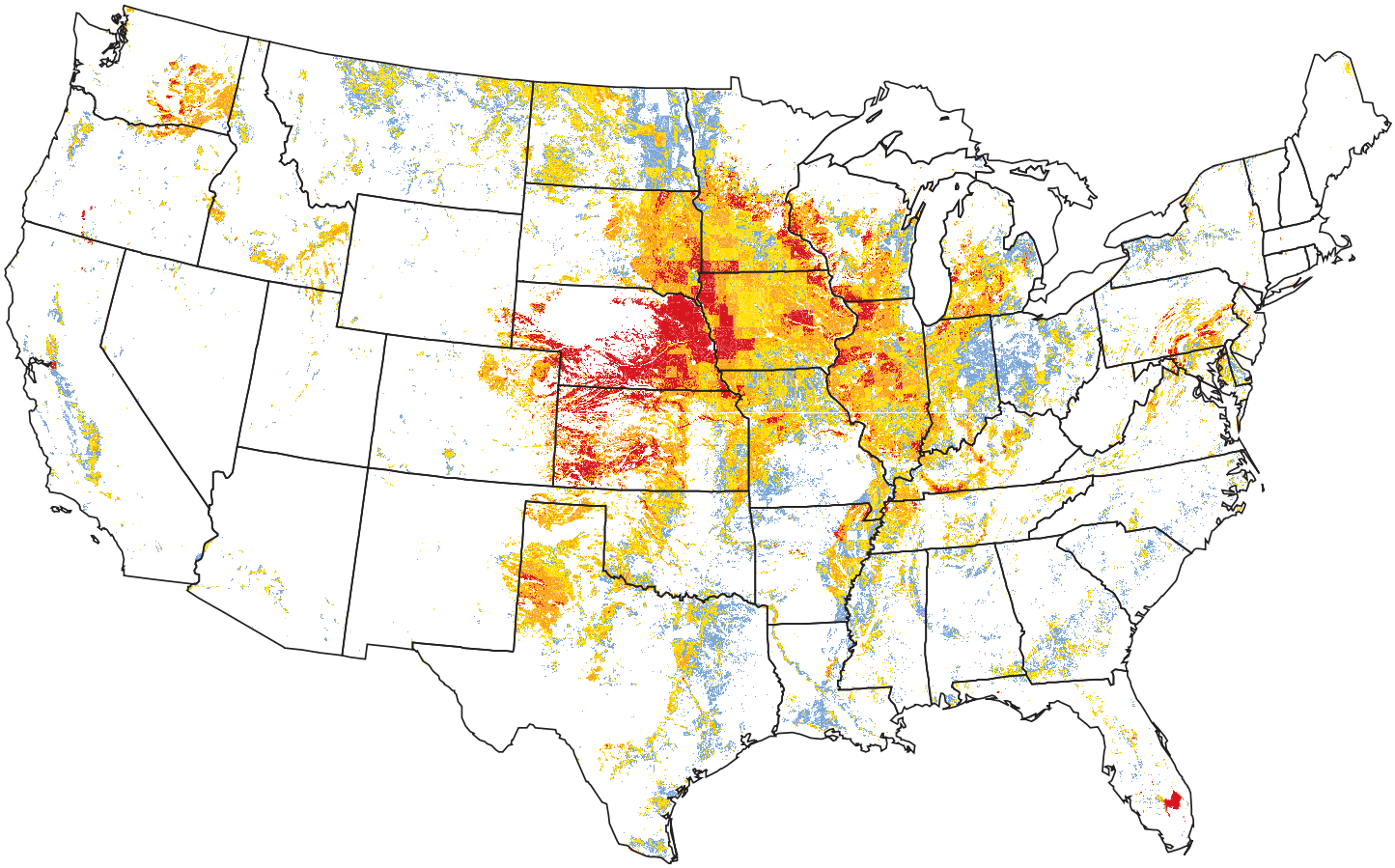
USGS
science for a changing world

National Water-Quality Assessment Program

Development and Application of a Regression Equation for Estimating the Occurrence of Atrazine in Shallow Ground Water beneath Agricultural Areas of the United States



Scientific Investigations Report 2005-5287

U.S. Department of the Interior
U.S. Geological Survey

Cover. Map shows the estimated frequency of atrazine occurrence in shallow ground water underlying agricultural areas throughout the conterminous United States.

Development and Application of a Regression Equation for Estimating the Occurrence of Atrazine in Shallow Ground Water beneath Agricultural Areas of the United States

By Paul E. Stackelberg, Robert J. Gilliom, David M. Wolock, and Kerie J. Hitt

National Water-Quality Assessment Program

Scientific Investigations Report 2005–5287

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
Gale A. Norton, Secretary

U.S. Geological Survey
P. Patrick Leahy, Acting Director

U.S. Geological Survey, Reston, Virginia: 2006

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment:
World Wide Web: http://www.usgs.gov
Telephone: 1-888-ASK-USGS

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this report is in the public domain, permission must be secured from the individual copyright owners to reproduce any copyrighted materials contained within this report.

Suggested citation:
Stackelberg, P.E., Gilliom, R.J., Wolock, D.M., and Hitt, K.J., 2006, Development and application of a regression equation for estimating the occurrence of atrazine in shallow ground water beneath agricultural areas of the United States: U.S. Geological Survey Scientific Investigations Report 2005-5287, 12 p., online only.

Foreword

The U.S. Geological Survey (USGS) is committed to serve the Nation with accurate and timely scientific information that helps enhance and protect the overall quality of life, and facilitates effective management of water, biological, energy, and mineral resources (http://www.usgs.gov/). Information on the quality of the Nation's water resources is of critical interest to the USGS because it is so integrally linked to the long-term availability of water that is clean and safe for drinking and recreation and that is suitable for industry, irrigation, and habitat for fish and wildlife. Escalating population growth and increasing demands for the multiple water uses make water availability, now measured in terms of quantity and quality, even more critical to the long-term sustainability of our communities and ecosystems.

The USGS implemented the National Water-Quality Assessment (NAWQA) Program to support national, regional, and local information needs and decisions related to water-quality management and policy. (http://water.usgs.gov/nawqa/). Shaped by and coordinated with ongoing efforts of other Federal, State, and local agencies, the NAWQA Program is designed to answer: What is the condition of our Nation's streams and ground water? How are the conditions changing over time? How do natural features and human activities affect the quality of streams and ground water, and where are those effects most pronounced? By combining information on water chemistry, physical characteristics, stream habitat, and aquatic life, the NAWQA Program aims to provide science-based insights for current and emerging water issues and priorities. NAWQA results can contribute to informed decisions that result in practical and effective water-resource management and strategies that protect and restore water quality.

Since 1991, the NAWQA Program has implemented interdisciplinary assessments in more than 50 of the Nation's most important river basins and aquifers, referred to as Study Units (http://water.usgs.gov/nawqa/nawqamap.html). Collectively, these Study Units account for more than 60 percent of the overall water use and population served by public water supply, and are representative of the Nation's major hydrologic landscapes, priority ecological resources, and agricultural, urban, and natural sources of contamination.

Each assessment is guided by a nationally consistent study design and methods of sampling and analysis. The assessments thereby build local knowledge about water-quality issues and trends in a particular stream or aquifer while providing an understanding of how and why water quality varies regionally and nationally. The consistent, multi-scale approach helps to determine if certain types of water-quality issues are isolated or pervasive, and allows direct comparisons of how human activities and natural processes affect water quality and ecological health in the Nation's diverse geographic and environmental settings. Comprehensive assessments on pesticides, nutrients, volatile organic compounds, trace metals, and aquatic ecology are developed at the national scale through comparative analysis of the Study-Unit findings (http://water.usgs.gov/nawqa/natsyn.html).

The USGS places high value on the communication and dissemination of credible, timely, and relevant science so that the most recent and available knowledge about water resources can be applied in management and policy decisions. We hope this NAWQA publication will provide you the needed insights and information to meet your needs, and thereby foster increased awareness and involvement in the protection and restoration of our Nation's waters.

The NAWQA Program recognizes that a national assessment by a single program cannot address all water-resource issues of interest. External coordination at all levels is critical for a fully integrated understanding of watersheds and for cost-effective management, regulation, and conservation of our Nation's water resources. The Program, therefore, depends extensively on the advice, cooperation, and information from other Federal, State, interstate, Tribal, and local agencies, non-government organizations, industry, academia, and other stakeholder groups. The assistance and suggestions of all are greatly appreciated.

Robert M. Hirsch
Associate Director for Water

Contents

Foreword iii
Abstract 1
Introduction 1
Methods 2
    Study Areas and Well Networks 2
    Sampling and Analytical Methods 3
    Statistical Methods 3
    Ancillary Data 4
Estimating Atrazine Occurrence in Shallow Ground Water 6
    Regression Equation 6
        Residuals Analysis 8
        Equation Performance 9
    Estimation of Atrazine Occurrence 9
Summary 11
Acknowledgments 11
References Cited 11

Figures

1. Generalized locations of the 52 agricultural ground-water studies of the NAWQA (National Water-Quality Assessment) Program, 1991–2001 2
2. Distribution of (A) studentized residuals, (B) studentized residuals and estimated atrazine detection frequencies, and (C) observed and estimated atrazine detection frequencies 9
3. (A) Estimated 1992 agricultural use of atrazine throughout the conterminous United States, and (B) estimated frequency of atrazine occurrence in shallow ground water underlying agricultural areas throughout the conterminous United States 10

Tables

1. Explanatory variables used for equation development 5
2. Statistics for ordinary least-squares regression equations 7

Conversion Factors (SI to Inch/Pound)

Multiply	By	To obtain
	Length	
foot (ft)	0.3048	meter (m)
	Area	
square kilometer (km^2)	247.1	acre
square kilometer (km^2)	0.3861	square mile (mi^2)
	Volume	
cubic centimeter (cm^3)	0.06102	cubic inch (in^3)
	Flow rate	
million gallons per day (Mgal/d)	0.04381	cubic meter per second (m^3/s)
	Mass	
gram (g)	0.03527	ounce, avoirdupois (oz)
kilogram (kg)	2.205	pound avoirdupois (lb)
	Density	
gram per cubic centimeter (g/cm^3)	62.4220	pound per cubic foot (lb/ft^3)

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

$$°F = (1.8 \times °C) + 32$$

Concentrations of chemical constituents in water are given either in milligrams per liter (mg/L) or micrograms per liter (µg/L).

Development and Application of a Regression Equation for Estimating the Occurrence of Atrazine in Shallow Ground Water beneath Agricultural Areas of the United States

By Paul E. Stackelberg, Robert J. Gilliom, David M. Wolock, and Kerie J. Hitt

Abstract

Results from 52 ground-water studies throughout the United States were used to examine relations between the occurrence of atrazine in shallow ground water in agricultural settings and explanatory variables that describe the natural setting, agricultural-management practices, and the type and amount of development in each area. The explanatory variables that were found to be correlated with atrazine occurrence were soil-infiltration rates, presence of artificial drainage (tile drains or trenches), available water-holding capacity of soils, soil permeability, amount of study area using ground water for irrigation source (as percentage of total area), amount of agricultural land (as percentage of total area), and intensity of atrazine use. Ordinary least-squares regression equations that used one or more of these explanatory variables describe as much as 58 percent of the variation in atrazine-detection frequencies. Application of a multivariate equation to unmonitored agricultural areas across the conterminous United States illustrates that atrazine use alone is insufficient for estimating the occurrence of atrazine in shallow ground water. Instead, areas in which soil characteristics and agricultural-management practices favor the movement of water from land surface to the water table and that also have intensive atrazine use are the most vulnerable to atrazine contamination.

Introduction

In 2001, the U.S. Geological Survey (USGS) completed the first decade of water-quality studies as part of its National Water-Quality Assessment (NAWQA) Program to describe the status of, and trends in, the quality of the Nation's ground-water and surface-water resources and to relate observed trends to natural and human-related factors that affect water quality (Hamilton and others, 2004; Gilliom and others, 1995). Water-quality investigations were conducted in 51 major river basins and aquifers throughout the Nation. The focus of the first decade (1991–2001) of ground-water studies in the NAWQA program was on water quality in major aquifers, with an emphasis on "recently recharged" (shallow) ground water in areas with consistent hydrogeologic and land-use settings (Gilliom and others, 1995). Samples of shallow ground water were collected from newly installed monitoring wells (usually 20 to 30) that were randomly located within each study area. Occasionally, preexisting wells that met study-design requirements were included in these ground-water networks. The guidelines and procedures for study design, sample collection, and ancillary data compilation were consistent among all study areas to ensure that the water-quality data would be comparable at the regional and national scale (Gilliom and others, 1995; Lapham and others, 1995; Koterba and others, 1995; Koterba, 1998).

Pesticide data from the first phase of the program (1991–95) were analyzed by Kolpin and others (1998; 2000) and Barbash and others (1999) to provide an initial summary of the concentrations and frequencies of detection of pesticide compounds in shallow ground water throughout the nation and to examine relations between pesticide occurrence and variations in the use and physical and chemical properties of each pesticide. In an effort to improve the understanding of why pesticides are found in shallow ground water on a national scale, Kolpin and others (2002) evaluated a set of factors that are most likely to affect the fate and transport of two herbicides (atrazine and metolachlor) in the subsurface. Results of their multivariate analysis explained 45 percent of the variation in atrazine-detection frequencies among the study areas examined.

The study described herein augments these previous studies by including data from the entire first decade of NAWQA ground-water investigations (1991–2001). Specifically, it builds upon the investigation by Kolpin and others (2002) by examining relations between the frequency with which atrazine was detected in each of 52 agricultural areas and a larger set of explanatory variables that describe the natural setting and the types and amounts of development within each study area. The enlarged set of explanatory variables was intended to provide additional information on the factors that control the fate and transport of atrazine to ground water. The purpose of this report is to document the development and application of a regression equation for estimating the occurrence of atrazine in shallow ground water beneath agricultural areas of the United States.

Methods

The consistent manner in which NAWQA studies are designed and samples are collected and analyzed allows direct comparison of results from studies collected in differing settings across the Nation. The following sections briefly describe study designs, methods of sample collection, methods of analysis, and the statistical techniques used for data analysis herein.

Study Areas and Well Networks

Atrazine data were collected from 1,294 wells within the 52 agricultural study areas during the first decade of NAWQA investigations (1991–2001); locations of the 52 study areas are shown in figure 1. Atrazine is used primarily as an agricultural herbicide; therefore, data from nonagricultural areas were excluded from this analysis. Atrazine-detection frequency reported from each study area was based on a single sampling of a set of randomly located wells in agricultural areas within a specifically defined geographic area and hydrogeologic setting. All wells tapped "recently recharged" (shallow) ground water and were installed or selected in accordance with protocols described in Lapham and others (1995) and Scott (1990). The median depth of wells was 29 ft and the median depth below the water table to the bottom of the well was 16 ft. Study areas with fewer than 10 well sites were excluded from this analysis to avoid calculating atrazine-detection frequencies from fewer than 10 samples. Also excluded were data from (1) three studies with domestic wells that were nested in the same area as shallow monitoring wells because the domestic wells were finished deeper and had longer screen intervals than the monitoring wells, and (2) two studies in areas of rice cultivation because the irrigation (flooding) practices altered the local hydrology and, thus, compromised the collection of ancillary data describing the natural setting (such as recharge rates).

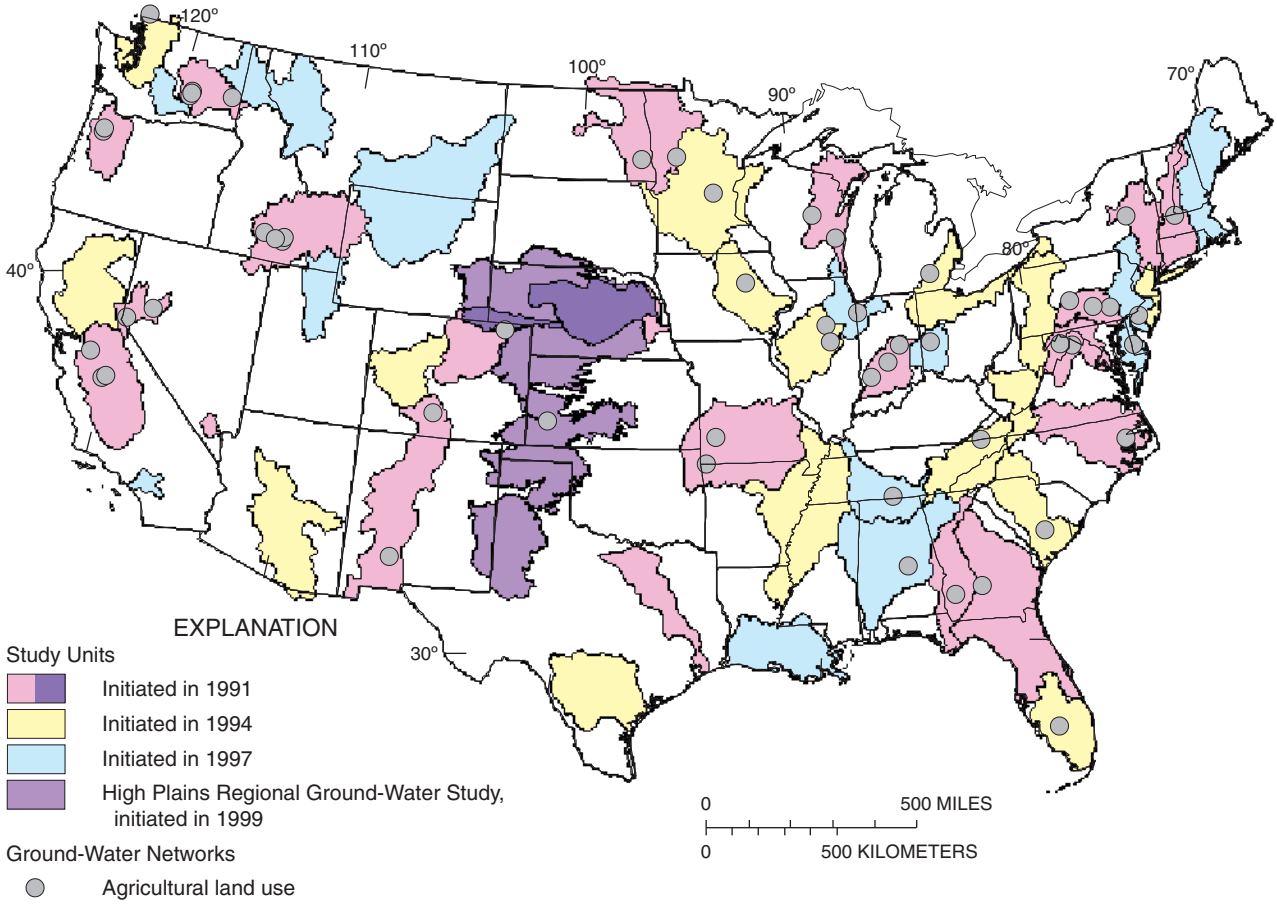


Figure 1. Generalized locations of the 52 agricultural ground-water studies of the NAWQA (National Water-Quality Assessment) Program, 1991–2001.

Sampling and Analytical Methods

Water samples were collected in accordance with procedures and protocols described in Koterba and others (1995). All samples were analyzed at the USGS National Water-Quality Laboratory in Denver, Colo., by gas chromatography/mass spectrometry with selected ion monitoring (Zaugg and others, 1995). The detection level for atrazine by this method is 0.001 µg/L; confirmed detections below this level were provided as described in Oblinger Childress and others (1999). The uncertainty associated with concentrations below the detection level is greater than for concentrations above that level; therefore, concentrations below the detection level were reported as estimates. The detection frequencies reported herein include these estimated values; however, only 1.5 percent of all atrazine detections were reported at or below the detection level (0.001 µg/L).

Statistical Methods

Multivariate ordinary least-squares (OLS) regression was used to explore relations between the frequency (percentage of samples) at which atrazine was detected in shallow ground water from each of the 52 agricultural study areas and selected explanatory variables that characterize aspects of the natural setting, agricultural-management practices, and the type and degree of development within each setting. (Explanatory variables are described in greater detail in the next section.) OLS regression theory and application are discussed in many texts (such as Draper and Smith, 1966; Helsel and Hirsch, 1992; and Ott and Longnecker, 2001). A combination of automated and manual best-subset and stepwise regression methods were used to examine potential equation formulations. Automated best-subset regression methods fit every possible regression equation involving p variables where $p = 1, 2, 3 \ldots n$ and n equals the number of explanatory variables. The best-subset method was used to identify optimum configurations of explanatory variables according to the value of each equations R^2, Mallow's C_p, and prediction error sum of squares (PRESS) statistics. The percentage of variation in atrazine-detection frequency that is explained by variation in the explanatory variables is described by R^2 and will always increase with the addition of new explanatory variables; therefore, Mallow's C_p statistic was also evaluated because this statistic balances the desire to explain as much of the variation in detection frequency as possible with the practicality of including as few variables as necessary. The PRESS statistic provides a measure of how reliably the equation predicts new values by iteratively regenerating the equation with n-1 observations, then estimating the value of the missing observation. Equations with low PRESS values produce less error in estimating missing observations than equations with higher PRESS values. Equations with high values of R^2 and low values of Mallow's C_p and PRESS represent optimal fits.

Automated stepwise regression methods allow the user to arbitrarily set significance levels for entering and removing explanatory variables. The method entails selecting the most significant explanatory variable, then iteratively adding or removing variables based on the significance of variables within and outside the equation at each step, until no more variables meet the criteria for inclusion or removal. Stepwise methods, although useful for exploring equations, have the disadvantage of not testing all possible formulations (Helsel and Hirsch, 1992).

A manual stepwise regression method was also used for equation development in accordance with procedures outlined in Larson and others (2004). With manual stepwise regression, a univariate relation was developed between the frequency of atrazine detection and the rate of atrazine application in each study area. Residuals from this relation were then correlated with the remaining explanatory variables, and the variable with the highest Pearson Product-Moment correlation coefficient added to the equation. This process was continued until improvement in the equations R^2 became negligible (2 percent or less). All equations obtained by the automated and manual procedures were confirmed to provide regression coefficients (slope and intercept) that differ significantly from 0 at a 95 percent confidence level ($\alpha = 0.05$).

Explanatory variables for equations identified through best-subset and stepwise procedures were subjectively evaluated for physical meaning and to ensure that their direction of influence was appropriate with regard to atrazine-detection frequency. Multicollinearity among explanatory variables was evaluated through the variance inflation factor (VIF). Variables associated with VIF values greater than $1/(1-R^2)$ are more closely related to other explanatory variables than to the dependent variable (SAS Institute Inc., 1991). VIF values for predictors in equations presented herein are all less than $1/(1-R^2)$.

Studentized residuals (Helsel and Hirsch, 1992) for equations selected from the best-subset and stepwise procedures were evaluated for normality and equal variation across the range of estimated values. Partial-residual plots were developed to evaluate the effect of individual cases on the estimation of a predictor's coefficient and to ensure a linear relation between response-variable and explanatory-variable residuals.

OLS regression equations with alternative response variables were considered during equation development. For example, an attempt was made to capture additional information on the transport of atrazine from land surface to the water table by developing equations from the frequency with which either atrazine or its primary degradate, deethylatrazine (DEA), was detected. Atrazine and DEA were detected together in 77 percent of samples in which either compound was detected, however, and DEA alone was found in only about 5 percent of samples in which only one of these compounds was detected; thus, the inclusion of DEA had little effect on detection-frequency calculations and did not improve the overall fit of exploratory regression equations. Additional exploratory equations were developed from the

95th percentile of the concentration of atrazine, or atrazine + DEA, for each study area as the response variable. Similar explanatory variables were selected by stepwise procedures for concentration-based equations and for the detection-frequency-based equations, but explained less of the variation in concentrations (lower R^2) because measured concentrations vary less widely among the study areas than do overall detection frequencies. Future work may include exploration at the site-specific and study-area scales of the relations between atrazine occurrence and variables that describe aspects of the natural setting, agricultural-management practices, and the type and degree of development.

Ancillary Data

A wide variety of ancillary data supports the NAWQA objectives of assessing current and future water-quality conditions and evaluating trends in water quality (Koterba, 1998). The present study compiled and evaluated more than 160 variables that describe aspects of the natural setting, agricultural-management practices, and the type and degree of development in each study area for consideration in equation development. Site-specific explanatory variables (such as dissolved oxygen concentration) that cannot be measured at unmonitored sites were not considered, but future work may incorporate such site-specific explanatory variables. Explanatory variables were transformed as needed to accommodate regression assumptions of linearity and to ensure that residuals were normally distributed with constant variance. The log transformation ln(x+1) was used to accommodate explanatory variables with values of 0. Other transformations of explanatory variables included the square, third, and fourth root. Criteria for elimination of explanatory variables were (1) they were poorly correlated (p-value ≥ 0.15) with atrazine-detection frequency, (2) the correlation was driven by outlying values, (3) the variable was collinear with other explanatory variables, or (4) the variable was sparsely populated. More than 160 explanatory variables were considered but only 33 were selected for equation development. These variables are listed in table 1 and described below.

Atrazine-use intensity: The amount of atrazine applied to agricultural lands in each study area annually (kg/km^2) was estimated from the (1) sum of county-level estimates of atrazine applications to agricultural crops, and (2) spatial distribution of agricultural land within each study area (Larson and others, 2004). Atrazine use on row crops, small grains, and fallow lands ("cgfuse" in table 1) and on row crops, small grains, fallow lands, pastures and hay, and orchards and vineyards ("alluse" in table 1) in each study area was estimated by (1) calculating areal weights for specific types of land in each county intersected by the study area, (2) multiplying the weighted value by atrazine use in the county on each type of agricultural land, and (3) summing the apportioned county estimates (Larson and others, 2004). These county estimates represent only agricultural applications because data on nonagricultural applications of pesticides are not currently (2005) available. The omission of nonagricultural applications is assumed to have only a minimal effect on the development of equations presented herein, however, because atrazine is used primarily as an agricultural herbicide. Variations in the size of the study areas, and in the amounts of agricultural land within each study area, required that the estimates of the amount of atrazine used in each study area be normalized to the total area of the targeted agricultural setting. This normalized atrazine-use variable is referred to herein as atrazine-use intensity (table 1).

Land use and population: Land-use data were compiled from the 1992 National Land Cover Data set (Vogelmann and others, 2001), which was enhanced with GIRAS land use/land cover data modified as described in Price and others (2003) to more accurately represent alpine tundra, orchards, vineyards, and residential areas. The following five land-use categories were combined for the agricultural-land variable ("ag" in table 1): row crops, small grains, fallow land, pastures and hay, and orchards and vineyards. Population-density data ("pden" in table 1) were compiled from block-group population counts from the 1990 census, and are expressed as the number of individuals per square kilometer (U.S. Bureau of the Census, 1991).

Climate: Mean annual temperature values ("temp" in table 1) for the period 1980–97 are from Thorton (2003).

Agricultural-management practices: Data on agricultural-management practices were obtained from the 1992 National Resources Inventory of the Natural Resources Conservation Service (U.S. Natural Resources Conservation Service, 1995) and include the percentage of each study area (1) that is irrigated ("irrig" in table 1), (2) utilizing ground water for source of irrigation water ("is01" in table 1), (3) under conservation tillage ("cp329" in table 1), and (4) with tile drains or trenches ("artdrain" in table 1). The variable "artdrain" was compiled using methods described in Schwarz and Alexander (1995). The ratio of the percentage of each study area utilizing ground water versus surface water as the source of irrigation water was calculated ("gwswirr" in table 1) and an indicator variable was included to distinguish study areas with less than, or more than, 25 percent total area in irrigation ("irriggrp" in table 1).

Soil and aquifer characteristics: Average soil characteristics were compiled from the 1994 U.S. Department of Agriculture's State Soil Geographic database (STATSGO) (U.S. Natural Resources Conservation Service, 1994) through methods described in Wolock (1997) and include the average (1) available water-holding capacity; (2) bulk density; (3) percentage organic-matter content; and (4) percentage clay, silt, and sand of soils ("awc", "bd", "om", "clay", "silt", and "sand" in table 1, respectively), the percentage of each study area with soils classified in hydrologic group D (very slow infiltration rates) or as hydric ("hgd" and "hydric" in table 1, respectively), the vertical permeability and permeability of the least permeable soils layer ("permv" and "permin" in table 1, respectively), erodibility of the uppermost soil layer ("kfactu"

Table 1. Explanatory variables used for equation development.

Variable name	Description
	Intensity of atrazine use
cgfuse92, cgfuse97	Sum of use (1992 or 1997) on row crops, small grains, and fallow land (kg)/area of row crops, small grains, and fallow land (km^2)
alluse92, alluse97	Sum of use (1992 or 1997) on row crops, small grains, fallow land, pastures and hay, and orchards and vineyards (kg)/area of row crops, small grains, fallow land, pastures and hay, and orchards and vineyards (km^2)
	Land use and population
ag	Percentage of area in agricultural land use [sum of NLCDE categories row crops, small grains, fallow land, pastures and hay, and orchards and vineyards (km^2)/study area (km^2)] x 100
pden	Population density from 1990 census (people/km^2)
	Climate
temp	Mean annual air temperature (1980–97) (ºC)
	Agricultural management practices
is01	Percentage of area using ground water for irrigation source [1992 NRI irrigation source is01 (km^2)/study area (km^2)] x 100
cp329	Percentage of area under conservation tillage [1992 NRI conservation practice cp329 (km^2)/study area (km^2)] x 100
artdrain	Percentage of area with artificial drainage [aggregate of 1992 NRI conservation practices cp606, cp607, cp608 (km^2)/study area (km^2)] x 100
irrig	Percentage of area that is irrigated [aggregate of 1992 NRI irrigation types it01, it02, it03 (km^2)/study area (km^2)] x 100
irriggrp	Indicator variable (0 = less than 25 percent of the study area irrigated; 1 = over 25 percent of the study area irrigated)
gwswirr	Ratio of ground water to surface water as irrigation source [1992 NRI irrigation source is01 (km^2)/study area (km^2)] x 100/ [aggregate of 1992 NRI irrigation sources is02, is03, is04 (km^2)/target area (km2)] x 100
	Soil and aquifer characteristics
awc	Average available water capacity (fraction)
bd	Average bulk density (g/cm^3)
om	Average organic matter content (percentage by weight)
clay	Average clay content (percentage)
silt	Average silt content (percentage)
sand	Average sand content (percentage)
hgd	Percentage of target area with soil hydrologic group D
wtdep	Average depth to seasonally high water table (ft)
kfactu	Soil erodibility factor of uppermost soil horizon
permv	Vertical permeability (in/hr)
permin	Permeability of least permeable soils layer (in/hr)
hydric	Percentage of target area in which hydric soils were identified
drain	Soil drainage class converted to numeric
recharge	Estimated annual natural recharge from precipitation (in)
dbwt	Median depth below the water table (ft)
	Irrigation water use
wgwtot	Total ground water withdrawals for irrigation (Mgal/d)
wswtot	Total surface water withdrawals for irrigation (Mgal/d)
wtotl	Total withdrawals for irrigation (Mgal/d)
cutot	Total irrigation consumptive use (Mgal/d)
irtot	Total irrigated acres (km^2)

in table 1), and the average depth to the water table ("wtdep" in table 1). The variable "drain" was compiled using methods described in Schwarz and Alexander (1995). Mean annual natural ground-water recharge ("recharge" in table 1) was estimated from base-flow analysis and is from Wolock (2003). The median depth to water for each network was subtracted from the median well depth for each network as an estimate of the depth below the water table at which samples were collected ("dbwt" in table 1).

Irrigation water use: Estimates of 1995 irrigation water use were obtained from Solley and others (1998) and include total withdrawals and total ground-water and surface-water withdrawals for irrigation ("wtotl", "wgwtot", "wswtot" in table 1, respectively), total irrigation consumptive use ("cutot" in table 1), and total irrigated acres ("irtot" in table 1).

Exploratory statistical analysis with estimates of local recharge from precipitation and irrigation indicated that recharge explained a significant proportion (p-value < 0.05) of the variation in atrazine-detection frequency. Recharge estimates for the entire conterminous United States that were derived from base-flow analyses and Gebert runoff ("recharge" in table 1) provided less explanatory power, most likely because these estimates tend to be biased low in areas where large volumes of ground water are withdrawn from the system and do not contribute to base flow (such as areas of irrigation). Efforts to improve the explanatory power of recharge estimates for the entire conterminous United States by incorporating information on consumptive use by irrigation resulted in only small improvements. Nationwide estimates of recharge that accurately account for consumptive losses from irrigation may be developed in the future.

Estimating Atrazine Occurrence in Shallow Ground Water

OLS regression equations relating the detection frequency of atrazine to explanatory variables that represent the natural setting, agricultural-management practices, and the type and degree of development in each study area are presented in table 2. Formulations that use alternative explanatory variables presented in table 1 are possible; however, the equations listed in table 2 represent optimal n-variable (where n = 1 through 5) equations as indicated by maximizing R^2 and minimizing Mallow's C_p and PRESS. Explanatory variables that describe the likelihood of detecting atrazine in shallow ground water beneath agricultural settings include the percentage of soils classified as hydrologic group D (very slow infiltration rates), the amount of artificial drainage (tile drains and trenches), the available water-holding capacity (AWC) of soils, average vertical permeability, source of irrigation (as percentage of study area), amount of agricultural land (as percentage of study area), and intensity of atrazine use in the study areas. These explanatory variables characterize soil conditions, agricultural-management practices, and atrazine use. Combinations of these variables describe (1) the appropriateness of the setting for cultivation and, thus, the likelihood that atrazine will be used; (2) the likelihood that water will infiltrate soils; (3) the likelihood that water that has infiltrated soils will be transmitted to the water table; and (4) atrazine-use intensity (table 2).

The 5-variable formulation presented in table 2 maximizes the percentage of variation in atrazine detection frequency explained by variation in the explanatory variables (R^2) while minimizing Mallow's C_p and PRESS. Automated and manual best-subset procedures both selected the 5-variable equation presented in table 2. This agreement between equation-building approaches further indicates the lack of multicollinearity between explanatory variables (Helsel and Hirsch, 1992). The 5-variable equation shown in table 2 was, therefore, selected for further discussion and application.

Regression Equation

The selected 5-variable equation (table 2) describes 58 percent of the variation in atrazine-detection frequencies from agricultural study areas across the Nation. The equation takes the form:

$$ADF = -118 + 10.9\ ln(alluse92+1) + 0.69(ag) - 15.0(artdrain)^{1/3} + 482(awc) + 36.2\ ln(permv+1) \quad (1)$$

where

ADF is atrazine-detection frequency,

alluse92 is the intensity of atrazine use on row crops, small grains, fallow lands, pastures and hay, and orchards and vineyards in 1992,

ag is the percentage of each study area consisting of row crops, small grains, fallow lands, pastures and hay, and orchards and vineyards

artdrain is the percentage of each study area with tile drains or trenches,

awc is the average available water-holding capacity of soils,

and

permv is the average vertical permeability of soils in each study area (table 2).

The effects of each of these five variables are described below.

Atrazine-use intensity (alluse92): This variable accounted for only a small amount of the overall variation in atrazine-detection frequencies in shallow ground water from agricultural areas, as indicated by a relatively low standardized coefficient value (table 2) and a univariate R^2 of only 7 percent. In contrast, a series of equations developed for estimating the distribution of atrazine concentrations in rivers and streams throughout the Nation (Larson and others, 2004) showed atrazine-use intensity to be the single most important

Table 2. Statistics for ordinary least-squares regression equations.

[PRESS, prediction error sum of squares; --, no data]

Number of explanatory variables	Variable[1]	Coefficient value, ($ß_1$)	Standardized coefficient value	*p*-value	VIF	Model *p*-value	Variance explained (R^2)	Mallow's Cp	PRESS
1	ln(hgd+1)	-12.2	-0.41	0.002	1.00	0.002	0.17	68.2	44,796
	intercept	64	0	<.0001	0	--	--	--	--
2	artdrain$^{1/3}$	-12.8	-.53	.0002	1.22	.0001	.31	51.4	38,968
	ln(alluse92+1)	11.7	.50	.0004	1.22	--	--	--	--
	intercept	40	0	<.0001	0	--	--	--	--
3	ln(is01+1)	8.5	.38	.004	1.30	<.0001	.42	38.0	33,673
	artdrain$^{1/3}$	-12.9	-.54	<.0001	1.22	--	--	--	--
	ln(alluse92+1)	16.1	.68	<.0001	1.53	--	--	--	--
	intercept	20	0	.04	0	--	--	--	--
4	ln(alluse92+1)	11.2	.48	.0003	1.38	<.0001	.50	29.6	31,437
	artdrain$^{1/3}$	-13.2	-.55	<.0001	1.24	--	--	--	--
	awc	527	.47	.002	1.92	--	--	--	--
	ln(permv+1)	33.6	.57	.0001	1.80	--	--	--	--
	intercept	-70	0	.02	0	--	--	--	--
5	ln(alluse92+1)	10.9	.46	.0002	1.38	<.0001	.58	19.5	27,447
	ag	.69	.30	.004	1.18	--	--	--	--
	artdrain$^{1/3}$	-15.0	-.63	<.0001	1.31	--	--	--	--
	awc	482	.44	.002	1.94	--	--	--	--
	ln(permv+1)	36.2	.62	<.0001	1.84	--	--	--	--
	intercept	-118	0	.0004	0	--	--	--	--

[1] hgd: percentage of area with soil hydrologic group D.

artdrain: percentage of area with artificial drainage [aggregate of 1992 National Resources Inventory conservation practices cp606, cp607, cp608 square kilometer (km^2)/study area (km^2)] x 100.

alluse92: sum of 1992 atrazine use on row crops, small grains, fallow land, pastures and hay, and orchards and vineyards kilogram (kg)/area of row crops, small grains, fallow land, pastures and hay, and orchards and vineyards (km^2).

is01: percentage of area using ground water for irrigation source [1992 NRI irrigation source is01 (km^2)/study area (km^2)] x 100.

awc: average available water capacity (fraction).

permv: vertical permeability inch per hour (in/hr).

ag: percentage of area in agricultural land use [sum of National Land Cover Data Set - Enhanced (Price and others, 2003) categories row crops, small grains, fallow land, pastures and hay, and orchards and vineyards (km^2)/study area (km^2)] x 100.

explanatory variable and to account for more than half of the observed variation in atrazine-concentration distributions. The difference between the explanatory power of atrazine-use intensity in surface-water systems and those of ground-water systems reflects, in part, differences in the nature of contaminant transport and transformation. Surface-water systems tend to respond rapidly to changing conditions in the watershed, whereas ground-water systems show a lag because of the time required for water to infiltrate soils, percolate to the water table, and travel along ground-water flow paths to points of discharge. The extended residence times associated with ground-water systems favor the sorption and degradation of pesticides, which further complicates attempts to relate ground-water chemistry to activities at land surface (Kolpin and others, 2004). Surface-water samples tend to integrate spatial and temporal factors more efficiently whereas ground-water samples from monitoring wells tend to represent water from a discrete location within the ground-water flow system that was recharged at a discrete time. Because the age and origin of ground-water samples is generally less well known than for surface-water samples, relating activities at land surface to ground-water chemistry commonly results in less explanatory power than provided in surface-water studies. In addition, as previously discussed, the atrazine-use data compiled for this study represent only agricultural applications; any nonagricultural use (although thought to be minimal) is not represented in the use-intensity variable.

Agricultural land (ag): The correlation between atrazine-detection frequency and agricultural land use reflects the use of atrazine primarily (although not exclusively) as an agricultural herbicide; it has been one of the most heavily applied agricultural herbicides throughout the Nation since the 1960's. The relatively low standardized coefficient value for the agricultural-land variable (table 2) probably reflects the inclusion of categories with little or no atrazine use (such as pastures and hay) in the compilation of this variable. The agricultural-land studies were designed to incorporate predominantly agricultural settings; therefore, nonagricultural settings are underrepresented. The amount of agricultural land ranged from about 40 percent to nearly 100 percent of the study areas. Other land-use categories were sparsely represented and, therefore, were not found to be significantly related to atrazine occurrence.

Artificial drainage (artdrain): The percentage of a study area drained by tile drains or trenches was found to be inversely related to atrazine-detection frequency. This negative correlation, and the magnitude of the standardized coefficient value for this variable (table 2), reflect the efficiency with which artificial drainage systems capture soil water that would otherwise recharge the ground-water system and divert it to nearby streams. Thus, the transport of atrazine to the water table in areas with extensive artificial drainage is disrupted, and the frequency with which atrazine is detected in shallow ground water in such areas is decreased.

Available water-holding capacity of soils (AWC): This variable was positively correlated with atrazine-detection frequencies, which corroborates findings of Kolpin and others (2002) and Burkart and others (1999). Low AWC values tend to be associated with sandy, permeable soils that readily transmit water from land surface to the water table and, therefore, AWC would be expected to be negatively correlated with atrazine-detection frequencies. Loamy soils, which naturally have higher AWC values than sandy soils, are generally preferred for crop cultivation; therefore, atrazine use is typically greater on soils with higher AWC values, which results in a positive correlation (Kolpin and others, 2002). In this way, AWC captures information not only on soil characteristics, but also on atrazine-use intensity.

Average vertical permeability of soils (permv): This variable was found to be positively correlated with atrazine-detection frequencies and to have the largest standardized coefficient value (table 2). This further corroborates the importance of soil characteristics in the transmission of water and of water-soluble pesticides from land surface to the water table. Less permeable soils will cause storm runoff to flow into streams rather than to infiltrate the soils and percolate to the water table.

Residuals Analysis

Studentized residuals from the 5-variable equation adhere to the assumptions of normality and constant variation across estimated values (figs. 2A–B). All studentized residuals were within 2.5 standard deviations of the regression line (fig. 2A), but three warrant explanation. Two studies with negative estimations are from southern Florida and the Nevada Basin and Range (fig. 2B). The study with the largest overestimation is from the White River Basin in Indiana (fig. 2B).

The southern Florida study was conducted in the surficial aquifer system. This study had the highest percentage of artificial drainage in the 52 study areas and a relatively low average AWC value. Values from this study for the other three explanatory variables in the equation were generally average in relation to those in the other 51 studies; thus, the negative estimation can be attributed to the large amount of artificial drainage and low average AWC (fig. 2B).

The Nevada Basin and Range study was in the Carson desert of Nevada, where no atrazine use was reported for row crops, small grains, or fallow lands in 1992. Accordingly, no atrazine was detected in ground-water samples. Values from this study for the other four explanatory variables in the equation, however, were average in relation to those in the other 51 studies; thus, the estimated negative atrazine-detection frequency can be attributed to the absence of atrazine use.

The White River Basin study was in a till plain in Indiana where ground-water samples were generally collected from stringers of sand embedded in poorly conductive till sediments. These sand stringers can be hydraulically isolated from land surface and some sand stringers are under confined conditions (Fenelon and Moore, 1996; 1998). Atrazine was detected in only 18 percent of samples from this study area. This study had the largest absolute studentized residual (-2.18),

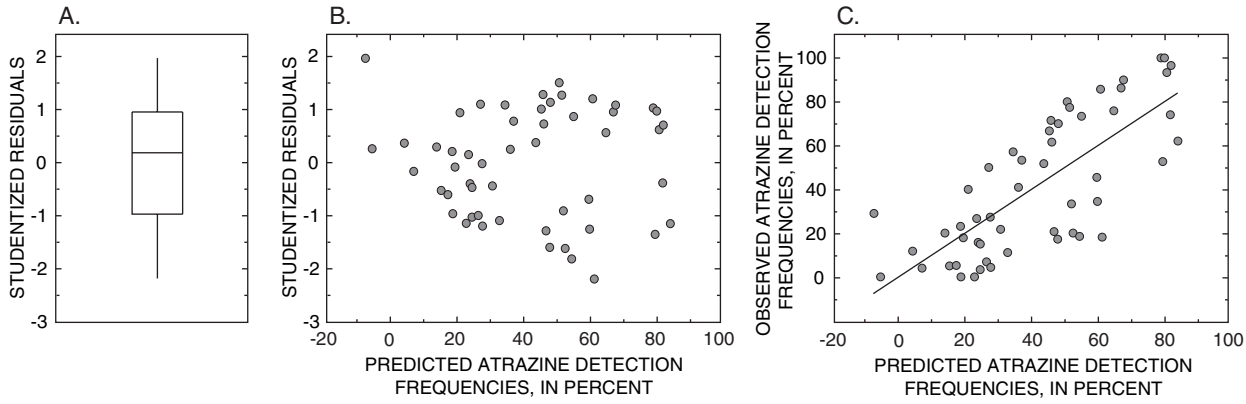


Figure 2. Distribution of (A) studentized residuals, (B) studentized residuals and estimated atrazine detection frequencies, and (C) observed and estimated atrazine detection frequencies.

which represents an overestimation of atrazine occurrence (fig. 2B). Values from this study for the explanatory variables "alluse92," "ag," and "awc" were among the highest 8 values from all 52 studies, and result in an estimation of atrazine-detection frequency of 61 percent.

Equation Performance

Equation performance (fig. 2C) depicts observed atrazine-detection frequencies in relation to estimated atrazine-detection frequencies. Ninety-six percent (50 of 52) of the estimations are within a factor of 10 of the observed atrazine-detection frequencies reported from the 52 agricultural-land studies, and 73 percent are within a factor of 2.5. The two estimated detection frequencies that differ from observed values by more than a factor of 10 are from locations where little or no agricultural use of atrazine was reported in 1992 and where atrazine was not detected in ground-water samples.

Estimation of Atrazine Occurrence

The selected 5-variable equation (table 2) was applied to a gridded map of agricultural areas across the conterminous United States to estimate the frequency of atrazine occurrence (percentage of samples) in areas not monitored in the 52 agricultural-land studies. The explanatory variables were quantified for each 1-km^2 cell in which agricultural land accounted for at least 50 percent of the total area, and an estimation was generated. The geographic distribution of atrazine use for agricultural purposes throughout the conterminous United States is depicted in figure 3A. The estimated atrazine-detection frequency in shallow ground water is depicted in figure 3B. The estimations illustrate that atrazine occurrence in shallow ground water in agricultural areas is not necessarily greatest in areas of greatest atrazine use because other factors can affect its transport from land surface to the water table as well as its transformation to degradates. For example, some of the greatest atrazine use is in Iowa, Illinois, and Indiana but soils in these areas tend to be poorly drained; thus, agricultural fields are drained through systems of subsurface tile drains or surface trenches. As a result, the concentrations in ground water of heavily applied pesticides such as atrazine are lower than they would be otherwise, and the concentrations in streams are higher. In contrast, areas with extensive atrazine use on soils that favor infiltration and do not require artificial drainage systems are estimated to have the highest rate of atrazine detection in shallow ground water (for example, western Iowa and eastern Nebraska; fig. 3B). Thus, atrazine use alone is an insufficient variable for estimating the occurrence of atrazine in shallow ground water (Barbash and others, 1999) because soil conditions and agricultural-management practices are important factors in determining the extent to which the ground-water system is vulnerable to contamination from surface sources.

The selected equation's greatest use to water-resource managers is its ability to identify areas of atrazine occurrence in shallow ground water (fig. 3B). Such estimations may also be useful in the design of future ground-water monitoring programs and for prioritizing research areas. The nationwide scale at which the equation was developed, however, could lead to erroneous estimations at the local scale because it may not contain explanatory variables that strongly affect atrazine occurrence in shallow ground water locally. Application of a broad-based equation to a local setting may either over- or underestimate the likelihood of atrazine occurrence; therefore, natural and human-related factors that affect atrazine transport to the water table locally must be considered before interpreting these results at the local scale.

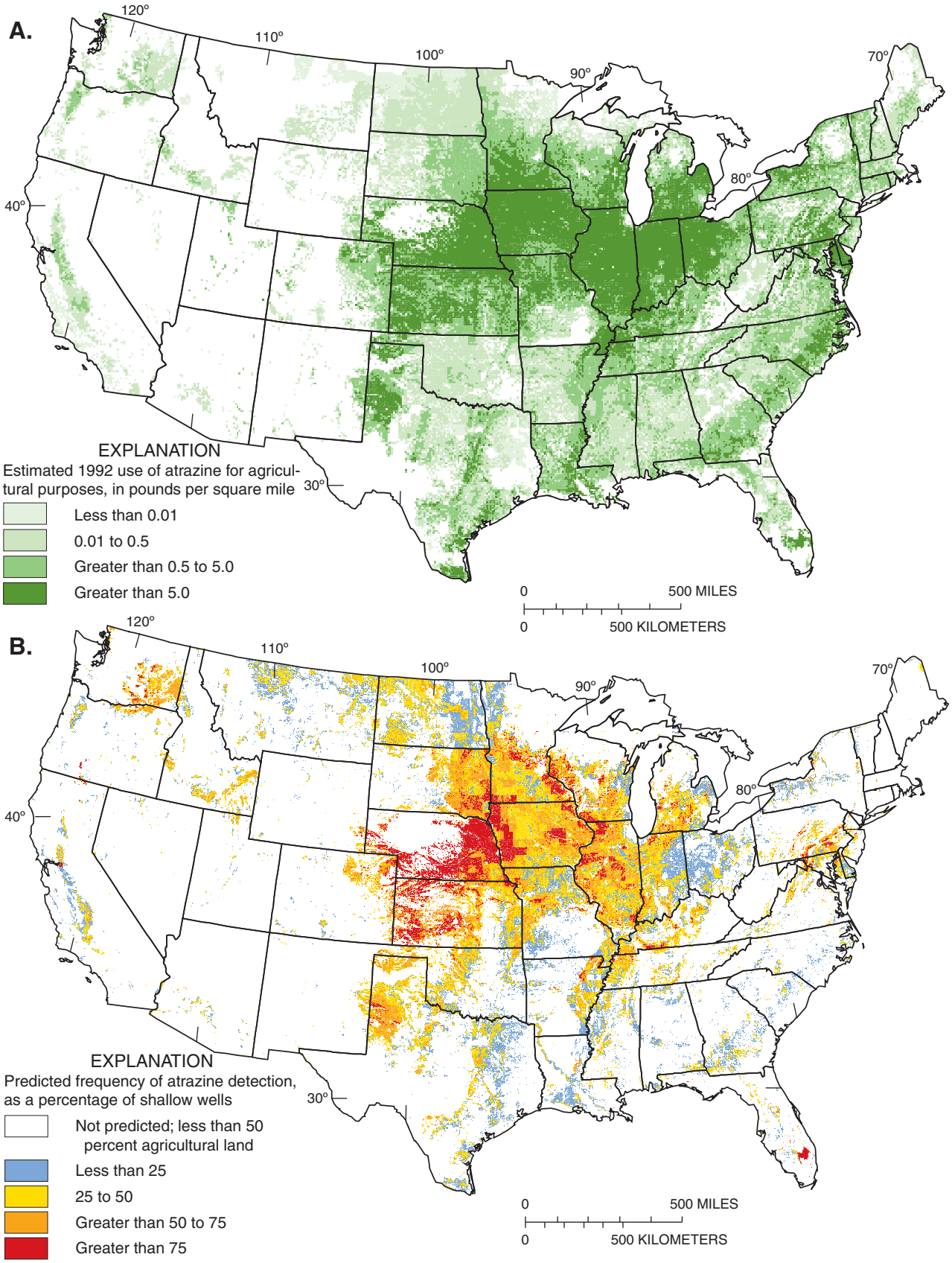


Figure 3. (A) Estimated 1992 agricultural use of atrazine throughout the conterminous United States, and (B) estimated frequency of atrazine occurrence in shallow ground water underlying agricultural areas throughout the conterminous United States.

Summary

Results from ground-water studies conducted in 52 agricultural areas throughout the Nation during 1991–2001 as part of the U.S. Geological Survey's NAWQA Program were analyzed to identify the major factors that determine the occurrence and distribution of atrazine in shallow ground water underlying agricultural land. OLS multivariate regression was used to examine relations between atrazine-detection frequency and a comprehensive set of explanatory variables that represent the natural setting, agricultural-management practices, and the type and amount of development in each study area. The frequency with which atrazine occurs in shallow agricultural ground water was found to correlate significantly with soil-infiltration rates, presence and extent of artificial drainage (tile drains and trenches), the available water-holding capacity of soils, soil permeability, amount of study area using ground water for irrigation source (as percentage of total area), the percentage of agricultural land within the study area, and the amount of atrazine applied per acre of agricultural land in the study area. Results indicate that variables that represent soil characteristics, agricultural-management practices, and rate of atrazine use are reliable estimators of the occurrence of atrazine in shallow ground water at the nationwide scale. Equations that incorporate these explanatory variables accounted for as much as 58 percent of the variation in atrazine-detection frequencies reported from the 52 studies. Application of a multivariate equation to unmonitored agricultural areas illustrates that atrazine use alone is an insufficient variable for estimating the occurrence of atrazine in shallow ground water because soil characteristics and agricultural-management practices are important factors in determining the vulnerability of shallow ground water to atrazine contamination.

Acknowledgments

The authors would like to thank all of the members of the NAWQA study units who conducted the ground-water studies and provided valuable insight into the hydrogeology and environmental setting of their respective study areas. The authors would also like to thank Michael Rosen and Anthony Tesoriero of the USGS for providing valuable reviews of the manuscript.

References Cited

Barbash, J.E., Thelin, G.P., Kolpin, D.W., and Gilliom, R.J., 1999, Distribution of major herbicides in the ground water of the United States: U.S. Geological Survey Water-Resources Investigations Report 98–4245, 57 p.

Burkart, M.R., Kolpin, D.W., Jaquis, R.J., and Cole, K.J., 1999, Agrichemicals in ground water of the Midwestern USA—relations to soil characteristics: Journal of Environmental Quality, v. 28, no. 6, p. 1908–1915.

Draper, N.R., and Smith, H., 1966, Applied regression analysis: New York, John Wiley, 407 p.

Fenelon, J.M., and Moore, R.C., 1996, Occurrence of pesticides in ground water in the White River Basin, Indiana, 1994–95: U.S. Geological Survey Fact Sheet 084–96, 4 p.

Fenelon, J.M., and Moore, R.C., 1998, Transport of agrichemicals to ground and surface water in a small central Indiana watershed: Journal of Environmental Quality, v. 27, no. 4, p. 884–894.

Gilliom, R.J., Alley, W.M., and Gurtz, M.E., 1995, Design of the National Water-Quality Assessment Program—Occurrence and distribution of water-quality conditions: U.S. Geological Survey Circular 1112, 33 p.

Hamilton, P.A., Miller, T.L., and Myers, D.N., 2004, Water quality in the Nation's streams and aquifers—Overview of selected findings, 1991–2001: U.S. Geological Survey Circular 1265, 22 p.

Helsel, D.R., and Hirsch, R.M., 1992, Statistical methods in water resources: New York, Elsevier, 522 p.

Kolpin, D.W., Barbash, J.E., and Gilliom, R.J., 1998, Occurrence of pesticides in shallow ground water of the United States—initial results from the National Water-Quality Assessment Program: Environmental Science and Technology, v. 32, no. 5, p. 558–566.

Kolpin, D.W., Barbash, J.E., and Gilliom, R.J., 2000, Pesticides in ground water of the United States, 1992–1996: Ground Water, v. 38, no. 6, p. 858–863.

Kolpin, D.W., Barbash, J.E., and Gilliom, R.J., 2002, Atrazine and metolachlor occurrence in shallow ground water of the United States, 1993 to 1995—relations to explanatory factors: Journal of the American Water Resources Association, v. 38, no. 1, p. 301–311.

Kolpin, D.W., Schnoebelen, D.J., and Thurman, E.M., 2004, Degradates provide insight to spatial and temporal trends of herbicides in ground water: Ground Water, v. 42, no. 4, p. 601–608.

Koterba, M.T., Wilde, F.D., and Lapham, W.W., 1995, Ground-water data-collection protocols and procedures for the National Water-Quality Assessment Program—Collection and documentation of water-quality samples and related data: U.S. Geological Survey Open-File Report 95–399, 113 p.

Koterba, M.T., 1998, Ground-water data-collection protocols and procedures for the National Water-Quality Assessment Program—Collection, documentation, and compilation of required site, well, subsurface, and landscape data for wells: U.S. Geological Survey Water-Resources Investigations Report 98–4107, 91 p.

Lapham, W.W., Wilde, F.D., and Koterba, M.T., 1995, Ground-water data-collection protocols and procedures for the National Water-Quality Assessment Program—selection, installation, and documentation of wells, and collection of related data: U.S. Geological Survey Open-File Report 95–398, 69 p.

Larson, S.J., Crawford, C.G., and Gilliom, R.J., 2004, Development and application of watershed regressions for pesticides (WARP) for estimating atrazine concentration distributions in streams: U.S. Geological Survey Water-Resources Investigations Report 03–4047, 68 p.

Oblinger Childress, C.J, Foreman, W.T., Connor, B.F., and Maloney, T.J., 1999, New reporting procedures based on long-term method detection levels and some considerations for interpretations of water-quality data provided by the U.S. Geological Survey National Water-Quality Laboratory: U.S. Geological Survey Open-File Report 99–193, 19 p.

Ott, R.L., and Longnecker, M., 2001, An introduction to statistical methods and data analysis: California, Duxbury, 1152 p.

Price, C., Nakagaki, N., Hitt, K., and Clawges, R., 2003, Mining GIRAS—Improving on a national treasure of land use data, in Proceedings of the 23rd ESRI International Users Conference, July 7–11, 2003: Redlands, Calif., 11 p., accessed July 2004, at http://gis.esri.com/library/userconf/proc03/p0904.pdf

SAS Institute Inc., 1991, SAS® system for regression, second edition: Cary, N.C.: SAS Institute, Inc., 210 p.

Schwarz, G.E., and Alexander, R.B., 1995, Soils data for the conterminous United States derived from the NRCS State Soil Geographic (STATSGO) data base: U.S. Geological Survey Open-File Report 95–449, accessed July 2004, at http://water.usgs.gov/lookup/getmeta?ussoils

Scott, J.C., 1990, Computerized stratified random site-selection approaches for design of a ground-water-quality sampling network: U.S. Geological Survey Water-Resources Investigations Report 90–4101, 109 p.

Solley, W.B., Pierce, R.R., and Perlman, H.A., 1998, Estimated use of water in the United States in 1995: U.S. Geological Survey Circular 1200, accessed July 2004, at http://water.usgs.gov/watuse/pdf1995/html

Thorton, P.E., 2003, DAYMET climatological summaries for average air temperature and total precipitation (18-year mean for 1980–1997) [digital data]: accessed at http://daymet.org

U.S. Bureau of the Census, 1991, Census of population and housing, 1990: Public Law 94-171 data (United States) [machine-readable data files on CD-ROM]: Washington, D.C., U.S. Census Bureau.

U.S. Natural Resources Conservation Service, 1994, State Soil Geographic (STATSGO) data base for the United States and Puerto Rico [CDROM]: Ft. Worth, Tex., accessed July 2004, at http://www.ncgc.nrcs.usda.gov/products/datasets/statsgo/.

U.S. Natural Resources Conservation Service, 1995, 1992 National Resources Inventory [CD-ROM reissued May 1995]: Ames, Iowa, Iowa State University, Statistical Laboratory.

Vogelmann, J.E., Howard, S.M., Yang, L., Larson, C.R., Wylie, B.K., and Van Driel, N., 2001, Completion of the 1990s national land cover data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources: Photogrammetric Engineering and Remote Sensing, v. 67, no. 6, p. 650–652, accessed July 2004, at http://www.asprs.org/asprs/publications/pe&rs/2001journal/june/highlight.html

Wolock, D.W., 1997, STATSGO soil characteristics for the conterminous United States: U.S. Geological Survey Open-File Report 97–656, accessed July 2004, at http://water.usgs.gov/lookup/getmeta?muid

Wolock, D.W., 2003, Estimated mean annual natural ground-water recharge in the conterminous United States: U.S. Geological Survey Open-File Report 03–311, accessed July 2004, at http://water.usgs.gov/lookup/getspatial?rech48grd

Zaugg, S.D., Sandstrom, M.W., Smith, S.G., and Fehlberg, K.M., 1995, Methods of analysis by the U.S. Geological Survey National Water-Quality Laboratory—determination of pesticides in water by C-18 solid-phase extraction and capillary-column gas chromatography/mass spectrometry with selected-ion monitoring: U.S. Geological Survey Open-File Report 95–181, 49 p.

For more information concerning the research in this report, contact:
Rafael W. Rodriguez, Director
U.S. Geological Survey
New York Water Science Center
425 Jordan Rd.
Troy, NY 12180-8349

or visit our Web site at:
http://ny.water.usgs.gov

Printed on recycled paper