Regression Models for Estimating Concentrations of Atrazine plus Deethylatrazine in Shallow Groundwater in Agricultural Areas of the United States

Paul E. Stackelberg,* Jack E. Barbash, Robert J. Gilliom, Wesley W. Stone, and David M. Wolock

Tobit regression models were developed to predict the summed concentration of atrazine [6-chloro-N-ethyl-N'-(1-methylethyl)-1,3,5-triazine-2,4-diamine] and its degrade deethylatrazine [6-chloro-N-(1-methylethyl)-1,3,5-triazine-2,4-diamine] (DEA) in shallow groundwater underlying agricultural settings across the conterminous United States. The models were developed from atrazine and DEA concentrations in samples from 1298 wells and explanatory variables that represent the source of atrazine and various aspects of the transport and fate of atrazine and DEA in the subsurface. One advantage of these newly developed models over previous national regression models is that they predict concentrations (rather than detection frequency), which can be compared with water quality benchmarks. Model results indicate that variability in the concentration of atrazine residues (atrazine plus DEA) in groundwater underlying agricultural areas is more strongly controlled by the history of atrazine use in relation to the timing of recharge (groundwater age) than by processes that control the dispersion, adsorption, or degradation of these compounds in the saturated zone. Current (1990s) atrazine use was found to be a weak explanatory variable, perhaps because it does not represent the use of atrazine at the time of recharge of the sampled groundwater and because the likelihood that these compounds will reach the water table is affected by other factors operating within the unsaturated zone, such as soil characteristics, artificial drainage, and water movement. Results show that only about 5% of agricultural areas have greater than a 10% probability of exceeding the USEPA maximum contaminant level of 3.0 μg L⁻¹. These models are not developed for regulatory purposes but rather can be used to (i) identify areas of potential concern, (ii) provide conservative estimates of the concentrations of atrazine residues in deeper potential drinking water supplies, and (iii) set priorities among areas for future groundwater monitoring.

Published March, 2012

The U.S. Geological Survey National Water-Quality Assessment (NAWQA) Program has monitored the occurrence of pesticides and other contaminants at thousands of sites throughout the United States and in a variety of media since 1992 (Gilliom et al., 2006; Dubrovsky et al., 2010). The resulting data provide a resource for the development of national- and regional-scale statistical models to estimate concentrations or frequencies of occurrence of pesticides and other contaminants in groundwater (Kolpin et al., 2002; Stackelberg et al., 2006a; Nolan and Hitt, 2006), stream water (Larson et al., 2004; Stone et al., 2008; and Stone and Gilliom, 2009), and fish tissue (Nowell et al., 2009). Development of these predictive capabilities is a vital component of a comprehensive national water quality assessment because, given the infeasibility of monitoring everywhere, the models provide a cost-effective means of understanding the chemical quality of the United States’ water resources and for estimating water quality conditions for unmonitored areas. This information is vital to improving water quality management and for setting efficient monitoring priorities.

Atrazine is one of the most extensively used agricultural herbicides in the United States, with an average annual use (1992–1995) of 29 million kg active ingredient applied to about 231,000 km² of cropland (primarily corn and sorghum) (Thelin and Gianessi, 2000). National models of atrazine occurrence in shallow groundwater were previously developed by Kolpin et al. (2002) and Stackelberg et al. (2006a) to evaluate atrazine detection frequencies in well networks (typically 20–30 wells each) associated with NAWQA land-use studies (Gilliom et al., 2006). Explanatory variables, values of which were averaged across each study area, were used to represent the natural setting, agricultural management practices, and the types and amount of development in each area. The model of Kolpin et al. (2002) was developed from data on the occurrence of atrazine in shallow groundwater underlying agricultural and urban settings, whereas the model of Stackelberg et
al. (2006a) focused solely on atrazine occurrence in agricultural settings. Both models used ordinary least-squares regression methods to identify and quantify the effects of factors that were significantly correlated with atrazine detection frequencies. Stackelberg et al. (2006a) also applied their model to predict the frequency with which atrazine would be detected in groundwater beneath unmonitored agricultural settings across the conterminous United States.

This study modified and expanded the atrazine model of Stackelberg et al. (2006a) by (i) using atrazine residue concentrations (i.e., the sum of the concentrations of atrazine and one of its degradates, deethylatrazine [DEA]) as the response variable rather than atrazine-detection frequencies and (ii) incorporating data for explanatory variables derived from site-specific information (such as water chemistry) that was available only for the locations at which water was measured. The reason for using atrazine residue concentrations as the response variable, rather than the concentration of atrazine alone, was to capture more information on factors that control the transport and fate of atrazine-derived compounds in the subsurface. Three other primary atrazine degradates—deisopropylatrazine, die delaikylatrazine, and hydroxyatrazine—were considered but were not examined because of considerations related to analytical performance (Mark Sandstrom, USGS, written communication, 2011). The specific objectives of this study were (i) to develop regression models that can predict atrazine residue concentrations as well as the probability that selected concentration thresholds will be exceeded in agricultural settings across the conterminous United States and (ii) to ascertain whether the incorporation of explanatory variables derived from site-specific information can improve our understanding of the relative importance of different factors that control the concentrations of atrazine and DEA in shallow groundwater beneath agricultural areas.

Materials and Methods

Data for Model Development and Evaluation

Estimates of annual atrazine use, in kilograms of active ingredient per square kilometer (kg km\(^{-2}\)) were available for agricultural settings with a wide range of use intensity, although most use is on corn and sorghum. Usage in nonagricultural settings is less than in agricultural settings (Gilliom et al., 2006), but reliable estimates of nonagricultural use are not available in most parts of the United States. Given the direct relation between atrazine use and occurrence in groundwater, all model-development and model-evaluation data were collected in areas dominated by agricultural land use. Because the models were developed from information collected solely through the NAWQA Program, the water quality data and explanatory variables represent only those settings monitored as part of the Program (Gilliom et al., 2006). The water quality data used in model development were collected between 1993 and 2001 from 1298 wells in 55 study areas representing regionally extensive combinations of agricultural land and hydrogeologic conditions (Gilliom et al., 2006). Model evaluation entailed the use of corresponding data from an additional 209 sites in 19 study areas that were sampled as part of the NAWQA Program during 2001–2008. Locations of the model-development and model-evaluation sites are shown in Fig. 1.

Sampling and Analytical Methods

All groundwater sampling sites were selected through methods described by Scott (1990), and the wells were installed in accordance with protocols described by Lapham et al. (1995) for the collection of “recently recharged” (young, shallow) groundwater. The median depth of the sampled wells was 9.5 m, and the median distance from the water table to the bottom of the wells was 4.6 m.

Water samples were collected in accordance with procedures and protocols described in Koterba et al. (1995). All samples were analyzed at the USGS National Water-Quality Laboratory in Denver, CO, by gas chromatography/mass spectrometry (GC–MS) with selected ion monitoring (Zaugg et al., 1995). The GC–MS analytical method does not have specified “detection limits” for each pesticide analyte; rather, compounds detected and conclusively identified by retention time and spectral characteristics are quantified and reported (Zaugg et al., 1995; Oblinger Childress et al., 1999; Stackelberg et al., 2006b; Martin, 2009). Analyses that do not meet identification criteria based on retention time and spectral characteristics are deemed “nondetections” and are reported as less than their reporting level. The types and numerical values of reporting levels used to report nondetections analyzed by GC–MS changed during the period represented by the samples used for this study (1993–2008). Removal of the resulting temporal structure in the reporting levels entailed reassigning the concentration value for each nondetection to the long-term method-detection level (LTMDL) for water years 1994 to 2006 (Martin, 2009), which for atrazine and DEA are 0.004 and 0.007 μg L\(^{-1}\), respectively. (A water year is the period 1 October through 30 September and is named for the year in which the water year ends.)

Analytical results for atrazine and DEA, including censored concentrations (those deemed to be nondetections) were adjusted to account for differences in analytical recoveries. This was done through methods described in Martin et al. (2009) and was necessary because of a pronounced disparity in the recoveries of the two analytes; the mean recovery of atrazine from groundwater matrix-spike samples during 1992–2006 was 99.6%, whereas the corresponding recovery for DEA was 47.2% (Martin et al., 2009).

Explanatory Variables

A total of 88 factors (explanatory variables) deemed likely to affect the concentration of atrazine or DEA in shallow groundwater were considered for model development. Data on each of these variables were compiled for a 500-m radius surrounding each well site; exceptions were (i) variables derived from site-specific information (such as water chemistry data) that can be measured only at a discrete location and (ii) the variable “artdrn” (prevalence of artificial drainage), for which the data were compiled for grid cells that were 1 km\(^2\) in area. Explanatory variables were transformed as needed to accommodate the assumption that the regressions were linear and to ensure that the model residuals were normally distributed with
Transformations that were considered included (i) raising individual explanatory variables to powers of 0.33 (cube root), 0.5 (square root), and 2 (square) and (ii) converting to their logarithm. (All logarithms used during this study were base 10 values.) The explanatory variables considered for use in development of the final models were selected using methods described in the next section. Those that were selected are grouped into seven categories in Table 1. Several of these variables are discussed below.

Intensity of Atrazine Use

The estimated use of atrazine for agricultural purposes in the 500-m-radius area around each well site (use9200 and use97 in Table 1) was obtained by combining county-level estimates of applications of atrazine to agricultural crops with information on the spatial distribution of agricultural land within each radius using methods described by Larson et al. (2004) and Thelin and Gianessi (2000). Estimates of atrazine use between 1992 and 2000 (use9200) were derived from proprietary-use data obtained from the Doane Marketing Research-Kynetec (DMRKYNETEC) Agro Trak database (Thelin and Stone, 2010), and estimates for 1997 use (use97) were derived from data from the National Center for Food and Agricultural Policy (Thelin and Gianessi, 2000) (Table 1). The source for the spatial distribution of agricultural land was the 2001 National Land Cover Data set, which was obtained at 30-m resolution (NLCD01; Vogelmann et al., 2001). “Agricultural land” in the calculation of atrazine use consists of the NLCD01 classifications of pasture/hay and the combined grouping of row crops/small grains/fallow land with orchards/vineyards. The intensity of atrazine use (kg km$^{-2}$) was calculated as the total amount of atrazine applied to the crops of interest divided by the radius area.

Land Use

The land-use variable agrcveg represents the combined categories of row crops and vegetables, and agcc represents cover crops (Table 1). These two variables represent ancillary data that are required for wells from which water quality data are collected for the NAWQA Program. Details on methods used to collect these ancillary data were provided by Koterba (1998).

Agricultural Management Practices

Data on subsurface tile drains and surface ditches were obtained from the 1992 National Resources Inventory of the Natural Resources Conservation Service. The variable artdrn (Table 1) represents the percentage of 1-km$^2$ grid cells with subsurface drains and graded ditches as represented by National Resources Inventory conservation practice codes CP606, CP607, and CP608 (NRCS, 1995).

Recharge

Mean annual groundwater recharge (recharge in Table 1) was computed from estimates of base flow and mean annual runoff as described by Wolock (2003) and modified to include groundwater withdrawn for use in irrigation (Karen Burow, USGS, written communication, May 2009).

Water Quality

Spatial patterns in groundwater quality were identified through principal component (PC) analysis of water quality data (Table 2). Principal components analysis was run on ranked data to accommodate censored values as described in Ator (2008). Two PCs that explain more than 50% of the variability in water quality data were considered in this study. The first PC explained 32% of the variability and was interpreted to represent overall ionic strength ($IS$ in Table 1). This PC showed strong positive loadings (which indicate the correlation coefficient between a variable and a PC) for specific conductance and the concentrations of major cations (Ca$^{2+}$, Mg$^{2+}$, and Na$^+$) and anions (Cl$^-$ and SO$_4^{2-}$) (Table 2). The second PC explained 20% of the variability and was interpreted to be an indicator of groundwater residence time in the saturated zone ($RTI$ in Table 1). A detailed discussion of $RTI$ is presented in the Results section.

Model-Development Approach

Measured atrazine and DEA concentrations, adjusted to approximate 100% analytical recovery, ranged from <0.004 to 4.7 μg L$^{-1}$ for atrazine and from <0.007 to 8.2 μg L$^{-1}$ for DEA. The sum of the concentrations of atrazine and DEA, referred to herein as “atrazine residue,” was transformed to its logarithm to make its relation with explanatory variables more linear, reduce heteroscedasticity, and shift the distribution of model residuals closer to normality.
Regression Analysis

In this study, 52% of the atrazine and DEA concentrations were censored (i.e., deemed to be nondetections); therefore, a Tobit model, which provides parameter estimates for a censored linear model (Judge et al., 1985; Tobin, 1958), was used to develop the regression models as described in Larson et al. (2004), Stone et al. (2008), and Nowell et al. (2009). Model parameters were estimated using maximum likelihood methods implemented by the SURVREG procedure (Therneau, 1999) in the statistical analysis program S-PLUS (TIBCO Software Inc., 2008). Maximum likelihood computations require lower and upper bounds be defined for left and interval censored samples (Helsel, 2005). The response variable in this study represents the summed concentration of atrazine and DEA (after both were adjusted to 100% analytical recovery, as described earlier). For samples in which neither atrazine nor DEA were detected, the response variable was left censored with a lower bound at 0 and the upper bound established as the sum of the LTMDLs (0.011 μg L⁻¹). For samples in which atrazine or DEA (but not both) was detected, the response variable was interval censored with the lower bound established by summing concentrations with the assumption that the censored value was equal to zero and the upper bound established by summing concentrations with the assumption that the censored value was equal to its LTMDL. For a sample in which atrazine concentration was reported as 0.009 μg L⁻¹ and DEA was not detected (and reported as less than the LTMDL of 0.007 μg L⁻¹), the response variable would be considered interval censored, and the lower and upper bounds of the interval would be 0.009 and 0.016 μg L⁻¹, respectively.

Explanatory variables for each model were screened through a stepwise, interval-censored Tobit regression procedure that used the Akaike Information Criterion statistic to select variables. The Akaike Information Criterion statistic balances the desire to explain as much variability as possible in the response variable (here, atrazine residue concentration) with the

<table>
<thead>
<tr>
<th>Table 1. Explanatory variables used for final model development.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td><strong>Intensity of atrazine use</strong></td>
</tr>
<tr>
<td>use9200†</td>
</tr>
<tr>
<td>use97‡</td>
</tr>
<tr>
<td><strong>Land use and population</strong></td>
</tr>
<tr>
<td>pasthay§</td>
</tr>
<tr>
<td>agrcvec‡</td>
</tr>
<tr>
<td>agcc‡</td>
</tr>
<tr>
<td>pden90†</td>
</tr>
<tr>
<td><strong>Agricultural management practices</strong></td>
</tr>
<tr>
<td>artdrn¶</td>
</tr>
<tr>
<td><strong>Soil and aquifer characteristics and well depth</strong></td>
</tr>
<tr>
<td>awcup§</td>
</tr>
<tr>
<td>awcave†</td>
</tr>
<tr>
<td>omup§</td>
</tr>
<tr>
<td>wtdep¶</td>
</tr>
<tr>
<td>permin¶</td>
</tr>
<tr>
<td>bdup†</td>
</tr>
<tr>
<td>bd§</td>
</tr>
<tr>
<td>siltup†</td>
</tr>
<tr>
<td>runoff‡</td>
</tr>
<tr>
<td>perdun†</td>
</tr>
<tr>
<td>recharge¶</td>
</tr>
<tr>
<td>welldepth‡</td>
</tr>
<tr>
<td><strong>Climate</strong></td>
</tr>
<tr>
<td>aitemp¶</td>
</tr>
<tr>
<td><strong>Water use</strong></td>
</tr>
<tr>
<td>wtot§</td>
</tr>
<tr>
<td>wswto†</td>
</tr>
<tr>
<td><strong>Water quality</strong></td>
</tr>
<tr>
<td>IS‡</td>
</tr>
<tr>
<td>RT¶</td>
</tr>
<tr>
<td>pH‡</td>
</tr>
</tbody>
</table>

† Nationally available variables considered in the development of the national-variable model.
‡ Site-specific variables considered in the development of the site-variable model (Table 3).
§ Nationally available variables considered in the development of the site-variable model.
¶ Nationally available variables considered in the development of both the national-variable model and the site-variable model.
practicality of including as few explanatory variables as possible (Helsel and Hirsch, 1992). Groups of explanatory variables were sequentially added before each step of the stepwise procedure, and variables were retained if their p values remained ≤0.25. A p value of 0.25 was used at this stage of model development to ensure that all explanatory variables that might contribute significantly to a final model formulation would be considered. Remaining variables were dropped from further consideration if they were redundant (derived from the same ancillary data set), highly correlated with one another (correlation coefficient >0.75), or dependent (as indicated by a condition index near 30 and variance decomposition proportions >0.5) (Belsley et al., 1980). In these cases, the variable with the lowest p value was retained, and the redundant, correlated, or dependent variables were eliminated from further consideration.

Optimal model formulations were identified by maximizing pseudo coefficient of determination ($p^2$) values (Laitila, 1993) and minimizing the scale parameter. The $p^2$ and scale parameter are alternatives to the $R^2$ and the standard deviation of the residual error, respectively, which are used in conventional least-squares analysis. As with conventional $R^2$, the $p^2$ ranges from 0 to 1 and is an estimate of the percentage of the variation in the response variable that is explained by the regression model. Thus, a $p^2$ value of zero indicates that no variation is explained, and a value of unity indicates that 100% of the variation is explained. The $p^2$ was calculated using methods described by Laitila (1993). Estimates of the standard deviation of the residual error from maximum-likelihood methods provide only asymptotically unbiased estimates of the standard deviation of the residual error when estimated from sample data (Aitkin, 1981). These estimates, on average, underestimate the true standard deviation. The bias is a function of the sample size and degree of censoring and is expected to be minimal for models with low percentages of censored observations. In this report, biased estimates of the standard deviation of the residual error are referred to as “scale.” Before final model formulations were established, all explanatory variables that had previously been dropped were reevaluated to confirm that they did not contribute significantly to the final model (i.e., that they were associated with a p value >0.05 and/or an increase in model $p^2$ of <1%). Finally, each pairwise combination of explanatory variables in the final models was considered as a cross-product term to assess whether interaction of explanatory variables contributed significantly to the final models.

Calculation of Prediction Intervals and Probabilities of Exceedance

Prediction intervals (PIs), which are a function of the fit of a model and the amount of variability explained by the model, represent the likelihood that a particular value for an individual site falls within a specified interval of the predicted value. The confidence level used for PIs in this study was 95%. Prediction intervals can be used to estimate the probability that the actual concentration of atrazine residues at a site exceeds a specified concentration, as described in Nowell et al. (2009).

A PI is computed as follows:

$$PI = \left[ \log_{10}(C_{est}) \pm SE_{est} \times t_{\alpha/2} \right]$$

where $C_{est}$ is the estimated atrazine residue concentration, $SE_{est}$ is the standard error of $\log_{10}(C_{est})$, and $t_{\alpha/2}$ is the point on the Student’s $t$ distribution with a probability of exceedance of one half of $\alpha$. (For the purpose of this study, $\alpha = 0.05$.)

As described previously, standard errors were estimated from the scale parameter through an adjustment suggested by Aitkin (1981). The effect of censoring of atrazine residue concentrations cannot be fully accounted for; therefore, PIs calculated for this study represent approximate values.

The probability ($P_{exc}$) that the actual concentration at a site will exceed a specified concentration ($C$) is approximated as follows:

$$P_{exc} = \frac{\left| \log_{10}(C) - \log_{10}(C_{est}) \right|}{SE_{est}}$$

Table 2. Loadings on the first two principal components from principal component analysis.

<table>
<thead>
<tr>
<th>Constituent or physical property</th>
<th>Principal component interpretation (percentage of variance explained)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component 1: Ionic strength (32%)</td>
<td>Component 2: Residence time indicator (20%)</td>
</tr>
<tr>
<td>Specific conductance</td>
<td>0.92†</td>
</tr>
<tr>
<td>Dissolved oxygen</td>
<td>–0.79</td>
</tr>
<tr>
<td>pH</td>
<td>0.38</td>
</tr>
<tr>
<td>NH₄</td>
<td>–0.27</td>
</tr>
<tr>
<td>NO₃⁻</td>
<td>0.68</td>
</tr>
<tr>
<td>P</td>
<td>–0.74</td>
</tr>
<tr>
<td>Ca²⁺</td>
<td>0.73</td>
</tr>
<tr>
<td>Mg²⁺</td>
<td>0.76</td>
</tr>
<tr>
<td>Na⁺</td>
<td>0.76</td>
</tr>
<tr>
<td>K</td>
<td>0.57</td>
</tr>
<tr>
<td>Cl⁻</td>
<td>0.73</td>
</tr>
<tr>
<td>SO₄²⁻</td>
<td>0.83</td>
</tr>
<tr>
<td>Silica</td>
<td>0.44</td>
</tr>
<tr>
<td>Fe³⁺</td>
<td>0.79</td>
</tr>
<tr>
<td>Mn²⁺</td>
<td>0.81</td>
</tr>
</tbody>
</table>

† Numbers in italics denote loadings with absolute values greater than 0.60. Loadings with absolute values less than 0.25 are omitted.
When the regression residual error is normally distributed, the error distribution about an estimated value of the dependent variable follows the Student's $t$ distribution. After antilogarithms have been computed, however, the error distribution around the regression line is no longer symmetric, and the variance is no longer constant across the full range of the independent variable. As a result, the magnitude of the uncertainty in $C_{est}$ is greater for values above the estimated concentration than for those below it and is greater for high concentrations than for lower concentrations.

Results

Many models could be created from subsets of the final set of explanatory variables (Table 1). The two optimal formulations that were obtained by maximizing $pR^2$ and minimizing scale, as described above, are presented in Table 3. One of the models was constructed using explanatory variables whose values are available nationwide (termed the "national-variable" model), and the second model was derived from a combination of nationwide and site-specific explanatory variables (termed the "site-variable" model). Inclusion of any of the previously discarded explanatory variables resulted in negligible improvement in the model fits, as indicated by an increase in $pR^2$ of <1%. Interaction between the variables $artdrn$ (artificial drainage) and $wtdep$ (depth to the seasonally high water table) was found to be significant in the national-variable model and thus required that both variables remain in the final model formulation (Kleinbaum and Klein, 2002), despite a $p$ value for $wtdep$ that exceeded 0.05 (Table 3).

The optimal model formulations explain 41% (national-variable model) and 50% (site-variable model) of the variability in the concentrations of atrazine residues measured in shallow groundwater beneath agricultural areas in this study (Table 3). Thus, the inclusion of site-specific variables increased the percentage of variance explained by about 22% (increased from 41 to 50%). The following sections examine the physical significance of each of the explanatory variables listed in Table 3.

Explanatory Variables

The 10 explanatory variables used in the two models (Table 3) are grouped below into three categories depending on whether they are associated with the sources, transport, or fate of atrazine and DEA in groundwater.

Source Variables

Three variables are related to amounts of atrazine applied within the 500-m radial areas around the sampled wells. Two of these variables are atrazine-use intensities ($use9200$ and $use97$ in Table 1), and the third represents the organic-matter (OM) content of the uppermost soil layer ($omup$ in Table 1), which, as explained below, has an influence on use.

The presence of atrazine in groundwater is assumed to result solely from its application for agricultural purposes because it is an anthropogenic compound used almost exclusively for agriculture in most parts of the United States and is thus unlikely to have come from any other source. Furthermore, although DEA may also be produced from the in situ dealkylation of the herbicides propazine and cyprazine, the use of these two compounds in the United States was negligible in the 1990s (Gianessi and Marcelli, 2000). Thus, the source of atrazine residues at each site of interest in this study was represented by the estimated intensity of atrazine use for agricultural purposes in that location. The intensity of atrazine use was calculated as the total amount of the herbicide applied to the crops of inter-

### Table 3. Statistics for Tobit regression models based on nationally available explanatory variables (National-variable model) and nationally available and site-specific explanatory variables (Site-variable model).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>$ p $ value</th>
<th>Contribution to overall $pR^2$† by addition of variable</th>
<th>$N$</th>
<th>Overall $pR^2$</th>
<th>Scale</th>
<th>Censored data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Left censored</td>
</tr>
<tr>
<td>$wtdep$</td>
<td>0.15</td>
<td>0.11</td>
<td>0.13</td>
<td>1298</td>
<td>0.41</td>
<td>0.82</td>
<td>51.6</td>
</tr>
<tr>
<td>$use9200^{1/3}$</td>
<td>0.46</td>
<td>&lt;0.0001</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$airtemp$</td>
<td>–0.05</td>
<td>&lt;0.0001</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$artdrn^{1/3}$</td>
<td>–0.1</td>
<td>&lt;0.0001</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$artdrn^{1/3} * wtdep$</td>
<td>0.04</td>
<td>0.0002</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$log_{10}(1 + recharge)$</td>
<td>0.62</td>
<td>&lt;0.0001</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$permin^{1/3}$</td>
<td>0.23</td>
<td>&lt;0.0001</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$intercept$</td>
<td>–3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$RTI$</td>
<td>–0.42</td>
<td>&lt;0.0001</td>
<td>0.25</td>
<td>1272</td>
<td>0.50</td>
<td>0.76</td>
<td>50.9</td>
</tr>
<tr>
<td>$log_{10}(1 + use97)$</td>
<td>0.69</td>
<td>&lt;0.0001</td>
<td>0.095</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$artdrn^{1/3}$</td>
<td>–0.06</td>
<td>&lt;0.0001</td>
<td>0.045</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$wtdep$</td>
<td>0.36</td>
<td>&lt;0.0001</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$log_{10}(1 + omup)$</td>
<td>1.18</td>
<td>&lt;0.0001</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$permin^{1/3}$</td>
<td>0.21</td>
<td>&lt;0.0001</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$log_{10}(1 + recharge)$</td>
<td>0.38</td>
<td>0.0002</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$log_{10}(1 + welldepth)$</td>
<td>–0.24</td>
<td>0.008</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$intercept$</td>
<td>–3.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† Pseudo coefficient of determination.
The greater explanatory power of pesticide-use intensity for stream concentrations compared with groundwater concentrations is consistent with the difference between contaminant transport and transformation processes in the two media. Surface waters tend to integrate water chemistry over space and time to a much greater extent than does groundwater because they exhibit relatively short flow paths, brief soil contact time, and rapid mixing. As a result, the concentrations of dissolved contaminants in surface waters respond rapidly to changing conditions within the watershed. In contrast, groundwater, which generally represents water entering the subsurface at a particular location and time (especially water from monitoring wells and other low-capacity wells), typically shows a more delayed response to changing conditions due to greater soil contact time as it percolates through the unsaturated zone to the water table and through the saturated zone to discharge areas (Kauffman et al., 2001). The longer residence times associated with groundwater systems than with surface runoff and the greater opportunity for physical, biological, and chemical effects make groundwater chemistry more difficult, compared with surface-water chemistry, to correlate with human activities (such as pesticide use) at land surface.

An unexpected observation from this study was that atrazine residues in groundwater were positively correlated with the OM content of the uppermost soil layer \( \text{omup} \) (Table 3). This is counter to the finding by other studies that lower amounts of soil OM tend to be associated with lower atrazine degradation rates (e.g., Fenner et al., 2007) and less sorption of organic compounds (Hamaker and Thompson, 1972), both of which should allow higher concentrations of atrazine residues to reach shallow groundwater. The intensity of atrazine use, however, depends in part on the soil characteristics of a given area. Loamy soils, which contain higher amounts of OM than sandy soils and are generally preferred for the cultivation of most crops, typically receive 25 to 50% greater applications of atrazine per unit area than sandy soils (Wisconsin Department of Agriculture, Trade and Consumer Protection, 2010). The significant correlation between \( \text{omup} \) and \( \text{use97} \) may therefore reflect an elevated use of atrazine in loamy areas (Spearman \( p = 0.39; p < 0.0001; n = 1298 \)), which may explain why \( \text{omup} \) is a significant explanatory variable in the site-variable model (Table 3). Inclusion of \( \text{omup} \) in the model appears to refine source estimates by identifying preferred areas for crop cultivation and indicating where atrazine might have been applied at higher rates. It is unlikely that OM content reflects preferential flow processes because it was found to be only weakly correlated with clay content (Spearman \( p = 0.17; p < 0.0001; n = 1298 \)), which has been found to be associated with structured soils and the rapid, downward transport of contaminants through preferential flow paths within the subsurface (e.g., Jarvis et al., 2009).

**Transport Variables**

Five variables are related to factors that may affect the transport of atrazine residues to the water table and to the sampled well. These are (i) prevalence of artificial drainage (\( \text{artdrn} \) in Table 1), (ii) depth of the seasonally high water table (\( \text{wtdep} \) in Table 1), (iii) rate of recharge (\( \text{recharge} \) in Table 1), (iv) permeability of the least permeable soil layer (\( \text{permin} \) in Table 1), and (v) depth of the well (\( \text{welldepth} \) in Table 1).

Artificial drainage (\( \text{artdrn} \)) and depth of the seasonally high water table (\( \text{wtdep} \)) interact significantly in the national-variable model to affect the concentrations of atrazine residues in shallow groundwater (Table 3). The water table beneath poorly drained agricultural soils tends to be close to the land surface during the growing season, often requiring the use of artificial drainage systems for reducing soil water content to facilitate agricultural activities. In this study, atrazine residue concentration was positively correlated with depth to the seasonally high water table and was inversely correlated with artificial drainage. This pattern reflects the efficacy of surface ditches and subsurface drains for capturing soil water and diverting it to streams rather than allowing it to enter the groundwater system. This diversion reduces the transport of atrazine to groundwater underlying the drained soils and thus may result in concentrations of atrazine residues in shallow groundwater beneath poorly drained areas that are lower than in well drained areas that do not require drainage systems; this pattern has been observed in previous studies (e.g., Keim et al., 1989; Kladiukvo et al., 2001). The significant interaction that was observed between \( \text{artdrn} \) and \( \text{wtdep} \) reflects the fact that artificial drainage systems are typically installed only in areas where soils are poorly drained.

The permeability of the least permeable soil layer, \( \text{permin} \), and the rate of recharge, \( \text{recharge} \), were positively correlated with atrazine residue concentrations (Table 3). This indicates that, as expected, high soil permeability and large water flux through the unsaturated zone favor the transport of atrazine residues to the water table.

The depth of wells, \( \text{welldepth} \), was inversely correlated with atrazine residue concentration (Table 3), indicating that shallower wells tended to have higher concentrations than deeper wells. Because the wells used in this study were selected or installed with the goal of providing samples of “recently recharged” (shallow, young) groundwater, the screened or open interval in most of the wells is located a short distance below...
the water table. As a result, the median distance from the water table to the bottom of the wells was relatively shallow (4.6 m). Given the fact that the wells were generally installed a short distance below the water table, well depth served as a surrogate for, and was significantly correlated with, unsaturated-zone thickness (Spearman \( p = 0.80; p < 0.0001; n = 1267 \)).

The thickness of the unsaturated zone is directly related to the amount of time available for atrazine residue concentrations to be diminished through sorption or degradation (Barbash and Resek, 1996; Tesoriero et al., 2007). The inverse correlation observed between atrazine residue concentrations and well depth during this study thus appears to support the hypothesis that the concentrations of atrazine and DEA are likely to diminish with increasing residence time within the unsaturated zone.

**Fate Variables**

Two variables reflect influences of fate processes in the subsurface: (i) residence-time indicator (RTI) in Table 1), and (ii) mean annual air temperature (airtemp in Table 1).

Of the explanatory variables considered in the present study, RTI is by far the most powerful single variable for predicting the concentration of atrazine residues in groundwater beneath agricultural areas. The incremental inclusion of this variable in the site-variable model accounted for 25% of the variability in atrazine residue concentrations, representing half of the total variability in atrazine residue concentrations accounted for by the full site-variable model (50%) (Table 3). The variable RTI was developed through a PCA of water quality data (Table 2) and provides an indication of the effect of time on geochemical conditions along groundwater flowpaths in the saturated zone. The PC identified as RTI was significantly correlated with the concentrations of several redox-active solutes—negatively with NO\(_{-}\) and dissolved oxygen (DO) and positively with Fe\(^{2+}\), Mn\(^{2+}\), and NH\(_{3}\) (Table 2). Negative RTI values thus represent oxic conditions, and positive values represent reducing conditions. The relation to time derives from the gradual depletion of DO through microbial respiration, the decrease of NO\(_{-}\) through denitrification, and the increase in concentration of reduced species (i.e., Fe\(^{2+}\), Mn\(^{2+}\), and NH\(_{3}\)) as groundwater becomes more reducing with increasing residence time in the saturated zone.

Although RTI was derived from redox-active solutes, it does not appear to reflect the effect of redox conditions on the persistence of atrazine residues. Results from several laboratory investigations indicate that atrazine is less persistent under oxic conditions than under reducing conditions (e.g., Kaufman and Kearney, 1970; Nair and Schnoor, 1992; Tesoriero et al., 1999; Larsen et al., 2000; Accinelli et al., 2001), although field evidence has provided mixed results (Barbash and Resek, 1996).

For example, Druliner (1989) observed an inverse correlation between atrazine and DO concentrations in groundwater from the High Plains aquifer in Nebraska, which is consistent with the above-cited laboratory study results. Similarly, atrazine has been reported to be relatively unreactive under reducing conditions in situ (e.g., Spillmann, 1989; Papiernik and Spalding, 1998; Rügge et al., 1999). In contrast, four national or regional monitoring programs have reported positive correlations between atrazine and DO concentrations (or other measures of oxidizing conditions) in groundwater (Kolpin et al., 1997, 2002; Ator, 2008; Saad, 2008), as was observed in this study. An explanation for this disparity was provided by Kolpin et al. (1997, 2002), who suggested that, within the saturated zone, DO concentration is a surrogate for residence time because oxygen tends to be consumed through biotic and abiotic processes as water moves downgradient through the saturated zone. Thus, a positive correlation between atrazine concentrations and DO concentrations reflects a simultaneous loss of both solutes as water moves downgradient, rather than a causal relation between redox condition and atrazine concentration. Kolpin et al. (1997, 2002) were unable to test this hypothesis, however, because estimates of groundwater age were not available at the time of their study. More recent studies that have found statistically significant correlations between DO concentrations in groundwater and residence time in the saturated zone have also observed inverse relations between the two variables (Tesoriero et al., 2005, 2007; Saad, 2008; Denver et al., 2010). This is consistent with the hypothesis that, within the saturated zone, DO concentration may often be viewed as a surrogate for residence time. The present study obtained estimates of groundwater age for 259 samples through an analysis of the concentrations of chlorofluorocarbons, SF\(_{6}\), and/or \(^{3}H/\(^{3}He\) (Hinkle et al., 2010) (246 of these samples were in the model-development data set; the remaining 13 were not used for model development because they lacked values for one or more of the explanatory variables). These age estimates were found to be positively correlated with RTI values (Spearman \( p = 0.46; p < 0.0001; n = 259 \)). The correlation between RTI and atrazine residue concentrations therefore appears to be primarily a reflection of the influence of residence time in the saturated zone on atrazine and DEA concentrations rather than the effect of redox conditions on the persistence of the two solutes.

The explanatory power of the RTI appears to derive from the fact that the residence time of sampled groundwater represents the combined influence of factors that can affect the concentrations of atrazine residues in groundwater through time. Another time-related factor that has been shown to be correlated with the concentrations of atrazine residues in groundwater is historical changes in atrazine use intensity (Tesoriero et al., 2007). The relation between atrazine residue concentration and the estimated year in which the sampled groundwater recharged the groundwater system is depicted in Fig. 2, along with boxplots showing the range and distribution of atrazine concentrations for each decade from 1950 through 2000. A bivariate Tobit model indicated a statistically significant relation between atrazine residue concentration and the estimated year of recharge (\( pR^2 = 0.15; p < 0.0001; \) scale = 0.93; \( n = 259 \)), suggesting that historical changes in atrazine use intensity may explain some of the lower concentrations of atrazine residues in older samples, relative to those in younger samples (Fig. 2). To evaluate the possibility that factors other than historical changes in atrazine use intensity (such as degradation, adsorption, and dispersion) diminish the concentration of atrazine residues through time, a second bivariate Tobit model was developed from only those samples of groundwater that entered the system between 1975 and 2000 (\( n = 208 \)), when national atrazine use was relatively stable, varying between 30 and 40 million kg \( \text{yr}^{-1} \) (Fig. 2). This model was significant at the 95% confidence level (\( p = 0.05 \)) but exhibited a weak correlation (\( pR^2 = 0.02 \)).
revealing little evidence that atrazine residue concentrations are diminished in shallow groundwater by processes occurring in the saturated zone such as degradation, adsorption, or dispersion. This conclusion is consistent with the findings from several laboratory studies that have observed negligible rates of atrazine mineralization in saturated, oxic aquifer sediments (e.g., Konopka and Turco, 1991; McMahon et al., 1992). Similarly, the mole fraction of atrazine plus DEA that was present as DEA did not vary as a function of groundwater age, in agreement with Tesoriero et al. (2007). These findings indicate that, at the national scale, residence time in the saturated zone is predictive of atrazine concentrations in groundwater primarily because it reflects temporal changes in the concentration of atrazine residues in recharge rather than the influence of redox controls on the persistence of these compounds or the effects of processes such as degradation, adsorption, or dispersion within the saturated zone.

Mean annual air temperature (airtemp) during 1980–1997 was inversely correlated with the concentration of atrazine residues (Table 3). Air temperature has been shown to correlate with soil temperature at the national scale (Toy et al., 1978; Zheng et al., 1993), indicating that airtemp may be viewed as a surrogate for soil temperature. As with all nonphotochemical reactions, the rate of atrazine transformation in soil decreases with decreasing temperature (e.g., Rocha and Walker, 1995), as does volatilization from soil (Thomas, 1990). Consequently, if all other factors are held constant, atrazine and DEA are more likely to persist long enough to reach groundwater if the atrazine is applied in cool climates or at cool times of the year than if the herbicide is applied in warm climates or during warm seasons. Several authors have suggested that the effect of temperature may be reinforced by the effects of recharge because cold seasons often coincide with periods of greater precipitation than warm seasons (e.g., Wyman et al., 1985; Jones et al., 1986; Priddle et al., 1989, 1992; Pickett et al., 1990; Porter et al., 1990; Traub-Eberhard et al., 1994). However, the present study did not observe a statistically significant interaction between airtemp and recharge (p = 0.94), indicating that the effects of temperature on atrazine residue concentrations in groundwater are independent of the amount of recharge received.

Model Performance and Evaluation

Residual errors were calculated by subtracting the logarithm of the predicted atrazine residue concentration for each site from the logarithm of the observed value. A residual error of less than zero thus indicated overprediction, and a residual error greater than zero indicated underprediction. Residual values of −1 and +1 indicate predicted concentrations that are 10 times and one tenth of the observed concentration, respectively. For a prediction with a censored observed concentration, the residual was computed in the following manner (Gregory Schwarz, USGS, written communication, 2010):

$$\epsilon_i = \sigma \Phi^{-1} \left[ \mu \frac{\log(\text{LTMDL}_{i,j}) - \log(C_{\text{est}})}{\sigma} \right]$$

where $\sigma$ is the scale parameter estimated from the Tobit model, $\Phi^{-1}$ is the inverse of the cumulative normal integral, $\mu$ is a random number between zero and unity drawn from a uniform (equiprobable) distribution, and LTMDL and $C_{\text{est}}$ are as defined earlier. The values of $\epsilon_i$ are randomly generated, normally distributed residuals that adhere to the lower and upper bounds of the censored observation and conform with the distribution of the residuals used to obtain the Tobit model estimates. These residuals were added to the predicted values for censored samples to obtain estimates of the observed concentrations for assessing model performance.

Model Performance

When all sites were examined, both models were found to be unbiased, with residuals that were normally distributed about zero across all predicted concentration values (Fig. 3A, 3B). Boxplots of the residuals from both models (Fig. 3C, 3D) suggest a similar conclusion when all sites were considered, but both models tended to underpredict concentrations at sites where neither atrazine nor DEA were detected (left-censored samples). Residuals from these sites were based on synthetic estimates rather than measured concentrations. At sites where both atrazine and DEA were detected (uncensored samples), both models tended to underpredict concentrations (Fig. 3C, 3D). Observed atrazine residue concentrations are plotted in relation to predicted concentrations in Fig. 3E and 3F, which indicate that 77% of the concentrations predicted by the national-variable model (Fig. 3E) and 81% of those predicted...
by the site-variable model (Fig. 3F) were within a factor of 10 of the observed values.

Model Evaluation

The models were applied to the independent set of 209 agricultural sites described earlier (Fig. 1), and model predictions and residuals were computed. Residuals for these sites were normally distributed but, in general, are biased toward overprediction (Fig. 4A–4D). The bias was primarily caused by overprediction of left-censored samples for which synthetic residuals were calculated (as noted earlier). The source of this overprediction is unknown but is not likely to have been overestimates in atrazine use because atrazine use appears to have been relatively stable from 1992 to 2006 (Fig. 2).

As was seen among the model development sites (Fig. 3C, 3D), for locations where both atrazine and DEA were detected (uncensored samples), the model tended to underpredict the atrazine residue concentrations at the model evaluation sites. However, the degree of underprediction (Fig. 4C, 4D) was less than that observed for the model-development sites (Fig. 3C, 3D). Overall, more than 70% of the predicted concentrations for either model were within a factor of 10 of the observed values at the model-evaluation sites (Fig. 4E, 4F).

Uncertainty in Model Predictions

Uncertainty in a predicted concentration is expressed in terms of the PI (Eq. [1]) in Fig. 5, which shows the atrazine residue concentrations for the model-development sites ordered by increasing predicted concentration. Comparison of Fig. 5A and 5B indicates that the PIs are larger for the national-variable model than they are for the site-variable model. This was expected because PIs are a function of the amount of variability explained by each model, and the site-variable model explained a higher proportion of variability in the observed atrazine residue concentrations than did the national-variable model (Table 3). The levels of uncertainty in model predictions are also compared in Fig. 6, in which the size of the PI at each site is represented as the ratio of the upper PI to the predicted atrazine residue concentration (which is the same as the ratio of the predicted atrazine residue concentration to the lower PI). Outliers in Fig. 6 represent sites where one or more explanatory variables were relatively extreme in value when compared with values from the rest of the sites. These outlier sites, therefore, exhibited greater uncertainty in their predicted atrazine residue concentrations than did the others. Figure 6 also indicates that, as anticipated from the results presented in Table 3 and Fig. 5, predictions from the national-variable model exhibited greater uncertainty than those from the site-variable model.

Model Limitations

The amount of variability in atrazine residue concentrations explained by the optimal Tobit model that used the site-specific variables was substantial. However, the fact that 50% of the observed variability in these concentrations remains unexplained (Table 3) indicates that important factors related to the sources, transport, or fate of atrazine may not be well represented by the explanatory variables for which data are available on a national basis. In addition, the national scale at which these models were developed could result in erro-
neous predictions if the models are applied at more local scales because they may not adequately reflect local variations in some of the explanatory variables that strongly affect atrazine residue concentrations in shallow groundwater. Local factors that may affect the transport of atrazine and DEA to wells require consideration before results from these models can be interpreted for any particular location. Also, the application of the models to sites with fundamental differences in hydrogeologic conditions or to sites whose values for explanatory variables differ from those used for model development will result in greater uncertainty in model predictions.

In this study, atrazine use estimates were for agricultural purposes only because estimates for nonagricultural use of pesticides are not available in most states. Although atrazine is known to have nonagricultural uses, they are considered to be minor relative to agricultural uses, especially in agricultural settings. Uncertainty arises, however, when examining correlations between atrazine residue concentrations in groundwater and estimates of atrazine use that were averaged for periods during which use was known to have changed, particularly for sites where groundwater age estimates are not available.

The observed uncertainty in the predictions from the models constructed for this study, as characterized in the previous section, indicate that these models should not be used to characterize absolute concentrations at specific locations. Instead, the models should be used for identifying areas where atrazine residues are likely to be highest and may therefore merit further investigation.

**Model Applications**

The national- and site-variable models (Table 3) were used to make two types of predictions for agricultural areas throughout the conterminous United States. The national-variable model was applied to predict concentrations in groundwater with a range of characteristics and residence times similar to those in the model-development data set. By contrast, the site-variable model was used to predict concentrations in recently recharged groundwater by setting the $RTI$ variable to the low end of its range. The models were applied by quantifying explanatory variables for every 1-km$^2$ grid cell of a gridded map of the conterminous United States in which agricultural land represented at least 50% of the total area. Only 4% of these gridded areas had values for one or more explanatory variables that were outside the range of values used in the model development dataset. This observation illustrates the high degree to which the groundwater quality assessment conducted by the NAWQA Program encompassed conditions that were representative of groundwater settings across the country. Virtually all of the explanatory variable values that were outside the model development dataset range were mean annual air temperature ($\text{airtemp}$) values—primarily in small areas of northern North Dakota and northwestern Minnesota where $\text{airtemp}$ values are lower than the minimum value used for model development. Because only a small portion of the agricultural areas in the United States was affected, the overall effect of the $\text{airtemp}$ outliers on the model estimates described in the following section was likely to have been minimal.

![Fig. 4. Distribution of residuals and observed and predicted atrazine residue concentrations for model-evaluation sites. All, all samples; IC, interval censored; LC, left censored; UC, uncensored.](image-url)
The national-variable model predicted atrazine residue concentrations in groundwater to be less than the sum of their LTMDL values ($0.011 \mu g \, L^{-1}$) in 53% of the grid-ded agricultural areas across the United States (Fig. 7A). The highest predicted concentrations were in parts of the High Plains aquifer system in Nebraska, Kansas, and Texas; the Driftless Area in Illinois, Iowa, Minnesota, and Wisconsin; and in southeastern Pennsylvania (Fig. 7A). Although these are areas where atrazine is known to be used, they do not generally represent the locations with the most intensive use of atrazine (Fig. 1). This suggests that, as noted previously by Barbash et al. (1999), Kolpin et al. (2002), and Stackelberg et al. (2006a), factors other than application intensity appear to exert a substantial effect on the distributions and concentrations of atrazine residues in shallow groundwaters beneath agricultural areas. Such factors include soil conditions, climate, recharge rates, and agricultural management practices. For example, some of the most intensive atrazine use during the study period was in parts of Iowa, Illinois, and Indiana, where soils tend to be poorly drained and where agricultural fields commonly have subsurface tile drains that capture soil water and divert it to nearby streams. Consequently, the concentrations of atrazine residues in groundwater would be expected to be lower in these locations than in areas with conditions that did not require tile drains, as suggested by previous researchers (e.g., Keim et al., 1989; Kladivko et al., 2001). In contrast, the national-variable model (Table 3) predicts that areas with extensive atrazine use on soils that do not require artificial drainage (e.g., parts of the High Plains aquifer and Driftless Area and southeastern Pennsylvania) would have the highest concentrations of atrazine residues in shallow groundwater (Fig. 7A).

The national-variable model predicted atrazine residue concentrations in groundwater to be less than the sum of their LTMDL values ($0.011 \mu g \, L^{-1}$) in 53% of the grid-ded agricultural areas across the United States (Fig. 7A). The highest predicted concentrations were in parts of the High Plains aquifer system in Nebraska, Kansas, and Texas; the Driftless Area in Illinois, Iowa, Minnesota, and Wisconsin; and in southeastern Pennsylvania (Fig. 7A). Although these are areas where atrazine is known to be used, they do not generally represent the locations with the most intensive use of atrazine (Fig. 1). This suggests that, as noted previously by Barbash et al. (1999), Kolpin et al. (2002), and Stackelberg et al. (2006a), factors other than application intensity appear to exert a substantial effect on the distributions and concentrations of atrazine residues in shallow groundwaters beneath agricultural areas. Such factors include soil conditions, climate, recharge rates, and agricultural management practices. For example, some of the most intensive atrazine use during the study period was in parts of Iowa, Illinois, and Indiana, where soils tend to be poorly drained and where agricultural fields commonly have subsurface tile drains that capture soil water and divert it to nearby streams. Consequently, the concentrations of atrazine residues in groundwater would be expected to be lower in these locations than in areas with conditions that did not require tile drains, as suggested by previous researchers (e.g., Keim et al., 1989; Kladivko et al., 2001). In contrast, the national-variable model (Table 3) predicts that areas with extensive atrazine use on soils that do not require artificial drainage (e.g., parts of the High Plains aquifer and Driftless Area and southeastern Pennsylvania) would have the highest concentrations of atrazine residues in shallow groundwater (Fig. 7A).

Atrazine-residue concentrations predicted by the national-variable model were expected to be lower than the actual concentrations in recently recharged, young groundwater because some of the samples used for model development had a decades-long residence time that predated the period of most intensive atrazine use, which began in the mid-1970s (Fig. 2). Predictions from the site-variable model cannot be obtained at unmonitored locations without assuming values for RTI and wellsdepth because national datasets for these variables are not available. To address this limitation, estimates of the concentration of atrazine residues in recently recharged groundwater were obtained using specific values for RTI and wellsdepth in the site-variable model. As previously described, RTI provides an indication of residence time in the saturated zone, where low values represent oxic, young groundwater and higher values represent older, more reduced water. Assigning the fifth percentile of RTI and the median depth of wells (wellsdepth) to all areas made it possible to use the site-variable model to predict atrazine concentrations for young, oxic groundwater drawn from shallow wells. The resulting atrazine residue concentrations in groundwater were lower than the sum of their LTMDLs ($0.011 \mu g \, L^{-1}$) in 10% of the agricultural areas across the United States (Fig. 7B), a percentage that was considerably lower than the estimate of 53% for the national-variable model. Predictions from the site-variable model are appropriate for comparison with those from process-based models of atrazine transport through the unsaturated zone that are not calibrated to field observations (e.g., Goss et al., 1998; Barbash and Voss, 2007). Such predictions are also useful for providing initial atrazine residue concentrations at the water table for models of atrazine transport and fate in the saturated zone. In the absence of a predictive model developed from supply-well data, predictions from the site-variable model may also provide conservative estimates of the potential concentrations of atrazine residues in deeper groundwater supplies that may be used for drinking. This is supported by the finding, noted here and by previous studies, that processes such as degradation, adsorption, or dispersion in the saturated zone do not appear to cause substantial decreases in atrazine residue concentrations over
time in groundwater underlying agricultural settings.

**Predicting the Probability of Exceeding a Specific Concentration**

The national- and site-variable models were also used to predict the probability that atrazine residues might exceed a specified concentration in groundwater beneath agricultural areas across the conterminous United States. The concentration thresholds considered were 3.0 and 0.3 $\mu$g L$^{-1}$. The 3.0 $\mu$g L$^{-1}$ value was selected as an exceedance threshold because it represents the maximum contaminant level established by the USEPA for atrazine in drinking water (USEPA, 2003). The 0.3 $\mu$g L$^{-1}$ value was selected to identify areas that may warrant further evaluation if groundwater is to be used as a drinking water supply. The USEPA has not established a regulatory concentration for DEA in drinking water supplies but has ascertained that atrazine and DEA share a common mechanism of toxicity and that their binary mixture therefore would have additive toxicity (USEPA, 2006).

Results from this analysis using the national-variable model indicate that atrazine residue concentrations in groundwater with characteristics and residence times similar to those examined for the model-development data set have greater than a 10% probability of exceeding the 0.3 and 3.0 $\mu$g L$^{-1}$ thresholds in 25% (Fig. 7C) and 2% (Fig. 8A) of the agricultural areas, respectively. In contrast, the results from this analysis using the site-variable model indicate that atrazine residue concentrations in recently recharged, young groundwater have a considerably greater probability of exceeding these thresholds: 60% of agricultural areas are predicted to have a greater than 10% probability of exceeding 0.3 $\mu$g L$^{-1}$ in 16 and 9% of agricultural areas, respectively. In contrast, 12% and 0.4% of agricultural areas had a greater than 10% probability of exceeding 0.3 $\mu$g L$^{-1}$ when using the 10th and 90th percentile values, respectively, for RTI.

Viewed with respect to the overall ranges spanned by these two parameters, predictions from the site-variable model are thus more sensitive to variations in RTI than they are to variations in weldepth.

**Conclusions**

Tobit regression models developed from explanatory variables whose values are available nationwide (the national-variable model) or from a combination of nationwide and site-specific explanatory variables whose values are available
only for the locations at which they were measured (the site-variable model) were found to explain up to half of the variability observed in the concentrations of atrazine and DEA in groundwater beneath agricultural areas across the conterminous United States. Explanatory variables that were found to be statistically significant were related to three factors that control atrazine residue (atrazine plus DEA) concentrations in groundwater: atrazine sources, transport processes, and degradation. The inclusion of explanatory variables derived from site-specific information (such as water chemistry data) improved predictive capability and suggested that, at the national scale, historical changes in use intensity have a greater effect on the distribution of atrazine residue concentrations in groundwater underlying agricultural settings than do the processes that control their dispersion, sorption, or degradation in the saturated zone. Results from this study indicate that the highest atrazine residue concentrations in groundwater underlying agricultural areas are likely to be in parts of the High Plains aquifer system, the Driftless Area of Illinois, Iowa, Wisconsin, and Minnesota and southeastern Pennsylvania. The degree of explanatory power exhibited by these newly developed models is notable, given the inherent difficulties associated with correlating the concentrations of pesticide compounds in a single groundwater sample with the land use, hydrologic, physical, biological, and chemical factors that can affect the concentrations of the compounds between their point of entry at the land surface (or formation in the subsurface) and the screened or open interval of the well from which they may be extracted.

These limitations notwithstanding, the Tobit models described here represent an approach that provides the most reliable predictions of atrazine residue concentrations in shallow agricultural groundwater across the United States. Predicted concentrations in individual wells are generally of greater value than network detection frequencies (Kolpin et al., 2002; Stackelberg et al., 2006a) because they can be directly compared with water quality benchmarks or with health-based screening levels (Tothcalino and Norman, 2006).

The site-variable model allowed predictions to focus on recently recharged (young) groundwater that is representative of water quality conditions near the water table. Such predictions may be useful for comparison with similar predictions from process-based models of transport through the unsaturated zone that are not calibrated to field observations and for providing initial concentration estimates at the water table for models of atrazine transport through the saturated zone. By focusing on recently recharged groundwater, the predictions from the site-variable model provide conservative estimates of the concentrations of atrazine residues in deeper portions of the aquifer that are being used, or considered, for drinking water supplies. Such comparisons can provide an early indication to water resource managers of whether or not pesticide concentrations in source waters in certain areas may warrant further study. The Tobit models can also be used to identify areas where the concentrations of atrazine residues, or their probability of exceeding specified concentrations, are likely to be relatively high or low. Such predictions can be used for the design of future monitoring programs or for setting priorities among different areas for further research.

References


Therneau, T.M. 1999. A package of survival routines for S. Mayo Clinic Department of Health Sciences Research, Rochester, MN.


TIBCO Software Inc. 2008. TIBCO Spotfire S+ 8.1 for Windows user’s guide. TIBCO, Palo Alto, CA.


