# 2. LITERATURE REVIEW

This chapter emphasizes research that is focused on GIS applications for agricultural non-point source pollution modeling. Models that include nutrients or pesticides are presented. Soil erosion and sediment transport applications are also mentioned here for completeness of the discussion.

The application of the GIS in modeling non-point source pollution can be grouped into three categories:

1) Existing pollution models that are linked with GIS software;
2) Pollutant transport is modeled entirely within the GIS;
3) GIS is utilized to extract spatial data required for analysis of non-point source pollution;

## *2.1 Linking GIS with Water Quality Models*

Water quality models linked with GIS programs are the dominant approach in modeling non-point source pollution. The GIS provides the data required for the model, then the model is executed. After calculations, the GIS is used for visual analysis of results. The most popular pollution model that is linked with GIS software is AGNPS (Agricultural Non-Point-Source) developed by Agricultural Research Service (ARS). The model source code is available through the WEB site: http://www.infolink.morris.mn.us/ars/download.html.

***AGNPS***

AGNPS is an event-based distributed parameter model, that is, it computes flow and pollutant loadings for a single rainfall event. It calculates runoff from agricultural watershed and transport processes of sediment, nitrogen, phosphorous, and COD (Chemical Oxygen Demand). The watershed is represented by square cells of 0.4 - 16 ha (1 - 40 acres). Each cell is characterized by twenty-two parameters that include: SCS curve number, terrain description, channel parameters, soil-loss equation data, fertilization level, soil texture, channel and point source indicators, and an oxygen demand factor. Sediment runoff is estimated from the modified version of USLE (Universal Soil Loss Equation) and its routing is performed for five particle size classes. Calculations of nutrient transport are divided into soluble and sediment-absorbed phases. Due to the large amount of input data required, the application of AGNPS is limited to watersheds not larger than 200 $km^2$ (Young, et al., 1989; DeVries and Hromadka, 1993; Engel, et al., 1993). However, it has been applied to larger basins, by representing the study area by a grid of cells larger than 16 ha. For example, Morse, et al., (1994) applied AGNPS with 100 ha cells to estimate concentrations in a 1645 $km^2$ watershed.

At least three interfaces between AGNPS and GRASS (Geographical Resources Analysis Support System) have been constructed: (1) at Michigan State University (He, et al., 1993), (2) by Srinivasan and Engel (Engel, et al., 1993a; Engel, et al., 1993b; Engel 1996; Mitchell, et al., 1993a; Mitchell, et al., 1993b), and (3) by the Soil Conservation Service as a watershed planning tool in the Hydrologic Unit Water Quality Project (HUWQ) (Cronshey, et al., 1993; Geter, et al., 1995; Drungil, et al., 1995).

GRASS is the major public domain GIS that supports a raster data structure with data conversion from vector data. It performs the basic GIS functions of data input, storage, manipulation, analysis, and display (Drungil, et al., 1995). Access to the source code of GRASS provides the flexibility to modify existing GRASS procedures or to add new ones. GRASS has a considerable ability to support hydrologic analysis.

AGNPS has also been linked with Arc/Info GIS. Jankowski and Haddock (1993) coupled AGNPS with PC-Arc/Info, a vector based GIS. The interface was constructed using Arc/Info macro language (SML), Pascal language, and batch programming. Vieux and Needham (1993) studied the AGNPS model sensitivity to grid-cell size. They used Arc/Info to generate AGNPS input files and to display model output. They demonstrated that the variation of channel erosion, sediment yield, and delivery ratio due to the cell size selection may introduce unacceptable errors or erroneous conclusions when analyzing nonpoint pollution using AGNPS.

Morse, et al., (1994) integrated AGNPS, Arc/Info and Oracle, a database system to estimate the nitrogen, phosphorus and COD concentrations for different management scenarios in the Bedford-Ouse catchment, UK. They represented the 1645 $km^2$ watershed by square cells of 100 ha (1 km * 1 km). Another AGNPS-Arc/Info integrated system was constructed by Tim and Jolly (1994) to evaluate effectiveness of several alternative management strategies in reducing sediment pollution in a 417 ha watershed located in southern Iowa.

AGNPS has also been linked to other GIS programs, such as: ERDAS (Earth Resources Data Analysis System), a grid cell-based system (Evans and Miller, 1988), Geo/SQL, a vector-based GIS (Yoon, et al., 1993), and IDRISI, a raster based GIS (Klaghofer, et al., 1993). This last interface has been used to evaluate erosion and sediment yields in a lower alpine drainage basin of area of 65 ha (located in Austria). The interface contained EPIC (Erosion/Productivity Impact Calculator, Williams, et al., 1990) a field scale comprehensive model developed to predict the long-term relationship between erosion and productivity. EPIC's components include weather simulation, hydrology, erosion-sedimentation, nutrient cycling, plant growth, tillage, soil temperature, economics, and plant environment control.

### *SWRRB*

Cronshey, et al., (1993) describe an interface that combines GRASS and a watershed scale water quality model SWRRB (Simulator for Water Resources in Rural Basins). SWRRB (Arnold, et al., 1990) uses a daily time step for calculations of sediment yield, flow routing, as well as pesticide and nutrient fate and transport studies. Basins are subdivided to account for differences in soils, land use, crops, topography, and weather. The soil profile can be divided vertically into ten layers. Basins of several hundred square miles can be studied, but the number of sub-basins is limited to 10. The hydrology component is based on a daily water balance equation that includes rainfall, runoff, evapotranspiration, percolation, and return flow. Rainfall intensity hyetographs are calculated by using a modified rational method. The Soil Conservation Service curve number technique is used to estimate runoff volume. The evapotranspiration component requires such data as daily solar radiation, mean air temperature, crop cover and snow cover. Daily precipitation as well as air temperatures and solar radiation can be supplied as model input or they can be simulated by a SWRRB weather generator. Sediment yield is computed for each sub-basin by using the Modified Universal Soil Loss Equation. SWRRB is physically based and is intended to be used for situations in which calibration data are not available (DeVries and Hromadka, 1993; Donigian and Huber, 1991).

### *SWAT-QUAL2E*

SWAT (Soil and Water Assessment Tool) is an extended version of the SWWRB model ( Arnold, et al., 1993). It has been linked with GRASS (Srinivasan and Arnold, 1994; Engel, et al., 1993) and with Arc/Info (Bian, et al., 1996).

The major components of SWAT are similar to SWWRB modules, which include weather, hydrology, erosion, soil temperature, crop growth, nutrients, pesticides, subsurface flow, and agricultural management. The model operates on a daily time step and is capable of simulating 100 or more years. The restriction of only being able to simulate 10 subbasins, in the case of SWRRB, has been removed in SWAT. The watershed can be divided into a practically unlimited number of cells and/or subwatersheds. In addition, each subbasin can be discretized into virtual areas that have a unique soil and land use combination. The hydrologic response is generated in each virtual area. The output of the subbasin is calculated as a weighted average of the virtual area hydrologic responses. The new features of SWAT include: routing of the flow through the basin streams and reservoirs, simulating lateral flow, groundwater flow, stream routing transmission losses, and modeling sediment and chemical transport through ponds, reservoirs, and streams (Mamillapalli, 1996).

The SWAT-GRASS model has been applied for small scale modeling as well as for continental scale hydrologic modeling. For example, Jacobson, et al., (1995) evaluated the water quality impacts of the diverse crops and management practices in a 4.6 $km^2$ subwatershed of the Herrings Marsh Run Watershed in the North Carolina Coastal Plains. On the other scale extreme, Srinivasan, et al., (1995) applied the SWAT-GRASS interface and such data as a map of soils (STATSGO), map of land use (USGS LULC) and a DEM to estimate the following features for the entire U.S.:

- average annual rainfall;

- average annual total water yield;
- average annual actual evapotranspiration (plant ET was calculated as a function of leaf area, root depth, and irrigation);
- average annual Penman-Montieth potential evapotranspiration; and
- annual grain yield and biomass production.

The U.S. was divided into 78,863 STATSGO polygons for this analysis.

In 1996 the QUAL2E (Enhanced Stream Water Quality Model) water quality component was incorporated into SWAT to simulate instream dynamics. The first-order decay relationships for algae, dissolved oxygen, carbonaceous biochemical oxygen demand, organic nitrogen, ammonium nitrogen, nitrate nitrogen, nitrite nitrogen, organic phosphorus, and soluble phosphorus used in QUALE2E were adopted in SWAT with various adjustments (Ramanarayanan, et al., 1996).

QUAL2E model uses a finite-difference solution of the advective-dispersive mass transport, reaction, and sink/source equation. The stream network is divided into headwaters, reaches, and junctions. The changes in flow conditions are represented as a series of steady- flow water profiles. Such parameters as velocity, cross-sectional area, and water depth that are required for the mass transport calculations are computed from the flow rate. For each river reach, QUAL2E requires specification of as many as 26 physical, chemical, and biological parameters. (DeVries and Hromadka, 1993; Camara and Randal, 1984; Schoellhamer, 1988). Compiling such data at a regional scale would take a very great investment of time and resources.

### *ANSWERS*

Rewerts and Engel (1991; Engel, 1993; Engel, 1996) integrated the GRASS GIS with ANSWERS (Aerial Nonpoint Source Watershed Environment Response Simulation). ANSWERS (Beasley, et al., 1982) calculates runoff, erosion, sedimentation and phosphorus movement from watersheds. The watershed is divided into a grid of square cells. For each cell the following parameters are defined: slope, aspect, soil porosity, moisture content, field capacity, infiltration capacity, USLE erodibility factor, crop and management factors. A channel is described by width and roughness. Runoff, erosion, sedimentation, and water quality related to sediment associated chemicals (for example, dissolved and sediment-bound ammonium, sediment bound total Kjeldahl nitrogen, and dissolved and sediment-bound phosphorus) are computed for each cell and routed (Beasley and Huggins, 1991; Donigian and Huber, 1991).

Typical cell sizes range from 0.4 to 4 ha with smaller cells providing more accurate simulations. During rainfall events the time step is 60 seconds. If there is no precipitation, the model uses a daily time step (Wolfe, et al., 1995).

De Roo (1993) applied the ANSWERS-GENAMAP GIS-PC-RASTER interface to calculate surface runoff and soil erosion in the Yendacott catchment, UK (147 ha), the Etzenrade catchment (225 ha) and Catsop catchment (46 ha) in The Netherlands. Engel, et al., (1993b) compared the results of GRASS-ANSWERS model, with two other NPS models that were integrated with GIS: AGNPS and SWAT. The simulated results matched observed values reasonably well. Wolfe, et al., (1995) created a user interface that links Arc/Info GIS with ANSWERS. The system has been designed for evaluating the overall effectiveness of selected best management practices at the farm scale.

### *HSPF*

Al-Abed and Whiteley (1995) linked PC-Arc/Info GIS with the Hydrologic Simulation Program-Fortran (HSPF) to simulate the effects of changes in land use and in resource management strategies on the irrigation water quality in the Grand River, Ontario, with a drainage area of 6,965 $km^2$.

HSPF simulates both watershed hydrology and water quality (Johanson, et al., 1980). The rainfall is distributed into interception loss, surface runoff, interflow, and flow into the lower soil zone or groundwater storage. Soil is divided into three moisture zones: an upper soil zone, that influences the rapid runoff, a lower soil zone, and a groundwater storage zone. Some of the water from the groundwater storage becomes stream base flow (DeVries and Hromadka, 1993).

The water quality component simulates silt, clay, and sand sediment transport, including resuspension and settling processes. It can also calculate nutrient and pesticide concentrations. The nutrient processes include DO (dissolved oxygen), BOD (biological oxygen demand), nitrogen, phosphorus, pH, phytoplankton, zooplankton, and benthic algae. HSPF simulates such transfer and reaction processes as hydrolysis, oxidation, biodegradation, volatilization, sorption, and chemical exchange between benthic deposits and the water column. The program user must supply parameters for each of the modeled processes (Donigian and Huber, 1991; DeVries and Hromadka, 1993).

The watershed is divided into segments--parcels of land that are exposed to weather conditions described by one set of meteorological time series. Hydraulic routing requires division of the major streams into modeling segments (Al-Abed and Whiteley, 1995).

## 2.2 GIS models of Water Quality

Some GIS programs are equipped with a macro language that allows the user to write models within the application. For example, Arc/Info has very powerful macro language, AML (Arc/Info Macro Language). In addition, external procedures written in such programming languages as C/C++ or FORTRAN can be executed by macro, thus the modeling process can be very efficient. This section discuss models of water pollution built using GIS tools.

White and Hofschen (1993) developed a spatial model for assessing nutrient loads in New Jersey rivers using Arc/Info. They used 3 arc-sec digital elevation models (DEM) to partition the study area (15,385 $km^2$) into 2,893 drainage basins (polygons) with a network of 10,916 stream segments (arcs). The time of travel was assumed as the basis for calculating predictors of water quality. A simple formula $v = 0.38 * Q^{0.24}$, which was estimated for New Jersey, was used to estimate the flow velocity in each reach. A first-order decay reaction was assumed to calculate the non-conservative downstream transport. White and Hofschen attempted to improve the model by representing the decay constant as a function of stream slope, and the nonpoint source yields as a function of subbasin gradient, but the model performance showed no improvement with these refinements. White and Hofschen found that the time of travel, which was calculated from the exponential velocity formula, underestimated by a factor of 0.57 the time of travel of dye-tracer, that is, the dye took approximately twice as long to traverse the stream as the formula suggested. This travel time underestimation was accommodated by assignment of higher values of pollutant decay than those reported in the literature.

Smith, et al., (1993) constructed a GIS model of total phosphorus concentrations in New Jersey streams. The core of this model is a regression equation that relates transformed (natural logarithm) total phosphorus concentration measured at a given point to transformed concentrations resulting from exponentially decayed phosphorous loads in the upstream watershed. In this study, the classical approach of modeling first-order reaction was modified. Instead of using the time of travel and time decay coefficient, the travel distance and a distance decay coefficient for phosphorus were applied in the model respectively. The data from 104 long term sampling stations, collected in the period from 1982 to 1987 were utilized to estimate regression coefficients. The area of the studied region was 15,401 $km^2$. The sources of phosphorous were represented by such variables as area of agricultural land, total human population, and total municipal effluent flow.

Zollweg, et al. (1995) constructed a GRASS model of the phosphorus transport for the 25.7 ha Brown Watershed, an upland agricultural watershed in Pennsylvania. The GRASS script language was used to describe the physical processes that originally were modeled by the Soil Moisture-based Runoff Model (SMoRMod). SMoRMod is a distributed spatially variable model. Such parameters as climatic variables, topography, land use, and soils distribution constitute the input. The watershed is divided into rectangular cells. For each cell, the infiltration, soil moisture, groundwater flow, and surface runoff are estimated. The surface runoff is translated through the channel system to the watershed outlet where the storm runoff hydrograph is calculated. The phosphorous module determines the P content of the storm runoff generated over the landscape and transport of this P to the watershed outlet (Zollweg, et al., 1995).

Hession and Shanholtz (1988) incorporated the Universal Soil Loss Equation (USLE) with delivery ratio into the Virginia Geographic Information System

(VirGIS). The model was used to estimate the potential sediment loading in Virginia's Chesapeake Bay drainage area.

## 2.3 GIS as a tool for spatial data extraction

The most basic application of the GIS is spatial data manipulation, data extraction for further analysis, and presentation of results in map form. This section discusses work in which the GIS tools have been utilized to support statistical analysis of surface water pollution.

Cressie and Majure (1994) used Arc/Info to determine explanatory variables for a statistical model of the variation in pollutant concentration from dairies in streams of the Upper North Bosque watershed located principally in Erath County, Texas. The Arc/Info GRID and Digital Elevation Models (DEM) were used to determine drainage basins and the lengths along flow paths. Cressie and Majure assumed a spatially constant flow velocity (0.5 m/s), and using simple map algebra, they determined a 3-day flow-time area of influence for each stream measurement site. Seventeen explanatory variables including a number of dairies per acre, a number of animals per acre, lagoons per acre, waste application method, soil hydrologic code, average slope, distance to basin outlet, and precipitation were considered. All variables, except one (seasonal variation), were determined using the GIS. The authors concluded that the GIS was an important tool in observational studies due to its ability to construct explanatory variables at the appropriate scale.

Mueller, et al., (1993) applied logistic regression to relate discrete categories of nitrate concentrations to such explanatory variables as land use in the drainage basins upstream from the sampling sites, percentile of daily streamflow at the time of sampling, acreage of the basin in corn, acreage in soybeans, density of cattle, and population density.

Logistic regression is used when the independent variable is discrete or categorical rather than continuous. It has the following form: $\log[p/(1-p)] = a + b_k X$, where $p$ is the probability of data value being in one of the possible categories, $a$ is the intercept, $b_k$ are $k$ regression coefficients, and X is the vector of explanatory variables. The percentiles ($P_j$) are computed using equation $P_j = A_{(n+1)*j}$, where: A is a data set ordered from smallest to largest ($A_i$, $i = 1, \ldots n$), $n$ is the sample size of A, and $j$ is the fraction of data less than or equal to the percentile value, e.g. for 25th percentile $j = 0.25$ (Helsel and Hirsh, 1995).

Mueller, et al., (1993) extracted data from GIS databases stored in 1:2,000,000- scale maps of the conterminous United States. The GIS software was used only to areally weight the extracted data and sum it by basin; their model did not include stream transport. Better classification of nitrate concentration was achieved by a model that included the flow percentile, the areal extent of corn and soybean production, the density of cattle, and the density of population, as compared to the model that contained percentile of flow, nitrogen fertilizer application, and population density. In addition, Mueller, et al., found that as the percentile of flow increased, the probability of nitrate concentration being in a higher category also increased. The logistic regression analysis results led these researchers to a conclusion that the level of nitrate contamination in midwestern streams is most strongly related to streamflow and to several characteristics of the upstream basin, including the areal extent of corn and soybean production, the density of cattle, and the population density.

From the observations made on the Mississippi River and four tributaries during a one - year period (from April 1, 1991 to March 31, 1992), Battaglin et al. (1993) estimated a single relationship between the annual use of nitrogen and nitrate transport: $N_{transport} = -0.2 + 0.1547\ N_{use}$ and a linear relationship between annual atrazine use and the atrazine transport: $A_{transport}$ = -12 * + 0.0156 * $A_{use}$, (in metric tons). They used a GIS to estimate the nitrogen and atrazine use within gauged watershed from the county level sales of nitrogen fertilizer and atrazine herbicide. In addition, Battaglin et al. estimated that 321 Mg (ton) of atrazine and 33.7 Mg (ton) of alachlor were discharged from the Mississippi River basin to the Gulf of Mexico in streamflow (from April 1, 1991 to March 31, 1992), while the amounts of these herbicides applied in the basin were approximately equal. This suggests that atrazine is much more persistent agrichemical than is alachlor.

Moody and Goolsby (1993) report the results of a large scale USGS study of herbicide transport in the Lower Mississippi River. Although they did not use a GIS, this work is mentioned here since it is one of the few large scale sampling studies available on the Mississippi River. During May 26-29, 1990, water samples for triazine herbicide analysis were collected every 16 km from Baton Rouge in Louisiana, upriver to the Mississippi-Ohio River confluence (distance of 1900 km). The measurements showed the background level of ~2.7 μg/L of triazine herbicides and an upriver concentration gradient of 0.2 μg/L per 100 km (concentration decreased going downstream). The authors suggest that the longitudinal spatial variability in concentration is a result of cross-channel gradients and the addition of 'slugs' of water from various upriver tributaries. A routing scheme was used to predict the location of water masses. This routing method was tested by using the measurements of the specific conductance. The average flow velocity was v = 6 km/h that gives about 13-day residence time of water in the Mississippi River over the distance sampled in this study. It is interesting that the measurements show about 50% decrease in the load whereas the reported atrazine half-life in water is about 140 days (Thurman, et al., 1992).

## 2.4 Comparison of the proposed method with previous studies

### 2.4.1 Time domain

The model presented in this dissertation is designed to represent average monthly values of flow rate, agrichemical concentration, and chemical load in streams of the Midwest. Introduction of a seasonal component to the model fills a gap in existing GIS models of pollutant transport. Most of the hybrid models that have been introduced in Section 2.1, are capable of performing continuous-time simulations (SWAT-GIS), event - related calculations (AGNPS), or daily computations (SWRRB-, ANSWERS-GIS). The models of surface water pollution that are constructed within GIS (discussed in Section 2.2) estimate annual average chemical concentrations for a study period. Even the USGS studies do not evaluate changes in chemical concentration in Midwest surface waters on a monthly basis. Seasonal variations are represented by usually three terms of the year: pre-planting, post-planting, and Fall low flow (Goolsby, et al., 1993a; Goolsby and Battaglin, 1993; Scribner, et al., 1993). The USGS model of agricultural chemical transport in the Midwest rivers relates the annual chemical load with annual agrichemical use (Battaglin, et al., 1993). The statistical model presented here explains seasonal variations of concentration on a monthly basis.

### 2.4.2 Spatial domain

The model is designed to predict the loads and concentrations in such large basins as the Upper Mississippi River basin ( drainage area 490,000 $km^2$), the Ohio

River basin (526,000 $km^2$), or the Upper Missouri-Mississippi-Ohio River basin (above Ohio-Mississippi River junction, about 2,400,000 $km^2$ of drainage area). A more detailed version of the model has been applied for evaluation of nitrate and atrazine concentrations in the Iowa-Cedar River basin, Iowa of 32,000 $km^2$ - an area larger than the reported limits of application of hybrid models discussed in Section 2.1 and models constructed within GIS, presented in Section 2.2.

Krysanova, et al., (1996), specifies limitations of selected pollution models: for example, AGNPS and ANSWERS are limited to watersheds of about 200 $km^2$, SWRRB was developed for agricultural basins as large as 600-800 $km^2$, and SWAT is intended to be applied in watersheds up to 25,000 $km^2$. Besides the present study the only models known to the author that can be applied for a such large area as the Midwest region are the annual agrichemical load functions presented in Section 2.3, which were estimated by Battaglin (1993).

### 2.4.3 Model formulation

The proposed model can be classified in the second category discussed previously in Section 2.2, namely models constructed within GIS. Because concentrations and loads estimated by the model are spatially and temporally distributed, the model can be characterized as a distributed system. On the other hand, the agrichemical concentration or load is calculated at a given location by applying average values that characterize the total upstream drainage area in a regression equation. Thus the model can also be considered as a lumped system.

The major differences between existing GIS models and the one presented here result from the spatial extent for which the model has been developed--the Upper Missouri-Mississippi River and the Ohio River basins. It utilizes data available for the whole US, i.e., digital elevation data and the agrichemical application rates, thus the model contains a limited number of parameters, not including either time-decay nor length-decay coefficients.