CHAPTER 2. Literature Review

2.1 Description of the Problem

Sources of pollution are broadly classified as either point or nonpoint sources (Krenkel, 1980). Point sources of pollution, as discrete identifiable locations, include sewered municipal and industrial effluents and discharges from solid waste disposal sites among others. On the other hand NPS, as the result of intermittent releases of pollutants over large areas, are difficult to identify and measure directly.

Nonpoint source pollution enters the receiving surface waters diffusely at intermittent intervals related mostly to the occurrence of meteorological events. There is correlation between the pollutant loading from a watershed and rainfall volume (Novotny and Chesters, 1981). Infiltration and storage characteristics of the basin, the permeability of soils, and other hydrological parameters also play an important role as driving forces of diffuse contamination.

The extent of NPS is also related to geographic, geological and land cover conditions differing greatly in space. The most important waste constituents from diffuse sources are suspended solids, nutrients, and pesticides. If agricultural chemicals such as pesticides and herbicides are placed on the land and surface overland flow is generated by a storm, a significant amount of these contaminants can be lost into surface waters.

The most severe concentrations for point source pollutants carried in surface waters are during low-flow conditions. In contrast, the highest pollutant loading, and in many cases the highest concentrations from diffuse sources, occur during high-flow and flood conditions. Therefore most of the models used for simulating NPS are linked to models of watershed hydrology.

One important issue in estimating nonpoint pollution load from a watershed is the type and extent of activities occurring on the land. Nonpoint source pollution is usually associated with land use. The relative importance and magnitude of the processes (i.e., hydrologic, physical, and chemical), in determining nonpoint loads, will vary among land use categories and associated activities.

The focus on the majority of nonpoint source estimation procedures and models has been on agricultural issues. The entrainment, transport, and fate of sediment, nutrients, and pesticides are largely controlled by the volume and rate of water movement through and across the soil surface. Precipitation, infiltration and surface runoff are the dominant processes.

2.1.1 Existing Nonpoint Source Models

As stated above, the development of models for NPS is linked to the hydrology of the watershed. There are basically two approaches to model diffuse pollution. The more widely used are lumped-parameter models, while more complex models are based on the distributed-parameter concept.

The lumped models were developed at a field-size scale using homogeneous areas. In order to apply them to larger areas "various characteristics of the watershed are often averaged together, and the final form and magnitude of the parameters are simplified to represent the model unit as a uniform system" (Novotny and Chesters, 1981). The distributed approach involves dividing the watershed into smaller homogenous units and adding up the results. While this is the next logical step, it means that calibration data for each field in the watershed is needed.

A real limitation is that runoff and water quality data are collected at only a few points across the watershed and normally at the outlet. The lumped-parameter approach treats the watershed as a hydrological unit using calibrated values for the involved parameters. This simplification tends to represent the model unit as a uniform system and thus limits its use in larger areas.

On the other hand, distributed models take into account spatial variability by dividing the watershed into smaller units with uniform characteristics. It is evident that distributed-parameter models require larger computer storage for performing comparable modelling tasks. A detailed description of the system parameters must be provided and stored for each element.

A key question in distributed modelling is the selection of the criteria for the discretization of the watershed into grid elements. The main difficulty in subdividing watersheds into areas or cells having uniform response is determining what constitutes a hydrological homogeneous area (Kouwen *et al*., 1993). One of the main characteristics of NPS is the spatial variability and its relation to land use. Therefore the common factor for grouping and selecting the cell size on most of the distributed models is the type of landcover. It should be recognized though, that land use management may cause different responses.

In the attempt to maintain a manageable number of grid elements as the watershed area increases, the assumption of uniformity is normally violated. The most common approach on hydrologic models is to obtain a response for each grid element by weighting the values of the parameters related to landcover area. This is known as the hydrologic response unit approach. Often this assumption of homogeneity, commonly related to landcover, dictates the grid size used to model the watershed.

At this point, special mention is required for the grouped response unit approach. It is based on calculating the response for each of the landcover classes within the element and then weighting the response by area. Grid cell response will then depend on the landcover fractions within the element.

Different types of landcovers can have a wide range of response characteristics; thus the grid elements are made up of different landcovers, each with its own response characteristics that are assumed to be uniform. It is this approach, grouping responses for different landcover classes, that can produce improvements in nonpoint source modelling. This grouped response unit approach forms the basis for the hydrologic model WATFLOOD (Kouwen, 1988).

Transferability, the capability to calibrate a model in one scenario and apply it in a different area, is also an important attribute of a such a scheme. If the model parameters are associated with landcover classes, they can be transferred to other watersheds that would have the same landcover classes but with a different distribution.

The size of the element is not restricted by the assumption of hydrologic homogeneity, only to a size where travel times within the element are small compared with the overall basin travel time. Thus the location of the responding units is not significant. Only the percent of each land class is necessary to characterize a grouped response unit (Kouwen *et al*., 1993).

Since the early 1970s, a large number of NPS models have been developed. Reviews of the available runoff-water quality models applicable to diffuse pollution modelling of urban and agricultural watersheds have been prepared by Giorgini and Zingales (1986), Rose *et al*. (1988), and Donigian and Huber (1990) among others.

Table 2.1 contains a summary of the features from some of the reviewed hydrologic and agricultural models. It can be noted that WATFLOOD is included to compare its hydrologic capabilities. Part of this research intends to add a contaminant component to WATFLOOD, based on the grouped response unit approach outlined above.

Following is a brief description of the reviewed models with the intention of determining what is already done, what could be improved, and what is still needed.

Table 2.1. Summary of Models Characteristics

Model	surface runoff	soil water	ground water	single event	continous	lumped	distributed (response)	sediment	nutrients	pesticides	Source
	Tasks, simulation, type and pollutants modeled										
AGNPS[1]	●			●	●		●h	●	●	●	USDA-ARS, Morris Minnesota
ANSWERS[1]	●			●			●h	●	●		University of Georgia, Tifton
ARM	●				●	●		●	●	●	EPA, Athens, Georgia
CREAMS-GLEAMS	●	●	●	●	●	●		●	●	●	ASDA-ARS, Tifton, Georgia
GAMES	●			●			●h	●			University of Guelph, Canada
HSPF[2]	●				●	●s		●	●	●	EPA, Athens, Georgia
SWAT	●				●	●s		●	●		USDA-ARS, Temple, Texas
SWRRB[2]	●				●	●s		●	●	●	USDA-ARS, Temple, Texas
WATFLOOD[1]	●	●	●	●	●		●g				University of Waterloo, Canada

1 - For distributed type response: (h)-hydrologic response unit, (g)-grouped response unit
2 - Watershed spatial domain allowing sub-basin division.

AGNPS - *Agricultural Nonpoint Source Pollution Model* was developed by the US Department of Agriculture (Young *et al*., 1986). It can simulate sediment, nutrient and pesticide loads from agricultural watersheds for a single storm event or for a continuous simulation. The watershed must be divided into uniform square cells where computations are done, and runoff, sediment, nutrients and chemicals are routed from cell to cell from the watershed boundaries to the outlet. The hydrology is calculated by the Soil Conservation Service (SCS) runoff curve number approach, combined with a unit hydrograph type for uniform rainfall. Soil erosion is based on the Universal Soil Loss Equation (USLE). Simple correlation for extraction of nutrients and pesticides in runoff and sediment forms the water quality component of the model.

ANSWERS - *Areal, Nonpoint Source Watershed Environment Response Simulation* was developed by the Agricultural Engineering Department of Purdue University (Beasley and Huggins, 1985). It is a distributed parameter and event oriented model. The watershed is divided into uniform square elements ranging from 1 to 4 hectares. Within each element the model simulates processes of interception, infiltration, surface storage, surface flow, sediment detachment (USLE) and transport. The output from one element becomes the input to the adjacent one. It is primarily a runoff and sediment model; the nutrient simulation is based on simple correlation between chemical concentrations, sediment yield, and runoff volume.

ARM - *Agricultural Runoff Management Model* is a version of the HSP-F *Hydrologic Simulation Program in Fortran* that was originally developed from the Stanford Watershed Model (Donigian and Davis, 1985). It is a large, lumped model and requires considerable effort when applied to a watershed. It is capable of simulating a hydrologic time series event, including hydrographs and conventional pollutants. The model uses a basin-scale analysis framework that includes fate and transport in one-dimensional stream channels. It integrates the simulation of land runoff processes (SCS) with in-stream hydraulics, sediment detachment (USLE), transport, and nutrients. Besides the complex and large amount of data needed, the model requires extensive calibration and application for large drainage systems is very limited.

CREAMS - *Chemicals, Runoff, and Erosion from Agricultural Management Systems* was developed by the US Department of Agriculture (Knisel, 1980). It is a field scale lumped approach model that uses separate hydrology, erosion, and chemistry submodels, connected by shared files. It can simulate continuous series, using the SCS runoff curve number, when daily rainfall data are available or single events with hourly rainfall data using the Green-Ampt equation. The erosion component of the model considers the basic processes of soil detachment (USLE), transport, and deposition. The basic concepts for nutrient modelling treat their transport as proceeding separately in adsorbed and dissolved phases where soil nitrogen is modified by nitrification-denitrification processes. The GLEAMS (Groundwater Loading Effects of Agricultural Management Systems) module is essentially a vadose zone component for CREAMS.

GAMES - *Guelph model for evaluating effects of Agricultural Management Systems on Erosion and Sedimentation* (Dickinson and Rudra, 1990). Developed to describe and predict soil loss by fluvial erosion and the delivery of suspended solids from agricultural fields. The analyisis of erosion is achieved throught the use of the USLE with modifications to the rainfall erosion index and to the soil erodibility factor for local and seasonal conditions. The SCS curve number method drives the hydrology of the model. The discretization of a watershed into field sized elements is done based on homogeneity of land use, soil type and slope.

SWAT - *Soil-Water Analysis Tool* developed by the USDA-ARS, Temple, Texas, to help water resource managers in assessing water supplies, soil erosion, and water and sediment transfers through watersheds (Arnold *et al*., 1995). The SWAT model estimates surface runoff volume using the Soil Conservation Service (SCS) curve number procedure. Overland sediment yield is computed using the Modified Universal Soil Loss Equation (MUSLE). Nutrient yield and nutrient cycling use the algorithms developed for the EPIC model (Arnold *et al*., 1993). SWAT allows for simultaneous computations on each sub-basin and routes the water, sediment and nutrients from the sub-basin outlets to the basin outlet.

SWRRB - *Simulator for Water Resources in Rural Basins* was developed for evaluating basin-scale quality in rural watersheds (Williams *et al*., 1985). It operates on a daily time step and simulates hydrology, crop growth, sedimentation, flood plain degradation, and nitrogen, phosphorus, and pesticide movement. The lumped approach model was developed by modifying the CREAMS model for applications to larger rural basins. Surface runoff is calculated using the SCS curve number technique and sediment yield is computed using the modified USLE. More information is not available to judge data and calibration requirements.

WATFLOOD - A full description of the hydrologic model can be found elsewhere (Kouwen, 1988 and 1993). Briefly, it is a distributed hydrologic modelling system that uses the grouped response unit concept. It can use spatially distributed meteorological data, for example from radar, for rainfall estimation. The model accounts for the dominant short duration rainfall-runoff processes including interception, surface storage, infiltration, interflow, and overland flow as well as the slower processes of snowmelt and evapotranspiration.

As each element response is based on land class area, the runoff is modeled in an element by adding the contributions from each land cover type and routing the results to the drainage system. It uses the storage routing technique to route the water through the channel system. The system includes a pattern search optimization technique to estimate the model parameters that cannot be previously assigned with standard values.

2.1.2 GIS and Decision Support Systems

Geographical Information Systems (GIS) can be thought as a means of storing and retrieving spatially varied data. In the strictest sense, a GIS is a computer system capable of storing, manipulating, and displaying geographically referenced information. They are ideally suited for studying the processes and impacts of diffuse pollution (Connors and Gardner, 1991).

In a GIS system, information about the spatial characteristics of a geographic area can be stored in a grid system (pixel value or vector-polygon object). Information is stored for each grid element and several layers can be used to account for different types of data (i.e., soil type, elevation, and land use). Manipulations, such as overlaying, can be used to extract additional information.

There have been several efforts using GIS technology in the field of nonpoint source pollution. Stuebe and Johnson (1990) used a GIS system for estimating runoff volumes. DeRoo *et al*. (1989) used a GIS system with ANSWERS to model soil erosion. Rewerts and Engel (1991) and Srinivasan and Engel (1994) used a public domain system GRASS to facilitate the input file creation and output visualization for the ANSWERS and AGNPS models. Bekdash *et al*. (1991) evaluated best management practices in agricultural lands using a linkage of a GIS and the CREAMS model.

One step further, to facilitate watershed management and planning, is to integrate different sources of information and knowledge into what is called spatial decision support systems. Lam *et al*. (1994) described an approach to build an environmental information system using RAISON (Regional Analysis Information System) as the base. It is expected that further development will allow it to become a system that will be part of the many research tools needed for better watershed management and planning.

While there are some similarities to GIS (Lam and Swayne, 1991), the RAISON system differs significantly as it emphasizes decision support and expert systems analysis that are difficult or impossible to achieve with traditional GIS. One of the important RAISON features to be used in this work, is its capability to incorporate modelling tools into the system.

This is done by building interfaces that interact with existing models and that can intercept the input and output to connect to the database in the system. Booty and Wong (1994) linked a water quality model within the RAISON system to simulate river flow, effluent advection and dispersion to study the effects on downstream concentrations in the Athabasca River.

2.1.3 Sensitivity Analysis

Sensitivity analysis can be defined as a procedure to determine the relative change in the results of a model due to changes in parameters values. If a small change in a parameter results in relatively large changes in the result, the model is said to be sensitive to that parameter. This may mean that the parameter has to be determined very accurately or that the model has to be redesigned. The most common approach in runoff models to analyze sensitivity is based on direct parameter sampling and normalized sensitivity coefficients (James, 1992), indicating the percentage change in the result due to individual parameter perturbations.

The two major drawbacks of this technique are: (a) the variations are referred to a base solution, e.g. precipitation input function for runoff models, and (b) different locations in space can have different responses to the same parameter perturbation. This results in a large number of runs required to assess the sensitivity of the model. Even with its limitations, it is a very powerful technique to detect the parameters that affect the results the most and others to which the model is less sensitive. The variations in the response of a model to changes in the parameters are also associated with the uncertainty in the values assigned to such parameters. There is an uncontrollable random component inherent to any parameter estimation. It is important to differentiate between risk and uncertainty analysis.

The distinction is that in a risky situation, the uncontrollable random event comes from a known probability distribution; whereas in an uncertain situation the probability distribution is unknown. Analysis of scenarios explores the effect on alternative strategies of changes in input. This is a "what-if" type of analysis with the “what-ifs” being external to the model parameters. The variation of the response to a parameter can be minimized if the variance (uncertainty) associated with such parameter can be reduced. However, if the sensitivity of the outcome to the parameter is small, reduction of its variance may not result in an improvement.

The basic concepts and numerical approaches for parameter sampling have been largely dominated by stochastic techniques such as Monte Carlo or Latin Hypercube sampling procedures. In the Monte Carlo analysis (Hornberger and Spear, 1980), the full distribution of the model output is accomplished through a very large number of simulations where the parameters are randomly selected according to their probability distributions.

The random sampling technique can assume uniform distributions and independence between parameters, requiring large number of samples to properly define the tails of the distributions. In the Latin hypercube procedure (McWilliams, 1987), the range of variation is partitioned into intervals of equal probability and uses a random selection of parameters within each of the intervals reducing the required number of samples.

A serious drawback of using deterministic models with stochastic sampling techniques is the failure to detect the random variation of the output. Only the uncorrelated variations of the input and system parameters can be adequately simulated. If this input and system parameters are cross-correlated, the sampling procedures must be modified to incorporate the cross-correlation, which is tedious and sometimes impossible. Recently, as a result of improvements in decision theory and artificial intelligence, a set of probabilistic approaches under high uncertainty has emerged (Shafer, 1990). These techniques, known as belief networks, are based on the principle of networking nodes representing conditional and locally updated probabilities.

This allows construction of large and densely coupled (interrelated) networks. Furthermore, without excessive growth in computation, such networks can be constructed to operate interactively and on-line. Varis (1995) suggested a methodology to use a belief network 'below' a deterministic model approach to deal with uncertainty in optimization and parameter estimation. This technique is spreading quickly to many application areas.

2.2 Integration in a Decision Support System

2.2.1 Selection of Models

NPS modelling is strongly affected by land use activities with its performance tied to the ability to model surface runoff and sediment erosion. Distributed models based on the hydrologic response unit consider each grid element to be homogeneous. As described in Section 2.1.1, there are numerous distributed models that simulate the fate and transport of diffuse pollutants at a watershed scale.

The AGNPS model is currently one of the proposed nonpoint source pollution models for use at the National Water Research Institute (NWRI) in Burlington, Ontario, Canada. As part of this research, work was done to link it with the RAISON system.

The AGNPS model was chosen for this study due to its accessibility, its multigrid division capability and the ability to simultaneously simulate water quantity and quality in different parts of the watershed. The AGNPS model is a well established and tested simulation event model (Mostaghimi *et al*., 1997, Bingner *et al.*, 1989, Finney *et. al.*, 1995). Due to its distributed scheme it is also a good choice for GIS integration and it is used to compare results of the addition of the proposed water quality component into the WATFLOOD model.

From the model review (see Table 2.1), it can be noted that the models with a distributed approach are based on the hydrologic response unit concept. AGNPS falls in this category. A distributed approach for which homogeneity in its elements is not required (ie. based on the group response for different land classes), can be a better choice for modelling NPS. A limitation of most models is that uniform precipitation is assumed during the event.

This is acceptable at field scale, but when the simulated area increases, the spatial variation of the rainfall must be taken into account. The availability of radar precipitation data and recent advances in the remote sensing of land cover characteristics, together with GIS tools to store and manipulate such data and a formulation that takes advantage of this information, could yield to an improvement in NPS modelling.

This research is based on selecting a water quality component to simulate the processes governing fate and transport of agricultural pollutants. This component would be incorporated into the distributed hydrologic model WATFLOOD, which uses the group response unit approach and was designed to account for the spatial variability of runoff.

2.2.2 Water Quality Component

To develop a water quality component that is appropriate at a watershed scale, a compromise must be established between the relationships that describe the processes at the microscale (such as adsorption, volatilization, and rainfall effect on soil erosion), and those that will be appropriate at the mesoscale group response unit approach.

Because most of these processes have being developed at a field scale, it is a concern that when applied to wider areas, the relationships used to simulate such processes, nevertheless correct, are taken out of context and stretched beyond their limits in most of the distributed models. By weighting the parameters and not the response for each land class places the effect of the variability on the averaged parameters.

Furthermore, the role that the distribution of landcover would have in the calibration of the parameters is hard to identify. This is the hypothesis to test in this part of the research. The main task was to identify and adapt the governing processes to calculate the response for each land class and then weight the response by area within the unit. In this way, by using the grouped response unit concept, the values of the parameters describing the response processes can be calibrated based on the landcover alone. This should lead to the possibility of transferring the model to other areas without the need of recalibration.

Soil Erosion.- Even though erosion, sediment transport, and deposition are to a large degree natural processes, sediment *per se* are considered a major pollutant in receiving waters. Soil erosion is the major cause of diffuse pollution and sediment is the most visible pollutant and a primary carrier of organic components, phosphates and metals (Beasley *et al*., 1984).

Soil erosion depends on particle size, soil texture, and the presence or absence of protective surface cover such as vegetation. Vegetative cover is extremely important since it provides additional resistance to shear stresses caused by falling and running water. Hydrologically the erosion processes are classified as overland and stream or channel erosion. Many factors, such as distance from source to streams, vegetative covers, slope, and roughness characteristics of the land, together with the presence of depositional areas during overland flow, affect the delivery of the sediment to the receiving body of water.

As noted from the reviewed models, most of them use the Universal Soil Loss Equation (USLE), or modifications of it, to estimate the soil loss caused by rainfall and runoff. Despite the empirical nature of the method, it is still the most widely used and validated technique to estimate soil erosion. The sediment component used in this research is based in the Simplified Process (SP) model developed by Hartley (1987). It considers the transport capacity of surface runoff and the soil erosion resulting from both runoff shear stress and rainfall impact. The sediment yield equals the smaller of the sediment transport capacity or the supply rate.

This method was selected due to a stronger physical basis than the USLE and therefore it was judged to be more suited to work coupled to the process oriented WATFLOOD model. Some modifications were required to take into account the GRU approach. The method was verified for unit consistency and modified to use the hydrology from WATFLOOD and incorporated as a subroutine into the system.

Nutrients.- In order to provide better plant production rates, fertilizers rich in nitrogen, phosphorus and potassium are applied on agricultural land. From the water quality point of view and as nonpoint source pollutants, nutrients are transported from the watershed by runoff, erosion and leaching. Soluble forms of nitrogen and phosphorus are transported in the runoff. Insoluble forms and forms adsorbed to the soil are moved by the sediments. Nitrate is the principal nutrient form leached to groundwater by percolation. The concentrations of nutrients and total loads depend on the amount of nutrient available for transport and on the conditions that affect the transport mechanisms.

Weather, soils, topography, and land uses all affect the transport capacity. Information about hydrology, erosion, and availability of nutrients should be considered as input data for any nutrient model. Previous calculations of runoff and erosion should be performed in the watershed in order to predict the nitrogen and phosphorus moving in runoff, with sediment, and by leaching.

Process descriptions have to be consistent with the hypothesis that a distributed approach based on land use response will improve the nonpoint source modelling predictions. As part of this research, simple relations (similar to those used in most of the reviewed models) to account for enrichment, solubility, adsorption and leaching (Frere *et al*., 1980 and Mills *et al*., 1985) are used. For example, to estimate the sediment transport of nutrients, the algorithms are based on a proportional factor that equates sediment loading to that of the contaminant. The potency factor is related to the concentration in the soil and the enrichment ratio for the contaminant. Calculations are limited to sediment only from overland erosion.

Pesticides.- Use of pesticides revolutionized agricultural production to the point that most agricultural practices formerly used to control weeds, insects, and disease shifted in favor of chemical control. To estimate the amount of pesticide that can be found in surface waters, properties of the applied chemical, amount applied, and the time of the application relative to the rainfall should be known.

Intensive research has been done on the mechanisms for decay of pesticides to define half life values, solubility, and partition coefficients (Wauchope and Leonard, 1980, Lyman, W.J., 1982, Nash, 1980, and Leonard, 1990). Application rates, efficiencies, and elapsed time between the storm event and application can be estimated from agricultural practices in the area of interest. Fortunately, application rates tend to fall within rather narrow ranges; "single applications of 1-5 kg/ha are typical for most herbicides and multiple applications totaling 10-20 kg/ha for insecticides" (Leonard and Knisel, 1989).

After initial application losses, such as canopy interception, volatilization and degradation, the remaining pesticide reaches the soil. The primary source of pesticide available to enter the runoff is from the surface layer of soil. Washoff applied to foliage is another source that may enter the runoff. Pesticide can dissipate from soil and foliar surfaces by degradation and volatilization.

During rainfall events, pesticides may move below the surface zone in the infiltrating water and across the surface in runoff. Pesticide can be extracted by water flowing over the surface, by dispersion and mixing of the soil material in the flow, and by raindrop impact. Once in the runoff, it can be either in solution or attached to the eroded sediment. Simple relations to describe these processes are used in the majority of the models (Knisel, 1980 and Young *et al.*, 1986) were used in the water quality component.

Some important relations that the model takes into account when dealing with the pesticide component are as follows:

(a) first order decay function to account for degradation based on the half-lives of the chemicals,

(b) Henry's law which describes the relation between vapor and solution phases to account for volatilization,

(c) isotherms for the sorption mechanism that controls the partitioning processes between the particulate and dissolved fractions based on a Freundlich equation and an octanol-water partition coefficient.

2.3 Chapter Summary

In this chapter a literature review has been presented. The problem of nonpoint source pollution was outlined and some of the existing models described in order to select the ones to be integrated in the decision support system. AGNPS and WATFLOOD with a water quality component to be incorporated into it were the selected models. A brief description of the GIS techniques used in the area of diffuse pollution and the decision support system RAISON were also described. The methods and algorithms to use in the water quality component for the WATFLOOD model were reviewed and a selection was done based on the physical processes of the yield sediment model. The nutrient relationships will be the same formulations as in the AGNPS that were adapted after the CREAMS model development.