uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Pascal Dumont
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Earth Sciences)
GRADE / DEGREE

Department of Earth Sciences
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Error propogation calculation in groundwater vulnerability models

TITRE DE LA THÈSE / TITLE OF THESIS

M. Robin
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

F. Michel

M. Sawada

Gary W. Slater
LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES /
DEAN OF THE FACULTY OF GRADUATE AND POSTDOCORAL STUDIES

Error propagation calculation in groundwater vulnerability models

Pascal Dumont

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies of the University of Ottawa
In partial fulfillment of the requirements
For the MSc degree in Earth Sciences

Ottawa-Carleton Geoscience Centre
and
University of Ottawa

Your file Votre référence
ISBN: 978-0-494-18134-8
Our file Notre référence
ISBN: 978-0-494-18134-8

AKNOWLEDGMENT

I would like to thank all those who made this thesis possible and an enjoyable experience for me.

First of all I wish to express my sincere gratitude to Michel J.L. Robin, who guided this work and helped me whenever I was in need. His endless patience, availability and comradeship contributed tremendously to the successful completion of my thesis.

I am also indebted to Michel J.L. Robin and Bahram Daneshfar (University of Ottawa), Alfred Jaouich and Gilbert Prichonnet (Université du Québec à Montréal) for the great teaching in advancing my knowledge to complete the master degree.

I am grateful to the thesis examiners, Fred Michel (University of Carleton) and Michel Sawada (University of Ottawa) for their great and extensive review of my thesis and in providing advice for improvement.

Finally, I would like to express my deepest gratitude for the constant support, patience, understanding and love that I received from my wife Nathalie and my children Marie-Lou & Léonard during the past years.

TABLE OF CONTENTS

1. INTRODUCTION

"The search for models of uncertainty must start with an understanding of how geographies are abstracted and geographical data are modelled." (Goodchild and Zhang, 2002).

"Uncertainty may be defined as a measure of the difference between the data and the meaning attached to the data by the current user." (Goodchild and Zhang, 2002).

"More input information must be known; more output information must be sorted through; and the model becomes more difficult to use and interpret. An ironical paradox arises: the more realistic has less credibility. The less the users knows of the subtleties involved, the more suspicious he/she becomes of the models as their difficulty increases" (Wauchope, 1992)

Introduction

In the past decade new social policies and public perception about sustainable development has increased the need to produce maps of environmentally sensitive parameters based on environmental modelling results. For instance, the principal target of the Millenium Development Goals (MDGs) initiative established by the United Nations stresses the need *to ensure environmental sustainability by halving the proportion of people without access to safe water by 2015* – where safe water is defined as a water supply that is devoid of disease-producing pathogenic bacteria/viruses and highly toxic substances (WHO/UNICEF 2000, Howard et al. 2001).

Two aspects to groundwater vulnerability can be considered: quantity and quality. Vulnerability of the supply (quantity) is usually addressed by water budgeting exercises such as those currently underway in a number of watersheds in Ontario as part of the Source Water Protection Planning Program (MOE TEC Report, 2004). The vulnerability of groundwater to the deterioration its quality can be characterized by a number of models that fall into two broad families: those that are based on a physical parameter and those that are based on an index system. While the issues of quantity and the interaction between quantity and quality are certainly important, they are beyond the scope of this study (only a brief description of the annual water budget will be given for the study area in the next section). We will instead focus on the issue of quality, and more specifically the vulnerability of aquifers to contamination. Aquifer vulnerability to contamination is usually expressed by some measure of the hydrogeologic system's ability to offer protection against potential contaminant sources (CH2M HILL, 2001).

This study investigates the propagation of uncertainty in aquifer vulnerability mapping. We applied three vulnerability models (defined shortly) to the aquifers of Eastern Ontario, Canada, where a considerable amount of information is already available as a result of the Eastern Ontario Water Resources Management Study (EOWRMS). Municipality and county administrations within this region in partnership with the Government of Ontario sponsored the EOWRMS in recognition of the need to develop a regional water resources information system on a watershed basis. The EOWRMS and other similar studies in Ontario produce (amongst many other things) aquifer vulnerability maps that are used by municipal planners without the benefit of an estimate of the uncertainty associated with the information. The purpose of this study is to develop a methodology that will examine the uncertainty in the vulnerability assessment.

An outcome of the present study is therefore to improve the drinking water risk management process for public health professionals and municipal planners by providing a means of assessing uncertainty in the vulnerability analysis. This is in line with the very important core challenge of improving the capability of small, rural and remote communities to assure clean and safe drinking water.

Objective and scope of the thesis

The scope of this thesis can be summarized by three (3) cornerstones:

1. Apply three (3) GW vulnerability models (ISI, SAAT & DRASTIC) to the aquifers of a portion of Eastern Ontario with data (GIS layers) produced by the Eastern Ontario Water Resources Management Study. Those GIS layers are the components of their Conceptualized Regional Aquifer System;
2. Highlight the data components' sources of uncertainty in the physical parameters associated with measurement and determine the variance (uncertainty) associated with each layer of data used in the three (3) GW vulnerability models.
3. Once the uncertainty values associated with the parameters are calculated, the final step is to identify and apply methodologies to track the propagation of uncertainty through each model

In GIS operations, typically, a number of map layers are combined according to the algebraic rules of the model to produce the final output map. The output map can appear to be extremely precise, giving the user a false sense of security about the reliability of the information it contains. In fact, the sources of uncertainty and error involved in the process are multiple.

First, there is measurement error at the specific points that are used to produce the basic layers (they can be of both type: attribute value and location value);

then the measured points are interpolated spatially to produce a map layer, and there is error produced by the interpolation process accordingly to the spatial variability structure and the estimation method used; Next, the layers are combined according to the model, which is a simplified representation of reality that introduces additional uncertainty which can be measurable through validation processes. And finally, there is the process of propagating the error associated with each layer through the model: additional uncertainty could arise from this process.. This study examines measurement error/uncertainty in three groundwater vulnerability models, DRASTIC, SAAT and ISI, highlighting their sources of uncertainty in the physical parameters (attribute values) but without considering geographical uncertainty associated with location value or spatial interpolation. The study focuses on the methodology used to track the propagation of uncertainty through each model rather than on the evaluation of the uncertainty of the components that enter into the model. Using similar error for the components the vulnerability model results are compared along with their uncertainty. The methodology is aimed at improving decision-making by producing vulnerability maps and vulnerability uncertainty maps, which may be used to evaluate the need for additional information in the risk assessment process.

The first step is to identify the sources of uncertainty associated with the layers that are used in the vulnerability calculations. The principal component is uncertainty associated with measured point estimates ("measurement error", linked to the process of abstraction or generalisation about real world phenomena). These uncertainty components are combined to give uncertainty maps, expressed as variance, for each input parameter that enters in the vulnerability calculations. The next step is to propagate the uncertainty of the input parameters through the vulnerability calculations using the Error Propagation Theorem (used in standard statistical applications). The result is a map of uncertainty, expressed as variance and coefficient of variation, associated with the vulnerability map.

While an uncertainty map is certainly valuable information from a scientific standpoint, it is difficult to use for planning purposes. Consequently, the final step for the process consists of producing vulnerability maps that are classified and that are accompanied with a map showing areas where classifications could change as a result of uncertainty, for a given confidence level. From the planner's perspective, a decision can then be made to make further field measurements in areas of high uncertainty depending on the risk assessment. In other words, the contribution of this thesis is the development of a simple methodology for private consultants and other users aimed to evaluate and include uncertainty in GW vulnerability modeling and the development of a practical application of uncertainty in decision-making.

The object of the thesis was not to develop the geomatics of error propagation (which would be speculative anyway because of the lack of hard data), but rather to develop (or try to) a simple means by which consultants can calculate and display some estimate of errors and more importantly to devise a way in which the end user (with little geomatics background) can use this information in a meaningful and practical way.

Data source descriptions and limitations

GW vulnerability modeling and uncertainty propagation calculation results issued from this thesis work are strongly influenced by the conceptual limitations related to the data. Models and associated uncertainties were calculated with datasets (GIS layers) already generalized/conceptualized by the EOWRMS and surface geology description were taken from the digital database provided along with the EOWRMS report.

The EOWRMS' regional aquifer system was conceptualized as :

- An upper overburden layer representing all surface deposits;
- a middle semi continuous impermeable layer; and
- a lower overburden aquifer referred to as the contact zone aquifer (consisting of higher conductivity materials, including basal sands and gravels and the fractured portion of the upper bedrock).

The EOWRMS conceptualized regional aquifer system is principally issued from water well drillers' logs (Ontario Ministry of the Environment Water Well Records). These records (point data) vary in quality and descriptive terminology and their quality remain generally poor. Less than 10% of the 90,000 were retained to model the conceptual system (and those data were rarely validated by geologist) Informed guesses from these records, with the help of geological reports & maps, produced a two-dimensional interpolation of point data to represents and models a three-dimensianal hydrogeologic system.

Without the presence of reliable and sufficient data it was not possible to achieve correct, efficient, reliable, and robust estimation of the spatial variability of variables (meaning the auto-correlation) and to evaluate the cross-correlation between variables:

- No model of spatial variation has been properly defined in the representation of EOWRMS' conceptualized data and then, spatial autocorrelation & cross-correlation were not taken into consideration in the calculation of error propagation.
- The covariance term has been dropped from the first order Taylor expression (technique of error propagation) because of lack of real data.

At the scale of this work, many of the variables of concern have:

a) Small scale (large area) regionalized trends (e.g. Regional geological formation): Which were treated as a 2nd order stationartity

b) Large scale (small area) variations that are self correlated & cross-correlated to other parameters: which were not considered because the correlation scale (when theoretically known) is less than the pixel size or is simply unknown.

We have to keep in mind that the analysis of spatial variation of a continuous field value (for instance hydraulic conductivity) should be done prior to any spatial interpolation processes.

A good geostatistical analysis is not only based on the observed data, but also on our scientific knowledge of the landscape under consideration in order to establish a physical understanding/framework where basic concepts issued from spatial analysis must be provided. Spatial dependency is a key concept for understanding and analyzing a spatial phenomenon: most of the occurrences, natural or social, present among themselves a relationship that depends on distance.

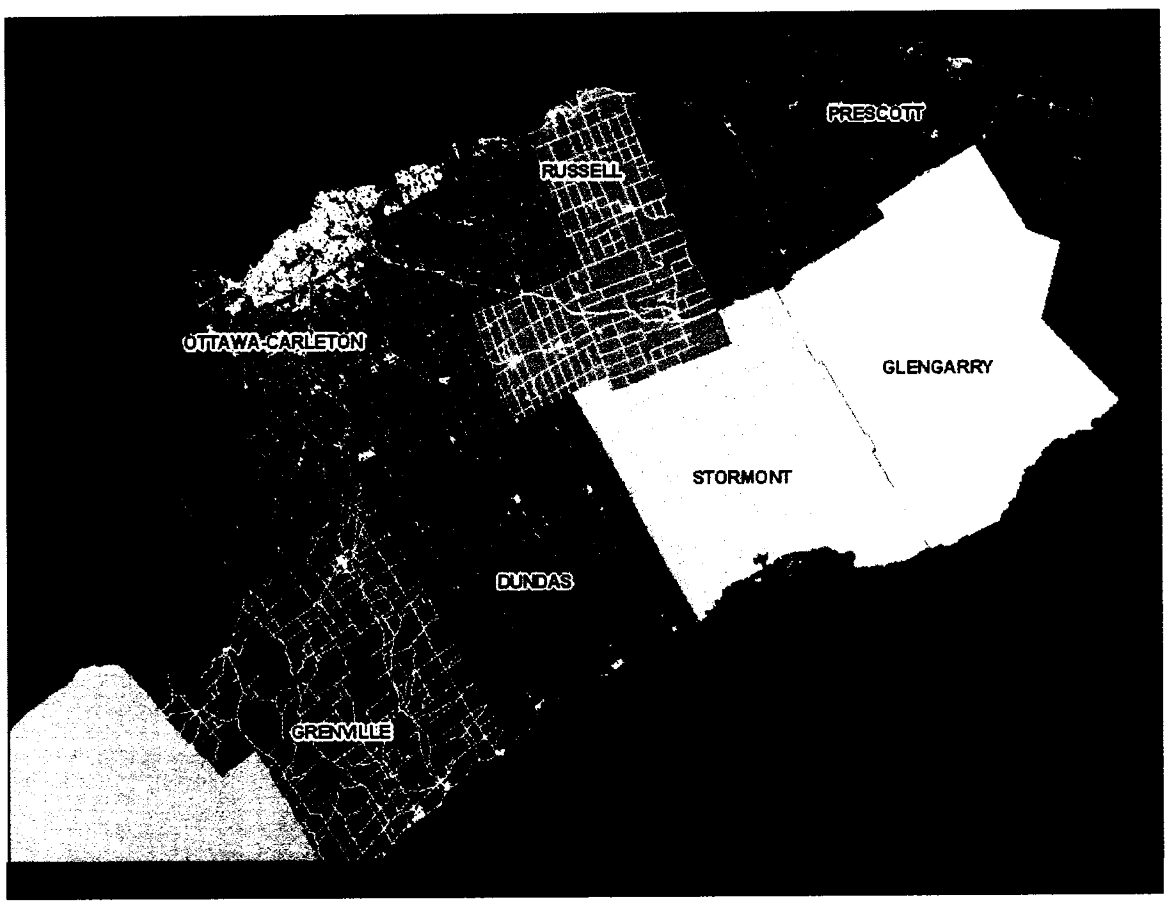
The computational expression of the concept of spatial dependence is the spatial autocorrelation (which is the measurement of the correlation done with the same variable, measured in different places in space). Different indicators (Moran's index, variogram) exist to measure the spatial auto-correlation, all of them based on the same idea: verifying how the spatial dependency varies by comparing the values of a variable in space. If observed autocorrelation reveals significant directional variation in spatial dependence then, It leads to a loss of explanatory power. This reflects on higher variances for the estimates

The main statistical concepts that define the spatial structure are the stationarity & isotropy. When the mean of a given variable is constant in space we have a stationarity of 1st order and when the autocovariance is constant in space we have a stationarity of 2nd order, meaning that both are constant in the whole region under study (no trend). A process is considered isotropic if the autocovariance depends only on the distance between the points and not on the direction between them.
"PHYSICAL LAWS ARE INVARIANT UNDER TRANSLATION AND ROTATION", this statement from Boogart summarized perfectly the concepts of stationarity and isotropy which must be assumed in order to estimate values at unknown points based on data observed.

Kriging is one of the possible estimation method: at each unknown point, a value of the random variable is estimated, using an estimator, which is a function of the data observed and of the spatial covariance structure obtained through the estimation of the variogram.
Once we have recognized that we have access to GIS layers without any knowledge on the spatial variability of the data, which way can we go to evaluate the (error) variance of each component (GIS layers) and propagate it through each GW vulnerability model ? A possible approach is to evaluate the variance of each GIS layers with the "best educated guesses" and to use the "first order Taylor method" as the error propagation equation.
The best educated guess is the estimation of a reasonable value of relative error to back calculate the variances for each GIS layers. Without any data correlation values available, the uncertainty of the GW vulnerability values will simply be the sum of the variances of the parameters.

Location of study

This study investigates a portion of Eastern Ontario including the United Counties of Prescott and Russell (P&R), the United Counties of Stormont, Dundas and Glengarry (SD&G) and a part of the City of Ottawa that falls within the South Nation River watershed. Administrations of these municipalities and counties initiated the EOWRMS (the Eastern Ontario Water Resources Management Study) in recognition of the need to develop a regional water resources information system on a watershed basis. The study area includes the South Nation River and Raisin Region watersheds and other minor adjacent watersheds; and covers approximately 6,800 square kilometres (CH2M HILL, 2001).



Regional water budget

One of the key components of the vulnerability model is the amount of recharge received. This can be estimated from a water budget, and a short review of the EOWRMS water budget calculation is presented here as background information.

A regional water budget provides general estimates of the quantity of water cycling through a study area. Most water budgets assume steady-state, meaning that over the period considered, there is no net gain or loss of water storage in the region. The budget exercise then becomes a matter of balancing three input and output components over the region: the annual precipitation, obtained by direct measurements; the annual evapotranspiration (ET), obtained by ET modelling and direct measurements; and the Net Available Water (NAW), which is simply the difference between the precipitation and the evapotranspiration. (CH2M HILL, 2001). The NAW is further divided into a surface water runoff component and a groundwater infiltration component, according to a partitioning model.

Our study area, receives an average of about 930 mm of precipitation annually. The regional water budget model calculates that approximately 420 mm is returned to the atmosphere through evapotranspiration and 510 mm of water is partitioned between the surface water drainage network (94 percent) and the deep groundwater reserves (6 percent) (CH2M HILL, 2001). Further details and information on data sources; limitations; assumptions; approach and methods regarding the calculation and modelling of the Eastern Ontario regional water budget can be found in the EOWRMS final report (CH2M HILL, 2001).

Based on the EOWRMS water budget, the NAW integrates over the region to 35 billion cubic meters of water annually, which is amply sufficient to sustain current demand for domestic, industrial and institutional uses as well as aquatic and terrestrial habitat. However, this water budget does not consider seasonal flow variability of precipitation and evapotranspiration, water quality status, aquifer transmissivity, etc., which leads to cyclic declines in the availability and quality of water resources. Therefore, we have to keep in mind that even if the regional water budget is annually (globally) positive, surface water resources and groundwater resources will be limited and more at risk during certain periods of the year, generally those associated with a lack of precipitation.

The amount of groundwater recharge is driven by seasonal fluctuations of precipitation. Further, as stated by Bélanger (2001), "Recharge occurs in the spring, beginning with the melting of the snow, and gradually decreases during May and June. Potential evapotranspiration exceeds the rainfall during approximately four months of the year, corresponding to the warmer season. Groundwater levels generally reach their lowest levels at the end of August. During autumn, there is some recharge before deep frost penetration occurs. Seasonal variations in groundwater levels can affect the supply of water for domestic use in shallow wells, but generally have no effect in deeper wells drilled into bedrock".

Geological stratigraphic units and aquifers.

Aquifer vulnerability is largely governed by the geological materials through which water has to travel to reach the aquifer. It is important, therefore, to have a good understanding of the geology of the region. The geological formations in Eastern Ontario represent three distinct geological times, which are different in terms of age and time span but also in the nature of the geological processes (events) involved and the type of materials produced. Three major elements constitute the geological setting for Eastern Ontario:

- Precambrian metamorphic and igneous rocks of the Canadian Shield;
- Cambrian and Ordovician sedimentary rocks of the continental platform;
- Quaternary unconsolidated sediments comprising the surface materials.

Except where noted, the following description is paraphrased from Bélanger (2001) and Scott & Aitken (1993).

Precambrian:

Precambrian rocks underlie the entire region and outcrop as the Laurentian Highlands, north of the Ottawa River, and as highlands to the southwest of the study region where they form a geological structure known as the Frontenac Arch. The geologic setting of Precambrian rocks is extremely complex with numerous faults, folds, and rapid successions of rock types.
As stated by Bélanger (2001), "Sound Precambrian rocks are relatively impermeable to groundwater and movement of water is restricted to cracks, joints and fractures, or solution channels in metasedimentary rocks like carbonate sedimentary rocks (limestone) transformed to marble or lime silicate rocks". The groundwater extraction regime of Precambrian aquifers is generally limited and depends on the recharge capacity of the overlying Quaternary deposits. The quality is highly variable and depends on the mineral content which increases with depth.

Paleozoic: Late Cambrian and Early Ordovician sedimentary rocks.

Eastern Ontario is part of the central St-Laurence continental shelf (platform) originating from the Late Cambrian and Early Ordovician marine transgression of the proto-Atlantic ocean called Iapetus following the Pangea split. At that time the Eastern Ontario area was located within a bay, called the Ottawa Embayment that extended into the continental mass (Canadian Shield), bounded by the Frontenac Arch and Adirondack mountains.

In terms of the first Paleozoic event and associated terrigenous sedimentary rocks, conglomerates and sandstones were formed from the erosion of Precambrian quartzites and conglomerates (Covey Hill and Nepean Formation). They correspond to environments like alluvial fan and braided fluvial deposits.

As the marine transgression increased, the fining-up of the sedimentation gave the *March Formation* (interbedded sandstones and dolostones with boulder-and-cobble-sized interclasts of quartzite) where the unit is in contact with Precambrian rocks followed by the clay-size sediments and carbonate precipitates (CaMg) of the *Oxford Formation* (dolostone) (Scott and Aitken, 1993).

The sandstone and conglomerate rocks of the Covey Hill and Nepean formations provide the best aquifers (regional aquifer) in the region (Bélanger, 2001)..The March and Oxford formations are relatively impermeable, consequently, water flows through joints at different depth giving several aquifers (Bélanger, 2001).

Paleozoic: Middle to late Ordovician sedimentary rocks.

With the Avalonian microcontinent approaching the Laurentian craton (convergent plates) the Appalachian orogeny occurred along the eastern margin of the Laurentian craton resulting in deformation and uplift of the underlying continental shelf, Late Cambrian and Lower Ordovician deposits. Therefore, an important regression of the sea started which resulted in subaerial exposure and erosion of a large part of the Early Ordovician rocks (erosonial discordance).

The erosional period was followed by a second period of marine transgression in the Ottawa Embayment that deposited sediments (the Rockliffe Formation which is graded from sandstone, with quartz-pebble conglomerate locally present, to shale and limestone with silty dolostone interbeds) derived from the surrounding uplifted Precambrian rocks (Bélanger, 2001).

A marine regression followed the *Rockliffe Formation* period which exposed the sediments to erosion, creating a stratigraphic break. This hiatus was followed by the third marine transgression which started a depositional sequence with the *Shadow Lake Formation* (nearshore sediments constituted of sandstone graded upward to sandy shale interlayered with carbonate rocks) and overlying formations reflecting

continuous deposition on a deepening shelf: shales, dolostones and limestones (*Gull River, Bobcaygeon, Verulam and Lindsay formations*).
Aquifers in these rocks are restricted to seams in the calcareous shale layers. The massive crystalline limestone layers generally contain little to no water (Bélanger, 2001).

Paleozoic: Late Ordovician sedimentary rocks.

This period is marked by a deep water sedimentation followed by a marine regression. The deep-water depositional environment of the late Ordovician gave the *Billings Formation* (shale with thin interbeds of dark grey limestone) and the transition from shallow to deep-water depositional environment is represented by the underlying *Eastview Formation* (calcareous nature). Due to the coarsening-up associated with marine regression (transition from deep to shallower depositional environment), the *Billings Formation* grades into the *Carlsbad Formation* (interbedded shale, fossiliferous calcareous siltstone, and silty bioclastic limestone) up to the Queenstone Formation (red to green-grey shales, siltstone and minor limestones) (Scott and Aitken, 1993).

From these formations, characterized by the predominance of fossiliferous and petroliferous shale and limestone and thick sections without water, the groundwater supply is generally sufficient for domestic use. (Bélanger, 2001).

Quaternary: Unconsolidated sediments and overburden aquifers

Eastern Ontario is generally covered by unconsolidated sediments that can reach 100 meters of thickness. These Quaternary deposits are mainly constituted of Champlain Sea marine sediment, and glacial and related materials from the *Late Wisconsinian* glaciation, and reworking of these sediments by subsequent geomorphic processes. Materials predating the Wisconsinan glaciation were mainly eroded and reworked by continental glaciers during the last glaciation. Older units predating the last glaciation occur in low-lying bedrock topography .

These glacial and marine deposits have a grain-size distribution ranging from granular materials which can act as important aquifers, to thick deposits of fine materials forming aquitards. Near the Precambrian shield, tills are sandy with high hydraulic conductivity but within the paleozoic formations of Eastern Ontario, clay content increase and tills are aquitards. The analysis and detailed mapping of stratigraphic sequences and associated geological events are necessary to fully understand the ways in which the various aquifers are connected and in order to determine local and regional patterns of groundwater flow.

The following descriptions of Quaternary deposits and associated aquifer potential are presented in stratigraphic order, from oldest to youngest. The Quaternary stratigraphy overall is not present everywhere, hence we have to identify local stratigraphic hiatus in order to complete a good assessment of a particular location.

Till and aquifers

This Quaternary deposit could be described as a compact layer of sand, silt and clay with some gravel (generally poorly sorted), overlain by a less compact, thus more permeable, layer of sand, gravel and boulders in some locations. The glacial till unconformably overlies the bedrock and has a structure that varies from relatively loose near the surface to very compact with depth. Thickness varies from 1 to more than 20 meters depending on the nature (Precambrian vs Paleozoic) and the relief of the bedrock.

The till is normally less permeable at depth as it becomes more compact and/or where clay content is more important. This is related to the origin of the till. The more compact and fine matrix till corresponds to lodgement till deposited under glacier, whereas the less compact and coarser till was deposited by melting of stagnant ice (Bélanger, 2001). Till is generally considered a good aquifer.

Glaciofluvial sediments and aquifers

Forming very permeable aquifers, glaciofluvial sediments were deposited by glacial melt-water during the last phase of glaciation, when glaciers were stagnant or retreating. Poorly to well sorted, they consist of coarse-to-medium-grained sand and gravel. They constitute very good recharge zones where they outcrop.

They can lie at the bottom and sides of valleys, occurring as eskers, kame terraces, and subaqueous outwash trains and fans deposited at the retreating ice margin in valleys and in low-lying areas in the rugged Precambrian terrain. Most glaciofluvial deposits in the lowlands have been buried by marine deposits or have been reworked by wave action of the Champlain Sea (Bélanger, 2001).

Glaciolacustrine sediments, marine offshore sediments and aquitard

Following glaciers retreating, fine grained sediments derived from nearby glacial sources were deposited in fresh water bodies which were rapidly affected by a marine invasion and top-dressed with associated marine offshore sediments. Glaciolacustrine and marine offshore sediments are both fine grained, clay and clayed silts, with thickness that can attain 100 meters. With very low to low permeability, they can form important aquitards between the underlying till aquifers and aquifers in overlying surficial deposits.

Marine nearshore sediments, Post-Champlain Sea deposits and surficial sand aquifers

Marine nearshore sediments are composed of 2 elements: (1) deposits near marine limits (beaches in topographic highs) and (2) reworking of existing sediment during the retreating of the sea (deltaic and estuarine sediments in lower levels). The first group of deposits reflect the nature of underlying materials. They consist of gravel, sand or boulders when the underlying materials are till and glaciofluvial deposits, and consist of slabs and shingles when sedimentary bedrock is the underlying material.
The second group of deposits consist of sheets and bars of fine-to-medium-grained sand issued from reworked glaciofluvial deposits or deltaic and estuarine deposits constituted of medium-to-fined-grained sand plains that were deposited at the mouth of rivers entering the sea and generally overlie fine-grained marine sediments (Bélanger, 2001).

Post-Champlain Sea deposits originate from marine and glacial deposits which were eroded and reworked by freshwater streams and lakes as the Champlain Sea retreated. Fluvial terraces and bars, which are the older deposits, consist of medium-grained stratified sand with some silt. Younger deposits are actually occupying floodplains and alluvial fans in areas of low relief where they consist of silty sand, silt, sand and clay. The Post-Champlain Sea deposits are in some locations stratified and include minor gravel, disseminated organic matter, marl, and peat bogs within the abandoned channels.

These surficial sand aquifers generally form good aquifer recharge zones. In the case of deltaic sandy deposits underlayed by marine clays and forming localized shallow flow systems, groundwater generally flows in the direction of topographic lows, discharges into the surface drainage system and provides little recharge to till or bedrock aquifers

Important definitions

Having to admit the presence of errors in the decision process presented to the public is probably the most difficult situation for natural resources managers. Part of the problem is the widely held perception that error or uncertainty is synonymous with "mistake". In fact, professionals dealing with natural resources and environmental assessments have to recognize that natural processes are inherently variable and instead of being a negative aspect of the data, uncertainty is an appreciation in the value of the data. It is important to recognise that all measurements have associated error (in the sense of uncertainty) and that it is not sufficient to specify the value of the measurement, one must also assess its quality by specifying its uncertainty.

To avoid uncertain and conflicting definitions among multiple disciplines and help alleviate negative connotations, we need to agree on consistent definitions for the terms we use to describe our work. The terms accuracy; precision; bias; error; resolution; scale; uncertainty and risk assessment were perfectly described or gathered by H.T. Mower (1999) and M. Goodchild and Zhang (2002). In addition to them, definitions of intrinsic and specific groundwater vulnerability from Brouyère and al. (2003) will be discussed in this section.

Accuracy and precision

While the term "accuracy" refers to "the closeness of computation or estimates to the exact or true values" (Marriot, 1990), precision measures the degree of conformity of measurements among themselves. It is a measure of how repeated observations conform to themselves. Thus, high precision may not guarantee good accuracy if measurements are biased (Goodchild and Zhang, 2002).

Numerical precision is a different matter. It is the exactness or degree of detail with which we measure, or the detail in recording the observed properties (Laurini and Thompson, 1992). Poor numerical precision may give the impression of high statistical precision and improved numerical precision might yield values which indicate a measurable lack of statistical precision (Mowrer, 1999).

Accuracy assessment is applied to discrete cases as well as for continuous variables. The classification of pixels into unique classes in remote sensing images (discrete case) involved the concept of classification accuracy with an associated error matrix which can be used to determine "producer's accuracy" and "user's accuracy" (Fenstermaker, 1994). For continuous variables (for example, hydraulic conductivity), accuracy can be described statistically as mean-squared error: the sum of the variance and a squared bias term (Mood et al., 1974).

Measurements of accuracy include components of statistical precision and of bias or systematic error, and require a "true" value of the quantity intended to be observed to correctly measure this bias. In many cases involving natural resource and environmental phenomena, the "true" state of nature is not known (Mowrer, 1999).

For these kinds of data which are derived from measurements and expressed on continuous scales, the error (ε) is defined as the deviation between the measured or computed data value (g) and the true value or the value accepted as being true (g_0):

$$\varepsilon = g_0 - g$$

This may, without loss of clarity, be re-written as the deviation between a measured value from the reference by placing a negative sign on both sides (Goodchild and Zhang, 2002).

Bias and random error

With regularity, bias deforms the representativeness of the measurement as opposed to random error which evens out on the average. If bias value is known accurately it could be removed from the measurement to correct this bias. However, we may obtain (random) variation in the results each time we make a given measurement. This is sample variance, or lack of precision, associated with the estimated measurement. For some authors, the measurement of accuracy include components of statistical precision and of bias or systematic error, and require a "true" value of the quantity intended to be observed to correctly measure this bias (Mowrer, 1999); while for others, random error is a measure of precision, and bias (or lack thereof) is a measure of accuracy.

Resolution and scale

All map objects have a limited spatial resolution. National Map Accuracy Standards (USA) state that not more than 10% of points tested may be in error more than 0.8 mm for map scales larger than 1:20,000, and 0.5 mm for smaller map scales (Wolk and Brinker, 1994). It is important to understand from this last affirmation that computing often permits a level of numerical precision that is much higher than the inherent accuracy with which geographical phenomena are described and geographical data are acquired (Goodchild and Zhang, 2002).

Uncertainty issued from map resolution (accuracy of location and shape of features that can be depicted for a given map scale) is certainly valuable information to put on a map or GIS product. Quattrochi and Goodchild (1997) suggest that metric scale should be augmented, or perhaps replaced by statements of positional accuracy and content giving information about the smallest observable object, or limiting spatial resolution. When consulting a map or using a GIS product, we have to keep in mind that the effect of generalisation is to introduce uncertainty into the representation of a real phenomenon that could only be mapped perfectly at a much larger scale (Quattrochi and Goodchild, 1997). Goodchild (1996) concludes that metric scale alone is not sufficient as a specification parameter for digital databases.

Uncertainty and risk.

The issue of spatial data uncertainty for decision-making is linked with the assessment of risk. The latent uncertainty inherent in spatial databases is important if it would cause a given decision to be changed or evaluated again (affect the decision process).

In the classical definitions of risk and uncertainty, these notions are totally different matters whereas uncertainty cannot be measured or quantified, while risk can: uncertainty occurs when outcomes can be enumerated, but the associated probabilities cannot be estimated, while risk incorporates explicit estimation of probabilities or likelihoods (Knight, 1921).

In a different context, if the uncertainty is seen as a manner to evaluate different levels of predictability of measures of risk, i.e., the probability of the probability, uncertainty is an inherent part of risk, but remains value-neutral (Cleaves and Haynes, 1996; Finkel, 1990). In other words, when a specific outcome cannot be predicted, but the possible outcomes can be defined and their probabilities estimated, the risk can be quantified. Measures of probability ("uncertainty") are then part of the "environmental risk assessment "process.

The evaluation of the calculated uncertainty for environmental modeling, and the associated environmental risk assessment, is confronted with the scale effect. An evaluation done at a local scale instead of a regional one often only takes into consideration the interaction of local parameters (different stressors acting individually and in combination). These could be impacted by regional exogenous parameters (such as climate change) that provide the contextual information necessary to assess long term cumulative impacts to the environment.

The interactions across scales are generally not simple and may reflect nonlinear, multivariate, and discontinuous processes which, in extreme circumstances, may result in irreversible bifurcation between given results (Moir and Mowrer, 1993).

Intrinsic and specific groundwater vulnerability

The development and application of the concept of groundwater vulnerability assessment date back to the seventies (Albinet & Margat, 1970). Since thenseveral methods have been developed in order to model and map the vulnerability of aquifers, the better known being certainly the DRASTIC method (Aller et al., 1987), but there is no generally recognized and accepted functional definition of aquifer vulnerability. However, there are a number of heuristic definitions, such as the definition of Aller et al., (1987) "[Aquifer] Vulnerability is an intrinsic property of a groundwater system that depends on the sensitivity of that system to human and/or natural impacts."; or the definition of Tripet et al. (1997), in which the sensitivity (or vulnerability) of a groundwater source to human impacts depends on the "purification capacity" of the aquifer and on the travel time of groundwater. Definitions such as these have been used in the development of semi-qualitative index-based approaches to assessing vulnerability (such as DRASTIC and ISI).

More quantitative approaches that are gaining acceptance distinguish between *Intrinsic* and *Specific* vulnerability. In general terms Intrinsic vulnerability describes the transport properties of the aquifer irrespective of the contaminant; whereas Specific vulnerability considers the transport properties of specific contaminants.

(*Intrinsic*)

For Brouyère et al. (2001)[i], *Intrinsic vulnerability* defines the vulnerability of groundwater to contaminants generated by human activities by taking into account the inherent geological, hydrological and hydrogeological characteristics of an area, and as just mentioned, without respect to the nature of contaminants. This approach results in a physically-based model that calculates transport travel time and delineates travel paths of the contaminants from ground surface to groundwater resource. Brouyère et al. (2001) propose a method (Intrinsic Vulnerability Cube) using a physically based approach to evaluate and validate vulnerability concepts where the influence of main intrinsic hydrodispersive mechanisms affecting the spatial and temporal evolution of a contaminant in the underground have been examined. Daly et al. (2002) directly use the Vulnerability Cube as a vulnerability assessment tool (model). The method is complex and has not gained wide acceptance.

A more recent approach proposed by the Ministry of the Environment of Ontario, the Surface to Aquifer Advective Time (SAAT) (MOE TEC Report, 2004) is to ignore the dispersive component and to simply express intrinsic vulnerability as advective times of travel and advective paths. The rationale is that although dispersivity is a property of the aquifer material its impact is relevant only when considering the specific contaminant. There is also a more restrictive definition of intrinsic vulnerability used in the ISI method which simply uses the aquifer's ability to transmit water without considering the hydraulic gradient (unlike the previous definition).

In all cases the definition of Intrinsic vulnerability is meant to provide a generic evaluation of the vulnerability of the aquifer material itself in its usual flow system, hence the term "intrinsic".

(*Specific*)

Specific vulnerability defines the vulnerability of groundwater to a particular contaminant or group of contaminants. Specific vulnerability modeling of groundwater takes into account the sorption properties of chemical contaminant and the dispersive behaviour of the system. Some chemicals have a tendency to adsorb

---

[i] Within the scope of the COST 620 Action "Vulnerability mapping for the protection of carbonate (karst) aquifer"

and desorb on and off sediments (sorbing contaminants). Strongly sorbing compounds will be soil or sediment pollutants and will tend to remain near surface in soil zones without producing extensive groundwater contamination. Nonsorbing compounds will move with water (advective transport). Low to medium sorbing compounds will move with water but at a slower velocity (retardation). A retardation factor must be estimated for sorbing contaminants.

Nonaqueous Phase Liquids (NAPLs) are fluids that are immiscible in water that will move as pure liquids through the subsurface. Light NAPLS (LNAPLs), have densities less than water and, will move downward through the unsaturated zone leaving behind small amounts of NAPL in the pore spaces and accumulate at the water table. LNAPLs will effectively "float" atop the water table. LNAPLs include contaminants like gasoline, heating oil, diesel fuel, Jet fuel, etc. Dense NAPLs (DNAPLs) on the other hand are denser than water and will keep on migrating downward past the water table into the saturated zone until an impermeable layer is encountered. Examples of DNAPLs include industrial solvents used as degreasers, TCE, PCE, etc. Both LNAPLs and DNAPLs will leave a residual phase of pure NAPL in their wake that will dissolve slightly in water for extremely long times and produce contamination plumes downgradient. NAPLs are extremely toxic even at concentrations lower than the solubility limit.

Real life cases make it difficult to apply these specific and intrinsic groundwater vulnerability models. For example heterogeneous conditions which occur when aquifers are made of many differing geologic materials (more the rule than the exception) and hydraulic gradients that may not be constant particularly when wells are pumping in the aquifer.

Lobo-Ferreira (1999) makes an important distinction between the concepts of *vulnerability* and *pollution risk*. Pollution risk depends not only on vulnerability but also on the existence of significant pollutant loading entering the subsurface environment. It is possible to have high aquifer vulnerability but no risk of pollution if there is no significant pollutant loading, and to have high pollution risk in spite of low vulnerability if the pollutant loading is exceptional. It is important to make clear the distinction between vulnerability and risk. This is because risk of pollution is determined not only by the intrinsic characteristics of the aquifer (relatively static and hardly changeable), but also on the existence of potentially polluting activities which are dynamic factors which can in principle be changed and controlled. Risk also must include a measure of the potential impact; i.e., is the contamination potentially lethal or is it simply an aesthetic concern?

Finally, to prevent any misinterpretation concerning uncertainty assessment, lets keep in mind the following statement of Goovaerts (1996), "...uncertainty is not intrinsic to the phenomenon under study: rather, it arises from our imperfect knowledge of that phenomenon, it is data-dependent and most importantly model-dependant...No model, hence no uncertainty measure, can ever be objective: the point is to accept the limitation and document clearly all aspects of the model."

Research topics and scope.

The first chapter of this thesis has introduced concepts related to the physical setting of the studied region (location of study; regional water budget; geological stratigraphic units setting and aquifers). It was followed by a review of important definitions in order to eliminate any uncertainties concerning the direction of the used concepts in this study. As mentioned earlier, the study develops a methodology for the calculation of the error and propagation of the error in groundwater vulnerability modelling with the assistance of a Geographic Information System (GIS).

The second chapter discusses the generalization of geographical realities with a GIS. It underlines the ambiguity, inexactness, and vagueness existing in many geodatabases, reflecting the inherent complexity of the real world and the subjectivity involved in a large number of GIS databases. After a brief description of the two models for geographical abstraction (fields and objects) an overview will follow of how uncertainty occurs in the field model through spatial interpolation (kriging and semivariogram concepts). This overview is followed by: (1) a discussion of how important uncertainty assessment is in the decision process; (2) a description of sources of uncertainty and of calculation of uncertainty in spatial data variables and, (3) a brief examination of the theory and available techniques of error propagation with local GIS operations.

Chapter 3 is a general overview of concepts associated with groundwater vulnerability modelling and various existing models. A full description is given for the three models chosen for this study (ISI, DRASTIC and SAAT) and other empirical method (GOD, EVARISK, MINESOTA, SINTACS, and VULNERABILITY CUBE). This retrospective is followed by a discussion of existing comparison studies between these methods and of prospects for further developments.

Chapter 4 serves to set the methodologies for the calculation of uncertainty propagation for each chosen model (ISI, DRASTIC, SAAT). It describes the methodologies used to track the propagation of these uncertainties through each model and compares the final results to improve decision-making.

Chapters 5 & 6 discuss and analyze the results and give some retrospective and prospective views. More work on characterization, description and modelling heterogeneity of geology will help to develop more physically-based quantitative approaches to the concept of groundwater vulnerability.

2. SOURCES, CALCULATION AND PROPAGATION OF UNCERTAINTY THROUGH ENVIRONMENTAL MODELS

Groundwater vulnerability mapping is carried out in a GIS (Geographic Information System) using the tools of cartographic algebra to carry out calculations that vary in complexity from the simple index-based models to the more complex physically-based models (Heuvelink, 1998). Typically, a number of map layers are combined according to the algebraic rules of the model to produce the final output map. The output map can appear to be extremely precise, giving the user a false sense of securiy about the reliability of the information it contains. In fact, the sources of uncertainty and error involved in the process are multiple. First, there is measurement error at the specific points that are used to produce the basic layers; then the measured points are interpolated spatially to produce a map layer, and there is error produced by the interpolation process; next, the layers are combined according to the model, which is a simplified representation of reality that introduces additional uncertainty; and finally, there is the propaggation of the error associated with each layer through the model.

Field values which are observation inputs for models are only known (with some error) at specific sample points. We must interpolate between these points to produce maps, and the quality of the data at interpolated locations is only as good as the interpolation process that produced them, which also depends strongly on the number of known field values. The accuracy of a reconstruction for a given regionalized parameter, for instance the hydraulic conductivity pattern, is affected by uncertainties in the (spatial) data originating from geographical abstraction, data acquisition, geoprocessing and use of data. Uncertainty may be of a conceptual nature, associated with the process of abstraction or generalisation of real world phenomena (Goodchild & Zhang, 2002), or it may occur in measuring the positions and attributes of geographical entities or in sampling. We must also keep in mind that there is always a possibility that our predictions issued from spatial interpolation (spatial modelling) processes turn out to be completely wrong at some locations since for no foreseeable reason the phenomena at some unknown places may be radically different than anything observed (Chilès and Delfiner, 1999). This is what Matheron calls the risk of a "radical error" and this is due to the spatial variability and continuity character of geophysical data.

Computational models are mathematical representations of a simplified reality. This simplified reality is called the conceptual model, and it often excludes processes or properties that may be of importance but about which little is known. For instance, intrinsic vulnerability models cannot easily account for the natural variability and complexity of geologic materials (particularly in fractured rock environments), and properties are often represented using an "equivalent porous medium" approach. Likewise, specific groundwater vulnerability models do not take into account all aspects of the important diversity and complexity of chemical and microbiological reactions or exchanges between the lithosphere, hydrosphere and atmosphere, and they rarely take into account the impacts of the contaminant source geometry and magnitude (mass loading and natural attenuation potential). The transient nature of proscesses is also often ignored, as are phenomena of a global nature such as global warming.

We must therefore keep in mind that models have their limits and represent reality only up to a certain point (Chilès and Delfiner, 1999). The uncertainty introduced by the choice of the conceptual model is impossible to evaluate per se. Only through the use of sensitivity analyses can we evaluate, in a very limited way, the impact of the simplifying assumptions.

GIS work with the field approach which represents the world simply as fields of attribute data (elevation; hydraulic conductivity; precipitation; etc.), without defining abstract geographical objects. Those fields only provide the value of an attribute at any location and do not link these values with any recognized objects (Heuvelink, 1998).

The final step of the process, which is the focus of this thesis, is to calculate how the sources of uncertainty in the various input layers propagate through our chosen groundwater vulnerability model.

This chapter describes in detail these various sources of uncertainty and how they propagate through the model.

Geographic Information System

Except where noted, much of the following section paraphrased from Bonham-Carter (1994).

A geographical information system (GIS) is a computer system for managing spatial data which are known or can be calculated in terms of geographic coordinates (Bonham-Carter, 1994). GIS can be described as a software package for the storage, analysis and presentation of geographical information providing a set of tools for collecting, storing, retrieving at will, transforming and displaying spatial data from the real world for a particular set of purposes (Burrough, 1986; Heuvelink, 1998). Ultimately, GIS provides support for making decisions based on spatial data or general research. GIS achieves these major goals through one or more of the following activities with spatial data: ***organization***, ***visualization***, ***query***, ***combination***, ***analysis*** and ***prediction*** (Bonham-Carter, 1994).

The efficiency and type of data ***organization*** affects all the other following activities and is therefore of fundamental importance (Bonham-Carter, 1994). Ideally, the data organization process should document uncertainties linked to the data acquisition and data processing processes. This information about the underlying data is called metadata. Errors introduced by data acquisition and data processing will be discussed in detail below (source and calculation of uncertainty in spatial data sources). Schemes for organizing data are sometimes called data models. Usually

GIS programs organize spatial data thematically categorizing data in *layers* (forest cover, soil classification, elevation, road network, ecological areas, hydrology, etc.). While spatial data layers are input one at a time, the process of entering attribute data is done one layer at a time. The data model used by the data storage subsystem will impact data organization formatting which is an essential required component of manipulation and analysis tasks. A clear understanding of a GIS project's needs and requirements is essential before any data input and/or layer definition. This is the key for success and could greatly improve the quality and reliability of the derived GIS outputs/products.

When producing conceptual models of the real world into GIS, data models abstract, simplify and discretise geographical realities to facilitate analysis and decision making. Geographical abstraction is selective, entails structuring of the web of reality, and leads to approximation (Chorley and Haggett, 1968). Goodchild (1989) describes data modelling as the process of abstracting and generalising real forms of geographical variation in order to express them in a discrete digital store. This process produces a conceptual model of the real world and it is highly unlikely that geographical complexity can be reduced to models with perfect accuracy. Thus the inevitable discrepancy between the modelled and real worlds constitutes inaccuracy and uncertainty (Goodchild and Zhang, 2002). In addition to the underlying phenomena of interest, uncertainties are an integral part of geographical information. Object and fields are the two differentiated yet closely related types of data models for geographers and will be discussed in detail further below.

The GIS ***visualization*** tools are based on graphical capabilities of computers. Visual display through video monitors and/or hardcopy displays allow the users to visually analyse or query complex spatial relationships without having to look at less inviting, huge tables of numbers (Bonham-Carter, 1994).

While visualization reveals a spatial pattern amongst collections of organized data items, ***spatial query*** is helpful for answering questions about special instances in the data. Spatial query is complementary to data visualization in helping to find the reason behind the spatial pattern. GIS provides tools for two types of interactive query. The first type is the question "What are the characteristics of this location?" The second type is "Whereabouts do these characteristics occur?" (Bonham-Carter, 1994).

Integration models (***combination***) create a new map from two or more existing maps using arithmetic and logical operations. This process can often lead to an understanding and interpretation of spatial phenomena that are simply not apparent when individual spatial data types are considered in isolation (Bonham-Carter, 1994). Integration models can reveal spatial relationship between spatial datasets from quite different sources. There is an increasing demand for GIS products issued from the integration and merging of multisource data. Nevertheless, spatial datasets from different sources can model the reality in such different ways that it can lead to incompatibility between the datasets. Before combining the data, it is therefore necessary to make them compatible together (Wang and Howarth, 1994):

- In the data structure: vector or raster coverages;
- In the data types: binary, nominal, ordinal, etc.;
- Regarding spatial context & geometry and levels of generalization.

Analysis infers meaning, and improves the understanding, knowledge and quality of the datasets. Through a GIS, analysis is carried out visually (video monitor and/or hardcopy display) or by measurements, statistical computations, fitting models to data values and other operations. Following are a few examples outlined by Bonham-Carter (1994):

- An analysis of areas on a map may lead to a table (histogram) summarizing the proportions of a region underlain by classes of surface materials.

- A statistical summary might be used to compare the mean and standard deviation of measurements by attribute type, for example, airborne uranium measurements by rock type.
- A regression model might be fitted to bivariate geochemical data and the resulting predicted values displayed as a map

Finally, GIS based systems were developed for mapping and predicting site characteristics using datasets. GIS has advantages over conventional methods in integrating various data sources, performing spatial analysis, modeling spatial processes, and mapping the results in land-use studies. For example a number of data layers indicative of geological and hydrogeological conditions may be combined together to predict groundwater vulnerability to contaminants. The purpose of a GIS study is often prediction. **Prediction** is sometimes a research exercise to explore the outcomes of making a particular set of assumptions, often with the purpose of examining the performance of a model (Bonham-Carter, 1994).

Generalization of geographical reality

There are two approaches to representing geographic information in a GIS. These are the ***field*** and ***object*** approaches (Goodchild, 1989, 1992). An understanding of this basic theoretical aspect is essential to the good comprehension of the GIS. The **field-based** perspective of geographical abstraction considers geographical reality as consisting of a set of single valued function defined everywhere while **object-based** is a collection of discrete objects. The selection of the type of model depends on the specific nature of the underlying phenomenon to input in the GIS and is limited by the data acquisition techniques and implementations of geographical databases. Field-based models are favoured in physical geography and environmental application, while object-based models are more suitable for cartography and facilities management (Goodchild and Zhang, 2002).

A field-based model deals with phenomenon which are distributed over geographical space and whose properties are a function of spatial coordinates and, in the case of dynamic fields, of time. Hydraulic conductivities, precipitation, temperature, etc., are natural and/or environmental phenomena measured on continuous scales. As observed by Zhang and Goodchild (2002), Goodchild (1988) and Bailey (1988), it is easier to model uncertainties in field-based models rather than in object-based models because the field approach permits a better integration of the roles of data acquisition and analysis (raw data) which are essential for deriving necessary information on errors in geo-processing.

Object data models were developed in the past largely for the purpose of digitising the contents of paper maps. Employing discrete points, lines and areas to represent spatial data seems to be compatible with the typology of real world entities (buildings, telecommunication cable networks, road networks etc.), whose positions and attributes may be acquired by direct or remote measurement, as is done by a surveying and mapping specialist (Goodchild and Zhang, 2002).

It is also important to briefly discuss two- and **three-dimensional hydrogeological modelling** in GIS.. Three dimensional interpretations of hydrogeological structures work with a sparse and irregular collection of data samples requiring considerable use of interpolation techniques, abstract visualisation of sub-surface units and the manipulation of data sets to test multiple working hypotheses.

Most GIS are restricted to data in two spatial dimensions, although some systems of particular interest to geologists, in particular hydrogeological systems, are in fact better represented by a three-dimensional system (Bonham-Carter, 1994). The problems encountered in modeling three-dimensional and time dependent groundwater flow in aquifers, are economy of computation, approximation and stability, which are especially demanding. At the present time the GIS environment is awkward for groundwater computation; it is however very well suited for developing conceptual models that can be used as input into numerical simulators based on finite-difference or finite-element approximations. The development of robust, GIS-based, three-dimensional conceptual models for groundwater flow and transport is of great interest for groundwater modellers particularly at the regional scale.

Hydrostratigraphic analysis was borrowed from the oil and gas industry where it is used to determine the best way to exploit underground reserves. It is fairly new to groundwater modeling but has proven to be an effective tool (Gélinas, 1996). However, even with a large body of characterization data that can be used to formulate the conceptual model the level of detail is often still insufficient to generate a highly resolved conceptual model at the scale of remedial efforts.

GIS can also be used in inverse modeling for back–calculating parameters from the behaviour of the system. For instance, measured patterns of groundwater flow and contaminants through the subsurface can be used to back-calculate soil properties with a variety of inverse solution methods (G. Blake, 1996; Chilès and Delfiner, 1999). This approach is of great interest to complete hydrostratigraphic analyses (G. Blake, 1996); however, it is not universally accepted because of problems of non-uniqueness of the inverse solutions, meaning that there are a number of parameters that can satisfy the observed behaviour.

Spatial interpolation: kriging and semivariogram concepts

Interpolation is a common task in many fields of physical science when there is a need for creating continuous surfaces or maps from point data. The general approach used in spatial interpolation is to estimate a function F(**x**), where **x** = (x,y), at one location $\mathbf{x}_p$ based on a limited number, m, of known values of F at neighboring points x_i. The interpolation function is a weighted sum, given by:

$$F(x_p) = \sum_{i=1}^{m} W_i * F(x_i)$$

The major issue is to determine the weighting value W_i of each neighboring point x_i. Many methods exist for choosing the weighting values. Two of the best-known methods are the Inverse Distance Weighting (IDW), which relies on linear interpolation between observations, and the cubic splines interpolation. The main shortcomings of these methods are:

- They do not account for data support (the methods use the nearest neighbors irrespective of the spatial density of measurement points);
- They do not account for the pattern of spatial variability (i.e. spatial correlation);
- They cannot provide estimation error.

When combined with more sophisticated approaches to density mapping, such as kernel density estimation, those estimation techniques can provide estimation error while taking into account spatial variability of the variables.

The last point is of particular importance in physical sciences: with the complexity of natural phenomena and the sparse information availability, error cannot be ignored. The need for a probabilistic method for spatial interpolation gave rise to what is arguably one of the most commonly used interpolation methods: kriging. The method was named after a geostatistician by the name of D. G. Krige from South

Africa, and is based on a geostatistical analysis of the existing data and the resulting variograms. The variogram gives the mean squared difference between known observations as a function of the distance (or lag) separating them (this quantity is called the semi-variance and a plot of the semi-variance as a function of lag is called a variogram, or sometimes a semivariogram). In essence, kriging uses the variogram to calculate the optimal weights to be assigned to the neighbouring values in the above equation. Since the semi-variance changes with lag, as described by the variogram, the weights depend on the spatial distribution of the known sample locations. The number of observations will not affect the spatial variability of the attribute, but it will reduce the uncertainty about it.

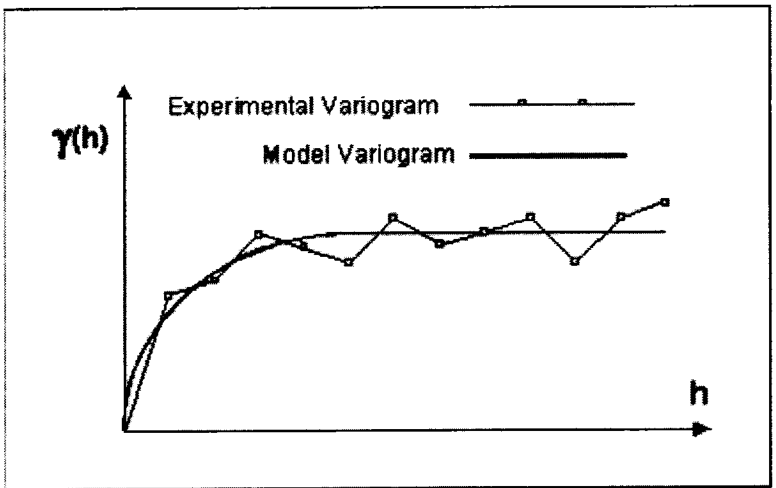
The main advantage of kriging is that it provides a geostatistically-based interpolation method that provides a measure of error and uncertainty when determining estimations. Details of the procedure are given below.

The semi-variance is calculated with the following equation for the $n(h)$ pairs of points $F(x_i)$ and $F(x_j)$ valuated at x_i and x_j, respectively, and separated by a lag distance $h = |x_i - x_j|$:

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} (F(x_i) - F(x_j))^2$$

As we can see from the above equation, for a given lag distance h, the semi-variance is equal to one half of the mean squared difference between observations separated by a distance h (hence the "semi" in "semi-variance"). If the mean and variance of function F are stationary, i.e. space invariant, then the semi-variance can be related to the autocovariance and autocorrelation functions, which offer some geostatistical advantages. The key notion here, is that the semi-variance does not depend on the point position but only on the distance between the known points.

Finally, to obtain a continuous function giving the semivariance as a function of lag, we have to fit an analytical function to the experimental variogram with a least squares estimate method. The choice of the function fitted to the semivariogram , called the model variogram, is the more critical challenge of the kriging process and it may involve expert judgment (subjective) operations. The model variogram is a simple mathematical function that represents the experimental variogram.



(Source: ESRI)

The above diagram is a schematic diagram of the typical variogram showing that points that are close together have similar F values (small semi-variance at small lags). The (average) difference between measurements increases as the distance separating them increases, up to a point, called the range, where the semi-variance becomes constant. The constant portion of the variogram is called the sill. At the sill, the semi-variance corresponds to the variance of the observations. Within the **range**, the locations are close enough that the values of the function are related to each other, and all known samples contained in this region, also referred to as the neighborhood, must be considered when estimating the unknown point of interest (Davis, 1973).

Technically, at a lag (separation distance) of zero, the semi-variance should be zero (the difference between an observation and itself is zero). However some variogram models produce a semi-variance value different than zero called a nugget.

Consequently, most variogram models require that three components be specified to describe the semivariogram: the **nugget**, the **sill** and the **range** .

The semivariograms can also be modeled while taking into account both the magnitude and the orientation (direction) of the lag vector between the points. This allows the modeller to search for statistical anisotropy. Statistical anisotropy is usually the result of physical anisotropy of the underlying function and is therefore a very useful piece of information. However, the directional semivariograms method increases the kriging processing time by 80% to 200%. Nevertheless, this method offers the modeler greater flexibility, and simulation or estimation results that are more representative of the site because the spatial variation of the data can be more precisely defined (Wingle and Poeter, 1998).

The next step in the kriging process consists in calculating the weights (W_i) of the main equation based on the distances separating the unknown interpolated point from a pre-selected number of neighboring points and the corresponding values of the model variogram function $\gamma(h)$.

Importance of uncertainty assessment in the decision process

"The fact that our descriptions of spatial phenomena are subject to uncertainty is now generally accepted, but for a time it met with much resistance, especially from engineers who are trained to work deterministically." (Chilès & Delfiner, 1998)

Because models replace experiments with an extrapolation from experience, they should always include probabilistic analyses and probable errors (uncertainties) of results in order to enhance the credibility of the process and avoid any suspicions towards the results. Even the most sophisticated groundwater vulnerability model can be set up and run for a fraction of the cost of a field experiment or survey, but this is only worth doing if the results will be believed (Wauchope, 1992).

No matter how sophisticated, all computer simulation models are simplifications of reality. Whether we are modeling a well-characterized system or trying to extend our models to predict behaviour in a lesser known system, we, at some point, surpass our knowledge and must estimate effects. Hence, suspicion of any result presented without a probable error estimate is normal (Wauchope, 1992). By assessing the sources of uncertainty in a model, we also work towards the goal of improving the model.

Zhang and Goodchild (2002) underline the fact that data may pass through many different transactions and custodians, each providing their own protocols and/or interpretations. This situation could lead to major issues of accuracy and data quality concerns if metadata (quality information description) are not attached with the data. In the last decade, the GIS research community has recognized the need and has vastly expanded the knowledge of many aspects of accuracy assessment and data quality more generally (Chrisman, 1999).

Uncertainty is not only a property of the data contents, but also a function of the relationships between data and users. In this context, Zhang and Goodchild (2002) defined uncertainty "as a measure of the difference between the data and the meaning attached to the data by the current user". In other words, it is essential to assess the uncertainty of a model to measure the difference between the content of a database, and the contents that the users would have created by a direct and perfectly accurate observation of reality.

One of the most pervasive forces in GIS technology and environmental modelling techniques/application/methodology is the tyranny in the way things have always been done. Despite the fact that the research community working on accuracy assessment in Natural Resources and Environmental Sciences met together each two years since 1994 (first International Symposium on this topic) and produced great advances, the practice of environmental modelling (like groundwater vulnerability modelling) did not really change its routine ways of doing things. Because of the general negative perception of uncertainty/error, there is reluctance from geoscientific researchers and engineers to show uncertainties/errors attached to their data/results/models. This problem is particularly critical in the case of non-expert users (who constitute a considerable and expanding proportion of the user-base) and for decisions that rely very much on environmental modelling results and/or for which the consequences of inappropriate use are severe (Agumya and Hunter, 1999). One approach (Assessing "fitness for use" of Geographic Information) was developed by Agumya and Hunter (1997) whereby such consequences are expressed in terms of risk during the decision-making process.

Sources and calculation of uncertainty in spatial data variables

The development and improvement of techniques for the modelling and mapping of groundwater vulnerability rests largely on the usage of the Geographic Information System (GIS) to import, handle, analyze and combine spatial data. For Oliveira and Lobo Ferreira (2002), given the ability to manipulate large spatial data sets, the major problems with vulnerability mapping are:

- Identifying the most appropriate data that will be used to define or model vulnerability and,
- Obtaining the spatial data and importing this in the correct format into a GIS.

Sources of uncertainty for geophysical data are generally divided into two categories: **Uncertainty from data measurement** and **uncertainty from data processing**. The "uncertainty from data processing" includes the following subcategories:

- Uncertainty in categorical variables;
- Uncertainty from integration and merging of multisource data; and
- Uncertainty from data modelling and geoprocessing.

The error or uncertainty propagation in modelling will be covered in the next section.

Some authors like M. Fortin (1998) use three different categories: Uncertainty in categorical variables, uncertainty from data acquisition and uncertainty from data processing. In this study, we preferred to address the *uncertainty in categorical variables* within the wider category of *data processing* because categorical variables are issued from raw data categorization/classification (processing). We also decided to include *uncertainty from data acquisition* in the category of *data measurement* because data acquisition could occur from already processed data sources.

Uncertainty automatically occurs from **data measurement** because all direct measures of a phenomenon are, at a certain point, influenced by effects of random events. The precision and the reproducibility of data is then dependant on these random events. These effects/variations in data measurement results could be linked, inter alias, to the following events/situations:

- Differences in physical conditions (meteorological, atmospheric, etc.);
- Different type of equipment, instrument calibration; sensor resolution (remote sensing), etc.
- Fluctuation in the level of attention, concentration and knowledge of the analysts/technician,
- Others.

From a statistical point of view, for all observation or measurement of a phenomenon, there is a deviation that exists between the measured value and the real or true value. The difference between these two values represents the error on the measurement. Because most of the time the true value is not known, we are rarely able to quantify exactly the error of a measurement value. Since, measurement techniques used to obtain the true value of a phenomenon are not perfect, we are instead representing the value with an associated uncertainty.

For instance, if we were to repeat a given measurement a large number of times (say an infinite number of times), under conditions that are identical as possible, then the measurements would follow a normal distribution because they are affected by a large number of small random variations (experimental error).From this population, the mean (which is equal to the median and the mode for a Normal Distribution) corresponds to the best representation of the real or true value.

$$\mu = \lim_{N\to\infty} (\tfrac{1}{N}\sum x_i)$$

Where x_i = Value of observation i

N = Total number of observations

Each deviation d_i between the mean value and the measured values x_i represent the uncertainty of this measure.

$$d_i = x_i - \mu$$

The mean absolute deviation, α, can be used to describe the error:

$$\alpha = \lim_{N\to\infty} (1/N\sum|x_i - \mu|)$$

But a more mathematically useful quantity is the mean squared deviation, called the variance:

$$\sigma^2 \equiv \lim_{N \to \infty} \left[1/N \sum (x_i - \mu)^2 \right]$$

The variance and its square root, the standard deviation, are measures of the dispersion of the observations around the mean of the population, which assesses the uncertainty associated with the limited experimental precision when trying to measure the real value of a phenomenon.

Data measurement is followed or sometimes replaced by the process of data acquisition, which is an additional source of uncertainty. For instance, data acquisition from remote sensing technologies involves automated on-the-fly analysis processes that produce classification of land cover features such as vegetation, soil, water and forest. The resulting classified raster images do not take into account the fact that the continuity and heterogeneity of classes in the real world are separated by undetermined boundaries with progressive zones of transition. Classified raster images are therefore subject to considerable uncertainty along and across boundary zones. This uncertainty is evaluated using the concept of fuzzy logic.

The theory of the fuzzy subsets was developed by L Zadeh in 1965 in order to mathematically represent the inaccuracy relating to certain classes of objects. This approach allows us to predict the class of a given geophysical variable/phenomenon along with a "certainty factor" (CF) associated with the classification process. The uncertainty is then managed by assigning degrees of certainty to the facts (geophysical variables) and the rules of deduction (representing the classification process) to lead to conclusions (categorization of variables) with more or less certainty. In addition to giving us the magnitude of the errors, this approach allows us to answer the following questions: How frequently is the category assigned the best choice for a given site/location? And how frequently is the category assigned acceptable?

Adding to uncertainty from data measurement and acquisition is uncertainty from **data processing.** This is an increasing source of uncertainty because GIS technology is more frequently used in the industry and among public services organizations, thereby increasing raw data manipulation processes. These manipulation processes include *raster image processing, formats conversion, digitization*, *generalization*, *spatial interpolation, integration and merging of multi-source data*, *analysis*, etc.

Geometric errors on raw raster data have to be corrected before their integration with other spatial data. Geometrical corrections (a type of *raster image processing*) are applied to the images to reduce the geometrical deformations that occur during the recording of the scene: distortions caused by the environment (curve of the ground, variation of altitude on the ground, atmospheric refraction...), distortions due to the errors of the systems of measurement and distortions coming from the movements of the platform. The geometrical corrections are also used to project and georeference the images. These techniques involve resampling processes with effects on the integrity of the original data that are not well understood (Fortin, 1998).

Format conversions processes also contribute to generate uncertainties/errors in the final outputs/results. Format conversion can alter data. The conversion software may or may not produce the intended results and conversion errors may be gross or subtle. It is then important to assess the uncertainty/risk inherent in conversion programs. This can be done by examining a file before and after migration to make sure that fields and field values of the original source file are properly reproduced in the target file.

While uncertainties/errors introduced during the *digitization* process of categorized data are generally negligible compared to uncertainties/errors already present in the data, multisource data integration is certainly a major source of uncertainty when dealing with different data structures (raster and vector coverages), data types (binary, nominal, ordinal, etc), spatial context and geometry and levels of generalization.

Cartographic *Generalization* often leads to simplified and smoothed maps, either as the technical consequences of procedures, or for aesthetic purposes (Douglas and Peucker, 1973; MacEachren, 1994). The difference between the modeled and real worlds represents the uncertainty due to geographical abstraction (Zhang and Goodchild, 2002). For instance, digitized data from maps will suffer any uncertainties that are already embedded in the source maps themselves, and errors introduced by the digitizing process (Bolstad et al., 1990). The former source of errors exists due to the fact that published maps have usually been subjected to a map generalization process, involving a variety of abstraction, selection, simplification, and approximation steps (Brassel and Weibel, 1988; Joao, 1998).

The problems (source of uncertainty) of *spatial interpolation* from scattered data (point data) as a method for prediction and representation of multi-variate fields was covered in the previous section "Geographic Information System". Interpolation is perhaps the most critical source of error in GIS. This aspect can impact all stages of the analysis process and still needs extensive study.

Theory of error propagation models with local GIS operations

The final section of this chapter looks into the available techniques to calculate error propagation through GIS operations. Multi sources spatial data stored in a GIS database are utilized as inputs for GIS operations. The errors in the input will propagate to the output of the operation (and so on if the output from one operation is used as input to an ensuing operation). This is because the resulting output is a function of the input values, and inaccurate input values automatically affect the computed results (Heuvelink et al., 1989). In addition, GIS operations which are computational model (spatial modelling), for instance the spatial interpolation of hydraulic conductivity for an entire watershed from limited sample points, produces an approximation of reality that contains errors. This is what we call the model error. In the calculation of the error propagation through GIS operations, both input and model error affect the output and both must therefore be accounted for (Heuvelink, 1998).

The error propagation calculation methodology presented below is a partial synthesis of the work of Heuvelink and Burrough (1989) and Heuvelink (1998), which was well summarized by Fortin (1998).

When an attribute, x, is not easily measurable or observable directly we can use modelling processes to estimate it. In the model, we use parameters u, v ..., that are measurable/observable in order to derive the targeted attribute, x, according to:

$$x = f(u, v, ...)$$

Now that we have a good idea of how we calculate the mean and uncertainty (standard deviation) values for our measurable variables (previous section), we must determine which is the most effective method for calculating the mean and uncertainty (standard deviation) associated with the derived attribute. The most probable mean value of x will be then given by:

$$\bar{x} = f(\bar{u}, \bar{v}, ...)$$

Which is obtained by the GIS operation $f(u, v, ...)$.

The uncertainty of the resulting value of $\bar{x}$ could be determined by taking into consideration the distribution of x_i values resulting from the combination of the individual measurements of $u_i, v_i, ...$

$x_i = f(u_i, v_i, ...)$

The variance will be then calculated as before:

$$\sigma^2 \equiv \lim_{N\to\infty} 1/N \sum (x_i - \bar{x})^2$$

For random error, we can express the deviations $(x_i - \bar{x})$ as the sum of the deviations of $(u_i - \bar{u})$, $(v_i - \bar{v})$... each expressed as a truncated Taylor Series expansion:

$$(x_i - \bar{x}) \cong (u_i - \bar{u})\left(\frac{\partial f}{\partial u}\right) + (v_i - \bar{v})\left(\frac{\partial f}{\partial v}\right) + ...$$

The variance of x will then be calculated with:

$$\sigma_x{}^2 \cong \lim_{N\to\infty} 1/N \sum \left[(u_i - \bar{u})(\partial f / \partial u) + (v_i - \bar{v})(\partial f / \partial v) + ... \right]^2$$

Expanding the square yields:

$$\sigma_x{}^2 \cong \lim_{N\to\infty} 1/N \sum \begin{bmatrix} (u_i - \bar{u})^2 (\partial f / \partial u)^2 + (v_i - \bar{v})^2 (\partial f / \partial v)^2 + \\ 2(u_i - \bar{u})(v_i - \bar{v})(\partial f / \partial u)(\partial f / \partial v) + ... \end{bmatrix}$$

The two first terms can be express in terms of the variances of u and v:

$$\sigma_u^2 = \lim_{N\to\infty} 1/N \sum (u_i - \bar{u})^2 \quad \text{and} \quad \sigma_v^2 = \lim_{N\to\infty} 1/N \sum (v_i - \bar{v})^2$$

The last term is actually the covariance between u and v, defined as:

$$\sigma_{uv}^2 \equiv \lim_{N\to\infty} 1/N \sum \left[(u_i - \bar{u})(v_i - \bar{v}) \right]$$

The final expression is obtained from these definitions of variance and covariance:

$$\sigma_x^2 \cong \sigma_u^2(\partial f/\partial u)^2 + \sigma_v^2(\partial f/\partial u)^2 + 2\sigma_{uv}^2(\partial f/\partial u)(\partial f/\partial v) + \ldots$$

The two first terms correspond to the uncertainty contribution from the parameters (variables) to the final uncertainty while the last term takes into account the relationship between the variables u and v (correlation). Without any correlation, the last term can be dropped and the variance of x will simply be the sum of the variances of the parameters weighted by their impact in the model, as expressed by the squared partial derivatives. It is important to note that if there are more than two input variables the above equation will have as many variance contributions as there are variables, and as many covariance terms as there are pairs of variables.

As underlined by Heuvelink (1998), "...not only will input error propagate to the output of a GIS operation, but model error will as well. This means that the output of a computational model may disagree with reality even when the input to the model is completely error-free, simply because the model is only an approximation of reality." The statement by Heuvelink (1998) is, in fact, somewhat incorrect: the above equation says that the input error will propagate in quadratic proportion to its contribution to the model (based on the squared partial derivative term); if the input variable error is zero or if the model does not depend on the variable (partial derivative equal to zero) then the input variable will not contribute additional error to the resulting dependent variable.

Not only does the input error propagate in the model, the spatial correlation structure must also be propagated. As pointed out by Heuvelink, (1998), the error model parameters should also take into consideration the spatial variability of the attribute and the mapping procedure used In order to decide which models of spatial variation to adopt in practical situation, when no prior information about the type of spatial variation is available, Heuvelink (1998) used three models of spatial variation:

1. The discrete model of spatial variation (*DMSV*),
2. The continuous model of spatial variation (*CMSV*), and
3. The mixed model of spatial variation (*MMSV*).

The *DMSV* assumes that the area of interest D can be divided into a finite number of more or less homogeneous units D_k, and that the value of a spatial attribute $Z(x)$ at any location x in unit D_k is the sum of a unit-dependent mean $\beta(k)$ and a location-dependant residual noise term $\varepsilon(x)$, which is uncorrelated with neighbouring values and whose variance is constant throughout the domain D (Heuvelink, 1998).

The *CMSV*, on the other hand, assumes that $Z(x)$ is a second order stationary random field meaning that there is no trend in the mean or variance, and that the autocovariance function of $Z(x)$, denoted here by $C_z(x)$, is a function only of the distance between the locations (Oliver and Webster, 1990; Cressie, 1991).

The assumptions underlying the *MMSV* are a combination of those underlying the *DMSV* and *CMSV*:

1. $\mu(x) = \beta(k)$ for all $x \in D_k$
2. $\text{cov}(\varepsilon(x), \varepsilon(x+h)) = C_\varepsilon(h)$ for all $x, x+h \in D$.

Second order stationarity is thus imposed on $\varepsilon(.)$ instead of on $Z(.)$. The *MMSV* is more general than the *DMSV* and *CMSV*, and in fact it contains both models (Heuvelink, 1998).

The error/uncertainty analysis process allows us to determine how much each individual variable/input and model error contributes to the output error. This information may be very useful, because it allows users to explore the degree to which the quality of the output improves, given a reduction of error in a particular output (Heuvelink, 1998). Also, in many situations model error will be a major source of error and should therefore be included in the error analysis to estimate the overall uncertainty in the model output.

3. GROUNDWATER VULNERABILITY MODELLING

General overview

Groundwater is a major source for public water supply and is often safer and more reliable than surface water sources. In terms of groundwater management strategies, contamination prevention and resource protection should be emphasized rather than remedial treatment. Thus, an assessment of groundwater vulnerability to pollution should always take place first before exploiting an aquifer for potable water supply.

The term vulnerability was used in the sixties in France. It was first introduced as a scientific term in the literature by Albinet and Margat (Doerfliger and others, 1999). Groundwater vulnerability to pollution was defined by Lobo-Fereirra and Cabral (1991) as "the sensivity of groundwater quality to an imposed contaminant load, which is determined by the intrinsic characteristics of the aquifer" (Lobo-Ferreira, 2001).

The likelihood for contaminants to reach some specific position in the groundwater after their introduction at some point above the top of the uppermost aquifer is determined by the natural attenuation processes occurring within the zone between the pollution source and the aquifer. In contrast to intrinsic vulnerability modelling which models groundwater vulnerability without respect to the nature of contaminants, specific vulnerability modelling takes into account some of the natural physical and chemical processes occurring between the contaminants and the soils. Chemical reactions that occur in the soil in the unsaturated zone and saturated zone can cause contaminants to change their physical states and their chemical forms, which can remediate the groundwater pollution or change the characteristics of contaminants (H. Guo and Y. Wang, 2004).

During the last two decades, groundwater vulnerability modelling and mapping was mainly developed through geographic information system (GIS). As mentioned in the introduction to this thesis, Lobo-Ferreira (1999) underlined three major problems with groundwater vulnerability mapping:

- Choosing the most appropriate approach
- Identifying the most appropriate data which will be used to define or model vulnerability
- Obtaining the spatial data and importing this in the correct format into a GIS

This chapter focuses on the first of these problems. The two other problems were discussed in Chapter 2 (Sources, calculation and modelling of uncertainties through environmental modelling).

There are four types of approach for modelling groundwater vulnerability (E. Artuso, M.M. Oliveira and J.P.C. Lobo-Ferreira, 2003): The **deterministic approach**, **probabilistic approach**, **stochastic approach** and **empirical approach**.

The **deterministic approach** is used for specific vulnerability modelling. An early example of this approach is due to Meeks and Dean (1990). They simplified the one-dimensional equation for the convective transport-dispersion-reaction process of solutes in a homogeneous porous medium to arrive at "leaching potential" indices (LPI) that ranked sites on the basis of their vulnerability to contamination. The LPI methodology mainly delineates the behaviour of degradable chemicals in the vadose zone, which is generally used to assess the groundwater vulnerability to pesticides (Guo and Wang, 2004), and it cannot be universally applied (Artuso, Oliveira and Lobo-Ferreira, 2003).

There is current progress and future potential of computer simulation models to predict pesticide risks. Wauchope (1992) underlines four current developments that provide a compromise between the two impossible extremes of total model validation and total field measurement:

1. As models develop more process details, the parameters become more fundamental in nature and these are generally more amenable to independent estimation or measurement;
2. The reporting of probabilistic analyses and probable errors of results by modelers will enhance credibility;
3. The development of data sets for specific crop/site/pesticide/weather scenarios will allow users to modify those parts of the input they are concerned with while retaining confidence that parameters with which they are unfamiliar will have reasonable values; and
4. Meso-scale and micro-scale physical simulation experiments can fill the gap between laboratory and field results.

The MLPI model (Modified Leaching Potential Index) is an improved method presented by H. Guo and Y. Wang (2004) that does not only focus on pesticides but also integrates more detailed physical processes for assessing, delineating and mapping groundwater vulnerability. The contaminants considered, included degradable contaminants, radioactive elements and non-degradable pollutants., Typically, more physically-based models have increased input requirements for parameters that are more fundamental in nature (Wauchope, 1992). As stated by Wauchope (1992), "In principle, more detailed, more physically-based models should ultimately be more accurate and more precise". As we learn more about fundamental processes, models evolve into more universal descriptions of reality and our dependence on "simple" empirical functions decreases (Wauchope, 1992). Each such model increase in explicitly accounted-for complexity should increase the accuracy of simulation of reality and the ability of the equation to describe different situations (i.e., the universality of the equation increases)" (Wauchope, 1992).

Another example of a deterministic approach is the Surface to Aquifer Advective Time calculation (SAAT). This approach will be discussed in detail shortly.

As stated by McNab et al. (1999), the **probabilistic approach** utilizes user-specified probability distributions of model variables to generate forecasts through multiple Monte Carlo realization. For instance "the Monte Carlo approach allows uncertainties in hydrogeological data (e.g., hydraulic conductivity, hydraulic gradient (magnitude and direction), nature of the source, and chemical data (e.g., degradation rates) to be translated into uncertainties regarding plume extents and rates of growth" (McNab and al., 1999).

Chilès and Delfiner (1999) explained the **probabilistic approach** and raised questions about the objectivity of probabilistic statements when modelling spatial uncertainty. The following statements, quoted from their book "Geostatistics, Modelling Spatial Uncertainty", underline the complexity of the probabilistic approach when applied to environmental modelling:

- The quantification of spatial uncertainty requires a model specifying the mechanism by which spatial randomness is generated.

- The clue is to carefully distinguish the model from the reality it attempts to capture. Probabilities do not exist in Nature but only in our models.
- Probabilities and their experimental foundation in the famous "law of large numbers" requires the possibility of repetitions, which are impossible with objects that are essentially unique (The objects we deal with, for instance the hydraulic conductivity or the rock porosity, are perfectly deterministic).

They continue on to say that despite these epistemiological constrains, the use of probabilistic statements (approaches) validates the environmental modelling methodology. Even though every groundwater watershed is unique, they all belong to classes of situation, regional water budget, geological stratigraphic units, etc., that are similar enough to give rise to specific methodologies, which over time can be validated objectively. This is "external objectivity" (Chilès and Delfiner, 1999). The use of probabilistic analyses in the presentation of environmental modelling results enhances credibility. Models do not have to be guilty of this. In the simplest deterministic models or equations, a propagation-of-errors analysis can be used to determine uncertainty in the predicted value (Wauchope, 1992). Heuvelink (1998) also underlines the fact that since a computational model through GIS operation is merely an approximation of reality, it also contains errors, here referred to as model error. Input and model errors both affect the output and both must therefore be accounted for (Heuvelink, 1998).

The **stochastic approach** is a kind of approach offering an alternative to subjective empirical judgments and data hungry deterministic assessment, using statistics to identify combinations of factors that determine the groundwater vulnerability for instance (E. Artuso, M.M. Oliveira and J.P.C. Lobo-Ferreira, 2003). In a stochastic model, the regionalized variable is regarded as one among many possible realizations of a random function. A stochastic model is analytically useful. It does not mean that we believe Nature to be random.

Finally, the **empirical approach** is based solely on direct observations on soil, geological and hydrogeological conditions where trends observed in the field data produce findings that are not reconciliated/linked with mathematical models of hydrogeologic variables (McNab and al., 1999). The mathematical representation of the process is simply a means of describing the relationship, without inference or insight into the physical process. The MINNESOTA's groundwater vulnerabilty model and DRASTIC described further below are examples of this type of approach.

This research compares three different methods for evaluating groundwater vulnerability, discusses their merits and drawbacks and evaluates their uncertainty. The three models, DRASTIC, ISI and SAAT, are described in detail below. The discussion of the three models is followed by a brief description of other existing models (GOD, EVARISK, MINNESOTA, EPPNA, SINTACS).

DRASTIC

This method was developed in the 1980's by the National Water Well Association at the request of the EPA (Environmental Protection Agency, USA). The name **DRASTIC** is an acronym from the seven parameters that enter into the method. The model uses a numerical ranking system assigning relative weights to the seven parameters selected to evaluate the relative groundwater vulnerability to contamination.

DRASTIC was developed based on 4 major assumptions (Aller et al., 1987)

i) the contaminant is introduced at the surface of the earth,

ii) the contaminant is flushed into the groundwater by precipitation,

iii) the contaminant has the mobility of water, and

iv) the area evaluated is 0.4 km^2 (400 000m^2) or larger.

The DRASTIC index is the weighted sum of 7 values corresponding to the following 7 parameters or hydrogeological indicators (ALLER *et al.*, 1987):

1. **D**epth to the water table: Shallow water tables pose a greater chance for the contaminant to reach the groundwater surface as opposed to deep water tables.
2. **R**echarge (Net): Net recharge is the amount of water per unit area of the soil that percolates to the aquifer. This is the principal vehicle that transports the contaminant to the groundwater. The more the recharge, the greater the chances of the contaminant to be transported to the groundwater table.
3. **A**quifer material: The material of the aquifer determines the mobility of the contaminant through it. An increase in the time of travel of the pollutant through the aquifer results in more attenuation of the contaminant.
4. **S**oil type: Soil media is the uppermost portion of the unsaturated / vadose zone characterized by significant biological activity. This along with the aquifer media will determine the amount of percolating water that reaches the groundwater surface. Soils with clays and silts have larger water holding capacity and lower conductivities, and thus increase the travel time of the contaminant through the root zone.
5. **T**opography: The higher the slope, the lower the pollution potential due to higher runoff and erosion rates. These include the pollutants that infiltrate into the soil.
6. **I**mpact of the unsaturated (vadose) zone: The unsaturated zone above the water table is referred to as the vadose zone. The texture of the vadose zone determines how long the contaminant will travel through it. The layer that most restricts the flow of water will be used.
7. **C**onductivity (Hydraulic): Hydraulic conductivity of the soil media determines the amount of water percolating to the groundwater through the aquifer. For highly permeable soils, the pollutant travel time is decreased within the aquifer.

A value (rating) from 1 to 10 (regrouping a range of "real" values or from an ordinal scale) is attributed to every parameter. High values correspond to a higher vulnerability according to the characteristic of the environment. Subsequently, each term of the DRASTIC sum is calculated by multiplying the value assigned to the parameter by its relative weight. Each parameter has a pre-defined weight consequent to its relative importance for the quantification of vulnerability. The most important parameters have a weight value of 5 and the less important have a weight value of 1 (E. Artuso, M.M. Oliveira and J.P.C. Lobo-Ferreira, 2003).

Two ranges of relative weights or weighting factors are available: one range for *general* or *standard* application (designated by DRASTIC) and another specific range for *pesticides* (designated as DRASTIC *pesticides*):

DRASTIC	**DRASTIC**	***pesticides***
(a) **D**epth to the water table	5	5
(b) **R**echarge (Net)	4	4
(c) **A**quifer material	3	3
(d) **S**oil type	2	5
(e) **T**opography	1	3
(f) **I**mpact of the unsaturated zone	5	4
(g) **C**onductivity (Hydraulic)	3	2

Finally, the seven partial products are added for each given location on a map to create the DRASTIC indices:

$$\text{DRASTIC INDEX} = D_R{*}D_W + R_R{*}R_W + A_R{*}A_W + S_R{*}S_W + T_R{*}T_W + I_R{*}I_W + C_R{*}C_W$$

where R=rating and W=weight. The resulting index has a minimum possible value of 23 and maximum possible value of 226.

ISI and AVI

The goal of the Aquifer Vunerability Index (AVI), developed in the early 1990's by the National Hydrology Research Institute (Canada), was to assess the vulnerability of local groundwater aquifers to surface contaminants. The Ontario Ministry of Environment modified this method with the calculation of the Susceptibility Index (ISI) at each well to create groundwater intrinsic susceptibility (GwIS) maps showing the areas of high, medium, and low intrinsic susceptibility and index values at point locations.

The ISI is calculated by summing the product of the thickness of each unit in the well log (**b_n**) with a corresponding K-Factor (**Kf_n**). The K-Factor is a dimensionless, relative number that can be loosely related to the negative exponent of the vertical hydraulic conductivity in m/s. This K-factor is therefore akin to the log hydraulic resistivity (**c**) (NHRI, 1992). This method was applied by van Stemvoort *et al.* (1992) to a part of the border zone of the Canadian regions of Saskatchewan-Alberta, and was adopted by the MOE of Ontario for its Province-wide large-scale groundwater studies.

The K-Factor of the MOE is given by:

$$Kf = \log \frac{1}{K}$$

Where "K", for a given group of geomaterial, is the highest "K" value obtained from a table of classified hydraulic conductivity values.

The calculation is performed from ground surface to the first significant aquifer (Ontario Ministry of the Environment, 2001). The mean ISI value at each well log can be calculated, as:

$ISI_{well} = \sum b_n * Kf_n$ For layers 1 to n

For *confined aquifers*, assuming that contaminant from the surface must migrate through the confining layer and reach the aquifer to cause potential impact; the ISI is calculated by summing the product of the thickness of each geological unit in the well and the corresponding K-Factor from ground surface to the top of the first significant aquifer (Ontario Ministry of the Environment, 2001). Thus the ISI can be loosely interpreted as the equivalent vertical hydraulic resistance.

For unconfined aquifers, semi-confined aquifers or where there is doubt about the integrity of the confining layer, assuming that contaminants from the surface must only migrate to the water table to cause potential impact, the ISI is calculated by summing the product of the thickness of each geological unit in the well and the corresponding K-Factor, from ground surface to the water table (Ontario Ministry of the Environment, 2001).

After calculation through the overall territory, the ISI indices are regrouped into thematic classes underlining different level of vulnerability. The Ontario MOE proposes the following assessment criteria:

- Low intrinsic susceptibility for values greater that 80;
- Low to moderate intrinsic susceptibility for values between 30 and 80; and
- High intrinsic susceptibility for values less than 30

SAAT

The Surface to Aquifer Advective Time (SAAT) model was described by the MOE Technical Experts Committee (MOE, TEC report, 2004). It addresses the aspect of vulnerability assessment that evaluates the urgency of an existing problem or the time available to respond to future threats; however, it does not account for mass loading and storage or cumulative effects.

The method is based on the same information that was used for ISI, plus the hydraulic gradient and porosity; alternatively, the infiltration rate, porosity and depth can be used. The advective time approach calculates the time it would take for a water tracer particle to travel from the ground surface to the aquifer of concern by the bulk motion of the water in the geologic material (advection).

As pointed out in the MOE Technical Expert Committee Report (2004) "this method is akin to methods used to evaluate supply well vulnerability using surface to well advective times (SWAT) thus providing continuity of methodology for aquifer and for well vulnerability. Further, because it is a physically-based quantity, instead of an index, the results are applicable and comparable province-wide. In addition, as a bonus, an intermediate step in the calculation produces maps of potential recharge and discharge areas that can also be used in the vulnerability analysis".

The SAAT model proposes a travel-time calculation in the saturated and non-saturated zone where the porosity and the residual water content are taken into account. The Darcy flux and the effective porosity are used to estimate vertical advective velocity. The advective time, i.e. time taken for a contaminant to travel from the ground surface to the saturated zone of an aquifer is then calculated based on the advective velocity and the thickness of the overlying material.

The equivalent effective porosity has to be evaluated in the calculation of the travel time for both saturated and unsaturated zone (value between the effective porosity and the water content at saturation). For the saturated zone, the equivalent effective porosity (n_e) is equal to the mean total porosity (n) weighted by the thicknesses of the layers. For the unsaturated zone, the model uses the mean between the volumetric water content at saturation and the residual moisture content of the given type of soil.

The potential surface to aquifer advection time is given by (MOE TEC Report, 2004):

$$\mathrm{SAAT} = \frac{\eta D}{q}$$

Where:

SAAT=potential surface to aquifer advection time [T];

η=Equivalent effective porosity of the material overlying the aquifer [L^3 L^{-3}]

D=Thickness to the material overlying the aquifer [L]

q= Potential vertical component of the water flux

(i.e. potential recharge/discharge) [L^3 L^{-2} T^{-1}]

The Equivalent effective porosity can be estimated as the arithmetic mean of the effective porosities of the individual layers above the aquifer, weighted by their respective thicknesses (MOE TEC Report, 2004):

$$\eta = \sum_{i=1}^{N} \eta_i D_i \Big/ \sum_{i=1}^{N} D_i = \sum_{i=1}^{N} \eta_i D_i \Big/ D$$

Where:

i= i^{th} layer overlying the aquifer

N=total number of layers overlying the aquifer

η_i=Effective porosity of layer *i* overlying the aquifer

D_i=Thickness of layer *i*

and

D, the overall thickness to the material overlying the aquifer is:

$$D = \sum_{i=1}^{N} D_i$$

The potential vertical component of the water flux (i.e. potential recharge/discharge) can be estimated from evapotranspiration models or from Darcy's Law (from Michel Robin personals notes, 2004):

$$q = K\,\Delta h / D$$

Where:
q= Potential vertical component of the water flux
(i.e. potential recharge/discharge) [$L^3\ L^{-2}\ T^{-1}$]
K=Equivalent hydraulic conductivity of the material overlying the aquifer [$L\ T^{-1}$]
Δh =Hydraulic head difference between the surface and the aquifer [L]
D=Thickness to the material overlying the aquifer [L]

Finally, as stated in the MOE TEC Report (2004): "If we can assume that the material overlying the aquifer forms uniform layers, then K can be calculated as the harmonic mean of the individual layer's K values, weighted by the layer thickness.

$$K = \left(\sum_{i=1}^{N} K_i^{-1} D_i \bigg/ \sum_{i=1}^{N} D_i \right)^{-1}$$

If the material is more like lenses than layers, then it is more appropriate to use a weighted geometric mean:

$$K = \log^{-1}\left(\sum_{i=1}^{N} \log(K_i) D_i \bigg/ \sum_{i=1}^{N} D_i \right)$$

The vulnerability map is then obtained with the classification of these advective times."

Other models

Finalizing this section is a brief description of some other existing models (GOD, EVARISK, MINNESOTA,) in order to get a better idea of other models available for modelling aquifer/groundwater vulnerability

GOD

As the DRASTIC method, GOD is a parametric empirical method (DRASTIC and GOD represent classic approaches in vulnerability assessment). However, the GOD method requires fewer parameters than the DRASTIC model. This method was developed in England in the late 1980's (Foster, 1987). The GOD index is a method that relates primarily to vulnerability to incoming pollutants in the saturated zone and with no regard to lateral transport within that zone. In accordance with Foster (1987), it is based on the following three factors:

a) **G**roundwater occurrence, *i.e.* whether the aquifer is unconfined, semi-confined, confined, etc. (if there is an aquifer);
b) **O**verall aquifer class in terms of degree of consolidation and lithological character;
c) **D**epth to groundwater table.

Thus the name **GOD** is an acronym resulting from the initials of the first letter of each parameter. The aquifer is classified, as referred to each of the three parameters, in a scale with a maximum value represented by the unit (E. Artuso, M.M. Oliveira and J.P.C. Lobo-Ferreira, 2003). In contrast to DRASTIC, the parameters are not weighted. The vulnerability index is obtained according to the following equation:

GOD index = $R_G * R_O * R_D$ where RG= Aquifer rating; R_O= Lithological rating; R_D= Depth to groundwater rating.

Values for the parameter G of the GOD index

None	0
Artesian confined	0.1
Confined	0.2
Semi-confined	0.3
Semi-confined (covered)	0.5
Unconfined	1

Values for the parameter O of the GOD index

Residual soils	0.4
Alluvial silts	0.5
Aeolian sands	0.6
Alluvial and fluvio-glacial sand-gravel	0.7
Colluvial gravels	0.8
mudstones shales	0.5
Siltstones	0.6
Volcanic tuffs	0.6
Sandstones	0.7
Chalky	0.8
Limestones	0.8
Calcarenites	0.9
Igneous/metamorphic formations	0.6
recent volcanic lava	0.8
calcretes and other limestones	1

Values for the parameter D of GOD index

>100 m	0.4
50-100 m	0.5
20-50 m	0.6
10-20 m	0.7
5-10 m	0.8

2-5 m	0.9
< 2 m	1

The computed GOD index values can be used to generate vulnerability maps. The Table below shows the ranges for each vulnerability class.

Ranges of the GOD Groundwater Vulnerability Index	
0.7 - 1	Extreme
0.5 – 0.7	High
0.3 – 0.5	Moderate
0.1 – 0.3	Low
0 – 0.1	Negligible

EVARISK

EVARISK is an analytical method based on analytical processes, that was developed to estimate the specific vulnerability of groundwater to non-point source pollutants. Like most analytical models EVARISK, can also be used to evaluate intrinsic vulnerability. Certain types of methods such as statistical methods using water quality information and hydrogeological or land management information are integrated to analytical methods because they constitute possible approaches to estimate groundwater vulnerability to pollutants.

The EVARISK model/program was developed in 1998 by the INRS-Eau by Olivier Banton. This program is easy to use and most of the needed information is available within the program through a database connection. EVARISK allows the user to simulate the transport of different compounds (transport of salt de-icing chemicals, migration and breakdown of miscible petroleum hydrocarbons, pesticides and N fertilizers), within a regional management environment. This program was built in order to study processes of natural attenuation and transport of contaminant within the first meter of the subsurface. To evaluate the intrinsic vulnerability, all chemical compounds should be initialized to zero in order to calculate only the quantity of water leached through the soil.

The theory basis of EVARISK is a mechanistic and stochastic model simulating the water flux and the pesticide and nitrogen breakdown pattern in the unsaturated section of the soil. The main methodological assumption of this model is that, from under the first meter of soil down to the water table (aquifer), it is only a matter of water vertical flow processes, without any significant influence of estimated groundwater vulnerability.

When evaluating the water flux reaching an aquifer, the EVARISK model requires only three parameters to be settled by the user (INRS-Eau, 1997):

1. Soil composition for each soil layer
2. Slope evidence
3. Vegetation type and associated volume of water consumption

All others parameters (18) are incorporated in the program and values are pre-settled for each region. The groundwater vulnerability is then defined for each location depending on the amount of water that leaches through the first meter of soil per year. Important amounts of water leached through the soil imply a weak protection of the aquifer. One disadvantage of this piece meal method is that it does not provide vulnerability indices. Vulnerability indices will be generated within the range of values (amount of water leached) obtained for a given region (Murat et al., 2004).

MINNESOTA

This contextual method was developed in the 1990's by the Minnesota Department of Natural Resources. The method identifies areas of sensitivity (vulnerability) using hydrogeological factors that are based on the general knowledge of the geological and hydrogeological setting of the region. Expert judgment is therefore an important source of information in this approach.

The most important step in this kind of method is the combination of different information layers with the superposition of thematic maps representing different parameters to take into account in the groundwater vulnerability modelling. In some cases, factors (parameters) could overlap together and create redundant parameters. For instance, use of aquifer thickness and aquifer permeability as input parameters is redundant if the model already includes aquifer transmissivity as a parameter. This kind of approach is primarily empirical in nature and the level of estimation and the bias could affect significantly the final results.

The MINNESOTA method uses three parameters or layers of information:

1. Soil map,
2. Surface formation map or borehole data; and
3. Water well records or geological maps

The second parameter helps in refining or validating the results from the first layer/parameter within areas where a lot of information is available. The third parameter/layer allows identification and processing using confined aquifer data.

4. RESULTS: MODEL AND ERROR CALCULATION FOR ISI, SAAT AND DRASTIC

This chapter presents model results for DRASTIC, SAAT and ISI (the methods were discussed in detail in the previous chapter) followed by the methods used for propagating the uncertainty (error) through each model. It is important to point out that the method propagates uncertainty from the sources of uncertainty in the physical parameters but without considering geographical uncertainty (uncertainty in geographical location). Furthermore, we focus on the methodologies used to track the propagation of uncertainty through each model rather than focusing on the evaluation of the sources of uncertainty of the components .

The above process produces maps of vulnerability indices and accompanying maps of error (expressed by the Coefficient of Variation, defined below). These maps are difficult to use by Planners and Officials without expert interpretation, and an additional map is proposed for each vulnerability model to help in the decision-making process. This map simply shows areas where classifications could change as a result of uncertainty for a given confidence level.

As mentioned above, we opted to describe quantitatively the uncertainty by the Coefficient of Variation (CV). The standard deviation is another possible way to express uncertainty but, because models use different variables, standard deviation values do not have comparable units, and are therefore not amenable to comparisons between models. To remedy this issue we have calculated coefficients of variation (which are unitless) for each model:

$$CV = \frac{\sigma_i}{\overline{X}}$$

Where $\sigma_i = s(\mathrm{Model\_result}_{grid})$ is the standard deviation, one possible measure of uncertainty, and $\bar{X} = \mathrm{Model\_result}_{grid}$ is the value of the model index at a given location.

The coefficient of variation is the degree to which a set of data points varies (also called the relative error or relative standard deviation).The larger this number is the greater the variability in the data. CV is typically displayed as a percentage.

The presentation of both models and uncertainty calculation results has to be enlightened in the context in which data were generated/obtained in order to understand the conceptual limitations related to the data. In this study, models and associated uncertainties were calculated from datasets already generalized/conceptualized in other studies, in particular:

- Surface geology and groundwater data were taken from the digital database accompanying EOWRMS report (2001).
- The Digital Elevation Model (DEM) was obtained from the Ontario Ministry of Natural Resources' Natural Resource Values Information System (NRVIS).

The conceptual model was borrowed from the EOWRMS:

- The regional aquifer system was conceptualized hydrostratigraphically as an upper overburden layer representing all surface deposits, a middle semi-continuous impermeable layer and a lower overburden aquifer referred to as the contact zone aquifer (consisting of higher conductivity materials, including basal sands and gravels and the fractured portion of the upper bedrock).
- The hydrostratigraphy unit thicknesses (key elements in the vulnerability calculation) were estimated by Robin and Daneshfar (in preparation) for the EOWRMS conceptual model.
- Saturated hydraulic conductivity values of Layer 1 (top layer) [m / mth] were estimated by Daneshfar & Robin (in preparation) from the soils map and from

typical values of hydraulic conductivity for the various soils that are published in the literature (Van der Leeden and Troise, 1991).

- Other sources of information specific to each model are given in the corresponding sections below.

For each of the three models (ISI, SAAT and DRASTIC) the following sections of this chapter link the model calculation formula (presented in previous chapter) to the conceptual model by identifying constant GIS layers (for each layers) used to calculate final results. For comparison purposes, all three methods used the contact zone aquifer (Layer 3) as the target aquifer for vulnerability calculations.

Calculation of the Intrinsic Susceptibility Index (ISI)

The ISI is calculated by summing the product of the thickness of each layer($\mathbf{b_n}$) with a corresponding K-Factor ($\mathbf{Kf_n}$) (MOE, 2001).

The following are constant GIS layers used in the calculations for layer 1 (Robin and Daneshfar, 2005):

b1TOT: Upper overburden thickness including saturated & unsaturated zones (meter).

b1: Thickness of the upper overburden that is saturated (meter)

K1: Saturated hydraulic conductivity of layer 1 (m/sec)

K_ratio: Ratio of unsaturated to saturated hydraulic conductivity for layer 1; to simplify the calculation, a single value of the unsaturated hydraulic conductivity was used, based on the soil type and on the saturated conductivity.

Thickness and K-Factor of layer 1

It was assumed that the surficial geology data (surficial deposit) were found across the entire thickness of layer 1 (b1TOT was greater than zero).

The upper overburden K-Factor was then evaluated from the Generic Representative Permeability (K-Factor) Table (Ontario Ministry of the Environment, 2001) by matching each "geomaterial description" and the associated representative K-Factor from the Generic Representative Permeability (K-Factor) table (See Table 1.0 and Map 1.0). Table 2.0, given in appendix A, presents the "K-Factor value associated to the geology description (upper overburden). The K-Factor values were added to the original surficial geology file (shapefile) and this feature was converted to a raster format using the Spatial Analyst tool.

Map 1.0 presenting the K-Factor values and Map 1.1 presenting the Upper Overburden thickness values are given in Appendix B. The K-Factor raster was then multiplied with the raster b1TOT (EOWRMS) to produce the Intrinsic Susceptibility Index of the Layer 1 (Upper Overburden) presented in Map 1.2 in Appendix B.

GIS layers (data) and ISI calculation for layer 2.

The following are constant GIS layers used in the calculations for layer 2 (Robin and Daneshfar, 2005, in preparation):

bm: Thickness of middle layer (meter)
km: Saturated hydraulic conductivity of middle layer (m/sec)

The saturated hydraulic conductivity for the semi-impermeable layer (clay) was established at $1X10^{-9}$ m/sec everywhere. The K-Factor value is 9 (log $1/1X10^{-9}$). Map 1.3 presenting the ISI values for the layer 2 and map 1.4 presenting the ISI values for the lower overburden aquifer (Layer 3: first significant aquifer) by summing the results of Layer 1 (b1TOT) & 2 (bm) are given in appendix B.

Calculation of the error (uncertainty) for the Intrinsic Susceptibility Index (ISI) Model.

The following development for the ISI uncertainty calculation and analysis is adapted from the "Renfrew County- Mississipi- Rideau Groundwater Study", Golder Associates Ltd, Final Report, 2003. The overall approach consists of (1) estimating the errors for the thickness and K-factor for each hydrostratigraphic unit; (2) calculating the error in the ISI value for each layer using the Error Propagation Theorem; and finally (3) calculating the error for the overall ISI index, again using the Error Propagation Theorem. The calculation involves the following variables:

b:	Thickness of EOWRMS layer
Kf:	K-Factor assigned by MOE
n:	Counter of number of layers from 1 to k
μ:	Mean layer thickness of population
α:	Relative error (%) of the layer thickness
σ :	Population variance of the layer thickness
s:	Sample variance of the layer thickness
$\mathrm{var}(b_n)$:	Variance in thickness of layer b_n
ISI_n:	Intrinsic Susceptibility Index for layer n
$\mathrm{var}(ISI_n)$:	Variance in ISI for layer n
$\mathrm{var}(Kf_n)$:	Variance of K factor for layer n

ISI_{grid} : Intrinsic Susceptibility Index at each grid

$\mathrm{var}(ISI_{grid})$: Variance in ISI at each well

Step 1: Estimate variances in bn, Kfn

For simplicity, we assume that the relative error is constant; which is typical of log-normally distributed variables. All calculations are carried out with the variance, so we need to estimate the variance from a reasonable range of values. If we assume that the constant relative error is based on the 95% confidence interval, then we can use the relative error to back calculate the variance as follows: The relative error (α) is given by:

$$\alpha = \frac{2\sigma}{\mu}$$

and estimated by:

$\alpha = \frac{2s}{\mu}$, which is estimated and assumed constant.

If the mean thickness of layer *n* is b_n, then the variance can be estimated by:

$$\mathrm{var}(b_n) = \left(\alpha b_n / 2\right)^2 \quad \textbf{(1)}$$

We assume that a reasonable value for the relative error (α) is 10% for the EOWRMS data set. If, in addition, we assume that it is independent of scale, then:

$$\mathrm{var}(b_n) = \left(\frac{0.1}{2}\right)^2 b_n^{\ 2} = 0.0025 b_n^{\ 2}$$

At the scale of this study, the hydraulic conductivity, K, can safely be assumed to be log-normally distributed, which translates into a constant relative error (α) on a linear scale or a constant absolute error on a log scale. Since the ISI K-factors are on a log scale, their absolute error will be constant. We assume that the K-factor can reasonably vary over +/- 1 unit at the 95% confidence level and therefore,

$2s = 1$ And thus $\text{var}(Kf_n) = \left(\frac{1}{2}\right)^2 = 0.25$ **(2)**

Step 2: Calculate variance of ISI for each layer

The ISI for each layer n is given by:

$$ISI_n = b_n Kf_n$$

If we assume that Kf_n is not correlated to bn then the variance can be obtained from the propagation of error theorem:

$$\text{var}(ISI_n) = \left(\frac{\partial ISI_n}{\partial b_n}\right)^2 \text{var}(b_n) + \left(\frac{\partial ISI_n}{\partial Kf_n}\right)^2 \text{var}(Kf_n)$$

Since,

$$\frac{\partial ISI_n}{\partial b_n} = Kf_n$$

And

$$\frac{\partial ISI_n}{\partial Kf_n} = b_n$$

We obtain,

$$\text{var}(ISI_n) = K_n^2 \text{var}(b_n) + b_n^2 \text{var}(Kf_n)$$

Substituting **(1)** and **(2)** into this equation produces:

$$\text{var}(ISI_n) = (\alpha^2 Kf_n^2 b_n^2 / 4) + (b^2{}_n \ / 4)$$

Simplifying:

$$\text{var}(ISI_n) = \left(b^2/_4\right)(\alpha^2 Kf_n^2 + 1)$$

Where α is the relative error on the thickness of the layer b_n.

Step 3: Calculate variance of ISI for each grid

For k layers, the ISI at the grid is:

$$ISI_{grid} = \sum_{n=1}^{k} ISI_n = \sum_{n=1}^{k} b_n kf_n$$

Since the ISI values are simply added up, the variances will also simply add up, based on the propagation of error theorem:

$$\mathrm{var}(ISI_{grid}) = \sum_{n=1}^{k} \mathrm{var}(ISI_n)$$

Map 1.5, given in Appendix B, presents the variance values for the ISI model. The following formula was used to calculate these values:

$$\mathrm{var}(ISI_{grid}) = \left[\left(b1TOT/4\right)\times\left(0.1^2 \times b1TOT_{kf} + 1\right)\right] + \left[\left(bm/4\right)\times\left(0.1^2 \times 9 + 1\right)\right]$$

The coefficient of variation is then calculated as follows

$CV = \frac{\sigma_i}{\bar{X}}$ Where $\sigma_i = s(ISI_{grid})$ and $\bar{X} = ISI_{grid}$

$$CV\% = \frac{\sigma_i}{\bar{X}} \times 100$$

Map 1.6, presenting the coefficient of variation for the ISI model, is given in Appendix B.

Calculation of the Surface to Aquifer Advective Time (SAAT)

The potential surface to aquifer advection time (SAAT) is given by (MOE TEC Report, 2004):

$$SAAT = \frac{\eta D}{q}$$

Where:

SAAT	=	potential surface to aquifer advection time [T];
η	=	Equivalent effective porosity of the material overlying the aquifer $[L^3 \quad L^{-3}]$
D	=	Thickness to the material overlying the aquifer [L]
q	=	Potential vertical component of the water flux (i.e. potential recharge/discharge) $[L^3\ L^{-2}\ T^{-1}]$

The following are additional GIS layer (see other ones within the ISI section) generated by Daneshfar & Robin (in preperation) or accompanying the EOWRMS reports (2001):

$sy1$ = Specific yield for Layer 1 [m3/m3] (represents the Volume of water that can be drawn or injected into the aquifer per unit volume of aquifer per unit change in hydraulic head), EOWRMS (2001).

$n1$ = Soil porosity of layer 1 (Esker=0.3, Peat=0.9, Clay=0.45, Sand=0.35, Till=0.35) estimated by Daneshfar & Robin (in preparation)

n_e = Effective porosity of the saturated zone estimated from Banton & Bangoy (1999).

nm: Porosity of the semi-permeable layer estimated estimated by Daneshfar & Robin (in preparation) = 0.45

infiltration: Flux of water across the ground surface [meter / sec]; if it is positive then we have groundwater discharge, and if it is negative then we have groundwater recharge (EOWRMS, 2001)

Advective Time calculation for Layer 1 (unsaturated zone & saturated zone) and layer 2 (semi-impermeable layer).

Layer 1: Unsaturated zone

The equivalent effective porosity of the unsaturated zone n was assumed to be approximately equal to field capacity (a value somewhat higher than the residual moisture content), which was approximated by the difference between the porosity and the specific yield :

n = porosity$(n1)$ - specific yield $(sy1)$

The thickness of the unsaturated zone was the difference between the total overburden thickness, b1TOT, and the saturated thickness, b1. The latter was estimated from the water table elevation by Robin and Daneshfar (in preparation):

D = Thickness of unsaturated zone (b1unsat) = b1TOT – b1 (meter)

The vertical flux was estimated as the infiltration rate, based on the evapotranspiration model presented in Robin and Daneshfar (in preparation):

q = Infiltration (meter/sec).

Then the advective travel time in the unsaturated zone can be estimated as follows:

$$SAAT_{b1unsat} = \frac{(n_1 - Sy_1) * b1TOT - b1}{Infiltration} \quad (1)$$

Layer 1: saturated zone

The vertical advective time in the saturated portion of Layer 1 was calculated from:

$$AT_{b1} = \frac{n_e * b1}{Infiltration} \quad (2)$$

where:

n = Effective porosity (n_e)

D = Thickness of saturated zone = b1 (meter)

and,

q = Infiltration (meter/sec).

Advective Time calculation for Layer 2: Middle Layer (Semi-Impermeable layer: clay)

The advective time through the middle layer was calculated from:

$$AT_{bm} = \frac{n_m * bm}{Infiltration} \quad (3)$$

where:

n = porosity (n_m) = 0.45 (assumed porosity for fine materials such as clay)

D = Thickness of middle layer = bm (meter)

and,

q = Infiltration (meter/sec).

<u>Surface to Aquifer Advective Time calculation</u>

The overall advective time from the surface to the aquifer (top of the contact zone aquifer) is the sum of the advective times in the layers above. Summing equations (1), (2) and (3) produces the SAAT values at each grid location:

$$SAAT_{grid} = \frac{[(n_1 - Sy_1) * b1TOT - b1] + n_e * b1 + nm * bm}{Infiltration}$$

The maps used to produce these calculations are given in Appendix B:

Map 2.0 : porosity values (n1) of layer 1 (Upper Overburden);
Map 2.1 : Specific Yield values of Layer 1;
Map 2.2 ; effective porosity of unsaturated zone of layer 1;
Map 2.3 : thickness of the unsaturated zone of layer 1;
Map 2.4 ; the thickness of the saturated zone of layer 1; and
Map 2.5 : annual infiltration (recharge) values (meter/year)

The results are also presented in Appendix B, in Maps 2.6.1 and 2.6.2 giving SAAT values using two different classifications.

Calculation of the error (uncertainty) for Surface to Aquifer Advective Time (SAAT) Model.

The uncertainty in the SAAT values was estimated by the variance and the coefficient of variation (given by the ratio of the variance to the mean). The SAAT variance was estimated by propagating the variances of the individual components of the above equation using the Propagation of Error Theorem. The underlying assumptions in this application are that the components are log-normally distributed and uncorrelated.

The variance components of the SAAT variance were estimated as follows:

Porosity:

Effective porosity values for the saturated zone of layer 1 were estimated from Banton & Bangoy (1999). Typical range of values for each geomaterial was established from different references (Banton & Bangoy, 1999; Freeze & Cherry, 1979; McWorther & Sunada, 1977). Tables presenting geomaterial descriptions with associated porosity, effective porosity & specific yield values used for the SAAT calculation (Table 2.1) and typical range values and associated variance values to be used in the uncertainty/error modeling (Table 2.2) are given in Appendix A.

The variance of the porosity values was approximated as the mean squared difference between the selected likely value and the minimum and maximum of a plausible range of values. This variance component was usually small relative to the other components.

$$\text{var (porosity)} = \frac{(\text{min} - \text{likely})^2 + (\text{max} - \text{likely})^2}{2}$$

Map 2.7 of Appendix B presents the porosity variance values $\mathrm{var(n)}$ and Map 2.8, the effective porosity variance values $\mathrm{var}(n_e)$ of geomaterials.

Thickness:

The variance for the thickness of each layer was estimated in the same way as for the ISI model, presuming a typical error of 10% and back calculating the variance.

$\mathrm{var(b_n)} = \left(\frac{0.1}{2}\right)^2 \mathrm{b_n}^2 = 0.0025\mathrm{b_n}^2$ (See calculation details in the ISI section)

Vertical Flux:

The vertical flux can be estimated directly from field measurements, or inferred from Darcy's Law. The EOWRMS (2001) had already produced maps of inferred flux from Darcy's law, which we used directly. The infiltration variance was calculated from the various components of Darcy's Law. It is important to point out that the error produced by this method is much larger than the error that would result from direct measurements; it is therefore highly recommended that these measurements be made directly in the field.

Darcy's Law gives the magnitude of the vertical flux:

Darcy flux = $\mathrm{Kz(dh/dz)}$

Where,

Kz = **Vertical hydraulic conductivity of the confining unit:** Taken directly from the EOWRMS (2001).

$\Delta\mathrm{h}$ = **Vertical hydraulic head difference:** Calculated from the water table elevations in the overburden aquifer and the piezometric surface in the Contact Zone Aquifer.

$\Delta\mathrm{z}$ = **Thickness of the confining unit**: Taken directly from the EOWRMS (2001).

Since Darcy's Law only involves products and quotients, it can be shown that the theorem of propagation of error simply states that the coefficient of variation (CV) or relative error, of the flux will be equal to the quadratic sum of the CV's of the components:

$$CV_q = \sqrt{(CV_{Kz})^2 + (CV_{\Delta h})^2 + (CV_D)^2} = \frac{S_q}{\bar{q}}$$

Assuming that the three components of the flux are log-normally distributed then their respective CV's are constant and the relative errors are constant. For the purpose of this exercise we assumed a relative error of 10% for each component, resulting in a CV of 17% for the infiltration flux. The variance of q was then back-calculated from the CV. This variance contribution was the largest in the overall SAAT variance.

The coefficient of variation is then calculated as follow:

$\mathrm{CV} = \frac{\sigma_i}{\bar{\mathrm{X}}}$ Where $\sigma_i = s(\mathrm{SAAT_{grid}})$ and $\bar{\mathrm{X}} = \mathrm{SAAT_{grid}}$

$$\mathrm{CV\%} = \frac{\sigma_i}{\bar{\mathrm{X}}} \times 100$$

Map 2.9 in Appendix B presents the SAAT's variance values and map 2.10 the coefficient of variation values.

Calculation of the DRASTIC model

The DRASTIC index (ALLER *et al.*, 1987) corresponds to the weighted sum of 7 values corresponding to the following 7 parameters or hydrogeological indicators:

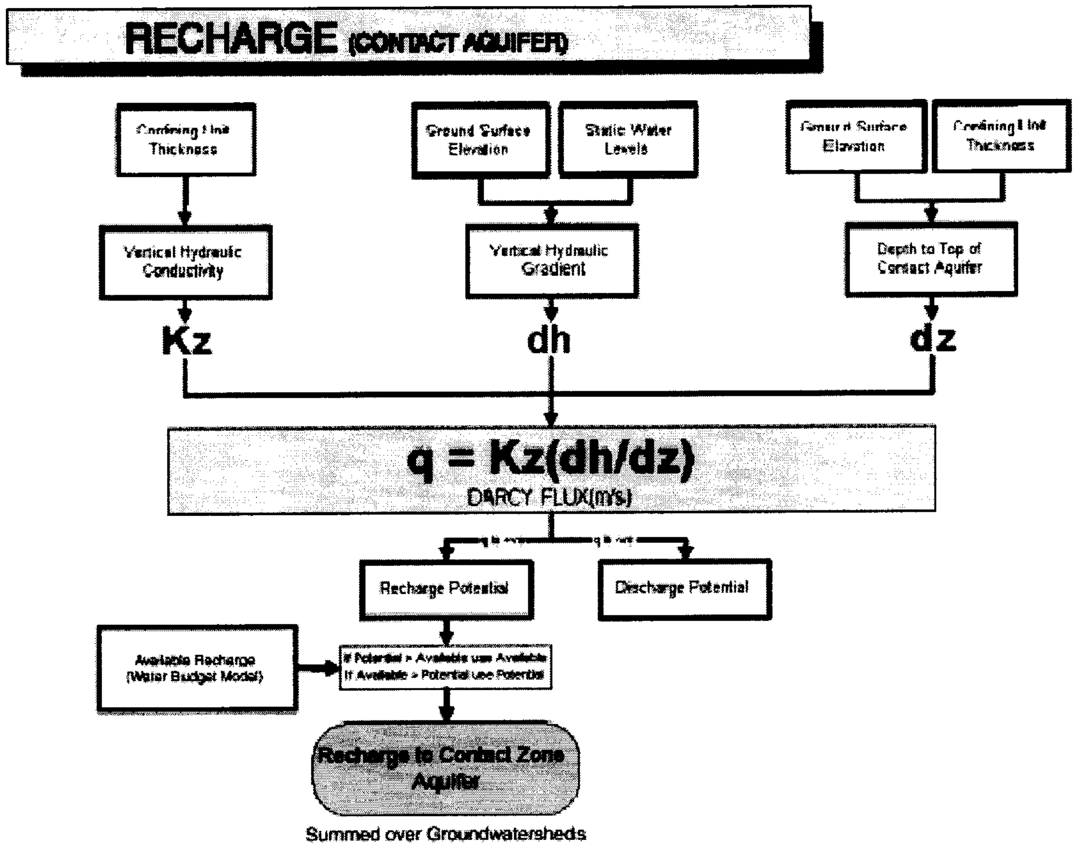
1. **D**epth to the water table
2. **R**echarge
3. **A**quifer material/media
4. **S**oil type
5. **T**opography
6. **I**mpact of the unsaturated (vadose) zone
7. **C**onductivity (Hydraulic conductivity)

A value (rating) from 1 to 10 (regrouping a range of "real" values or from a nominal scale) is attributed to every parameter. High values correspond to a higher vulnerability according to the characteristic of the environment. Each parameter has a pre-defined weight consequent to its relative importance the quantification of vulnerability. The most important parameters have weight 5 and the less important have weight 1.

Depth to the water table: DRASTIC uses Depth to the water table for this rating because it assumes that the aquifer of interest is the shallowest, phreatic aquifer. In this study however we interpret D as the depth to the aquifer of concern, that is the contact zone aquifer. D was therefore calculated by summing the thickness of the upper overburden (b1TOT) layer and the middle layer (bm). Table 3.0, presenting the range and rating values, is given in Appendix A and Map 3.0, presenting the depth to water table ratings, is given in Appendix B.

Net **recharge to the water table** data were taken from the EOWRMS data: Recharge to the Contact Zone Aquifer.

Recharge to the Contact Zone Aquifer was derived from the aquifer mapping, the water budget analysis, and the water well records (EOWRMS, 2001). The following figure, taken from the EOWRMS report (2001), presents/summarizes the process of estimating groundwater recharge to the contact zone aquifer (layer 3: Lower Overburden).



(source: EOWRMS, 2001)

Table 3.1 presenting the range and rating values is given in Appendix A, and Map 3.2 presenting the recharge ratings, is given in Appendix B.

The **aquifer media** composition was taken from the Aquifer Contact Zone raster accompanying the EOWRMS report defining areas where lower overburden (sand and gravel aquifer) or bedrock (the upper zone of bedrock) aquifers are found. Aquifer areas located in the fractured bedrock were assigned a rating of 7/10 while aquifers composed of sand and gravel were assigned a rating of 9/10.

Table 3.2, presenting the range and rating values, is given in Appendix A; and Map 3.2, presenting the aquifer media ratings, is given in Appendix B.

Surficial geology/deposits (**Soil type**) data were taken from the digital database accompanying EOWRMS report (2001). Each soil type was assigned a fixed rating predetermined by the DRASTIC model. Table 3.3, presenting the range and rating values, is given in Appendix A, and Map 3.3, presenting the soils (surficial deposits) ratings, is given in Appendix B.

The percentage slope incidence (**topography**) was generated from the DEM (Digital Elevation Model) taken from the digital database accompanying EOWRMS report (2001). The slopes (%) were derived from the DEM with the Surface Analysis function of the Spatial Analyst and reclassified/rated with Spatial Analyst. Table 3.4, presenting the range and rating values, is given in Appendix A; and Map 3.4, presenting the slope ratings, is given in Appendix B.

The **Impact of the unsaturated zone** (vadose zone) was generated from the surficial geology/deposits (Soil type) taken from the digital database accompanying EOWRMS report (2001). Table 3.5, presenting the range and rating values, is given in Appendix A, and Map 3.5, presenting the vadose zone ratings, is given in Appendix B.

The **Hydraulic Conductivity** (Layer 1: Upper Overburden) were estimated by Daneshfar & Robin (in preparation) from the soils map (taken from the digital database accompanying EOWRMS report, 2001) with typical values of hydraulic conductivity for the various soils (Van der Leeden and Troise, 1991). Table 3.6, presenting the range and rating values, is given in Appendix A, and Map 3.6, presenting the hydraulic conductivity ratings, is given in Appendix B.

Finally, the seven partial products are added for each given location to create the DRASTIC indices with a minimum possible value of 23 and maximum possible value of 226. The DRASTIC index at each location can be calculated, as:

DRASTIC INDEX = $D_R*D_W + R_R*R_W + A_R*A_W + S_R*S_W + T_R*T_W + I_R*I_W + C_R*C_W$

where R=rating and W=weight.

The DRASTIC indices are regrouped into thematic classes underlining different levels of vulnerability from low (less than 79) to high (more than 200). MAP 3.7, presenting the DRASTIC Pollution Potential Index Range, is given in Appendix B.

Calculation of the error (uncertainty) for DRASTIC Model.

The uncertainty in the DRASTIC index at each location can be calculated from the error propagation theorem, which gives, for the variance in the DRASTIC Index:

$$Var(DRASTIC) = d^2.(S^2_D)+r^2(S^2_R)+a^2.(S^2_A)+ t^2.(S^2_S)+i^2(S^2_T)+ i^2.(S^2_i)+ c^2(S^2_C)$$

Where the lower case letters, d, r, a, s, t, i, c refer to the weighting values and the upper case indices $_{D, R, A, S, T, I, C}$ refer to the rating values.

Variance estimation for each rating was based on the range of reasonable values for the physical parameter upon which the rating is based. This range was assumed to represent the 95% confidence interval. The impact of this uncertainty on the corresponding rating variance depended on the rating and how it was coded in the DRASTIC methodology.

For **D**epth to the water, **R**echarge, **T**opography and Hydraulic **C**onductivity, the DRASTIC ratings are ascribed by classifying the physical parameter on a discrete ordinal scale. For these factors the variance was set to zero except near the class limits, where uncertainty in the physical parameter can cause the DRASTIC rating to be in one class or the adjacent one. If the physical parameter value fell within one confidence interval of the class limit its variance was set equal to the difference between the rating values of the two adjacent classes.

Example: **D**epth to the water values between 1.5 and 4.5 meters have a predetermined rating value of 9 and values between 4.5 and 9 meters have a predetermined rating value of 7. Assuming a 95% confidence interval of ± 1 meter depth to water, then values of depth to water within ± 1 meter of the class limit of 4.5 were ascribed a variance value of 2 (= 9-7), and a value of zero elsewhere .

Table 3.7 and Map 3.8, presenting the variance values of **D**epth to the water values, are given respectively in Appendix A and B; Table 3.8 and Map 3.9, presenting variance values of **R**echarge ($\pm 10mm/yr$), are given respectively in Appendix A and B; Table 3.9 and Map 3.10, presenting variance values of **T**opography ($\pm 2units$), are given respectively in Appendix A and B; and Table 3.10 & Map 3.11, presenting variance values of Hydraulic **C**onductivity ($\pm 10\%$), are given respectively in Appendix A and B.

For **A**quifer media and **I**mpact of vadose zone, the DRASTIC method specifies a range of ratings and a likely value. For each of these factors, the mean squared difference between the likely value and the limits of the range was used to approximate the variance.

Example: **A**quifer media consisting of sand (sand plain) has a rating's range varying between a minimum of 4 and a maximum of 9; with a predetermined typical value of 8. For the sand the A rating variance was calculated as:

$$\text{var(sand)} = \frac{(\text{min} - \text{likely})^2 + (\text{max} - \text{likely})^2}{2} = \frac{(4-8)^2 + (9-8)^2}{2} = 8.5$$

Table 3.11 and Map 3.12, presenting the variance values of **A**quifer media, are given respectively in Appendix A and B; Table 3.12 and Map 3.13, presenting variance values of **I**mpact of vadose zone, are given respectively in Appendix A and B.

Finally, for **S**oil type, the variance was estimated from the range of DRASTIC ratings that could reasonably be inferred from the geomaterial layer (by matching each "geomaterial description" and the associated representative range of rating values proposed by the DRASTIC method).

Table 3.13 and Map 3.14, presenting variance values of soil type, are given respectively in Appendix A and B.

The DRASTIC's (potential pollution) variance is then calculated from the equation given above and presented in Map 3.15. given in Appendix B. Map 3.16, presenting the Coefficient of variance of the DRASTIC's potential pollution, is given in Appendix B.

5. DISCUSSION

From a **qualitative** comparison of the three models, we can underline the main limitations of each one.

ISI limitations:

- For a semi physically based approach, too many parameters are ignored. For instance the hydraulic gradient, the porosity and the water content of the porous media.
- The estimates of K-factor for various sediment types used in the ISI method are poor estimation. In reality, K may vary by several orders of magnitude for each sediment type (Freeze and Cherry, 1979)

- As in the AVI method, the ISI method does not consider lateral continuity or discontinuity of aquifers rigorously. Two vertical overlapping vulnerable aquifers would also appear as one continuous zone of high vulnerability (Lobo-Ferreira, 2001).

DRASTIC limitations:

- Semi-confined or confined aquifers corresponding to the presence of clay layers are not properly evaluated.
- The weight attributed to the Impact of vadoze zone parameter/variable (5) is too high for the study region where impact of vadoze zones are generally weak mainly due to northern climatic conditions.
- There is too much subjectivity involved in the choice of the ranking when definitions of input data are complex (for instance geomaterial description of surficial data).

SAAT limitations (they also apply to ISI) :

- This model does not take into account the sorption properties of chemical contaminants assuming that every contaminant moves with water (advective transport). Sorbing compounds will move with water but at a slower velocity (retardation). A retardation factor must be estimated for sorbing contaminants.
- This model does not take into account the groundwater horizontal flow components in the saturated and unsaturated zones.
- The distribution of moisture in the vadose (unsaturated) zone is poorly integrated in the model equation.

Quantitatively, all three models present comparable results with general similarities in assessing the vulnerability of regional aquifers. Because DRASTIC's model takes into account the topography (higher are the slope values, lowest are the infiltration values and then the vulnerability is decreased) the most notable difference is between DRASTIC results and the other two models‘ results within zones/sectors with the presence of relatively important slopes.

At first glance, the most notable things in the different vulnerability models' error/uncertainty results are the facts that the most accurate physically based approach model in terms of quality and quantity of physical parameters utilized, SAAT, is also the one with the highest values of coefficient of variation (a statistical measure of the deviation of a variable from its mean) and the less accurate model of this type (ISI) is the one with the lowest values of coefficient of variation. This observation arises from an interesting paradox: simple models have fewer parameters and therefore fewer sources of error, but they produce less realistic (accurate) results; conversely, more complex models have more sources of error but are a better representation of reality. Because more input information must be taken into account and then more output information must be sorted through in the SAAT model, the model has more apparent uncertainty even if it's the most realistic model. As stated by Wauchope (1992): "An ironical paradox arises: the more realistic models have less credibility. The less the users know of the subtleties involved, the more suspicious he/she becomes of the models as their difficulty increases".

The DRASTIC model presents the highest amount of uncertainty because of the vagueness/subjectivity involved in the choice of the ranking when definitions of input data are complex (for instance soils description: Table 3.13, Appendix A) and because of the wide range of possible ratings for a given rating to be assigned (for instance Recharge values: Table 3.8 in Appendix A).

Map 4.0 presenting the ISI's, DRASTIC's and SAAT's classified map of Coefficient of variation demonstrate clearly this observation (Appendix B).

Having calculated the mean CV for each map of CV (see equations below), we can affirm that the method with the lowest amount of variability (uncertainty) is the ISI model (2%), followed by the SAAT model (4%) and the DRASTIC model (14%).

$$MeanCV_{ISI} = \frac{\sqrt{MeanVar_{ISI}}}{MeanISI_{index}} * 100 * = 2\%$$

$$MeanCV_{DRASTIC} = \frac{\sqrt{MeanVar_{DRASTIC}}}{MeanDRASTIC_{index}} * 100 = 14\%$$

$$MeanCV_{SAAT} = \frac{\sqrt{MeanVar_{SAAT}}}{MeanSAAT_{index}} * 100 = 4\%$$

Values of mean CVs only depict the amount of variability between a model and its output/results' uncertainties. To evaluate/compare together the reliability of each model, first we must assess the quality of the models themselves. This is why, when analysing the uncertainty of the models' outputs, we should only compare each model with its own uncertainty/variance. The use of uncertainty analysis to judge the quality of a model relative to other models is therefore ill-advised.

The last step in this error analysis process is to produce maps that are more useful from the practical stand point of the planner. If we compare each model with its own uncertainty/variance, we can evaluate where, when added to the vulnerability index, the standard deviation (we assume that the values can vary over ± 1 standard deviation) make the vulnerability' classification, for a given model, change from low to medium, low to high or medium to high. This produces a map showing areas where classifications could change as a result of uncertainty, for a given confidence level. Map 5.0, presenting the DRASTIC vulnerability index classification's modification after adding or subtracting the standard deviation, is given in Appendix B.

The following map (map 6.0, Appendix B) presents the zones where DRASTIC vulnerability can jump from low vulnerability to medium vulnerability and from medium to high vulnerability when the standard deviation is added (and vice versa). To compare DRASTIC result with other models' results, the 9 classes of the DRASTIC vulnerability's model were regrouped under three classes (top 3, middle 3 and bottom 3) to create HIGH , MEDIUM and LOW vulnerability.

The same process was used with SAAT. Map 7.0, presenting the SAAT vulnerability index classification's modification after adding or subtracting the standard deviation, is given in Appendix B and the following map (Map 8.0) presents the zones where SAAT vulnerability can jump from low vulnerability to medium vulnerability and from medium to high vulnerability when the standard deviation is added (and vice versa). The 5 original classes were regrouped as follow:
0 – 10 years (HIGH)
10 – 100 years & 100 – 1000 years (MEDIUM)
1000 – 10,000 years & 10,000 and more (LOW).

Due to the low amount of uncertainty in the ISI modelling, relatively few pixels present enough classification modification to have significant increase in the vulnerability index.

With these type of analytical results taking into account the uncertainty in the results, from the planner's perspective, this approach is a good decision-making tool. With these maps of uncertainties showing classification modification, a decision can then be made to make further field measurements in areas of high uncertainty depending on the risk assessment.

6. CONCLUSION

"Geographical databases have the appearance of precision, the Human context in which they are created is often extremely vague"

(Michael Goodchild, Uncertainty In Geographical Information, 2002)

The most important conclusion of this study is that, for the following reasons, uncertainty should NOT be used as a basis of comparison between vulnerability models:

- Simple (or simplistic) models offer low uncertainty BUT low usefulness.
- As models increase in complexity the representation of the physical reality can improve (accuracy can improve) but the number of sources of uncertainty also increases (precision decreases).
- Example: mean Coefficient of Variation (CV)

$$\text{MeanCV} = \frac{\sqrt{\text{MeanVar}}}{\text{Mean}} * 100$$

ISI model, mean CV= 2%

SAAT model, mean CV= 4%

DRASTIC model, mean CV= 14%

Furthermore, values of mean CVs only depict the amount of variability between a model and its output/results' sources of uncertainty. To evaluate/compare together the reliability of each model, first we must assess the quality of the models themselves. This is why, when analyzing the uncertainty of the models outputs, we should only compare each model with its own uncertainty/variance. Vulnerability models should be compared based on their ability to represent physical reality. Ideally, several vulnerability indices should be used, compared, and areas showing discrepancy should be explained

If we compare each model with its own uncertainty/variance, we can evaluate where, when added to the vulnerability index, the standard deviation make the vulnerability' classification, for a given model, change from low to medium, low to high or medium to high (see maps mentioned above for DRASTIC and SAAT). Consequently, this is the final step of the process of evaluating the uncertainty of a groundwater vulnerability model. In other words, this consists of producing vulnerability maps that are classified and that are accompanied with a map showing areas where classifications could change as a result of uncertainty, for a given confidence level.

The assessment of groundwater vulnerability should always be presented with the associated range of uncertainties, which are an integral part of physical data. We should also stress the need/importance for improvement in field test/sampling for validating such models.

The database used in this study is the same used in the Eastern Ontario Water Resources Management Study (EOWRMS). This database is principally issued from water well drillers' logs. These records vary in quality and descriptive terminology and their quality remain generally poor (they were not always validated by geologists and/or engineers). While continuing work on vulnerability assessment/modeling in order to reduce the risk of contamination of water supply, the process of modeling and developing an understanding of hydrogeologic systems, including field-validation process, should be emphasized. For instance, there is a need for an in-depth characterization of hydraulic properties and a scale effects investigation of regional aquifers of Eastern Ontario within a groundwater resources assessment project. Furthermore, we should stress the need for better development of methods to account for heterogeneity (in all hydrogeological related processes), in order to represent it and model it more accurately (De Marsily et al., 2005).

REFERENCES

Agumya, A. and Hunter. G.J.

1999 : Accuracy (Re)assurance : Assessing "Fitness for Use" of Geographic Information: What Risk Are We Prepared to Accept in Our Decisions?, Third International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences in Québec City, Ann Arbor Press Chelsea, Michigan, pp. 35-45

Alesheikh, A.A., Blais, J.A.R., Chapman, M.A. and Karimi, H.

1999: Rigorous Geospatial Data Uncertainty Models for GISs, Third International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences in Québec City, Ann Arbor Press Chelsea, Michigan, pp. 195-203

Allan, R.C.

1999: Characterizing Local Spatial Uncertainty in the Optimization of Thematic Class Areas, Third International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences in Québec City, Ann Arbor Press Chelsea, Michigan, pp. 105-113

Aller, L., Bennet, T., Lehr, J.H. and Petty, R.J.

1987: "DRASTIC: a standardized system for evaluating groundwater pollution potential using hydrogeologic settings", U.S. EPA Report 600/2-85/018.

Artuso, E., Oliveira, M.M., Lobo Ferreira, J.P.C.

2003: Assessment of Groundwater Vulnerability to Pollution Using Six Different Methods: AVI, GOD, DRASTIC, SI, EPPNA and SINTACS ? Application to the Évora Aquifer". Relatório 184/02 GIAS/DH, Laboratório Nacional de Engenharia Civil, Lisboa, Junho/2002, 77 pp.

Bailey, R.G.

1988: Problems with using overlay mapping for planning and their implications for geographic information systems. Environmental Management, 12, pp. 11-17.

Blake, R.G.

1996: Groundwater Cleanup Using Hydrostratigraphic Analysis. Science & Technology Review January/February, pp. 7-15.

Banton, O. and Bangoy, M.,

1999 : Hydrogéologie multiscience environnementale des eaux souterraines, Québec : Presses de l'Université du Québec

Bélanger, J.R.

2001: Urban Geology of the National Capital Area, Beta Released

Bohnam-Carter, G.F.

1994: Geographic Information Systems for Geoscientists: Modelling with GIS. Pergamon Press, Oxford, 398 pp.

Bolstad, P., Gessler, P., and Lillesand, T.M.

1990: Positional uncertainty in manually digitized map data. International Journal of Geographical Information Systems, 4, pp. 399-412.

Brassel, K.E. and Weibel, R.

1988: A review and conceptual framework of automated map generalization. International Journal of Geographical Information Systems, 2, pp. 229-244..

Brouyère, S., Jeannin, P-Y., Dassargues, A., Goldscheider, Popescu, I-C., Sauter, M., Vadillo, I. and Zwahlen, F.

2001: Evaluation and validation of vulnerability concepts using a physically based approach, Working Group I of the COST 620 ACTION

Champagne, L. et Chapuis, RP.

1993 : Évaluation de la vulnérabilité à la pollution des formations aquifères de la MRC de Montcalm selon la méthode DRASTIC. Rev Sc Tech Eau; 26 : 169-76

Chilès, J.-P. and Delfiner, P.

1999: Geostatistics: Modeling Spatial Uncertainty, Wiley series in probability and statistics.

Chorley, R.J. and Haggett, P.
1968: Models in geography (London:Methuen)
Cressie, N.
1991: Statistics for spatial data, New York, Wiley.
Davis, J.C.
1973: Statistics and Data Analysis in Geology, Kansas Geological Survey, John Wiley & Sons Inc.
De Marsily et al.
2005 : Dealing with spatial heterogeneity. Hydrogeology Journal, vol. 13 no. 1, pp. 161-183
Douglas, D.H. and Peucker, T.K.
1973: Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. The Canadian Cartographer, 10, pp. 112-122.
Freeze, R.A. and Cherry, J.A.
1979: Groundwater. Prentice-Hall, Inc., New Jersey.
Fortin, M.
1998 : Analyse de la propagation des erreurs spatiales induites par l'intégration des données multisources dans le modèle de vulnérabilité de la nappe d'eau souterraine DRASTIC. Mémoire présenté à la Faculté des études supérieures de l'Université Laval pour l'obtention du grade de maître ès Sciences (M.Sc.)
Fortin, M., Edwards, G. and Thomson, K.P.B.
1999: The Role of Error Propagation for Integrating Multisource Data within Spatial Models: The Case of the DRASTIC Groundwater Vulnerability Model, Third International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences in Québec City, Ann Arbor Press Chelsea, Michigan, pp. 437-445
Gadd, N.R.
1987: Geological setting and Quaternary deposits of the Ottawa region, in Quaternary Geology of the Ottawa Region, Ontario and Québec, Fulton, R.J. (ed.), Geological Survey of Canada, paper 86-23
Gélinas, R.J.
1996: Groundwater Modeling: More Cost-Effective Cleanup by Design. Science & Technology Review, pp. 13-22.
Godfrey, S., and Smith, M.
2005: Improved microbial risk assessment of groundwater. Hydrogeology Journal, vol. 13 no. 1, pp. 321-324
Golder Associated Ltd
2003: Renfrew County – Mississipi – Rideau Groudwater Study, final report.
GSC web page (Quebec)
(http://www.cgq-qgc.ca) Comparison of four methods (EVARISK, GOD, DRASTIC and Minnesota)
Gottsegen, J., Montello, D. and Goodchild, M.
1999: A Comprehensive Model of Uncertainty in Spatial Data, Third International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences in Québec City, Ann Arbor Press Chelsea, Michigan, pp. 175-183
Goodchild, M.F.
1989: Modelling error in objects in fields. In *Accuracy of Spatial Databases*, edited by Goodchild, M.F. and Gopal, S. (Basingstoke: Taylor and Francis), pp 107-113.
Goodchild, M.F. and Zhang, J.
2002: Uncertainty in Geographical Information, London: Taylor and Francis.
Goovaerts, P.
1999: Combining Minimum Error Variance and Spatial Variability in the Mapping of Environmental Variables, Third International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences in Québec City, Ann Arbor Press Chelsea, Michigan, pp. 299-309
Gratton, Y.
2002 : Le krigeage : la méthode optimale d'interpolation spatiale. Les articles de l'Institut d'analyse géographique (www.iag.asso.fr)

Griffith, D.A., Haining, R.P. and Arbia, G.
1999 : Accuracy (Re)assurance : Uncertainty and Error Propagation in Map Analyses Involving Arithmetic and Overlay Operations: Inventory and Prospects, Third International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences in Québec City, Ann Arbor Press Chelsea, Michigan, pp. 11-27
Guo, H., Wang, Y.
2004: Specific vulnerability assessment using the *MLPI* model in Datong city, Shanxi province, China. Environmental Geology, Volume 45, Number 3, January 2004, pp. 401-407
Heuvelink, G.B.M.
1998: Error Propagation in Environmental Modelling, London: Taylor and Francis.
Heuvelink, G.B.M.
1999: Aggregation and Error Propagation in GIS, Third International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences in Québec City, Ann Arbor Press Chelsea, Michigan, pp. 219-227

Hughes, M., Bygrave, J., Bastin, L. and Fisher, P.
1999: High Order Uncertainty in Spatial Information: Estimating the Proportion of Cover Types Within a Pixel, Third International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences in Québec City, Ann Arbor Press Chelsea, Michigan, pp. 319-325
Isabel, D., Gélinas, P. et Aubre, F.
1987 : Cartographie de la vulnerabilité des eaux souterraines au Québec, Rev Sc Tech Eau; 23 : 255-8
Joao, E.M.
1998: Causes and Consequences of Map Generalisation, London: Taylor and Francis
Journel, A.G. and Huijbregts, C.J.
1978: Mining Geostatistics. Academic Press, 600 p.
Knight, F.H.,
1921: Risk, Uncertainty and Profit. Boston, MA: Hart, Schaffner and Marx, Houghton Mifflin Company.
Lobo-Ferreira, J.P.
1999: The European Union Experience On Groundwater Vulnerability Assessment And Mapping, in COASTIN A Coastal Policy Research Newsletter, Number 1, Sep. 1999. New Delhi, TERI, pp. 8-10.
MacEachren, A.M.
1994: Some Truth with Maps: A Primer on Symbolization and Design (Washington, D.C.: Association of American Cartographer)
Mavko et al.
1978: The rock physics Handbook. Cambridge University Press.
Marsily, G. de
1986: Quantitative hydrogeology- Groundwater hydrology for engineers. Academic Press, San Diego, 440 p.
McNab and al.
1999: Historical Case Analysis of Chlorinated Volatile Organic Compound Plumes. Lawrence Livermore National Laboratory, University of California.
Murat and al.
2004: Comparing vulnerability mapping methods in two Canadian hydrogeological settings. Proceedings of the 57th Canadian Geotechnical Conference and 5th Joint CGS/IAH-CNC Conference.
Mowrer, H.T.
1999 : Accuracy (Re)assurance : Selling Uncertainty Assessment to the Uncertain, Third International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences in Québec City, Ann Arbor Press Chelsea, Michigan, pp. 3-11.

Menami, MR
1999 : Using matrices for mapping the groundwater pollution potential. Case of the DRASTIC method applied to the El Madher plain (Eastern Algeria). 4th Intern Conf on Geochemistry, 15-16 Sept. 1999, Alexandria, Egypt.

Menami, M.R.
2001: Evaluating and mapping the groundwater pollution susceptibility of the El Madher alluvial aquifer, Eastern Algeria, using the Drastic method, Science et changements planétaires / Sécheresse. Number 12, volume 2, 95-101.

MOE (Ministry of the Environment, Ontario),
2001: Groundwater Studies 2001/2002. Technical Terms of Reference.

Oliver, M.A. and Webster, R.
1990: Kriging: a method of interpolation for geographical information systems, International Journal of GIS, 4, pp. 313-332.

Robin, M.J. and Daneshfar, B.
(in preparation) Generating hydrogeological vulnerability maps based on a simplified hydrogeological partition model within a GIS: an example from Southeastern Ontario, Canada

Scott, D.F. and Aitken, J.D.
1993 : (Géologie du Canada no 5) La couverture sédimentaire du craton au Canada, Ottawa : Commission Géologique du Canada.

Technical Experts Committee, Ontario Ministry of the Environment
2004: Watershed-based source protection planning, science-based decision-making for protecting Ontario's drinking water resources: A Threats Assessment Framework.

Van der Leeden, F., Troise, F.L., Todd, D.K.,
1991: The water encyclopedia, 2nd edition, Lewis Publishers.

Van Seempvoort, D., Ewert, L., Wassenaar, L.
1992: AVI: A Method for Groundwater Protection Mapping in the Prairie Provinces of Canada. PPWD pilot project, Sept. 1991 – March 1992. Groundwater and Contaminants Project, Environmental Sciences Division, National Hydrology Research Institute. 23 pp.

Wang, M., Howarth, P.J.,
1994: "Multi-Source Spatial Data Integration: Problems and some Solutions", Canadian Journal of Remote Sensing, Vol. 20, No. 4, pp. 360-367.

Wauchope, R.D.
1992 : Environmental Risk Assessment of Pesticides : Improving Simulation. Weed technology. Volume 6: 753-759

Wingle, W.L. and Poeter, E.P.,
1998: Directional Semivariograms: Kriging Anisotropy Without Anisotropy Factors, American Association of Petroleum Geologist (AAPG) '98 Annual Convention, Advances in Geostatistics, Salt Lake City, Utah, May 17-20, 1998.

Zadeh, Lofti
1965: Fuzzy Sets. Information and control 8: 338-353.

APPENDIX A : TABLES

Table 1.0: Generic Representative Permeability (K-Factor) Table (Ontario Ministry of the Environment, 2001)

Geomaterial	Representative K-Factor (dimensionless)*	K-Value (m/s) @75% range**	Highest K-Value (m/s)
gravel	1	1.00^{E}-01	0.1
weathered dolomite/limestone		1.00^{E}-06	
karst		1.00^{E}-03	
permeable basalt		1.00^{E}-03	
sand	2	0.01	1.00^{E}-02
peat (organics)	3	1.00^{E}-03	1.00^{E}-03
silty sand		1.00^{E}-04	
weathered clay (<5m below surface)		1.00^{E}-04***	
shrinking/fractured & aggregated clay		1.00^{E}-04***	
fractured igneous metamorphic rock		1.00^{E}-05	
weathered shale		1.00^{E}-05***	
Silt	4	1.00^{E}-06	1.00^{E}-06
loess		1.00^{E}-06	
limestone/dolomite		1.00^{E}-06	
weathered/fractured till	5	1.00^{E}-07	1.00^{E}-07
diamicton (sandy, silty)		1.00^{E}-07***	
diamicton (silty, clayey)		1.00^{E}-08***	
sandstone		1.00^{E}-07	
clay till	8	1.00^{E}-09***	1.00^{E}-09
clay (unweathered marine)		1.00^{E}-10	
unfractured igneous and metamorphic rock	9	1.00^{E}-13	1.00^{E}-13

- Representative K-Factors are relative numbers and do not correspond directly to the exponent or index of the observed K-Values for the geomaterial in the group.

** Correspondence with descriptors of observed K-Values in Freeze & Cherry 1979, Prentice-Hall. Derived using the length of the line to determine the 75% value and rounding to the highest K-Value.
*** Estimated value based on field studies in Ontario

Table 2.0: K-Factor value associated to the geology description (upper overburden).

FREQUENCY	TYPE_1	K_factor
2318		
6	Beach, bar, and related deposits	2
262	Clay and silt	5
15	Clay ridges	8
982	Clay, silty clay and silt	5
34	Combination of Gravel, sand and boulders and Fine-to medium-grained sand	1
76	Combination of Silty sand, silt, sand and clay and Medium grained stratified sand with some silt	5
367	Fine-to medium-grained sand	3
138	Gravel, □ravely sand, sand	1
950	Gravel, sand and boulders	1
751	Intrusive and metamorphic rocks	9
588	Limestone, dolomite, sandstone	4
117	Medium-to fine-grained sand	2
177	Medium grained stratified sand with some silt	3
102	Medium to medium fine sand	2
2513	Organic deposits	3
521	Sand and gravel	1
1	Sand, gravel, boulder gravel	1
2760	Sandy and silt	3
194	Silty sand, silt, sand and clay	5

Table 2.1: Geomaterial descriptions with associated porosity, effective porosity & specific yield values used for the SAAT calculation

FEATURE	Specific Yield	POROSITY	Effective porosity (unsaturated zone) (n-sy)	Effective Porosity (saturated zone)
Clay Plain	0.03	0.45	0.42	0.06
Esker	0.16	0.30	0.14	0.26
Limestone Plain	0.14	0.30	0.16	0.14
Peat and Muck	0.44	0.90	0.46	0.55
Sand Plain	0.28	0.35	0.07	0.33

Till Plain (Drumlinized)	0.16	0.35	0.19	0.1
Till Plain (Undrumlinized)	0.10	0.35	0.25	0.15

Table 2.2: Geomaterial's typical range values and associated variance values to be used in the SAAT's uncertainty/error modeling

FEATURE	Range (Porosity)	Range (Effective Porosity)	Variance (effective porosity)	Variance (porosity)
Clay Plain	0.34-0.57 (1)	0.01 – 0.18 (4)	0.09	0.12
Esker	0.25-0.45 (2)	0.01 – 0.46 (4a)	0.23	0.11
Limestone Plain	0.07-0.56 (1)	0.00 – 0.36 (4)	0.18	0.25
Peat and Muck	0.60-0.90 (2a)	0.45 – 0.70 (2)	0.13	0.21
Sand Plain	0.25-0.45 (2)	0.16 – 0.46 (4)	0.15	0.10
Till Plain (Drumlinized)	0.10-0.35 (2a)	0 – 0.15 (2)	0.08	0.18
Till Plain (Undrumlinized)	0.10-0.35 (2a)	0.05 – 0.2 (2)	0.08	0.18

(1) McWorther and Sunada (1997)

(2) Banton & Bangoy (1999)

(2a) Banton & Bangoy (1999). The highest value of the range was increased to include the value used in the model calculation

(3) Estimated from range values of porosity and effective porosity

(4) McWorther and Sunada (1977)

(4a) McWorther and Sunada (1977): estimated with silt, sand and gravel values

Table 3.0: Depth to the water's range and rating values

DEPTH TO THE WATER (meter)

WEIGHT=5

Range	Rating
0 -1.5	10
1.5 – 4.5	9
4.5 – 9.0	7
9.0 – 15.1	5
15.1 – 22.7	3
22.7 – 30.2	2
30.2 and more	1

Table 3.1: *Recharge's range and rating values*

RECHARGE (mm/year)
WEIGHT=4

Range	Rating
0 – 50.8	1
50.8 – 101.6	3
101.6 – 177.8	6
177.8 – 254	8
254 and more	9

Table 3.2: Aquifer Media 's range and rating value

AQUIFER MEDIA
WEIGHT=3

Geomaterial	Range	Typical Rating
Massive Shale	1	2
Metamorphic/Igneous	2 – 5	3
Weathered Metamorphic/igneous	3 – 5	4
Glacial Till	4 – 6	5
Bedded sandstone, limestone and Shale sequences	5 – 9	6
Massive Sandstone	4 – 9	6
Massive Limestone	4 – 9	8
Sand and Gravel	4 – 9	8
Basalt	2 – 10	9
Karst Limestone	9 – 10	10

Table 3.3: Soil type's range and rating value

SOIL TYPE
WEIGHT=2

Range	Rating
Thin or Absent	10
Gravel	10
Sand	9
Peat	8
Shrinking and/or Aggregated Clay	7
Sandy Loam	6

Loam	5
Silty Loam	4
Clay Loam	3
Muck	2
Clay	1

Table 3.4: Topography's range and rating value

TOPOGRAPHY (% SLOPE)

WEIGHT=1

Range	Rating
0 – 2	10
2 – 6	9
6 – 12	5
12 – 18	3
18 and more	1

Table 3.5: Impact of vadose zone's range and rating value

IMPACT OF VADOSE ZONE

WEIGHT=5

Geomaterial	**Range**	**Typical Rating**
Confining Layer	1	1
Silt/Clay	2 – 6	3
Shale	2 – 6	3
Limestone	2 – 5	3
Sandstone	2 – 7	6
Bedded Limestone, Sandstone, Shale	4 – 8	6
Sand and Gravel with significant Silt and Clay	4 – 8	6
Sand and Gravel	4 – 8	8
Basalt	2 – 10	9
Karst Limestone	8 – 10	10

Table 3.6: Hydraulic conductivity's range and rating value

HYDRAULIC CONDUCTIVITY (m^2/sec)

WEIGHT=3

Range		**Rating**
1.44^{E}-06	1.44^{E}-05	1
1.44^{E}-05	4.32^{E}-05	2
4.32^{E}-05	1.01^{E}-04	4
1.01^{E}-04	1.44^{E}-04	6
1.44^{E}-04	2.88E-04	8
2.88E-04 and more		10

Table 3.7: Variance of Depth to the water's values

Values with overlapping rating (meter)		Rating assigned	Range of rating		Variance
0.5	1.5	10	9	10	0.50
1.5	2.5	9	9	10	0.50
3.5	4.5	9	7	9	2.00
4.5	5.5	7	7	9	2.00
8	9	7	5	7	2.00
9	10	5	5	7	2.00
14.1	15.1	5	3	5	2.00
15.1	16.1	3	3	5	2.00
21.7	22.7	3	2	3	0.50
22.7	23.7	2	2	3	0.50
29.2	30.2	2	1	2	0.50
30.2	31.2	1	1	2	0.50

Table 3.8: Variance of Recharge's values

Values with overlapping rating		Rating assigned	Range of rating		Variance
40.8	50.8	1	1	3	2.00
50.8	60.8	3	1	3	2.00
91.6	101.6	3	3	6	4.50
101.6	111.6	6	3	6	4.50
177.8	187.8	8	6	8	2.00
244	254	8	8	9	0.50
254	264	9	8	9	0.50

Table 3.9: Variance of Topography's values

Values with overlapping rating		Rating assigned	Range of rating		Variance
0	2	10	9	10	0.50
2	4	9	9	10	0.50
4	6	9	5	9	8.00
6	8	5	5	9	8.00
10	12	5	3	5	2.00
12	14	3	3	5	2.00
16	18	3	1	3	2.00
18	20	1	1	3	2.00

Table 3.10: Variance of Hydraulic Conductivity's values

Values with overlapping rating		Rating assigned	Range of rating		Variance
1.30^{E}-05	1.44^{E}-05	1	1	2	0.50
1.44^{E}-05	1.58^{E}-05	2	1	2	0.50
3.89^{E}-05	4.32^{E}-05	2	2	4	2.00
4.32^{E}-05	4.75^{E}-05	4	2	4	2.00
9.09^{E}-05	1.01^{E}-04	4	4	6	2.00
1.01^{E}-04	1.11^{E}-04	6	4	6	2.00
1.30^{E}-04	1.44^{E}-04	6	6	8	2.00
1.44^{E}-04	1.58^{E}-04	8	6	8	2.00
2.59^{E}-04	2.88^{E}-04	8	8	10	2.00
2.88^{E}-04	3.17^{E}-04	10	8	10	2.00

Table 3.11: Variance of Aquifer media's values

Geomaterial	**Rating assigned**	**Range of rating**		**Variance**
Fractured bedrock (limestone)	7	4	9	6.50
Sand	9	4	9	12.50

Table 3.12: Variance of Impact of vadose zone's values

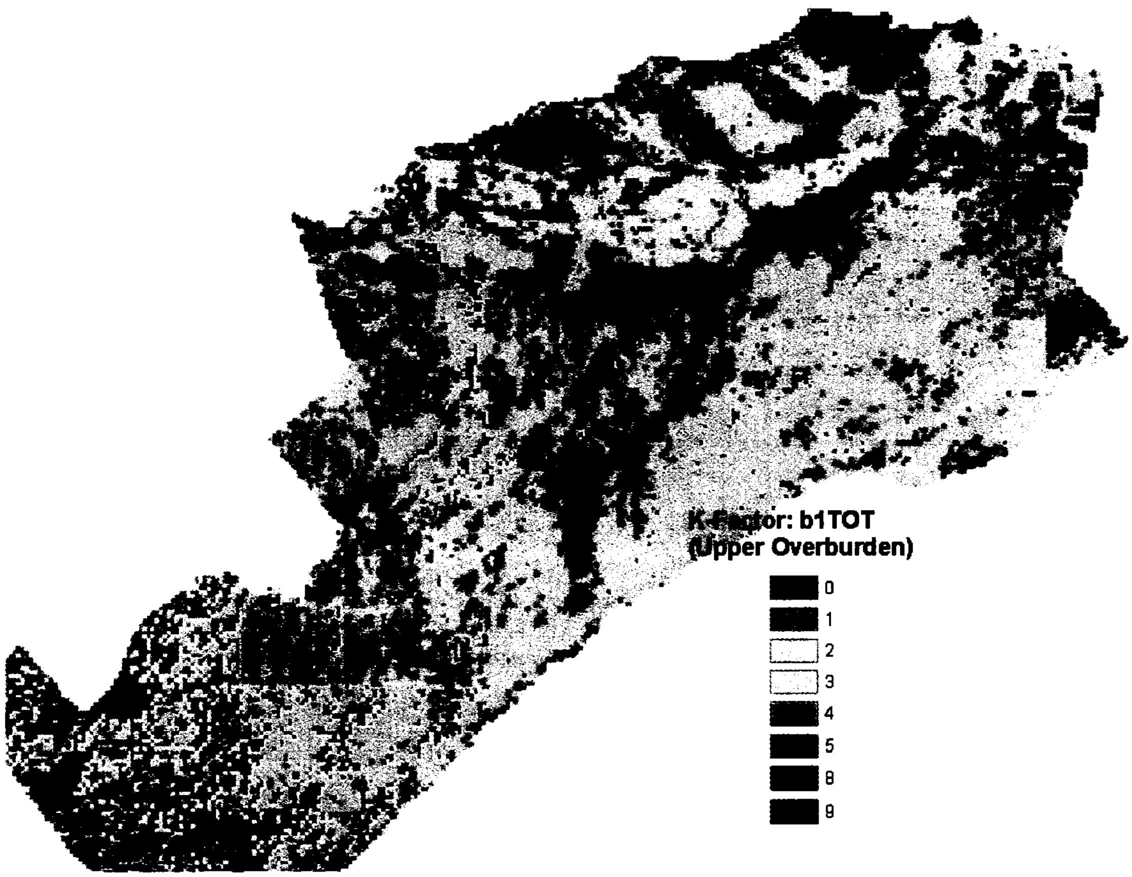
Geomaterial	**Rating assigned**	**Range of rating**		**Variance**
Beach, bar, and related deposits	8	4	8	8.00
Clay and silt	4	2	6	4.00
clay ridges	4	2	6	4.00
Clay, silty clay, silt	4	2	6	4.00
silty sand,sand, clay and silt	6	2	8	10.00
Fine-to medium-grained sand	8	4	8	8.00
Gravel, □ravely sand, sand	8	4	8	8.00
Gravel, sand and boulders	8	4	8	8.00
Intrusive and metamorphic rocks	6	2	10	16.00
Limestone, dolomite, sandstone	9	2	9	24.50
medium-to fine-grained sand	8	4	8	8.00
Medium grained stratified sand with some silt	6	4	8	4.00
medium to medium fine sand	8	4	8	8.00
Organic deposits	1	1	3	2.00
sand and gravel	8	4	8	8.00
sand, gravel, boulder gravel	8	4	8	8.00
sand and silt	6	4	8	4.00
Silty sand, silt, sand and clay	6	4	8	4.00

Table 3.13: Variance of soil's values

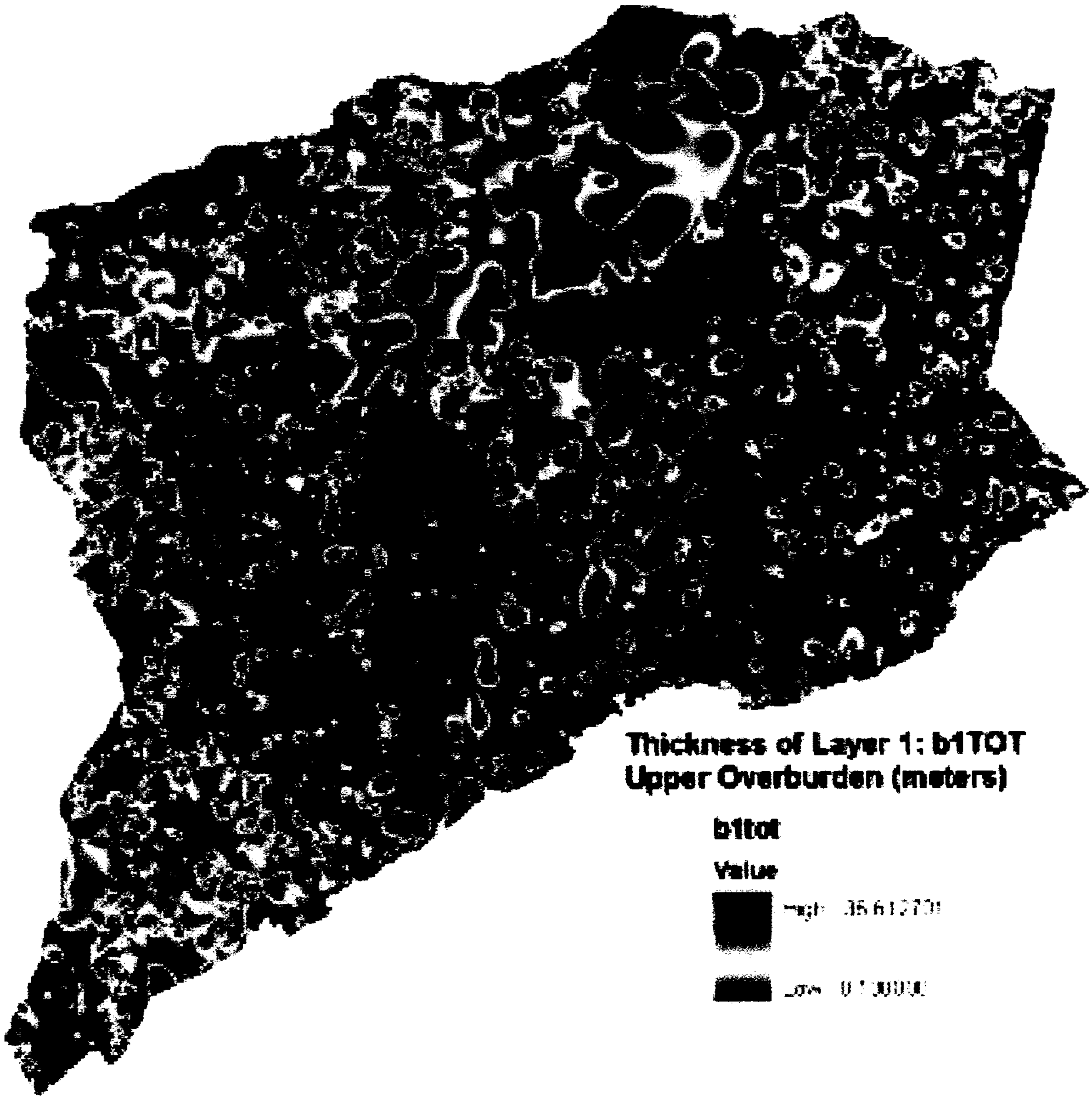
Geomaterial	Rating assigned	Range of rating		variance
Beach, bar, and related deposits	9	9	10	0.50
Clay and silt	2	1	7	13.00
clay ridges	1	1	7	18.00
clay, silty clay and silt	2	1	7	13.00
silty sand,sand, clay and silt	10	9	10	0.50
Fine-to medium-grained sand	9	9	9	0.00
Gravel, □ravely sand, sand	10	9	10	0.50
Gravel, sand and boulders	10	9	10	0.50
Intrusive and metamorphic rocks	10	10	10	0.00
Limestone, dolomite, sandstone	10	10	10	0.00
medium-to fine-grained sand	9	9	9	0.00
Medium grained stratified sand with some silt	9	4	9	12.50
Organic deposits	8	8	8	0.00
sand and silt	6	4	9	6.50
Silty sand, silt, sand and clay	6	1	9	17.00

APPENDIX B : MAPS

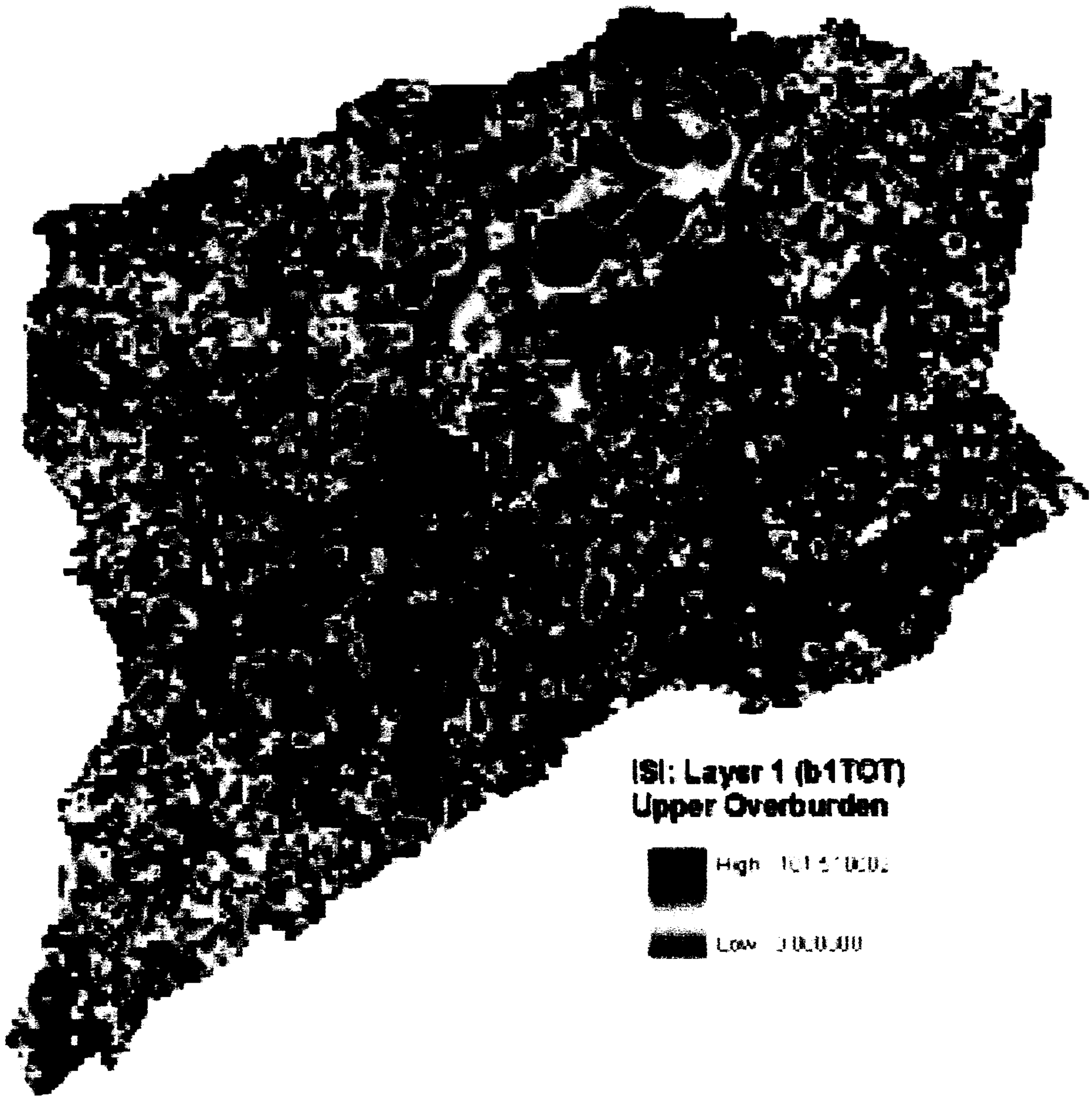
MAP 1.0: ISI K-factor values for b1TOT



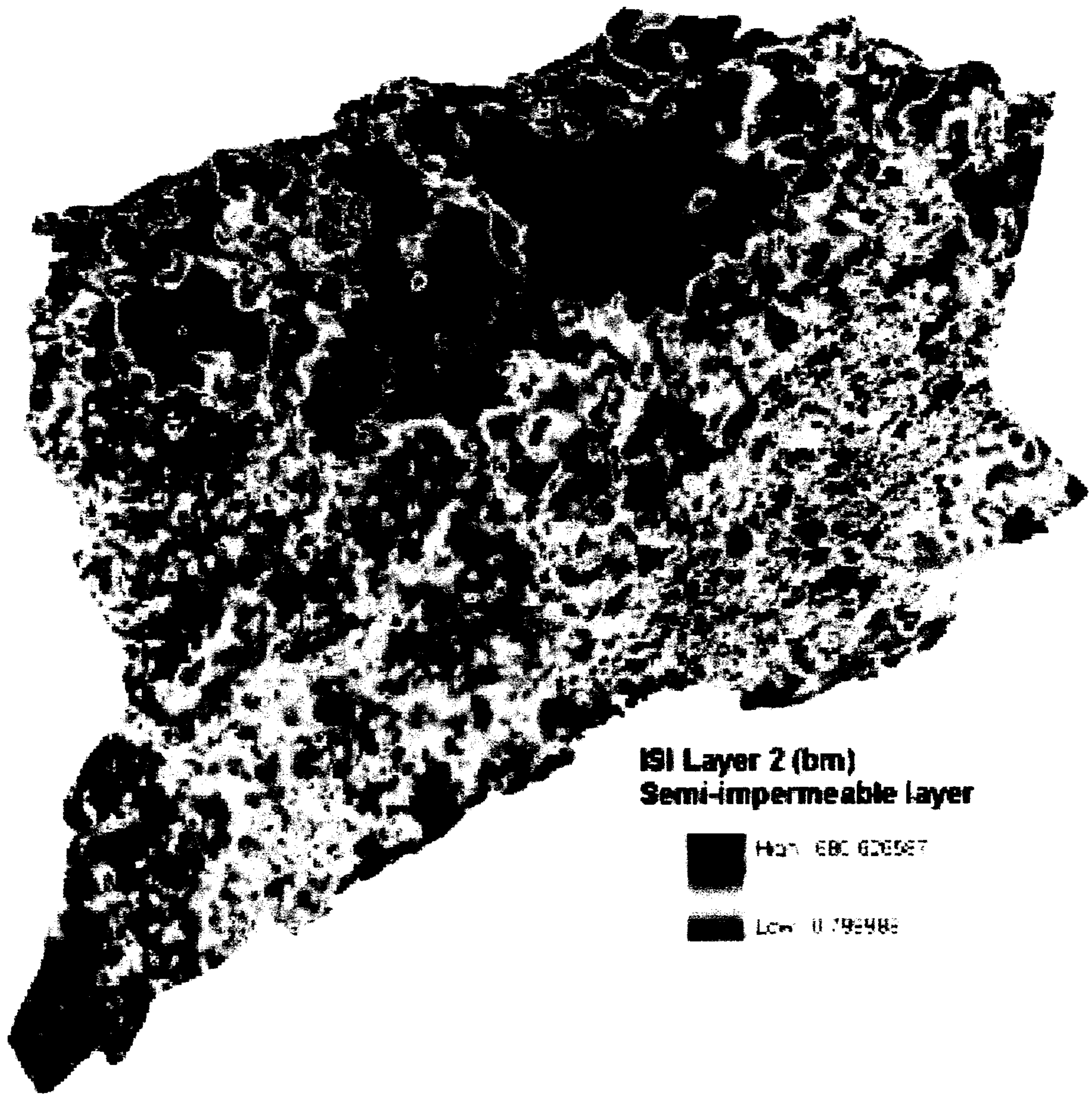
MAP 1.1: Thickness of Upper Overburden (b1TOT)



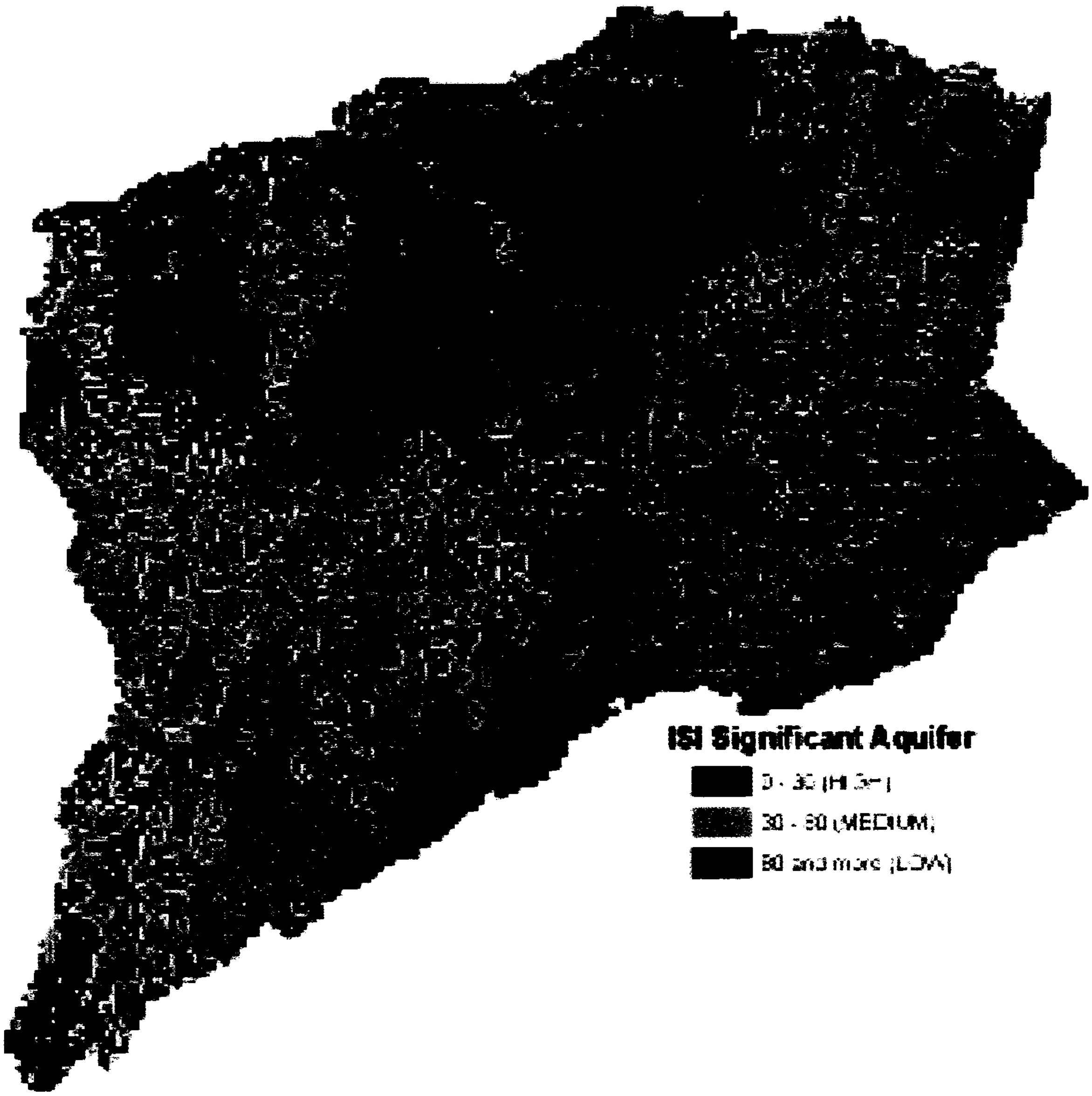
MAP 1.2: ISI for Upper Overburden



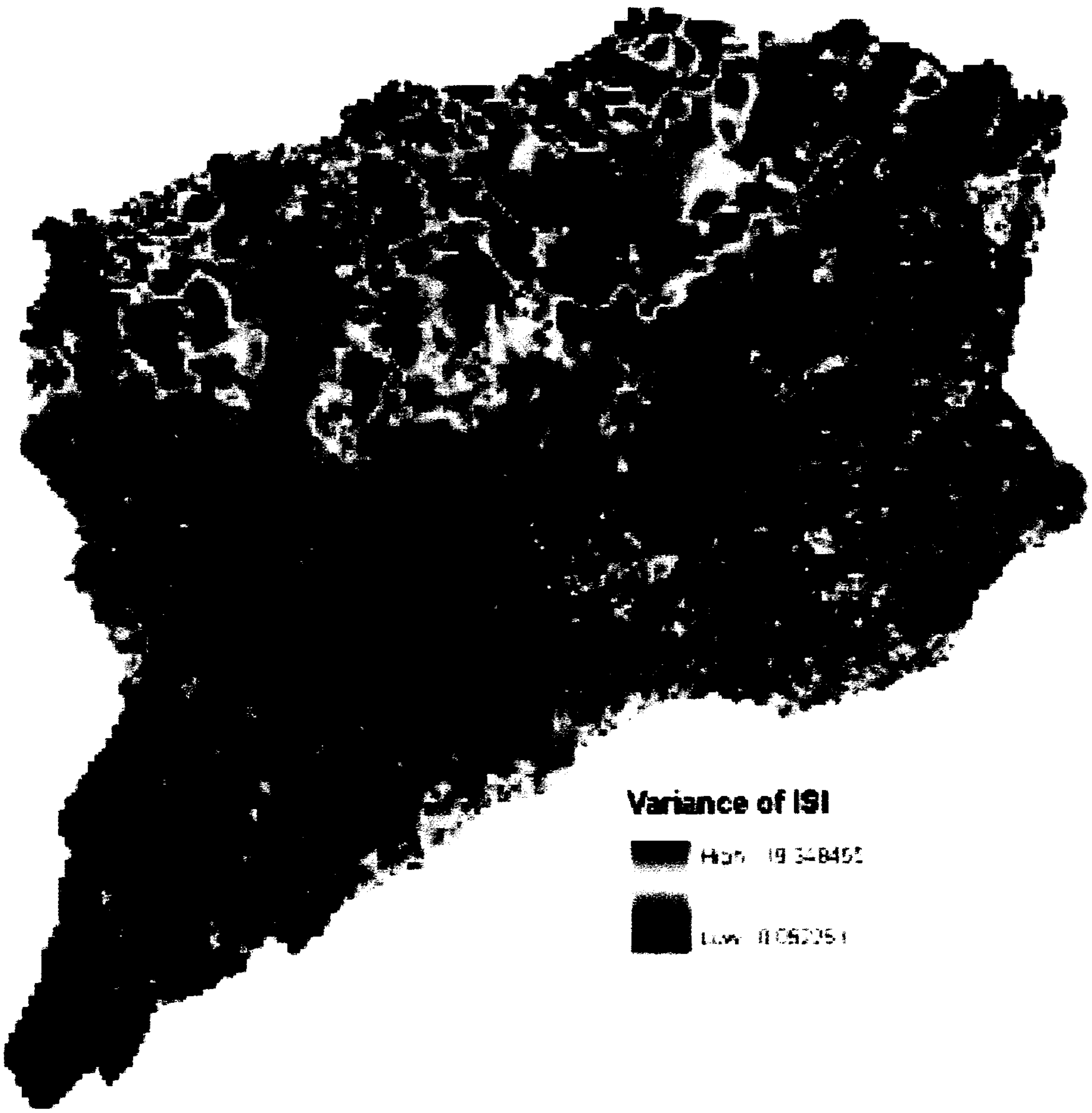
MAP 1.3: ISI for Layer 2 (semi-impermeable layer)



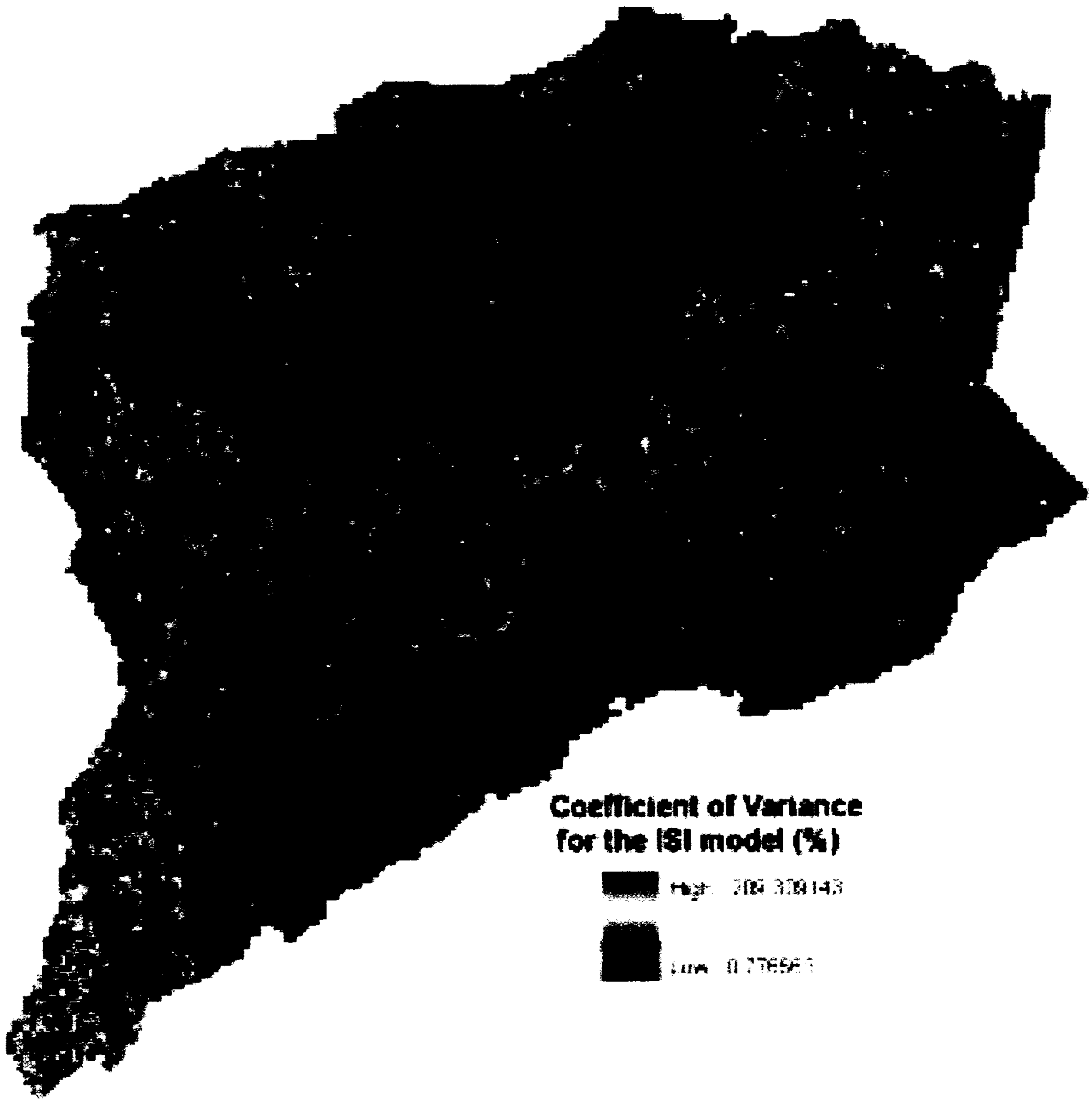
MAP 1.4: ISI for first significant aquifer



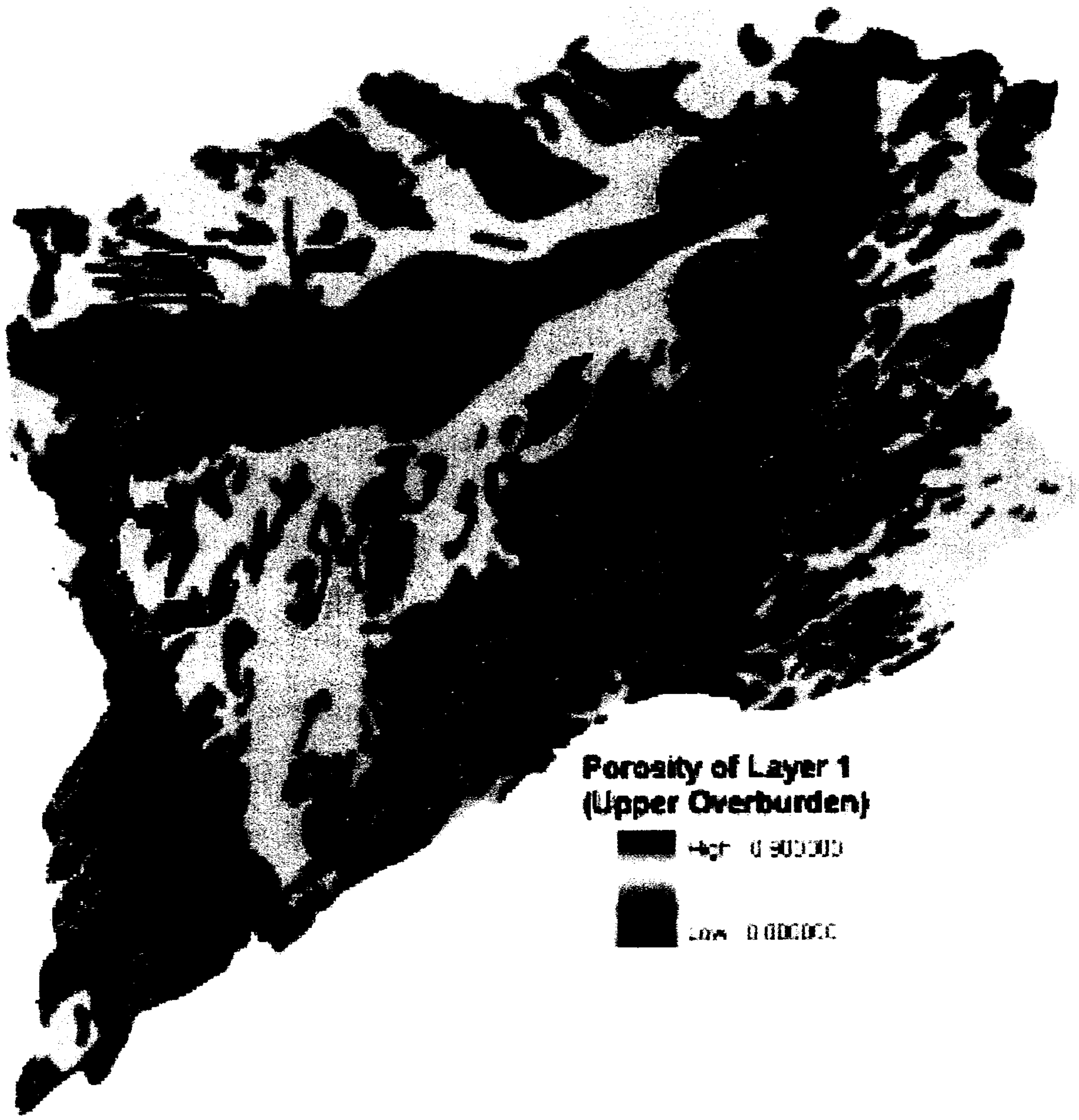
MAP 1.5: Variance of ISI (first significant aquifer)



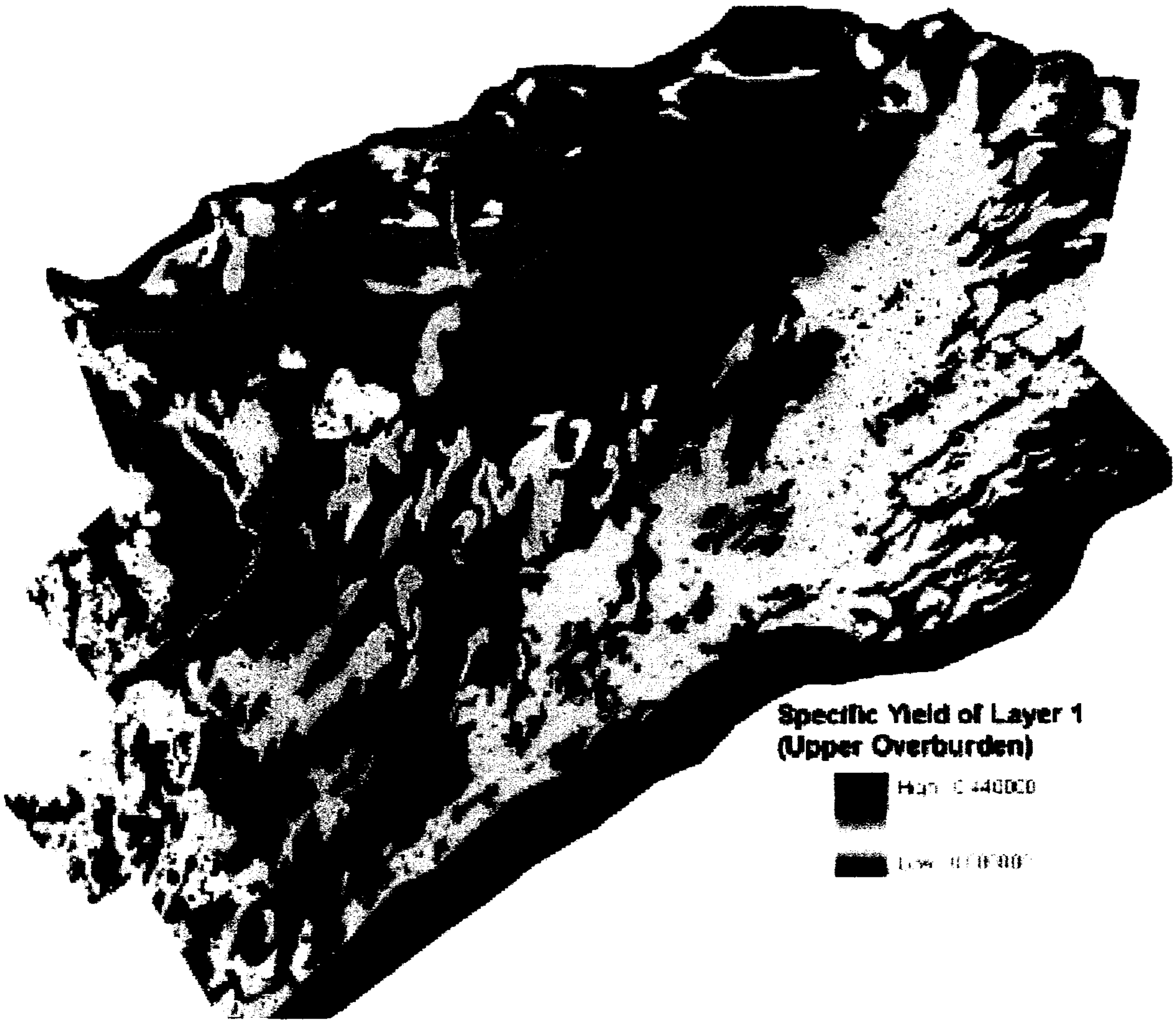
MAP 1.6: Coefficient of variation for ISI (first significant aquifer)



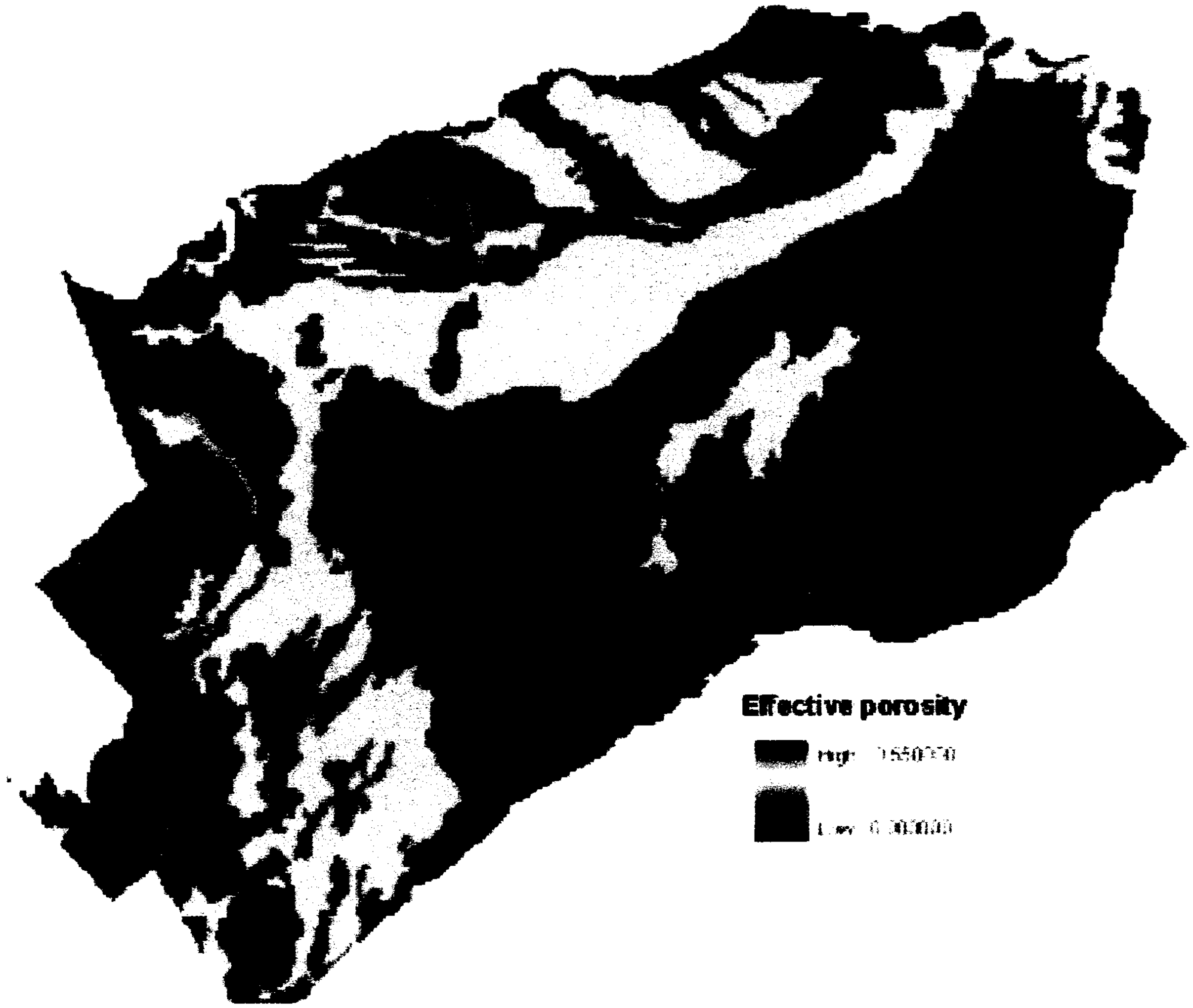
MAP 2.0: Porosity values of Upper Overburden



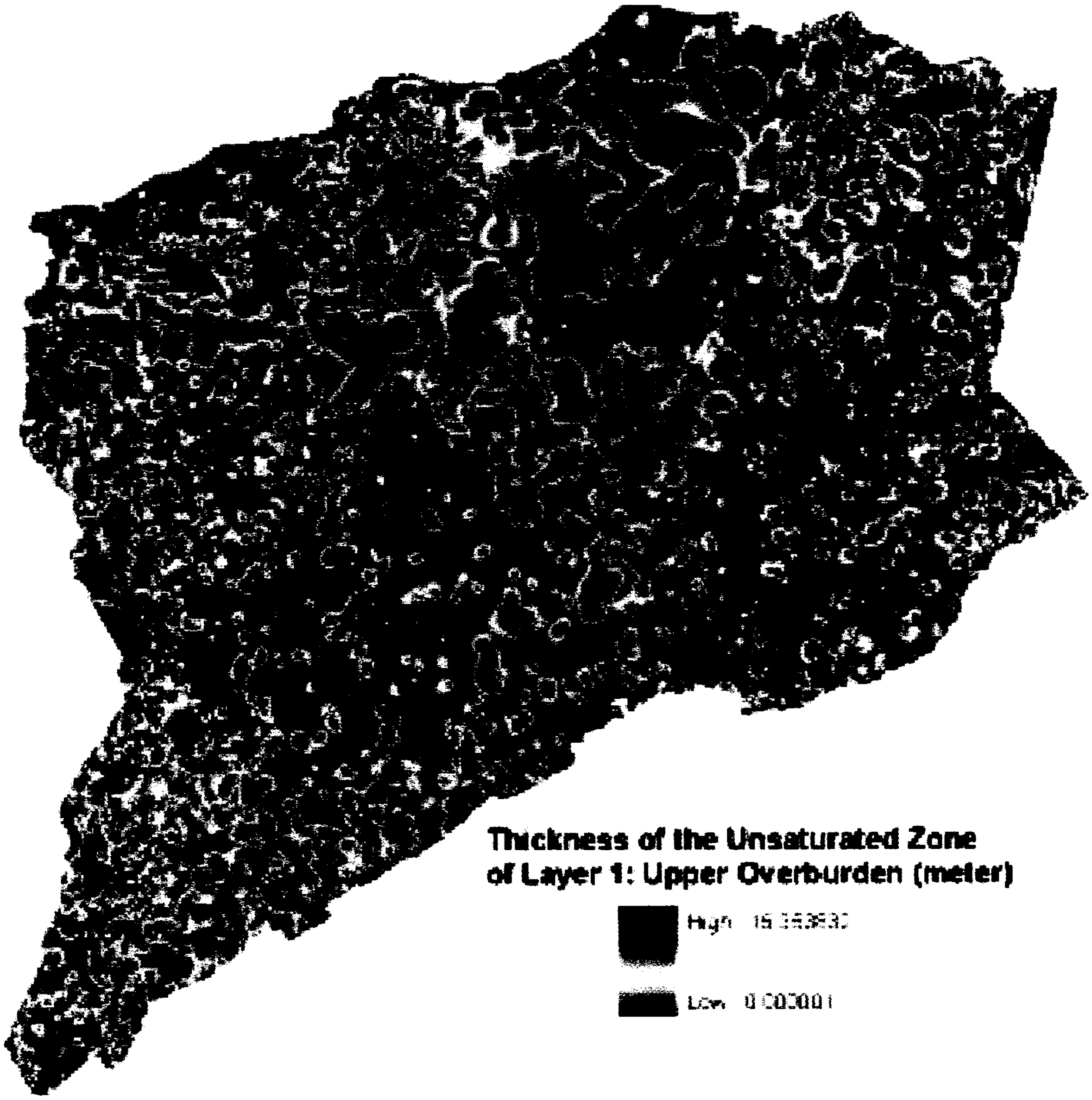
MAP 2.1: Specific yield values of Upper Overburden



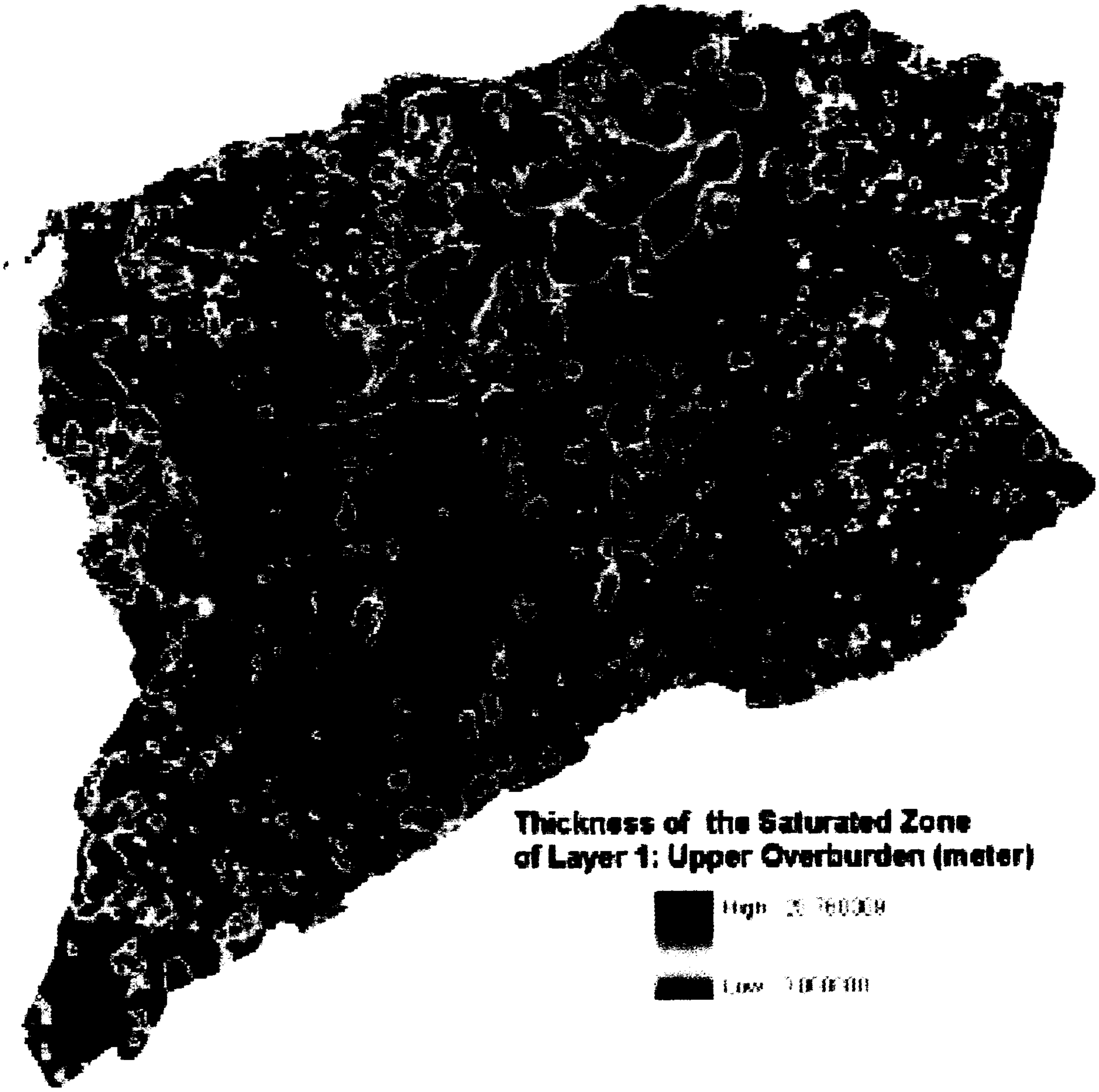
MAP 2.2: Effective porosity values of Upper Overburden



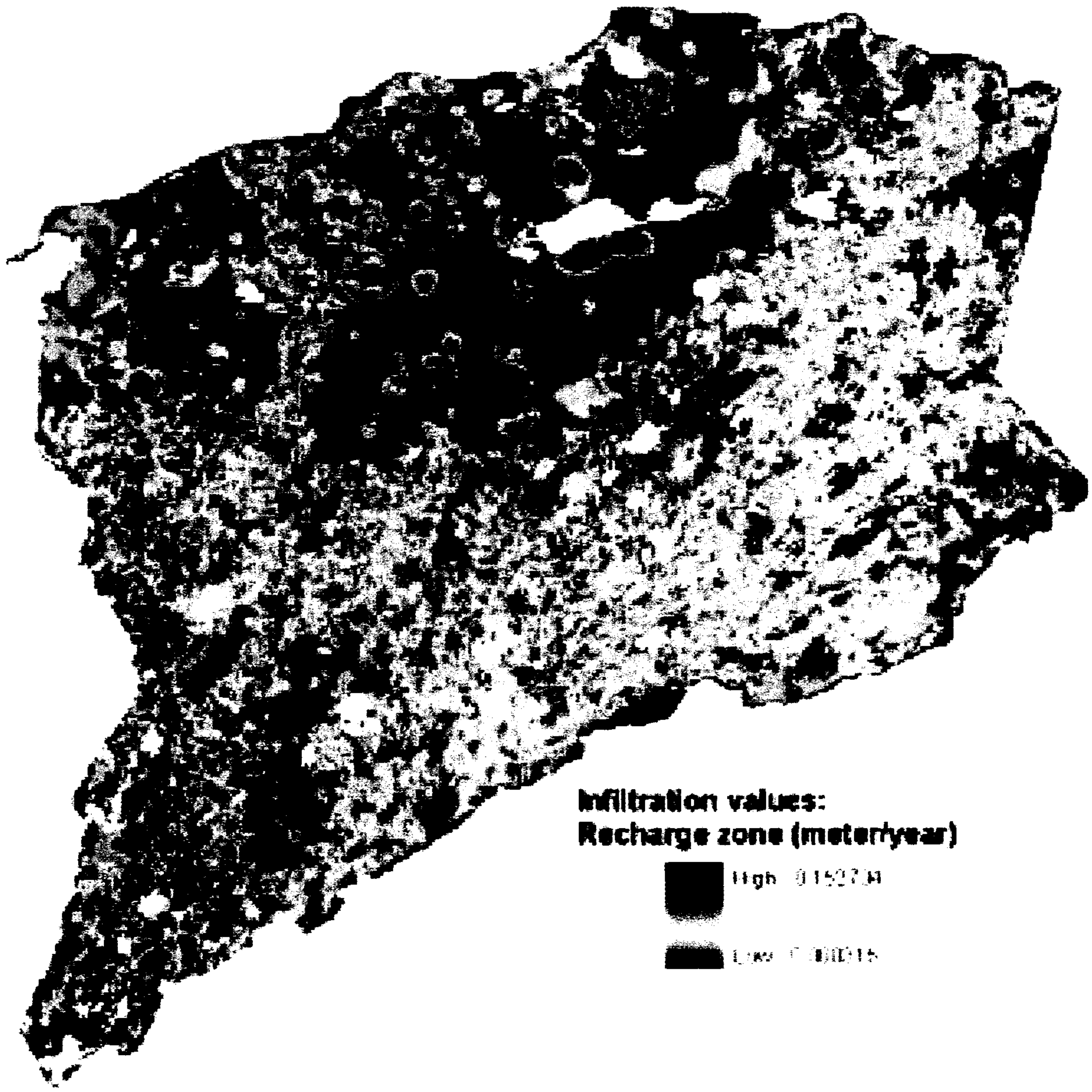
MAP 2.3: Thickness of the unsaturated zone of the Upper Overburden



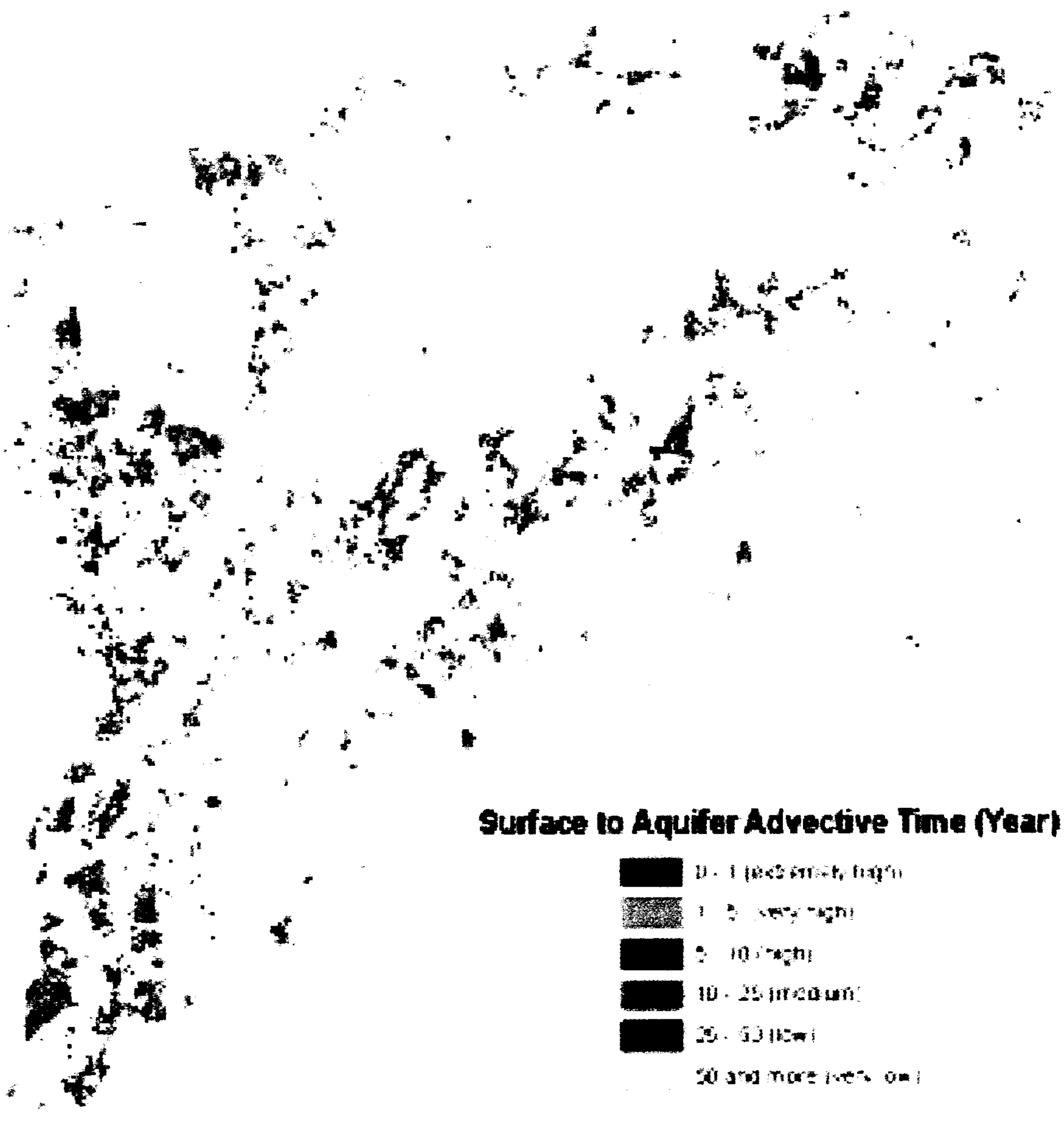
MAP 2.4: Thickness of the saturated zone of the Upper Overburden



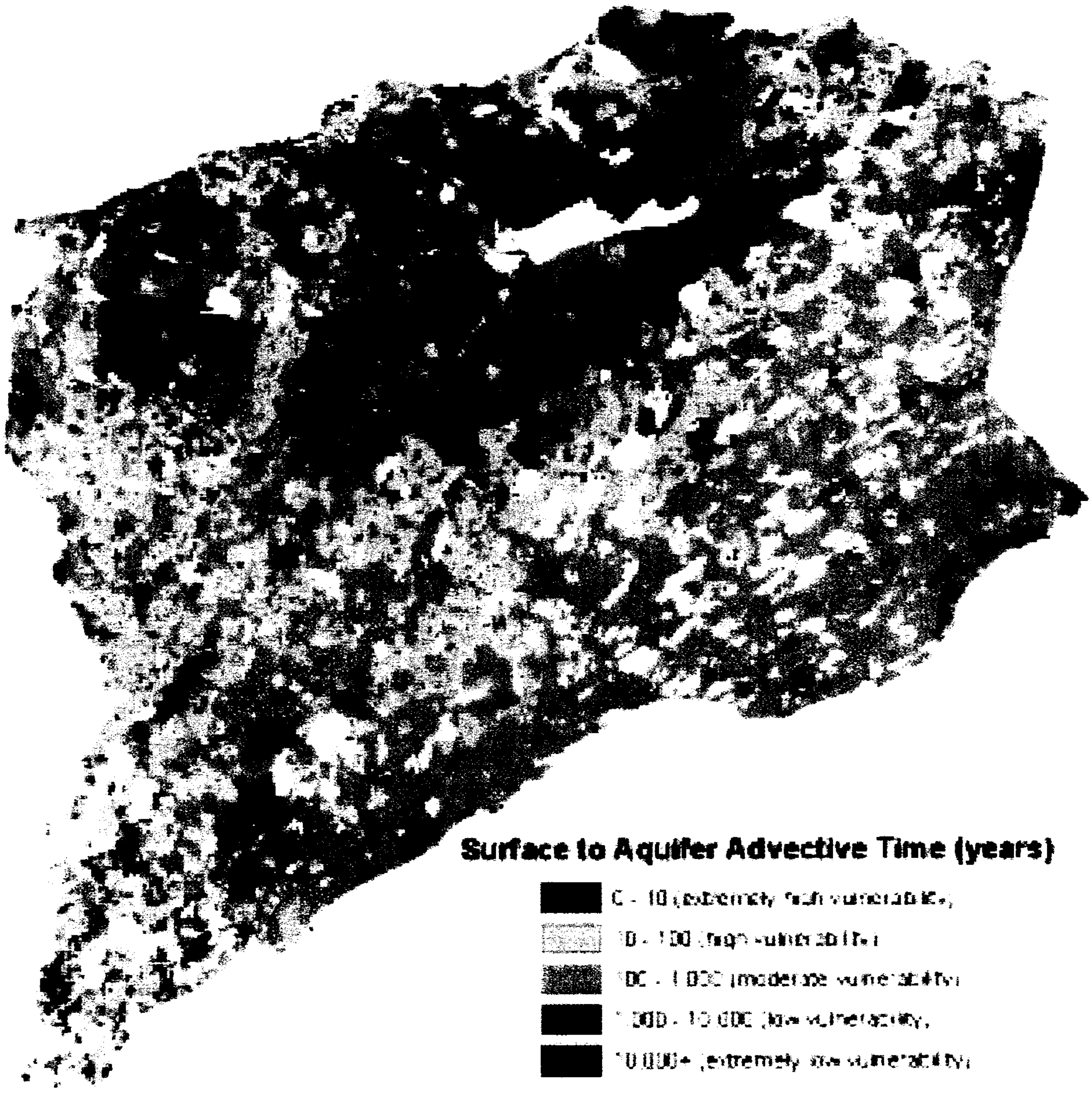
MAP 2.5: Infiltration values



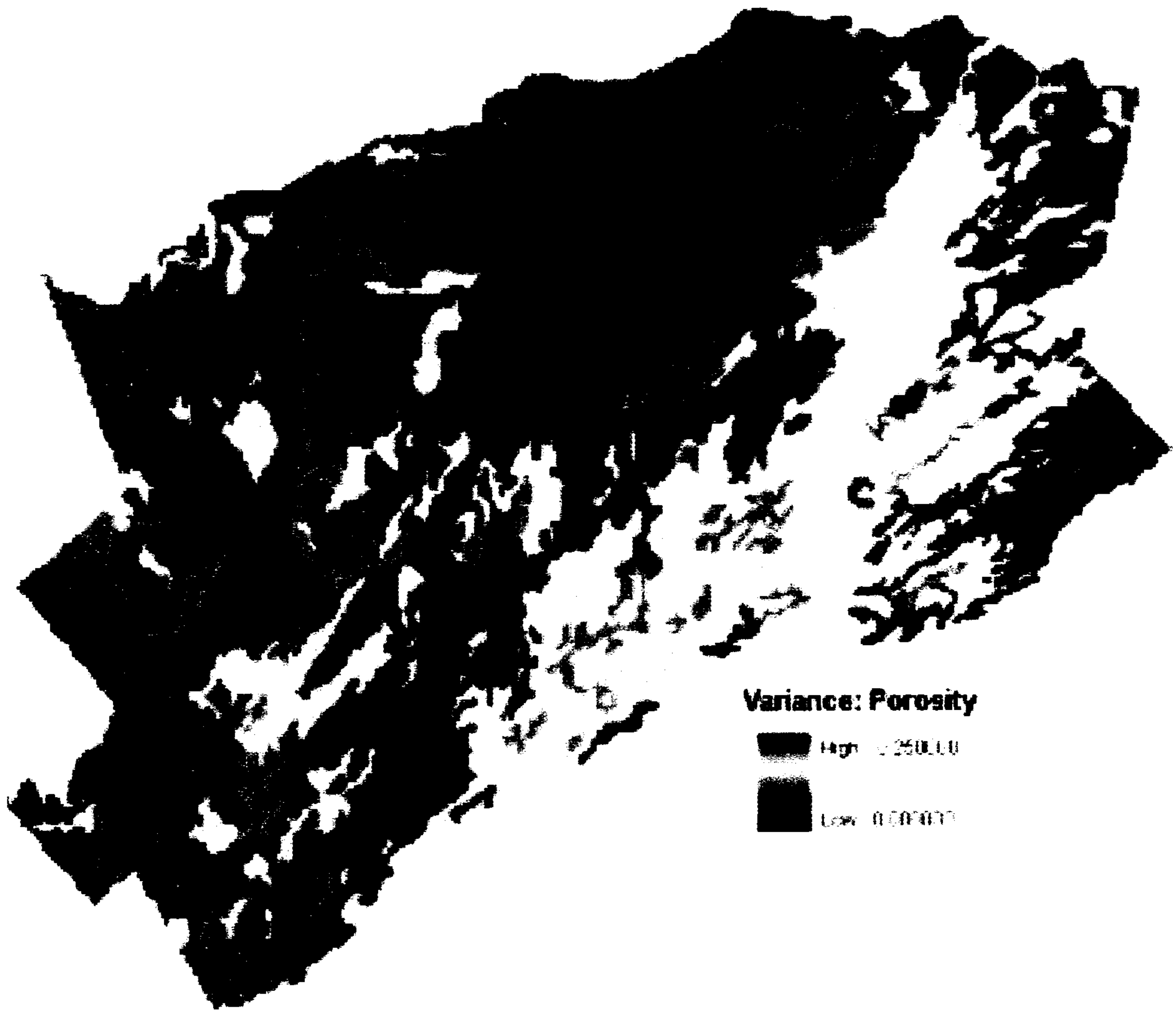
MAP 2.6.1: Surface to Aquifer Advective Time (0 – 50 years' range classification)



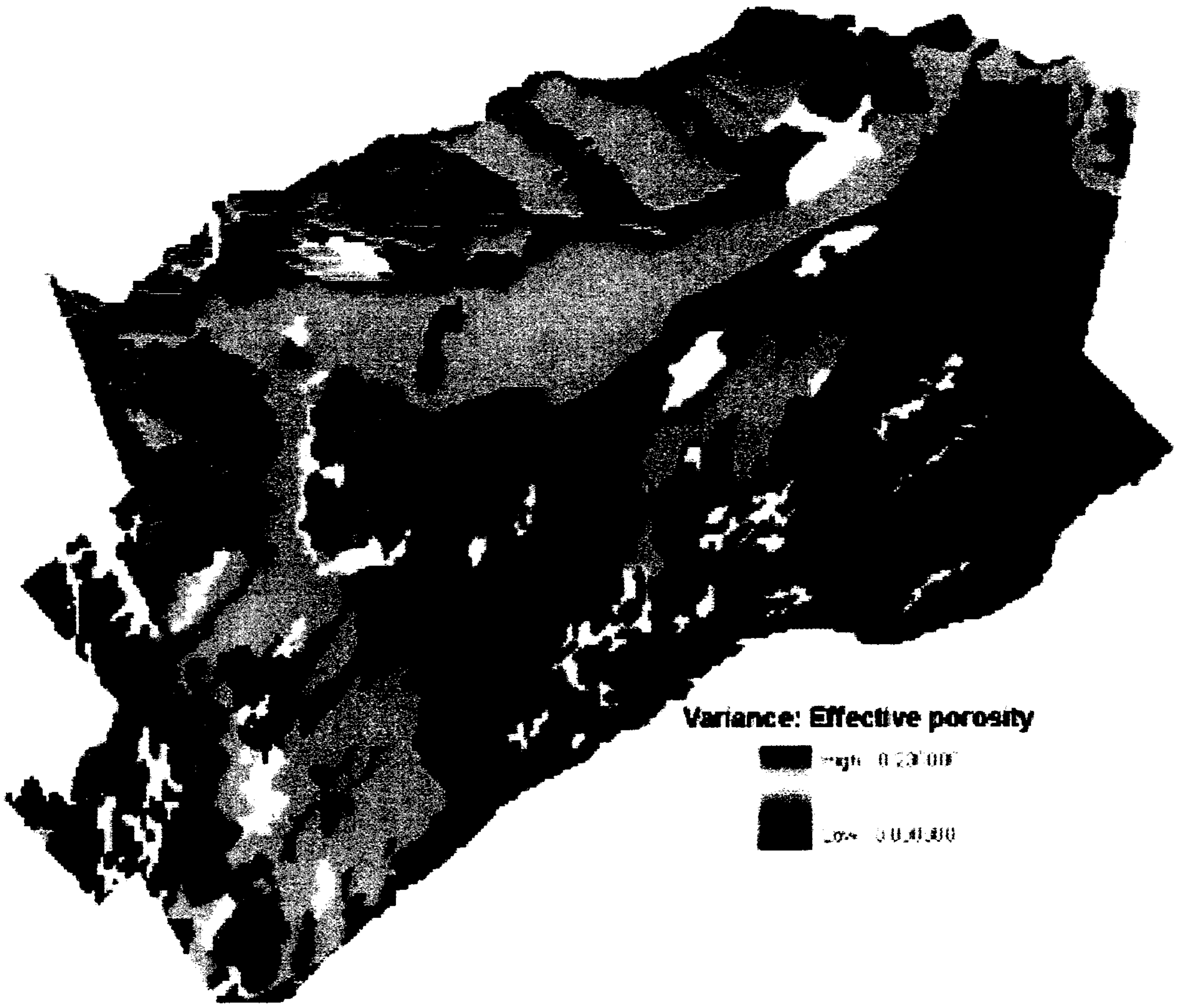
MAP 2.6.2 Surface to Aquifer Advective Time (0 – 10,000 years' range classification)



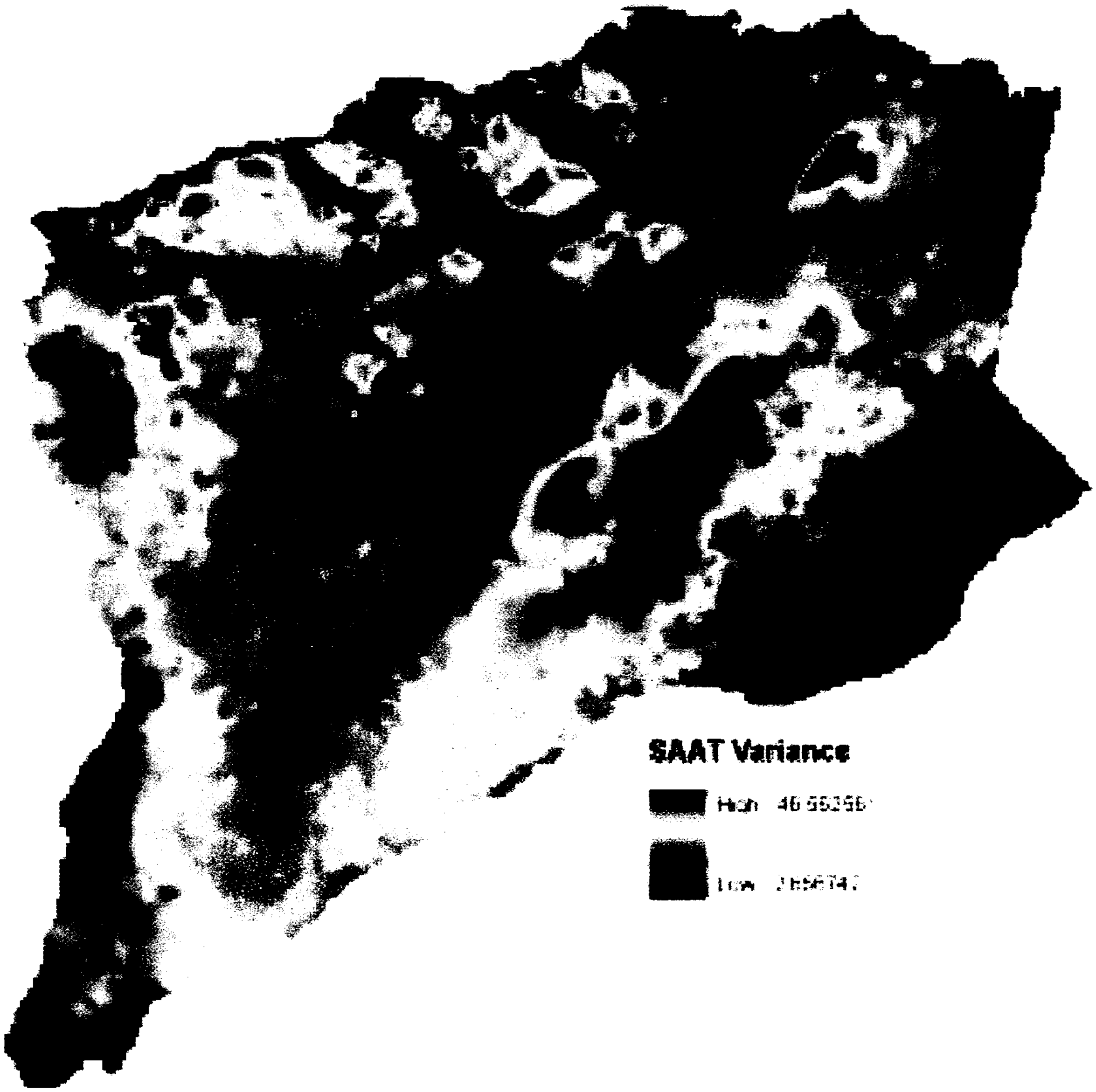
MAP 2.7: Variance of porosity



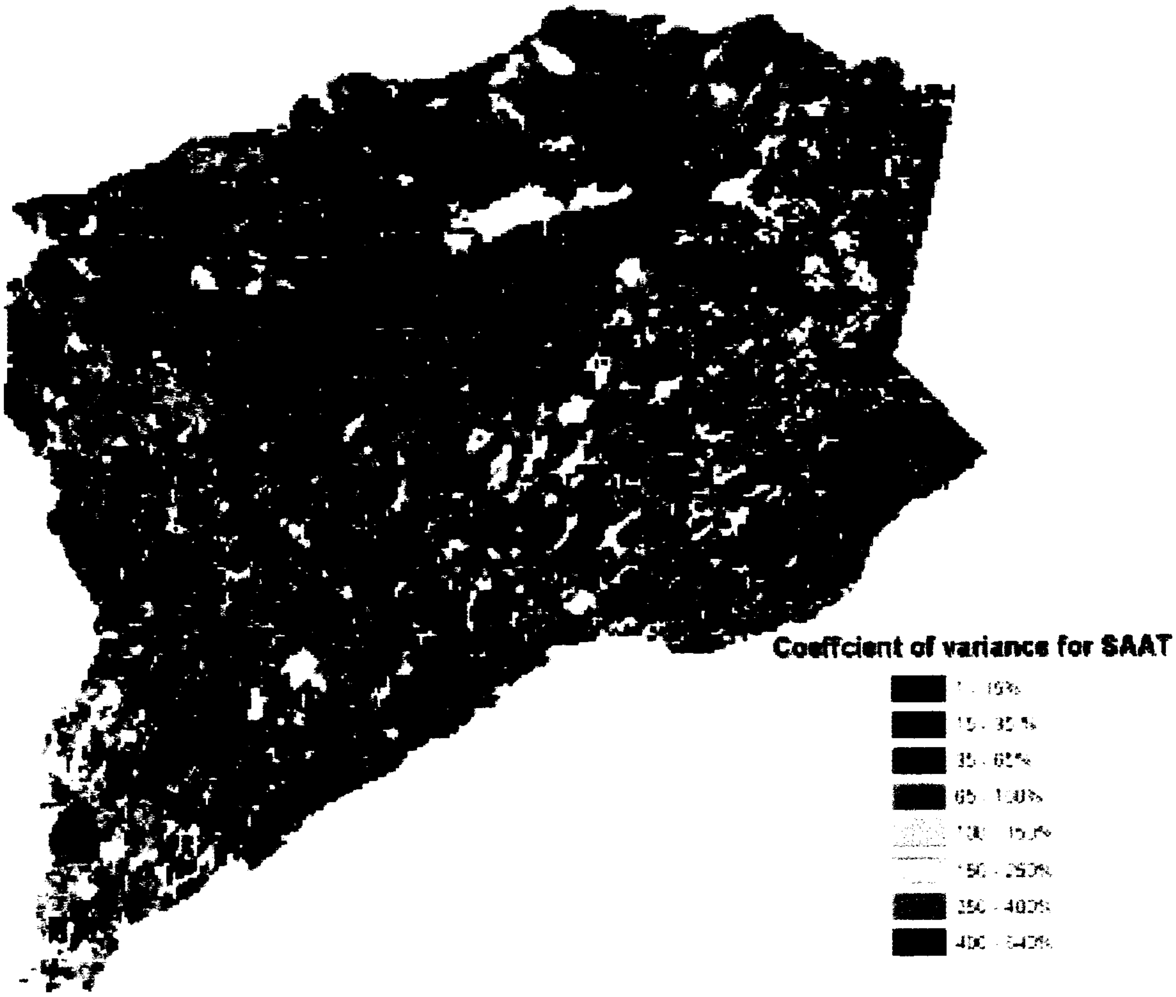
MAP 2.8: Variance of Effective Porosity



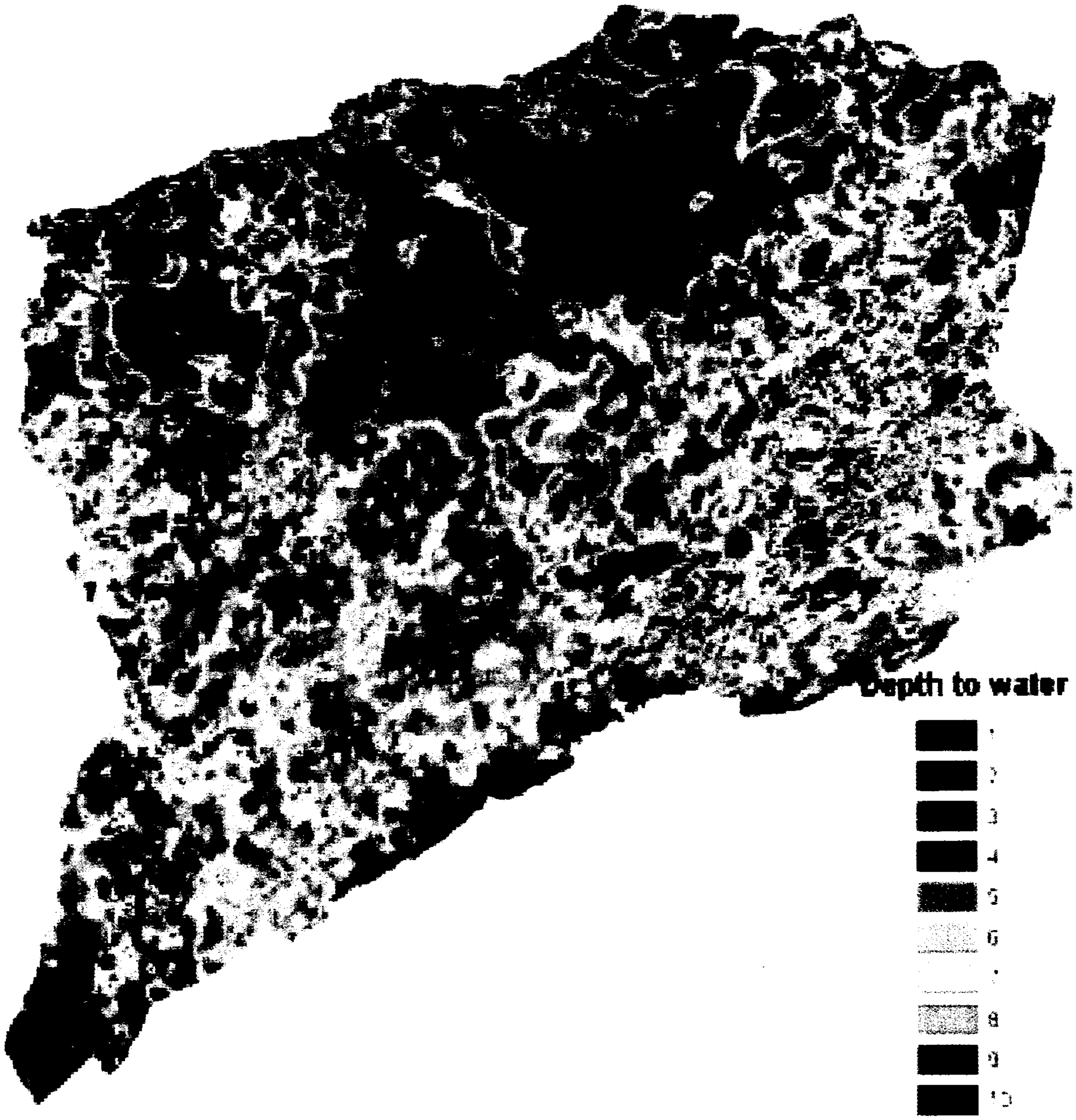
MAP 2.9: SAAT variance



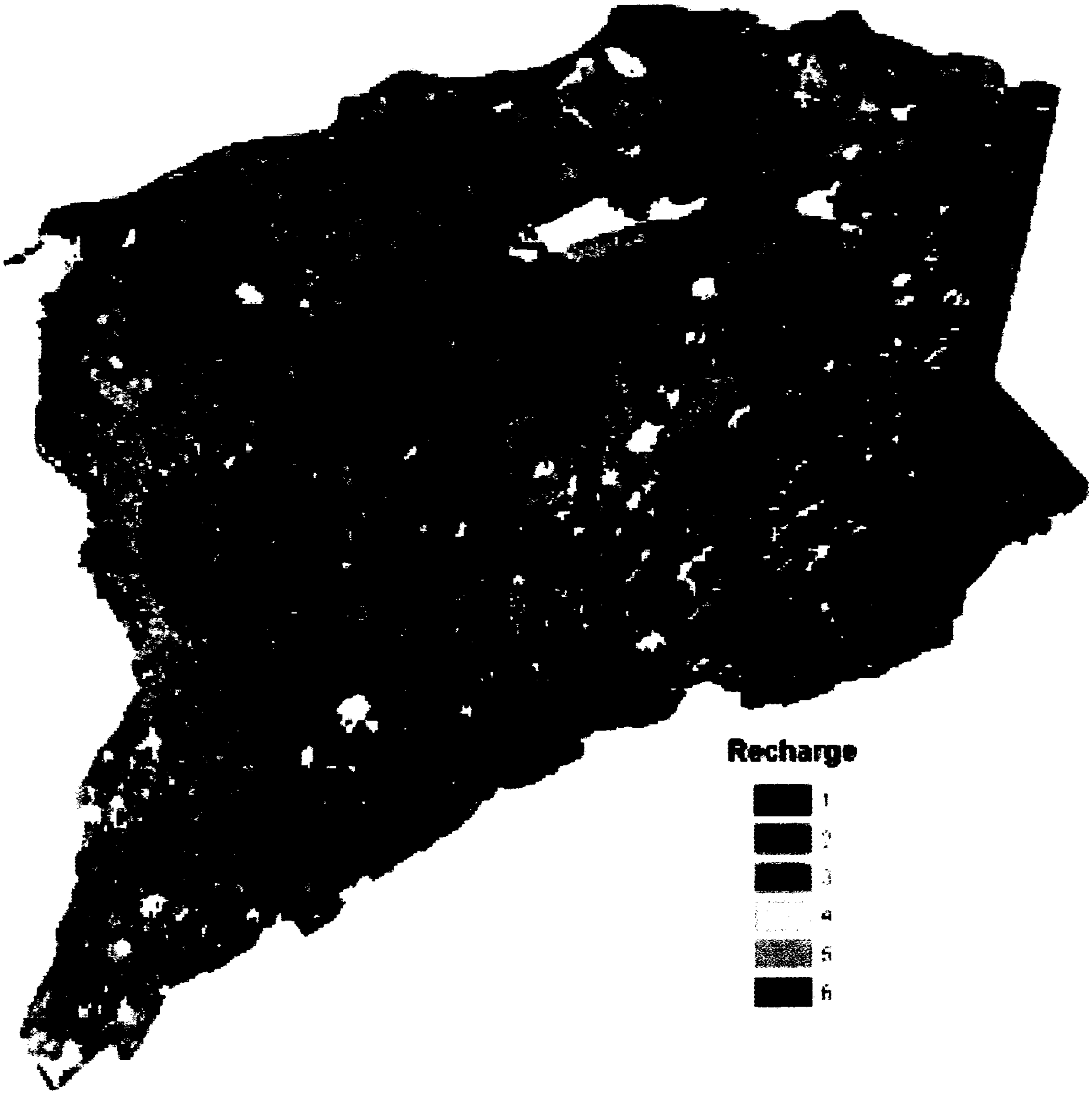
MAP 2.10: SAAT Coefficient of Variation



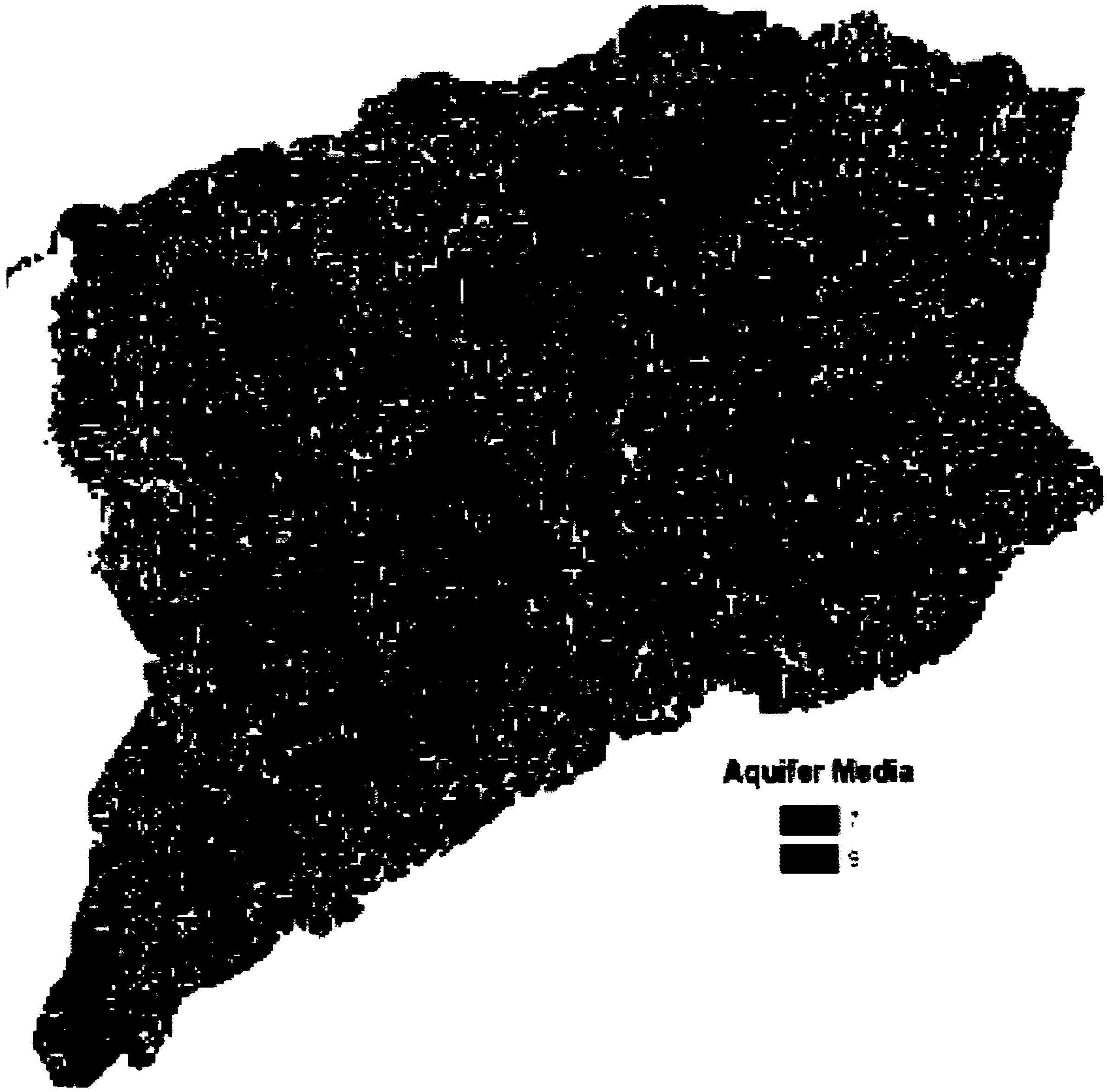
MAP 3.0: Depth to the Water DRASTIC's rating



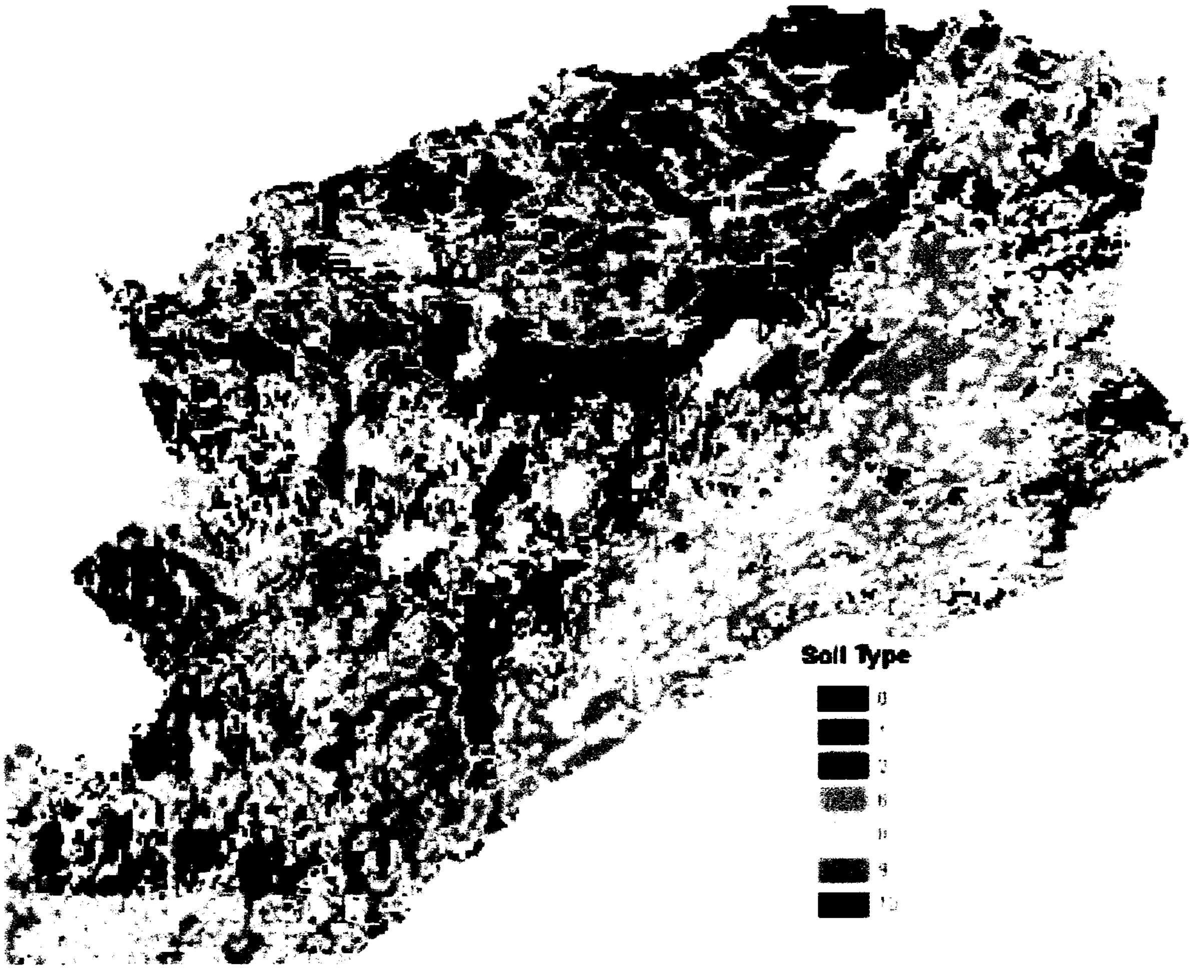
MAP 3.1: Recharge DRASTIC's rating



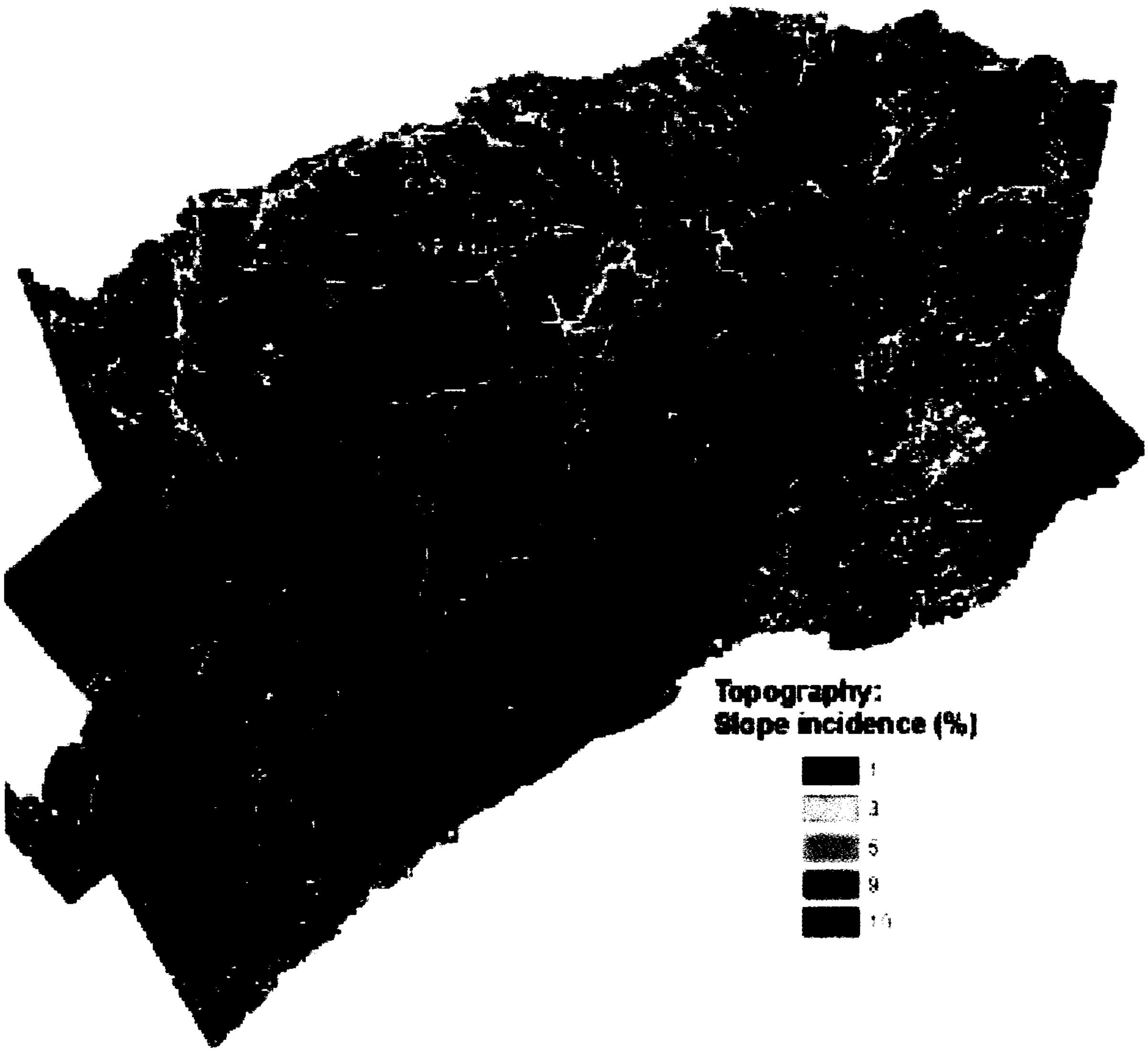
MAP 3.2: Aquifer Media DRASTIC's rating



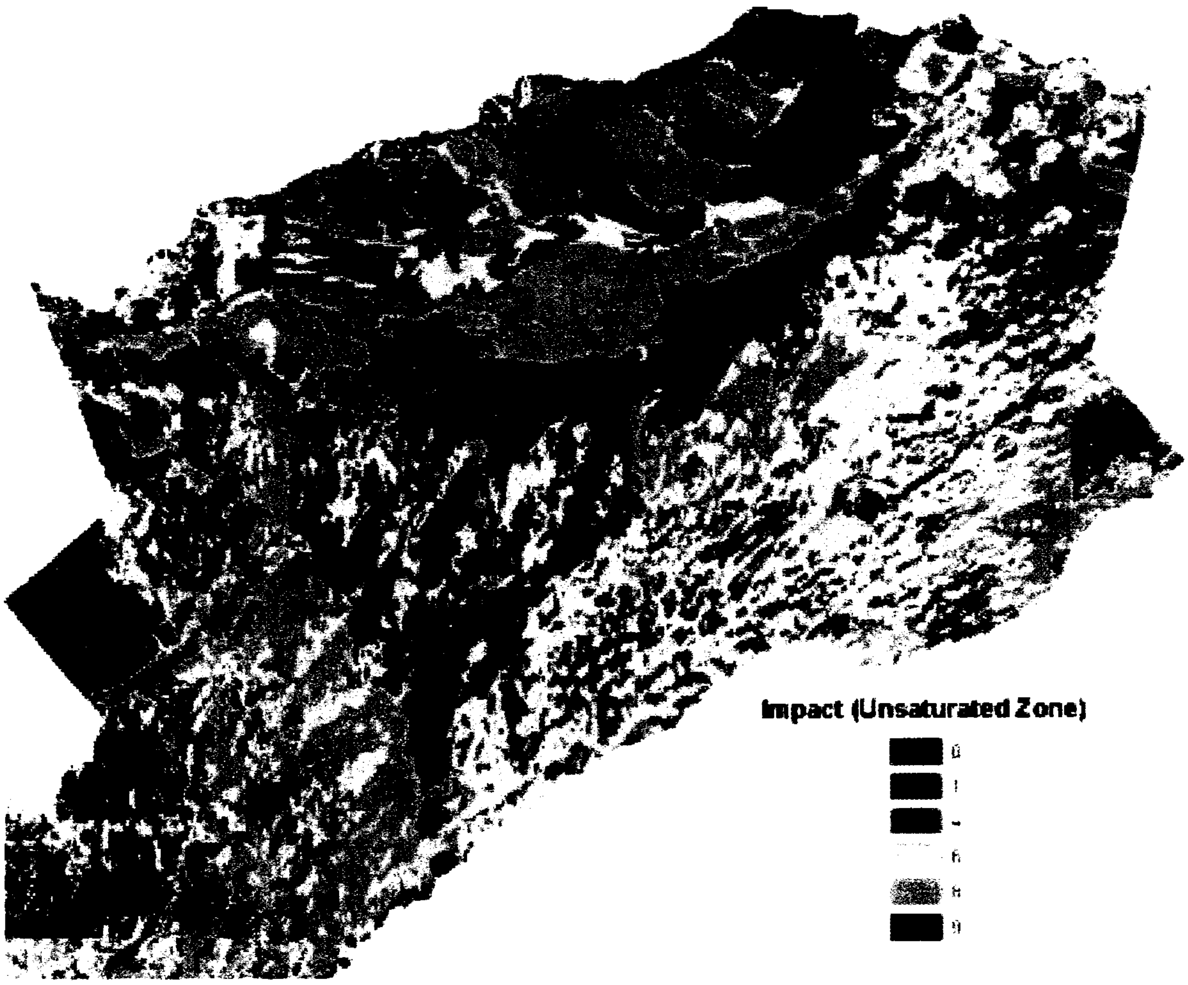
MAP 3.3: Soil Type DRASTIC's rating



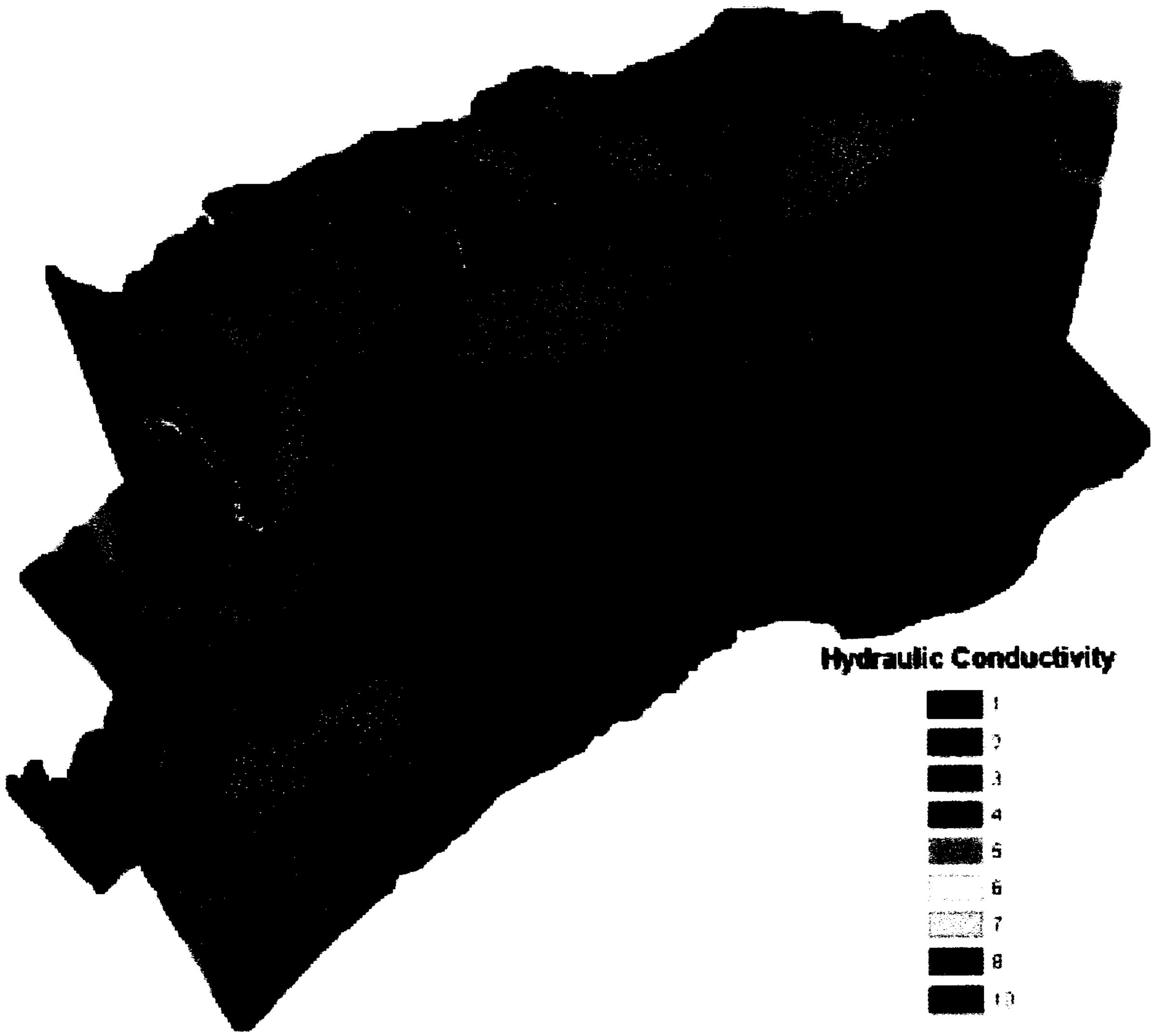
MAP 3.4: Topography DRASTIC's rating



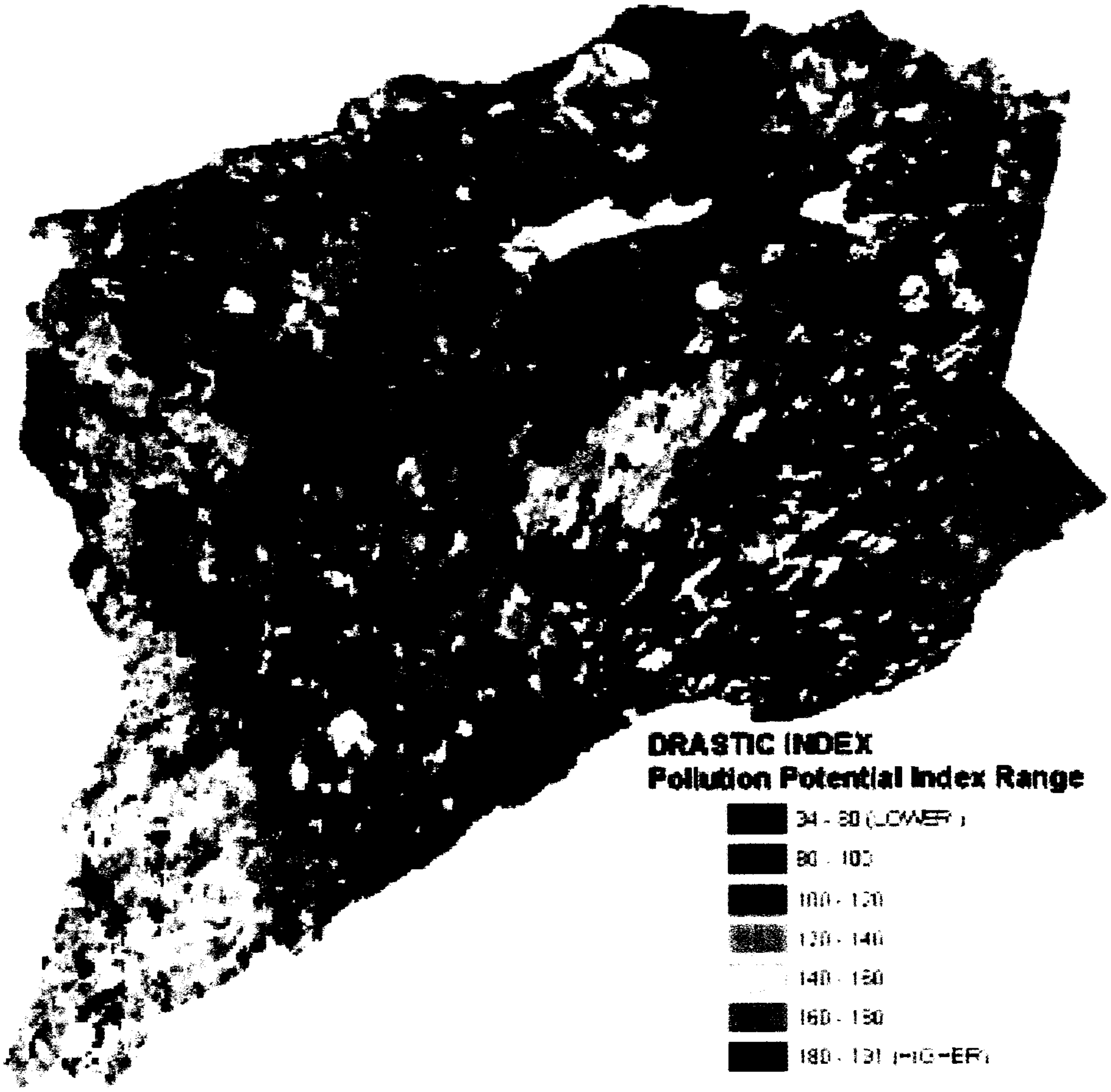
MAP 3.5: Impact of vadose zone DRASTIC's rating



MAP 3.6: Hydraulic Conductivity DRASTIC's rating

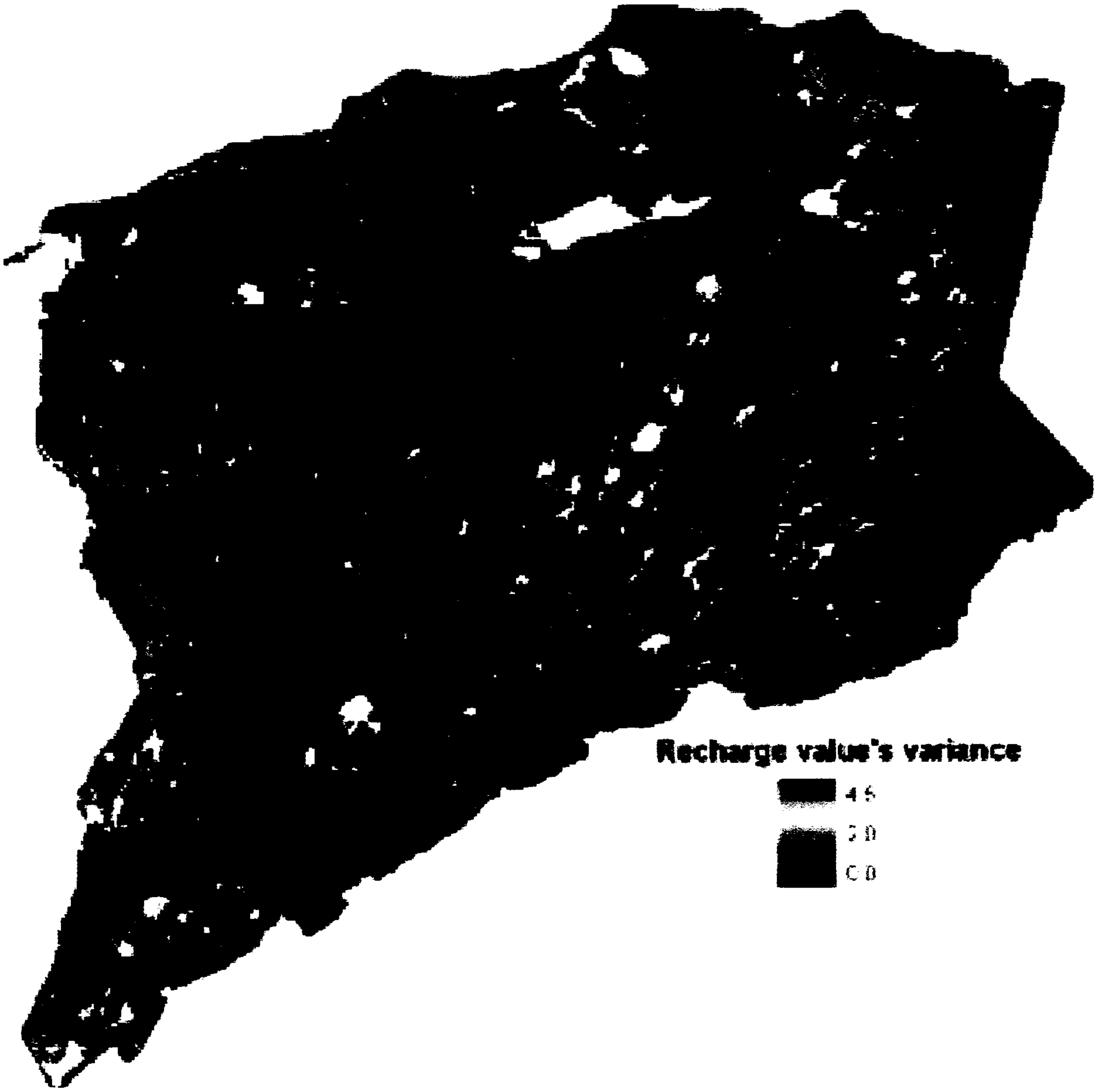


MAP 3.7: DRASTIC INDEX (Pollution Potential Index Range)

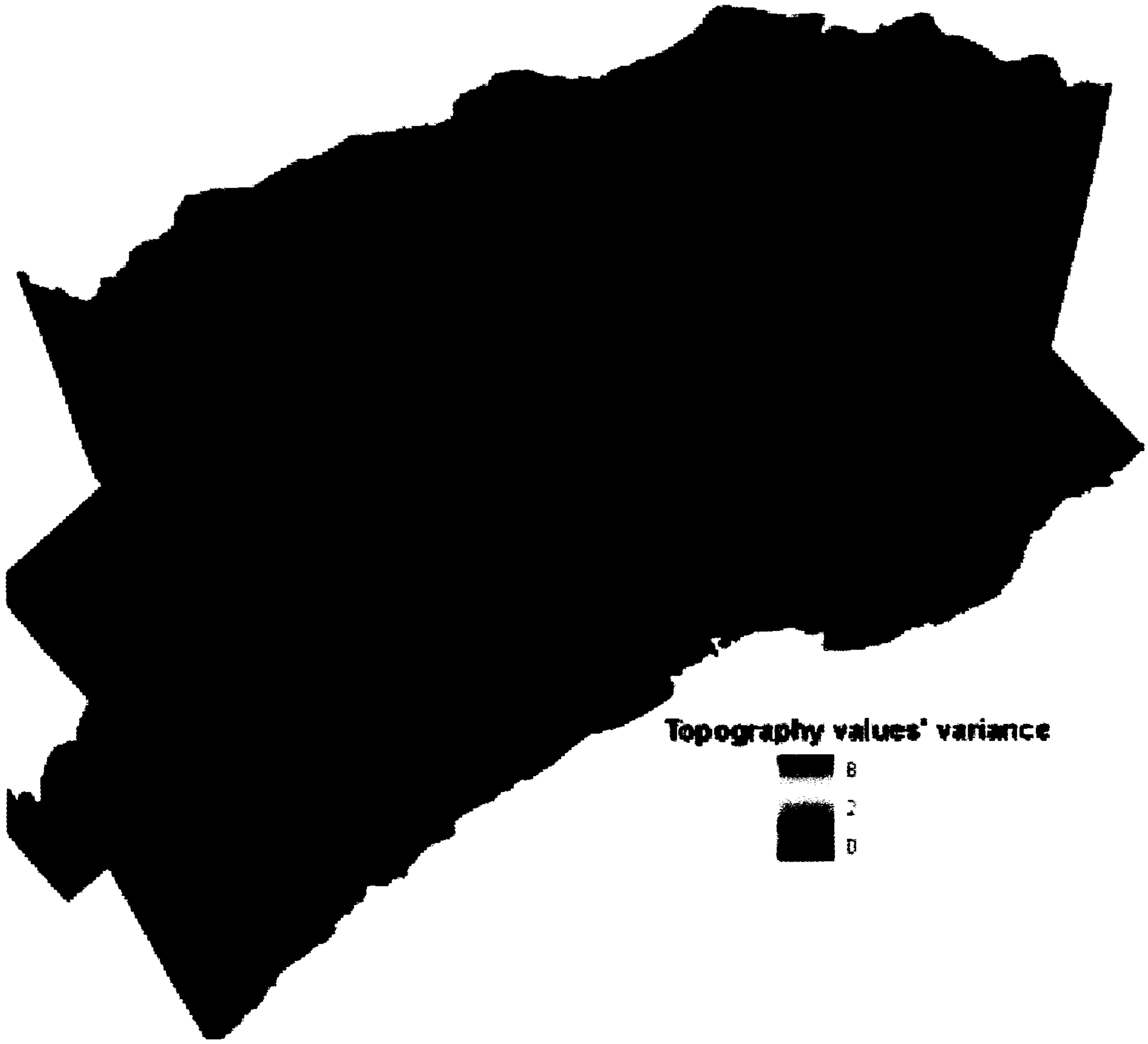


MAP 3.8: Depth to the Water's Variance values (DRASTIC)

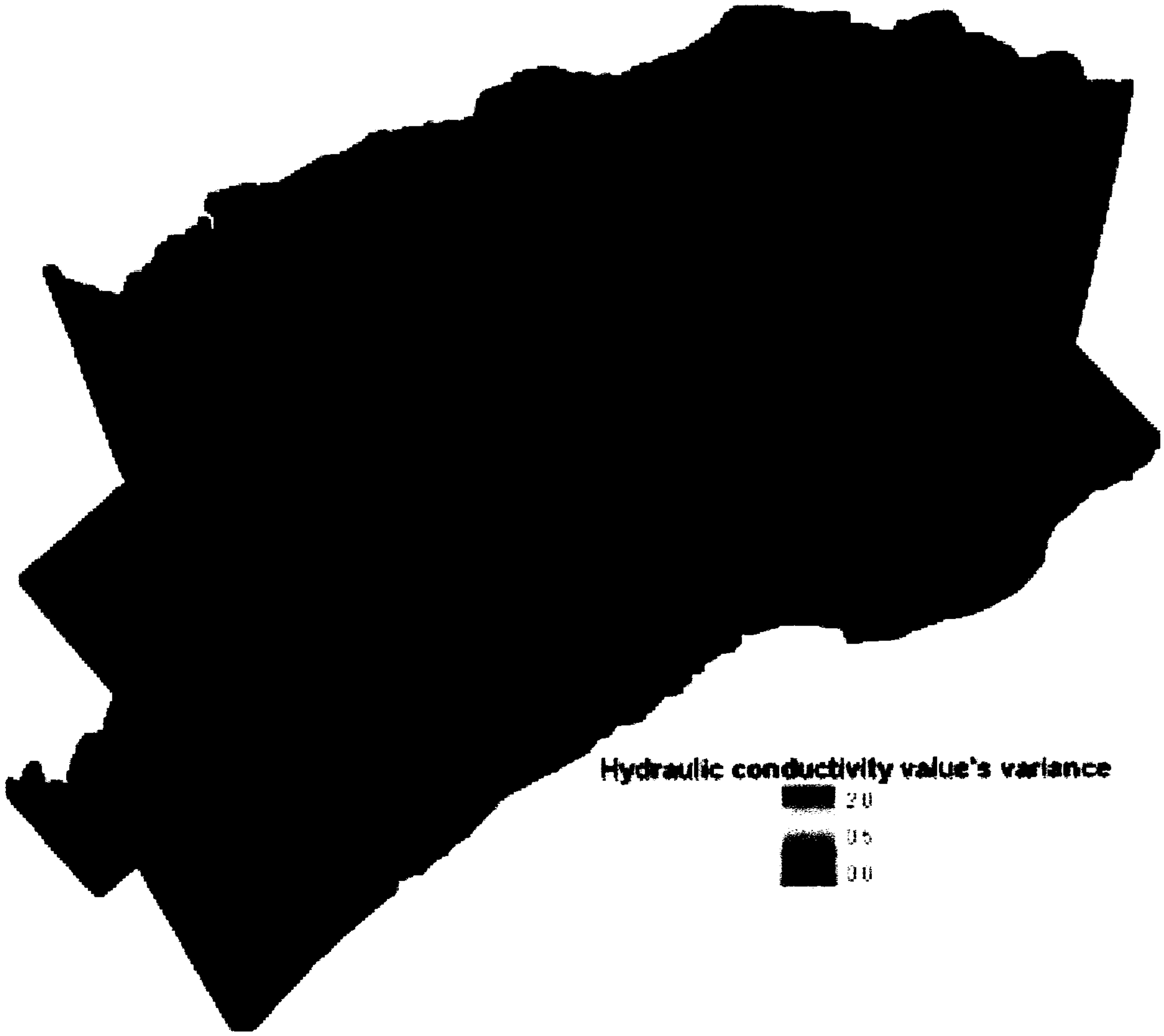
MAP 3.9: Recharge's Variance values (DRASTIC)



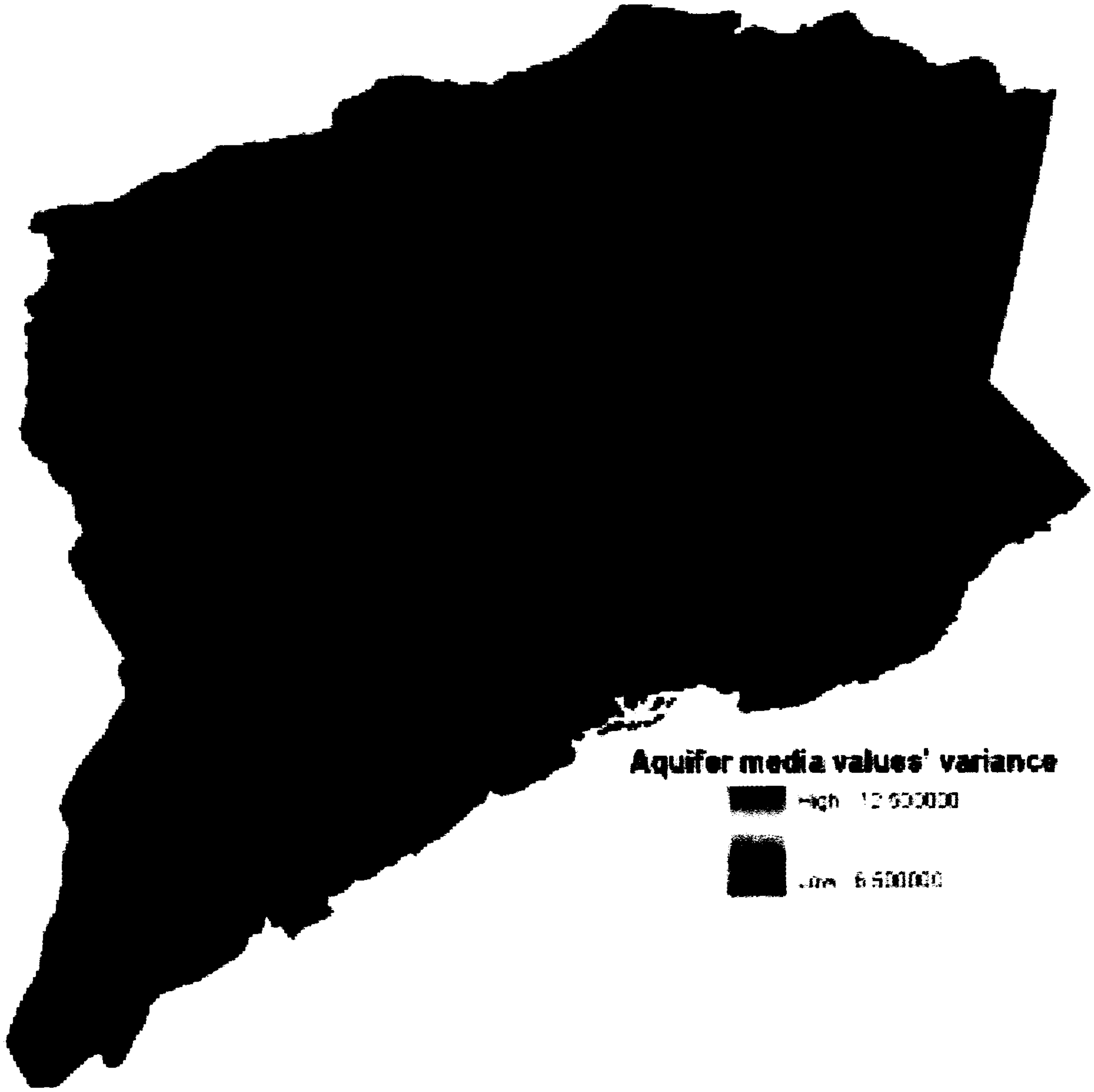
MAP 3.10: Topography's Variance values (DRASTIC)



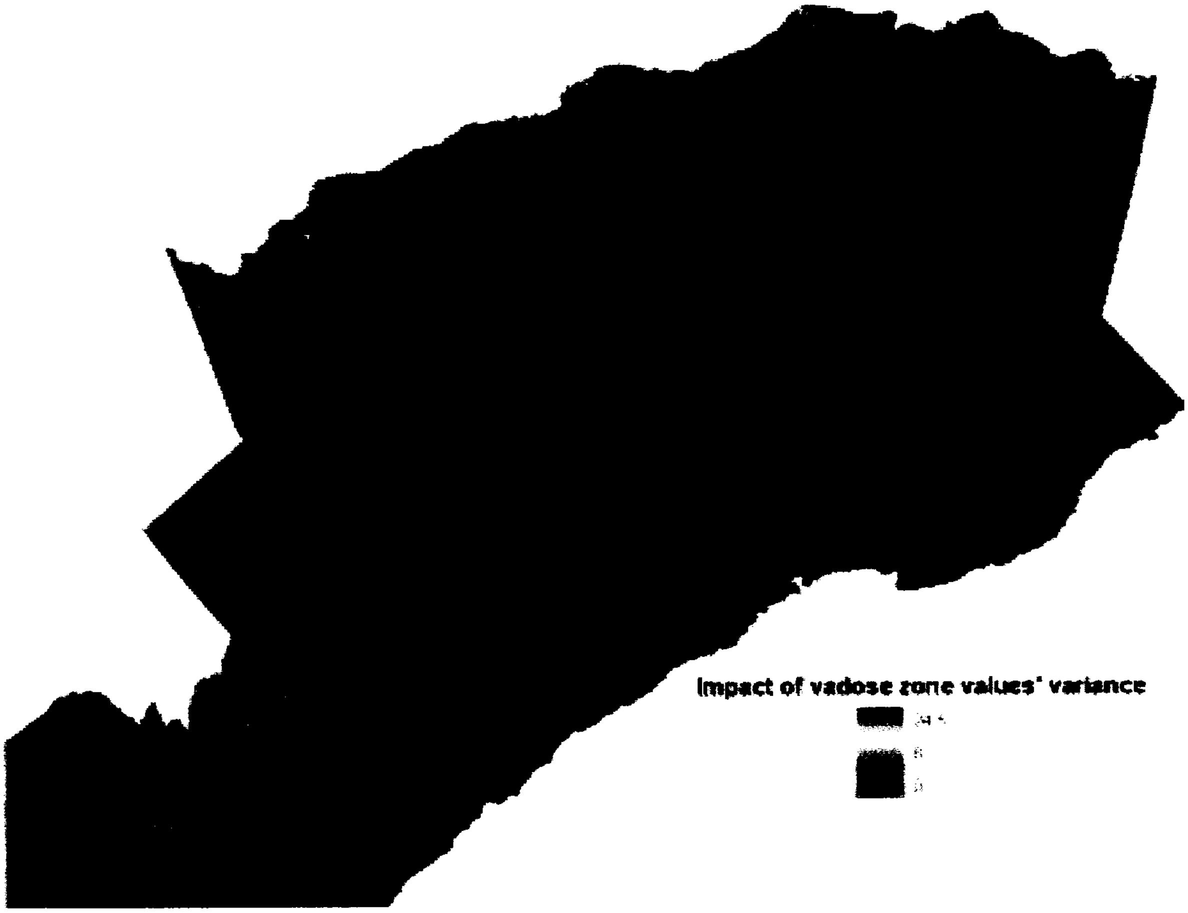
Map 3.11: Hydraulic Conductivity's Variance values (DRASTIC)



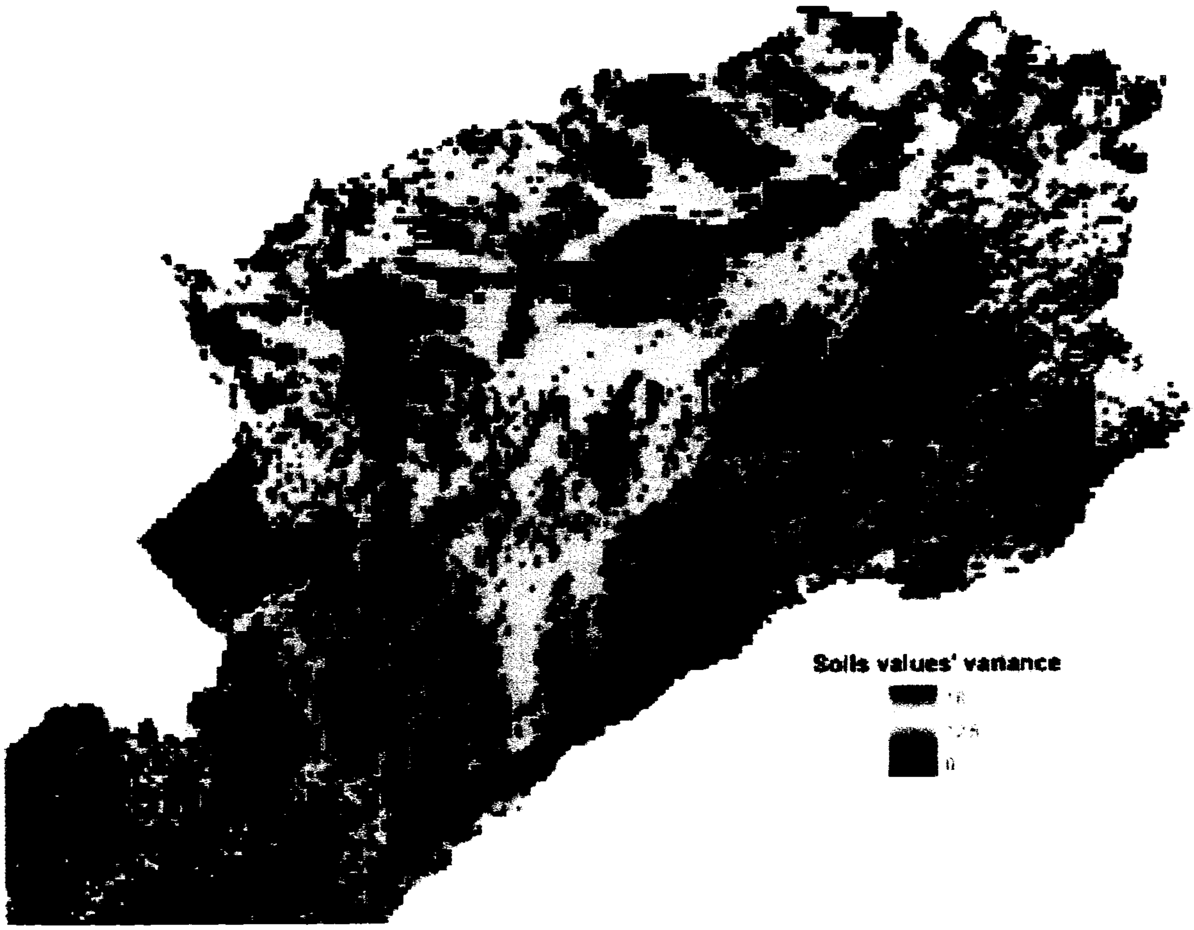
Map 3.12: Aquifer Media's Variance values (DRASTIC)



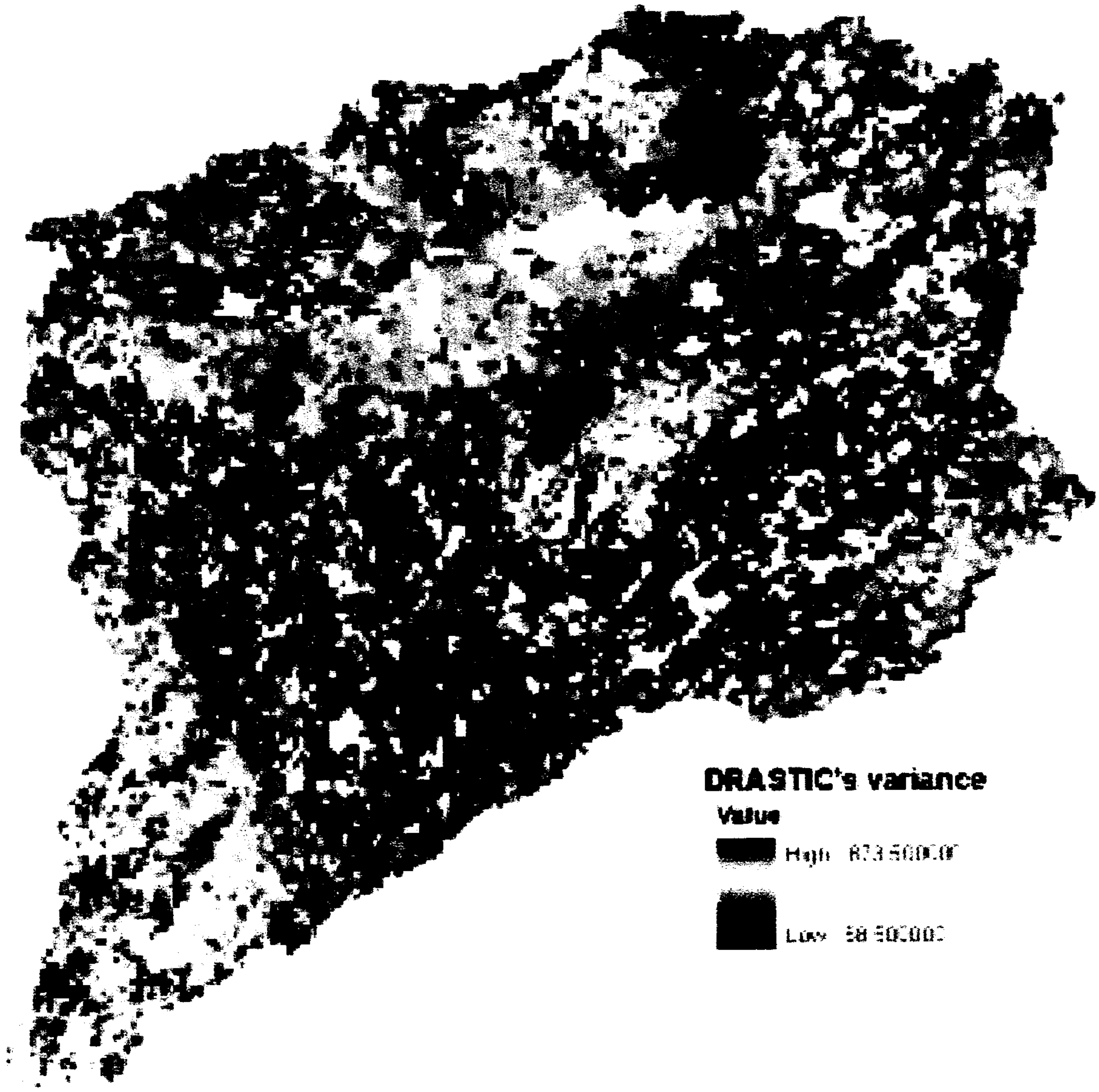
MAP 3.13: Impact of vadose zone's Variance values (DRASTIC)



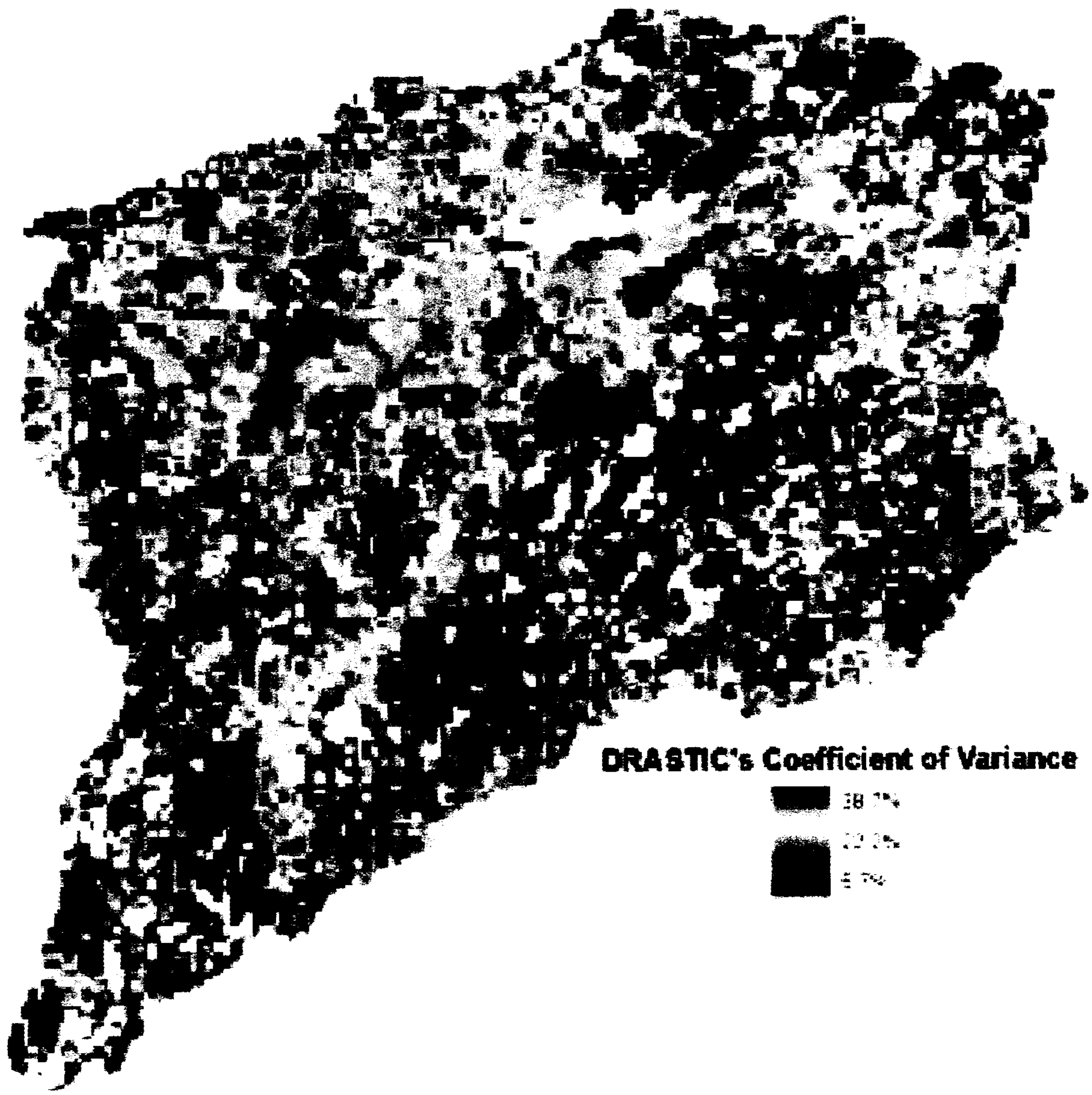
MAP 3.14: Soils' Variance values (DRASTIC)



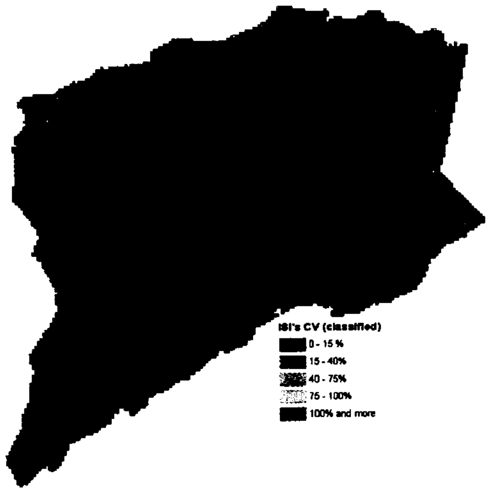
MAP 3.15: DRASTIC's Variance

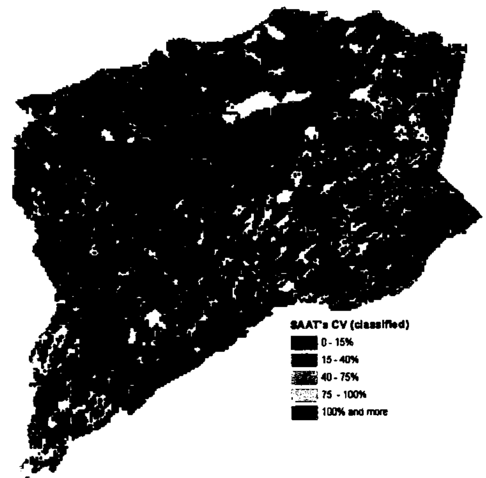


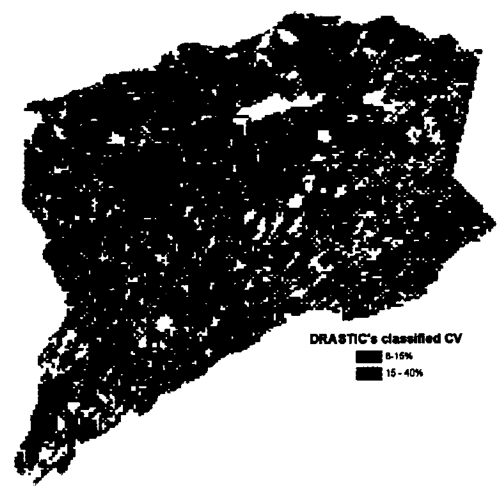
MAP 3.16: DRASTIC's Coefficient of Variation



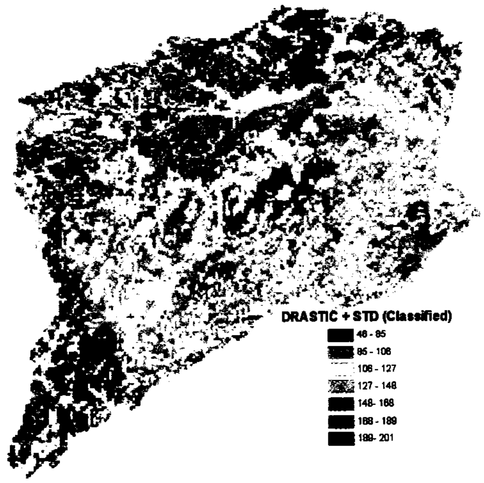
MAP 4.0: Comparison of CVs (ISI, SAAT & DRASTIC)

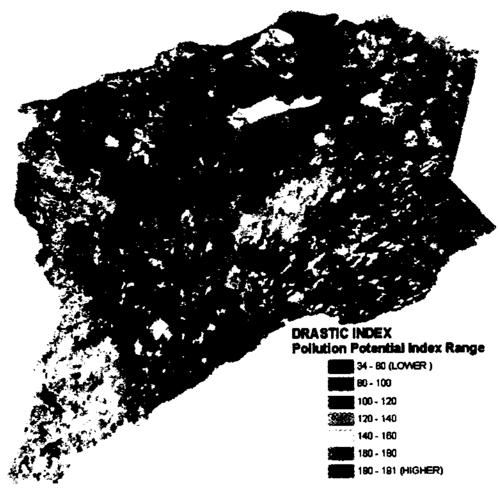


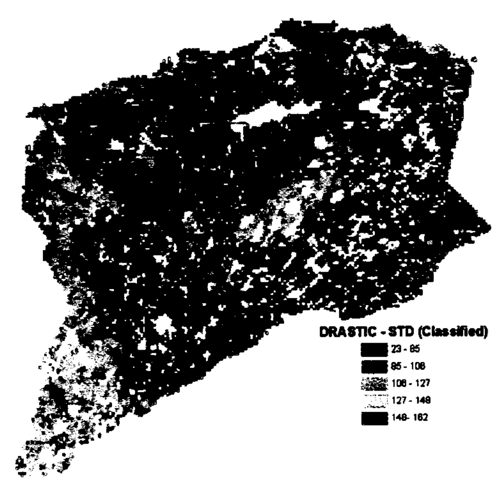




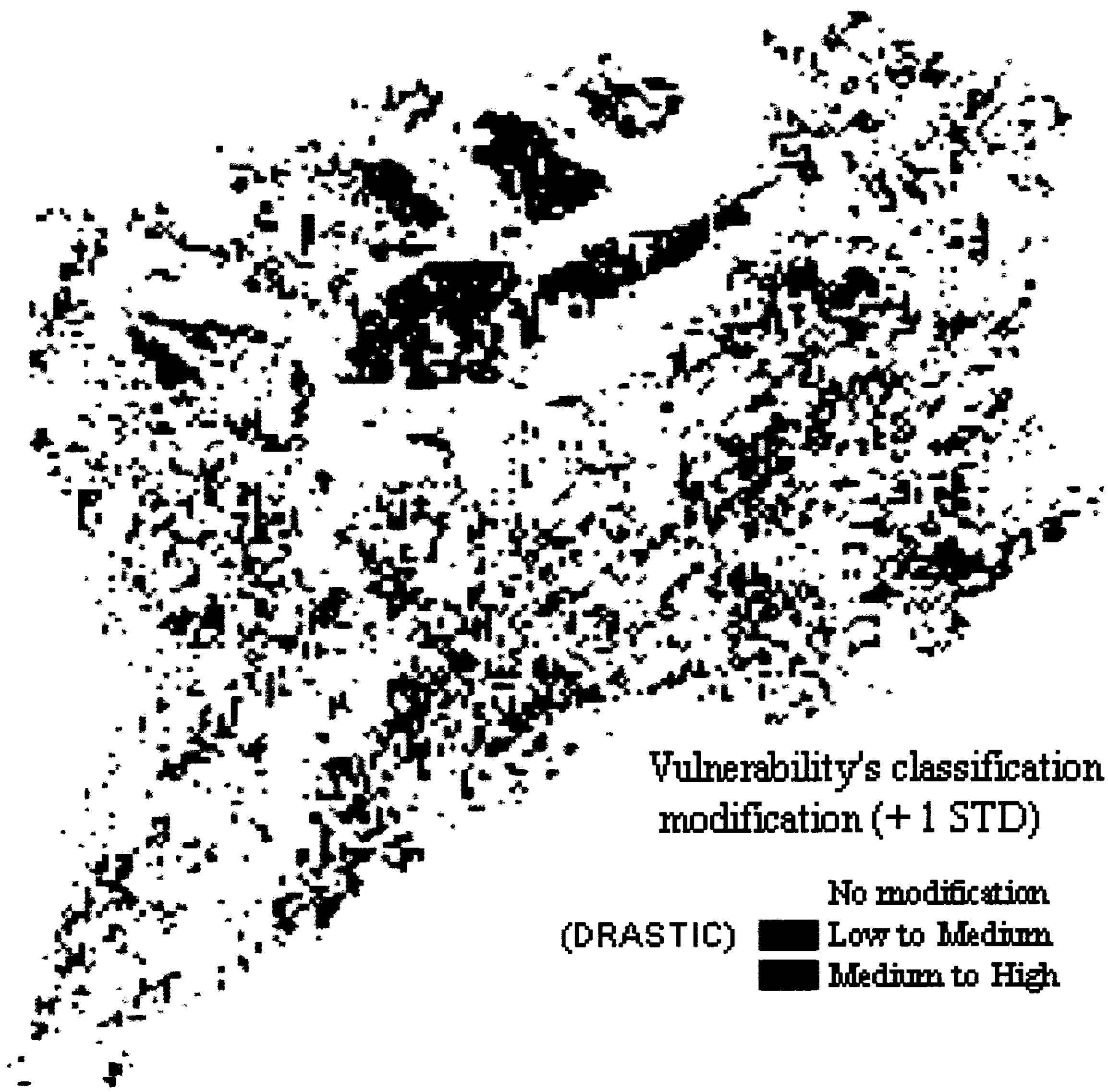
MAP 5.0: DRASTIC vulnerability index classification's modification (+or-1std)



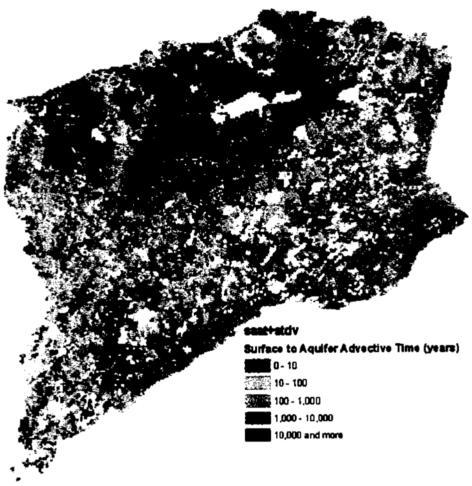


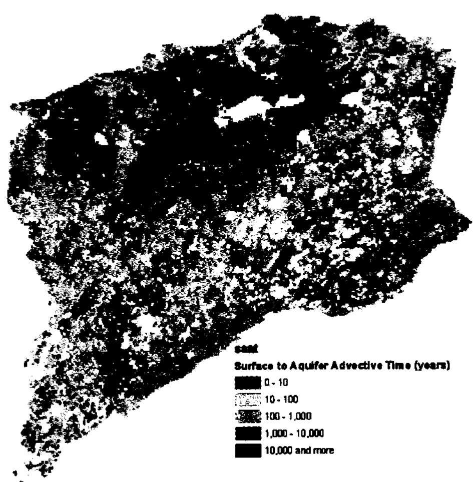


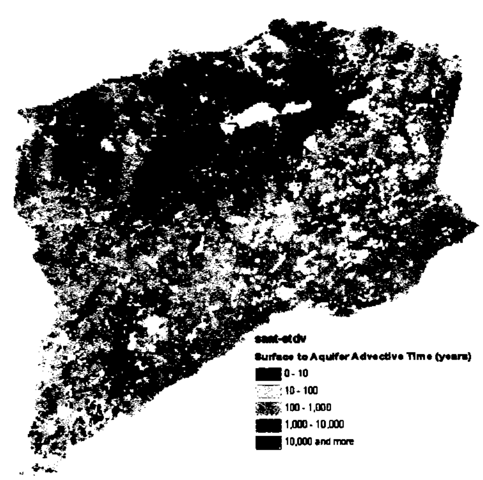
MAP 6.0: DRASTIC Vulnerability's classification modification issued from Standard deviation addition



MAP 7.0: SAAT vulnerability index classification's modification (+or-1std)







MAP 8.0: SAAT Vulnerability's classification modification issued from Standard deviation addition