulate the hydrological balance of watersheds. Because traditional models tend to lump processes at the catchment or occasionally the sub-catchment scale, the degree of variation expressed in the model is limited by the degree of averaging of hydrological properties into homogenous units at that scale.

Currently both GIS users and hydrologists alike have begun to acknowledge the wide variety of benefits associated with integrating the two disciplines (Sui et al, 1999), which has been made largely apparent by the explosion of literature on the subject (DeVantier and Feldman, 1993; Maidment, 1996; McDonnell, 1996; Moore, 1996; Tsihrintzis et al., 1996; Correia et al., 1998; Sui et al., 1999). Integration of GIS with hydrological systems however poses certain challenges as well as benefits. In many cases the separate ontologies between GIS and traditional or mathematical hydrological conceptions of the world pose problems in the co-evolution of coupled GIS and hydrological applications.

All aspects of modelling from conception to application however, involve a simplification of the system under study and consequently the introduction of numerous sources of uncertainty cannot be avoided. Uncertainties in a model are the result of error within the model or deviation from reality and best management practices for traditional hydrological modelling suggest that uncertainty in model performance should be identified, understood, communicated and, where possible, reduced. While numerous studies and methods have been developed that manage uncertainty associated with hydrological modelling, the effects of uncertainty propagation in hydrological models based on GIS are not as widely understood and may have a significant impact on the reliability of GIS-based hydrological models.

One popular approach used to model rainfall-runoff relationships and the effects of future land use changes on a hydrological regime is in the application of the United States Natural Resources Conservation Service (NRCS) Curve Number method. At the center of this method is a series of Curve Numbers relating soil and land use conditions within a watershed to physical rainfall-runoff relationships. Once a suitable Curve Number is defined for a particular area, prediction of future development impacts rely in conversion tables, relating the current Curve Number to some future value based on a prediction of increases in impervious surface area. The Curve Number method is relatively popular because of its inherent simplicity of application. Further it is often used in conjunction with a GIS as relevant soil and land use data can easily be processed in this environment to derive either lumped or distributed Curve Number parameters. As a result, this method provides an excellent opportunity to study the impacts of uncertainty propagation in GIS-based hydrological modelling, and is the focus of this research.

In order to deal effectively with uncertainties derived from the application of GIS to hydrology, comprehension of how uncertainty may be introduced in these types of models is of significant importance. The scope of this research first and foremost is to examine the role of uncertainty propagation within the context of GIS, hydrological, and coupled models. In order to comprehend this, an understanding of hydrological and GIS model ontology is necessary as this investigates how each model deals with the simplification of reality and thus directly influences the ways in which uncertainty is introduced. Following this, various sources of potential uncertainty are identified, as are paths of error propagation.

Understanding of the role of uncertainty and its effect on model performance has been achieved in traditional models through the application of global model uncertainty analysis (UA) and parameters specific sensitivity analysis (SA). While uncertainty analysis can provide an indication of the overall uncertainty or tolerance of a models predictive capacity, parameters specific sensitivity analysis quantifies the degree of impact each factor of the underlying model contributes to the overall model performance. While GIS-based models have not typically been subjected to this type of analysis, application of UA in conjunction with SA can provide guidance on the utility of a GIS-based models performance and application of these methods to GIS-based hydrological models can provide a strong indication of the effects of uncertainties resulting from geographic data.

The overall scope of this research is to examine the role of uncertainty in spatial and a-spatial data and how this uncertainty impacts the reliability of GIS-based hydrological models. This research does not attempt to suggest a better model, but rather examines a modeling approach that is widely applied within the field of hydrology. Investigation of the implication of data uncertainty on model performance will take place is a series of steps. In Chapter 1 a general review is provided of basic modelling theory. This is first in the general sense and later in the chapter with specific emphasis on the various types of hydrological modelling. Chapter 2 provides the reader with a review of GIS and coupled models. Various challenges with respect to the evolution of the coupling of GIS with hydrological modelling is reviewed, as are concepts of best management practices applied to a GIS based hydrological model. In Chapter 3, a comprehensive discussion of uncertainty and error is provided. Topics reviewed include error, uncertainty, error propogation in modelling as well as the various types of uncertainty in spatial and a-spatial data. Various methodologies for identifying and quantifying sources of uncertainty in data will also be reviewed in Chapter 3. These techniques will be evaluated with respect to published applications and a suitable methodology for performing case study is presented. The case study in Chapter 4 focuses on the development of a lumped SCS Curve Number based hydrological model. Model parameters are derived through loosely coupling the mathematical model with a GIS. The uncertainty for a number of model factors is identified and used in a global uncertainty analysis. This is followed with a one-at –a time sensitivity analysis of model factors as well as a quantitative sensitivity test. The final sensitivity analysis provides a quantitative measure of the sensitivity of the model to each model factor and their n^{th} order interactions with other model factors. Finally in Chapter 5, results of the case study are evaluated with respect to model performance, global uncertainty of the model, and the sensitivity of the model to various model factors. Conclusions and recommendations regarding the results of the case study are presented in Chapter 6.

This section is intended to provide the reader with a working background of environmental modeling. The development of modelling best management practices (BMP's) as well as various modelling approaches and terminology is discussed. The role of Geographic Information Systems (GIS) in the field of hydrology is reviewed, as is the most common approaches to merging GIS with physical based hydrological models. The reader will finish the chapter with an understanding of environmental modelling as well as the role of physically based models, GIS and challenges facing the integration of the two approaches.

1.1 Environmental Modelling

Models are used in virtually every discipline, from physics to philosophy, for hypothesis formulation and testing, as heuristic tools for understanding system structure and function, and in assessing the effects of future control or management strategies (Bobba et al., 2000). While evolving separately, both hydrological modelling and GIS share many common elements application of these models to a decision-making process requires a comprehensive understanding of modelling tolerance and limitations.

The term *model* is generically considered to mean a conceptualization of reality. Despite an extensive diversity of applications and developments, models share many common, and one pivotal element in that they are all in some way devised to simplify a real world pattern or process too vast or complex to study or manipulate at natural scales and complexities. The subject of interest, for a modelling study is commonly referred to as a *system*. Within a system, an object of interest may be referred to as an *entity*, and an *attribute* is considered to be a property of an *entity*. Because a model is designed to simplify the study of a system, the user must weigh the benefits of creating a model simple enough to represent only those aspects of the system of interest but on the other hand, should be sufficiently detailed such that valid conclusions can be drawn about that system (Banks, 2000).

1.2 Model Classification

Models can be categorized into numerous classes and subclasses (see Figure 1.1). The primary distinction within model classification is made between *physical* and *mathematical* models. A physical model is a scaled down physical structure representing a subject of interest. Some examples may include an architect's building model. A map can also be considered a particular subclass of physical model defined as a graphic representation of the *milieu* or all aspects of the cultural and physical environment (Dent, 1996). Because maps have a distinct relationship with geographic information systems, and are the sources of much primary geographic data, they are discussed to a certain extent.

1.2.1 Mental Models

Although not often classified, *mental models,* are also an important class of modelling. The roots of model structure discussed in the next section are derived from the mental constructs of a real world system and hence mental models could theoretically be placed higher in the classification system than physical and mathematical models as both physical and mathematical models can be expressed as the result of mental models. For example, one famous example of a mental mathematical model is that of Einstein's Special Theory of Relativity or SPTOR (Einstein, 1920). Since technology at the time of the development of the SPTOR limited bench scale physical modelling of its principles, Einstein developed his principles according to what are now considered, "thought experiments". From a series of though experiments,

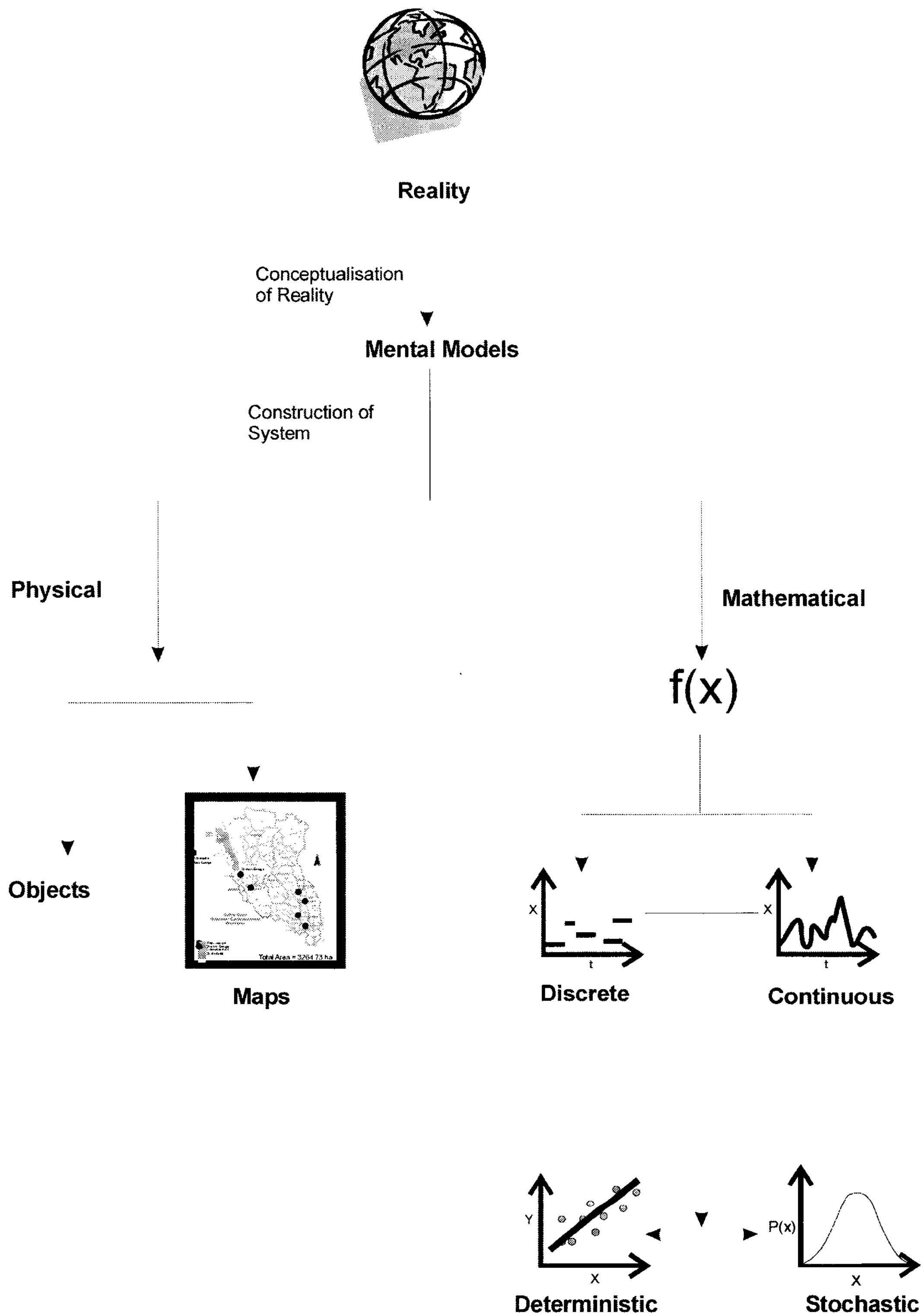


Figure 1.1. The creation of a model begins with a conceptualization of reality, Often described as a mental model, the manifestation of such may take on various classes and sub-classes

Einstein developed the SPTOR and later through mathematical models extended this to the General Theory of Relativity (Einstein, 1920).

1.2.2 Mathematical Models

The mathematical model is the primary means with which a scientist may investigate an environmental system. A mathematical model uses symbolic notation and mathematical equations to represent the physical processes of the system of interest. Usually hierarchical in its development, many mathematical models are complex enough to require relating a series of mathematical equations describing and integrating the behavior of various *components* or *factors* of a single system. In this respect, the results of computing a solution for one mathematical model are often used to drive another model.

The term *factors* refer to those aspects of the system that are described mathematically and are commonly categorized as *model structure*, *model attributes*, *model inputs*, *model variables*, *model parameters* and *model outputs*.

Model Structure

Lei, (1996) has defined model structure as a group of hypotheses consisting of a set of general laws $L_1, L_2, \ldots L_n$ and a set of statements pertaining to empirical circumstances $C_1, C_2, \ldots C_n$. Laws governing the dynamics of model structure include physical laws such as the laws of thermodynamics and gravity. As an example, a popular model structure used in hydrological modelling is the mass balance approach introduced by Thornthwaitee, (1944). The mass balance approach conceptualizes the water catchment as a series of inputs, outputs and storages (Figure 1.2). Each physical process conceptualized in the model structure is defined by a series of statements operating under a set of physical laws where:

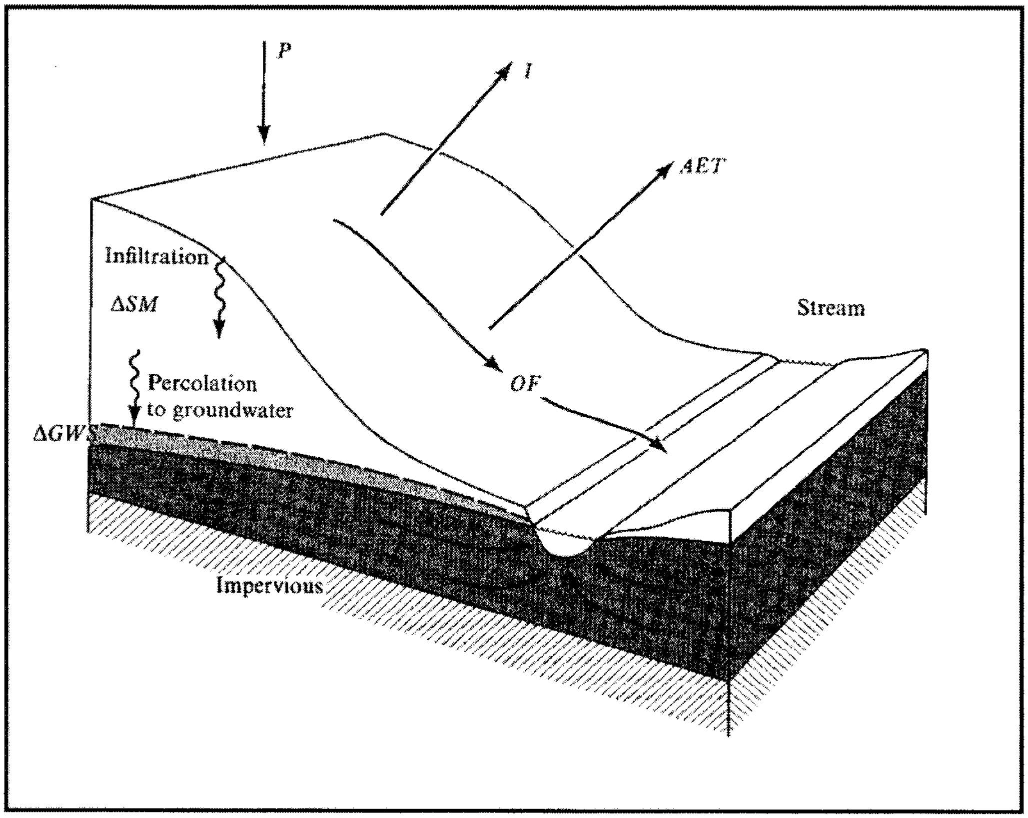


Figure 1.2. Thornthwaitee, (1944) conceptualized a rainfall-runoff relationship as a series of storages and linkages. Source: Dunne and Leopold, (1998).

$$OF = P - I - AET + \Delta GWS \tag{1.1}$$

and:

OF	=	**Overland Flow**
P	=	**Precipitation**
I	=	**Interception Loss**
AET	=	**Actual Evapotranspiration**
ΔGWS	=	**Change in Ground Water Storage**

As previously identified, mathematical models can be hierarchical especially when dealing with complex systems. In that respect, mathematical models are often composed of a series of hierarchical model structures. Various common mathematical approaches to water balance variables are vast and include but are not limited to general interception loss (Dunne et al., 1998); regional interception loss (Helvey and Patrice 1965), grass interception (Stoltenberg et al., 1950; Merriam, 1961; Lull, 1964; Crouse, 1966) Modelling of evaporation studies include those treated by the USGS (1954) and include the water budget method, as well as the energy budget method and the Blaney Criddle Formulae (USSC, 1970). Evaporation models are dependent on variables of solar radiation and have been addressed in the form of mean monthly solar radiation (Chang, 1968). The movement of groundwater is most often accounted for using Darcey's Law, methods concerning the determination of zones of recharge and discharge have been conducted by Meyboom (1962); and Toth (1966) and common methods for modelling groundwater infiltration include those of Horton, (1940) as well as Green and Ampt (1911).

Model Attributes

A model attribute can be described as a single or set of properties associated with a model entity. The number of entities in a system of interest may be related to the complexity of a model. For example, in the case of a single sub-watershed, at the macro scale the attributes used to describe the watershed may reflect area, maximum length, maximum width, average elevation, slope etc. Other attributes used to describe the watershed may relate to types of vegetation class such as agricultural, forest or wetland. Further, types of land use may also constitute attributes of the watershed and include categories such as urban, rural, industrial, etc. The number and types of attributes as well as entities within a system are a function of the complexity of the model as well as the complexity of the system at a given scale of interest.

The set of attributes for a given entity can be described as the vector of $\boldsymbol{e}$ entity attributes where:

$$\vec{\boldsymbol{a}} = [\, \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots \boldsymbol{a}_e] \tag{1.2}$$

Such that $\vec{a} \in \vec{A}$ and $\vec{A}$ is the vector of s system attributes described as:

$$\vec{A} = \{A_1, A_2, \ldots A_s\} \tag{1.3}$$

Model Inputs

Model inputs can be described as external variables that drive the system and are usually defined according to a time series. Because a model may have more than a single input, model inputs are often described as a vector denoting a series of n model inputs where:

$$\vec{I}_{(t)} = [\, I_{1(t)}, I_{2(t)}, \ldots I_{n(t)} \,] \tag{1.4}$$

Typical examples of model inputs within the context of environmental models included temperature, precipitation and evaporation.

Model Variables

Model variables, sometimes referred to as system state variables; refer to physical properties of a system that may vary within the context of the system and for each discrete time step. Further, a model variable can be measured directly and can also be denoted in vector form for a set of model variables where:

$$\vec{X}_{(t)} = [X_{1(t)}, X_{2(t)}, \ldots X_{n(t)}] \tag{1.5}$$

Model Parameters

Model parameters are typically coefficients that appear in mathematical functions and differ from model variables insofar as they are not directly measurable. A model parameter also does not typically vary with respect to time. Further, parameters are ideally statistically independent of one another. In cases where parameters are not statistically independent, the model structure is said to be poor (Lei, 1996). The set of model parameters can be described in vector notation as

$$\vec{P}_{(t)} = [P_{1(t)}, P_{2(t)}, \ldots P_{n(t)}] \tag{1.6}$$

Model Outputs

The model structure, attributes, variables and parameters all contribute to the model output. The degree to which each one of the system components contribute to the model output relates to the sensitivity of the model components as well as the model structure. The temporal resolution of a model output is often normally a function of the temporal resolution of the model inputs. A vector of model outputs in time series can be described as:

$$\vec{O}(t) = [O_{1(t)}, O_{2(t)}, \ldots O_{(t)n}] \tag{1.7}$$

Models may further be classified as either *discrete* or *continuous*. A discrete model is one in which the system state of the model is examined at only discrete steps in time, while continuous modelling describes a system whose state varies continuously over a period of time of a given resolution. The resolution of time intervals in a continuous model will depend on the application of the model and the scale of the system of interest. Most modelling in practice however is neither fully discrete nor fully continuous (Banks et al., 2000; Law and Kelton 2000). Whether a model is discrete or continuous depends largely on the time interval under which the phenomenon of interest is being investigated. In some cases, a phenomenon, which can be easily conceptualized as continuous such as streamflow, will behave more and more discretely given a coarse enough temporal resolution. Precipitation on the other hand, could be easily modeled as a series of discrete events and is often modeled as a continuous phenomenon again depending on the temporal resolution of the investigation.

An environmental model may also be further classified as either *deterministic* or *stochastic*. A deterministic model is one that possesses a known set of model inputs resulting in a set of unique outputs while stochastic modelling possesses some model inputs which are random in nature, having stochastic properties and typically are estimates of the true values of inputs based on inferential statistics, thus resulting in a set of model outputs that can only be considered estimates of the true characteristics of the system (Banks et al., 2000)

1.3 Hydrological Modelling

Application of mathematical modelling for an extensive number of water related issues has been in use over the last 50 years (Bobba et al, 2000; Freidman, et al 1984) while notions of modelling tools at the local and regional watershed scales have been presented in what Newson, (1998) refers to as a "modern" context for the last 300 years. Further, one historical account by Nace, (1974) puts hydrological prediction and management within a 3-6000 year context thus, application of mathematical models to hydrology is not a new science the applications of which have been documented by Singh (1996; 1989; 1988).

Singh, (1996) proposes a classification scheme for hydrological models reflecting five characteristics common to all watershed models including *watershed characteristics*, *initial system state*, and the model's *governing equations* and the model *inputs* and *outputs*. Each of these elements can be directly compared to elements found in all mathematical models including model structure and model parameters, the system state or model variables, model inputs, model outputs and model system attributes (see Figure 1.3). The governing equations of the hydrological model relate specifically to the model structure and parameters while the initial system state of the watershed can be described by system state variables and model parameters at the onset of the modelling scenario. Watershed characteristics can be described by the set of system attributes and the relationship between model inputs and outputs and this likewise has been used to classify models in the hydrological sense.

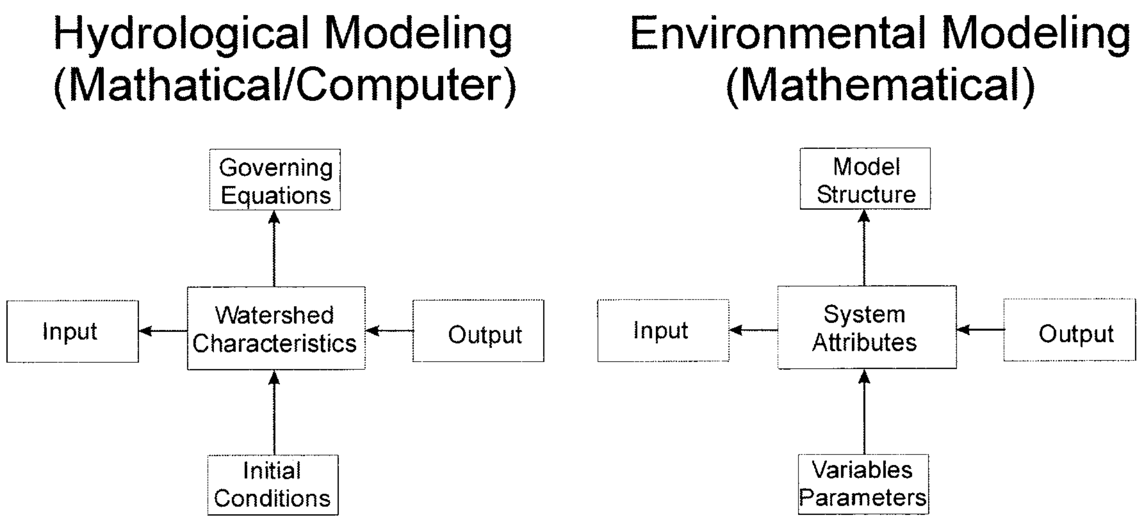


Figure 1.3. Adapted from Singh (1995), the classification scheme used to group hydrological models is strongly related to model factors in mathematical environmental modelling,

1.3.1 Models Classified by Process

Models classified by process are described as being *lumped*, *distributed*, or *quasi* (semi) *distributed.* A model is said to be lumped if takes into account no spatial variability in the watershed process. The lumped approach is typical of numerous models because of its simplicity in capturing and utilizing only dominant watershed characteristics. Further lumped models typically are less computationally demanding as the models will depend on a smaller set of attributes, inputs, and variables. For example, models using the SCS curve approach will be dependant on Curve Numbers, which relate rainfall to runoff transformation by a set of numbers that correspond to various soil and land-use combinations. In the case of a lumped model, a single Curve Number is used to describe the dominant characteristics of the entire watershed. Lumped models, however, are commonly criticized for over generalizing the watershed characteristics, possibly leading to poor model performance.

In 1973, O'Neil expressed seminal findings that error in predictions from modelling should decrease with a decreasing degree of model aggregation. *Distributed* modelling, as opposed to lumped modelling explicitly attempts to capture spatial variation in model inputs, variables, and attributes of the system of interest through the representation of the system as a series of locally representative homogenous units. These units can be in the form of a continuous tessellation of cells such as grid squares or they can be of variable geometry such as division by sub-watersheds or some other meaningful land-parceling scheme.

While in theory, distributed models preserve a higher amount of information about the system of interest by avoiding generalization, in reality a lack of field or laboratory data prevents such a general formulation of distributed models (Singh 1996; Vachaud et al., 2002). In these cases, the terms *mixed*, *semi-distributed* or *quasi-distributed* are often used. The term distributed can also be misleading as a model is only as distributed as its smallest homogenous unit and most often, models which claim to be distributed only incorporate distributed elements such as soil type or land use while still incorporating a lumped input scheme. Dependent upon the scale of the units, a model can behave more like a lumped model or more like a distributed model although because the resolution of the land parcels is feasibly finite such is the degree to which a model will be distributed. Appropriate scales for grid elements in distributed modelling have been investigated by Wood et al (1990; 1988).

The performance of mixed models is varied. In a study by Carpenter et al, (2001), a typical lumped modelling approach was examined in combination with NEXRAD distributed radar-rainfall data. The results of the study indicated that the NEXRAD distributed precipitation produced data that was no more accurate, or reliable than traditional lumped precipitation input using area-weighted rain gauge data. Obled et al, (1994) found that the hydrological model TOPMODEL, was sensitive to the changes in input volume from distributed versus lumped rainfall inputs, but that the particular model was insensitive to spatial patterns of rainfall. Shah et al., (1996), has experimented with distributed rainfall input versus that of spatially averaged rainfall and found that in cases where antecedent conditions and wet-lumped rainfall produced comparable results to that of distributed modelling provided at least one rain gauge lies within the catchment. For antecedent conditions that are dry, however, better results are found with distributed rainfall data and the spatial pattern of rainfall is found to be an important factor. It was further noted that this comparability might lie in the ability to compensate for inaccuracy in model prediction through the calibration of the model.

1.3.2 Models Classified by Scale

Watershed models in this class can be categorized by temporal or spatial scale (see Figure 1.4). In the temporal case, a model may be described as continuous or discrete (event-based) as is described in the first section. Singh (1996), makes a further distinction by separating classification schemes based firstly on temporal resolution of input and processing and secondly on the temporal resolution of the model output. In this case, a model may be either set to perform for daily, monthly, seasonally or yearly time scales. Models can also be set to shorter time intervals such as hourly or by the minute. Typically models that are continuous, and of short time intervals are more difficult to calibrate and often produce reasonable results only for longer periods of time.

The issue of appropriate temporal scales for hydrological modelling has been discussed by a number of investigators. Many authors have demonstrated that appropriate temporal resolution for continuous rainfall-runoff simulation falls between 1-5 minute intervals (Niemczynowics, 1984; Gujer and Krejci, 1988; Eicher, 1990; Zhu and Schilling 1996; Burckhardt-Gammeter and Fankhauser, 1998.) Further, it is likely that temporal resolution of 5-minute intervals will become the widely accepted paradigm over the next decade (Ostrowski, 2000). Despite this paradigm shift however, the vast majority of hydrologic data available to modelers fails to meet the temporal resolution standards of 5-minute intervals. As a result of this, numerous authors have examined the possibility of developing appropriate disaggregation procedures for larger time interval data or synthetic time series generation; see reviews by Burckhardt-Gammeter and Fankhauser, (1998) and Burian et al (2001). Despite the move towards a fixed time scale for hydrological modelling, it is widely understood that natural variability in the temporal scales of rainfall as well as spatial scales have a strong effect on the runoff produced for a system (Singh, 1997) and therefore, modelling with a fixed temporal scale may be inappropriate for some circumstances.

Models can also be classified based on spatial scale. Singh (1996), identified three classes of small (micro), medium (meso) and large (macro) being of ≤ $100km^2$, between 100 – $1000km^2$, and >$1000km^2$, respectively. Song and James (1992), reviewed models according to five scales of classification including the laboratory or bench scale, the hillslope scale, the catchment scale, basin scale and, the continental/global scale, while Ostrowski (2002), classified models according to hydrologic scales and corresponding geographic definitions and includes a relevant temporal context (see Figure 1.4).

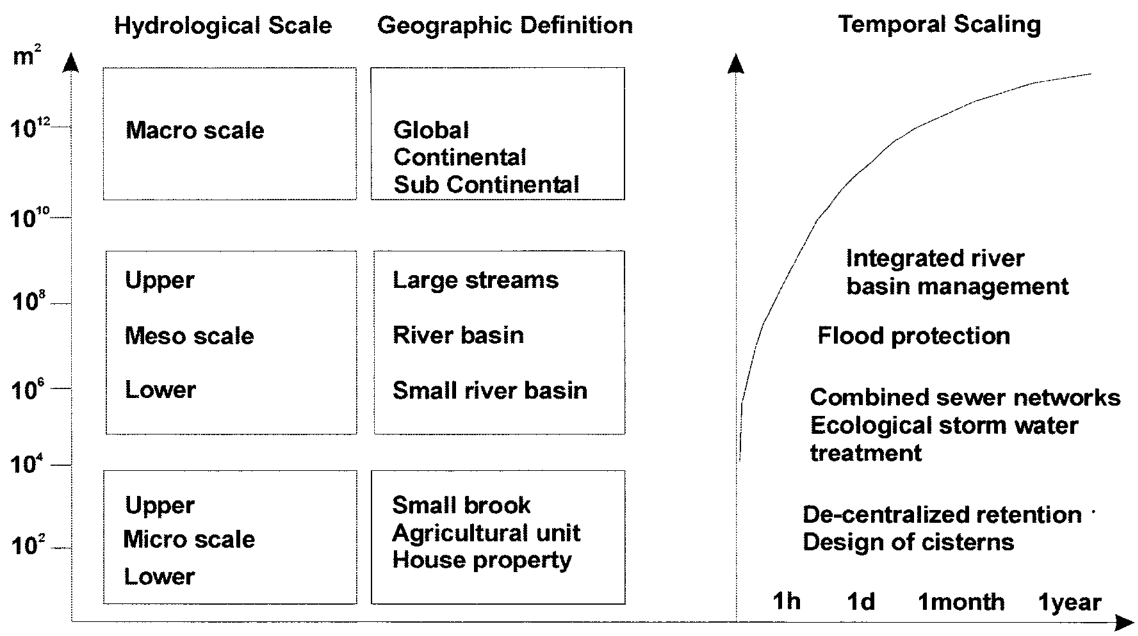


Figure 1.4. Hydrological models can be classified according to various spatial and temporal scales. Source: Ostrowski, (2002).

There is an abundance of literature available for the interested reader dealing with hydrological modelling at specific scales. Models performing at the micro scale include RENSIM and SWIFT Ostrowski, (2002) which operate at scales of 10^2 m and are geared towards rainfall-runoff for single house properties and small ecological storm water management schemes respectively. At the meso scale, examples of models for natural environments are plentiful and include RHESsys (Tague et al, 2001), and ANSWERS (Vachaud et al, 2002), but also exist for anthropogenic environments such as the urban systems models URBAN (Rodriguez et al, 2000), WBrM and TALSIM (Ostrowski, 2000). Some examples of models geared towards regional scales include SiSPAT (Boulet, 1999; Braud et al., 1995). Many hydrological models developed for macro or global scales are used in conjunction with what are referred to as Global Circulation or Climate models (GCM's), or Soil Vegetation Atmosphere Transfer (SVAT) models. For reviews of these applications the reader should see Kabat et al, (1997) and Liang and Xie, (2001).

1.4 Appropriate Model Scale

Ostrowski (2002), discusses a number of factors affecting the scale of a model including the scope of investigation and system homogeneity. While somewhat intuitive, the notion that a macro scale system may require a macro scale model may be redundant. While the scope of any specific investigation will likely define the spatial scale of the model in question, models are often created and tested on smaller catchments with the intention of being applied to larger systems. In cases where models that are developed for small areas are applied over large areas, a lack of available data at comparable detail may result. This, in turn, almost invariably means that the amount of information available to run the model is very much less than the ideal (Vachaud et al., 2002), resulting in poorer model performance and a greater emphasis on model calibration to achieve desirable results.

The homogeneity within a system refers to variability of spatial characteristics of the system attributes. Depending on the scale of observation of a system, variability within the system may appear as homogenous or heterogeneous. The concept of using a distributed versus a lumped approach to modelling systems was also discussed in preceding sections. While the lumped model approach has a model structure with parameters and variables representative of an averaged set of system attributes, the distributed model is broken up into a series of continuous homogenous subunits which attempt to capture the spatial variability of attributes within the system. These units, in the most ideal case, would show little or no natural within-unit variability of attributes consisting of uniform combinations of land use and soil conditions (Ostrowski, 2002). However, in reality, the finer the resolution the more heterogeneity is typically apparent, a concept described by the phenomenon based on fractal theory (Mandelbrot, 1983).

Given a hypothetical feasibility of infinite resolution, one of the disadvantages to distributed modelling is easily recognized as, for a given distributed model with infinite resolution capability, the number of hydrological response units will be representative of infinite numbers of attribute combinations. Therefore, such a model would require an infinite set of parameters and variables to describe the system attribute variation. This limitation defined largely by computing power and methods of field investigation and validation leads to the necessity for a quasi-distributed approach where some averaging is inevitable. The scale of the model should, therefore, be chosen in such a way as to maximize naturally definable homogeneity while balancing the need for detailed information against a finite availability of computing power. The scale should also be carefully considered because neglecting the issue of spatial variation of attributes through inadequate averaging can result in a loss of spatial detail and possibly a loss of hydrological extremes in distributed simulation (Ostrowski, 2002).

1.5 Best Management Practices

Because of the wide use, development and application of hydrological models, the engineering and scientific community has adopted a series of Best Management Practices or BMP's for the application of modelling to management and decision support. These BMP's are considered to be an evolved code of conduct with respect to the development and application of modelling in modern planning and have been iterated in many forms throughout the literature while maintaining a consistent premise. One comprehensive outline of BMP's for hydrological modelling encountered are those determined by the U.S. Office of Technology Assessment (OTA) in 1982 and include the following statements and guidelines interspersed with further considerations:

- *Models are often the most available alternative for analyzing complex resource problems.*
- *Models have the potential to provide even greater benefits for future water resource decision-making*
- *Water resource models vary greatly in their capabilities and limitations and must be carefully selected and applied by knowledgeable professionals. Selecting appropriate procedures for analyzing a particular problem is a major part of the modelling activity*

- *Mathematical models should be based on a fundamental representation of physical mechanisms and incorporate, to the extent possible state of the art understanding of the problem.*

- *Selection of mathematical models for regulatory applications requires a thorough understanding of the capabilities and limitations of the available models.*

- *Proper development of models requires modellers that are well trained in the underlying physical principles of the environmental system as well as in the computational procedures for modelling*

- *Models used for regulation applications must be subject to a two stage confirmation with field data consisting of 1) Calibration, where the parameters of the model are estimated to allow a good match between the model predictions and an observed field data set; and 2) verification where the model is compared to an independent data set without further modifying the parameter values and relationships in the model.*

- *Comparison of the model predictions and observed field data as part of the confirmation process should include qualitative graphical comparisons and if appropriate quantitative measures of the goodness of fit.*

- *Sensitivity and uncertainty analysis of environmental models should be performed to provide decision makers with an understanding of the level of confidence in model predictions and to identify key areas for future study.*

- *The models initial test applications and the modelling applications should be peer reviewed.*

Further, Bobba et al., (2000), in a review of model application to a variety of hydrological systems states:

> *The art of modelling lies in determining which watershed and surface water processes and data are essential for inclusion in the model. The challenge is to develop that art into rigorous scientific methodology suitable for the assessment...However, sophisticated models are not always the necessity. Frequently, the simpler the approach the better.*

The notion that simplistic models are often found to have superior performance to complex models has also been stated in an earlier review by Jackson, (1975). Further, simpler models consisting of less complex model structure often require fewer parameters and variables and, therefore, there is typically less uncertainty inherent within the model as all parameters and variables bring about some degree of uncertainty as absolute values cannot be known to within absolute accuracy and precision.

The use of comprehensive uncertainty analysis (UA) and sensitivity analysis (SA) is a modelling practice that is not often applied to modelling applications. This is largely because it can be a fairly arduous task and one that may result in the necessity to expend more resources in order to build confidence in the model. The topic of SA and UA are explored in greater detail in later parts of this thesis.

Until recently, geographic information systems (GIS) have evolved completely separately from the fields of hydrology or hydrological modelling. Currently, there is a considerable amount of interest in bridging the gap between these two disciplines, a gap resulting from the evolution of two separate conceptualizations not sharing a common ontology (Sui et al, 1999). The topics of this chapter include an introduction to the conceptualization of GIS model structure as well as a discussion of the benefits and challenges posed in integrating GIS and hydrological models.

2.1 GIS Defined

While the components and capabilities of a GIS may vary to some degree from one application to another, most GIS's possess certain fixed or required elements described by Marble (1984) and including:

- *A data input subsystem, which collects and/or processes spatial data derived from existing maps, remote sensors, etc. The data input is usually accomplished using digitizers, scanners or manual encoding of geographic information grid cells, points, lines or polygons.*

- *A data storage and retrieval subsystem which organizes the spatial data in a form that permits it to be quickly retrieved by the user for subsequent analysis, as well as allows for rapid and accurate updated and corrections to be made to the spatial database. Typical directories include: land cover, soils imagery, topography and water information.*

- *A data manipulation and analysis subsystem which converts data through user-defined aggregation rules, or produces estimates of parameters and constraints for various space-time optimizations of simulation models.*

- *A data reporting subsystem, which displays all, or part of the original database, as well as manipulated data, and the output from spatial models in tabular or map form.*

The heart of a GIS is in its computer platform. Typically three types of computer platforms are used to run a GIS including in chronological order of development (Tsihrintzis et al., 1996), mainframes, personal microcomputers (PCs), and workstations.

While this distinction may be true from a historical development perspective, the differences between the capabilities of a stand alone PC and what has been deemed the GIS "workstation" have been blurred. Most of today's PCs are exponentially more powerful than those available to the common user in the early 1980's and come minimally equipped with enough processing and storage power to drive all but the most powerful and demanding of GIS applications. In this respect, perhaps it is more useful to make a distinction between the PC, being the most common platform and capable of easily managing any "*personal geodatabase*" while the mainframe system, still in use for large so called "*enterprise spatial databases*" consisting largely of decades worth of attribute data and complex processing algorithms.

While the distinction between what has become known as a personal versus an enterprise geodatabase is somewhat vague, the distinction may be classified based on the particular database management system engine. Personal geodatabases are typically smaller and perform well through the use of less powerful database engines such as Microsoft Access, while enterprise databases are large, have many interrelationships, multiple concurrent or individual users, possess dedicated data management and validation managers, and are typically housed in more robust database software packages such as Oracle, dBase or Microsoft Sequel Server.

2.2 GIS Data Structure

Data within a GIS are stored in what is known as a geodatabase or a spatial database management system (SDBMS). The SDBMS has a number of features that make it attractive for use in GIS applications. These have been described by Longley et al, (2001) and outlined initially by Date (1995) as:

- *Collecting all data at a single location reduces redundancy and duplication.*
- *Maintenance costs decrease because of better organization and decreased data duplication*
- *Applications become data independent so that multiple applications can use the same data and they can evolve separately over time.*
- *User knowledge can be transferred between applications more easily because the database remains constant.*

- *Data sharing is facilitated and a corporate view of data can be provided to all managers and users.*
- *Security and standards for data and data access can be established and enforced.*

A DBMS can be classified according to the way in which it stores and manipulates data (Longley et al., 2001) and typically in a GIS consists of one of three formats including the relational database management system (RDBMS), the object database management system (ODBMS), and the object-relational database management system (ORDBMS).

The RDBMS is the most popular data structure model used in conjunction with GISystems. Typically, relational databases consist of one or many two dimensional *arrays* within which attributes are organized as a series of *fields* (columns). The fields within an array are typically classified or grouped in various tables according to a hierarchy of system attributes. Each array also contains a list of *records* (rows), and each record represents an *entity* within the system of interest.

The ODBMS evolved out of RDBMS largely to compensate for the fact the RDBMS was not designed to store complex objects such as geographic relationships, sounds, video or other complex media. The ORDBMS largely evolved out of the need for RDMS vendors to address the issues of limitations in standard RDBMS's. As a result, RDBMS's have in many cases have been equipped to handle geographical objects and other media through the evolution of data storage and management algorithms.

2.3 GIS Model Structure

The physical model structure or *spatial structure*, is similar to any model structure in that it represents a set of relationships that describe or conceptualize reality. In the case of a GIS, reality is inherently spatial and model structures within a GIS conceptualize or simplify special relationships in a form that can be easily encoded into a digital format.

Within GIS there are two commonly accepted model structures, raster and vector, each with separate advantages and disadvantages dependent upon the context of application. While the two model structures vary considerably, most conventional GIS packages are capable of processing spatial data in both forms although integration is often slow and takes large amounts of processing that is subject to many sources of error and, consequently, uncertainty.

2.3.1 Raster Based GIS

In the *raster* model structure, also commonly referred to as the field model, space is conceptualized as a continuous tessellation of cells. While normally square, occasionally raster fields are stored as hexagonal grids or some other geometrically and optimally packable shape. Spatial variation within raster-based modelling is conceived through the assignment of numerical values to each cell in the tessellation and each cell is of a fixed scale or dimension for example 10 m^2 or 10m x 10m, which defines its scale or resolution (Figure 2.1).

A raster model may represent more than one theme in which case each cell having a fixed location in space can have one or more cell values each belonging to mutually exclusive and exhaustive classes and each representing a separate theme within the GIS data model. As an example, a single cell for a raster model may represent an elevation in one theme, a soil type in another and a land use in a third theme. In the case of the land use class, the given cell number may be ordinal and represent one in a set of mutually exclusive and exhaustive numbers such that:

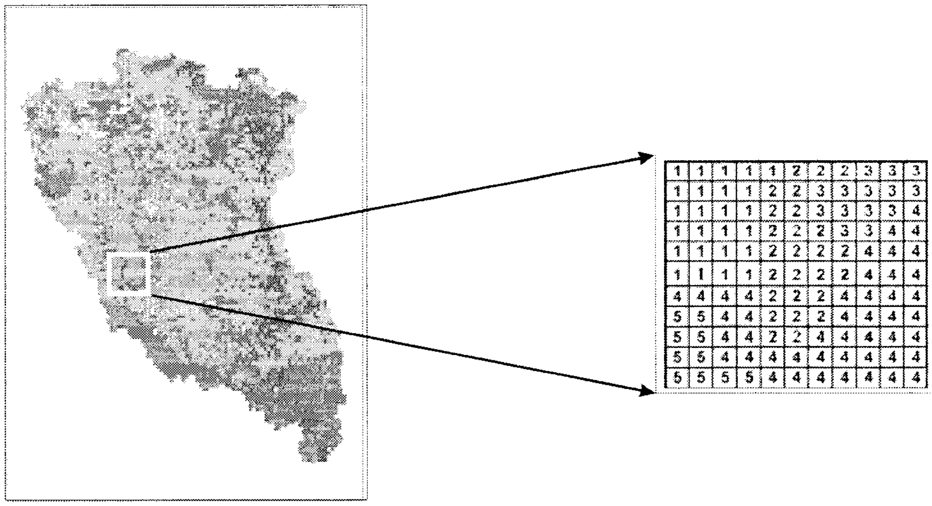


Figure 2.1. A 100 m^2 resolution raster model for the Duffin's Creek watershed depicting land use classes

$$X_{ij} = \mathbf{n} \tag{2.1}$$

and $$n \in \{1,2\ldots N\}$$

where X represents a given cell and i and j represent the row and column of the cell respectively.
n represents the numerical value of the cell and
where possible values and n are whole numbers such that $1 \leq n \leq N$

If a numeric integer field represents a soil classification and, we are modelling a soil system based on the Soil Conservation Service (SCS) HSG soil grouping criteria, then each cell will be classified according to four possible values where, 1 may represent soil type "**A**", 2 type "**B**", 3 type "**C**" and 4, type "**D**". Likewise, a number of discrete land use classes also identified by the SCS could be represented uniquely in a separate theme such that 1 may represent "Permanent Meadow" , 2 "Forested Wetland", 3 "Urban Open Space" etc until all classes in each theme are uniquely identified by a numerical value.

In some cases, however, classes may not be so discrete and may represent a range of values. Data may be ordinal or quantitative in nature and be associated with one of a set of mutually exclusive classes each belonging to a single theme and each representing a range of possible values. In a third case, say elevation for example, data may be numerically continuous, and as such any real number within a range of values may exist.

Raster formats have a number of advantages and disadvantages. One major advantage is that a large amount of digital data are gathered in raster format. For example, remote sensing satellite images are collected and stored in raster format, as are digital aerial photos, or a scanned topographic map. Further, because of the quantitative nature of raster images where each cell contains numerical information, they are considered infinitely more computationally powerful and can be treated as a large matrix with the number of cells equal to the number of rows multiplied by the number of columns. This structure allows for more advanced processing and manipulation capabilities than the vector object model.

Rasters however also have a number of disadvantages. Because rasters are essentially square tiles representing a round surface, they inherently become distorted in small-scale system modelling (Longley et al, 2001). Because rasters are of fixed resolution, and typical classified into mutual exclusivity discrete classes, they tend to generalize the true heterogeneous nature of the system attributes and therefore are highly sensitive to issues of scale. Other problems relate to the fixed resolution of raster cells, as there is a high likelihood that a given cell in the system model may occupy the region of more than one class or theme creating the need for more advanced rule-based classification schemes. Finally, rasters may be disadvantageous in that they require significantly more computer storage space for storage and are slower to process.

The size of cell is arbitrary and could be expected to increase the distributed nature of hydrologic modelling. However, given its data shortcomings, the vector model has tended to dominate GIS usage to date. Much more environmental data (soils, land use, drainage networks and catchments, etc) are currently available in vector format, so this research has focused on this approach. The interested reader may however reference topic related to uncertainty propagation in raster models by Heuvelink (1998) and Zhang et al, (2002).

2.3.2 *Vector Based GIS*

The *vector* or object model of GIS also has certain advantages and disadvantages. Like the raster model, the vector model is a spatial model structure that attempts to describe geographical phenomena. The vector model is often referred to as the object model because of its superior ability to represent geographical objects. In the vector model, an object is defined according to a hierarchy of topologically related geometry. A given entity may be defined by a set of points or nodes, when connected these nodes form a set of lines or arcs. Finally arcs may be joined to form a set of polygons (see Figure 2.2). Using this procedure, the shape of any object can theoretically be modeled given enough point, line and area information.

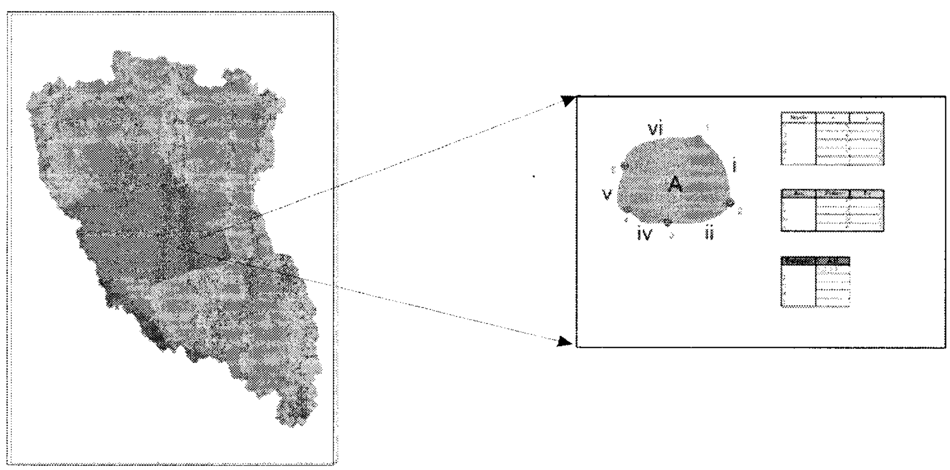


Figure 2.2. A vector data model of the Duffin's creek watershed depicting land use.

Topology in the vector model is stored in a series of hierarchical tables. Nodes are stored according to the x and y coordinates of each node with additional digitized points rendering fidelity to boundary and line shapes. Other topological information such as polygons to the left and right of a source polygon, or polygons that share a single arc as a boundary may also be defined (Longley et al, 2001).

Vector models have an advantage over rasters in that they are scale independent and, therefore, require much less information to define a given object to a comparable level of precision. This allows for much less of a demand on computer storage and computing power. Vector models however are more limited in terms of mathematical processing and are considered less powerful than rasters in this respect. Vector models also have one distinct disadvantage in that because of their scale properties, maps with a fixed level scale may be subject to misrepresentation from highly precise yet inaccurate vector representation. This may cause further errors and uncertainties when vector models are subject to analysis such as overlay as over-precision causes the formation of undesirable sliver polygons and other problems related to logical consistency.

2.4 Model Hierarchy and Inheritance

A GIS on its own is a model of spatial relationships capable of performing complex analysis. The analytic capabilities of a GIS vary depending on the complexity of the software. However most common GIS systems are capable of at least performing rudimentary spatial analysis based on the topological relationships of the points, lines and polygons. Algorithms used for the spatial analysis of points, lines and polygons are typically programmed within the primary functionality of the GIS. While the details of these algorithms are beyond the scope of this research, in a general sense they are based on fundamental geometry and set theory including the concepts of intersection, union, proximity, trigonometry and distance.

While GISystems are most often capable of powerful analysis, in order to extend the functionality of a GIS, the model structure is extended through integration with the external model (see Figure 2.3). Once this occurs, the limits of what constitutes GIS model or external model factors become more difficult to define as the lines between the external model and the GIS grow fuzzy dependent upon the degree of coupling. The extending of GIS functionality through the incorporation of external models introduces a concept we will define as *inheritance* whereby, the GIS and the external model inherit parameters and variables associated with each independent model structure.

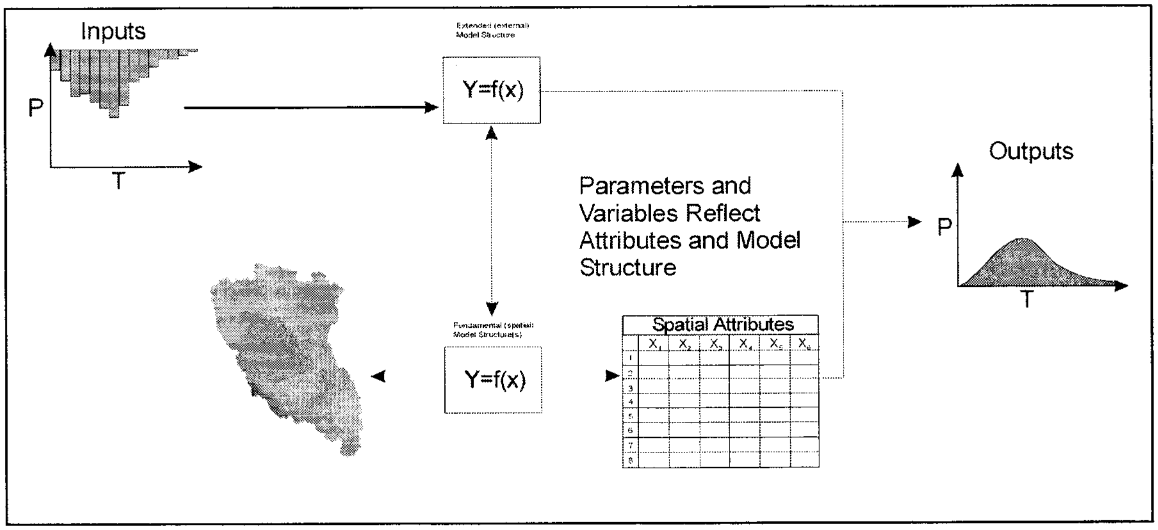


Figure 2.3. During coupling of GIS and external models, model factors such as parameters, variables and attributes are inherited by the GIS and external model concurrently.

2.5 Coupling GIS and Hydrological Models

The concept of coupling GIS and hydrological mathematical models has been discussed in detail by Sui et al. (1999). They propose four classes of coupled GIS hydrological models including a GIS embedded within a hydrological model, a hydrological model embedded within a GIS, loose coupling and tight coupling (Figure 2.4), each with different advantages and disadvantages.

2.5.1 GIS Embedded within a Hydrological Model

In this first case, modellers attempt to embed GIS functionality within a pre-existing hydrological model. Here, the use of GIS is primarily limited to cartographic visualisation. This technique has the advantage of allowing developers freedom of design in that, implementation of the coupling strategy and design of the hydrological component is not dependent or limited by GIS data structure. The limitation, however, is that visualisation and data management capabilities of these packages do compare to the potential presented by stand-alone GIS systems, the advantages of which have been outlined by Singh (1996) and discussed earlier in this chapter. Further, programming efforts required to achieve this coupling strategy are seen as intensive and occasionally redundant. Some examples of this approach include a variety of the HEC series of models developed by the US Army Corps of Engineers, the LDMS system as well as the application MODFLOW (Sui et al., 1999), the MIKE-SHE model, (Feyen et al., 2000) and automated mapping and facilities management AM/FM (Shamis, 2000).

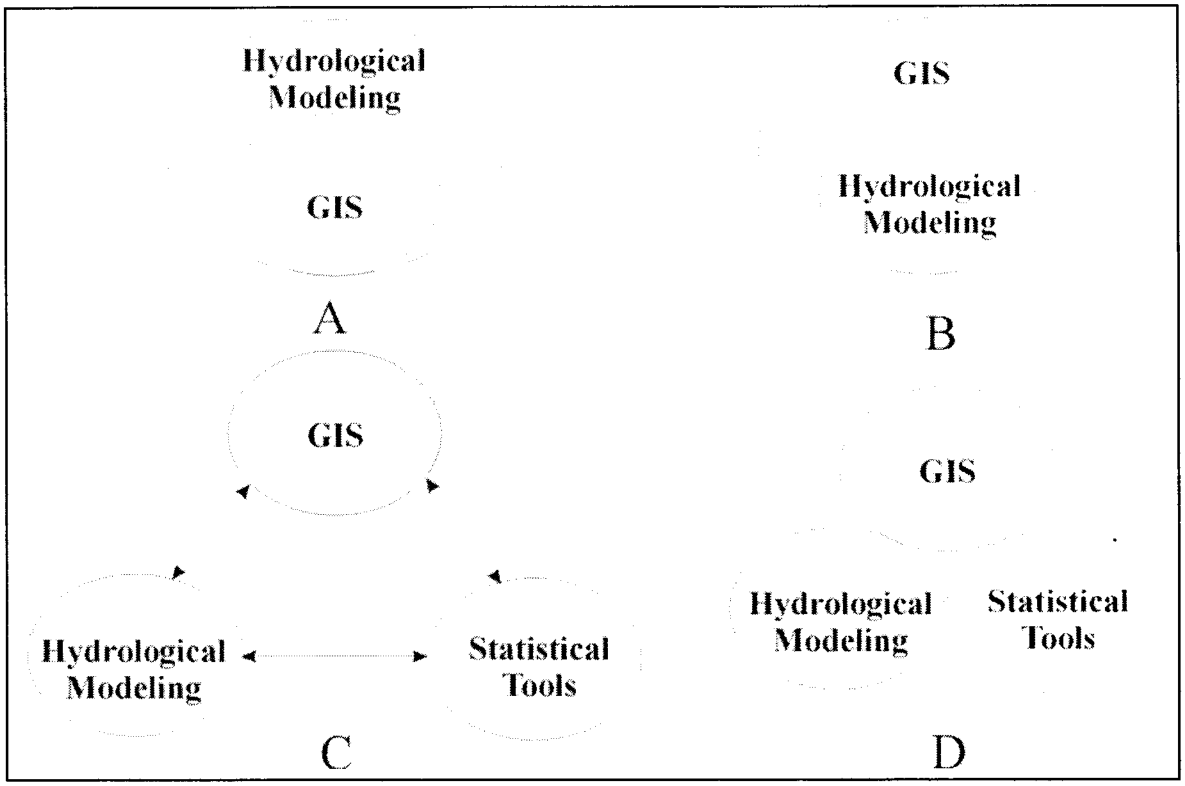


Figure 2.4. Coupling strategies for the integration of GIS and hydrological modelling. Source:, Sui et al., 1999.

2.5.2 *Hydrological Models Embedded within a GIS*

As a result of the drive to improve the analytical capabilities of a GIS over the last 10-15 years, many software vendors have introduced stand alone GIS packages with a variety of hydrological application embedded within. One example of this includes the Environmental Systems Research Institute's (ESRI's) ArcHydro. While these modules take advantage of the suite of collection, storage, analysis and visualisation capabilities of commercial stand alone GIS packages, the hydrological functionality is often criticized in that it, does not often conform to the set of long-standing BMP's associated with conventional hydrological modelling. This is especially true in the case of issues relating to model calibration and validation where often models of this type must be calibrated outside of the existing model (Sui et al., 1999).

2.5.3 Loose Coupling

The loose coupling approach typically involves the linking of stand-alone GIS package such as ArcView, IDRISI, GRASS or Arc INFO and a stand alone hydrological modelling package such as TR-55, HEC, OTTHYMO, and potentially a statistical package STATA, SPSS, STATISTICA through some common interface. Model data is integrated through data exchange where output from one model becomes the input to another model (Figure 2.3). In the loose coupling approach inheritance of model parameters and variables is less strictly defined as models are still functional as stand-alone packages and the paths of model input and outputs are still traceable and largely controlled by the user.

The advantage behind loose coupling is that intensive and often redundant programming can be avoided while maintaining the full potential of each stand-alone package. Disadvantages, however, are that data exchange between model inputs and outputs is highly prone to errors associated with the manipulation and conversion of data formats and the tedious exchange of data. This approach, however, has been supported as the most feasible for the majority of investigators as programming requirements are considered minimal and less time consuming (Sui et al., 1999).

Because loose coupling is the most popular method for the above mentioned reasons, numerous examples exist some of which include a study by Frankenberger et al., where the GRASS GIS package was linked with TOPMODEL to examine the effects of soil heterogeneity in a small watershed. As well, Storck et al. (1998) is another example where the Distributed Soil Hydrology Vegetation Model (DHSVM) was loosely coupled with GIS package to provide pre and post-processing of model dependent variables to examine the effects of land surface change. Further, numerous examples exist which utilize

GIS in conjunction with the AGNPS model such as Yagow (1997), Al-Smadi (1998) and a recent study by the TRCA (2003).

2.5.4 Tight Coupling

Finally, the tight coupling approach is utilised when hydrological models, a stand-alone GIS and perhaps statistical packages are strictly integrated with each other through custom programming interfaces. This allows users to develop custom functionality and provide proprietary user interfaces. Historical limitations of this strategy stem from the programming languages used not often being powerful enough to maintain total and desirable functionality of software packages, resulting in less sophisticated hydrological or GIS functions. This, however, is changing through the evolution of software and programming languages becoming more and more component object model (COM) compliant.

Through COM, many of even the most complex models can be made available as subroutines in the form of DLL's or dynamic link libraries, as well as ADO's or Active X data objects. An example of a GIS platform that readily adopts COM standard is the new suite of GIS products associated with ESRI's ArcGIS platform. Through this platform developers have the opportunity to program in many of the most powerful and popular languages such as or visual basic (VB), or visual basic for applications (VBA), C, C++ and Visual C, as well as incorporate any COM compliant scripting for expansion of the analytical or performance capabilities of the GIS. Some examples using the tight-coupling approach include BASINS, a tightly coupled GIS hydrological model developed within ArcView as well as the new ESRI ArcHydro platform.

While literally hundreds of examples of integrated GIS/hydrological models exist (see Tsihrintzis et al, 1996), few have outlined how best management practices established for system modelling are extended or simply viewed within the context of GIS. In the next section, the BMP's outlined in the first chapter are put in the context of GIS-based system modelling.

2.6 Geographic Information Best Management Practices

Like any model, GIS allows for various representations or conceptualization of reality and inherent in this is the notion that GIS permits a certain number of hypotheses regarding the system of interest. For example, a model may be more accurate if it takes into consideration the distributed nature of geographic versus lumped data. Like any tool, the application of GIS requires certain assumptions on the part of the user. For example, how does the user know that the information presented within a GIS is accurate or more importantly, how will the assumptions made in applying a potentially inaccurate GIS model affect the outcome of spatial decisions? GIScience is geared towards the study and management of these and other issues through the scientific investigation issues concerning the creation, handling, storage and use of geographic information (Longley et al. 2001). While the notion of GIScience has been presented in a variety of different specific as well as general contexts, perhaps an appropriate definition of GIScience would be the scientific pursuit of best management practices concerning the identification, transfer and application of spatial information in support of spatial analysis and decision-making.

In 1996, the Unites States University Consortium for Geographic Information Science (USCGIS) held an assembly to determine the most important issues facing the study of GIScience. These topics include:

- *Cognition of geographic information*
- *Spatial data acquisition and integration*
- *Spatial analysis in a GIS environment*
- *Interoperability of geographic information*
- *Distributed computing*
- *The future of spatial data infrastructure*
- *Uncertainty in geographic data and GIS-based activities*
- *Extensions to geographic representations*
- *Issues of scale*
- *GIS and society*

The remainder of this research will focus primarily on issues of uncertainty in geographic information and geographic based activities and to a certain degree how uncertainty in data affects spatial analysis in a GIS.

Previously, the notion was emphasised that GIS is becoming less a means of providing support to traditional or less spatial hydrological modelling and is now evolving into the primary modelling environment for hydrological systems. While the evolution of GIScience has provided strong support for development of the best management practices for spatial data, in the absence of a clearly defined set of criteria, the notion that GIS for hydrology should adhere to already established modelling best management practices seems to be a logical step. This is examined in the following through reference to BMP's identified for hydrological models in the previous chapter.

- *Models are often the most available alternative for analyzing complex resource problems.*
- *Models have the potential to provide even greater benefits for future water resource decision-making*

In the case of GIS, the diversity of applications is nearly limitless. It has been said that on the order of 80-90 percent of all data is spatial in one context or another and, therefore, the use of a model specifically evolved to deal with the complexities of spatial data seems a reasonable application especially given the shear quantity of data involved in spatial decision making. Further, it was previously noted that hydrological data is not only complex but is primarily considered spatial by nature. While diversity of applications are too numerous to discuss at present, the topic has been examined by Longley et al. (2001) where the variety of GIS applications can be classified according to the five M's including mapping, measurement, monitoring, modelling and management.

- *Water resource models vary greatly in their capabilities and limitations and must be carefully selected and applied by knowledgeable professionals. Selecting appropriate procedures for analyzing a particular problem is a major part of the modelling activity*
- *Selection of mathematical models for regulatory applications requires a thorough understanding of the capabilities and limitations of the available models*

- *Proper development of models requires modellers that are well trained in the underlying physical principles of the environmental system as well as in the computational procedures for modelling*

The availability of GIS technology has become increasingly decentralized over the last 15 years. This has transformed the application of GIS models from what was once the domain of experts to a now relatively commonplace activity (McKendry, 2000). In this respect, the novice user can now accomplish what were once complex analysis tools with the press of a button, and with a complete disregard for the applicability of the results.

The ability of GIS to provide convincing, if not completely misleading information, should not be neglected/overlooked and the ability to properly interpret information from a GIS can be twofold. Spatial data, like any data is subject to various degrees of uncertainty and, therefore, no spatial model stored within a GIS should be considered to be "error free" (Heuvelink, 1998), although, naivety on the part of the user often perceived it as such. In this respect, the unskilled user may be prompted to make spatial decisions based on incomplete knowledge of the model outcome.

Additionally, in many cases the result of modelling in GIS consists of at least one if not many maps. The ability of the GIS to provide interesting and informative visualisation of model processes has also been discussed as an advantage over traditionally hydrological modelling. The ability of maps, however, to convey inappropriate spatial information is significant as maps are considered a cognitive media strongly capable of communicating spatial information (Harely, 1990; Wood, 1992; Tyner 1982; Henrickson, 1975). In this respect, the unskilled interpreter may become victim to performing decisions on a spatial miscommunication of information.

Finally, the GIS user should not be in a position to suggest a course of actions based on the results of data analysis beyond his or her ability to effectively interpret. Some aspects of spatial data are, however, complex and difficult to interpret even with the support of quantitative measures of data accuracy and completeness. In these cases, the user should have enough experience to impart wisdom about those aspects of the data, which cannot be further measured. This wisdom, in the case of hydrological modelling is often attributed to engineering judgement or expertise. In this respect, the appropriate judgement should come from a user possessing ample geographic expertise or judgment and may be defined as "*Geographers Judgment.*"

- *Mathematical models should be based on a fundamental representation of physical mechanisms and incorporate, to the extent possible a state of the art understanding of the problem.*

Model structures within a GIS are often supported in a hierarchical fashion in part by other physical models. As is often the case with hydrological models, the given input to one model is often necessarily the output of another model (Beck, 1991). In GIS, as previously noted, the mathematical relationships which define model structure and functionality will be dependent on the way hydrological modelling will coupled with the GIS the type of hydrological structure, and the degree of inheritance. In most cases, the hydrological functionality embedded within a GIS will be inherited from external model structures and in this respect ought to be the most up-to-date and representative hypothesis of the appropriate physical relationships. In the case where spatially specific processing is required to execute the model, the same should apply to the algorithms that define the spatial processing governed by the GIS.

- *Models used for regulation applications must be subject to a two stage confirmation with field data consisting of 1) Calibration, where the parameters of the model are estimated to allow a good match between the model predictions and an observed field data set; and 2) verification where the model is compared to an independent data set without further modifying the parameter values and relationships in the model.*
- *Comparison of the model predictions and observed field data as part of the confirmation process should include qualitative graphical comparisons and if appropriate quantitative measures of the goodness of fit.*

While intuitive to hydrological modelling the processes of calibration and verification may not be directly obvious to the GIS user. In an ideal environment, spatial data, which has been processed, must be subject to an extensive process of ground truthing. This may occur as part of the data acquisition phase or may in part be validated at later stages of data processing and manipulation. In cases where data is gathered through processes of remote sensing often, data portrayed in the GIS is verified on the ground. For example, an air photo may provide data for a GIS on vegetation and/or land use. In cases where the information entered into a GIS is subject to interpretation, randomly selected sites are validated through a series of field procedures. Ideally, metadata is then recorded which documents the reliability of the interpreted data.

In other cases, data within the GIS may be interpolated. This procedure involves the field measuring of some attribute such as soil at particular points within the area of interest. These points are then entered into the GIS and through geostatistical procedures; the points are interpolated into a continuous field. This process is also subject to calibration and validation. Once the resulting field of data has been interpolated, randomized selections of points within the study site that are estimated as part of the interpolation procedure are verified through a series of field measurements and investigations. In some cases the model may be calibrated through additional statistical techniques or abandoned in favour of another model. Unfortunately, confirmation of spatial data is often not performed may be a source of error that warrants documentation.

- *Sensitivity and uncertainty analysis of environmental models should be performed to provide decision makers with an understanding of the level of confidence in model predictions and to identify key areas for future study.*

One of the most intensively researched fields in GIScience is in developing methods for understanding the effects of uncertainty and uncertainty propagation within spatial data. Because of the recent dependence of hydrological models on GIS, the exploration of the affects of uncertainty as well as the sensitivity of the model output to uncertainty in model input is the subject of significant research and is the focus of the remainder of this thesis. Because significant research has been focused on the quantifying, handling and minimizing of spatial uncertainty in GIS, and hydrological models separately, the integration of GIS with hydrological models should also warrant that these uncertainties are taken into consideration in the overall uncertainty and sensitivity analysis of the model output.

Aside from placing GIS within the context of system modelling BMP's there are other technical and philosophical challenges related to the coupling of the two approaches. The final section of this chapter will be devoted to this topic.

2.7 Challenges of GIS and Hydrological Model Integration

Despite the numerous advantages of integrating GIS with hydrological modelling, there exist a number of challenges in linking the two approaches.

2.7.1 Technical Challenges

Providing an effective set of tools to the environmental modeller is paramount to the evolution of modelling practice. While the integration of GIS with hydrological modelling certainly is aimed at facilitating this notion, the technical challenges are often great. As mentioned earlier in the chapter, various advantages and disadvantages with respect to integration strategies are often technical in nature and a primary inhibition to using GIS in modelling are the difficulties presented in integrating an external model with a GIS (Karimi et al, 1996). Further, since the interfacing of various standalone programs requires in many cases the development of additional programs to accomplish the integration task, a human element is introduced which can be the source of a large degree of error and uncertainty where data transfer or integration is concerned. This challenge leans towards a need for a more consistent and powerful means of automating integration tasks (Tsihrintzis et al, 1996).

Karimi et al. (1996), discuss and compare the various technical challenges in the loosely coupled versus tightly coupled strategy with respect to the speed of data transfer, the level of GIS specificity, and the level of integration. The speed of data transfer is an important challenge as many different data are passed through separate models and model structures often requiring a large amount of computing power.

GIS specificity deals with transferability or dependency of the hydrological component on single GIS platform, limiting the broadening of the GIS application through the introduction to a variety of platforms with varying analytical capabilities.

Finally, the level of integration deals with the synergism of all hydrological and GIS components and models. Conclusions of this study indicated that tight coupling was preferential to loose coupling in addressing the previously mentioned issues. Although, the limitations of GIS programming made successful coupling difficult. This aspect, however, has evolved since the study was published and COM technology has opened the door to new integration capabilities.

While various technical challenges to integrating GIS and hydrological models exist, perhaps one of the most significant challenges is the tendency for individuals to focus on the technical challenges as opposed to other more conceptual or philosophical limitations.

2.7.2 Concepts of Space and Time

Sui et al, (1996), have identified issues concerning the separate ontology of GIS and traditional modelling, a philosophy also discussed by Clark (1998). In the classification of hydrological models, it was noted that depending on how space and time are dealt with, a model can either be stochastic or deterministic and lumped or distributed. The GIS model structure on the other hand adopts a layer-based approach where the representation of space and the associated analytical model structure is limited to a series of map layers occupying the same space. As a result of this, space within a GIS is conceptualized as a geometrically indexed representation and simulation of time is limited in conceptualization to discrete slices (Sui et al, 1996). Because of these differences, integration of hydrological models with GIS must be conceptually limited requiring the future development of coupling to address issues surrounding varying conceptualizations of space and time.

2.7.3 Data Availability

Also touched on earlier in this chapter is the notion that GIS provides numerous advantages to the design of geographically distributed hydrological models. While the power of GIS lies in the ability to spatially manage higher degrees of resolution and consequently, integrate higher degrees of heterogeneity and spatial variation of system attributes, high-resolution GIS data requires highly accurate data acquisition methods. Unfortunately, the current power of precision modelling available with GIS exceeds the standards of available field data resulting in the simulation of highly detailed system attributes from coarser, less variable but more accessible field data. In this respect, the evolution of hydrological modelling is limited by the availability of spatially accurate GIS field data. Tsihrintzis et al, (1996), suggests that as the use of GIS increases, spatial information should also be readily available and local authorities should establish centralized data banks providing the most up-to-date and compatible stores of digital maps, remote sensing images and spatial metadata. The availability of such data is also consistent with improvements in the collection and processing of remotely sensed imagery, which is vastly becoming a superior means of data collection and quickly replacing extensive field analysis.

2.7.4 Handling Uncertainty

Finally, of great importance to the integration of GIS and hydrological models is the control and communication of uncertainties. As is explored in detail in the next chapter, every stage involved in the integration of GIS with hydrological modelling is potentially full of uncertainties. While uncertainties will never be fully eliminated from modelling and while quantifying and addressing all possible uncertainties is highly unlikely, Sui et al, (1996) have proposed a two-tiered approach for dealing with uncertainty in coupled models. In the first case, it is proposed that while there is a proliferation of statistical models available for the analysis and quantification of uncertainty, there are currently no techniques in wide practise which address issues of imprecision and incompleteness. Although, evidence theory and fuzzy set theory have been suggested as providing a significant amount of potential with respect to these problems (Sui et al, 1996).

The second issue surrounds the method of visualisation and communication of uncertainty. Numerous methods in the cartographic literature such as the use of Epsilon error bands as well as manipulation of primary graphic elements such as shape, hue, value saturation and texture can be employed. These cartographic techniques have a long-standing history and numerous publications exist on the use of these techniques in cartographic communication.

Paramount to investigations of model structure, data structure and inheritance, is the concept of uncertainty. All models, whether they are geographic or mathematical are derived ultimately from a mental models or a simplified conceptualization of some system of interest. In the case of GIS, these models conceptualize space and in the case of mathematical system modelling they conceptualize the physical functions of a system and indirectly time as many of these systems are inherently time dependent. Further conceptualization of the world is largely biased by experiences and influences. While knowledge of a system can over time improve through observation and experimental inquiry, knowledge will never be perfect and, thus, explanations of systems of interest will contain inherent uncertainty.

This chapter reviews the concepts of error, uncertainty and sensitivity within hydrological modelling in a geographic environment. The following sections review sources of error and, consequently, causes of uncertainty within spatial and aspatial systems. Aspects of model structure uncertainties and model factor uncertainties relating to positional and attribute uncertainty are discussed. The final sections introduce methods for quantifying and dealing with uncertainty in data in the form of global model uncertainty and parameters specific sensitivity analysis.

3.1 Uncertainties and Error

Uncertainty is induced ultimately from our abstraction of reality and since it is impossible to conceptualize a perfect representation of the world, introducing uncertainty about it is inevitable (Longley et al, 2001). Inherent in the concept of uncertainty is the notion of error. The process of materializing an abstraction of a system in the form of a model (map or mathematical system) will contain error or a certain number of inaccurate assumptions, misconceptions or generalizations about that system. While error and uncertainty are often discussed as synonyms within the literature, this is not the case and a drawing a distinction between the two is necessary. Heuvelink, (1998), defines error as the deviation between reality and our representation of reality, and draws on the following example to illustrate.

> *Let the true value of some attribute at a location x be a(x), and let our approximation of that value be b(x) then the error or deviation from reality v(x) is described as the arithmetic difference between a(x) and b(x) such that v(x) = a(x) – b(x). The error v(x) is not known exactly and but we should have some knowledge about its range or distribution of values that it is likely to take. For example, we may know that the chances are equal that v(x) is positive or negative, or we may be 95% confident that v(x) lies within a given interval.*

Building knowledge about the value or range of values of v(x) is the subject of defining what is known as an error model. This will be discussed in further detail later in this chapter but for now it simply acknowledged that error can be considered as a deviation from reality and since we can not know the deviation exactly it may be preferable to describe it stochastically as opposed to deterministically.

The term uncertainty should not be considered the same as error although the two are tightly related. Given that error can be defined as the deviation from reality, then *uncertainty* may be thought of as a lack of confidence as to whether something in fact deviates from reality or rather the degree with which something deviates given that perfect simplification is impossible. Put another way, the stochastic nature of uncertainty may justify its definition as being the probability that something is incorrectly represented.

The term uncertainty has been used under numerous contexts to represent a generalized description of data quality. Longley et al (2001), note distinctions between types of uncertainty including *ambiguity* in which various perspectives established by a variety of individuals on a common element may not share the same definition leading to an indefinite assignment of properties or other identifying features. Further, it was noted that ambiguity is often introduced under circumstances where imperfect indicators of a phenomenon are used to describe or identify an entity as opposed to the entity itself. To draw an example from the SCS-method, the set of soil types initially identified in the 1950's by Musgrave, (1955) and used in conjunction with local land use classes for the determination of relevant Curve Number are classified as 4 (A,B,C,D) mutually exclusive hydrological soil groups (HSG's). Although various physical soil types have been identified to loosely correspond to each soil class (see Chapter 4) the HSG remains an ambiguous classification and is largely used as a generalized measure for the hydraulic conductivity of a broad category of soils where class "A" has the highest propensity for infiltration while class "D" is the most resistant. The classes are not assigned using any quantitative measures, and the data used to support the initial class separations by Musgrave, is not described in the literature. In this respect the HSG soil classification scheme is a broad and ambiguous generalization, loosely based on local soil types.

Vagueness, is another type of uncertainty relating to classification of entities within a system. While it is commonplace to take a series of discrete observations about a system of entities and transform or interpolate these observations into a generalization of an area, this raises questions about the validity of boundary determination among classes of entities whose class assignment is fuzzy at best. In the hydrological context, this may relate to the determination of areas of influence for a series of point estimations of rainfall as is often accomplished through the assignment of Thiessen polygons (Thiessen, 1911). While Thiessen polygons are used to discretely identify the areas influenced by a series of points, the degree of discreetness conveyed in this type of tessellation can be misleading as the true spatial properties of the phenomena being tessellated is not known. To draw a further example and building on the previous example illustrating the ambiguous nature of the HSG, the degree of heterogeneity associated with soil makes the classification of soil into discrete classes extremely challenging. Consequently, a great deal of scepticism with respect to the reliability of soil classes as well as the confidence placed in the relevance of boundaries displayed in soil maps is often wrought with uncertainty.

3.2 Sources of Error and Uncertainty

Error derived at one stage of a models conceptualization may directly influence how error is introduced through other sources of the models abstraction. For example, scale is an important aspect of designing any spatial model and choice of a particular scale of observation directly influences the degree of spatial generalization within a system. This generalization can then lead to other sources of spatial error resulting from generalization which can again lead error in positional inaccuracy, attribute inaccuracy and so on. This process known as *error propagation,* is the primary source of uncertainty and derived from the abstraction of GIS based hydrological models the process of which is mapped in Figure 3.1.

One of the reasons identifying sources of model uncertainty is challenging is that identifying all sources of uncertainty quantitatively involves a comprehensive understanding of error propagation and consequently a complete understanding of the lineage in the models development, model abstraction, parameter estimation, data capture, measurement and analysis as well as storage, transfer and manipulation. This is further complicated by the notion that while identification of some of the error within a model may be possible, it is difficult if not impossible to quantify this stochastically in totality through ordinary statistical procedures.

Figure 3.1 divides the propagation of uncertainty into four levels of abstraction. The first relates to the way in which in which a phenomena of interest is initially perceived. This is followed by a conceptualization phase during which physical relationships are mathematically identified effectively assigning a model structure to the phenomena of interest. The abstraction of the entity is then fulfilled through spatial and a-spatial attribute assignment. Each of these stages coincides with a certain degree of abstraction. As a result error is introduced from a number of different processes that then in turn propagate throughout the model. The specific types and sources of error leading to propagation within the model are the subject of the following section.

3.3 Uncertainty and Error from Interpretation and Abstraction

3.3.1 Interpretive Uncertainty

It was proposed in the first chapter that all models begin as mental models or conceptualizations of the world. This may materialise in the case of a system governed by laws and physical relationships as a mathematical model or may materialize in the form of map, digital or otherwise. The ability to conceptualize the world around us is based on experiences and culture and is predicated by our current state of knowledge about the world and the laws that govern it. This is further influenced by the fact that we can observe so little of the earth directly and are reliant on a host of methods for learning about its other parts (Longley et al, 2001). As a result, bias is inherent in any abstraction process. We interpret the world with a certain perspective and, in so doing, we introduce uncertainty into our mental image of the way the world works. This error leads to what can be identified as *interpretive uncertainty*.

3.3.2 Abstraction Uncertainty

Abstraction uncertainty is the uncertainty introduced by imperfect conceptualization of the world. Ideally systems could be conceptualized comprehensively enabling fidelity with the reality they represent. However, knowledge is incomplete and, therefore, our abstraction of systems will ultimately be based on incomplete knowledge and is inherently linked to ideas of *ambiguity* and *vagueness* already discussed.

3.4 Model Uncertainty

Dealing with error in model structure is a complex problem and one that is normally avoided through refinement of confidence in other sources of error such as the estimation of model parameters and increased accuracy in measuring model attributes. Model structure was defined in chapter one as a group of hypotheses consisting of a set of general laws $L_1, L_2, \ldots L_n$ and a set of statements pertaining to empirical circumstances $C_1, C_2, \ldots C_n$. As discussed, a model is an abstraction of reality and in the absence of a perfect knowledge, no abstraction (which is predicated on simplification) can be without error or departure from reality. This error introduced with the application of model structure, manifests itself in what is identified as *model structure uncertainty*.

In Figure 3.1, opposing views are presented for both GIS and again for hydrological modelling. In the GIS case, two opposing model structures, the raster or field model, and the vector or object model. In the hydrological case, two structures are presented largely adopted as a means of computing runoff from a given mass balance. In the first case, the Horton physically-based infiltration model is depicted as one model structure while the other, the empirically based SCS-curve method is presented as an alternative. In either

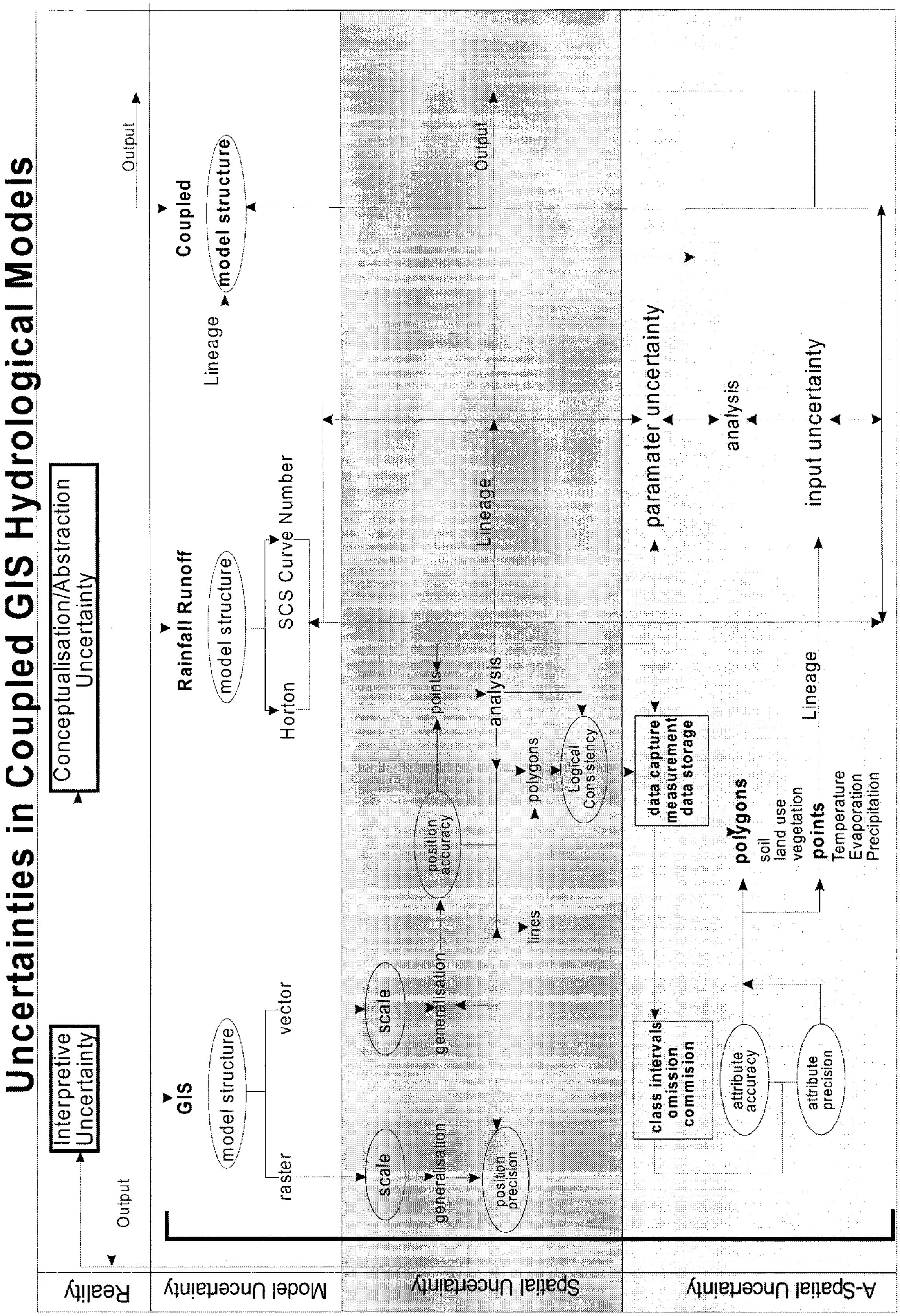


Figure 3.1. Error and consequently uncertainty is introduced into GIS, hydrological models, and coupled models through various levels abstraction. Abstraction exists as abstraction from reality, abstraction of model structure, spatial abstraction and, a –spatial abstraction.

case, both model structures are equally applicable in a number of cases. However, both alternatives contain uncertainty as a result of diverse abstraction.

As a means of dealing with uncertainties in model structure, Beck (1991), suggests that given the ability to reduce or quantify uncertainty in other sources of error, uncertainty in model structure can be dealt with through precise verification (validation) of the model, a process discussed in Chapter 1. In this respect, it is necessary to note that in validating the model, there is no control for error or uncertainty in model structure, nor is there any attempt to quantify that uncertainty. However, in testing the ability of the structure to be validated, there is an increase in the confidence in a hypothesis that the model is a reasonable representation of reality.

The notion that model validation provides a means of reducing uncertainty in model structure has been criticized. It has been recognized that the greater the number of parameters involved in defining model structure, the greater is the likelihood in achieving an acceptable model calibration (Aitkin, 1973). Beck, (1991) however, reiterated earlier findings by O'Neil, (1973) suggesting that, in adopting a more complex model, the need is inherent to determine more variables each having an associated uncertainty and, thus, increasing the likelihood of error being introduced into the model.

Issues of uncertainty in model parameters have also been discussed by Beven et al. (1992). Here, it is identified that in a given calibration exercise it is highly likely that a number of correct parameter values may be chosen which provide an equally favourable set of model outcomes. It was also noted that these problems tend to get worse with complex models containing a large number of parameters.

Assuming that parameter uncertainty can be dealt with, the remaining error in the underlying physical abstraction defining the structure of a model is suggested to be a philosophical debate that may be overlooked in the urgency of reaching a decision (Beck, 1991). Inherent in this argument is the application of the precautionary principle, whereby in the absence of a clearly better understanding of the problem, one is obligated to draw upon an existing hypothesis until evidence supports a more suitable prediction. This has been further asserted by Lei (1996) who suggests that unless disproved through observation, the uncertainty in model structure is negligible or at the very least liveable.

3.5 Scale

Scale is an important component of uncertainty and error propagation in models. Often, the chosen scale for a model is necessarily a tradeoff between the level of spatial resolution needed to reasonably represent the processes of interest and the degree of manageable detail. Therefore, the scale at which data are collected, stored, manipulated, visualized and displayed will determine the level of generalization in the data. This ultimately will lead to other sources of error and uncertainty associated with generalization.

3.5.1 Soil Generalization

Various authors have examined the role of scale-based generalizations on GIS-based hydrological modelling. Liang et al, (2001) demonstrated the importance of detailed soil information for the simulation of large scale Global Circulation Modelling (GCM) while Brath et al (2003), discussed the effect of soil resolution on infiltration excess predicted by a distributed SCS-Curve Number on peak flows, finding that prediction of lower peak flows is more sensitive to the scale with which the Curve Numbers are resolved. Zhu et al (2001) examined the effects of high-resolution soil data on measured hydrological and ecological responses of a catchment identifying that modeled responses such as peak runoff and net photosynthetic activity of forested areas are sensitive to the spatial detail of soil when a lumped model is used but increased soil detail has less of an impact on model performance when a distributed approach is used.

3.5.2 Rainfall

Various authors have also examined the role of rainfall heterogeneity and distribution on watershed response. Arnaud et al., (2002) investigated the effects of rainfall resolution on a 2000 km^2 densely monitored area surrounding Mexico city and found that lumping rainfall as opposed to using a distributed field led to problems in model calibration. Koren et al. (1999), examined the effect of precipitation data at different scales on lumped runoff modelling and found less surface runoff and greater evapotranspiration at larger scales as well as noting that scale issues were the primary cause for reduction in runoff prediction. Shah et al, (1996) investigated the effects of distributed versus lumped rainfall on an experimental catchment finding under 'wet' conditions, good predictions of runoff can be obtained with a spatially averaged rainfall input. However, for 'dry' catchment conditions, the runoff prediction errors are seen to be considerably larger than for the 'wet' case, suggesting that there is interaction between the spatial variability in rainfall and the spatial distribution of soil moisture which controls runoff production.

Many authors have examined the relationship between the heterogeneity of rainfall input and the density of rain gauges. For instance, Faurier et al., 1995 examined the effect of high-resolution rain data gathered from a densely sampled small catchment with an area of 4.4 ha and concluded that the spatial pattern of rainfall data captured at higher resolution has a measurable impact on predicted runoff. Because rainfall is highly spatially variable, especially in climates subject to frequent brief localized downbursts, rainfall estimated using a single rain gauge for a catchment can not reasonable represent the heterogeneity of rainfall processes within a system resulting in significant errors in model prediction. For example, Mustzner, (1991) suggests that achieving runoff predictions of greater than 20% of the known values for a given time period may not be possible when using a single rain gauge. The effects of the spatial and temporal resolution of rainfall was discussed in detail by Lei (1996), who found it was apparent after a review of a number of studies, even for small catchments with a relatively high density of rain gauges errors in runoff prediction ranged from 20 –300%. Further, Lei also suggests that it is not reasonable to expect that better than 20 percent accuracy can be achieved in predicting runoff when only a single rain gauge is used to define precipitation for a catchment.

3.6 Spatial and A-Spatial Uncertainty

The National Committee for Digital Cartographic Data Standards (NCDCDS, 1988) has identified five types of error and uncertainty related to objects including positional accuracy, logical consistency, attribute accuracy, completeness, and lineage.

3.6.1 Logical Consistency

Logical consistency essentially relates to a lack of conveyed consistent logical features and represented topological relationships among entities. Often the result of positional errors, objects with poor logical consistency will often convey erroneous meaning about their intrinsic relationships. For example, errors in positional accuracy in streets that meet at a corner may over or undershoot one another offering a misimpression about the reality of a transportation network. Alternatively, separate and discrete objects may in fact overlap or share boundaries giving the impression of a unified structure where in reality separate and discrete entities in fact exist. Maintenance of logical consistency can, therefore, be managed through verification of topological relationships and procedures used in the management and documentation of logical consistency within spatial databases have been discussed at length by Servigne et al (1999), as well as Guptil and Morrison (1995).

3.6.2 Positional Accuracy Defined

Spatial uncertainty exists where error about an object is directly related to the accuracy of its position in space. Assuming that spatial features within a system are ultimately discrete entities and assuming further that one has the ability to investigate this discreteness with infinite scale. Consequently, infinite precision then given an infinite number of observations and a hypothesis can be drawn where the discrete spatial nature of a set of objects could be realized with perfect accuracy. This is however completely impractical and even if a set of entities could be observed independent of the scale of observation, it is likely that upon closer inspections many entities would exhibit a continuous transition and would likely not be discrete at all. As such, it could be argued that the degree of discreteness in set of objects is a scale-dependent property and, as such, when we record position of entities at a given scale into a digital representation, a process known as data modelling (Goodchild, 1989), those positions are subject to error or deviation from reality as the true positions of a given object is not known. How we represent or rather where we represent entities within the spatial realization of a system is, therefore, subject to *positional inaccuracy* or error.

Assessment of error in position or the positional accuracy in vector data begins with the measurement of points since the positions of points in space are subject to numerous sources of random and systematic error (Zhang et al, 2002). In the vector object model, polygons are defined from the position of a set of arcs (or lines), and the position of lines is inherited by the location of a set of points used to define the extend of the line. Error in the location of points then will be inherited by lines and, consequently create uncertainty in the geometrical representation of polygons (see Figure 3.2). This geometrical uncertainty often

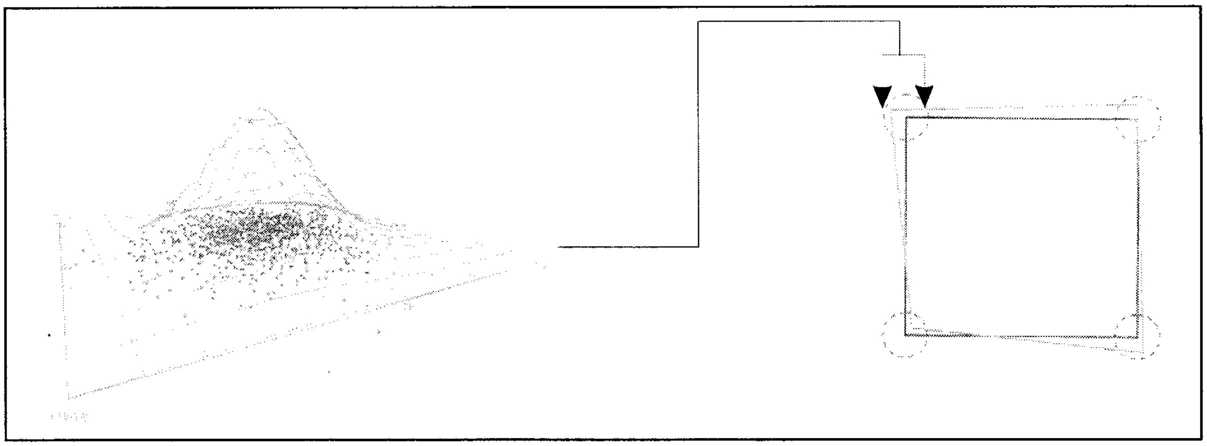


Figure 3.2. Error describing the position of a point can be described stochastically using a mean and variance. Error in points is inherited by lines and consequently affects the geometry of polygons. Source: Longley et al, 2001.

translates to errors in attributes derived from the spatial properties of features. For example, many attributes used in hydrological modelling are directly dependent on the analytical results of position dependent measurements within the GIS such as area, perimeter, elevation and slope. Other important aspects may include proximity to regions of influence or the area of influence associated with rain gauges and determined from a Thiessen polygon or other such interpolation procedure, which are largely dependent on the measured distance between two points in an area. As such, many uncertain attribute values inherited by the hydrological components of coupled models are directly related to positional errors within the GIS.

Various authors have examined the influence of the stochastic nature of points on the propagation of positional error in higher objects such as lines and polygons (Zhang et al, 2002; Shi et al., 2002), most of which are largely related to the concept of epsilon error bands, (Shi, 1998; Perkal, 1956) where the error in the position of lines is expressed based on the probability of its actual location occupying a region around its assumed position.

3.6.3 Statistical Measures of Positional Accuracy

The error in points is often determined through comparison with errors in known locations of at least three times the accuracy (Zhang et al., 2002), and is often quantified using descriptive statistics. Some of the descriptive statistics used to describe this error include the sample mean error, the Root Mean Square Error (RMSE), the standard deviation of error, the mean displacement and the standard displacement. The first three statistics are used to approximate error in the X and Y coordinates separately where, the mean error, RMSE and standard deviation of error in X and Y is independently represented by :

$$m_e = 1/n \sum e_i \tag{3.1}$$

$$RMSE_e = (1/n \sum e_i^2)^{1/2} \tag{3.2}$$

$$Std_e = [1/n \sum (e_i - m_e)^2]^{1/2} \tag{3.3}$$

where,

e_i is the difference between an error prone point, in the **X** or **Y** coordinate, and a known point of higher accuracy and
n is the sample size.

The mean displacement and standard displacement differ from previous measures as they consider error in both the X and Y simultaneously as:

$$m_{dis} = [m_e(x)^2 + m_e(y)^2]^{1/2} \tag{3.4}$$

$$S_{dis} = [stdx(x)^2 + stdy(y)^2]^{1/2} \tag{3.5}$$

where:

> ***$m_e(x)$*** and ***$m_e(y)$*** represent the mean error in the **X** and **Y** coordinate of a series of points and
> ***stdx(x)*** and ***stdy(y)*** represent the standard deviation in error from a series of points again when compared to a series of points of known and higher accuracy.

In cases where the error in the position of points is purely random, the error in a point can be conceptualized as a set of circular bands represented by a series of standard displacements about a central location where 99.9 percent of all error theoretically occurs within three standard displacements assuming the error is normally distributed. However, as is often the case, error in points is not purely residual random error and is largely systematic or relative thus error bands for a point more closely approximate an ellipse.

Extending this to the concept of error regions for a line would follow the same logic. Given that error in the points that define a line is random an error band would be formed around the line evenly being the same width on either side of the line and circular at the end. However, given that the error is partly systematic, the regions around the line will be distorted (Figure 3.3).

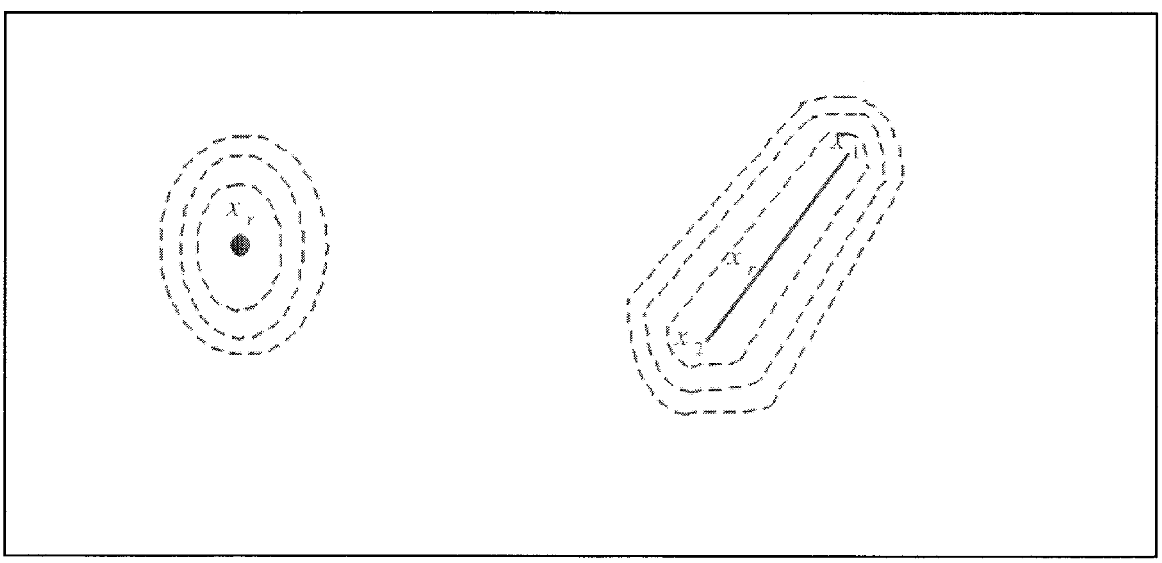


Figure 3.3. Error bands delineating positional error in points and lines under systematic as well as stochastic error can be conceptualized as an ellipse as opposed to a circle. Source: Zhang et al, (2002).

3.6.4 Sources of Positional Uncertainty

Numerous sources of positional error are known to occur during stages of data capture, data manipulation. In the case of data capture, data for GIS-based hydrological models comes from various sources. Model attributes such as land use and soil type are often digitised from pre-existing paper maps, or scanned into digital form or constructed from measured field data. In all cases, the input of positional information concerning features of interest is subject to numerous sources of error such as measurement error either from poor field practices, or the tolerances of field and lab equipment. For example, a digitizer will only enter data as accurate as the person who is taking the measurements. Further, a GPS receiver can only provide digital information to within a specified margin of error. If secondary sources are being used, certainly error was introduced during its creation that will undoubtedly be transferred to the new application. In many cases, remotely sensed (RS) data are used to create digital GIS files on land use and perhaps soil. Because RS data is in raster form, to be used with a vector model, it must be processed into a vector format, consequently, introducing positional errors in addition to those associated with the capture process, its scale and resolution.

3.6.5 Lineage and Completeness

Aspatial uncertainty refers to sources of error that are not directly influenced by spatial error and include attribute accuracy, completeness and lineage. *Completeness* refers to degree to which data of interest include all relevant information. For example are all objects in a theme present? Has all data been classified exhaustively into a series of relevant and exhaustive categories. *Lineage* on the other hand refers primarily to the history of a data source. For example, was the data in a vector theme digitized by hand? What inappropriate projections were used or, are the data a product of numerous sources each with separate levels of accuracy.

Often it is assumed that error within various layers of a GIS are independent. However, this is strongly influenced by the lineage of the data. Goodchild (2002) suggests:

> ...geographic data is often derivative in the sense that many stages of processing, interpretation, compilation and transformation occurred between the initial stage of measurement and the final product.

Understanding lineage means comprehending the history of data and, thus, its potential uncertainties.

3.6.6 Attribute Accuracy Defined

Attribute accuracy deals fundamentally with error in the aspects of objects that are separable from the descriptions of the positions of those objects (Zhang et al, 2002). Although arguably the position of objects in and of itself can be considered an attribute, attributes of objects more often deals with the properties associated with entities. Typically attribute error is discussed in terms of the accuracy and precision with which values of a property are associated or defined for an object. Attributes for an object may be associated with a set of nominal properties as in the case of distinguishing between a house or a school in an urban model or may distinguish between soils and land uses such as the Hydrological Soil Grouping or (HSG) classification scheme or in distinguishing land uses such as agriculture versus woods in the case of a hydrological application. If properties of objects are defined in such a way as having rank, any number of ordinal attributes may be considered or, in the case of quantitative or continuous attributes such as elevation or soil moisture, values may be any real number within a logically acceptable range.

3.6.7 Sources of Attribute Uncertainty

In many cases, attribute uncertainty may be related to the process of membership classification where attribute accuracy refers to the degree with which an object is correctly classified. In this case, the vagueness of class assignment discussed earlier, can have an effect on the attributes associated with a particular entity that is given membership in that class. Uncertainty of this sort is often related to errors of omission and errors of commission where objects classified based on properties are left out of an appropriate class or added to an inappropriate class, respectively.

Errors associated with data capture are introduced through imprecise digitising of paper-based data, imprecise data entry by individuals populating a database, or imprecise measurement of properties by field or lab persons and equipment. There are numerous sources of attribute accuracy which like positional accuracy, are related to the processing or manipulation of data. For example, the classification of data from RS imagery into various land use classes is often error prone and imprecise. Various authors have discussed these issues in detail. The interested reader may refer to Longley et al, 2002) Zhang et al, 2002; Shi et al, 2002; Heuvelink, 1998; Lodwick et al, 1990 and Goodchild et al, 1989)

3.7 Sensitivity Analyses and Uncertainty Analysis

The term uncertainty analysis may be related to error propagation (Heuvelink, 1998), error modelling (Goodchild et al., 1992) or geographical error analysis (Lodwick et al., 1990) within the literature. Conceptually, uncertainty analysis can be considered a quality test for a particular model and its input data by consideration of all quantifiable sources of error simultaneously (Crosetto et al., 2001). Sensitivity analysis (SA), often referred to as error analysis (Heuvelink, 1998), quantification of error contribution (Arbia et al., 1998), or geographic sensitivity analysis (Lodwick et al, 1990), can be defined as the study of all information flowing in or out of a model (Saltelli et al, 2000). More specifically however, SA refers to one or a series of processes carried out in order to identify how much total (global) model uncertainty can be attributed to the uncertainty associated with each individual model factors including, all model parameters, inputs, variables, attributes, and outputs.

3.7.1 Usefulness of Sensitivity Analysis

There are numerous reasons to perform a sensitivity analysis. Saltelli et al, (2000) offers the following reasons:

- *To identify if a model resembles the system or processes under study*
- *To identify the model factors that mostly contribute to the output variability of the model and which may require additional research to strengthen the knowledge base.*
- *To identify the model parameters that are insignificant and, that can be eliminated from the model.*
- *To identify if and which group of factors interact with each other*
- *To identify if there is some region in the space of the input factors for which the model variation is a maximum and*
- *To identify the optimal regions within the space of the factors for use in a subsequent calibration study.*

It is intended that through application of UA and SA in this study that further knowledge can be gained as to the applicability of GIS and SCS based hydrological models as predictive tools for hydrological responses to land use change. Further, it is intended through sensitivity analysis of identifiable sources of uncertainty in the model, that an overall measure of importance for the model factors examined will be produced. This, subsequently, provides an indication of areas requiring further research or more careful attention prior to model application.

3.7.2 Methods for performing UA and SA

Uncertainty Analysis

While a variety of methods for uncertainty estimation exist within the literature (see Helton, 1993 for a review), the most popular methods for analysing total model (global) uncertainty are those related to the Monte Carlo simulation (MC). The Monte Carlo simulation is belongs to the family of global uncertainty methods often grouped under the auspices of "sampling-based techniques" all of which involve the generation and exploration of uncertain from analysis inputs to analysis results (Helton, 1993). The goal of these types of analysis is essentially an understanding of the how uncertainty in a model output will depend on the combined uncertainties of model inputs. The term input in this case however should be used loosely as depending on the analysis this could potentially be applied to all model factors by the definitions established in Chapter 1. Typically a MC analysis relies on the repeated sampling of model factor values according to some sampling strategy and the computation of model outputs. The number of iterations (trials) used in a MC analysis reflect the complexity of the model, the number of factors, their stochastic properties and typically is reflected in the sample size necessary for insuring statistical significance of the simulation results.

While numerous other UA methods have been differentiated from MC analysis within the literature, these methods are fundamentally no different in application but differ largely in the sampling strategy used. While random sampling is the most common method used, Campolongo et al (2000) differentiate between numerous sampling strategies including stratified sampling, latin-hypercube sampling and simple random sampling, all of which are unique sampling methods used in conjunction with the MC approach.

Justification for use of different sampling strategies may be made in cases where the sampling space of model factors necessitates a more complex strategy to insure that that the number of trials used in a simulation is statistically representative of the overall sampling space. For example, stratified random sampling differs from simple random sampling in that the sampling space is divided into a series of disjointed strata within which a random sample is obtained whereby the likelihood of a unified coverage of the sampling space is increased. Latin Hypercube Sampling, considered a special case of stratified sampling, divides the range of model factors into N equal intervals of marginal probability whereby a sample set is created by drawing one observation from each interval (Saltelli et al, 2000).

The use of one sampling over another will have an impact on the outcome of the analysis and should be justified based on the type of model and the number of parameters. For example, in a study by Yuh et al, (2001), examined various sampling strategies used to evaluate uncertainty in an event-based distributed rainfall-runoff model. Strategies used included a simple random MC approach, the Latin-Hypercube strategy (LHS), Rosenbleuth's Point Estimation Method (RPEM) and the Harr's Point Estimation Method (HPEM). Results indicated that the LHS strategy required only ten percent of the number of trials to produce similar results to the simple Monte – Carlo method but that the other two methods produced results suggesting that they are not suitable methods where small numbers of model parameters are used for calibration.

Other sampling strategies focus on the series of methods including those related to Markov Chain sampling (Gilks et al., 1996), include the Metropolis sampling method (Metropolis et al., 1953) have also been widely applied within the literature as a means of evaluating uncertainty among equally effective parameters sets. This application was demonstrated in a study by Mailhot et al. (1997), where a Metropolis based MC analysis to examine the effects of data uncertainty on a lumped catchment model created using the Storm Water Management Model (SWMM). This study identified the Metropolis MC approach as a useful approach for evaluating acceptable parameters used in model calibration given uncertainties in calibration data. Further, in a study by Kuczera and Parent (1998), the Metropolis MC method was compared to importance sampling, a strategy demonstrated in a paper by Beven and Binley (1992) on the Generalized Likelyhood Uncertainty Estimation (GLUE) technique, another Monte Carlo based technique for examining uncertainty in parameter calibration sets. Results identified that the Metropolis method produced better results with fewer sampling trials when compared with the GLUE method.

The Monte Carlo method of uncertainty analysis is the most popular means of performing global model uncertainty. While a highly diverse set of sampling algorithms exist, should be applied with caution as these methods can have significantly different results. Care should be taken to justify the application of a more complex method over simplest methods, especially when a small number of parameters are being examined and computing power is not of crucial concern.

Sensitivity Analysis

Pure sensitivity analysis has been classified into two generalised methodological approaches being local SA or global SA, where local SA measures typically involve a concentration on the localized impact of a single factor within a model. The use of local sensitivity analysis is, however, primarily carried out through the computation of partial derivatives of model output functions with respect to small changes in model input. Applications are often used to solve problems of chemistry and physics, and are used to solve issues such as the inverse problem which relates to the back calculation of kinetic constants for outputs whose measurements are not directly measurable (Saltelli, 2000).

Global sensitivity analysis is primarily concerned with the process of apportioning a model's total output uncertainty to uncertainty in a models input factor in order to ascertain those input factors which contribute most to the overall uncertainty of the model. Saltelli (2000) has defined global uncertainty analysis based on two properties:

- *The global property, including the influence of scale and shape where the sensitivity of estimates of individual factors incorporates the effect of the range and shape of their probability density functions and,*
- *Multidimensional averaging where the sensitivity estimates of individual factors incorporates the effect of the range and the shape of their probability density function.*

With respect to global sensitivity, numerous methods have been reviewed by Saltelli, (2000) and Hamby, (1994) and are described in the literature. Such techniques include ANOVA and other variance-based techniques, the use of bootstrap (Archer et al, 1997), non linear methods (Sobol, 1993) the Fourier Amplitude Sensitivity Test or FAST, (Cukier et al, 1979); Extended FAST, (Saltelli et al, 1999), as well as a series of one at a time (OAT) approaches (Daniel, 1958). While there are multitudes of applications of sensitivity and uncertainty analysis within modelling in general, a comprehensive review is largely beyond the scope of this thesis.

3.8 Application of UA and SA to GIS and Hydrological models.

Numerous applications of sensitivity and uncertainty analysis have been performed in the fields of hydrology, although, to a lesser extend in GIS and include the application of OAT techniques, local measures of sensitivity as well as variance based methods.

The one-at-a-time approach (OAT) is the simplest class of techniques used in performing SA. In standard OAT designs, the impacts on model output are observed through changing each model factor in turn. Typically nominal values for all model factors are taken from the literature or derived *in situ* . The combination of all nominal model factors is considered the control scenario and is either represented by the midway of two extreme values for each factor (Campolongo et al, 2001), or in some cases represents the combination of factors used to achieve a best fit for the calibration/validation of a model of interest. The factors are then adjusted about the range of extreme values and the corresponding model response is "mapped" for each factor independently. While OAT designs do not allow for an investigation of the interaction between model factors (Campolongo et al, 2001), they are a useful means of sensitivity analysis where the purpose of the analysis is to perform screening for more complex analysis or when it is expected that random error is small compared to systematic error within model factors.

Some examples of OAT analysis applied to hydrological models include those of Ibbitt, (1972) who examined the effects of random data on a conceptual rainfall-runoff model finding that random errors caused no significant change in parameters values while systematic bias as a result of missing data could be considered as the cause of variability in final parameter values. Singh et al, (1976) and Singh, (1977) used a one at a time approach to examine the sensitivity of linear and non-linear rainfall-runoff model structures to systematic errors in rainfall. Findings included that while the behaviour of watersheds are commonly accepted to be non-linear, systematic errors in rainfall tended to overpower the robustness of non-linear systems suggesting the use of linear models would be preferable in some cases. This was expanded on in the latter study where five linear and non-linear models were compared. Findings suggested that a perfectly identified non-linear model cannot be uniformly better than an optimally identified linear model when error in rainfall is considered, but that the degree of non-linear behaviour of the system will very the rigidity of the conclusion. Paturel at al, (1995) also utilised OAT sensitivity to examine the effects of systematic and random errors on model inputs to a simplified catchments model and found that under various circumstances systematic as well as random error tended to amplify the response of peaks and troughs but that these effects did not magnify over longer time periods. However, the errors were mimicked in model output such that optimal solutions could be achieved through calibration.

The use of local sensitivity analysis has also been used in a number of cases to examine models associated with evaporation estimation. Approaches of this type evaluate sensitivity of model factors through calculation of rates of change of model outputs based on small changes in each model factor. Typically, these approaches do not attempt to derive sensitivity estimates for more than first order interactions due to difficulties in deriving the first order partial derivatives of the system known as the Taylor series (Campolongo et al, 2001).

A number of studies utilizing this method have however been applied to a variety of sensitivity studies within the hydrological context. McCuen, (1974) used a First order Taylor local sensitivity method to examine the effects of variation in meteorological factors on evaporation estimates produced in conjunction with three evaporation estimation model structures including the Fractional Factoral Method, the Weather Bureau Method and the Penman model. Studies indicated that evaporation estimates resulting from measurement error in meteorological factors were significant and that use of the Penman model was less sensitive to these errors and preferable to the other two.

Saxton (1975) also utilized the First order Taylor method to evaluate sensitivity of the combined aerodynamic-energy budget method for estimating evapotranspiration. Results indicated that the model was most sensitive to errors in net radiation. Beven, (1979) performed sensitivity analysis of the Penman-Montieth equation on actual evapotranspiration using also using the First order Taylor indicating that the Penman-Montieth method is sensitive to values of aerodynamic and canopy resistance parameters which introduce influences of vegetation type into predictions more than climate parameters. Finally, a First order Taylor was used by Piper, (1989) to examines the sensitivity of input errors on the Penman-Montieth equation suggesting that this model was most sensitive to temperature.

While there have been some recent attempts at application of sensitivity and uncertainty analysis on GIS-based models, these have been largely non-comprehensive and fail to address sensitivity / uncertainty in a way appropriate to addressing the needs of successful coupling of GIS-based hydrological models. While Lodwick et al, 1990 examined the numerous potential sources of uncertainty involved in map operations, this is primarily applicable to aspects of site suitability assessment and tended to focus exclusively on raster operations. Mendicino et al, 1999 provides a sensitivity analysis of specific GIS procedures but concentrates primarily on sensitivity of multiple algorithms in the approximation of a single attribute. Mckenney et al., (1999) conducted an OAT sensitivity analysis on a multiple scales within a GIS solar radiation model. However, input specific sensitivities were not exhaustively dealt with.

One very promising comprehensive UA/SA analysis has been proposed and demonstrated by Crosetto et al., (2000), Crosetto et al., (2001). These studies emphasise the use and benefits of performing UA and SA in conjunction with spatially based GIS modelling and suggest the use of a stepwise method for SA. The method, based on an approach suggested by Crosetto et al, (2001), is augmented to support the complexities associated with performing sensitivity analysis on GIS based models. After identifying a suitable test site as well as sources in the model that could contribute to overall uncertainty, a global UA is performed utilising a simple MC approach and suitably determined error models for each factor of concern. The list of model factors is then analyzed in combination with the results of the MC and an OAT screening test proposed by Morris (1991). Once important model factors are determined through application of the screening test, a more complex variance-based SA approach is then applied to compute the first and n^{th} order interactions, without concern for lack of computing power. In the earlier study, the use of a screening test was not necessary as the model considered was dependent on a small number of factors however, in the later study a series of fifteen model factors were first ordinally ranked using the OAT screening test of Morris (1991). This followed with the application of the Extended FAST technique and an analysis of first and n^{th} order sensitivities.

The use of this approach suggests a promising framework for performing quantitative SA on GIS based hydrological models. However, in many cases error models for many of the model factors are assumed (necessarily) using expert judgment. This may be easier to do for model factors associated with physically-based modelling as parameter values can more easily be bounded however. In the case of empirically based models such as the SCS-curve number method, model parameters are representative of the combined effects of numerous physical processes and more analytical methods may be necessary for the provision of suitable error model definition. In the following chapter on methodology, the approached outlined by Crosetto et al (2000; 2001) is suitably adapted to applications dealing with vector GIS and the SCS-Curve Number method. A MC approach in combination with descriptive spatial statistics will also be used to determine a suitable error model for the SCS-Curve Number as a result of spatial error in soil boundaries and ambiguity of soil and land use classes. The results of this will be used in conjunction with other appropriately identified error models and the stepwise approach previously described will be utilized to perform a combined UA/SA on a GIS-based hydrological model developed using a revised SCS-Curve Number Method.

In the introduction as well as in sections 3.1-3.2, it was identified that the SCS-Curve Number technique is used as a core model structure for a number of different predictive hydrological models. The SCS method has further been utilised in a number of models that rely on GIS systems for managing model attributes and data input. Various methods for recognizing potential uncertainties in the application of this approach have been identified in the previous section. Identifying how these uncertainties impact model performance is an important process for evaluating the model as a predictive tool. The following case study and accompanying methodology will examine various approaches that can be used for determining the implications of uncertain data when used in conjunction with a GIS based Curve Number approach to hydrological modelling.

4.1 Procedure for Performing Uncertainty and Sensitivity Analysis

Reasons for performing UA/SA have been discussed as a commonly accepted BMP for hydrological modelling in Chapter 1, and have been proposed as a necessary BMP for hydrological models that rely on GIS procedures in Chapter II. Further, in the previous chapter, the specific motivations for performing UA/SA have been reviewed in conjunction with various methods and their relative advantages and disadvantages. In this Chapter, a series of methods for performing UA /SA on GIS based hydrological models is introduced as well as the modelling methodology for use in the case study described in Chapter 5.

Saltelli et al, (2001) and Crosetto et al, (2000; 2001) outline a useful procedure for performing sensitivity analysis of spatial modelling based on GIS. The procedure utilizes a combination of global model uncertainty analysis in the form of a comprehensive Monte Carlo simulation in combination with OAT-based screening test and a quantitative n^{th} order extended FAST (Saltelli et al, 1999), to identify total model uncertainty and quantify that uncertainty by partitioning it to each model factor and its n^{th} order interaction for which a suitable error model has been identified (see Figure 4.1).

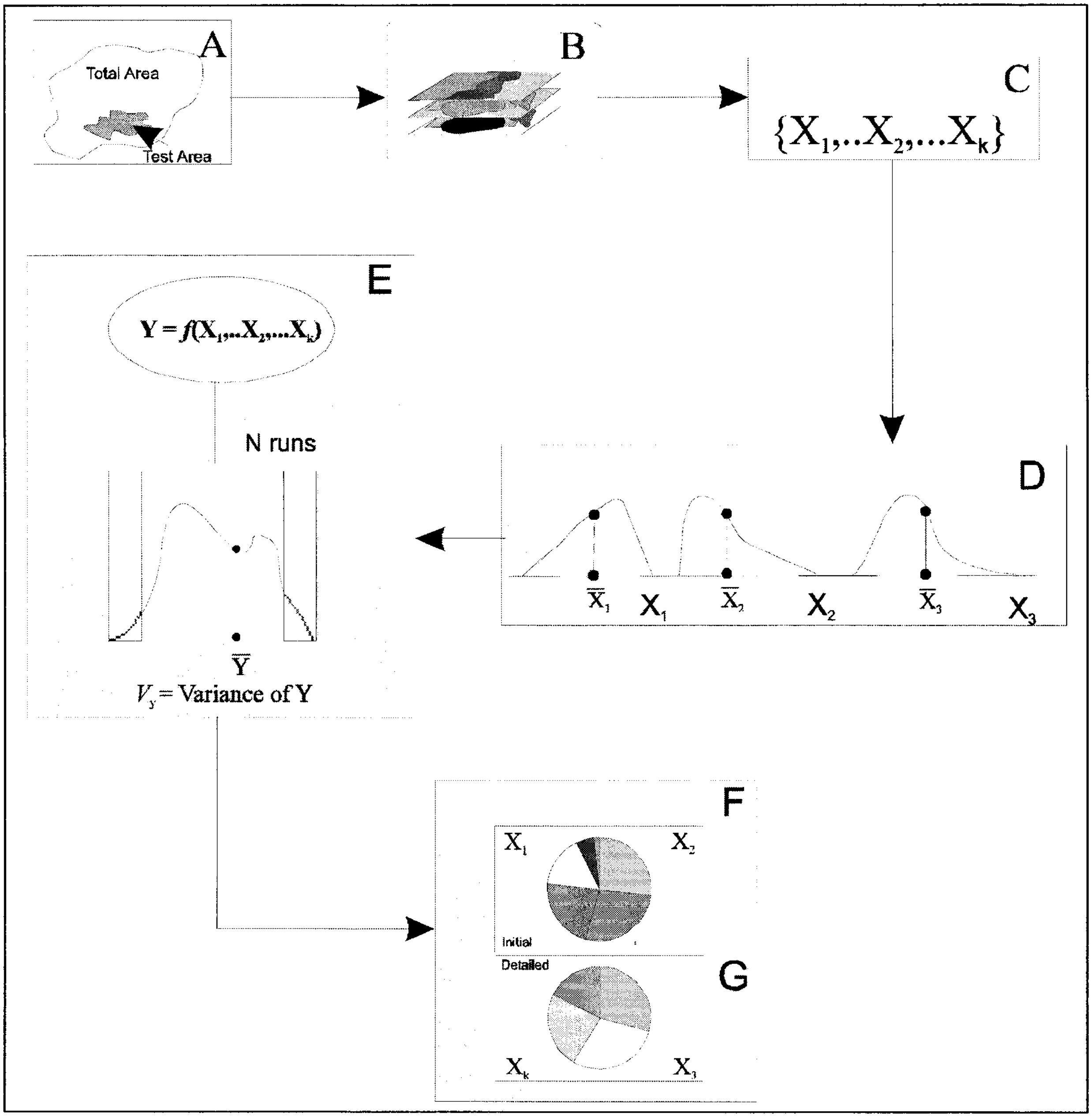


Figure 4.1. Illustration of a comprehensive series of steps for performing an uncertainty and global sensitivity analysis on GIS-based hydrological models. Source: Crosetto et al, (2001)

The procedure is performed according to a series of steps (Crosetto et al 2001), including:

- **A**, Identification of an appropriate test area
- **B**, Development of the model
- **C**, Identification of areas of model factors that are likely to contribute to the total model uncertainty.
- **D**, Development of an appropriate stochastic error model for each identified factor of interest capable of generating a series of distorted realizations.
- **E**, Performance of a pre-assessment in order to determine major sources of sensitivity
- **F**, Performance of global model uncertainty analysis.
- **G**, Performance of a detailed sensitivity analysis.

These steps offer a suitable methodology to follow for the demonstration of the case study. While some steps will be altered to suit the needs of the investigation, the methodology used is discussed throughout the remainder of this chapter according to the various steps identified in the previous procedure.

4.1.1 Identification of the Test Site

In identification of suitable test site, the investigator must give careful thought as to whether the chosen site represents a suitable system for performing the investigation. In such cases, it is highly desirable to have a test site that is representative of the larger environment of interest but also small enough so as not to impose unnecessary complexity.

The environment chosen for the case study is the Duffins Creek Watershed (Figure 4.2) The site is considered suitable for a number of reasons including its proximity to the greater Toronto area (GTA) and the availability of land use and soil data. Further, the Duffins Creek watershed is the subject of a number of other studies supporting the development of predictive and descriptive models for management purposes.

The Duffins Creek watershed is located within the north-eastern limits of the GTA and extends from just north of the community of Whitchurch-Stouffville, south to Lake Ontario. The majority of the watershed lies within the regions of Durham and York.

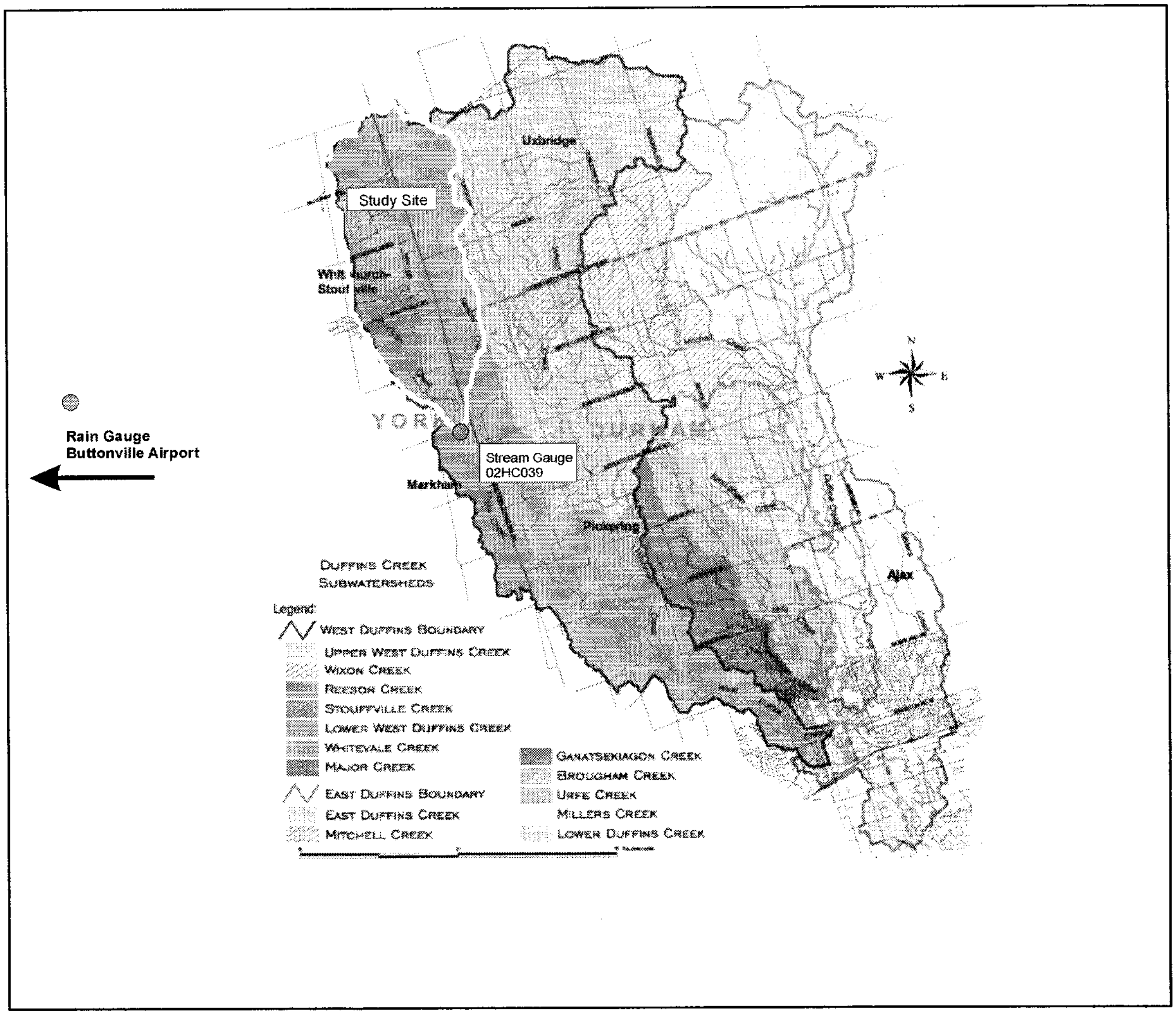


Figure 4.2. The Duffins Creek watershed with associated tributaries. The Reesor Creek sub-watershed is identified in the top left corner in yellow as are associated rain and stream, gauges. Adapted from: TRCA, 2003b.

Populated centres within the watershed include the communities of Whitchurch-Stouffville, Markham, Uxbridge, Pickering and Ajax (TRCA, 2003b).

Land use within the Duffins Creek watershed remains primarily agricultural (TRCA, 2003a) with approximately 50 percent of land currently in public ownership (TRCA, 2003b). Land use within Duffins Creek is however changing rapidly as numerous pockets of urban and commercial development are beginning to emerge and recent population growth statistics for the area suggest that pressure for development will likely continue for the next 25 years (TRCA, 2003b.)

Physiographic characteristics within the northern extent of the Duffins Creek watershed is dominated by the Oak Ridges Moraine, an area typified by a higher elevation than other areas of the watershed and consisting of glacial deposits consisting of sand, till and gravel. The central regions of the watershed are represented by the Halton Till plain and are typified by a physiography of rolling hills and stream valleys. Soil in this region is a mixture of agriculturally valuable loams as well as some surficial clay deposits within proximity to the communities of Whitchurch-Stouffville and Clarimont. The southern portions of the Halton Till plain consist of steeper gradients corresponding to the historical shoreline of glacial Lake Iroquois. The remaining bottom third of the watershed is considered to be part of the glacial Lake Iroquois plain where soils are typified by various mixtures of sand, silt and clay.

Duffins Creek has a number of significant tributaries, including Reesor Creek, Stouffville Creek, Wixon Creek, Whitevale Creek, Major Creek, Urfe Creek, Brougham Creek, Ganatsekiagon Creek and Mitchel Creek. Annual flow characteristics within the watershed are considered to be the last in the GTA that are typical of an agricultural as opposed to an urban watershed (TRCA, 2003b). Each sub-watershed within Duffins Creek is subject to local and regional characteristics and contains a variety of flora and fauna described at length by the TRCA, (2003a,b).

Given the complexity of modelling the entire Duffins Creek watershed, it was decided for the case stud that a portion of the overall watershed corresponding to the Reesor Creek and Stouffville Creek sub-watersheds would be modeled as a suitably representative test site (see Figure 4.2).

This particular area was chosen for three reasons:

- A reasonable stream gauge and precipitation record was available for a three-year time span.
- The land-use and soil characteristics were representative of the overall watershed and possessed pervious rural as well as impervious urban land use characteristics.
- A more complex model had been already been developed without sensitivity and uncertainty analysis for the same sub-watershed, the results of which were available for comparison.

4.1.2 Model Development

As identified in Chapter 3, there have been numerous hydrological models that rely on the SCS-Curve Number technique as a primary model structure used in the prediction of surface runoff. The SCS-Curve Number technique is frequently used in conjunction with GIS-based hydrological models such as the AGNPS model (TRCA, 2003; Al-Smadi, 1998; and Yagow, 1997), the OTHYMO model and XSRAIN, (Correia, 1998) and, is a popular model structure adopted for use with GIS loosely coupled models for a number of reasons including:

- The model's simplified structure
- The relative availability of spatially distributed land use and soil data
- The lack of field work required and
- The relative ease in loosely coupling the Curve Number structure with GIS output

The SCS-Curve Number technique was developed largely as a response to the United States 1935 Soil Conservation Act as well as Flood Control Act of 1936 and was largely developed over a period of 20 years between 1930 and 1950. Based on the analysis of empirical data gathered throughout experimental catchments throughout the United States, the SCS-Curve Number technique was developed to provide engineers with a simplified means of estimating runoff excess from a watershed given a measurable amount of rainfall, a set of identifiable soil classes and a dominant classification of land use types.

The SCS-method is applied using a set of empirically derived relationships equating rainfall depth to excess runoff where:

$$Q = (P\text{-}Ia)^2 / (P\text{-}Ia) + S \tag{4.1}$$

and:

$$S = 254(100\text{-}CN)\text{-}CN \tag{4.2}$$

and:

$$Ia = S^{0.55} \tag{4.3}$$

Q = depth of runoff
P = depth of incoming precipitation
Ia = Initial abstraction losses
S = The amount of soil water storage available
CN = The SCS- Curve Number

Cover Description		Curve Number for Hydrological Soil Groups				
Cover Type	**Treatment**	**Hydrologic Condition**	**A**	**B**	**C**	**D**
Fallow	Bare Soil	-	77	86	91	94
	Crop Residue Cover (CR)	Poor	76	85	90	93
		Good	74	88	83	90
Row Crops	Straight Row (SR)	Poor	72	81	88	91
		Good	64	75	82	85
	SR + RC	Poor	71	80	87	90
		Good	64	75	82	85
	Contoured (C)	Poor	70	79	84	88
		Good	65	75	82	86
	C + CR	Poor	69	78	83	87
		Good	64	74	81	85

Table 4.1. The SCS-curve number procedure relies on a set of curve numbers relating rainfall to runoff for a variety of soil and land use classes. Source: Mayes (2001).

The SCS-Curve Number parameter is the only parameter that needs to be identified in order to apply the model. Curve Numbers (CN's), are typically derived from a series of tables (see Table 4.1) relating ranges of Curve Numbers to various local combinations of soil corresponding the HSG soil classification

scheme and local land uses. Curve Numbers are based on empirical data derived from long periods of observation about the rainfall-runoff relationships throughout numerous experimental catchments each expressing a particular soil and land use combination that is depicted in the CN lookup table. The empirical procedure used is dependent on the derivation of a rainfall-runoff curve for a single catchment during its peak seasonal rainfall event. The CN is established through fitting empirically derived rainfall-runoff curves from the experimental catchment with a series of CN curves based on Equation 4.1.

Although these methods can theoretically be applied to any catchment for the empirical derivation of a catchment specific Curve Number, this is seldom done as it defeats the underlying simplicity for which the CN lookup method was created. Further, the empirical approach relies on the gathering of a number of years of data. This is necessary as correct approximation of representative rainfall-runoff relationships depends on empirical data for a catchment's maximum seasonal rainfall. Further, in theory, the empirical approach can only be performed on catchments with homogenous soil and land use features, a highly unrealistic assumption.

While land use types depicted in CN lookup tables are for the most part ambiguous numerous classes exist which represent a broad range of anthropogenic and natural cover types. Soil type variation however, is restricted to four macro categories of soils known as the Hydrological Soil Grouping classification (HSG) as introduced in Chapter 3. The HSG is divided into four categories A, B, C, and D, which are loosely associated with the following descriptions (Mayes, 2001).

A = *deep sand, deep loess or aggregated silts*
B = *shallow loess or sandy loam*
C = *clay loams, shallow sandy loams, soils low in organic content or clay soils*
D = *soils the swell upon wetting such as heavy plastic clays or saline soils.*

In order to facilitate simplicity without sacrificing validity, the proposed model subscribes to a number of approaches that have been reviewed in Chapters 2 and 3. The model consists of two components, a GIS watershed model utilizing soil and land-use data as well as a mathematical mass balance hydrological component which will rely on certain model parameters identified through GIS analysis. The mathematical component of the model is loosely coupled with the GIS and is developed through the Microsoft Excel environment. The Excel environment offers a number of advantages including the ease with which data exchange can be facilitated between the mathematical model and the GIS, its user friendly environment and its compatibility with other software used for more advanced analysis.

The mass balance approach adopted in this research uses a simplification of a two-tiered rainfall-runoff model originally developed in 2000 by Clarifica consultants. The model was initially developed for analyzing future development scenarios within watersheds in the GTA and has been adopted by the Toronto and Region Conservation Authority as a predictive tool. The hydrological rainfall-runoff model uses a single table and is dependent on the input of daily averaged pan evaporation and precipitation data for continuous simulation as well as measured daily average streamflow for calibration and validation.

Continuous stream gauge data from the test site was used for the period of April to October 1997 and 1998 and 2000. The time period selected was a function of the available data, data quality and correspondence with available continuous precipitation and evaporation records. While the streamflow data were considered good quality and were continuously available for longer time periods, these were limited by the availability of rainfall records that coincided with stream records.

Streamflow was measured at TRCA station number 02HC039 just south of the confluence between Reesor and Stouffville Creeks at the southern tip of the study site (see Figure 4.2 and Appendix III). In order for a logical comparison to be made with rainfall data, the measured streamflow in cubic meters per day was converted to an average excess depth over the sub-watershed area, which produced a continuous series of daily average excess depth in millimeters.

Two continuous rainfall records from Buttonville airport and Cherrywood meteorological station were available from January 1997 – January 1999 and January 2000 to December 2000. Unfortunately, due to the lack of monitoring for meteorological variables, neither of these stations are within a suitable distance to the study site. Further, the Cherrywood station, although closer to the site, was not used as it was previously shown to produce poor results when used in a similar modelling approach while, rainfall events measured at the Buttonville rainfall station was shown to conform reasonably well with peak flow events when plotted against the corresponding stream gauge (Clarifica 2002).

Evaporation records for the time series were retrieved from Clarifica (2002) and were originally taken from pan evaporation records. According to Clarifica, historical lake evaporation data were only available at three southern Ontario stations:

- Hamilton RGB station (in Burlington)
- Lindsay Frost, and
- Peterborough (Trent University)

The Hamilton station contained data from 1986 to 1996 and was the closest to the study area however, operation of all three stations ceased as of 1996. In order to derive an approximate measure of daily potential evaporation, an averaged seasonal evaporation function was derived by Clarifica and used to produce evaporation data for the period of calibration from 1997 to 1998. The algorithm used produces a seasonal trend based on

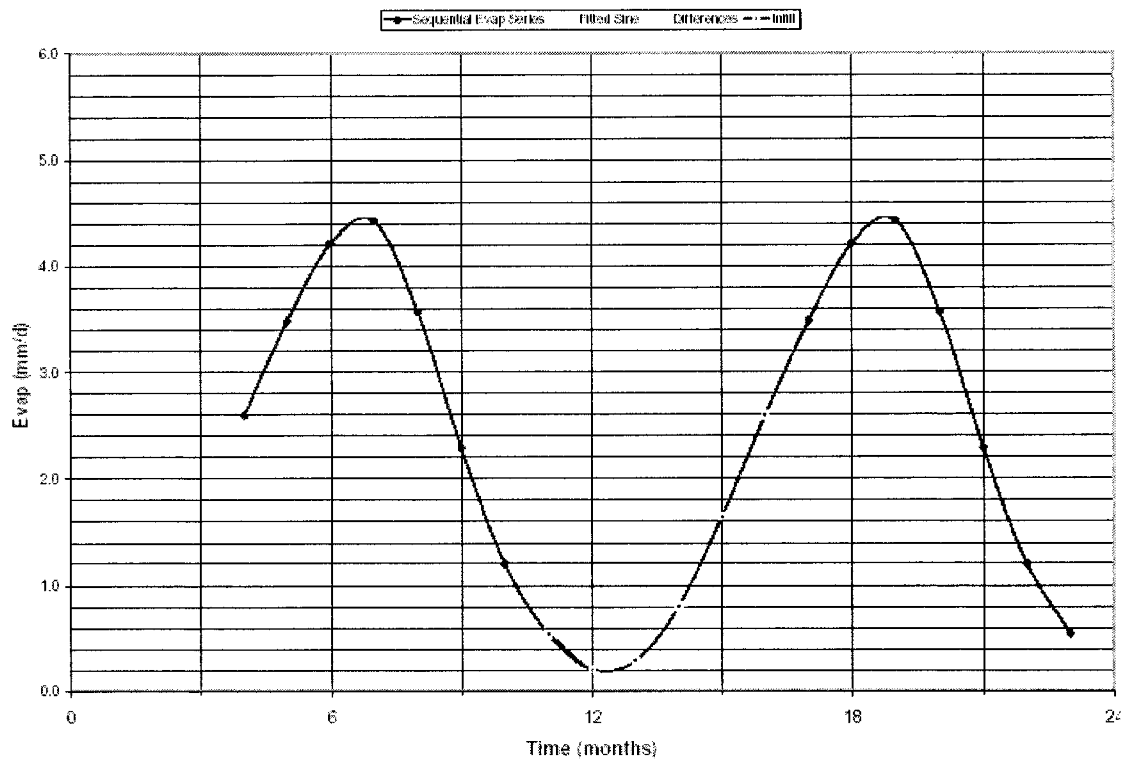


Figure 4.3 Average evaporation estimates and corresponding sine function used to extract daily evaporation estimates. Source: Clarifica 2002.

average values collected during operational periods for which a sine-function curve (Figure 4.3) is derived and from which daily average evaporation for the period of study was extracted and used as model input (Appendix IV).

The research model uses a mass balance approach and calculates runoff from precipitation falling on impervious and pervious areas separately (see Figure 4.4, 4.5). Precipitation falling on impervious areas can be divided into two types, directly connected and indirectly connected impervious areas. An area that is directly routed to streamflow such as a drainage ditch or other storm water infrastructure is considered to be a *directly connected impervious* area since conceptually this input is channelled directly to

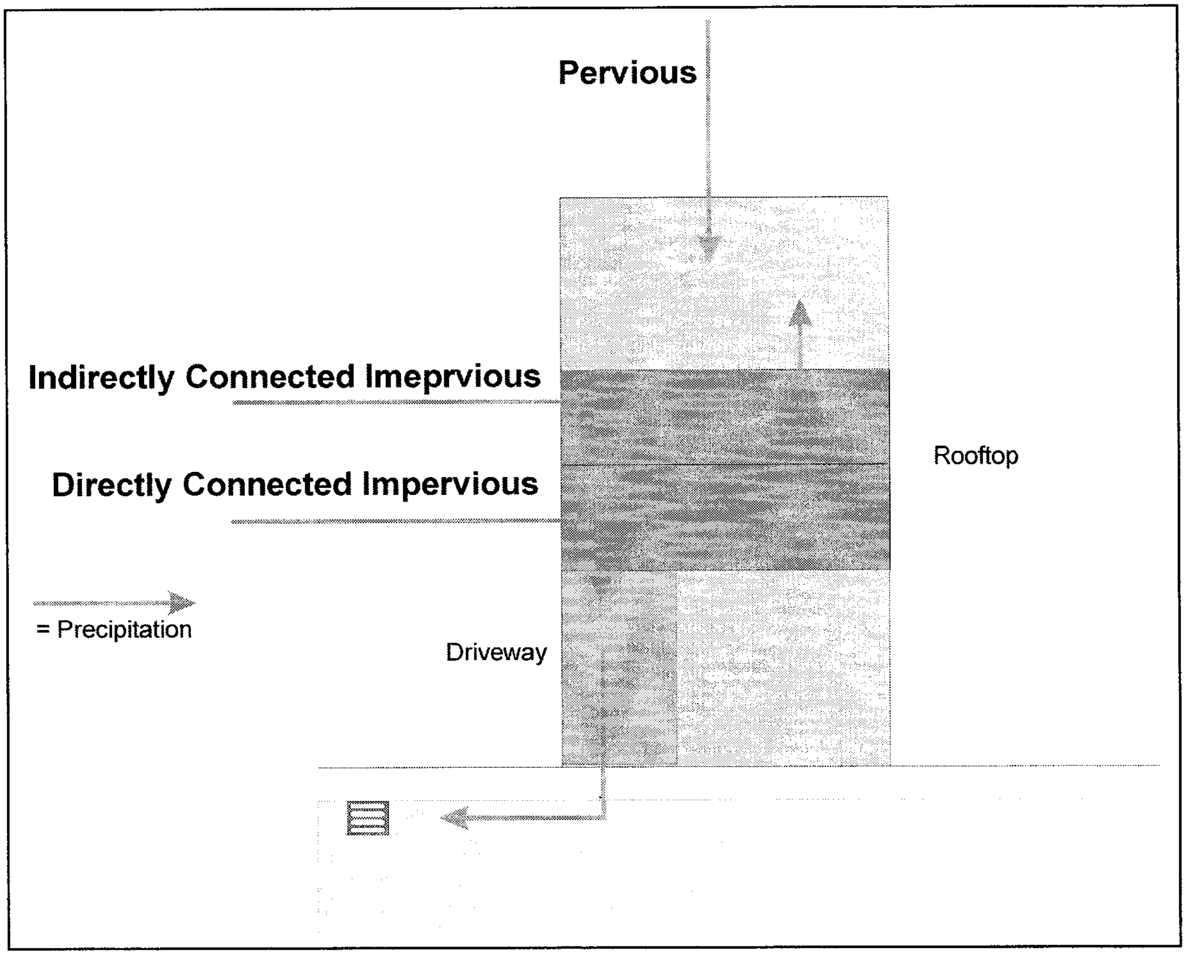


Figure 4.4. Precipitation entering a system and falling on impervious surfaces may be either directly or indirectly connected to streamflow.

streamflow and, there are no further losses to this input prior to it exiting the system. Directly connected impervious areas are typically indicative of an urban landscape and precipitation falling within this type of a catchment is intentionally and quickly routed directly to local natural surface waters by way of conduits or storm drains. Precipitation falling on directly connected surfaces is quickly routed away so as not to impose a hazard or, otherwise impact the daily activities of the urban community. Runoff from urban infrastructure is typically not subject to any losses prior to being introduced into the natural environment thus precipitation falling on directly connected impervious areas, is assumed to be converted directly to runoff.

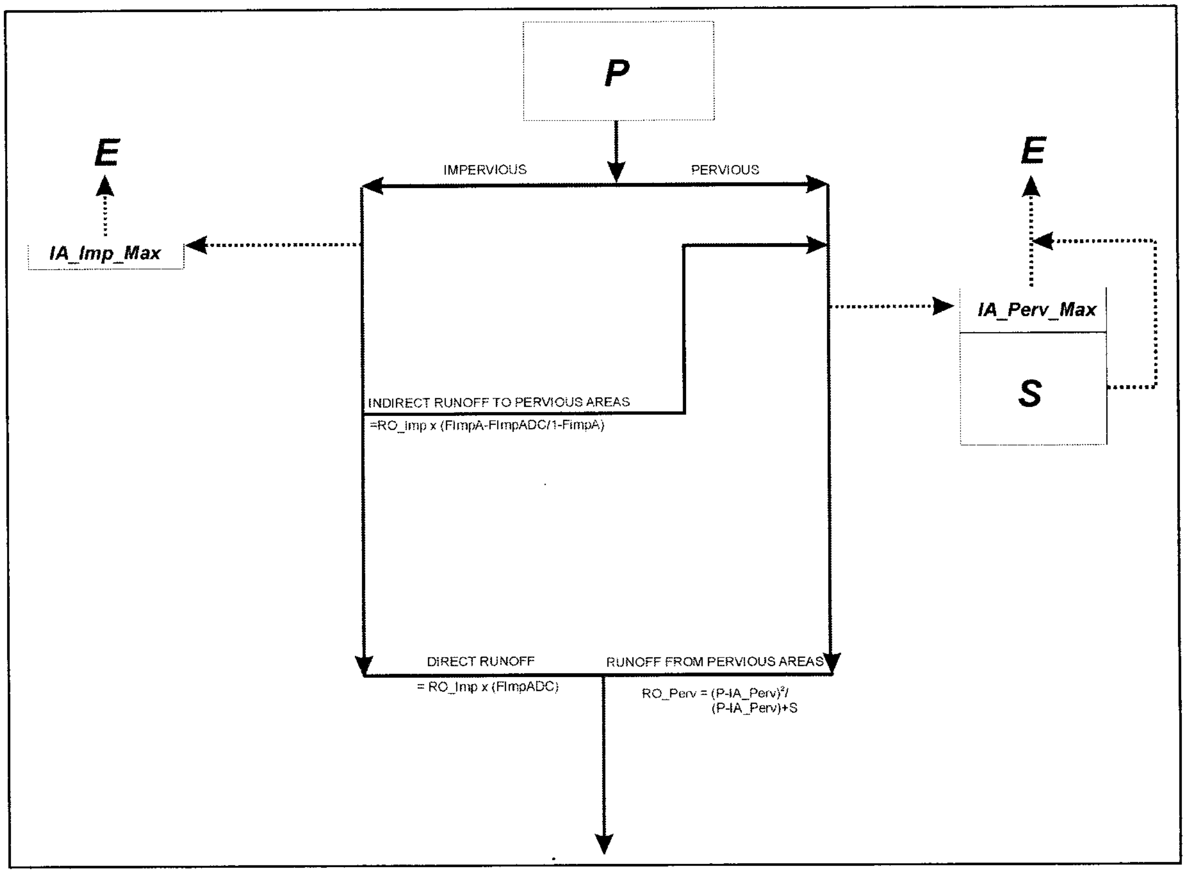


Figure 4.5. The rainfall-runoff model uses a two-tiered approach for computing runoff from pervious to impervious areas separately.

Precipitation falling on impervious areas can also be routed to pervious areas by flowing off a rooftop or sidewalk and onto a lawn or other previous area. Precipitation that falls on impervious surfaces is routed to pervious areas is said to be falling on *indirectly connected impervious* area. For precipitation falling on indirectly connected impervious area, it is assumed that this in turn is directly added to the input precipitation falling onto pervious areas and, is subject to corresponding losses from those areas. It is assumed that in a typical urban watershed, approximately 75% of the impervious area is considered directly connected while, 25% is considered indirectly connected.

Precipitation falling onto a surface prior to becoming runoff is subject to losses in the form of interception from vegetation, storage in surface depressions and losses to the atmosphere in the from of evaporation. Collectively, the combination of depression storage as well as interception by vegetation are known as *initial abstraction losses* and, are a necessary parameter for use in application of the SCS procedure. The SCS method provides a formula for estimating the quantity of initial abstraction for a system based on equations 4.2 and 4.3. The approach described by Clarifica (2000), as well as the approach adopted in the case study however, uses an adjusted SCS procedure in that, published physical estimates for both depression storage and interception loss from pervious and impervious areas respectively are used to derive values for initial abstraction within the study site. The basis behind this follows the assumption that physical values should be used in cases where feasible over and above empirical estimates.

It is assumed that water falling on impervious areas is subject to initial abstraction losses in the form of depression storage only, as strictly impervious area is considered paved and free of all naturally occurring vegetation. Once depressions in pervious areas are filled they are subject to further losses in the form of evaporation. Runoff is assumed to occur when the rate of input to impervious areas in the form of precipitation exceeds the capacity to be stored as depression storage or lost to evaporation over the study's time increment. In the case of indirectly connected impervious areas, a fraction of the runoff generated from impervious surfaces is routed to pervious areas is considered part of the input to pervious areas.

In the research model, precipitation falling on pervious surfaces is converted to runoff through the use of the SCS-method identified in Equations 4.1- 4.3. Through the application of this method, precipitation falling on pervious areas has the opportunity of being subject to losses in the form of initial abstraction as well as being stored in the soil as soil moisture. The total amount of available soil moisture storage for pervious area (***S***), is calculated using Equations 4.1 and 4.2. and, is based on the Curve Number defined for the area. Like impervious surfaces, runoff from pervious areas is assumed to occur when precipitation exceeds the storage capacity of the soil as well as the maximum potential initial abstraction consisting of both depression storage and interception losses by surface vegetation. In order for runoff to be generated from pervious surfaces, the amount of precipitation falling in a given day must also exceed the amount of moisture lost from the system due to evaporation. Total runoff calculated for the study site is a function of daily runoff produced from both pervious and impervious areas. Full details on the mass balance procedure and an example of the spreadsheet, are described in Appendices I and II.

The GIS model was developed using the ArcGIS 8.2 platform created by the Environmental Systems Research Institute (ESRI). For simplicity, data within the GIS was developed using a vector model structure. All the spatial data for the study site was provided by the Toronto and Region Conservation Authority, under a Memorandum of Understanding (2001) with Ryerson University. The data had been digitized from a number of sources including city plan maps, soil maps as well as land-use and vegetation maps, and conformed with the current data standards of the TRCA. Data sources were of diverse lineage, having a variety of map projections and input was controlled by a number of users suggesting that sources of positional error in data were of an independent nature. GIS data layers of the Duffins Creek watershed were provided by the TRCA and included such themes as watershed boundaries, soil types, land use classification, surface hydrography and transportation networks. Of these themes, the soil type and land use data was considered the most important for the study.

Each polygon in the soils data was classified according to the HSG soil classification as well as being identified as either clay, clay loam, organic, sand, sand loam and variable soil types (see Table 4.2). Soil types described as "variable" and "organic" by the TRCA were associated with the HSG class of "B' or soils having good drainage and consisting of shallow loess and sandy loams (Mayes, 1998). Further, each polygon in the land use theme provided by the TRCA was classified as agriculture/rural, meadow, urban, urban open space, federal airport lands, forest or wetlands. The attributes of each polygon also classified each area as either previous or impervious respectively. In most cases, impervious areas corresponded to the urban land use classification although some exceptions did exist.

Soil Types Classifications of the Duffins Creek Watershed and Published Descriptions of the Soil Types Corresponding to the HSG Classification Scheme			
TRCA Soil Types	Corresponding HSG	HSG Classification (Mayes, 1998)	
Clay	D	**A**	*Deep Sands, Deep Loesses, Aggregated Soils*
Clay Loam	C	**B**	*Shallow Loess and Sandy Loams*
Organic	B	**C**	*Clay Loams, Shallow Sandy Loams, Soils Low in*
Sand	A		*Organic Matter, Soils High in Clay*
Sand Loam	B	**D**	*Soils that Swell Significantly When Wet, Heavy*
Variable	B		*Plastic Clays, Saline Soils*

Table 4.2. Soil Classes used by the TRCA and corresponding HSG soil classifications.

Watershed parameters used by the mathematical model developed using the Excel platform are derived using a GIS analysis followed by an export of relative information to the Microsoft Excel environment for use with the mathematical model. The necessary parameters used by the model and determined using the GIS are the total area of the watershed, the total impervious area and the SCS-Curve Number. In order to derive these parameters, three layers of GIS data are necessary including, the watershed boundary, a soils layer and a land-use layer. The first step of the process involved isolating the GIS data relevant to the test area. This was achieved through the application of a series of GIS procedures. First, the database corresponding to the sub-watersheds theme was activated and updated by adding new field. The three sub-watersheds corresponding to the test site were then selected using an on screen selection method. The shape records in the corresponding sub-watersheds database now activated through the selection procedure were then updated under the new field with a singe value "1." Each of the polygons corresponding to the three sub-watersheds within the test site were then merged into a single polygon identifying the area of interest. This was achieved using the *Dissolve* function that merges all selected polygons having equal field values into a single shape. The soil and land use themes for the test site were then isolated from the entire watershed by using the *Clip* function. This procedure is used to create a theme by removing all information in one theme that falls outside of the boundary identified in another theme. Using the boundary theme created for the test site, the *Clip* procedure was applied to the Duffins Creek soil and land use themes provided by the TRCA such that soil and land use pertaining only to the test site was isolated into two new themes (see Figures 4.6, 4.7). These themes were then used to extract model parameters.

The extraction of total area was achieved through a simple area analysis of the watershed boundary and corresponded to an area of 3264.73 ha. The land-use theme provided by the TRCA identified each land use polygon as consisting of pervious or impervious area. Total impervious area for the test site was derived by performing an area analysis on each polygon in the land use theme.

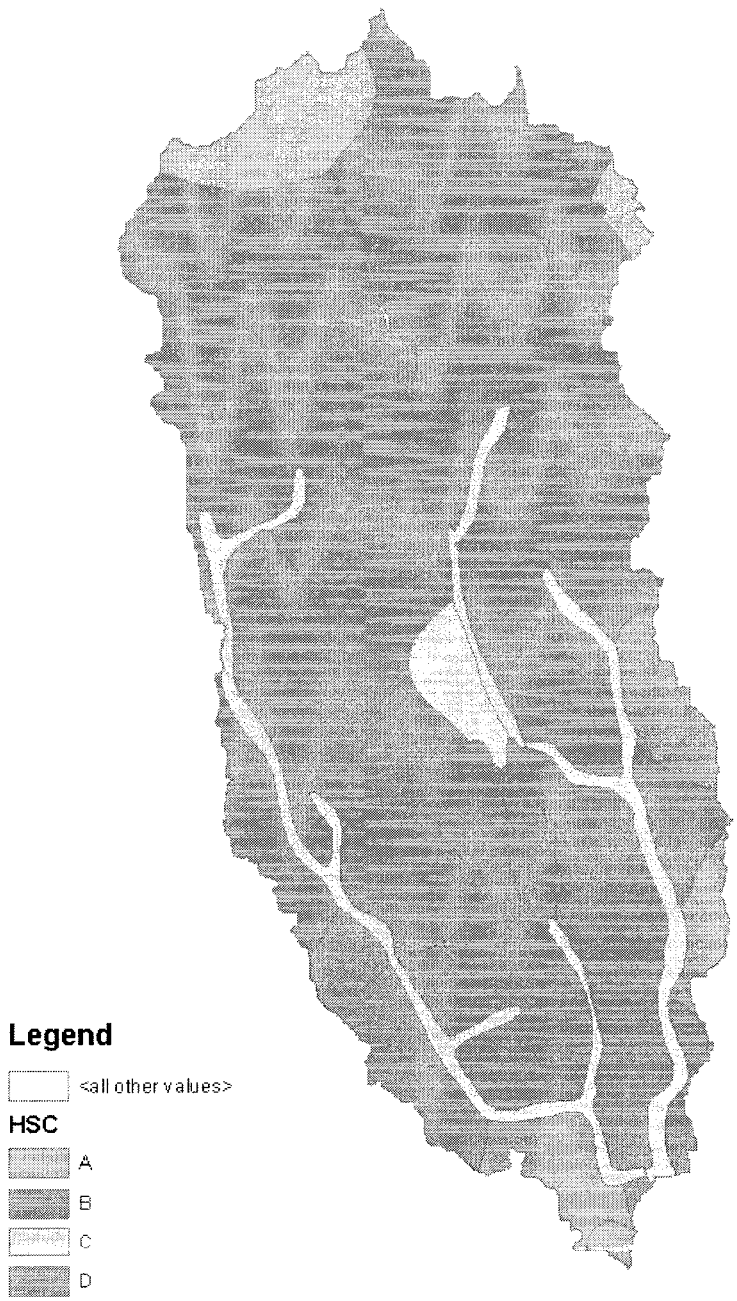


Figure 4.6. Corresponding TRCA Hydrological Soil Grouping soil classes for the Reesor and Stouffville creek sub-watershed extracted using a clipping procedure in ArcGIS.

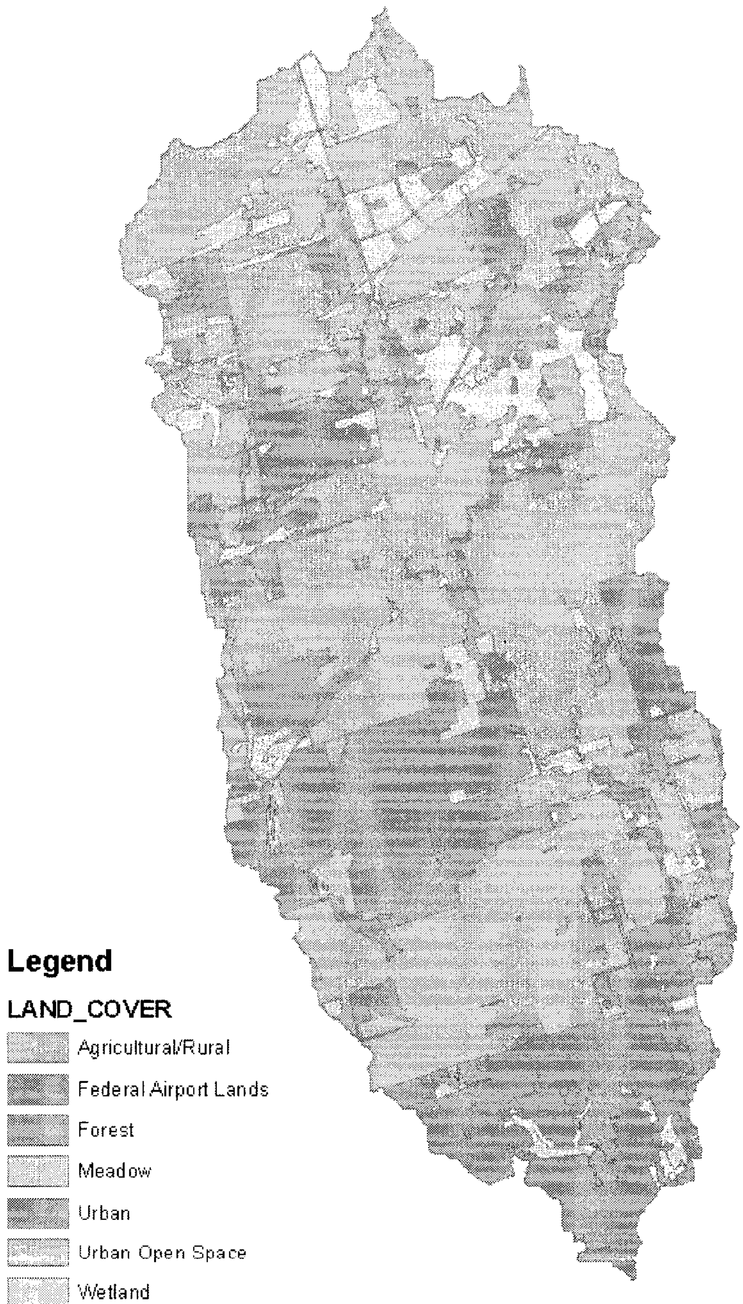


Figure 4.7. Corresponding TRCA land use classes for the Reesor and Stouffville creek sub-watershed extracted using a clipping procedure in ArcGIS.

This was performed by first selecting polygons identified as impervious by the TRCA. The selection procedure was performed by entering a query procedure using an SQL statement in ArcGIS. The total area for each impervious polygon was summarized, the results of which indicated a total impervious area for the test site equal to 146.91 ha or approximately 4.55 percent of the total area.

Identification of a lumped CN was not as straight forward. In order to determine the CN for the test area, a CN must be identified which represents the overall soil and land use characteristics of the study site. This was determined using a GIS overlay procedure. The soil and land use maps for the test area were combined using an *Overlay* procedure which resulting in a continuous and exhaustive set of polygons identifying where each unique soil type and land use co-occurred in space. The various land use characteristics provided by the TRCA however did not correspond directly to the land use classes used to identify Curve Numbers in available SCS lookup tables. In order to identify appropriate CN values, the land use classes provided by the TRCA were cross referenced with the range of similar classes identified through a survey of published Curve Number lookup tables. In all cases, however, numerous classes within published SCS-Curve Number lookup tables could, in theory, correspond to each of the TRCA land use classes. This necessitated the creation of a range of CN's which could be associated with each TRCA land use and associated HSG class. For example, the TRCA land use class of agricultural can in theory represent a wide diversity of cover types a few of which are depicted in Table 4.1. This results in a large range of potential CN's that could in theory correspond to the TRCA land use class of agriculture. Further, the TRCA land use class of Wetland without further field investigations could also be associated with a series of SCS-classes corresponding to forested and non-forested wetland (see Table 4.2).

Once a satisfactory lookup table was created, a CN was associated with each polygon record created using the overlay procedure and based on its relevant soil-land use. This was achieved by referencing each record's attributes followed by a cross-referencing of the appropriate CN range in the lookup table (Table 4.3), and assigning the median CN value to the corresponding record. Once this process was completed, an area analysis was performed on each individual polygon using the GIS. An area weighted average technique was then applied by multiplying the proportion of area for each unique polygon by its associated CN and summing across the entire set of results for each polygon. This resulted in a lumped CN value for the test site equal to 80.08.

Appropriate CN Ranges Corresponding to TRCA Land Use Classes					
TRCA LAND USE	**SCS-LAND USE(S)**	**A**	**B**	**C**	**D**
Agricultural Rural	Agricultural (Various)	6-77	35-86	70-91	80-94
Meadow	Meadow (Various)	30-68	58-79	71-86	78-92
Forest	Forest (Various)	25-36	55-60	70-73	77-79
Federal Airport Lands	Meadow (Various)	30-68	58-79	71-86	78-92
Urban	Residential 1/8-1 Acre Lots	51-77	68-85	79-90	84-92
Urban Open Space	Urban Open Space (Various)	39-49	61-69	74-79	80-84
Wetlands	Wetland (Forested and Non-Forested)	45-49	66-69	77-79	83-84

Table 4.3. Lookup table identifying appropriate ranges of CN's based on all SCS land use classes potentially associated with land use classes identified by the TRCA.

Two sources are used to define the maximum accepted values for initial abstraction in pervious and impervious areas. Clarifica (2002) defines the maximum initial abstraction values for pervious and impervious areas within the Duffins Creek watershed to be 8 mm and 0.8 mm, respectively while Viessman et al., (1974) defined maximum depression storage values for pervious and impervious areas to be 0.403 inches (10.23 mm) and 0.093 inches (2.36 mm), respectively. Further Viessman et al (1974) identify maximum interception losses for pervious areas to be approximately 0.03 inches (0.762 mm). It was assumed that the values for depression storage as well as interception loss presented by Viessman et al (1974) and Clarifica (2002) were the best physical estimates of a range of initial abstraction losses for impervious and pervious areas within the study site. In order to define a reasonable maximum value, the average of values identified by Clarifica and Viessman was calculated and the average of the two values established the upper range of initial abstraction for pervious areas equal to 9.9 mm and impervious areas equal to 1.6 mm.

4.1.3 Areas of the Model likely to Contribute to Uncertainty

While all sources of data contribute some uncertainty to a model's performance, understanding how this uncertainty arises within the data can significantly aid in the identification of those factors which will likely impact model performance the most. The methodology behind this process deals primarily with an understanding of the lineage behind the data of concern.

Geographic data in this study is extracted from secondary data sources identifying soil and land-use within a selected study site and then analyzed within a GIS environment. The results of this analysis identify various model parameters that are used in conjunction with a mass-balance hydrological model. Because the chosen model structure is based on the SCS-Curve Number approach, the Curve Numbers identified in conjunction with the GIS analysis is based on the local soil and land-use conditions, and therefore, is a reflection of the quality of soil and land use data analyzed and visualized within the GIS environment.

The HSG classification of soils into discrete classes has been under criticism within the literature since Musgrave (1955), first proposed the classification in the U.S. Yearbook of Agriculture. Further, assignment of local soil variation to hydrological groups specified as part of the HSG for use with the SCS method has been based largely on published criteria that are subjectively interpreted and applied loosely by soil scientists (Neilson et al, 2002). Further, the abstraction of land uses to various cover types also introduces error through the generalizations associated with classification and the ambiguity with which local authorities identify their land uses. Consequently, there is a high probability that an individual who is applying procedures outlined by the SCS rainfall-runoff method and not intimately familiar with the local landscape characteristics of a study site, will not likely be able to differentiate between the SCS classes and local land classification schemes. This results in a wide range of CN's that can be identified for any one soil-land use combination.

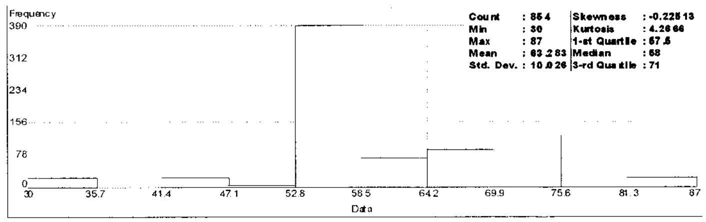
Digital models of soil and land use are based on a number of data sources such as field surveying, air photographs and preexisting soil maps. In this respect, issues of uncertainty may also be exacerbated through positional and classification inaccuracies as a result of problems with data capture. Further it is quite feasible that the mixed lineage of the land use and soil data has created data sets that encompass contrasting levels of uncertainty. The identification of Curve Numbers for the study site is dependent on the results of GIS overlay procedures as previously described in section 4.1.2. The result of the *Overlay* procedure is a composite map reflecting a continuous but varying theme of polygons reflecting the combination of soil type and land use (see Figure 4.8). Given the mixed lineage of source data for the soil and land-use themes, boundaries identifying classes of each theme will likely be subject to positional uncertainty possible the result of poor field techniques, air photo interpretation or the inherent vagueness associated with the class boundaries of natural physiographic phenomena.

While the ranges of initial abstraction values for the case study have been derived using published physical estimates, these values possess a large degree of inherent generalization in that they are do not take into consideration any local field measurements, and are based on widely spread geographical surveys the validity of which to the current test site could easily be questioned. In this respect, it is likely that the range of acceptable values for impervious and pervious initial abstraction is the source of considerable uncertainty in the model.

The model inputs identified in Section 4.1.2 are also subject to a large degree of both spatial and attribute uncertainty. Attribute error for the precipitation will likely be introduced through the degree with which the rain gauge used can accurately measure volume of precipitation over a particular unit of time. Given that data available is also in average daily values, uncertainty is also likely to be introduced as; precipitation is known to be much more temporally variable then can be accurately depicted in a twenty-four hour time period (see Section 1.3.2). Estimating potential evaporation from a series of pan evaporation networks is also a known source of uncertainty. Further, pan evaporation was reconstructed from previously recorded time periods using a curve fitting technique as discussed in Section 4.1.2, representing a considerable source of uncertainty in the data quality. Error in model input data is also introduced as a result of the distance the precipitation and the evaporation monitoring stations are from the study site.

While the range of uncertainties in loosely coupled modelling procedures is highly diverse, it is relatively improbable that all sources of uncertainty within a SCS-Curve Number and GIS-based hydrological model can be identified. This would require a significant quantification of historical error propagation as well as the computation of numerous multidimensional joint distributions for which the estimation of numerous first and second order moments could be realized. In this respect, a series of measurable or at least justifiable sources of error are identified which, based on discussions throughout the literature review, undoubtedly contribute to uncertainty in the overall model.

Histogram
Transformation: None



Data Source:
Layer: Soil-Landuse Overlay
Attribute: CN

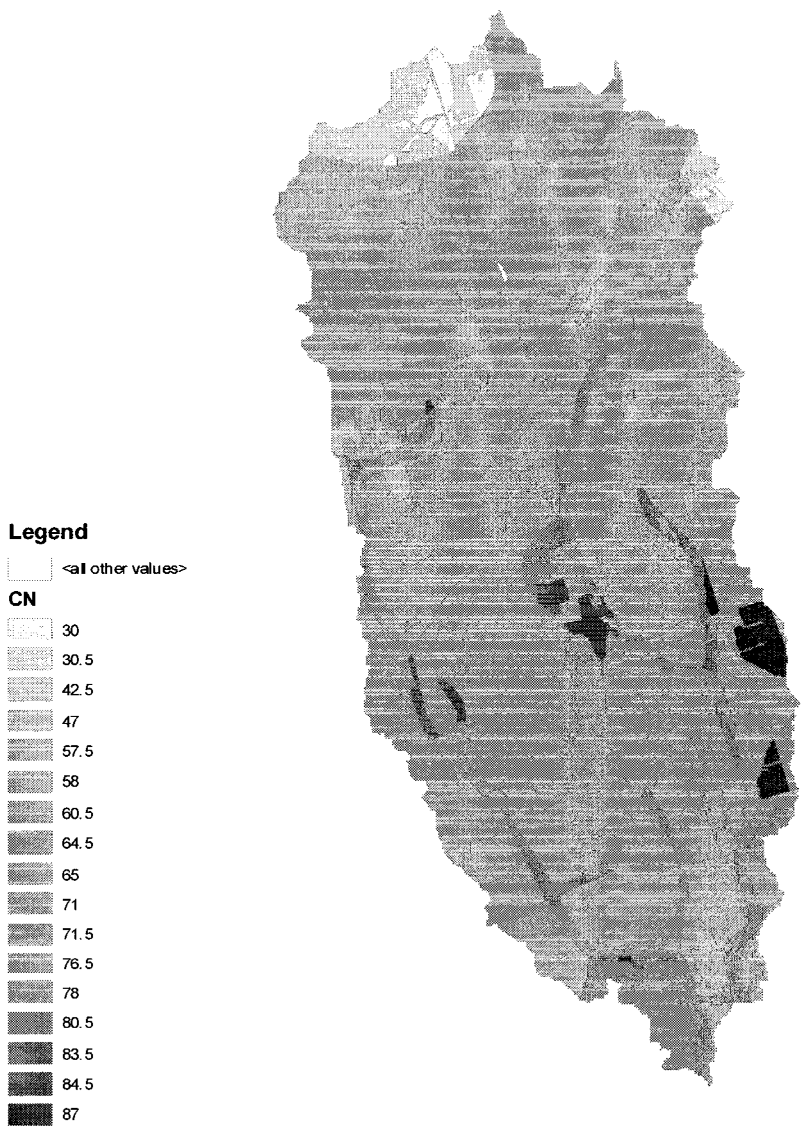


Figure 4.8 Map of curve numbers for each soil and land use combination within the study area. As can be seen from the attached frequency table, dominate CN's lie within a range of 52 – 58. Values in this range are consistent with agricultural land having reasonably good drainage.

4.1.4 Error Model Development

The definition of the distributions used to characterize uncertainty is, in many ways, the most important single part of sampling-based uncertainty and sensitivity analysis (Saltelli et al, 2001.) In the case study, information on the lineage of the data described in subsection 4.1.3, is used in conjunction with a series of methods to identify appropriate and justifiable error models for use with the uncertainty and sensitivity analysis.

Establishing the error models associated with a particular model factor can involve three basic approaches.

- *a priori knowledge*
- *expert judgment or rule-of- thumb*
- *estimation procedures.*

In the first case, the first and second order moments of an error distribution as well as the appropriate probability distribution function (pdf), (Guassian, Uniform etc) may be identifiable through published estimates. The validity of second hand information however, is dependent on the source of data, its lineage, and may be subject to scrutiny dependent upon the conditions or applications used to determine relevant statistical information.

In the second case, expert judgment or rule-of-thumb may be used to constitute a formalized estimate of a suitable error model. The application of a rule-of-thumb approach can be subject to scrutiny however; this approach is often necessary in cases where in depth knowledge of the ranges and distributions of model factors is absent, or when UA/SA is considered to be an exploratory exercise. Further, as long as ranges are not considered unreasonable, the establishment of error models through application of expert judgment can lead to considerable insights into the behavior of the system in question (Saltelli et al 2001.)

When expert judgment is applied, the use of the assumption of normality is often used when first and second order moments characterizing a factors uncertainty are known, but further information on its statistical distribution is not. This assumption is most often justifiable based on the conditions of the central limit theorem (CLT) where, randomly influenced errors of significant sample size subscribes to a Gaussian or Normal pdf (Heuvelink, 1998). In cases where first and second order moments describing uncertainty in a model factor are not known, the use of a uniform or log normal distribution may be assumed and physical plausibility arguments may be used to establish appropriate ranges (Beven and Binley 1992; Saltelli et al, 2001). While care must be taken when using expert judgment to assign an appropriate distribution for a model factor, some liberties may be taken as; the results of UA/SA are typically more sensitive to the range of uncertainty identified as opposed to the particular distribution (Crosetto, 2001).

In the final case, quantitative statistical procedures can be used to directly estimate the uncertainty for a particular model factor. This procedure is the most rigorous of the three and should be used where constraints permit. While a variety of methods exist for quantitatively deriving error estimates, the simples of procedures would involve the comparison of an error prone data set (the experiment), with that of a controlled data set (the control) for which all known sources of error contributing to uncertainty can be removed have been removed. Once a significant number of realizations of experimental and control data have been iterated so as to preserve statistical significance, the difference between each realization of controlled and experimental data is the error leading to uncertainty in the experiment. As an example, lets assume a scenario in which two tipping bucket rain gauges are set up so as to measure precipitation from a single event. It can be assumed that without proper calibration, tipping buckets are sensitive to error in the volume of precipitation measured as a result of between tip wetting losses as small amounts of water adhere to the surface of the bucket as well as between tip evaporation. Given the assumption that both buckets are subject to the same errors prior to calibration, if one gauge is calibrated and the other left uncalibrated, then the difference in volume of precipitation measured for each event captured by the two gauges should be representative of the error in the gauge due to the previously described losses. This error than can be described statistically through calculation of its mean and standard deviation.

Initial Abstraction

As previously mentioned, in cases where an error model cannot be quantified and a range of values is present, often a uniform distribution is assumed. Thus, error models for the range of values corresponding to initial abstraction for pervious and impervious areas were identified as approximating a uniform distribution having a range of 0-9.15 and 0-1.6 mm, respectively.

Precipitation

It is reasonable to assume that measurement error of rain gauges combined with the lack of any gauges within the test site and the overall distance of the rain gauge from its geographical centre, will contribute to the overall error in the rainfall time series. Unfortunately, estimating the effect of distance and lack of appropriate gauge densities with respect to the study site is very difficult. Thus, an error model was necessarily assumed which attempts to capture error caused by measurement inaccuracies of the rain gauge. This is a reasonable identifiable source of error and one that has a generally accepted maximum error margin of +/- 10% (Lei 1996). This 10% error can be related to a miscalibration of the rain gauge, wetting losses associated with a tipping bucket apparatus or between-tip evaporation of water in the tipping bucket. This percentage estimate however does not provide us with a first and second order moment of rainfall error and thus, an assumed value is used which approximates the error in measurement equal to a standard deviation of plus or minus 0.215 mm/day from daily value.

Due to logistical reasons, some of the apparent sources of uncertainty identified in the previous section was not incorporated into the determination of the error model for precipitation. Some of these concerns included the distance of the rain gauge from the study site as well as the positional accuracy of rain gauges. Given more appropriate data, the uncertainty due to any or all of these concerns could be identified, or at least approximated using the following techniques.

The case of distance from the study site can be examined if a number of gauges in a particular area that is likely to be exposed to the same synoptic rain events given an identifiable lag period. In this case a rainfall time series for a single event recorded at a series of gauges could be identified by equating peak rainfall within the acceptable lag period. The distance between each gauge and its nearest neighbour could then be calculated and the difference values for each paired gauge could then be divided by the respective distance between gauges. The average value could then be calculated for each paired gauge for a single event. A set of data for a series of events could then examined and analyzed for a mean and standard deviation. This would then represent the mean and standard deviation of rainfall uncertainty per average unit distance between gauges, and these values in conjunction with a suitable distribution could be used to produce a set of error prone rainfall values based on distance from an ideal location.

Of course this simplifies the relationship somewhat as it assumes a suitable lag period could be determined so as to equate rainfall peaks recorded at each gauge. This also ignores complications of elevation and orientation of one gauge with respect to another as rainfall is known to be directionally variant in space as well as variant with respect to elevation. Some of these factors could however given enough investigation be accounted for using a weighting function.

In the case of the positional accuracies of the rain gauge, it is suggested that this will only be of concern when a number of gauges is being used to identify an average rainfall for an area such as is performed through the use of a Thiessen polygon weighted average technique. The Thiessen polygon technique is performed by first tessellating each point representing a rain gauge into an area of influence for that point, essentially delineating all area closest to a single point than to all other points. Once this is performed then the proportion of total area belonging to each gauge is used to weight the recorded rainfall value for each time step. Finally the average rainfall value is calculated by summing across all values for each time step calculated for each rain gauge and its corresponding area of influence.

It is hypothesised that in cases where there may be considerable positional error associated with the location of each rain gauge that the uncertainty introduced into rainfall by uncertainty in the polygon tessellation will resulting from uncertainty in point locations.

Assuming the positional error associated with each gauge is calculated based on field measurements than a number of error prone realizations of the location of each gauge can be simulated. For each simulation the corresponding Thiessen polygons can also be realized. Assuming a number of these realizations are calculated, the proportional areas for each polygon may well change slightly based on the positional uncertainty of each gauge. The uncertainty this contributes to rainfall can then be simulated using a Monte Carlo approach. Once a number of area proportions are recorded for each error prone Thiessen polygon realization, each realization can be associated with a number in chronological order from 1 to N realizations. The MC simulation could then choose according to a uniform distribution about the range of numbers (1 to N) a set of error prone polygon proportions according to a user defined number of trials. In each case the proportions could be used in conjunction with measures or simulated rainfall at each respective gauge to come up with an error prone realization of a Thiessen weighted average rainfall value. The number of realizations is then equal to the number of steps in a rainfall time series and could then be used as an error prone rainfall input to a rainfall-runoff model. Thiessen based modelling following this type of approach could also be used in assessing the uncertainty of any point-based observations related to the model such as evaporation estimates.

Evaporation

Again with the evaporation measurements it is logical to assume that error is introduced by distance from the site and microclimate variation at the location of the evaporation pan as well as the deviation from actual values by relying on values extracted from climate reconstruction. Like the rainfall, however, estimating an error model based on these contributions is very difficult. Literature, however, suggests that pan evaporation values when used to predict potential evaporation for a region are typically subject to a maximum error of 10 % (Chin et al, 1995). Again, due to the lack of published first and second order moments describing predictive error, the corresponding error model for evaporation was assumed to have zero mean and plus or minus one standard deviation of 0.003 mm/ day.

Due to logistical limitations, the uncertainty associated with the curve fitting technique was not hypothesized as part of the evaporation error model. This could in theory be done however if a number of days during each month were sampled for evaporation using a pan evaporation station within the test site. Once a suitable number of days were sampled then the difference between the sampled values and the extracted values could be calculated. The mean and standard deviation could then be derived from the differences and used in conjunction with a suitable distribution to describe and error model for that particular source of uncertainty.

Curve Number

In order to identify a probability distribution function for the SCS-Curve Number, two random processes were considered. The first random process concerned the association of a particular Curve Number to a given soil a polygon entity within the GIS. Because of the ambiguous classification of specific land use classes in the SCS-method, a given user could in theory associate any number in a range of Curve Numbers that would loosely correspond to a particular local land use classification scheme. Secondly, an additional source of error identified relates to the spatial properties of the soil classes. Soil classification is vague and can be misleading as classes are not mutually exclusive and spatial boundaries of classes as illustrated in soil maps are often subject to considerable positional error and misinterpretation. In order to determine the effect of these sources of error on Curve Number assignment to each polygon within the GIS the following procedure was utilised. However, if the position of soil boundaries is uncertain, then the results of the overlay procedure used in Section 4.1.2 will also produce uncertain results with respect to the soil classes identified in conjunction with each land use polygon. In order to identify a CN model that reflected this uncertainty, it was necessary to produce a GIS record set that reflected the uncertainty of soil types that corresponded to each land use type as well as the range of CN's used to define each land use class given a particular soil land use combination.

The range of CN's identified in the lookup table can be used to model uncertainty resulting from ambiguity in land use classification. In the lookup table, TRCA land use classes are associated with a range of potential CN's such that an appropriate CN could be derived for model calibration. As previously discussed, the median CN value of the corresponding range was assumed for each soil and land use combination. This was followed by the application of a weighed averaging of each polygon and its respective CN value to come up with a single representative parameter estimate. However, given the uncertainty in ambiguous land use classification, it is assumed that each value in the range of CN's identified for each TRCA land use could potentially be associated with each land use and soil combination. Given this assumption, the selection of an appropriate CN for each record could be associated using a randomized process whereby for each soil and land use combination, a CN value could be randomly selected from the range of values in the associated lookup table. The weighted average CN value could then be calculated for the study site and the process could be repeated a user defined number of times as part of a MC simulation. The data set produced could then be analyzed statistically to identify the mean and variance of the CN values. This in turn would describe the error model for the CN's based on ambiguity of land use classification, but would ignore any uncertainty in soil boundary determination.

In order to assess the effect of uncertainty in soil boundaries it is assumed that if the location of soil boundaries in the soils layer are uncertain, then the area of each polygon used to delineate each of the four HSG soil classes will also be uncertain. In this respect, the classes of soil associated with each land use type in the GIS overlay procedure described will also be uncertain, as the true area associated with each HSG class is not known. One way to approximate this uncertainty would be to identify a region of uncertainty based on positional inaccuracies in soil boundaries based on comparing positions of boundaries in the GIS with those measured in the field. As discussed in Chapter 3, the boundaries of polygons are often identified using a series of points. Therefore, the positional error in point measurements can be extended to polygon boundaries and consequently be used to infer uncertainty in areas. However, because the measurement of the true locations of soil boundaries would require more time and resources that the current study allowed for, the positional accuracy of boundaries within the data set is generalized through the assessment of positional error of a more easily verifiable data set. For this purpose, verification of positional error within the study data was measured based on the locations of a series of street intersections corresponding with the transportation network theme provided by the TRCA. It is necessarily assumed in this case that any error detected in the position of points in the streets layer is representative of point error in other themes. Justification of this lies in the assumption that sources of error leading to positional inaccuracies in the digitizing of the streets layer will also have a similar effect on the digitizing of information from other sources.

The location of 30 street intersections corresponding to the test site was recorded in the field with a Trimble hand held GPS receiver set to record in UTM coordinates corresponding to a NAD 83 projection. Typically positional error is determined for a given point or points through comparison to a set of points whose locations are known to at least three times the accuracy (Zhang et al., 2002). Given that at least some of the primary sources of data used to create the GIS files are derived from soil and topographic maps, with various projections and unknown lineage, the GPS is assumed to be at least three times as accurate in measuring the locations of points as the digital GIS data. In order to increase accuracy, the GPS receiver was also set to record each location 200 times and output the average of the value to a stored waypoint.

As can be seen from Figure 4.9, the measured points are displaced by a considerable amount. Further analysis of the data using the mean displacement statistic described in Equation 3.4, indicated that the layers were biased by what seems to be a

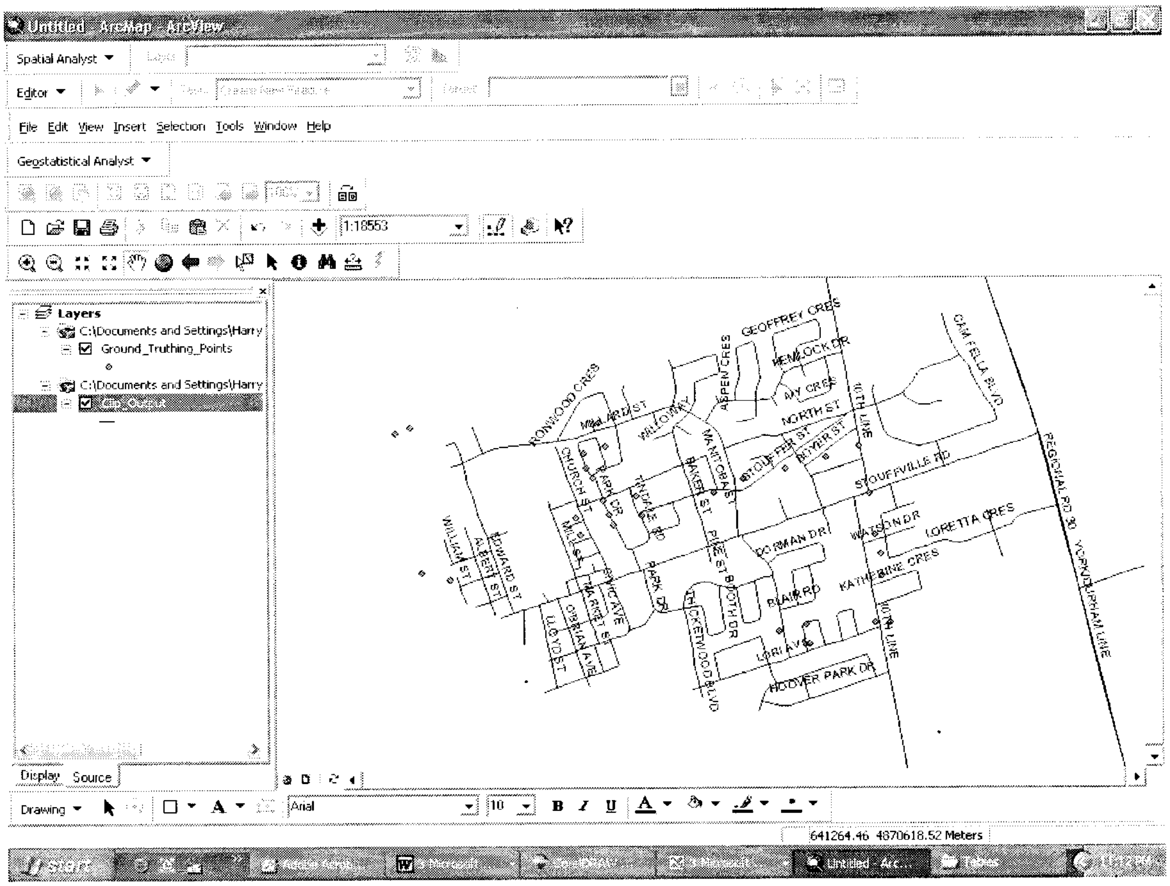


Figure 4.9 Co-display of field sampled points using a hand held GPS and the corresponding streets file layer for the Reesor and Stouffville creek sub-watersheds. Results of a mean displacement statistic indicate a potential systematic error of 229 m.

systematic error of 229.7 m (see Appendix VI). While easily correctable, systematic positional error in this case is assumed to proliferate through other data layers and is used as a standard for positional error in boundaries within the GIS data. This does not suggest that error of this type should could not be easily corrected, but that this type of an error when not identified using ground-truthing techniques could and does go unrecognized. This may especially be true in the case of boundary determination for objects such as soil classes whose boundaries are notoriously vague and difficult to validate using field techniques. Based upon this assumption, it is highly probably that error of this type could go unnoticed and therefore be allowed to propagate through various types of application and analysis.

In order to model the uncertainty in soil boundaries using error identified for the streets layer, the *Buffer* function was applied using the GIS to approximate a maximum region of uncertainty corresponding to 229 m for each soil class in the soils layer. This had the effect of producing a number of overlapping regions of uncertainty where more than one soil type could theoretically occur. In order to assess the regions where a particular land use class overlaps with uncertain soil regions, an *Overlay* procedure was used similar to that described in Section 4.1.1. Because regions of uncertainty in the soils layer had been modeled using the *Buffer* procedure, the results of the overlay produced records where each land use class co-occurred with the set of possible soil classes (see Figure 4.10), such that up to four soil classes could be identified as corresponding to a single land use for each polygon.

In order to model the effects of the uncertain soil regions in conjunction with the ambiguous land use classes, it was assumed that each soil type co-occurring with a land use class for a single polygon had an equally likely chance of occurring in that particular region. While this assumption could have been theoretically improved given further information on the probability of a particular soil type occurring in a given area, the absence of this information necessitated a more simplified approach.

The error model for the curve number based on uncertainties in land classification and soil boundaries was then evaluated using a MC technique written in VBA (see Appendix V). This first involved the exporting of the polygon records produced using the *Overlay* procedure to Microsoft Excel in conjunction with the CN lookup table discussed previously. The record in the record set produced using the *Overlay* procedure identifies each land use class and a set of soil classes that potentially occurred within each uncertainty region. For every record, the MC program identifies each soil class associated with a particular land use type and chooses one at random. Once this is done, the program then identifies the land use class for the same record and using that class in conjunction with the randomly identified soil type, references an appropriate range of CN's in the corresponding lookup table. One an appropriate range is identified; the program randomly

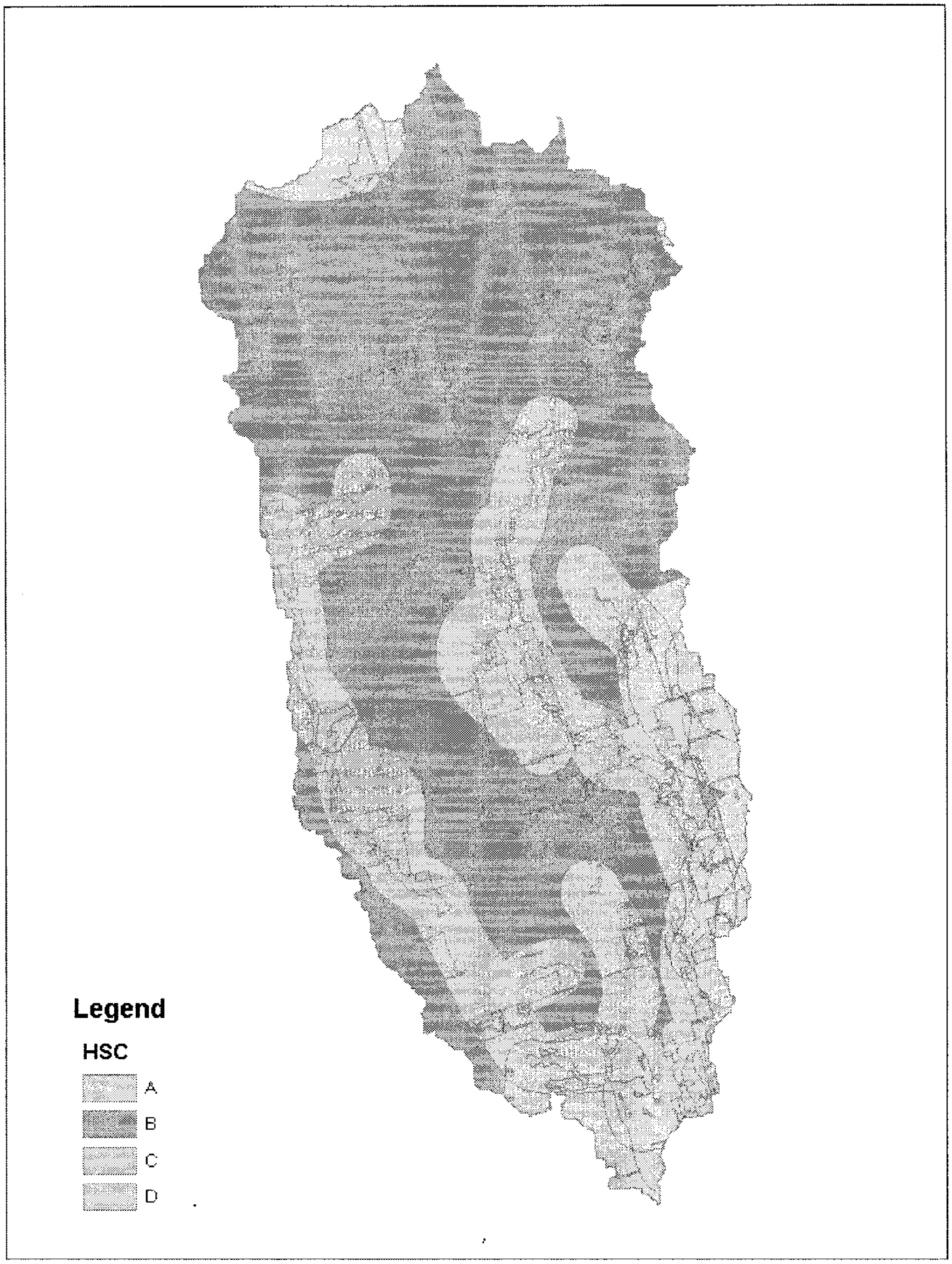


Figure 4.10. Result of combining the clipped and buffered soil layer with the land use layer in ArcGIS.

assigns a CN value based on the appropriate range to the corresponding record in the record set. After this has been repeated for each record, the corresponding CN's are multiplied by their corresponding proportion of total area. The weighted average random CN is then calculated by summing the values across the entire record set and this value is exported to an output file. This procedure is consistent with a single trial of the MC approach. Once this has been completed for a user-defined number of trials (15,000 in this case), the first and second order moments describing the CN error model summarized resulting in a mean CN of 66 and a standard deviation of 3.1. It was also assumed for simplicity that the set of CN's produced using the MC procedure corresponded to a normal distribution.

4.1.5 Screening Sensitivity Analysis

The application of sensitivity analysis on many model factors is often time consuming and in cases where numerous model factors are used in conjunction with complex model structures, application can be very computationally demanding. In order to minimize computation time in cases when many model factors are being evaluated, the use of what has been referred to as a screening test (Saltelli et al, 2001) is recommended. The screening process is typically used to provide an ordinal ranking of model factors such that the top few most influential factors can be evaluated using the more rigorous and quantitative techniques. In this instance, certain non-influential model factors can be identified early in the process.

The method of screening test used depends on the nature of the study however the method recommended and applied by Crosetto et al, (2000; 2001) is the method of Morris (1991). The Morris method represents what can be referred to as a variance-based OAT approach in that, model factors are varied one at a time with all other factors constant. An ordinal ranking is then established based on the relative influence each factor has on the total variance of the model output.

While straightforward in its approach, the method of Morris will not be used in the case study and is replaced by a more traditional OAT method. The primary motivation for this is that the model used in the case study is fairly simple and straightforward. As well, the number of model factors is low to begin with such that the computational efficiency is not a concern limiting the performance of more quantitative methods. Further, the method of Morris, while an OAT technique, does not allow the user to visualize the effects of perturbations on model factors by means of any graphical analysis. Traditional OAT techniques on the other hand, typically record the outcome of each perturbation. Consequently, the use of the Morris screening test results in less information that can be visualized by the user and thus, interpretation of the outcome of the procedure is limited.

In order to identify those model factors that may be most influential in contributing to model output uncertainty, as well as to interpret the range of model response to perturbations in each factor, an OAT procedure is performed as part of the case study. This test is modeled after similar applications discussed in Chapter 3, where each model factor is adjusted in the positive and negative direction by a small percent error relative to its calibrated value. For each step, the resulting model output is visualized by means of graphical analysis. This information is then used to support more complex analysis.

4.1.6 Global Uncertainty Analysis

Execution of the global uncertainty analysis is performed in the case study using a global Monte Carlo simulation. This approach has been applied in various hydrological modelling applications reviewed in Section 3.8, and forms the basis for the identification of joint uncertainty in CN estimation as described in Section 4.1. Further, the MC analysis forms the basis of similar GIS sensitivity studies using Crosetto et al.'s (2001) recommended stepwise method, and is necessary as the chosen methods for performing sensitivity analysis discussed in the next section can only be performed in conjunction with a Monte Carlo approach. The Monte Carlo approach is considered a popular method over other approaches largely as a result of the simplicity of execution. The Monte Carlo procedure is applicable irrespective of model complexity. In this respect, the method has been dubbed a "black box" where simulation can occur regardless of the nature of the dependency relationships of model factors.

The Monte Carlo method is based on the generalized expression of total model uncertainty described through the function of a simplified system equation where

$$\mathbf{Y} = f(\mathbf{X}_1, ..\mathbf{X}_2 .. \mathbf{X}_n), \tag{3.5}$$

In this case, the variance in model output **Y**, is a function of the total system response to a series of variances in model inputs, $\mathbf{X_1 \ldots X_n}$. The Monte Carlo procedure using a series of stochastic model factors executes a user defined series of random trials where, a set of values for each factor is selected for every trial according to the first and second order moments as well as the corresponding probability distribution function identified for each factor. The corresponding model output then consists of one or more values for each model output based on the values chosen for each model factor corresponding to the number of trials. The overall model uncertainty is then described through statistical analysis of the first and second order moments derived from the output series.

The Monte Carlo analysis used in conjunction with the Case Study is performed using the proprietary software SimLab. This software, not yet currently available has been developed by members of the European Union's Joint Research Center (JRC) applied statistics (APPST) research group which functions under the Institute for Protection and Security of the Citizen (IPSC). The software was given to the author through a memorandum of understanding as a means of supporting the research contained within this study. The SimLab software is useful as it is designed to work in conjunction with Microsoft Excel such that all data input and output is contained within the Excel environment. In this respect, the types of models that can be analysed using SimLab, are limited only by that which can be designed within the Excel environment.

Once appropriate error models have been identified through methods discussed in Section 4.1.4, the first and second order moments are specified for each model factor in conjunction with an appropriate probability distribution function and the set of error models for identified for each factor can then be saved as a SimLab profile for future use. Once an appropriate profile has been created, the user must choose to execute the simulation in conjunction with an external model designed in the Excel environment. The user enters the appropriate path where the Excel model is located into the SimLab environment and then specifies the number of trials for the Monte Carlo procedure. Once initiated, SimLab scans the Excel file associated with the model of interest for a table called "Inputs." Once located, a series of error prone values for each model factor is created starting in row one of the table and each factor occupying a single column in the same order as the factors are identified in the error model profile. The number of records of error prone data produced coincides with the number of user-defined trials. The user then passes the data through the model and stores the series of output values in a table called "Outputs." Once saved, the user then exits the Excel environment and SimLab continues with further analysis. The user can then perform further analysis within the SimLab environment or continue to analyse the results of the procedure within the Excel environment by performing graphing procedures or descriptive statistics.

4.1.7 Detailed Sensitivity Analysis

The extended FAST (Saltelli et al., 1999) is one of a series of sensitivity techniques known as a variance-based method as their result is to quantitatively partition the total variance identified through the global uncertainty analysis to each model factor based on its relative sensitivity. The basis of the extended FAST approach is a transformation that converts a multidimensional integral over all the uncertain model factors to a one-dimensional integral via a search curve that scans the entire parameter space. The scanning is done so that each axis of the factor space is explored with a different frequency. A Fourier decomposition is then used to obtain the fractional contribution of the individual input factors to the variance of the model prediction (Saltelli et al. 2001).

Total variance in model output is related to all model factors such that:

$$V = \sum_{i} V_i + \sum_{i<k} V_{ij} + \sum_{i<j<m} V_{ijm} + \ldots + V_{12\ldots k} \quad (3.6)$$

as:

$$V_i = V[E(Y|X_i = x_i^*)]$$
$$V_{ij} = V[E(Y|X_i = x_i^*, X_j = x_j^*)] - V[E(Y|X_i = x_i^*)] - V[E(Y|X_j = x_j^*)]$$

etc and, $\mathbf{[E(Y|X_i = x_i^*)]}$ relates to the expected value of $\mathbf{Y}$ conditional on $\mathbf{X_i}$ having a fixed value of $\mathbf{x_i^*}$, and where the operator $\mathbf{V[\cdot]}$ denotes a conditional variance. The first order sensitivity index for the factor $\mathbf{X_i}$ can then be denoted as:

$$S_i = V_i / V \tag{3.7}$$

In cases where the interactive effects of inputs with respect to one another are not negligible or at least have a measurable effect on model output uncertainty, an additional sensitivity analysis that accounts for the total relational sensitivity $\mathbf{S_{Ti}}$, or "total sensitivity" (Saltelli et al 1999; Sobol 1993) offers an analytical advantage to the extended FAST technique.

The total sensitivity is defined as (Crosetto et al., 2001):

> the sum of all indices (S_i and higher orders) where X_i is included, thus concentrating in a single term all the interactions involving X_i.

Given then a GIS coupled model with multiple factors $\mathbf{X_i}$, the sensitivity indices $\mathbf{S_i}$, $\mathbf{S_{Ti}}$ can be used to provide a measure for both the isolated impact of $\mathbf{X_i}$ on the model output $\mathbf{Y}$, as well as the overall impact of factor $\mathbf{X_i}$ through interactions between other model inputs on $\mathbf{Y}$ and, consequently, a comparison between the two effects could also be investigated. The total sensitivity index for a 3-factor model can then be described by the following

$$\begin{aligned} S_{T1} &= S_1 + S_{12} + S_{13} + S_{123} \\ S_{T2} &= S_2 + S_{12} + S_{23} + S_{123} \\ S_{T3} &= S_3 + S_{13} + S_{23} + S_{123} \end{aligned} \tag{3.8}$$

Where, $\mathbf{S_{12} = V_{12}/V, S_{123} = V_{123}/V} \ldots$

In this case, the total model uncertainty can be related to the influence of either a single factor or the result of that factor and its n^{th} order interactions with all other factors.

The extended FAST technique represents a particularly suitable approach for application to the case study of this research. The technique is based on the results of a Monte Carlo analysis approach and in this respect is suitable, as the Monte Carlo technique has been shown to be an appropriate and preferable method of model uncertainty analysis for reasons identified in Section 4.1.6. The extended FAST is a preferential method for performing sensitivity analysis in that; it is capable of calculating the n^{th} order interactions between model factors representing a significant advantage over other variance-based methods such as the ordinary FAST (Cukier, 1973). Further, this method has been shown to be more computationally efficient than other variance based methods such as those of Sobol (Saltelli et al, 2001), and requires relatively few trials of the Monte Carlo analysis to produce reliable results, approximately 100 trials for each model factor (Crosetto et al, 2001). The extended FAST approach also represents a number of advantages over local and differential based sensitivity analysis such as the first order Taylor as these methods are more complex in their application and lack straightforward and effective methods for determining higher order sensitivities among model factors. Further, problems often arise in the appropriate determination of the Taylor series used to approximate the model often having an effect on the estimates for the model expected value and variance (Saltelli et al, 2001). It is for all these reasons that the extended FAST technique has been identified as a suitable method for application to GIS based modelling in previous studies reviewed in Chapter 3 and is used to support the case study in Chapter 5.

The extended FAST technique is applied in conjunction with the uncertainty analysis discussed in Section 4.1.6, and applied to the GIS based model discussed in Section 4.1.2. Like the global uncertainty analysis, the extended FAST, is performed using the proprietary software SimLab. Prior to initiating the Monte Carlo procedure the user can specify one of a number of sensitivity procedures including the methods of Morris, (1991) as well as the normal or extended FAST procedure. At this point, the use must also specify the number of trials to be used with the uncertainty procedure. Once data from the Monte Carlo procedure is stored, SimLab then scans the relevant Excel file for a table identified as "Output", the software then uses the output series in conjunction with the input data generated as part of the Monte Carlo procedure and stored in the corresponding Excel table identified as "Inputs", to compute both the first and total order sensitivity values for each model factor the results of which can then be visualized in SimLab by way of pie graphs or data plots.

5.1 Model Calibration and Validation

The model was subject to a two-stage calibration and validation procedure. Model calibration occurred during the wet weather period from 1997-1998 and validation occurred during the wet weather period for 2000 (see Figures 5.1 – 5.3). The initial parameter estimated for the SCS-Curve Number was taken from the weighted average Curve Number calculation procedure outlined previously and corresponded to a value of 80.01 for the study area. The initial parameter calibration values for initial abstraction for pervious and impervious areas were chosen to be half the theoretical maximum for pervious (4.75) and impervious areas (0.8), respectively. Final calibration values of these parameters used to achieve the best model performance were 0.35 mm and 4 mm for initial abstraction from impervious and pervious areas. The final Curve Number value of 71.5 was also used. Other parameters based on the area of the study site as well as the fractions of impervious area directly connected were not manipulated during the calibration. The parameters used were then adjusted by small amounts in the positive and negative direction one at a time until the lowest possible percent error with respect to the measured streamflow was achieved on a seasonal and monthly basis for the calibration period.

In both the calibration years and the validation year, models predicted to within 15 percent of the measured streamflow and was equal to 14% and 0% for the calibration years and 2% for the validation year (see Table 5.1). Monthly results were more variable, and ranged from 2.4-139% and 4-56% for calibration years and 13% to 83% percent for the validation period. In all cases, more than half of all monthly values fell within 25% to 35% of the measured streamflow.

One of the goals of this thesis was to produce a model with performance reasonable enough to justify the model structure and to allow for further testing of the effects of uncertainty. Given numerous problems identified with respect to the use of a single rain gauge and in general the spatial variability of precipitation, the results achieved with the calibrated model were viewed as meeting these expectations.

Measured vs Predicted 1997

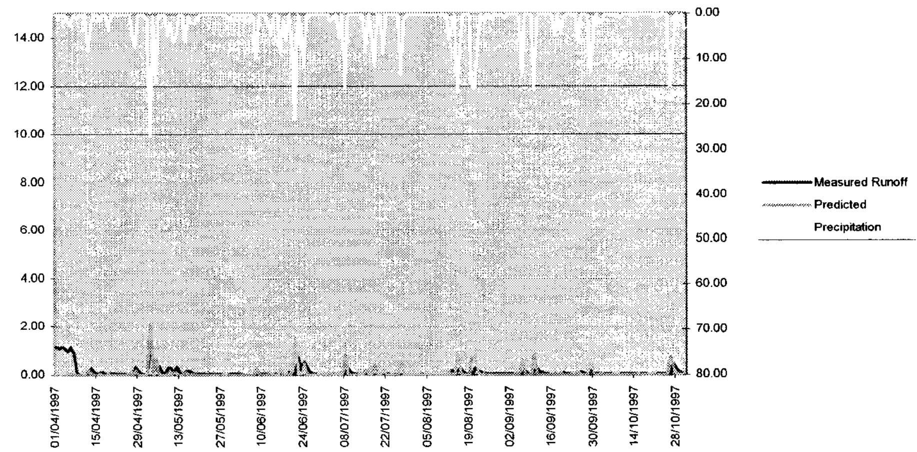


Figure 5.1. Calibration results for the 1997 period including daily average precipitation as well as daily average measured and predicted streamflow.

Measured vs Predicted 1998

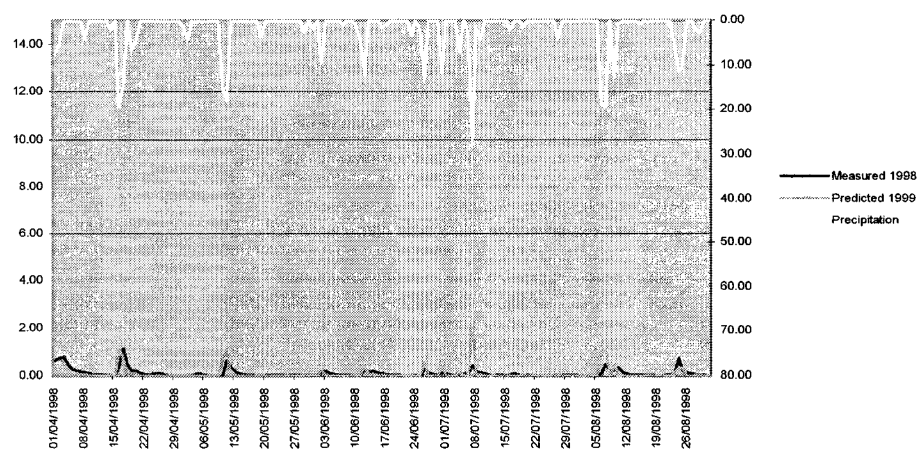


Figure 5.2. Calibration results for the 1998 period including daily average precipitation as well as daily average measured and predicted streamflow.

Measured vs Predicted 2000

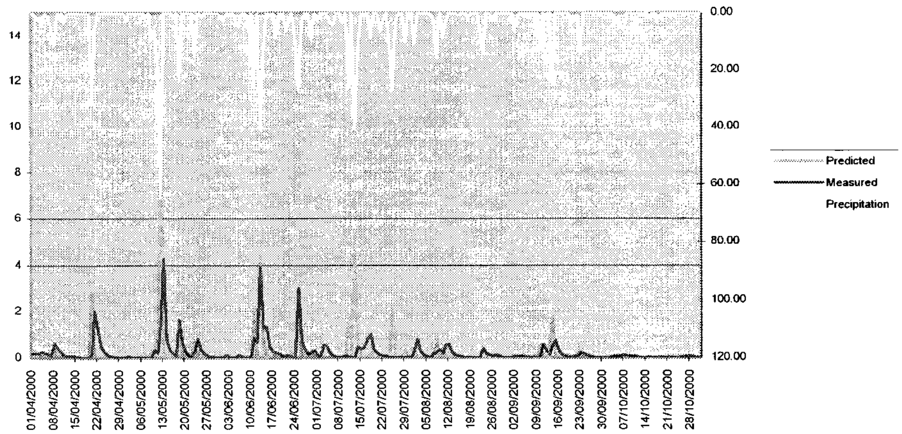


Figure 5.3. Validation results for the 2000 period including daily average precipitation as well as daily average measured and predicted streamflow.

Year	May	June	July	August	Sept.	Oct	Total
1997	-2.4	8.0	139.0	55.0	37.0	-21.0	14.0
1998	-6.0	-22.0	56.0	4.0	-21.0	-10.0	0.0
2000	13.0	-30.0	83.0	-24.0	-18.0	-68.0	-2.0

Table 5.1. Percent error between predicted and actual values for each year of model simulation.

5.2 One-At-A Time Sensitivity Analysis

Prior to performing the global uncertainty analysis and parameter-specific sensitivity analysis or extended FAST, a generalized pre-assessment of sensitivity was computed for all model parameters using an OAT analysis that introduced a systematic error in parameter estimates and graphed this against percent error introduced in modelling results.

Model parameters which directly affected model output included precipitation, evaporation, total area, the total pervious and impervious area and corresponding fractions of pervious and impervious area, the total directly and indirectly connected impervious area, the initial abstraction values, and the SCS-Curve Number. To proceed with the analysis, a VBA script (see Appendix VII) was created that systematically altered each model parameter individually by an error of one percent from increments of –100% to –30% and 30% – 100% percent of its calibrated value and, 0.5 percent increments between –30% and 30% of its calibrated value in order to evaluate the effect of smaller error imposed around the baseline calibrated values.

In some cases, however, it was not practical to evaluate systematic error up to 100 percent as this would result in parameter values outside the acceptable or possible boundaries. For example, the Curve Number value used to calibrate the model had a value of 72 and the corresponding systematic error did not exceed approximately 38% that would correspond to the maximum possible Curve Number of 100. A CN in excess of this would physically imply that a fixed depth of precipitation could be translated to a greater depth of runoff. Results of the analysis are depicted in the following tables.

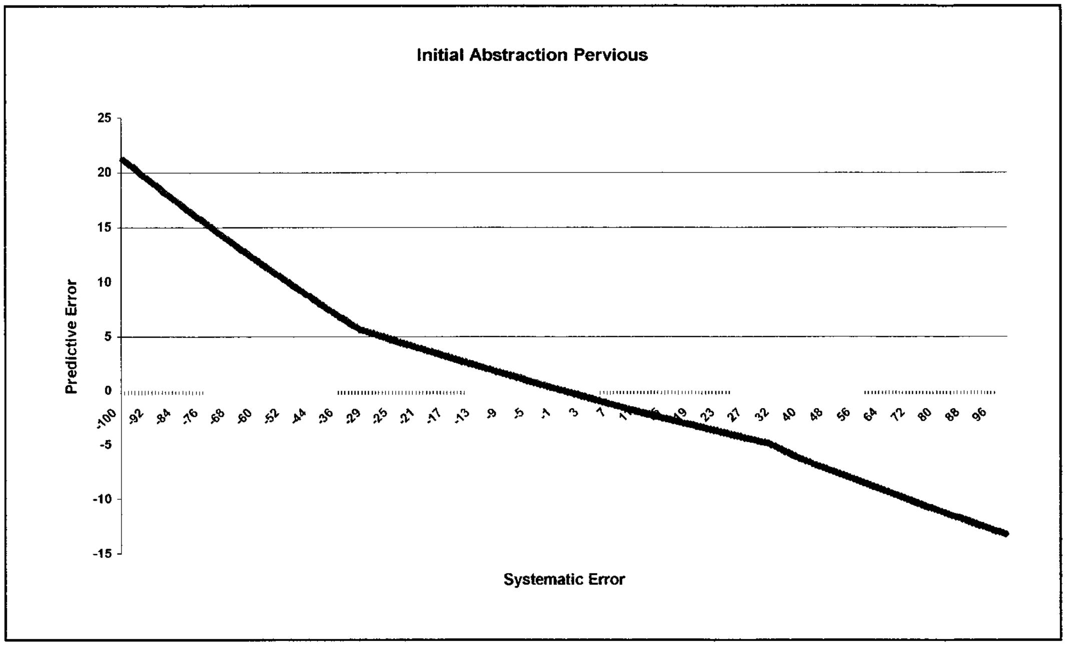


Figure 5.4. Response of the model to a systematic error of plus or minus 100% of the calibrated value for initial abstraction in pervious areas.

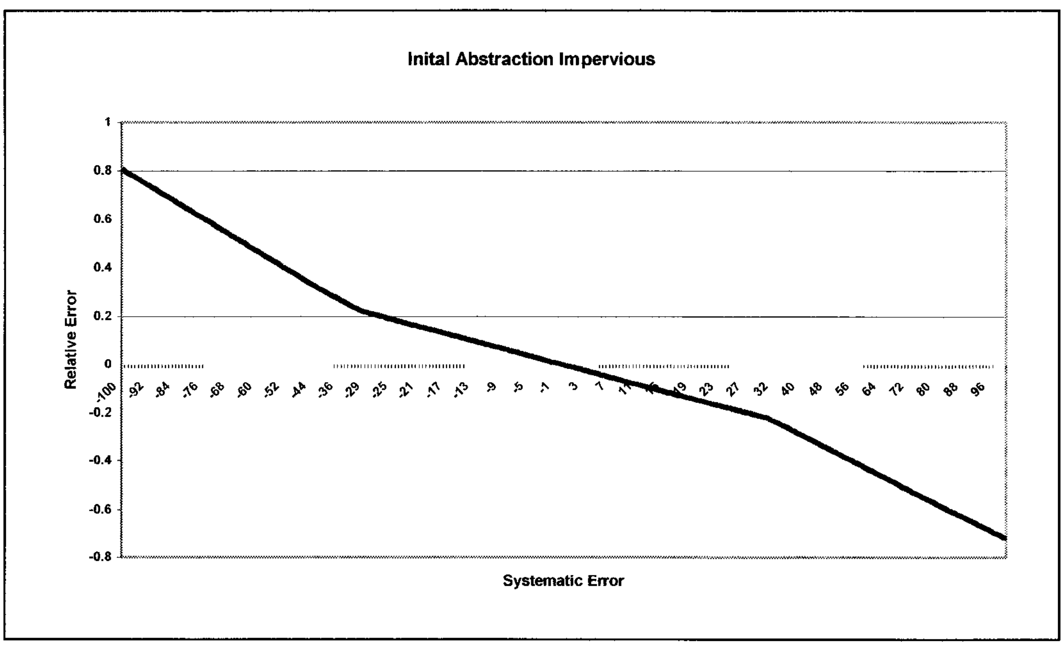


Figure 5.5. Response of the model to a systematic error of plus or minus 100% of the calibrated value for initial abstraction in impervious areas.

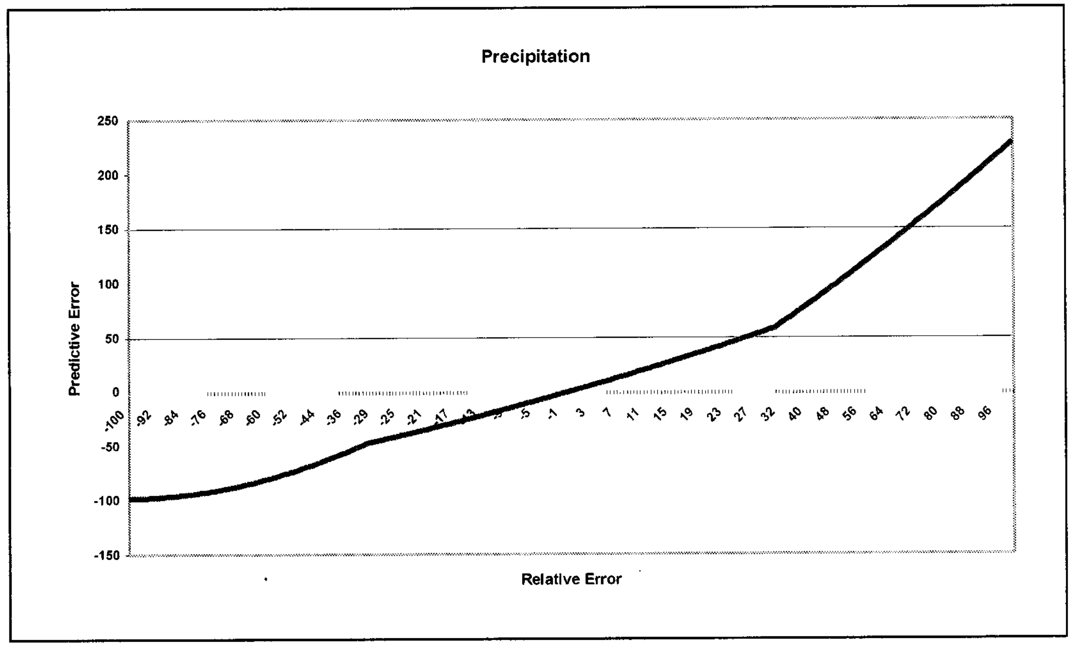
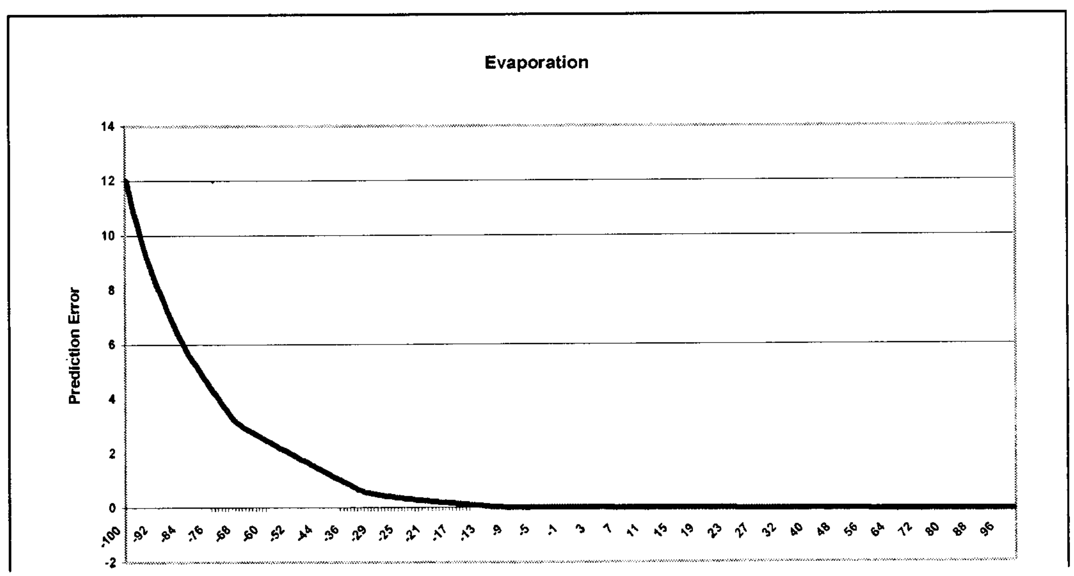


Figure 5.6. Response of the model to a systematic error of plus or minus 100% of the calibrated value for precipitation.



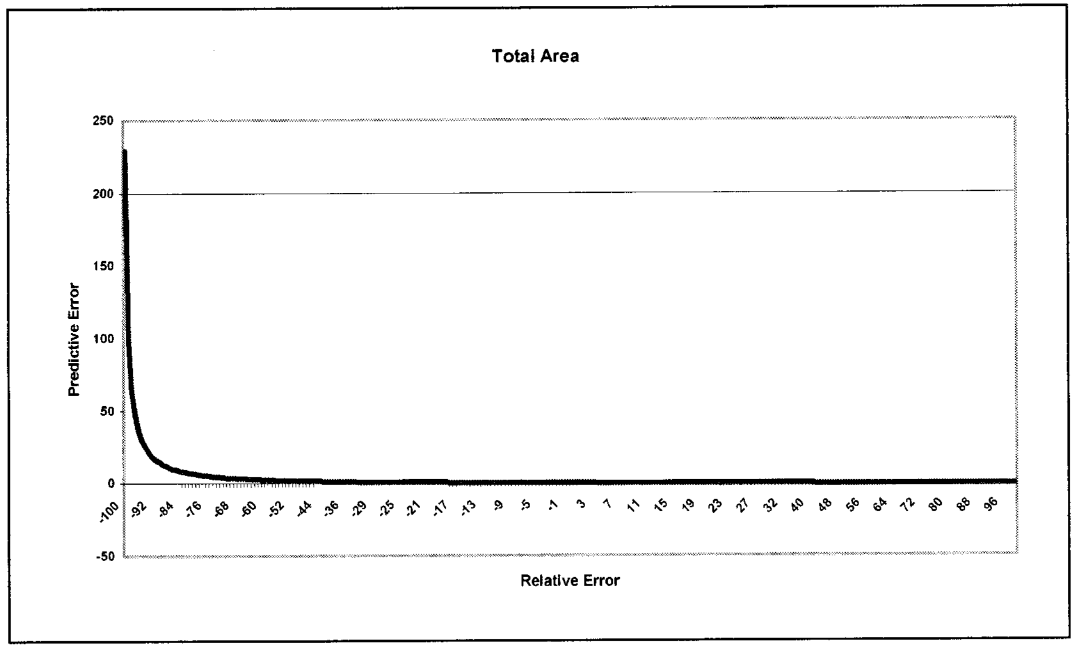
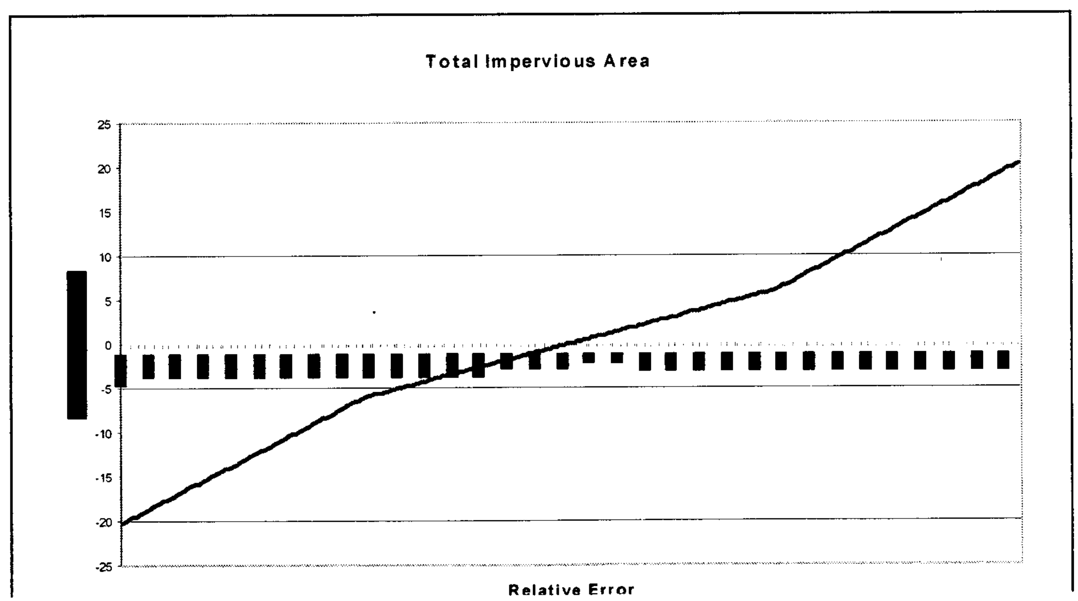


Figure 5.8. Response of the model to a systematic error of plus or minus 100% of the calibrated value for total area.



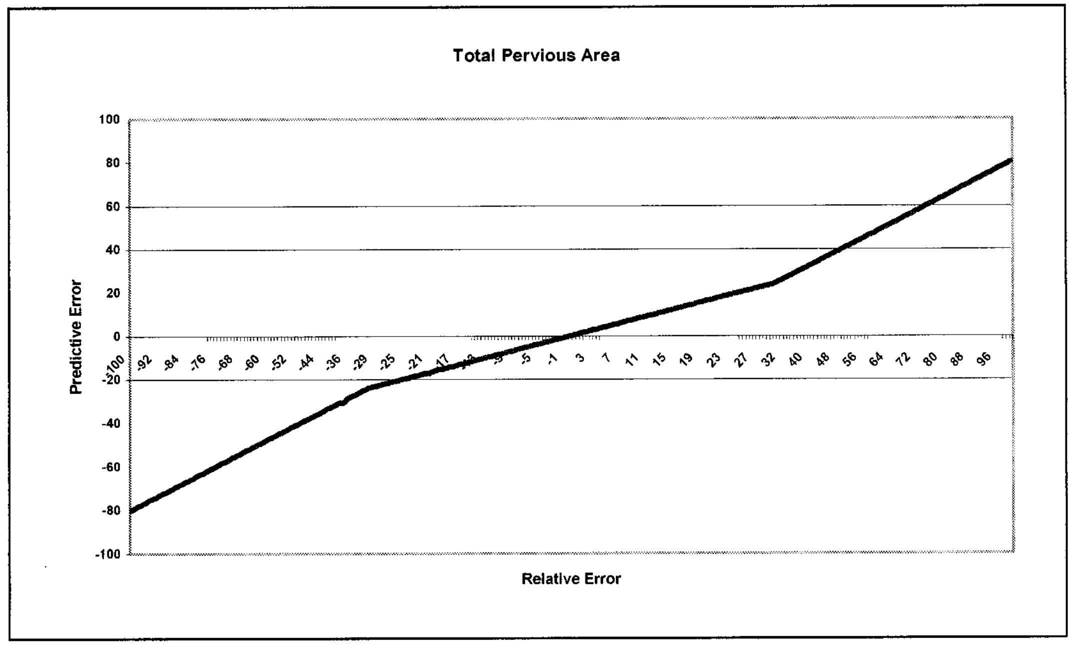
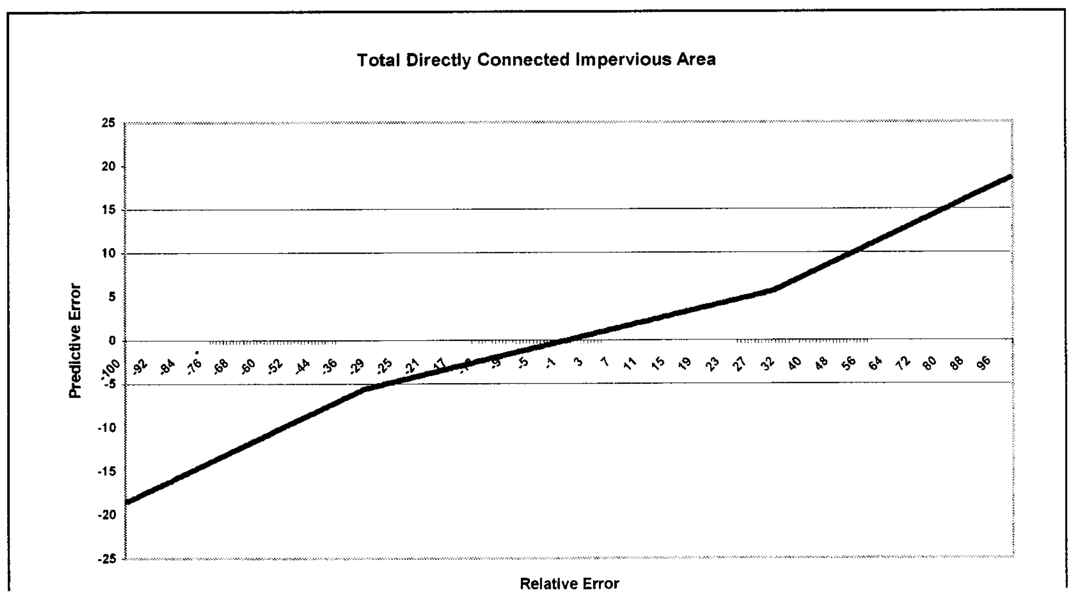


Figure 5.10. Response of the model to a systematic error of plus or minus 100% of the calibrated value for total pervious area.



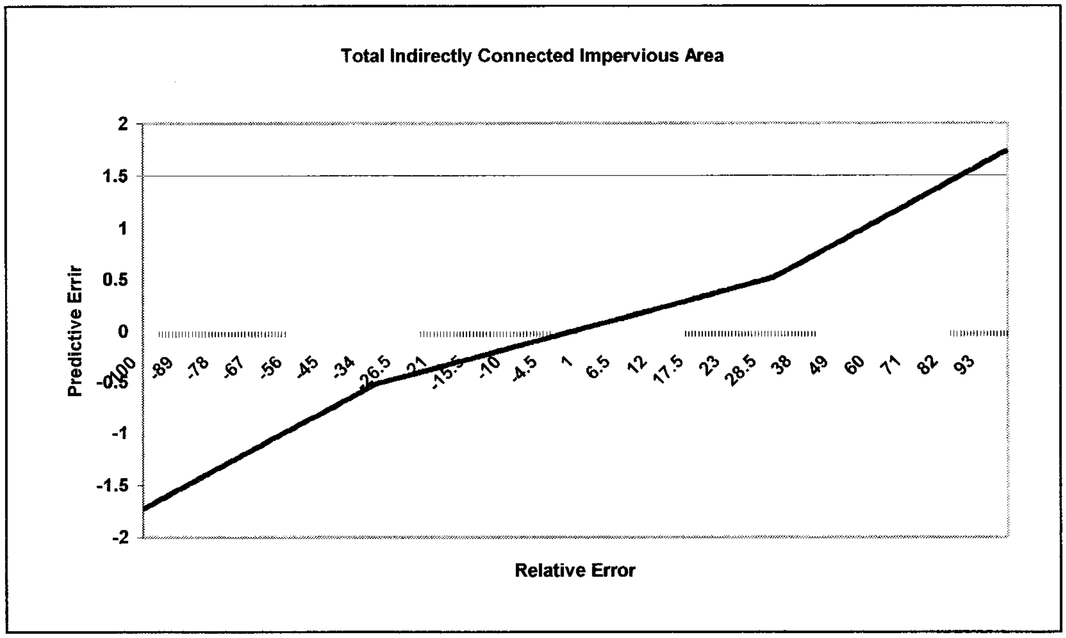


Figure 5.12. Response of the model to a systematic error of plus or minus 100% of the calibrated value for total indirectly connected impervious area.

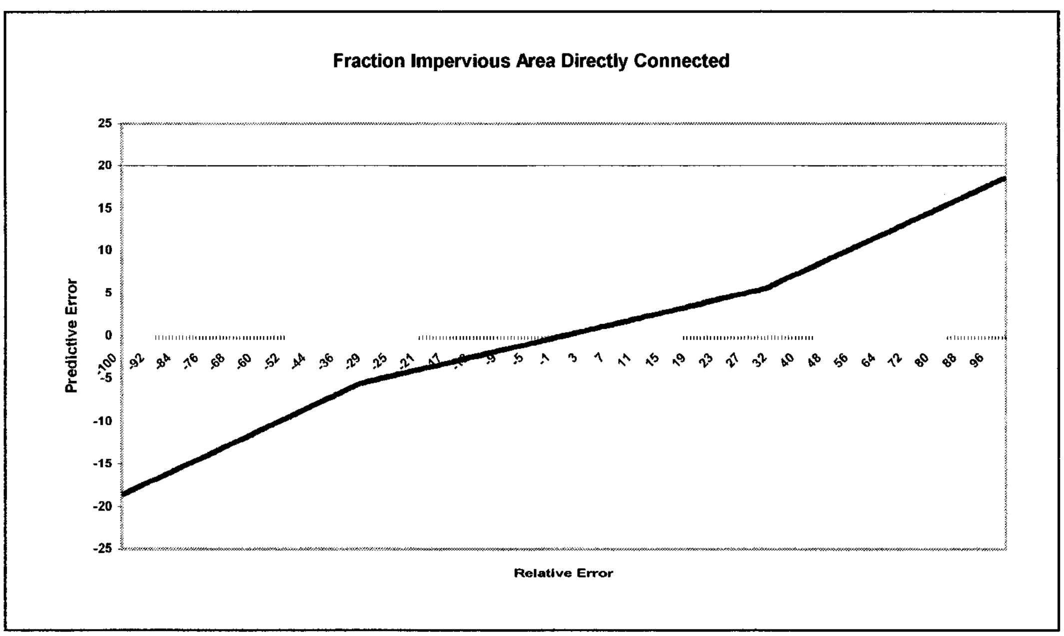


Figure 5.13. Response of the model to a systematic error of plus or minus 100% of the calibrated value for the fraction of impervious area directly connected.

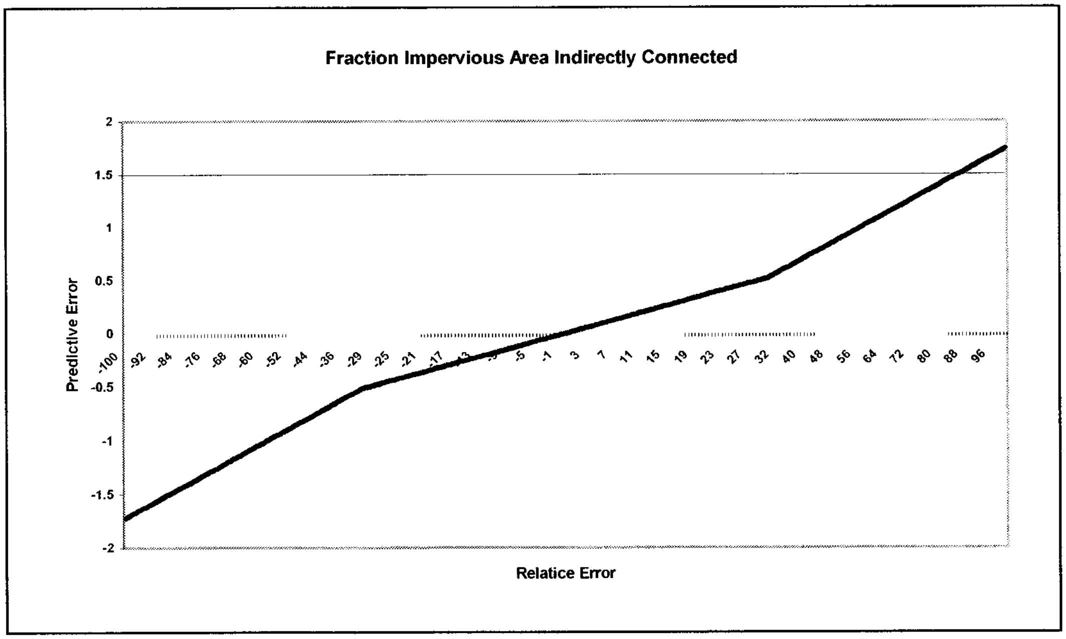


Figure 5.14. Response of the model to a systematic error of plus or minus 100% of the calibrated value for the fraction of impervious area indirectly connected.

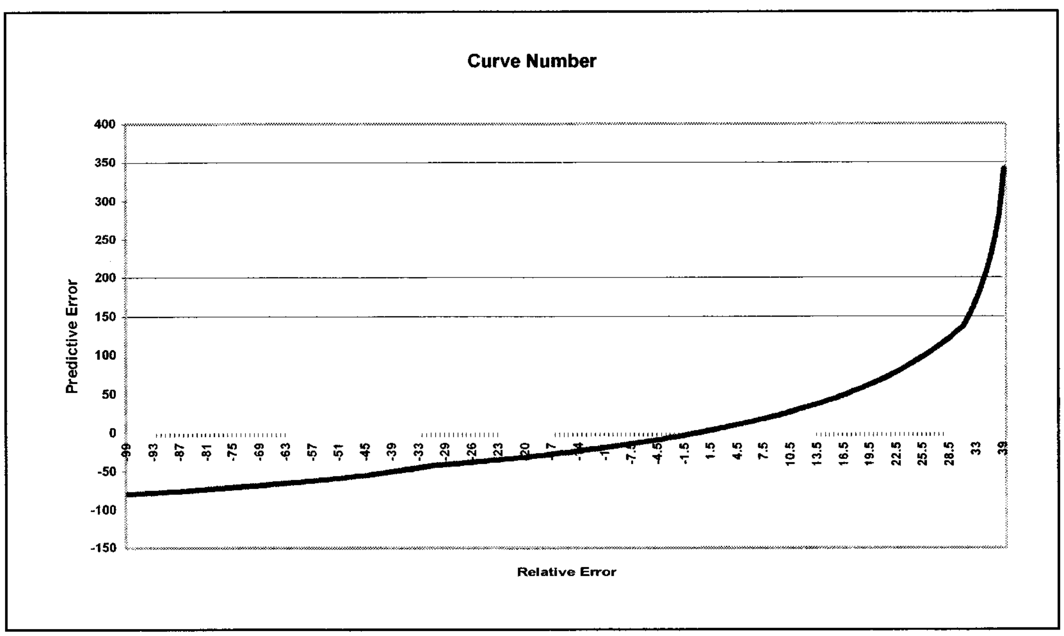


Figure 5.15. Response of the model to a systematic error of -100% to + 39% of the calibrated value for the CN.

Systematic error bias in precipitation was shown to be fairly linear within the -100% to 0% error range such that a systematic bias of –100 percent resulted in a predictive error of approximately 100 percent. When positive error was applied however, the model responded in a non-linear fashion such that for a 100 percent error bias in rainfall more than a two hundred percent error is realized. This indicates the higher sensitivity of the Curve Number technique to positive rainfall error versus negative.

Evaporation contrasted sharply with rainfall however in that error increases logarithmically in the negative direction from approximately -6 to –100 percent. Error in the positive direction had no response in the model past the –6 percent point. The zero response in the model can be explained by the invariance in model response after evaporation increased to the point where the losses imposed accounted for all input to the system.

Total area showed very little response except in the very extreme negative bias case where error sharply increased as a result of available sources of losses such as soil moisture storage and initial abstraction approach zero. Results for the other area parameters showed very similar results where total impervious and pervious areas gradually increased or decreased in predictive error with corresponding increases and decreases of systematic bias. The results indicated a sensitivity of plus or minus approximately 25 percent and 75 percent respectively for corresponding systematic error of plus and minus 100 percent. Given that impervious area only accounts for approximately 5 percent of the total area with 95 percent of the area being pervious, results support the fact that the watershed model is more sensitive to impervious areas as opposed to pervious.

The fractions of directly and indirectly connected impervious areas demonstrated a similar and non-descriptive pattern. However, the model was shown to be approximately 10 times more sensitive to the fraction of directly connected impervious area as a maximum systematic error resulted in a ten percent predictive error for the directly connected impervious portion versus only one percent for the indirectly connected impervious.

Perhaps the most significant results of the OAT analysis were for the Curve Number parameter where, for negative bias the values in model error gradually dropped off to approximately –90 percent in model error. While after approximately 2 percent positive bias, the model error rose dramatically. This indicates that in cases where high infiltration and high soil moisture storage is available the SCS method is less sensitive. However, as Curve Numbers increase towards a theoretical maximum sensitivity drastically increases such that a 37 percent bias in Curve Number (from a CN of 70 to 100) error in model predictions rose by in excess of 400 percent.

This has interesting consequences for studies in which Curve Number models are used to predict future runoff as a result of changing land uses. Typically, once a given model is calibrated, the Curve Number identified for the area is used in conjunction with an adjustment table which provides a means of predicting a CN given an increase in predicted CN based on current conditions CN and a future increased percent imperviousness scenario (see Figure 5.16).

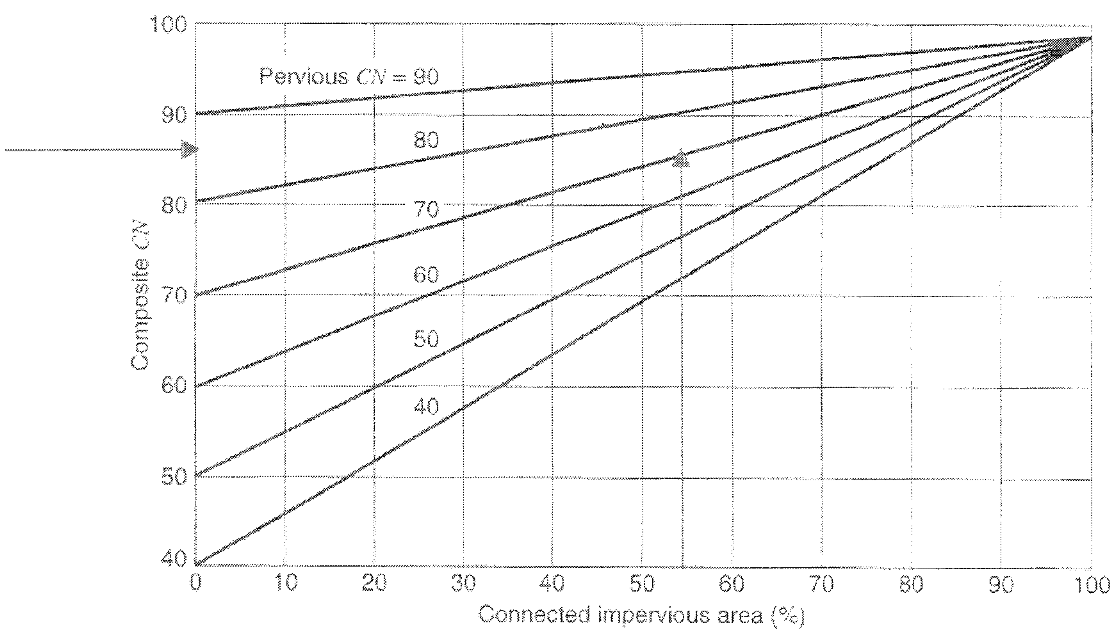


Figure 5.16. Composite CN graph for calculating an adjusted CN based on a predicted increase in total impervious area resulting from urbanization or other developments. Source: Mayes, 2001.

It was determined through quantitative and stochastic means that the error associated with the derivation of a lumped Curve Number within the study site as a result of uncertain soil boundaries as well as vagueness in land use classification had a mean of 66 and a standard deviation of 3.1. In most cases, this error would be seen as small and somewhat insignificant, however, this may not be the case if future development scenarios are concerned.

Assuming a future development scenario where percentage impervious area within the test site is expected to increase by a factor of 55%, this would produce a composite CN of approximately 86% (see Figure 5.16). Given that the CN used to predict the future development scenario had a standard deviation of plus or minus 3.1, the resulting composite CN will be subject to the same error as it is dependent on the first value for its determination.

If this new value of CN equal to approximately 88 is put into the model, results of the OAT sensitivity analysis demonstrated that a CN of 88 is approximately equal to the calibration parameter plus approximately 25.5 percent. The model in this range is has also been shown to be much more sensitive to small changes in CN (see Figure 5.17). Thus a

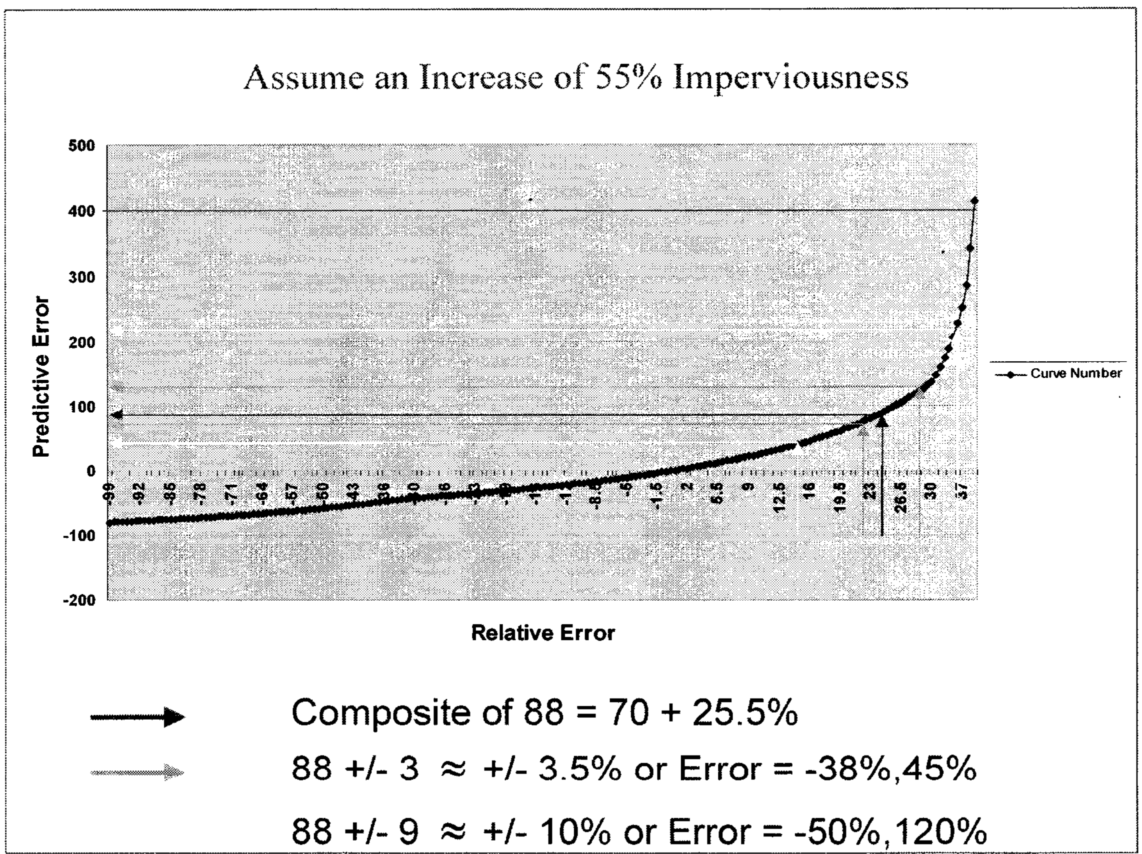


Figure 5.1.7. Assuming a standard deviation is carried forward in the calculation of a composite CN from an existing value, error in predictions could result in the order of –38% and +120% when considering three standard deviations.

standard deviation of plus or minus 3.1 at this range can produce much higher predictive errors than at the lower value of 70, up to a maximum of approximately 120 percent.

5.3 Global Uncertainty

The global Monte Carlo analysis was performed using the statistical package SimLab. The procedure was performed for 500 trials or approximately 100 trials for every model factor being evaluated in the corresponding SA as recommended by Crosetto et al., (2001). In order to perform the Monte Carlo simulation using the error model estimates for rainfall previously discussed, a specialized procedure was used to accommodate the rainfall.

In order to reduce the potential for a left skewed rainfall time series, a synthetic rainfall time series was generated using a Monte Carlo simulation and using an exponential rainfall distribution in combination with the measured mean and an approximated standard deviation from the measured wet weather time series. The resulting synthetic rainfall generation was then assumed to be representative of a wet weather time series for the study site. This was then used as input to the model in conjunction with a daily error adjustment based on the calculated error value. The error value was generated using a normal distribution with a mean of 0 and a STD of 0.215 mm per day that corresponded to the assumed measurement error associated with precipitation.

Results of the Monte Carlo indicated an extremely left skewed distribution despite attempts to standardise rainfall values (Figure 5.18). Statistical analysis of model output indicated that when all values are considered the corresponding model uncertainty has a mean of 0.65 and a STD deviation of 0.96 mm. The overall model uncertainty then can be interpreted as accurate to within plus or minus a daily average excess rainfall of 0.96 mm.

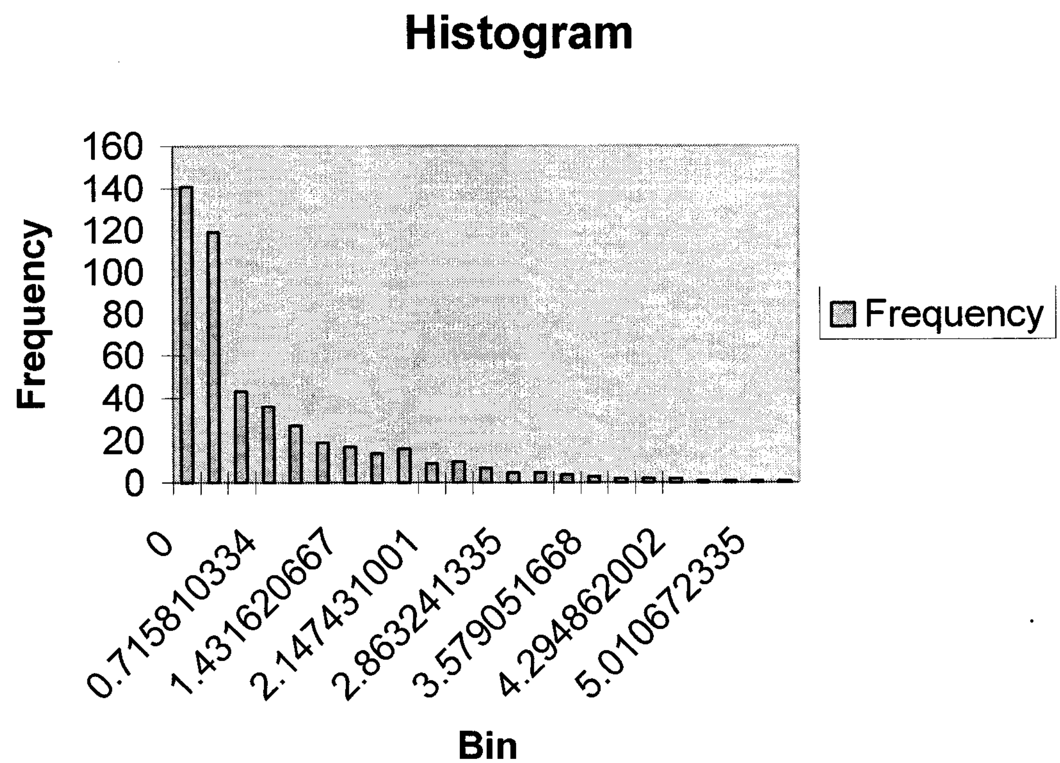


Figure 5.18. Results of the global uncertainty MC analysis. Note: Bin relates to the VBA command used to identify histogram classes in Microsoft Excel. .

5.5 Extended FAST

After the completion of the Monte Carlo simulation, the extended FAST was calculated using the sensitivity analysis software package SimLab. Graphical Results are depicted in Figure 5.19. Results of the corresponding extended FAST indicated that for first order analysis, the model was most sensitive to initial abstraction estimates followed by evaporation, the CN estimates, precipitation and initial abstraction for pervious areas. For total indices, most sensitive was the initial abstraction from impervious areas, followed by initial abstraction for pervious areas, precipitation, evaporation and finally the Curve Number. While it may seem reasonable for precipitation and evaporation to be highly sensitive, the much larger sensitivity of the initial abstraction estimates are somewhat surprising, this can be explained by the large range of possible values accepted as feasible for the initial abstraction estimates. This wide range of uncertainty in many ways can be related to the lack of more precise field data concerning losses as a result of depression

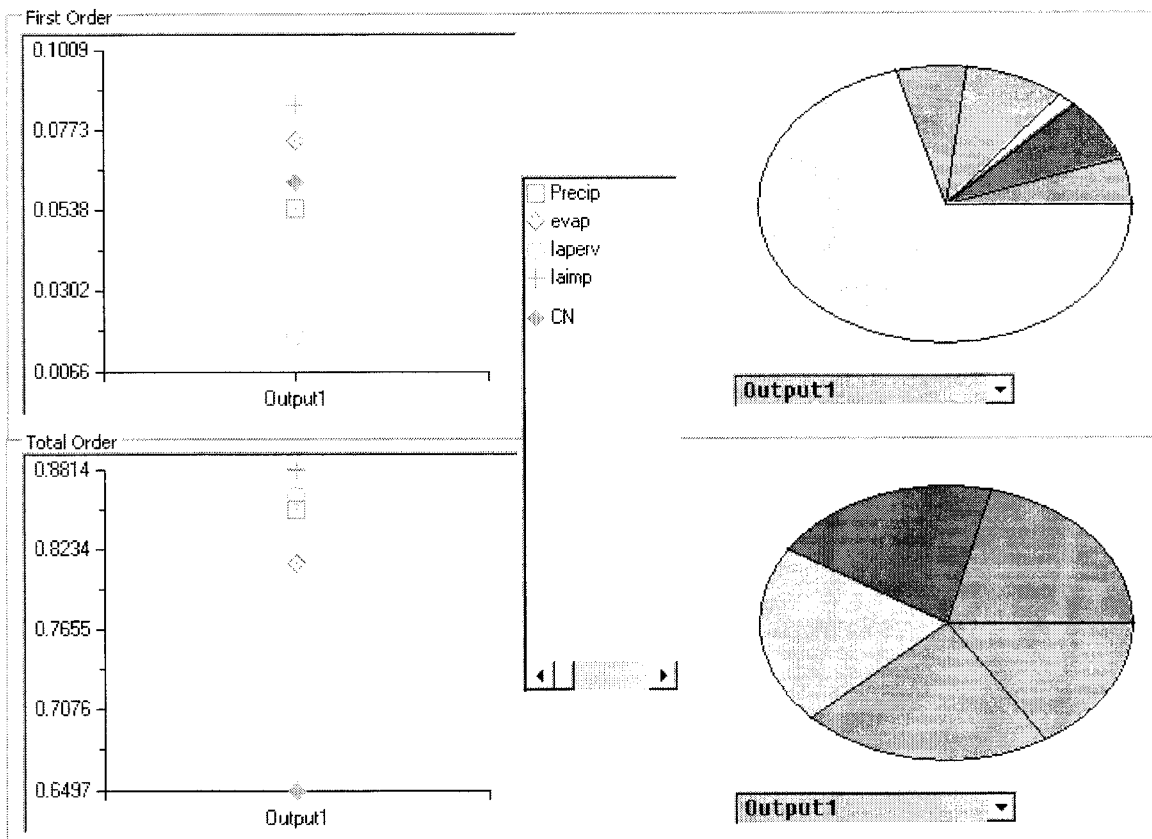


Figure 5.19. Results of the first and n^{th} order extended FAST sensitivity analysis. Note, gray area in the first pie graph indicates the proportion of sensitivity that can be attributed to additional n^{th} order interactions.

storage and interception loss. While again somewhat surprising, the low total sensitivity rank of the Curve Number can largely be attributed to the relatively narrow range of errors predicted by the Monte Carlo simulation (plus or minus approximately 3 STD) compared to the overall ranges of the IA values. Further, given the results of the OAT analysis, SA demonstrated the lower sensitivity of the model to small changes in Curve Numbers within the agricultural land use range of 65-75. Finally, the extended FAST has demonstrated the overall sensitivity of the model to small changes in precipitation and evaporation. This is apparent when first and nth order sensitivities are considered. It is likely that these model factors may prove more sensitive if errors due to rainfall and evaporation temporal aggregation, and distance from the study site were considered.

The goal of this research is to quantitatively demonstrate the effect of uncertainties arising from the use of spatial data in hydrological models in conjunction with GIS. Results have shown that in lumped models, error predictions resulting from uncertainty in geographic data can be somewhat small as a result of spatial averaging but that even small variations can have measurable effects when future scenarios are concerned. This is largely due to the increased sensitivity of Curve Number techniques for land uses undergoing transformation from agricultural landscapes to a more urban land use classification.

Through the use of the extended FAST technique it is demonstrated that model parameters causing the largest influence to total model uncertainty include those used to approximate initial abstraction losses. This suggests that more field research on the physical ranges that contribute to initial abstraction in urban and non-urban environments are needed if similar revised SCS-procedures are used. Further, although not surprising, it was shown that even small errors such as those associated with the measurement of climate input data can have a significant influence on model output even when n^{th} order interactions as well as issues concerning the lack of reasonable meteorological data are not considered.

While sensitivities due to scale were not examined in this research it is recommended that this be an area for future study. Specifically, where the derivation of a representative Curve Number is concerned. While the Curve Number in this case was shown to have rather small uncertainty, it is hypothesised that if a more distributed approached is taken, the lack of spatial averaging may produce larger degrees of uncertainty in the estimate of the Curve Number. Further, selection of an optimally scaled hydrological response unit will also be of consideration.

Due to practical implications only a small number of GIS related uncertainties were examined in this thesis. It is recommended that future research support the identification of stronger methods for determining the effect of GIS operations on the derivation of model outputs. This is especially recommended in the case of continuous modelling, as simulation of uncertainty for precipitation during periods of zero or near zero precipitation is difficult as propagating error for these periods often results in the error prone realization of days exhibiting negative rainfall.

Finally, the distance of rain and evaporation gauges, the aggregation of input data into daily average values and the positional accuracies in rain and evaporation stations all are likely sources contributing to model uncertainty. It is suggested that future investigations adopt some of the scenarios described in Section 4.1.4 for evaluating the contribution of these factors to uncertainty in model data.

REFERENCES

Al Smadi M. 1998. Incorporating Spatial and Temporal Variation of Watershed Response in a GIS Based Hydrological Model. Phd Thesis. Virginia Polytechnic University

Arbia G., Griffith D., Haining R., 1998. Error propagations modelling in raster GIS: overlay operations. *International Journal of Geographic Information Science* 12(2): 145-167.

Archer G.E.B., Saltelli A., Sobol I.M., 1997. Sensitivity measures, ANOVA-like techniques and the use of Bootstrap. *Journal of Statistics and Computer Simulation*, 58: 99-120.

Arnell N.W. 1999. A simple water balance model for the simulation of streamflow

Banks J., Carson J.S.(II), Nelson B.L., Nicol D.M., 2001, Discrete Event System Simulation. Prentice Hall, NJ.

Beck M.B. 1991. Principles of Modelling. Water Science and Technology. 24(6): 1-8.

Beven K, Binley A. 1992. The Future of Distributed Models: Model Calibration and Uncertainty Prediction. Hydrological Processes 6: 279-298.

Beven K. 1979. A Sensitivity Analysis of the Penman-Monteith Actual Evapotranspiration Estimates. Journal of Hydrology 44:169-190.
Bill James (ed) CHI Guelph.

Bobba A.G., Singh V.P., Bengtsson L., 2000. Application of environmental models to different hydrological systems. *Ecological Modelling* 125: 15-49.

Bouleta,, J. D. Kalmab, I. Brauda, M. Vauclina 1999. An assessment of effective land surface parameterisation in regional-scale water balance studies Journal of Hydrology 217: 225–238.

Brath A., Montanari A. 2000. The Effects of Spatial Variability of Soil Infiltration Capacility in Distributed Flood Modelling. Hydrological Processes 14: 2779-2794.

Braud, I., Dantas-Antonino, A.C., Vauclin, M., Thony, J.L., Ruelle,P., 1995. A simple soil-plant-atmosphere transfer model (SiSPAT), development and field verification. Jounal of Hydrology. 166, 231–250.

Burian S.J., Rocky D. 2001. Short Time Interval Disaggregation for Continuous Hydrological Simulation In:, Best Modelling Practices for Urban Systems Monograph 10

Burkhardt-Gammeter S., Fankhauser R. 1998. Analysis of Rainfall Time Series with Regard to Temporal Disaggregation in For the Use in Urban Hydrology. Water Science and Technology. 37(11): 65-72.

Campolongo F., Saltelli A., Sorensen T., Tarantola S. 2000. Hitchhiker's Guide to Sensitivity Analysis. In: Sensitivity Analysis. Saltelli A., Chan K., Scott E.M. (eds). Toronto Wiley.

Chang J.H., 1968. Climate and agriculture; an ecological survey. Aldine, Chicago.

Chin D., Sifang Zhao. 1995. Evaluation of Pan Evaporation Networks. Journal of Irrigation and Drainage. Sept Oct: 338-346.

Clarifica 2002. Water Budget in Urbanizing Watersheds: Duffins Creek Watershed. Clarifica Inc Water Resources Engineering and Consulting. For: Toronto and Region Conservation Authority.

Clark M.J. 1998. Putting Water in Its Place: A Perspective on GIS in hydrology and water management. Hydrological Processes. 12(6), 823-834.

Correia F.N., Filipe C.R., Da Graca Saraiva M., Ramos I., 1998. Coupling GIS with hydrological and hydraulic flood modelling. *Water Resources Management* 12: 229-249.

Crosetto M., Tarantola S., 2001. Uncertainty and sensitivity analysis: tools for GIS-based model implementation. *International Journal of Geographic Information Science* 15(5): 415-437.

Crosetto M., Tarantola S., Saltelli A., 2000. Sensitivity and uncertainty analysis in spatial modelling based on GIS. *Agriculture, Ecosystems and Environment* 81: 71-79.

Crouse R.P., Corbett E.S. Seegrist, D.W., 1966. Methods of measuring rainfall interception by grass. *International Association of Scientific Hydrology Bulletin* 11(2): 110-120.

Cukier R.I., Fortuin, C.M., Schuler K.E., Petschek A.G., Schailbly J.H., 1973. Sensitivity of coupled reaction systems to uncertainties in rate coefficients. Part I. *Theory. J. Chem. Phys.* 59(8), 3873-3878.

Daniel C. 1958. On varying one factor at a time. Biometrics 14: 430-431.

Date C.J. 1995. Introduction to Database Systems 6th Edition. Reading Mass. Addison Wesley.

Dent Borden 1996. Cartography: Thematic Map Design Forth Edition. Wm. C Brown. Toronto. 434 pp.

DeVantier B.A. and Feldman A.D. 1993. Review of GIS applications in Hydrologic Modelling. Journal of Water Resources Planning and Management. 119, 246-261.

Eicher C. 1990. Selection of Design Storms – Time Resolution Considerations. International Workshop on Urban Rainfall and Meteorology, St. Moritz, Switzerland. 2-5.

Einstein A. 1920 Relativity; The Special and General Theory. New York, Henry Holt.

Faures J.M., Goodrich D.C., Woolhiser D.E., Sorooshian S. 1995. Impact of Small Scale Spatial Variability on Runoff Modelling. *Journal of Hydrology* 173: 309-326.

Frankenberger J.R., Brooks E.S., M. Walter M.T., Walter M.F., Steenhuis M.S. 1999 A GIS-based variable source area hydrology model Hydrological Processes 13: 805-822.

Friedman R., Ansell C.S., Diamond S., Haimes Y.Y., 1984. The use of models for water resources management planning and policy. *Water Resources Research* 20: 793-802.

Goodchild M.F, Gopal S. 1989. The Accuracy of Spatial Databases. National Center for Geographic Information and Analysis. Taylor and Francis, Santa Barbara. CA.

Goodchild M.F., 1992. Geographic Information Science. *International Journal of Geographic Information Systems*. 6: 31-45.

Goodchild M.F., 2002 Introduction to Theoretical Models for Uncertain GIS. In: Spatial Data Quality. Shi W., Fisher P., Goodchild M.F. (ed). Taylor and Francis NY.

Goodchild, M.F., Guoqing, S., Shiren Y., 1992. Development and test of an error model

Goovaerts P., 2001. Geostatistical modelling of uncertainty in soil science. *Geoderma* 103: 3-26.

Green, W. H., and Ampt, G. A., 1911, Studies on soil physics, I--Flow of air and water through soils: *Journal of Agricultural Research*, v. 4, p. 1-24.

Gujer, W. Krejcik V. 1989. Niederschlag und Siedlungsentwasserung: Von Holer/Rein Zum historischen Ereignis, VSA-Report No. 388, Zurich

Guptil S.C. Morrison J.L. (ed) 1995. Elements of Spatial Data Quality. Elsevier NY.

Hamby D.M., 1994. A review of techniques for parameter sensitivity analysis of environmental models. *Environmental Monitoring and Assessment* 32: 135-154.

Harley J.B. 1990. Cartography Ethics and Social Theory. Cartographica 27: 1-23

Helton J.C.,1993. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliab. Eng. Syst. Safety* 42, 327-367.

Helvey J.D., Patrice J.H., 1965. Canopy and litter interception of rainfall by hardwood of eastern United States. *Water Resources Research.* 1: 193-206.

Henrickson A.K. 1975 The Map as an Idea: The Role of Cartography During the Second World War. American Cartographer 2: 19-53.

Heuvelink, G.B.M., 1998. Error Propagation in Environmental Modelling with GIS. Taylor and Francis, London.

Horton, R.E., 1940. An approach towards a physical interpretation of infiltration capacity. *Soil Science Society of America Proceedings* 4: 399-417.
hydro-ecological models of forest hydrology at the small catchment scale. Hydrological Processes 15: 1415-1459.
Hydrological Symposium 3. , Groundwater National Research Council of Canada.

Ibbit R.P. 1972. Effects of Random Data Errors on the Parameter Values for a Conceptual Model. Water Resources Research 8(1): 70-78
Jackson B.B., 1975. The use of streamflow models in planning. *Water Resources Research* 11(1): 54-63.

Kabat P, Hutjes R.W.A., Fedes R.A. 1997 The Scaling Charachteristics of Soil Paramters From Plot Scale Heterogeneity to sub Grid Scale Parameterization. Jounral of Hydrology 190: 363-396.

Karimi H, Houston B. 1998. Evaluating Strategies for Integrating Environmental Models with GIS: Current Trends and Future Needs. Computers Environment and Urban Systems. 20: 413-425.

Koren V.I., Finnerty B.D, Schaake, J.C., Smith, M.B. Seo D.J., Duan Q.Y.1999
Law A.M. and Kelton, W.D. 2000 Simulation Modelling and Analysis. 3rd Edition. McGraw Hill New York.

Kuczera G., Parent E. 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *Journal of Hydrology*. 211: 69-85

Lei J.H. 1996. Uncertainty Analysis of Urban Rainfall-Runoff Modelling. Phd Thesis., Faculty of Science, Norwegian University of Science and Technology. Trodenheim Norway. 171 pp.

Liang X., Xie Z. 2001. A New Surface Runoff Parameterization with Sub-Grid Scale Soil Heterogeneity. Advances in Water Resources 24: 1173-1193

Liang Xu, Xie, Z. 2001. A New Surface Runoff Parameterization with Sub-Grid Scale Heterogeneity for Land Surface Models. Advances in Water Resources 24: 1173-1193.

Lodwick W., Monson W., Svoboda L., 1990. Attribute error and sensitivity analysis of map operations in geographic information systems: suitability analysis. *International Journal of Geographic Information Systems*. 4(4): 413-428.
Longley, P.A., Goodchild M.G., Maguire D.J., Rhind D.W., 2001. Geographic Information Systems and Science. Toronto: Wiley 454 pp

Lull, H., 1964. Ecological and silvicultural aspects; in Handbook of Applied Hydrology. (ed) Ven te Chow, Mcgraw Hill, New York

Maidment D.R. 1996. GIS and Hydrological Modelling An Assessment of Progress. Paper presented at The Third Annual Conference on GIS and Environmental Modelling, Santa Fe NM

Mandelbrot B. 1983 The Fractal Geometry of Nature. San Francisco: Freeman.

Mayboom, P., 1962. Patterns of groundwater flow in the prairie profile. Proceedings of

Mays L.W. 2001. Water Resources Engineering. John Wiley and Sons, NY.

McCuen R.H. 1974. A Sensitivity and Error Analysis of Procedures Used for Estimating Evaporation. Water Resources Bulletin 10(3): 486-497.

McDonnell R.A. 1996. Including the Spatial Dimensions: Using Geographical Information Systems in Hydrology. Progress in Physical Geography 20: 159-177.

McKendry J.E. 2000. The Influence of Map Design on Resource Management Decision Making. Cartographica 37(2): 13-25.

Mckenney D.W., Mackey B.G., Zavitz B.L., 1999. Calibration and sensitivity analysis of a spatially-distributed solar radiation model. *International Journal of Geographic Information Science*.13(1): 49-65.

Mendicino G., 1999. Sensitivity analysis on GIS procedures for the estimate of soil erosion risk. *Natural Hazards* 20: 231-253.

Merriam R.A., 1961. Surface water storage on annual ryegrass. *Journal of Geophysical Research* 66: 1833-1838.

Moore, I.D. 1996. Hydrologic Modelling and GIS. In: M.F. Goodchild, B.O. Parks and L.T. Stayaert (ed), GIS and Environmental Modelling: Progress and Research Issues. Fort Collins CO: GIS World Books

Musgrave, G.W. 1955. How Much of the Rain Enters the Soil? Water Yearbook of Agriculture. US Department of Agriculture. Washington DC.

Mutzner H. 1991. The Significance of Aerial Rainfall Distribution for Flows from a Very Small Urban Catchment. Atmospheric Research. 27: 99-107.

Nace R., 1974. General evolution of the concept of the hydrological cycle., in Three Centuries of Scientific Hydrology (UNESCO: Paris).

NCDCDS 1988. The Proposed Standard for Digital Cartographic Data. The American Cartographer 15: 9- 140.

Newson M., 1994. Hydrology and the River Environment. Clarendon Press, Oxford

Nielson R.D., Hjelmfelt A.T. 2002. Hydrological Soil Group Assignment. NRCS. http://www.info.usda.gov/CED/

Niemczynowicz J. 1984. An Investigation of the Aerial and Dynamic Properties of Rainfall and its Influence on Runoff Generating Processes. Dept of Water Resources Engineering Lunds Institute of Technology, University of Lund, Report number 1005. Lund Sweden.

O'Neil, R.V. 1973 Error Analysis of Ecological Models. In Nelson D.J. (ed) Radionuclides in Ecosystems Conf. 710501 NTIS: Springfield, VA, 898-908.

Obled, C, Wendling, J. and Beven, K., 1994. The sensitivity of hydrological models to spatial rainfall patterns: an evaluation using observed data. J. Hydrol., 159: 305-333.

Office of Technology Assessment (OTA), 1982. Use of models for freshwater resource management, planning and policy, GPO Stock 052-003-00880-7, U.S. Government Printing Office, Washington D.C.

Ostrowski M.W. 2001. Modelling Urban Hydrological Processes and Management Scenarios at Different Temporal and Spatial Scales. In:, Best Modelling Practices for Urban Systems Monograph 10 Bill James (ed) CHI Guelph.
over a large geographic domain. Journal of Hydrology 217: 314-335.

Paturel J.E., Servat E., Vassiliadis A. 1995. Sensitivity of Conceptual Rainfall-runoff Models to Errors in Input Data; Case of the GR2M Model. Journal of Hydrology168: 11-25
.

Perkal J. 1956. On Epsilon Length. Bulletin de l'Academie Polonaise Des Sciences 4: 399-403.

Pipier B.S. 1989. Sensitivity of Penman Estimates of Evaporation to Errors in Input Data. Agricultural Water Mangement 15: 279-300.

Rodriguez, F., Andrieu H. and Zech Y. 2000. Evaluation of a distributed model for urban catchments using a 7-year continuous data series. Hydrological Processes 14: 899-914.

Rahman A., Weinmann P.E., Hoang T.M.T., Laurenson E.M. 2002. Monte Carlo Simulation of flood frequency curves from rainfall. *Journal of Hydrology*. 256: 196-210.

Saltelli A., Chan K., Scott E.M. (ed) 2000. Sensitivity Analysis. John Wiley and Sons, NY.

Saltelli A., Tarantola S., Chan K.P.S., 1999. A quantitative model independent method for global sensitivity analysis of model output. *Technometrics*, 41(1): 39-56.

Saxton K.E. 1975. Sensitivity Analyses of the Combination Evapotranspiration Equation. Agricultural Meteorology 15: 343-353.

Scale dependencies of hydrologic models to spatial variability of Precipitation. Journal of Hydrology 217: 285-302.

Servigne S., Ubeda T., Puricelli A., Laurini R. 1999. A Methodology for Spatial Consistency Improvement of Geographic Databases. Geoinformatica 4(1): 7-34.

Shah S.M.S., O'Connell P.E., Hosking J.R.M. Modelling 1996. The effects of spatial variability in rainfall on catchment response. 2. Experiments with distributed and lumped models. Journal of Hydrology. 175 89-111.

Shi W. 1998 A Generic Statistical Approach for Modelling Errors of Geometric Features in GIS. International Journal of Geographic Information Science. 12:131-143.

Singh V.P and Fiorentino M., 1996. Geographical Information Systems in Hydrology. Kluwer, Dordrecht Netherlands.

Singh V.P. 1997. Effect Of Spatial And Temporal Variability In Rainfall And Watershed Characteristics On Stream Flow Hydrograph. Hydrological Processes 11: 1649-1669.

Singh V.P. 1977. Sensitivity of Some Runoff Models to Errors in Rainfall Excess. Journal of
Hydrology. 33:301-318.

Singh V.P., 1996. Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, Colo.

Singh V.P., 1996. Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, Colo.

Singh V.P., Woolhiser D.A. 1976. Sensitivity of Linear and Nonlinear Surface Runoff Models to Input Errors. Journal of Hydrology. 29:243-249.

Singh, V.P. 1988 Hydrologic Systems Vol. 1. Rainfall-runoff Modelling. Prentice-Hall Englewood Cliffs New Jersey.

Singh, V.P. 1989 Hydrologic Systems Vol. 2. Watershed Modelling. Prentice-Hall Englewood Cliffs New Jersey.

Sobol I.M., 1993. Sensitivity analysis for nonlinear mathematical models. Mathematical *Modelling and Computational Experiment I.* 407-414.
Stoltenberg N.L., Wilson T.V., 1950. Interception storage and rainfall by corn plants, EOS. *American Geophysical Union Transactions* 31: 443-448.

Storck P., Bowling L. Wetherbee P., Lettenmaier D. 1998. Application of a GIS-based distributed hydrology model for prediction of forest harvest effects on peak streamflow in the Pacific Northwest. Hydrological Processes 12: 889-904.

Sui D.Z., Maggio R.C., 1999. Integrating GIS with hydrological modelling: practices, problems and prospects. *Computers Environment and Urban Systems* 23:33-51.

Tague C. L., Band L.E. 2001. Evaluating explicit and implicit routing for watershed

Thiessen A.H. 1911. Precipitation Averages for Large Areas. Monthly Water Review 39: 1082-1084.

Toth J., 1966. Mapping and interpretation of field phenomena for groundwater reconnaissance in a prairie environment, Alberta, Canada. *International Association of Scientific Hydrology Bulletin* 11(2): 20-68.

TRCA, 2003a Agricultural Non-Point Source Modelling of the Duffins and Carruthers Creek Watersheds. Toronto and Region Conservation Authority.

TRCA, 2003b Duffins Creek: State of the Watershed Report. Toronto and Region Conservation Authority.

Tsihrintzis V., Rizwan H, Fuentes H., 1996. Use of geographic information systems (GIS) in water resources: a review. *Water Resources Management* 10: 251-277.

Turanyi T., 1990. Sensitivity analysis of complex kinetic systems, tools and applications. *Journal of Mathematical Chemistry*, 5:203-248.

Tyner J. 1982. Persuasive Cartography. Journal of Geography 84: 140-44.
U.S. Geological Survey, 1954. Water-loss investigations: Lake Mead Studies. U.S. Geological Survey Professional Paper 298.

U.S. Soil Conservation Service, 1970. Irrigation water requirements. U.S. Department of Agriculture, Technical Release 21.

UCGIS 1996. Research Priorities for Geographic Information Science. Cartography and Geographic Information Systems. 23: 115-127.

Vachaud G., Chen T. 2002. Sensitivity of a large-scale hydrologic model to quality of input data obtained at different scales; distributed versus stochastic non-distributed modelling. Journal of Hydrology 264. 101-112.

Viessman W., Terence E.H., Knapp J.W. 1974. Introduction to Hydrology. Intext Educational Publishers NY.

Wood D. 1992. The Power of Maps. Guilford, NY.

Wood E.F., Sivapalan, M., Beven K.J. 1990. Similarity and Scale in Catchment Storm Response. Reviews Geophysics, 28 1-18.

Wood, E.F. Sivapalan, M., Beven K.J. Band, L. 1988. Effects of spatial variability and scale with implications to hydrological modelling. Journal of Hydrology. 102. 29-47.

Yagow E.R. 1997 Auxiliary Procedures for the AGNPS Model in Urban Fringe Watersheds. Phd Thesis. Virginia Polytechnic University

Zhang J., Goodchild M. F. 2002. Uncertainty in Geographic Information. Taylor and Francis NY.

Zhu H.J., Schilling W. 1996. Simulation Errors due to Temporal Rainfall Resolution Annual Combined Sewer Overflow. Atmospheric Research 42: 19-32.

Date Column A:

Time series in days ranging from April 1, 1997 to Oct 31 1997 for the calibration period and for the periods between April – October and from 1997-2000 for the entire input time series.

Measured Streamflow Column B:

The measured streamflow collected by HYDROMET station 02HC309 for Reesor Creek just south of confluence with Stouffville. Measured values are for daily average m^3/s.

9 Day Minimum Column C:

This Column calculates the nine day minimum value of discharge used for separating the baseflow from the runoff. The formulas used to calculate the nine day minimum returns the minimum value for the range of daily values from the previous six days, the current day and two days after the current day. For example, if the current day is April the 7, the formula for April 7, returns the minimum average daily streamflow from April 1 – April 9th. Values are for daily average m^3/s

Q Baseflow Column D:

This Column returns the baseflow for the day by taking the minimum of the lesser of the average value for the values in Column C ranging from four days previous to the current day to four days after the current day or the value for the current day.

As an example, the baseflow for April 5th would be calculated by returning the minimum value of either the average of Column C from April 1 to April 9 or the value of Column C for the current day. Values are for daily average m^3/s

Runoff Column E:

The runoff in Column E is calculated by taking the measured streamflow for each day in Column B minus the baseflow value for the same day calculated in Column D (B-D). Values are in daily average m^3/s

Precipitation Column F:

The precipitation data used is the measured daily average precipitation in average mm/day. Of the three initial datasets, rainfall data from Buttonville airport was found to be the most suitable in that it was geographically close to the watershed geographically. Further, time-series rain data was shown to graphically coincide with peaks in streamflow.

Evaporation Column G:

Pan evaporation data from Hamilton Harbor was utilized as a potential evaporation input to the model. This site was largely chosen out of necessity, as no other station data were available for the associated rainfall runoff time periods. Evaporation is in average daily mm of evaporation.

Balance for Impervious Areas

The balance for impervious areas is more simplified than that for pervious areas. Input in the form of precipitation is assumed to be lost due to initial abstraction in the form of depression storage where it can either be further lost to evaporation or become part of the runoff. In the case where input which surpasses storage in depressions or losses to evaporation, surplus water can be directly routed as stream runoff from areas which are directly connected or indirectly routed to pervious areas for impervious areas which are not directly connected to streamflow. In this case, indirectly connected runoff is considered routed to pervious areas and becomes part of water input to pervious areas. The percent of runoff which is routed directly to streamflow is calculated using the watershed parameter value representing the total fraction of impervious area within the entire study area (FimpA = 0.0455 or 4.6 percent of the total area). The total amount of area which is considered directly connected impervious uses a rule-of-thumb approach assuming approximately 75 percent of developed impervious areas is usually considered directly connected. This is represented by the watershed parameter value (FImpADC) or Fraction of Impervious area directly connected which = 0.0335, 3.4 percent of the total watershed area or approximately 75% of the total impervious area. The remaining fraction of impervious area indirectly connected is calculated by taking the total impervious area minus the total impervious area directly connected which leaves approximately 1.2% of total area (39 ha) as indirectly connected impervious area. Other assumptions made about the water balance for impervious areas are that water losses to evaporation only occur over a day for water which has initially been lost (stored) as initial abstractions. Further, the only available initial abstraction losses in pervious areas are those which are the result of depression storage. In impervious areas, a rule-of-thumb is used for calculating the maximum depression storage which the average value for the entire impervious proportion of the watershed and is equal to 0.8mm. The balance accounting for this process is as follows:

Dstor_avail_beg Column I:

This represents the total amount of abstraction storage available at the beginning of the day and is equal to the remaining depression storage left at the end of the previous day. The abstraction loss or depression storage in impervious areas at the end of the day is accounted for in Column O or Dstor_avail_end. For example, the value of Column I for April 2nd equals the value of Column O for April 1st. Values are in mm

Water_Prev_day Column J:

This Column represents the total amount of water left in depression storage at the end of the previous day. The amount of water left in depression storage at the end of the day is accounted for in Column N or Water_Final_day. For example, the water left in storage at the beginning of April 2nd will equal the water left in storage at the end of April 1st. This will also be a maximum represented by the watershed parameter IA_Imp_Max or maximum amount of abstraction loss averaged over the study area in mm. In this case, the value is equal to 0.8mm. The value of 0.8 mm is considered a standard value of abstraction losses over impervious areas. Thus, at no time can the water stored in abstraction exceed this depth, as it would become runoff.

Watercapt_cur_day Column K:

The water captured over the current day represents the amount of water captured as initial abstraction loss for a given day. This is dependent on the amount of incoming precipitation for a given day as well as the amount of storage available. In the case, that incoming precipitation is greater than the amount of storage available as specified in Column I, then the value captured will be the remainder of storage available in Column I. However in the case that the amount of incoming precipitation is not greater than the storage available in Column I, then the amount of water captured will equal the incoming precipitation. At no time can the amount of water captured for a given day exceed the maximum storage available or 0.8 mm for impervious areas. Values are in mm.

Water_Tot_day Column L:

The values in this Column represent the total water captured for the current day. It is calculated by adding the values in Column J and Column K. At no point can the total water captured into storage exceed the maximum abstraction of 0.8 for impervious areas. Values are in mm.

Water_loss_Evp Column M:

This Column calculated water lost to evaporation that has been initially abstracted. If the value for potential evaporation over the day in Column G is greater than the total water captured for the day in Column L then the value will equal the total water captured for the day in Column L. If the value for potential evaporation in Column G is less than the total amount of water captured for the day in Column L then the value returned in this Column is equal to the value for potential evaporation. Values cannot exceed the maximum initial abstraction storage of 0.8 for impervious areas. Values are in mm.

Water_Final_day Column N:

The value in this Column represents the total amount of water left in abstraction storage at the end of the day. This value is equal to the total amount of water captured for the day in Column L minuswater lost to evaporation represented by Column M. Values are in mm and cannot exceed the maximum value of 0.8mm.

Dstor_avail_end Column O:

This Column represents the amount of initial abstraction storage available at the end of the days balance. It is calculated by taking the maximum storage value of 0.8 minus the value for Column N. Values are in mm.

Runoff from Pervious Areas

RO_Imp Column P:

Runoff for impervious areas is calculated in Column P by returning the total amount of precipitation in mm in Column F minus the total amount of water captured in Column K. values are in average mm/d.

RO_vol_DCImp Column Q:

This Column represents the total volume of runoff generated for each day (d) from directly connected impervious areas in m^3. This total runoff is computed using the depth of input which exceeds the depth of initial abstraction in mm/1000 to give the value in meters and multiplied by the total area of the watershed that is considered impervious. The value is computed by taking the depth of runoff from Column P multiplied by the total watershed area which is directly connected impervious (A_DCImp = 109.37 ha) * 10000 to return the value in m^2. The formula is then:

Column P (mm)/ 1000(mm/m)*A_ACImp (ha)*10000(m_2/ha) = Average discharge in m^3/d.

RO_vol/s_DCImp Column R:

This Column represents the average daily discharge from directly connected impervious areas per second. It is calculated by taking the value in Column R in Average m^3/day divided by 86400 s/day. Values are in average daily discharge in m^3/s.

RO_vol_IDCImp Column S:

This Column represents the total runoff volume generated from indirectly connected impervious areas. It is computed similar to the values for directly connected impervious areas only the area value is represented by the total area in the watershed which is considered indirectly connected impervious area or (A_IDCImp = 39.2 ha) The formula is then: Column P (mm)/1000(mm/m)*A_AIDCImp (ha)*10000(m_2/ha) = Average discharge in m^3/d.

RO_vol/s_IDCImp Column T:

Similar to the value in Column R for directly connected impervious areas, the value in this Column equals the value in the previous Column/ 86400 s/day. Values are in average daily discharge in m^3/s.

Input to Pervious Areas

Pervious Input Column U:

Input to pervious areas is a function of the total runoff from indirectly connected impervious areas as well as the depth of precipitation falling in pervious areas. As the depth of rainfall falling in pervious areas is assumed to be uniform throughout the watershed this is equal to the depth of rainfall in Column F. The amount of water routed to pervious areas from impervious areas in mm is equal to the fraction of total impervious area directly connected (F_ImpADC = 0.0335 percent of watershed area) multiplied by the depth of runoff generated for the total amount of impervious area in the watershed (Column P). Thus the formula is [RO_Imp (mm)*F_ImpADC (%ha)] + Column F (mm) = Total depth of input to pervious areas (mm)

Pervious Areas

The water balance for pervious areas is very similar to the water balance for impervious areas in the manner it treats the balance of abstraction losses. However, initial abstraction losses in pervious areas are not only a product of depression storage but area also a function of initial infiltration and interception losses as a result of vegetation. Further, the fundamental difference between losses in pervious vs impervious areas is that a significant amount of moisture can be stored in the unsaturated zone of the soil horizon. The amount of water which can be stored is a function of the type of soil, its porosity, depth and the type of vegetation on the surface of the soil. These parameters and others make balancing the amount of loss to soil moisture storage difficult. The rainfall runoff equation developed by the Soil Conservation Service has developed a rainfall runoff equation which related the physical land conditions and their influence on the rainfall runoff relationship by use of a curve number (CN). This curve number, is used in conjunction with Equations 1 and 2 as well as with the initial abstraction (IA) loss balance to calculate the amount of soil moisture storage for a given area as well as to compute the rainfall runoff mass balance for pervious areas. This is achieved through the following steps within the simplified model.

IA_avail_perv Column V:

This represents the total amount of initial abstraction storage available at the beginning of the day and is equal to the remaining initial abstraction storage remaining at the end of the previous day. The abstraction loss or in pervious areas at the end of the day is accounted for in Column AC or IA_avail_end. For example, the value of Column V for April 2nd = the value of Column AC for April 1st. Values are in (mm) and cannot exceed the maximum initial abstraction loss for pervious areas specified as a model parameter where, IA_Perv_Max = maximum initial abstraction losses over pervious areas = 8 (mm).

Water_prev_day Column W:

This Column represents the total amount of water left in initial abstraction storage at the end of the previous day. The amount of water left in storage at the end of the day is accounted for in Column AB or Water_Final_day. For example, the water left in storage at the beginning of April 2nd will equal the water left in storage at the end of April 1st. This will also be a maximum represented by the watershed parameter IA_Perv_Max or maximum amount of abstraction loss averaged over the study area in (mm)

Watercapt_cur_day Column X:

The water captured over the current day represents the amount of water captured as initial abstraction loss for a given day. This is dependent on the amount of incoming input from precipitation and runoff from indirectly connected impervious areas (Column U), for a given day as well as the amount of initial abstraction storage available. In the case, that input is greater than the amount of storage available specified in Column V, then the value captured will be the remainder of storage available in Column V. However in the case that the amount of incoming precipitation is not greater than the storage available in Column V, than the amount of water captured will equal the input specified in Column U. At no time can the amount of water captured for a given day exceed the maximum storage available for pervious areas. Values are in mm.

Water_Tot_day Column Y:

The values in this Column represent the total water captured for the current day. It is calculated by adding the values in Column W and Column X. At no point can the total water captured into storage exceed the maximum abstraction of 8 mm for pervious areas. Values are in mm.

Water_loss_Evp Column Z:

This Column calculates water lost to evaporation that has been initially abstracted. If the value for potential evaporation over the day in Column G is greater than the total water captured for the day in Column Y than the value will equal the total water captured for the day in Column Y. If the value for potential evaporation in Column G is less than the total amount of water captured for the day in Column Y, then the value returned in this Column is equal to the value for potential evaporation. Values cannot exceed the maximum initial abstraction storage of 8 mm for pervious areas nor can they exceed either the amount of water in initial abstraction or the potential evaporation for the day. Values are in mm.

Water_loss_Inf Column AA:

This Column represents the depth of water initially lost to abstraction losses which will infiltrate over the course of the day. This is dependent on whether there is water left in storage after evaporation has occurred. The values in this Column are calculated using the calibration parameter IAPerv_Inf. This parameter described the amount of water depth in initial abstraction which will infiltrate over the day as a fraction of initially abstracted water depth which remains after evaporation. If there is more water in storage then can be removed as a result of evaporation then the remaining value is multiplied by the fraction of the remainder that will infiltrate divided by 100. The amount that returns equals the depth of abstraction storage which will infiltrate at the end of the day. The formulae for this is: Water_Tot_day (mm)- Water_loss_Evp (mm)]*IAPerv_Inf(%)/100 =
Depth of initially abstracted water lost to infiltration in (mm)

Water_Final_day Column AB:

The value in this Column represents the total amount of water left in abstraction storage at the end of the day. This value is equal to the total amount of water captured for the day in Column Y minus water lost to evaporation represented by Column Z and minus any water remaining in initial abstraction loss that infiltrated as specified in Column AA. Values are in mm and cannot exceed the maximum value.

IA_avail_end Column AC:

This Column represents the amount of initial abstraction storage available at the end of the days balance. It is calculated by taking the maximum storage value of for pervious areas minus the value for Column AB. Values are in (mm).

Runoff From Pervious Areas

Runoff from pervious areas is initially calculated as the depth of input which is remains after initial abstraction losses. This is achieved using the area weighted average curve number for the study area as defined in the first section and equations 1 and 2 where:

RO_Per Column AD:

Is calculated as:

$IA = (U - X)^2/(U\text{-}X) + S$

And:

$S = 254 *(100 - CN)/CN$

Where:

IA = initial abstraction losses for pervious areas (mm)

U = Column U or amount of input to pervious areas (mm)

X = Column X or the depth of initial abstraction available (mm)

S = Amount of soil moisture storage in (mm)

CN = the area weighted curve number for the study site = 81

RO_vol_Perv Column AE:

This Column represents the volume of runoff from pervious areas generated for the total pervious area in the study area. It is calculated by taking the depth of runoff for pervious areas calculated in Column AD in mm/1000 to give a value in m and multiplying it by the total area in the watershed that is pervious A_Perv = 3117.8 (ha)*10000 to convert it to a value of (m^2) giving a final value of Average (m^3/day).

RO_vol/s_Perv Column AF:

This Column simply divides the value in the previous Column by 86400 (s/day) to return a rate of daily average (m^3/s).

Total Runoff

Throughout the mass balance the amount of abstraction losses have been calculated as a depth in mm for impervious and pervious areas separately. The amount of runoff for each has also been balanced separately as depth followed by a conversion to volume based on the areas of directly connected impervious, indirectly connected impervious and pervious area. The final three Columns calculated total runoff by adding these values for each day.

RO Column AG:

This Column adds the depth of runoff from directly connected impervious areas in Column P to the depth of runoff generated from pervious areas in Column AD. Values are in mm.

RO_vol Column AH:

This is similar to the previous Column but instead ads the daily volume runoff for directly connected impervious areas and pervious areas in Columns Q and AE respectfully. Values are in (m^3/day).

RO_vol/s Column AI:

This final Column returns a value in daily average (m^3/s) by dividing the value in Column AH by 86400 (s/day).

Date	Streamflow			Runoff	Precipitation	Evaporation	Predicted Ref
			0.24				
01-Apr-97	1.15	#REF!		1.15	0.00	2.153	0.00
02-Apr-97	1.38	#REF!	0.30	1.08	0.00	2.185	0.00
03-Apr-97	1.44	#REF!	0.30	1.14	0.00	2.216	0.00
04-Apr-97	1.38	1.15	0.30	1.08	0.00	2.248	0.00
05-Apr-97	1.27	1.15	0.30	0.97	2.00	2.279	0.02
06-Apr-97	1.43	0.79	0.30	1.13	2.40	2.311	0.03
07-Apr-97	1.21	0.55	0.30	0.91	0.20	2.342	0.00
08-Apr-97	0.79	0.51	0.67	0.12	1.20	2.374	0.01
09-Apr-97	0.55	0.48	0.55	0.00	0.00	2.405	0.00
10-Apr-97	0.51	0.48	0.51	0.00	0.00	2.437	0.00
11-Apr-97	0.48	0.48	0.48	0.00	0.00	2.468	0.00
12-Apr-97	0.56	0.48	0.48	0.08	8.00	2.500	0.19
13-Apr-97	0.73	0.48	0.48	0.25	2.60	2.531	0.03
14-Apr-97	0.62	0.48	0.47	0.15	0.00	2.563	0.00
15-Apr-97	0.53	0.48	0.47	0.06	0.00	2.594	0.00
16-Apr-97	0.54	0.48	0.47	0.07	1.60	2.624	0.02
17-Apr-97	0.56	0.46	0.46	0.10	0.00	2.653	0.00
18-Apr-97	0.49	0.46	0.46	0.03	0.00	2.683	0.00
19-Apr-97	0.46	0.46	0.45	0.01	3.40	2.713	0.04
20-Apr-97	0.51	0.44	0.43	0.07	0.00	2.742	0.00
21-Apr-97	0.48	0.43	0.42	0.06	0.00	2.772	0.00
22-Apr-97	0.44	0.41	0.41	0.04	0.00	2.802	0.00
23-Apr-97	0.43	0.39	0.39	0.04	0.00	2.831	0.00
24-Apr-97	0.41	0.36	0.38	0.03	0.00	2.861	0.00
25-Apr-97	0.39	0.34	0.37	0.02	0.00	2.891	0.00

IMPERVIOUS AREAS						
Dstor_avail_beg	Water_Prev_day	Watercapt_cur_day	Water_Tot_day	Water_loss_Evp	Water_Final_day	Dstor_avail_end
					0	0.8
0.8	0	0	0	0	0	0.35
0.35	0	0	0	0	0	0.35
0.35	0	0	0	0	0	0.35
0.35	0	0	0	0	0	0.35
0.35	0	0.35	0.35	0.35	0	0.35
0.35	0	0.35	0.35	0.35	0	0.35
0.35	0	0.2	0.2	0.2	0	0.35
0.35	0	0.35	0.35	0.35	0	0.35
0.35	0	0	0	0	0	0.35
0.35	0	0	0	0	0	0.35
0.35	0	0	0	0	0	0.35
0.35	0	0.35	0.35	0.35	0	0.35
0.35	0	0.35	0.35	0.35	0	0.35
0.35	0	0	0	0	0	0.35
0.35	0	0	0	0	0	0.35
0.35	0	0.35	0.35	0.35	0	0.35
0.35	0	0	0	0	0	0.35
0.35	0	0	0	0	0	0.35
0.35	0	0.35	0.35	0.35	0	0.35
0.35	0	0	0	0	0	0.35
0.35	0	0	0	0	0	0.35
0.35	0	0	0	0	0	0.35
0.35	0	0	0	0	0	0.35
0.35	0	0	0	0	0	0.35
0.35	0	0	0	0	0	0.35

RUNOFF IMPERVIOUS					
RO_Imp	RO_vol_DCImp	RO_vol/s_DCImp	RO_vol_IDCImp	RO_vol/s_IDCImp	Pervious Input
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
1.7	1818.01125	0.021041797	606.00375	0.007013932	2.01856214
2.1	2258.74125	0.026142839	752.91375	0.00871428	2.423062053
0	0	0	0	0	0.2
0.9	936.55125	0.010839714	312.18375	0.003613238	1.209562314
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
7.7	8428.96125	0.097557422	2809.65375	0.032519141	8.08606083
2.3	2479.10625	0.028693359	826.36875	0.009564453	2.625312009
0	0	0	0	0	0
0	0	0	0	0	0
1.3	1377.28125	0.015940755	459.09375	0.005313585	1.614062227
0	0	0	0	0	0
0	0	0	0	0	0
3.1	3360.56625	0.038895443	1120.18875	0.012965148	3.434311834
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0

PERVIOUS AREAS							
IA_avail_perv	Water_prev_day	Watercapt_cur_day	Water_Tot_day	Water_loss_Evp	Water_loss_Inf	Water_Final_day	IA_avail_end
						0	0
0	0	0	0	0	0	0	3
3	0	0	0	0	0	0	3
3	0	0	0	0	0	0	3
3	0	0	0	0	0	0	3
3	0	2.01856214	2.01856214	2.01856214	0	0	3
3	0	2.423062053	2.423062053	2.310812131	0	0.112249922	2.887750078
2.887750078	0.112249922	0.2	0.312249922	0.312249922	0	0	3
3	0	1.209562314	1.209562314	1.209562314	0	0	3
3	0	0	0	0	0	0	3
3	0	0	0	0	0	0	3
3	0	0	0	0	0	0	3
3	0	3	3	2.499615596	0	0.500384404	2.499615596
2.499615596	0.500384404	2.499615596	3	2.531082841	0	0.468917159	2.531082841
2.531082841	0.468917159	0	0.468917159	0.468917159	0	0	3
3	0	0	0	0	0	0	3
3	0	1.614062227	1.614062227	1.614062227	0	0	3
3	0	0	0	0	0	0	3
3	0	0	0	0	0	0	3
3	0	3	3	2.712617966	0	0.287382034	2.712617966
2.712617966	0.287382034	0	0.287382034	0.287382034	0	0	3
3	0	0	0	0	0	0	3
3	0	0	0	0	0	0	3
3	0	0	0	0	0	0	3
3	0	0	0	0	0	0	3
3	0	0	0	0	0	0	3

RUNOFF PERVIOUS			TOTAL RUNOFF		
RO_Per	RO_vol_Perv	RO_vol/s_Perv	RO	RO_vol	RO_vol/s
0	0	0	0.0	0	0
0	0	0	0.0	0	0
0	0	0	0.0	0	0
0	0	0	0.0	0	0
0	0	0	0.0	1818.01125	0.021042
0	0	0	0.0	2258.74125	0.026143
0	0	0	0.0	0	0
0	0	0	0.0	936.55125	0.01084
0	0	0	0.0	0	0
0	0	0	0.0	0	0
0	0	0	0.0	0	0
0.249057	7765.148573	0.089874405	0.2	16194.10982	0.187432
0.00016	4.980641196	5.76463E-05	0.0	2484.086891	0.028751
0	0	0	0.0	0	0
0	0	0	0.0	0	0
0	0	0	0.0	1377.28125	0.015941
0	0	0	0.0	0	0
0	0	0	0.0	0	0
0.001901	59.2774849	0.000686082	0.0	3419.843735	0.039582
0	0	0	0.0	0	0
0	0	0	0.0	0	0
0	0	0	0.0	0	0
0	0	0	0.0	0	0
0	0	0	0.0	0	0
0	0	0	0.0	0	0

Date	Streamflow
01-Apr-97	1.15
02-Apr-97	1.38
03-Apr-97	1.44
04-Apr-97	1.38
05-Apr-97	1.27
06-Apr-97	1.43
07-Apr-97	1.21
08-Apr-97	0.79
09-Apr-97	0.55
10-Apr-97	0.51
11-Apr-97	0.48
12-Apr-97	0.56
13-Apr-97	0.73
14-Apr-97	0.62
15-Apr-97	0.53
16-Apr-97	0.54
17-Apr-97	0.56
18-Apr-97	0.49
19-Apr-97	0.46
20-Apr-97	0.51
21-Apr-97	0.48
22-Apr-97	0.44
23-Apr-97	0.43
24-Apr-97	0.41
25-Apr-97	0.39
26-Apr-97	0.36
27-Apr-97	0.34
28-Apr-97	0.64
29-Apr-97	0.50
30-Apr-97	0.41
01-May-97	0.36
02-May-97	0.35
03-May-97	1.36
04-May-97	0.95
05-May-97	0.61
06-May-97	0.72
07-May-97	0.47
08-May-97	0.44
09-May-97	0.66
10-May-97	0.67
11-May-97	0.52
12-May-97	0.70
13-May-97	0.52

14-May-97	0.40
15-May-97	0.52
16-May-97	0.54
17-May-97	0.47
18-May-97	0.41
19-May-97	0.40
20-May-97	0.38
21-May-97	0.34
22-May-97	0.31
23-May-97	0.31
24-May-97	0.29
25-May-97	0.28
26-May-97	0.28
27-May-97	0.26
28-May-97	0.22
29-May-97	0.23
30-May-97	0.26
31-May-97	0.25
01-Jun-97	0.24
02-Jun-97	0.23
03-Jun-97	0.21
04-Jun-97	0.19
05-Jun-97	0.17
06-Jun-97	0.18
07-Jun-97	0.19
08-Jun-97	0.19
09-Jun-97	0.20
10-Jun-97	0.21
11-Jun-97	0.21
12-Jun-97	0.22
13-Jun-97	0.22
14-Jun-97	0.21
15-Jun-97	0.22
16-Jun-97	0.27
17-Jun-97	0.27
18-Jun-97	0.28
19-Jun-97	0.32
20-Jun-97	0.29
21-Jun-97	0.32
22-Jun-97	0.94
23-Jun-97	0.42
24-Jun-97	0.77
25-Jun-97	0.52
26-Jun-97	0.37
27-Jun-97	0.31
28-Jun-97	0.27
29-Jun-97	0.24
30-Jun-97	0.23
01-Jul-97	0.23
02-Jul-97	0.21

03-Jul-97	0.22
04-Jul-97	0.24
05-Jul-97	0.20
06-Jul-97	0.20
07-Jul-97	0.27
08-Jul-97	0.27
09-Jul-97	0.42
10-Jul-97	0.27
11-Jul-97	0.24
12-Jul-97	0.21
13-Jul-97	0.18
14-Jul-97	0.16
15-Jul-97	0.18
16-Jul-97	0.18
17-Jul-97	0.18
18-Jul-97	0.16
19-Jul-97	0.16
20-Jul-97	0.14
21-Jul-97	0.21
22-Jul-97	0.19
23-Jul-97	0.15
24-Jul-97	0.13
25-Jul-97	0.15
26-Jul-97	0.13
27-Jul-97	0.13
28-Jul-97	0.15
29-Jul-97	0.13
30-Jul-97	0.12
31-Jul-97	0.12
01-Aug-97	0.12
02-Aug-97	0.13
03-Aug-97	0.14
04-Aug-97	0.12
05-Aug-97	0.12
06-Aug-97	0.12
07-Aug-97	0.12
08-Aug-97	0.11
09-Aug-97	0.11
10-Aug-97	0.11
11-Aug-97	0.14
12-Aug-97	0.15
13-Aug-97	0.27
14-Aug-97	0.18
15-Aug-97	0.29
16-Aug-97	0.34
17-Aug-97	0.25
18-Aug-97	0.20
19-Aug-97	0.16
20-Aug-97	0.19
21-Aug-97	0.41

22-Aug-97	0.32
23-Aug-97	0.28
24-Aug-97	0.24
25-Aug-97	0.23
26-Aug-97	0.23
27-Aug-97	0.22
28-Aug-97	0.27
29-Aug-97	0.27
30-Aug-97	0.24
31-Aug-97	0.23
01-Sep-97	0.24
02-Sep-97	0.22
03-Sep-97	0.22
04-Sep-97	0.20
05-Sep-97	0.19
06-Sep-97	0.24
07-Sep-97	0.32
08-Sep-97	0.21
09-Sep-97	0.20
10-Sep-97	0.35
11-Sep-97	0.45
12-Sep-97	0.30
13-Sep-97	0.28
14-Sep-97	0.24
15-Sep-97	0.25
16-Sep-97	0.23
17-Sep-97	0.24
18-Sep-97	0.25
19-Sep-97	0.25
20-Sep-97	0.29
21-Sep-97	0.26
22-Sep-97	0.23
23-Sep-97	0.22
24-Sep-97	0.21
25-Sep-97	0.29
26-Sep-97	0.31
27-Sep-97	0.27
28-Sep-97	0.25
29-Sep-97	0.35
30-Sep-97	0.32
01-Oct-97	0.27
02-Oct-97	0.27
03-Oct-97	0.27
04-Oct-97	0.26
05-Oct-97	0.24
06-Oct-97	0.24
07-Oct-97	0.21
08-Oct-97	0.22
09-Oct-97	0.24
10-Oct-97	0.25

11-Oct-97	0.25
12-Oct-97	0.24
13-Oct-97	0.25
14-Oct-97	0.26
15-Oct-97	0.24
16-Oct-97	0.24
17-Oct-97	0.26
18-Oct-97	0.27
19-Oct-97	0.24
20-Oct-97	0.25
21-Oct-97	0.25
22-Oct-97	0.28
23-Oct-97	0.27
24-Oct-97	0.27
25-Oct-97	0.28
26-Oct-97	0.29
27-Oct-97	0.68
28-Oct-97	0.47
29-Oct-97	0.41
30-Oct-97	0.35
31-Oct-97	0.33
01-Apr-98	0.95
02-Apr-98	1.02
03-Apr-98	1.13
04-Apr-98	0.84
05-Apr-98	0.69
06-Apr-98	0.60
07-Apr-98	0.57
08-Apr-98	0.56
09-Apr-98	0.55
10-Apr-98	0.47
11-Apr-98	0.50
12-Apr-98	0.45
13-Apr-98	0.44
14-Apr-98	0.41
15-Apr-98	0.41
16-Apr-98	0.77
17-Apr-98	1.55
18-Apr-98	0.85
19-Apr-98	0.64
20-Apr-98	0.64
21-Apr-98	0.54
22-Apr-98	0.48
23-Apr-98	0.44
24-Apr-98	0.48
25-Apr-98	0.51
26-Apr-98	0.47
27-Apr-98	0.42
28-Apr-98	0.38
29-Apr-98	0.38

30-Apr-98	0.36
01-May-98	0.36
02-May-98	0.37
03-May-98	0.39
04-May-98	0.43
05-May-98	0.41
06-May-98	0.36
07-May-98	0.34
08-May-98	0.34
09-May-98	0.33
10-May-98	0.32
11-May-98	0.98
12-May-98	0.66
13-May-98	0.48
14-May-98	0.39
15-May-98	0.35
16-May-98	0.32
17-May-98	0.30
18-May-98	0.28
19-May-98	0.28
20-May-98	0.27
21-May-98	0.25
22-May-98	0.24
23-May-98	0.23
24-May-98	0.22
25-May-98	0.19
26-May-98	0.21
27-May-98	0.20
28-May-98	0.16
29-May-98	0.19
30-May-98	0.18
31-May-98	0.18
01-Jun-98	0.19
02-Jun-98	0.33
03-Jun-98	0.37
04-Jun-98	0.24
05-Jun-98	0.21
06-Jun-98	0.20
07-Jun-98	0.20
08-Jun-98	0.19
09-Jun-98	0.18
10-Jun-98	0.18
11-Jun-98	0.17
12-Jun-98	0.46
13-Jun-98	0.32
14-Jun-98	0.36
15-Jun-98	0.32
16-Jun-98	0.27
17-Jun-98	0.26
18-Jun-98	0.25

19-Jun-98	0.25
20-Jun-98	0.25
21-Jun-98	0.25
22-Jun-98	0.25
23-Jun-98	0.27
24-Jun-98	0.29
25-Jun-98	0.26
26-Jun-98	0.54
27-Jun-98	0.36
28-Jun-98	0.30
29-Jun-98	0.28
30-Jun-98	0.35
01-Jul-98	0.32
02-Jul-98	0.24
03-Jul-98	0.21
04-Jul-98	0.32
05-Jul-98	0.28
06-Jul-98	0.25
07-Jul-98	0.67
08-Jul-98	0.39
09-Jul-98	0.33
10-Jul-98	0.28
11-Jul-98	0.22
12-Jul-98	0.19
13-Jul-98	0.19
14-Jul-98	0.18
15-Jul-98	0.15
16-Jul-98	0.20
17-Jul-98	0.23
18-Jul-98	0.17
19-Jul-98	0.14
20-Jul-98	0.16
21-Jul-98	0.15
22-Jul-98	0.13
23-Jul-98	0.13
24-Jul-98	0.13
25-Jul-98	0.13
26-Jul-98	0.13
27-Jul-98	0.13
28-Jul-98	0.15
29-Jul-98	0.15
30-Jul-98	0.13
31-Jul-98	0.14
01-Aug-98	0.12
02-Aug-98	0.11
03-Aug-98	0.12
04-Aug-98	0.12
05-Aug-98	0.12
06-Aug-98	0.22
07-Aug-98	0.59

08-Aug-98	0.31
09-Aug-98	0.37
10-Aug-98	0.50
11-Aug-98	0.28
12-Aug-98	0.22
13-Aug-98	0.19
14-Aug-98	0.19
15-Aug-98	0.19
16-Aug-98	0.19
17-Aug-98	0.16
18-Aug-98	0.18
19-Aug-98	0.15
20-Aug-98	0.14
21-Aug-98	0.16
22-Aug-98	0.18
23-Aug-98	0.22
24-Aug-98	0.88
25-Aug-98	0.38
26-Aug-98	0.30
27-Aug-98	0.24
28-Aug-98	0.21
29-Aug-98	0.21
30-Aug-98	0.19
31-Aug-98	0.16
01-Sep-98	0.15
02-Sep-98	0.24
03-Sep-98	0.20
04-Sep-98	0.19
05-Sep-98	0.17
06-Sep-98	0.16
07-Sep-98	0.21
08-Sep-98	0.18
09-Sep-98	0.18
10-Sep-98	0.16
11-Sep-98	0.16
12-Sep-98	0.15
13-Sep-98	0.17
14-Sep-98	0.15
15-Sep-98	0.39
16-Sep-98	0.33
17-Sep-98	0.20
18-Sep-98	0.19
19-Sep-98	0.18
20-Sep-98	0.21
21-Sep-98	0.21
22-Sep-98	0.20
23-Sep-98	0.17
24-Sep-98	0.16
25-Sep-98	0.17
26-Sep-98	0.21

27-Sep-98	0.30
28-Sep-98	0.24
29-Sep-98	0.19
30-Sep-98	0.20
01-Oct-98	0.30
02-Oct-98	0.21
03-Oct-98	0.20
04-Oct-98	0.20
05-Oct-98	0.21
06-Oct-98	0.19
07-Oct-98	0.42
08-Oct-98	0.40
09-Oct-98	0.30
10-Oct-98	0.26
11-Oct-98	0.24
12-Oct-98	0.23
13-Oct-98	0.22
14-Oct-98	0.27
15-Oct-98	0.26
16-Oct-98	0.23
17-Oct-98	0.23
18-Oct-98	0.25
19-Oct-98	0.25
20-Oct-98	0.25
21-Oct-98	0.23
22-Oct-98	0.23
23-Oct-98	0.23
24-Oct-98	0.24
25-Oct-98	0.25
26-Oct-98	0.23
27-Oct-98	0.24
28-Oct-98	0.26
29-Oct-98	0.24
30-Oct-98	0.24
31-Oct-98	0.24
01-Apr-00	0.41
02-Apr-00	0.41
03-Apr-00	0.42
04-Apr-00	0.54
05-Apr-00	0.51
06-Apr-00	0.45
07-Apr-00	0.44
08-Apr-00	1.00
09-Apr-00	0.79
10-Apr-00	0.66
11-Apr-00	0.54
12-Apr-00	0.53
13-Apr-00	0.53
14-Apr-00	0.49
15-Apr-00	0.47

16-Apr-00	0.47
17-Apr-00	0.44
18-Apr-00	0.43
19-Apr-00	0.41
20-Apr-00	0.56
21-Apr-00	2.41
22-Apr-00	1.66
23-Apr-00	0.91
24-Apr-00	0.64
25-Apr-00	0.51
26-Apr-00	0.45
27-Apr-00	0.42
28-Apr-00	0.40
29-Apr-00	0.37
30-Apr-00	0.37
01-May-00	0.38
02-May-00	0.39
03-May-00	0.35
04-May-00	0.34
05-May-00	0.34
06-May-00	0.32
07-May-00	0.32
08-May-00	0.31
09-May-00	0.31
10-May-00	0.63
11-May-00	0.50
12-May-00	1.63
13-May-00	4.59
14-May-00	1.12
15-May-00	0.74
16-May-00	0.58
17-May-00	0.50
18-May-00	2.08
19-May-00	1.11
20-May-00	0.76
21-May-00	0.59
22-May-00	0.51
23-May-00	0.74
24-May-00	1.25
25-May-00	0.78
26-May-00	0.59
27-May-00	0.49
28-May-00	0.44
29-May-00	0.40
30-May-00	0.37
31-May-00	0.37
01-Jun-00	0.34
02-Jun-00	0.43
03-Jun-00	0.35
04-Jun-00	0.33

05-Jun-00	0.31
06-Jun-00	0.40
07-Jun-00	0.35
08-Jun-00	0.30
09-Jun-00	0.29
10-Jun-00	0.30
11-Jun-00	1.15
12-Jun-00	0.97
13-Jun-00	4.26
14-Jun-00	1.64
15-Jun-00	1.72
16-Jun-00	0.87
17-Jun-00	0.68
18-Jun-00	0.65
19-Jun-00	0.61
20-Jun-00	0.51
21-Jun-00	0.57
22-Jun-00	0.53
23-Jun-00	0.48
24-Jun-00	0.42
25-Jun-00	3.45
26-Jun-00	1.15
27-Jun-00	0.77
28-Jun-00	0.56
29-Jun-00	0.64
30-Jun-00	0.72
01-Jul-00	0.53
02-Jul-00	0.43
03-Jul-00	0.96
04-Jul-00	0.94
05-Jul-00	0.58
06-Jul-00	0.45
07-Jul-00	0.41
08-Jul-00	0.37
09-Jul-00	0.36
10-Jul-00	0.39
11-Jul-00	0.34
12-Jul-00	0.29
13-Jul-00	0.29
14-Jul-00	0.77
15-Jul-00	0.64
16-Jul-00	0.76
17-Jul-00	1.08
18-Jul-00	1.31
19-Jul-00	0.66
20-Jul-00	0.50
21-Jul-00	0.43
22-Jul-00	0.40
23-Jul-00	0.37
24-Jul-00	0.35

25-Jul-00	0.33
26-Jul-00	0.30
27-Jul-00	0.31
28-Jul-00	0.32
29-Jul-00	0.31
30-Jul-00	0.32
31-Jul-00	0.32
01-Aug-00	0.61
02-Aug-00	1.09
03-Aug-00	0.59
04-Aug-00	0.42
05-Aug-00	0.36
06-Aug-00	0.33
07-Aug-00	0.50
08-Aug-00	0.54
09-Aug-00	0.70
10-Aug-00	0.55
11-Aug-00	0.91
12-Aug-00	0.93
13-Aug-00	0.61
14-Aug-00	0.47
15-Aug-00	0.43
16-Aug-00	0.40
17-Aug-00	0.37
18-Aug-00	0.36
19-Aug-00	0.36
20-Aug-00	0.33
21-Aug-00	0.33
22-Aug-00	0.32
23-Aug-00	0.70
24-Aug-00	0.51
25-Aug-00	0.41
26-Aug-00	0.38
27-Aug-00	0.40
28-Aug-00	0.37
29-Aug-00	0.30
30-Aug-00	0.28
31-Aug-00	0.28
01-Sep-00	0.28
02-Sep-00	0.32
03-Sep-00	0.33
04-Sep-00	0.35
05-Sep-00	0.30
06-Sep-00	0.29
07-Sep-00	0.28
08-Sep-00	0.29
09-Sep-00	0.29
10-Sep-00	0.34
11-Sep-00	0.85
12-Sep-00	0.61

13-Sep-00	0.44
14-Sep-00	0.84
15-Sep-00	1.04
16-Sep-00	0.64
17-Sep-00	0.49
18-Sep-00	0.43
19-Sep-00	0.39
20-Sep-00	0.37
21-Sep-00	0.44
22-Sep-00	0.41
23-Sep-00	0.60
24-Sep-00	0.54
25-Sep-00	0.46
26-Sep-00	0.41
27-Sep-00	0.39
28-Sep-00	0.36
29-Sep-00	0.34
30-Sep-00	0.35
01-Oct-00	0.35
02-Oct-00	0.35
03-Oct-00	0.36
04-Oct-00	0.41
05-Oct-00	0.40
06-Oct-00	0.44
07-Oct-00	0.46
08-Oct-00	0.44
09-Oct-00	0.41
10-Oct-00	0.41
11-Oct-00	0.39
12-Oct-00	0.37
13-Oct-00	0.37
14-Oct-00	0.37
15-Oct-00	0.38
16-Oct-00	0.37
17-Oct-00	0.36
18-Oct-00	0.36
19-Oct-00	0.37
20-Oct-00	0.36
21-Oct-00	0.39
22-Oct-00	0.36
23-Oct-00	0.35
24-Oct-00	0.35
25-Oct-00	0.37
26-Oct-00	0.37
27-Oct-00	0.40
28-Oct-00	0.43
29-Oct-00	0.38
30-Oct-00	0.37
31-Oct-00	0.37

Date	Precipitation	Evaporation
01-Apr-97	0.00	2.153
02-Apr-97	0.00	2.185
03-Apr-97	0.00	2.216
04-Apr-97	0.00	2.248
05-Apr-97	2.00	2.279
06-Apr-97	2.40	2.311
07-Apr-97	0.20	2.342
08-Apr-97	1.20	2.374
09-Apr-97	0.00	2.405
10-Apr-97	0.00	2.437
11-Apr-97	0.00	2.468
12-Apr-97	8.00	2.500
13-Apr-97	2.60	2.531
14-Apr-97	0.00	2.563
15-Apr-97	0.00	2.594
16-Apr-97	1.60	2.624
17-Apr-97	0.00	2.653
18-Apr-97	0.00	2.683
19-Apr-97	3.40	2.713
20-Apr-97	0.00	2.742
21-Apr-97	0.00	2.772
22-Apr-97	0.00	2.802
23-Apr-97	0.00	2.831
24-Apr-97	0.00	2.861
25-Apr-97	0.00	2.891
26-Apr-97	0.00	2.920
27-Apr-97	3.80	2.950
28-Apr-97	7.20	2.979
29-Apr-97	0.00	3.009
30-Apr-97	0.00	3.039
01-May-97	0.40	3.068
02-May-97	1.20	3.098
03-May-97	27.20	3.128
04-May-97	0.50	3.157
05-May-97	15.00	3.187
06-May-97	0.00	3.217
07-May-97	0.00	3.246
08-May-97	0.40	3.276
09-May-97	4.60	3.306
10-May-97	1.80	3.335
11-May-97	6.20	3.365

12-May-97	4.40	3.395
13-May-97	0.00	3.424
14-May-97	0.20	3.454
15-May-97	6.10	3.484
16-May-97	0.60	3.508
17-May-97	1.80	3.532
18-May-97	2.20	3.557
19-May-97	0.80	3.581
20-May-97	0.00	3.606
21-May-97	0.00	3.630
22-May-97	0.00	3.655
23-May-97	0.00	3.679
24-May-97	0.00	3.703
25-May-97	0.00	3.728
26-May-97	0.00	3.752
27-May-97	0.00	3.777
28-May-97	0.00	3.801
29-May-97	0.60	3.826
30-May-97	1.00	3.850
31-May-97	0.00	3.874
01-Jun-97	0.00	3.874
02-Jun-97	0.00	3.899
03-Jun-97	0.00	3.923
04-Jun-97	0.00	3.948
05-Jun-97	0.00	3.972
06-Jun-97	0.00	3.997
07-Jun-97	1.00	4.021
08-Jun-97	9.40	4.045
09-Jun-97	0.00	4.070
10-Jun-97	0.00	4.094
11-Jun-97	0.00	4.119
12-Jun-97	2.50	4.143
13-Jun-97	4.20	4.168
14-Jun-97	0.00	4.192
15-Jun-97	0.00	4.216
16-Jun-97	8.40	4.223
17-Jun-97	1.40	4.230
18-Jun-97	5.80	4.237
19-Jun-97	0.00	4.244
20-Jun-97	7.80	4.251
21-Jun-97	23.20	4.258
22-Jun-97	0.00	4.265
23-Jun-97	3.20	4.272
24-Jun-97	11.80	4.279
25-Jun-97	0.00	4.286
26-Jun-97	0.00	4.293
27-Jun-97	0.00	4.300
28-Jun-97	0.00	4.307

29-Jun-97	0.00	4.314
30-Jun-97	0.00	4.321
01-Jul-97	0.00	4.328
02-Jul-97	0.00	4.335
03-Jul-97	1.70	4.342
04-Jul-97	0.80	4.349
05-Jul-97	0.00	4.356
06-Jul-97	5.00	4.363
07-Jul-97	0.00	4.370
08-Jul-97	16.40	4.377
09-Jul-97	0.60	4.384
10-Jul-97	0.00	4.391
11-Jul-97	0.00	4.398
12-Jul-97	0.00	4.405
13-Jul-97	0.00	4.412
14-Jul-97	0.00	4.419
15-Jul-97	8.40	4.426
16-Jul-97	0.00	4.397
17-Jul-97	3.80	4.369
18-Jul-97	11.50	4.340
19-Jul-97	0.00	4.312
20-Jul-97	0.00	4.283
21-Jul-97	8.80	4.255
22-Jul-97	0.00	4.226
23-Jul-97	0.00	4.198
24-Jul-97	0.00	4.169
25-Jul-97	0.00	4.141
26-Jul-97	0.00	4.112
27-Jul-97	13.20	4.084
28-Jul-97	1.60	4.055
29-Jul-97	0.00	4.027
30-Jul-97	0.00	3.998
31-Jul-97	0.00	3.970
01-Aug-97	0.00	3.970
02-Aug-97	0.00	3.941
03-Aug-97	0.00	3.912
04-Aug-97	0.00	3.884
05-Aug-97	0.00	3.855
06-Aug-97	0.00	3.827
07-Aug-97	0.00	3.798
08-Aug-97	0.00	3.770
09-Aug-97	0.00	3.741
10-Aug-97	0.00	3.713
11-Aug-97	0.00	3.684
12-Aug-97	5.50	3.656
13-Aug-97	3.00	3.627
14-Aug-97	0.00	3.599
15-Aug-97	17.40	3.570

16-Aug-97	7.00	3.527
17-Aug-97	0.00	3.485
18-Aug-97	0.00	3.442
19-Aug-97	0.00	3.399
20-Aug-97	15.70	3.356
21-Aug-97	16.40	3.313
22-Aug-97	5.60	3.270
23-Aug-97	0.20	3.228
24-Aug-97	1.60	3.185
25-Aug-97	0.00	3.142
26-Aug-97	0.00	3.099
27-Aug-97	1.20	3.056
28-Aug-97	0.00	3.013
29-Aug-97	0.00	2.971
30-Aug-97	0.00	2.928
31-Aug-97	0.80	2.885
01-Sep-97	0.00	2.885
02-Sep-97	0.00	2.842
03-Sep-97	0.00	2.799
04-Sep-97	0.00	2.757
05-Sep-97	1.60	2.714
06-Sep-97	14.00	2.671
07-Sep-97	0.00	2.628
08-Sep-97	0.00	2.585
09-Sep-97	0.00	2.542
10-Sep-97	16.80	2.500
11-Sep-97	0.80	2.457
12-Sep-97	0.20	2.414
13-Sep-97	0.00	2.371
14-Sep-97	0.00	2.328
15-Sep-97	0.00	2.285
16-Sep-97	0.00	2.249
17-Sep-97	5.00	2.213
18-Sep-97	0.00	2.177
19-Sep-97	4.40	2.141
20-Sep-97	1.00	2.105
21-Sep-97	0.00	2.069
22-Sep-97	0.00	2.033
23-Sep-97	0.40	1.997
24-Sep-97	0.00	1.961
25-Sep-97	6.00	1.925
26-Sep-97	0.00	1.889
27-Sep-97	0.00	1.853
28-Sep-97	0.60	1.817
29-Sep-97	13.00	1.781
30-Sep-97	3.00	1.745
01-Oct-97	0.00	1.709
02-Oct-97	1.80	1.672

03-Oct-97	1.80	1.636
04-Oct-97	0.00	1.600
05-Oct-97	0.00	1.564
06-Oct-97	0.00	1.528
07-Oct-97	0.00	1.492
08-Oct-97	0.00	1.456
09-Oct-97	1.20	1.420
10-Oct-97	0.00	1.384
11-Oct-97	0.00	1.348
12-Oct-97	0.00	1.312
13-Oct-97	0.00	1.276
14-Oct-97	0.40	1.240
15-Oct-97	0.00	1.204
16-Oct-97	0.00	1.182
17-Oct-97	0.00	1.160
18-Oct-97	0.00	1.138
19-Oct-97	0.00	1.116
20-Oct-97	0.00	1.094
21-Oct-97	0.80	1.072
22-Oct-97	0.00	1.050
23-Oct-97	0.40	1.028
24-Oct-97	0.00	1.006
25-Oct-97	0.00	0.984
26-Oct-97	16.00	0.962
27-Oct-97	9.40	0.940
28-Oct-97	0.00	0.918
29-Oct-97	1.80	0.896
30-Oct-97	0.00	0.874
31-Oct-97	3.40	0.852
01-Apr-98	7.40	2.153
02-Apr-98	5.40	2.185
03-Apr-98	0.00	2.216
04-Apr-98	0.00	2.248
05-Apr-98	0.00	2.279
06-Apr-98	0.00	2.311
07-Apr-98	0.00	2.342
08-Apr-98	4.20	2.374
09-Apr-98	0.00	2.405
10-Apr-98	0.00	2.437
11-Apr-98	0.00	2.468
12-Apr-98	0.00	2.500
13-Apr-98	0.00	2.531
14-Apr-98	2.00	2.563
15-Apr-98	0.20	2.594
16-Apr-98	19.40	2.624
17-Apr-98	15.00	2.653
18-Apr-98	0.00	2.683

19-Apr-98	6.40	2.713
20-Apr-98	4.60	2.742
21-Apr-98	0.00	2.772
22-Apr-98	0.00	2.802
23-Apr-98	0.00	2.831
24-Apr-98	0.00	2.861
25-Apr-98	0.00	2.891
26-Apr-98	0.00	2.920
27-Apr-98	0.00	2.950
28-Apr-98	0.00	2.979
29-Apr-98	0.00	3.009
30-Apr-98	0.00	3.039
01-May-98	1.40	3.068
02-May-98	4.40	3.098
03-May-98	0.00	3.128
04-May-98	0.40	3.157
05-May-98	0.00	3.187
06-May-98	0.00	3.217
07-May-98	0.00	3.246
08-May-98	0.00	3.276
09-May-98	0.00	3.306
10-May-98	15.40	3.335
11-May-98	18.10	3.365
12-May-98	0.00	3.395
13-May-98	0.00	3.424
14-May-98	0.00	3.454
15-May-98	0.00	3.484
16-May-98	0.00	3.508
17-May-98	0.00	3.532
18-May-98	0.00	3.557
19-May-98	3.60	3.581
20-May-98	0.00	3.606
21-May-98	0.00	3.630
22-May-98	0.00	3.655
23-May-98	0.00	3.679
24-May-98	0.00	3.703
25-May-98	0.00	3.728
26-May-98	0.00	3.752
27-May-98	0.00	3.777
28-May-98	0.00	3.801
29-May-98	2.80	3.826
30-May-98	0.00	3.850
31-May-98	1.80	3.874
01-Jun-98	0.00	3.874
02-Jun-98	9.60	3.899
03-Jun-98	0.00	3.923
04-Jun-98	0.00	3.948
05-Jun-98	0.00	3.972

06-Jun-98	0.00	3.997
07-Jun-98	1.00	4.021
08-Jun-98	0.00	4.045
09-Jun-98	0.00	4.070
10-Jun-98	1.00	4.094
11-Jun-98	4.30	4.119
12-Jun-98	12.00	4.143
13-Jun-98	0.00	4.168
14-Jun-98	0.00	4.192
15-Jun-98	0.00	4.216
16-Jun-98	1.50	4.223
17-Jun-98	1.00	4.230
18-Jun-98	0.00	4.237
19-Jun-98	0.00	4.244
20-Jun-98	0.00	4.251
21-Jun-98	0.00	4.258
22-Jun-98	0.00	4.265
23-Jun-98	3.40	4.272
24-Jun-98	0.00	4.279
25-Jun-98	2.00	4.286
26-Jun-98	13.40	4.293
27-Jun-98	0.00	4.300
28-Jun-98	0.00	4.307
29-Jun-98	0.00	4.314
30-Jun-98	11.60	4.321
01-Jul-98	0.00	4.328
02-Jul-98	0.00	4.335
03-Jul-98	0.00	4.342
04-Jul-98	7.20	4.349
05-Jul-98	0.00	4.356
06-Jul-98	2.50	4.363
07-Jul-98	28.00	4.370
08-Jul-98	1.50	4.377
09-Jul-98	4.00	4.384
10-Jul-98	0.00	4.391
11-Jul-98	0.00	4.398
12-Jul-98	0.00	4.405
13-Jul-98	0.00	4.412
14-Jul-98	0.00	4.419
15-Jul-98	0.00	4.426
16-Jul-98	2.20	4.397
17-Jul-98	0.00	4.369
18-Jul-98	0.00	4.340
19-Jul-98	1.60	4.312
20-Jul-98	0.00	4.283
21-Jul-98	0.00	4.255
22-Jul-98	0.00	4.226
23-Jul-98	0.00	4.198

24-Jul-98	0.00	4.169
25-Jul-98	0.00	4.141
26-Jul-98	0.00	4.112
27-Jul-98	4.20	4.084
28-Jul-98	0.40	4.055
29-Jul-98	0.00	4.027
30-Jul-98	0.00	3.998
31-Jul-98	0.00	3.970
01-Aug-98	0.00	3.970
02-Aug-98	0.00	3.941
03-Aug-98	0.00	3.912
04-Aug-98	0.00	3.884
05-Aug-98	0.00	3.855
06-Aug-98	18.60	3.827
07-Aug-98	19.40	3.798
08-Aug-98	0.00	3.770
09-Aug-98	12.80	3.741
10-Aug-98	1.00	3.713
11-Aug-98	0.00	3.684
12-Aug-98	0.00	3.656
13-Aug-98	0.00	3.627
14-Aug-98	0.00	3.599
15-Aug-98	0.80	3.570
16-Aug-98	0.00	3.527
17-Aug-98	0.50	3.485
18-Aug-98	0.00	3.442
19-Aug-98	0.00	3.399
20-Aug-98	0.00	3.356
21-Aug-98	0.00	3.313
22-Aug-98	0.00	3.270
23-Aug-98	5.20	3.228
24-Aug-98	11.40	3.185
25-Aug-98	10.00	3.142
26-Aug-98	0.00	3.099
27-Aug-98	0.00	3.056
28-Aug-98	2.00	3.013
29-Aug-98	2.60	2.971
30-Aug-98	0.00	2.928
31-Aug-98	0.00	2.885
01-Sep-98	0.00	2.885
02-Sep-98	3.60	2.842
03-Sep-98	0.60	2.799
04-Sep-98	0.00	2.757
05-Sep-98	0.00	2.714
06-Sep-98	4.20	2.671
07-Sep-98	0.00	2.628
08-Sep-98	2.00	2.585
09-Sep-98	0.00	2.542

10-Sep-98	0.00	2.500
11-Sep-98	0.00	2.457
12-Sep-98	0.00	2.414
13-Sep-98	0.00	2.371
14-Sep-98	3.00	2.328
15-Sep-98	13.40	2.285
16-Sep-98	0.00	2.249
17-Sep-98	0.00	2.213
18-Sep-98	0.00	2.177
19-Sep-98	0.00	2.141
20-Sep-98	0.00	2.105
21-Sep-98	0.00	2.069
22-Sep-98	1.20	2.033
23-Sep-98	0.00	1.997
24-Sep-98	0.00	1.961
25-Sep-98	0.00	1.925
26-Sep-98	2.80	1.889
27-Sep-98	1.80	1.853
28-Sep-98	0.00	1.817
29-Sep-98	0.20	1.781
30-Sep-98	10.00	1.745
01-Oct-98	0.40	1.709
02-Oct-98	0.00	1.672
03-Oct-98	0.00	1.636
04-Oct-98	0.00	1.600
05-Oct-98	0.00	1.564
06-Oct-98	0.00	1.528
07-Oct-98	18.80	1.492
08-Oct-98	1.20	1.456
09-Oct-98	0.00	1.420
10-Oct-98	0.00	1.384
11-Oct-98	0.00	1.348
12-Oct-98	0.00	1.312
13-Oct-98	0.00	1.276
14-Oct-98	3.20	1.240
15-Oct-98	0.00	1.204
16-Oct-98	0.00	1.182
17-Oct-98	0.00	1.160
18-Oct-98	0.00	1.138
19-Oct-98	0.00	1.116
20-Oct-98	0.00	1.094
21-Oct-98	0.00	1.072
22-Oct-98	0.00	1.050
23-Oct-98	0.00	1.028
24-Oct-98	0.00	1.006
25-Oct-98	0.00	0.984
26-Oct-98	0.00	0.962
27-Oct-98	0.20	0.940

28-Oct-98	0.60	0.918
29-Oct-98	0.00	0.896
30-Oct-98	0.00	0.874
31-Oct-98	0.00	0.852
01-Apr-00	0.00	2.153
02-Apr-00	0.80	2.185
03-Apr-00	2.20	2.216
04-Apr-00	4.80	2.248
05-Apr-00	0.20	2.279
06-Apr-00	1.80	2.311
07-Apr-00	5.40	2.342
08-Apr-00	11.40	2.374
09-Apr-00	0.00	2.405
10-Apr-00	0.20	2.437
11-Apr-00	3.20	2.468
12-Apr-00	0.60	2.500
13-Apr-00	0.00	2.531
14-Apr-00	0.00	2.563
15-Apr-00	0.00	2.594
16-Apr-00	0.00	2.624
17-Apr-00	0.00	2.653
18-Apr-00	0.20	2.683
19-Apr-00	0.00	2.713
20-Apr-00	32.00	2.742
21-Apr-00	6.40	2.772
22-Apr-00	0.20	2.802
23-Apr-00	0.00	2.831
24-Apr-00	0.00	2.861
25-Apr-00	0.00	2.891
26-Apr-00	0.00	2.920
27-Apr-00	0.00	2.950
28-Apr-00	0.00	2.979
29-Apr-00	0.00	3.009
30-Apr-00	0.00	3.039
01-May-00	3.00	3.068
02-May-00	0.00	3.098
03-May-00	0.00	3.128
04-May-00	0.00	3.157
05-May-00	0.00	3.187
06-May-00	0.00	3.217
07-May-00	1.40	3.246
08-May-00	0.00	3.276
09-May-00	9.40	3.306
10-May-00	6.20	3.335
11-May-00	17.80	3.365
12-May-00	64.80	3.395
13-May-00	0.00	3.424
14-May-00	0.00	3.454

15-May-00	0.00	3.484
16-May-00	1.20	3.508
17-May-00	0.20	3.532
18-May-00	24.00	3.557
19-May-00	0.00	3.581
20-May-00	0.00	3.606
21-May-00	0.00	3.630
22-May-00	0.00	3.655
23-May-00	15.40	3.679
24-May-00	0.20	3.703
25-May-00	1.40	3.728
26-May-00	0.00	3.752
27-May-00	0.00	3.777
28-May-00	0.00	3.801
29-May-00	0.00	3.826
30-May-00	0.00	3.850
31-May-00	1.20	3.874
01-Jun-00	2.80	3.874
02-Jun-00	0.40	3.899
03-Jun-00	0.00	3.923
04-Jun-00	0.00	3.948
05-Jun-00	3.20	3.972
06-Jun-00	4.20	3.997
07-Jun-00	0.00	4.021
08-Jun-00	0.00	4.045
09-Jun-00	3.60	4.070
10-Jun-00	0.00	4.094
11-Jun-00	26.00	4.119
12-Jun-00	0.20	4.143
13-Jun-00	41.00	4.168
14-Jun-00	14.40	4.192
15-Jun-00	0.00	4.216
16-Jun-00	0.80	4.223
17-Jun-00	0.00	4.230
18-Jun-00	8.60	4.237
19-Jun-00	0.00	4.244
20-Jun-00	0.00	4.251
21-Jun-00	6.20	4.258
22-Jun-00	2.80	4.265
23-Jun-00	0.00	4.272
24-Jun-00	9.60	4.279
25-Jun-00	27.40	4.286
26-Jun-00	3.20	4.293
27-Jun-00	0.20	4.300
28-Jun-00	0.00	4.307
29-Jun-00	9.80	4.314
30-Jun-00	0.00	4.321
01-Jul-00	2.80	4.328

02-Jul-00	0.40	4.335
03-Jul-00	0.00	4.342
04-Jul-00	0.00	4.349
05-Jul-00	3.20	4.356
06-Jul-00	4.20	4.363
07-Jul-00	0.00	4.370
08-Jul-00	0.00	4.377
09-Jul-00	3.60	4.384
10-Jul-00	0.00	4.391
11-Jul-00	26.00	4.398
12-Jul-00	0.20	4.405
13-Jul-00	41.00	4.412
14-Jul-00	14.40	4.419
15-Jul-00	0.00	4.426
16-Jul-00	0.80	4.397
17-Jul-00	0.00	4.369
18-Jul-00	8.60	4.340
19-Jul-00	0.00	4.312
20-Jul-00	0.00	4.283
21-Jul-00	6.20	4.255
22-Jul-00	2.80	4.226
23-Jul-00	0.00	4.198
24-Jul-00	9.60	4.169
25-Jul-00	27.40	4.141
26-Jul-00	3.20	4.112
27-Jul-00	0.20	4.084
28-Jul-00	0.00	4.055
29-Jul-00	9.80	4.027
30-Jul-00	0.00	3.998
31-Jul-00	0.00	3.970
01-Aug-00	11.00	3.970
02-Aug-00	0.00	3.941
03-Aug-00	0.00	3.912
04-Aug-00	0.00	3.884
05-Aug-00	0.00	3.855
06-Aug-00	0.20	3.827
07-Aug-00	2.20	3.798
08-Aug-00	15.80	3.770
09-Aug-00	0.00	3.741
10-Aug-00	8.60	3.713
11-Aug-00	3.60	3.684
12-Aug-00	0.00	3.656
13-Aug-00	0.20	3.627
14-Aug-00	0.00	3.599
15-Aug-00	0.80	3.570
16-Aug-00	0.00	3.527
17-Aug-00	0.00	3.485
18-Aug-00	1.80	3.442

19-Aug-00	0.20	3.399
20-Aug-00	0.00	3.356
21-Aug-00	0.00	3.313
22-Aug-00	6.00	3.270
23-Aug-00	14.20	3.228
24-Aug-00	0.20	3.185
25-Aug-00	0.00	3.142
26-Aug-00	2.60	3.099
27-Aug-00	0.40	3.056
28-Aug-00	0.00	3.013
29-Aug-00	0.00	2.971
30-Aug-00	0.00	2.928
31-Aug-00	0.00	2.885
01-Sep-00	0.00	2.885
02-Sep-00	6.40	2.842
03-Sep-00	0.80	2.799
04-Sep-00	0.00	2.757
05-Sep-00	0.00	2.714
06-Sep-00	0.00	2.671
07-Sep-00	0.00	2.628
08-Sep-00	0.00	2.585
09-Sep-00	0.00	2.542
10-Sep-00	13.40	2.500
11-Sep-00	0.20	2.457
12-Sep-00	0.80	2.414
13-Sep-00	0.00	2.371
14-Sep-00	25.20	2.328
15-Sep-00	0.00	2.285
16-Sep-00	0.00	2.249
17-Sep-00	0.00	2.213
18-Sep-00	0.00	2.177
19-Sep-00	0.00	2.141
20-Sep-00	4.20	2.105
21-Sep-00	1.20	2.069
22-Sep-00	1.20	2.033
23-Sep-00	14.60	1.997
24-Sep-00	0.00	1.961
25-Sep-00	0.00	1.925
26-Sep-00	0.00	1.889
27-Sep-00	0.00	1.853
28-Sep-00	0.00	1.817
29-Sep-00	0.00	1.781
30-Sep-00	0.00	1.745
01-Oct-00	0.00	1.709
02-Oct-00	0.00	1.672
03-Oct-00	0.00	1.636
04-Oct-00	4.20	1.600
05-Oct-00	1.40	1.564

06-Oct-00	4.60	1.528
07-Oct-00	5.20	1.492
08-Oct-00	0.00	1.456
09-Oct-00	0.00	1.420
10-Oct-00	0.00	1.384
11-Oct-00	0.00	1.348
12-Oct-00	0.00	1.312
13-Oct-00	0.00	1.276
14-Oct-00	0.00	1.240
15-Oct-00	1.80	1.204
16-Oct-00	0.00	1.182
17-Oct-00	0.00	1.160
18-Oct-00	2.00	1.138
19-Oct-00	0.00	1.116
20-Oct-00	0.00	1.094
21-Oct-00	0.00	1.072
22-Oct-00	0.00	1.050
23-Oct-00	0.20	1.028
24-Oct-00	0.20	1.006
25-Oct-00	0.00	0.984
26-Oct-00	0.00	0.962
27-Oct-00	0.60	0.940
28-Oct-00	0.00	0.918
29-Oct-00	0.00	0.896
30-Oct-00	0.00	0.874
31-Oct-00	0.00	0.852

APPENDIX V Curve Number Error Model Monte Carlo

```
Sub Output_Processor()
'
' Macro1 Macro
' Macro recorded 6/25/2003 by Harry Manson
'

'The GIS export will contain duplicate shape objects because during the buffer process, some regions with varying
'Soil types will inevitably overlap. As a result it is neccesary to reduce the records such that there is one
'for each shape but that each possible soil and landuse is documented for each record rather that haveing a series
'of Records for each possible outcome. Further Scripts will examine the stochastic nature of which region will
'posses which soil type. (Manson, June 2003)


Dim rngCurrent As Range
Dim dblfirstvalue As Double
Dim intArea As Integer, intTrials As Integer
Dim strArea As String, strRange As String, strSoil As String, strLanduse As String
Dim rngArea As Range, rngMonte As Range
Dim i As Integer, j As Integer, k As Integer, l As Integer, m As Integer, n As Integer

Set rngCurrent = Cells(1, 3)


Do While rngCurrent.Offset(0, -1) <> ""

   If rngCurrent.Cells.Offset(1, 0).Value <> rngCurrent.Cells.Value Then
     rngCurrent.Cells.Offset(0, 1).Value = rngCurrent.Cells.Offset(0, -2).Value
     Set rngCurrent = rngCurrent.Cells.Offset(1, 0)

   ElseIf rngCurrent.Value = rngCurrent.Cells.Offset(1, 0).Value And rngCurrent.Value <>
rngCurrent.Cells.Offset(2, 0) Then
      rngCurrent.Cells.Offset(0, 1).Value = rngCurrent.Cells.Offset(0, -2).Value
      rngCurrent.Cells.Offset(0, 2).Value = rngCurrent.Cells.Offset(1, -2).Value
      Set rngCurrent = rngCurrent.Cells.Offset(2, 0)

   ElseIf rngCurrent.Cells.Value = rngCurrent.Cells.Offset(1, 0).Value And
rngCurrent.Value = rngCurrent.Cells.Offset(2, 0) And _
        rngCurrent.Cells.Value <> rngCurrent.Cells.Offset(3, 0).Value Then
```

```
        rngCurrent.Cells.Offset(0, 1).Value = rngCurrent.Cells.Offset(0, -2).Value
        rngCurrent.Cells.Offset(0, 2).Value = rngCurrent.Cells.Offset(1, -2).Value
        rngCurrent.Cells.Offset(0, 3).Value = rngCurrent.Cells.Offset(2, -2).Value
        Set rngCurrent = rngCurrent.Cells.Offset(3, 0)
    ElseIf rngCurrent.Cells.Value = rngCurrent.Cells.Offset(3, 0).Value Then
        rngCurrent.Cells.Offset(0, 1).Value = rngCurrent.Cells.Offset(0, -2).Value
        rngCurrent.Cells.Offset(0, 2).Value = rngCurrent.Cells.Offset(1, -2).Value
        rngCurrent.Cells.Offset(0, 3).Value = rngCurrent.Cells.Offset(2, -2).Value
        rngCurrent.Cells.Offset(0, 4).Value = rngCurrent.Cells.Offset(2, -2).Value
        Set rngCurrent = rngCurrent.Cells.Offset(4, 0)
    End If

Loop

' This Section removes the unecessary Records

dblfirstvalue = Range("C2").Value
ActiveCell = Cells(2, 3).Select

Do While ActiveCell.Offset(0, -1) <> ""
    If IsEmpty(ActiveCell.Offset(0, 1)) Then
        ActiveCell.EntireRow.Delete
    Else
        ActiveCell.Offset(1, 0).Select
    End If
Loop

Range("C2").Value = dblfirstvalue ' This is neccesary for some reason otherwise a value of
"TRUE" is returned as
                        ' Apposed to the Area




' This Section Sums the total area in Hectares and fills a column with proportions for each
shape record

i = 3
j = Range("C2").End(xlDown).Row
strArea = "(" & "C" & i & ":" & "C" & j & ")"
intArea = Evaluate("SUM" & strArea)
Set rngTest = Cells(1, 2)
rngTest.Cells.Value = intTest

Range("H1").Value = "Area Proportion"
Set rngCurrent = Cells(2, "H")
```

```
   Do While rngCurrent.Cells.Offset(0, -4) <> ""
      rngCurrent.Cells.Value = (rngCurrent.Cells.Offset(0, -5).Value) / (intArea)
      Set rngCurrent = rngCurrent.Cells.Offset(1, 0)
   Loop

End Sub

'This Section will birandomize values based on the lookup table random functions and will
incorporate
'A randomized selection of the lookup table based on the 1 to 4 possible soil types occuring
in each shape

Sub MonteCarlo()

Dim rngCurrent As Range
Dim dblfirstvalue As Double, dblCN As Double
Dim intArea As Integer, intTrials As Integer
Dim strArea As String, strRange As String, strSoil As String, strLanduse As String, strCN
As String
Dim rngArea As Range, rngMonte As Range
Dim i As Integer, j As Integer, k As Integer, l As Integer, m As Integer, n As Integer

Range("I1").Value = "CN Range"


Set rngMonte = Worksheets("Stochastic Modeler").Cells(1, 1)
strCN = "(I2:I672)"
intTrials = Application.InputBox(Prompt:="Number of Trails", Type:=1)
n = Range("I2").End(xlDown).Row
strCN = "(" & "I" & "2" & ":" & "I" & n & ")"

For i = 1 To intTrials




   For j = 2 To 672
      Set rngCurrent = Cells(j, 9)

      If rngCurrent.Cells.Offset(0, -5) <> "" And rngCurrent.Cells.Offset(0, -4) = "" Then
         strSoil = rngCurrent.Cells.Offset(0, -5).Value
            If strSoil = "A" Then
               k = 14
            ElseIf strSoil = "B" Then
               k = 15
            ElseIf strSoil = "C" Then
```

```
            k = 16
          ElseIf strSoil = "D" Then
            k = 17
          End If

      strLanduse = rngCurrent.Cells.Offset(0, -7).Value
          If strLanduse = "Agriculture/Rural" Then
            l = 5
          ElseIf strLanduse = "Meadow" Then
            l = 6
          ElseIf strLanduse = "Forest" Then
            l = 7
          ElseIf strLanduse = "Federal Airport Lands" Then
            l = 8
          ElseIf strLanduse = "Urban" Then
            l = 9
          ElseIf strLanduse = "Urban Open Space" Then
            l = 10
          ElseIf strLanduse = "Wetlands" Then
            l = 11
          End If
      strRange = Cells(l, k).Value

    ElseIf rngCurrent.Cells.Offset(0, -4) <> "" And rngCurrent.Cells.Offset(0, -3) = ""
Then
      m = Evaluate("RANDBETWEEN(-5,-4)")

      strSoil = rngCurrent.Cells.Offset(0, m).Value
          If strSoil = "A" Then
            k = 14
          ElseIf strSoil = "B" Then
            k = 15
          ElseIf strSoil = "C" Then
            k = 16
          ElseIf strSoil = "D" Then
            k = 17
          End If

      strLanduse = rngCurrent.Cells.Offset(0, -7).Value
          If strLanduse = "Agriculture/Rural" Then
            l = 5
          ElseIf strLanduse = "Meadow" Then
            l = 6
          ElseIf strLanduse = "Forest" Then
            l = 7
          ElseIf strLanduse = "Federal Airport Lands" Then
```

```
            l = 8
          ElseIf strLanduse = "Urban" Then
            l = 9
          ElseIf strLanduse = "Urban Open Space" Then
            l = 10
          ElseIf strLanduse = "Wetlands" Then
            l = 11
          End If
      strRange = Cells(l, k).Value


    ElseIf rngCurrent.Cells.Offset(0, -5) <> "" And rngCurrent.Cells.Offset(0, -2) <> ""
Then
      m = Evaluate("RANDBETWEEN(-5,-2)")

      strSoil = rngCurrent.Cells.Offset(0, m).Value
          If strSoil = "A" Then
            k = 14
          ElseIf strSoil = "B" Then
            k = 15
          ElseIf strSoil = "C" Then
            k = 16
          ElseIf strSoil = "D" Then
            k = 17
          End If

      strLanduse = rngCurrent.Cells.Offset(0, -7).Value
          If strLanduse = "Agriculture/Rural" Then
            l = 5
          ElseIf strLanduse = "Meadow" Then
            l = 6
          ElseIf strLanduse = "Forest" Then
            l = 7
          ElseIf strLanduse = "Federal Airport Lands" Then
            l = 8
          ElseIf strLanduse = "Urban" Then
            l = 9
          ElseIf strLanduse = "Urban Open Space" Then
            l = 10
          ElseIf strLanduse = "Wetlands" Then
            l = 11
          End If
      strRange = Cells(l, k).Value

    End If
```

```
        rngCurrent.Cells.Value = (Evaluate("RANDBETWEEN" & strRange)) *
(rngCurrent.Cells.Offset(0, -1).Value)
        Set rngCurrent = rngCurrent.Cells.Offset(1, 0)

    Next j



dblCN = Evaluate("SUM" & strCN)
rngMonte.Cells.Value = dblCN
Set rngMonte = rngMonte.Cells.Offset(1, 0)


Next i

End Sub
```

APPENDIX V I Mean Displacement of Street Intersections

Point Num	Location	Measured Easting	Measured Northing	GIS Easting	GIS Northing	Diff Easting	Diff Northing	RMSE
1		641948	4869816	641937.56	4869599.93	10.44	216.07	216.3220712
2		641885	4869818	641881.95	4869581.56	3.05	236.44	236.4596712
3		641600	4869725	641591.47	4869484	8.53	241	241.150909
4		641464	4869781	641452.46	4869550.75	11.54	230.25	230.5390078
5		641583	4869807	641561.72	4869586.57	21.28	220.43	221.4547884
6		641912	4870022	641900.38	4869799.19	11.62	222.81	223.1127977
7		641908	4870114	641881.34	4869891.22	26.66	222.78	224.3695256
8		641883	4870197	641865.02	4869973.74	17.98	223.26	223.9828297
9		641860	4870375	641831.02	4870156.44	28.98	218.56	220.4729326
10		641816	4870577	641792.03	4870358.64	23.97	218.36	219.671688
11		641674	4870530	641650.12	4870310.13	23.88	219.87	221.1629971
12		641495	4870483	641498.25	4870257.99	-3.25	225.01	225.03347
13		641312	4870438	641289.7	4870191.35	22.3	246.65	247.6560367
14		641188	4870379	641162.76	4870149.73	25.24	229.27	230.6551333
15		640870	4870279	640861.28	4870046.27	8.72	232.73	232.8933045
16		640845	4870361	640826.82	4870137.85	18.18	223.15	223.8893363
17		640717	4870580	640700.79	4870357.73	16.21	222.27	222.8603083
18		640683	4870683	640658.17	4870454.75	24.83	228.25	229.596584
19		640619	4870551	640593.79	4870323.28	25.21	227.72	229.1112012
20		640629	4870484	640618.27	4870261.62	10.73	222.38	222.6387147
21		640659	4870441	640652.73	4870188.63	6.27	252.37	252.4478754
22		640707	4870345	640694.47	4870124.71	12.53	220.29	220.6460627
23		640730	4870284	640717.11	4870055.34	12.89	228.66	229.0230288
24		640749	4870240	640734.34	4870002.3	14.66	237.7	238.1516441
25		640606	4870196	640603.31	4869958.32	2.69	237.68	237.6952219
26		640587	4870270	640582.91	4870015.45	4.09	254.55	254.5828561
27		640038	4870004	640030.27	4869769.72	7.73	234.28	234.4074899
28		639914	4870033	639884.28	4869806.45	29.72	226.55	228.4910959
29		639804	4870637	639775.93	4870413.5	28.07	223.5	225.2557988
30		639864	4870663	639850.28	4870435.71	13.72	227.29	227.7037165

Mean	229.7146
Standard	9.604919

```
Sub Relative_Error()

Static sngCalibresult As Single

Dim rngStart As Range, rngRestore As Range
Dim rngCurrent As Range, rngMbcell
Dim sngPercent As Single, intDiv As Integer
Dim i As Integer, j As Integer, sngResults As Single, intCount As Integer
Dim sngX As Single


sngCalibresult = Worksheets("Totals for Validation Year").Range("J17").Value
Set rngStart = Worksheets("OAT Sensitivity").Range("B3")
sngDiv = 100

For i = 3 To 10        ' Column numbers corresponding to each parameter column
Set rngStart = Worksheets("OAT Sensitivity").Cells(3, i)
    If rngStart = Worksheets("OAT Sensitivity").Cells(3, 2) Then 'OK Precip
      Set rngCurrent = Worksheets("OAT Sensitivity").Cells(3, 2)
      intCount = Worksheets("Mass Balance").Range("BE4").End(xlDown).Row
         Do While rngCurrent.Offset(0, -1).Value <> ""
           sngPercent = rngCurrent.Offset(0, -1).Value / sngDiv

           For j = 4 To intCount
              Set rngMbcell = Worksheets("Mass Balance").Cells(j, "BF")
              sngX = rngMbcell.Cells.Offset(0, -1).Value * sngPercent
              rngMbcell.Value = rngMbcell.Cells.Offset(0, -1).Value + sngX
           Next j

           sngResults = Worksheets("Totals for Validation Year").Range("J17").Value
           rngCurrent.Value = sngResults - sngCalibresult
           Set rngCurrent = rngCurrent.Cells.Offset(1, 0)
         Loop
           For j = 4 To intCount
              Set rngMbcell = Worksheets("Mass Balance").Cells(j, "BF")
              rngMbcell.Value = rngMbcell.Cells.Offset(0, -1)
           Next j

    ElseIf rngStart = Worksheets("OAT Sensitivity").Cells(3, 3) Then 'OK Evap
      Set rngCurrent = Worksheets("OAT Sensitivity").Cells(3, 3)
      intCount = Worksheets("Mass Balance").Range("BE4").End(xlDown).Row
         Do While rngCurrent.Offset(0, -2).Value <> ""
          sngPercent = rngCurrent.Offset(0, -2).Value / sngDiv
```

```
        For j = 4 To intCount
          Set rngMbcell = Worksheets("Mass Balance").Cells(j, "BH")
          sngX = rngMbcell.Cells.Offset(0, -1).Value * sngPercent
          rngMbcell.Value = rngMbcell.Cells.Offset(0, -1).Value + sngX
        Next j

         sngResults = Worksheets("Totals for Validation Year").Range("J17").Value
         rngCurrent.Value = sngResults - sngCalibresult
         Set rngCurrent = rngCurrent.Cells.Offset(1, 0)


     Loop
        For j = 4 To intCount
          Set rngMbcell = Worksheets("Mass Balance").Cells(j, "BH")
          rngMbcell.Value = rngMbcell.Cells.Offset(0, -1)
        Next j




ElseIf rngStart = Worksheets("OAT Sensitivity").Cells(3, 4) Then 'OK A
   Set rngCurrent = Worksheets("OAT Sensitivity").Cells(3, 4)
     Do While rngCurrent.Offset(0, -3).Value <> ""
        sngPercent = rngCurrent.Offset(0, -3).Value / sngDiv
        Set rngMbcell = Worksheets("Mass Balance").Cells(8, "AN")
        sngX = rngMbcell.Cells.Offset(0, -3).Value * sngPercent
        rngMbcell.Value = rngMbcell.Cells.Offset(0, -3).Value + sngX
        sngResults = Worksheets("Totals for Validation Year").Range("J17").Value
        rngCurrent.Value = sngResults - sngCalibresult
        Set rngCurrent = rngCurrent.Cells.Offset(1, 0)
     Loop
rngMbcell.Value = rngMbcell.Cells.Offset(0, -3)


ElseIf rngStart = Worksheets("OAT Sensitivity").Cells(3, 5) Then 'OK A_Imp
  Set rngCurrent = Worksheets("OAT Sensitivity").Cells(3, 5)
     Do While rngCurrent.Offset(0, -4).Value <> ""
        sngPercent = rngCurrent.Offset(0, -4).Value / sngDiv
        Set rngMbcell = Worksheets("Mass Balance").Cells(9, "AN")
        sngX = rngMbcell.Cells.Offset(0, -3).Value * sngPercent
        rngMbcell.Value = rngMbcell.Cells.Offset(0, -3).Value + sngX
        sngResults = Worksheets("Totals for Validation Year").Range("J17").Value
        rngCurrent.Value = sngResults - sngCalibresult
        Set rngCurrent = rngCurrent.Cells.Offset(1, 0)
     Loop
```

```
 rngMbcell.Value = rngMbcell.Cells.Offset(0, -3)

 ElseIf rngStart = Worksheets("OAT Sensitivity").Cells(3, 6) Then 'OK A_Perv
 Set rngCurrent = Worksheets("OAT Sensitivity").Cells(3, 6)
     Do While rngCurrent.Offset(0, -5).Value <> ""
        sngPercent = rngCurrent.Offset(0, -5).Value / sngDiv
        Set rngMbcell = Worksheets("Mass Balance").Cells(10, "AN")
        sngX = rngMbcell.Cells.Offset(0, -3).Value * sngPercent
        rngMbcell.Value = rngMbcell.Cells.Offset(0, -3).Value + sngX
        sngResults = Worksheets("Totals for Validation Year").Range("J17").Value
        rngCurrent.Value = sngResults - sngCalibresult
        Set rngCurrent = rngCurrent.Cells.Offset(1, 0)
     Loop
 rngMbcell.Value = rngMbcell.Cells.Offset(0, -3)

ElseIf rngStart = Worksheets("OAT Sensitivity").Cells(3, 7) Then 'OK A_DCImp
 Set rngCurrent = Worksheets("OAT Sensitivity").Cells(3, 7)
     Do While rngCurrent.Offset(0, -6).Value <> ""
        sngPercent = rngCurrent.Offset(0, -6).Value / sngDiv
        Set rngMbcell = Worksheets("Mass Balance").Cells(11, "AN")
        sngX = rngMbcell.Cells.Offset(0, -3).Value * sngPercent
        rngMbcell.Value = rngMbcell.Cells.Offset(0, -3).Value + sngX
        sngResults = Worksheets("Totals for Validation Year").Range("J17").Value
        rngCurrent.Value = sngResults - sngCalibresult
        Set rngCurrent = rngCurrent.Cells.Offset(1, 0)
     Loop
rngMbcell.Value = "=0.75 *(AN9)"


 ElseIf rngStart = Worksheets("OAT Sensitivity").Cells(3, 8) Then 'OK A_IDCImp
   Set rngCurrent = Worksheets("OAT Sensitivity").Cells(3, 8)
     Do While rngCurrent.Offset(0, -7).Value <> ""
        sngPercent = rngCurrent.Offset(0, -7).Value / sngDiv
        Set rngMbcell = Worksheets("Mass Balance").Cells(12, "AN")
        sngX = rngMbcell.Cells.Offset(0, -3).Value * sngPercent
        rngMbcell.Value = rngMbcell.Cells.Offset(0, -3).Value + sngX
        sngResults = Worksheets("Totals for Validation Year").Range("J17").Value
        rngCurrent.Value = sngResults - sngCalibresult
        Set rngCurrent = rngCurrent.Cells.Offset(1, 0)
    Loop
 rngMbcell.Value = "=.25*AN9"

 ElseIf rngStart = Worksheets("OAT Sensitivity").Cells(3, 9) Then 'OK  FImpADC
   Set rngCurrent = Worksheets("OAT Sensitivity").Cells(3, 9)
     Do While rngCurrent.Offset(0, -8).Value <> ""
```

```
        sngPercent = rngCurrent.Offset(0, -8).Value / sngDiv
        Set rngMbcell = Worksheets("Mass Balance").Cells(14, "AN")
        sngX = rngMbcell.Cells.Offset(0, -3).Value * sngPercent
        rngMbcell.Value = rngMbcell.Cells.Offset(0, -3).Value + sngX
        sngResults = Worksheets("Totals for Validation Year").Range("J17").Value
        rngCurrent.Value = sngResults - sngCalibresult
        Set rngCurrent = rngCurrent.Cells.Offset(1, 0)
    Loop
rngMbcell.Value = "=AN11/AN8"

ElseIf rngStart = Worksheets("OAT Sensitivity").Cells(3, 10) Then 'OK FImpAIDC
  Set rngCurrent = Worksheets("OAT Sensitivity").Cells(3, 10)
    Do While rngCurrent.Offset(0, -9).Value <> ""
       sngPercent = rngCurrent.Offset(0, -9).Value / sngDiv
        Set rngMbcell = Worksheets("Mass Balance").Cells(15, "AN")
        sngX = rngMbcell.Cells.Offset(0, -3).Value * sngPercent
        rngMbcell.Value = rngMbcell.Cells.Offset(0, -3).Value + sngX
        sngResults = Worksheets("Totals for Validation Year").Range("J17").Value
        rngCurrent.Value = sngResults - sngCalibresult
        Set rngCurrent = rngCurrent.Cells.Offset(1, 0)
    Loop
rngMbcell.Value = "=AN12/AN8"


ElseIf rngStart = Worksheets("OAT Sensitivity").Cells(3, 11) Then 'OK CN
 Set rngCurrent = Worksheets("OAT Sensitivity").Cells(3, 11)
    Do While rngCurrent.Offset(0, -10).Value <> ""
        sngPercent = rngCurrent.Offset(0, -10).Value / sngDiv
        Set rngMbcell = Worksheets("Mass Balance").Cells(16, "AN")
        If sngPercent = -1 Then
       sngPercent = -0.0009999
        Else
        sngX = rngMbcell.Cells.Offset(0, -3).Value * sngPercent
        rngMbcell.Value = rngMbcell.Cells.Offset(0, -3).Value + sngX
        sngResults = Worksheets("Totals for Validation Year").Range("J17").Value
        rngCurrent.Value = sngResults - sngCalibresult
        End If
        Set rngCurrent = rngCurrent.Cells.Offset(1, 0)


    Loop
rngMbcell.Value = "=AK16"

ElseIf rngStart = Worksheets("OAT Sensitivity").Cells(3, 12) Then 'OK IAPerv_Inf
  Set rngCurrent = Worksheets("OAT Sensitivity").Cells(3, 12)
    Do While rngCurrent.Offset(0, -11).Value <> ""
```

```
            sngPercent = rngCurrent.Offset(0, -11).Value / sngDiv
            Set rngMbcell = Worksheets("Mass Balance").Cells(18, "AN")
            sngX = rngMbcell.Cells.Offset(0, -3).Value * sngPercent
            rngMbcell.Value = rngMbcell.Cells.Offset(0, -3).Value + sngX
            sngResults = Worksheets("Totals for Validation Year").Range("J17").Value
            rngCurrent.Value = sngResults - sngCalibresult
            Set rngCurrent = rngCurrent.Cells.Offset(1, 0)
        Loop
    rngMbcell.Value = "AK18"
  End If

Next i



End Sub

Sub Testcomp()

Dim rng As Range

Set rng = Cells(3, 11)
rng.Value = Worksheets("Mass Balance").Range("BE4").End(xlDown).Row

End Sub
```