SETAC PRESS

Environmental Toxicology and Chemistry, Vol. 28, No. 6, pp. 1346–1358, 2009

Printed in the USA
0730-7268/09 $12.00 + .00

REGRESSION MODELS FOR EXPLAINING AND PREDICTING CONCENTRATIONS OF ORGANOCHLORINE PESTICIDES IN FISH FROM STREAMS IN THE UNITED STATES

LISA H. NOWELL,*† CHARLES G. CRAWFORD,‡ ROBERT J. GILLIOM,† NAOMI NAKAGAKI,† WESLEY W. STONE,‡ GAIL P. THELIN,† and DAVID M. WOLOCK§

†U.S. Geological Survey, Placer Hall, 6000 J Street, Sacramento, California 95819
‡U.S. Geological Survey, 5957 Lakeside Boulevard, Indianapolis, Indiana 46278
§U.S. Geological Survey, 4821 Quail Crest Place, Lawrence, Kansas 66049

(*Received* 6 *October* 2008; *Accepted* 6 *January* 2009)

Abstract—Empirical regression models were developed for estimating concentrations of dieldrin, total chlordane, and total DDT in whole fish from U.S. streams. Models were based on pesticide concentrations measured in whole fish at 648 stream sites nationwide (1992–2001) as part of the U.S. Geological Survey's National Water Quality Assessment Program. Explanatory variables included fish lipid content, estimates (or surrogates) representing historical agricultural and urban sources, watershed characteristics, and geographic location. Models were developed using Tobit regression methods appropriate for data with censoring. Typically, the models explain approximately 50 to 70% of the variability in pesticide concentrations measured in whole fish. The models were used to predict pesticide concentrations in whole fish for streams nationwide using the U.S. Environmental Protection Agency's River Reach File 1 and to estimate the probability that whole-fish concentrations exceed benchmarks for protection of fish-eating wildlife. Predicted concentrations were highest for dieldrin in the Corn Belt, Texas, and scattered urban areas; for total chlordane in the Corn Belt, Texas, the Southeast, and urbanized Northeast; and for total DDT in the Southeast, Texas, California, and urban areas nationwide. The probability of exceeding wildlife benchmarks for dieldrin and chlordane was predicted to be low for most U.S. streams. The probability of exceeding wildlife benchmarks for total DDT is higher but varies depending on the fish taxon and on the benchmark used. Because the models in the present study are based on fish data collected during the 1990s and organochlorine pesticide residues in the environment continue to decline decades after their uses were discontinued, these models may overestimate present-day pesticide concentrations in fish.

Keywords—Organochlorine pesticides Models Fish tissue Pesticide use Water quality

INTRODUCTION

Environmental residues of organochlorine pesticides such as DDT, chlordane, and dieldrin have declined nationally in whole freshwater fish from U.S. rivers [1] ([2]; http://pubs.usgs.gov/circ/2005/1291) since uses of these pesticides were discontinued decades ago [3]. For total DDT (the sum of DDT plus its degradates), the steepest declines occurred during the 1970s, followed by a gradual leveling off. Similar trends were observed for predator fish from the Great Lakes [4] and in sediment cores from lakes and reservoirs [5]. These trends are consistent with historical DDT use in the United States, which peaked during the 1960s and was discontinued in 1972 [6]. For chlordane and dieldrin in fish, declines were more recent (late 1970s to 1980s) and gradual [2], and chlordane trends in lake sediment cores were variable [5]—consistent with the regulatory history of these pesticides. Agricultural uses of chlordane, dieldrin, and aldrin (which degrades to dieldrin in the environment) were cancelled during the early 1970s, but their use for termite control continued through the late 1980s [3,7] ([8]; http://www.epa.gov/greatlakes/bns/pesticides/finalpestreport.html). As discussed for lake sediment cores ([9]; http://pubs.usgs.gov/circ/2005/1291), the apparent exponential decline in organochlorine pesticide concentrations in fish does not represent a specific fate process but reflects the rate of change over time as a consequence of reduced input, chemical transformation, and dilution. The slow rate of recent decline in organochlorine pesticides residues in fish suggests the continued existence of sources, such as contaminated soil transported to streams by runoff [6,10,11].

The legacy organochlorine pesticides had their registrations cancelled because of their persistence, tendency to bioaccumulate, carcinogenicity, and hazard to wildlife [3,12]. They also induce the monooxygenase enzymes [13] and have been associated with fish diseases [14], endocrine and reproductive changes [15], and immunosuppressive effects [16]. Despite nationally declining trends, there is potential for remaining organochlorine residues to adversely affect aquatic life and wildlife, especially given the uncertainty in thresholds for adverse effects [17,18]. Organochlorine pesticides were widely detected in fish and sediment from U.S. streams during the 1990s, and concentrations in some areas were high compared with environmental benchmarks [2] (*Supporting Information* Table S1; http://dx.doi.org/10.1897/08-508.S1). Over 80% of whole-fish samples and over 50% of bed sediment samples collected by the U.S. Geological Survey (USGS) National Water Quality Assessment (NAWQA) Program from 1992 to 2001 contained organochlorine pesticide residues. Moreover, concentrations in whole fish exceeded benchmarks for the protection of fish-eating wildlife in approximately 20 to 75% of streams, depending on which wildlife benchmarks were used. Most exceedances were caused by total DDT (*Supporting Information* Table S1; http://dx.doi.org/10.1897/08-508.S1).

The continued occurrence of organochlorine pesticides in

* To whom correspondence may be addressed (lhnowell@usgs.gov).
Published on the Web 2/12/2009.
Any use of trade, firm, or industry names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

streams, combined with potential for adverse effects, makes it important to assess present-day exposure and evaluate concentrations in relation to levels of potential concern. However, the high cost of sampling and analysis is a deterrent to needed studies and monitoring, especially as organochlorine pesticide residues continue to decline nationally and other priorities in contaminant monitoring and assessment compete for available resources.

The use of models to predict pesticide concentrations offers an alternative to monitoring for broad-scale assessment and a guide for selective monitoring. Previous studies have investigated the relationship between land use and the presence of organochlorine pesticides in sediment or biota [19–21]. Several studies used regression models to predict concentrations of dissolved pesticides in streams from watershed characteristics, estimates of past use, or both [22–24] ([25]; http://pubs.usgs.gov/wri/wri034047). A previous report from the present study describes a preliminary regression model for dieldrin in whole fish from U.S. streams ([26]; http://pubs.usgs.gov/sir/2006/5020).

The purpose of the present study is to describe the development and application of regression models for predicting concentrations of dieldrin, total chlordane, and total DDT in whole fish from U.S. streams. The specific objectives are to evaluate relationships between concentrations in fish tissue and various explanatory factors, such as historical use patterns, watershed characteristics, and geographic location; derive best-fit regression models to quantify the relationships; and apply the regression models to produce a national assessment of the predicted distributions of pesticide concentrations in whole fish and the areas of potential concern that merit more detailed examination.

MATERIALS AND METHODS

Regression models were developed from data on organochlorine pesticide concentrations in fish collected from 1992 to 2001 as part of the NAWQA Program, combined with nationally available data on potential sources and watershed characteristics. Detailed information is provided elsewhere on pesticide sampling methods [2] and model-development procedures [26].

The present study focuses on the three most commonly detected pesticides or pesticide groups, which include degradates and by-products as well as parent pesticides, in fish. The most commonly detected was total *p,p′*-DDX, defined as the sum of *p,p′* isomers of DDT and its degradates dichlorodiphenyldichloroethylene (DDE) and dichlorodiphenyldichloroethane (DDD); one or more components of *p,p′*-DDX was detected in approximately 80% of fish samples in the model-development data set. The next most commonly detected were total chlordane (the sum of *cis*-chlordane, *trans*-chlordane, *cis*-nonachlor, and *trans*-nonachlor) and dieldrin, which were detected in 47 and 36% of fish samples, respectively. Use of these three response variables in regression analysis provides the most complete accounting of residues from the application of DDT and tetrachlorodiphenylethane (TDE; also known as DDD) combined; technical chlordane; and aldrin and dieldrin combined, respectively. Total DDT was represented by the sum of only the *p,p′* isomers of DDT and its degradates because the *o,p′* isomers had low detection frequencies and, on average, constituted only 1% of summed *o,p′*- and *p,p′*-DDX concentrations in NAWQA fish samples. Models also were developed for the individual components of total chlordane and total *p,p′*-DDX. These component models helped to identify outliers in the model-development data and evaluate consistency in explanatory variables but were not used for extrapolation and are not discussed in detail.

A second set of fish regression models was developed in which the measured pesticide concentration in bed sediment (collected at the same site as fish) was considered as a potential explanatory variable. The objective was to evaluate whether data on pesticides in sediment improved the explanatory power of the fish model, compared with using data on watershed characteristics alone. These sediment-based fish models cannot be used to predict fish concentrations in unmonitored streams because measured sediment data are not available to support a national extrapolation.

Model-development data for pesticides in fish

The NAWQA data used for model development consist of chemical analyses of one composite whole-fish sample from each of 648 stream sites in 43 hydrologic basins (study areas) in the conterminous United States (Fig. 1). These sites represent a variety of land-use and environmental settings [2]. Typically, each site was sampled once between 1992 and 2001. For sites where multiple fish samples were collected, only one sample was selected. Preference was given to the fish sample collected on the same date as bed sediment, to commonly sampled fish taxa, or both [26]. Each fish sample was a composite of five to eight individual whole fish of the same taxon collected on a single date. Fish taxa were selected from a National Target Taxa List [27]. The national data represent 57 fish taxa, of which the most common are common carp (29% of samples) and white sucker (26%). Fish sampling sites are shown in Figure 1, with sites color-coded by fish taxon.

Fish samples were analyzed for organochlorine pesticide compounds at the USGS National Water Quality Laboratory in Denver, Colorado. Fish samples were homogenized, and a subsample was dried and Soxhlet-extracted with dichloromethane. An aliquot of the extract was removed for lipid determination. The extract was cleaned by gel permeation chromatography, solvent-exchanged to hexane, fractionated using alumina/silica adsorption chromatography, and analyzed by gas chromatography with electron capture detection [28]. The reporting level for individual organochlorine compounds in fish was 5 μg/kg. The reporting levels for total *p,p′*-DDX (15 μg/kg) and total chlordane (20 μg/kg) were calculated by summing reporting levels for their individual components.

Bed sediment samples were collected on the same date as fish samples for 49% of sites, within one year for 86% of sites, and within three years for the remainder. Sediment samples were composites of fine-grained sediment (sieved to <2 mm) collected from multiple depositional areas within a stream reach on a single date, typically during low-flow conditions ([29]; http://water.usgs.gov/nawqa/pnsp/pubs/ofr94-458), and analyzed for organochlorine pesticides by gas chromatography with electron capture detection ([30]; http://nwql.usgs.gov/OFR-95-140.shtml).

Three slightly different model-development data sets were used, one each for dieldrin, chlordane-group, and DDT-group models. With the original data set (648 sites), preliminary models were developed, and regression diagnostics (described below in the *Model development* section) were used to identify sites that were outliers in one or more models. Outliers were dropped from each data set as appropriate. A few outlier sites were deleted from all three model-development data sets if

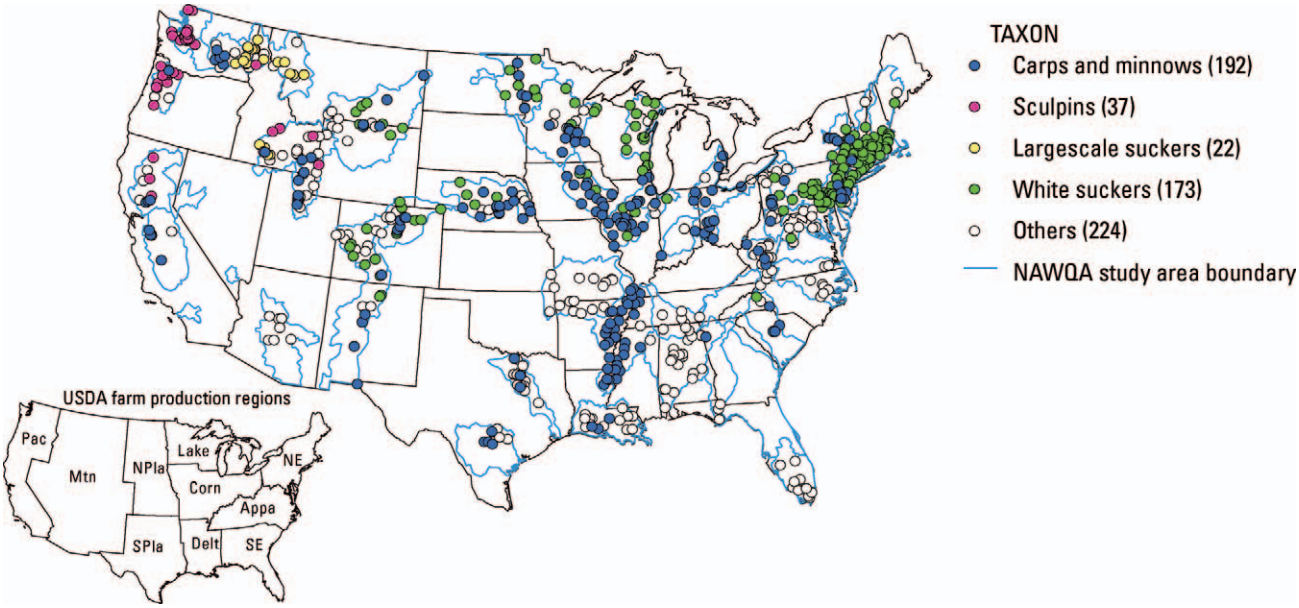

Fig. 1. National Water Quality Assessment (NAWQA) Program sites in the model-development data sets for pesticides in whole fish. The 648 sites are color-coded by the taxon of the fish sampled at each site. The *others* category represents 53 taxa. The number of samples for each taxon is shown in parentheses. Study unit boundaries are shown in blue. Inset, U.S. Department of Agriculture (USDA) farm production regions [32]. Appa = Appalachian; Corn = Corn Belt; Delt = Delta States; Lake = Lake States; Mtn = Mountain; NE = Northeast; NPla = Northern Plains; Pac = Pacific; SE = Southeast; SPla = Southern Plains.

warranted by extreme sample or site characteristics (such as extremely high or low fish lipid content or a history of dredging in the stream). The dropped sites made up less than 3% of the total stream sites. Restoring the outliers to the data sets did not change the regression coefficients substantially (relative percent difference <10, usually $<5\%$). Without the outliners, the model fit improved slightly (the proportion of the variation explained increased by 6%, and the standard deviation of the residual error decreased by 6%, on average).

Explanatory variables

Factors that could affect pesticide use or transport were considered as potential explanatory variables. Potential variables are listed with their data sources in *Supporting Information* Table S2 (http://dx.doi.org/10.1897/08-508.S2) and described in more detail in Nowell et al. [26]. Some potential explanatory variables were characteristics of the fish sample; others represented past sources of the pesticide in the basin, watershed characteristics, or geographical region. Variable names are shown in italics.

Fish sample characteristics. Three types of variables represent fish sample characteristics:

Fish lipid content (in percent) was considered because of its generally high correlation with organochlorine pesticide residues in fish ($p \leq 0.0001$).

Time is defined as the number of years elapsed from 1966 (representing the period of peak organochlorine use) to the sampling date. Because the present study spanned a 10-year period, samples collected late in the study period may have undergone greater pesticide degradation or dissipation than samples collected earlier.

Eight taxon variables (e.g., *carp*) were created to represent the most common fish taxa in the model-development data set. Each is a dummy variable, populated by either 1 or 0, depending on the sample taxon. Because organochlorine pesticide bioaccumulation in fish varies with taxon [6], taxon variables were intended to represent species-specific characteristics that could affect pesticide concentrations and were not already accounted for by differences in *fish lipid content.*

Source variables. Several variables represent estimates (or surrogates) of past sources of pesticides in the drainage basin.

Historical *agricultural use intensity* in the basin was estimated for aldrin and dieldrin combined (for use in the dieldrin model), chlordane (for chlordane-group models), and DDT and TDE combined (for DDT-group models) on the basis of regional application rates and county-level harvested crop acreage (G.P. Thelin, U.S. Geological Survey, Sacramento, CA, July 13, 2004, personal communication), as described in *Supporting Information* S3 (http://dx.doi.org/10.1897/08-508.S3) (with references in *Supporting Information* S4; http://dx.doi.org/10.1897/08-508.S4). Estimates are designed to represent use during the 1960s, so regional application rates are based on U.S. Department of Agriculture (USDA) data on pesticide use by farmers in 1966 or 1971 [31,32]. Harvested crop acreage data are from the 1964 and 1969 Census of Agriculture [33]. Agricultural use estimates for pesticides in the present study are shown in Figure 2. Distinct changes in *agricultural use intensity* sometimes occur along state boundaries (Fig. 2A–C) because regional application rates are based on USDA farm production regions (Fig. 1, inset), which follow state boundaries.

The *termite-urban score* is a surrogate representing past termiticide use [26]. Aldrin, dieldrin, and chlordane were used in subterranean termite control through the late 1980s [7,8,34], but quantitative use data are not available at the county or regional scales. The *termite-urban score* was calculated by multiplying the urban land within the basin (1970s) by a weighting factor corresponding to the zone of subterranean termite density (*Supporting Information* S3; http://dx.doi.org/10.1897/08-508.S3). The 1970s land-use information was derived from the USGS's Land Use and Land Cover data (http://edc.usgs.gov/products/landcover/lulc.html, as enhanced

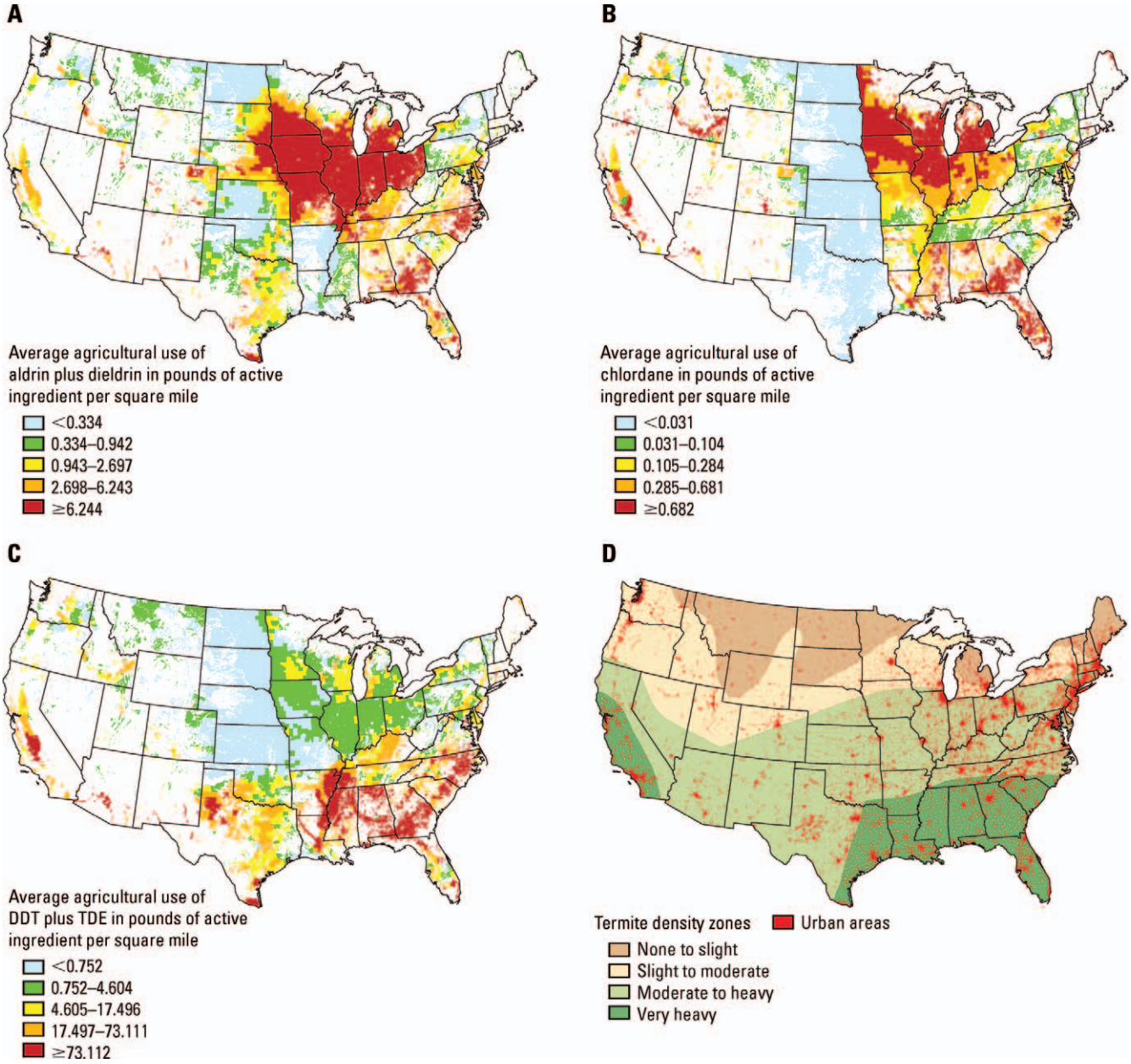

Fig. 2. National distribution of some key explanatory variables from organochlorine pesticide models in fish. Historical *agricultural use intensity* of (**A**) aldrin plus dieldrin, 1966; (**B**) chlordane, 1971 (1966 for AR, LA, MS); (**C**) DDT plus TDE, 1966; (**D**) urban land during the 1970s and zone of subterranean termite density, which are combined in the *termite-urban score*. Italics indicate variable names. See *Supporting Information* for variable derivation (*Supporting Information* S3; http://dx.doi.org/10.1897/08-508.S3) and source references (*Supporting Information* S4; http://dx.doi.org/10.1897/08-508.S4). TDE = tetrachlorodiphenylethane.

by http://pubs.usgs.gov/ds/2006/240), which henceforth is called the 1970s Geographic Information Retrieval and Analysis System (GIRAS) data (after the digital file storage format and software system used to develop the data). Subterranean termite distribution zones are from Beal et al. [35]. Figure 2D shows the national distribution of urban land (1970s GIRAS data) and the subterranean termite density zones.

Other source variables include various surrogates for agricultural use (e.g., *cropland and pasture*); urban use (e.g., *population density*); and minimal use (e.g., *forested land*), which represent parts of the basin where pesticide application was expected to be lower than that in developed areas. All land-use variables are based on 1970s GIRAS data (described above), and *population density* is from the 1990 Census of Population and Housing [36].

Watershed variables. Many watershed characteristics that may affect pesticide fate and transport were considered as potential explanatory variables (*Supporting Information* Table S2; http://dx.doi.org/10.1897/08-508.S2). These include soil characteristics such as mean *soil organic matter* content in the basin and *R factor* (rainfall erosivity from the Universal Soil Loss Equation), physical features such as *drainage basin area*, hydrologic parameters such as mean annual *runoff*, climate variables such as mean annual *precipitation*, and agricultural management practices such as *irrigated land* in the basin. All watershed variables are available for the conterminous United States.

Geographical region. Regional variables were used to represent any regional differences not accounted for by previously selected source and watershed variables. Ten regional variables

(e.g., *Southern Plains*) correspond to the 10 USDA farm production regions [32] (Fig. 1, inset and *Supporting Information* Table S2; http://dx.doi.org/10.1897/08-508.S2). Regional variables are dummy variables, populated by either 1 or 0; each site was assigned to one primary region, defined as the region containing the largest percentage of the watershed area for that site.

Sediment variables. Sediment variables (which reflect measured pesticide data in sediment collected at the same site as the fish) were considered only in sediment-based fish models. There are three types of sediment variables: The *fish-sediment date* (the difference between the fish and sediment sampling dates, in days); *sediment organic carbon* content (in grams per kilogram); and a series of dummy variables reflecting the pesticide concentration measured in sediment (*not detected, detection <10 μg/kg, 10 to <100 μg/kg*, and *≥100 μg/kg* dry wt in sediment), each populated by either 1 or 0. Although approximately half of the sites had their corresponding fish and sediment samples collected on different dates, the *fish-sediment date* variable was not selected in any sediment-based models.

Model development

In many samples, pesticide concentrations were reported as less than a detection threshold (nondetections), resulting in censored data for the response variable. Because conventional least-squares regression methods yield biased and inconsistent estimates with censored data [37], Tobit regression methods (also called maximum likelihood methods) were used. These methods provide parameter estimates of a censored linear model when the regression residual errors are independent and identically and normally distributed and have finite variance and a mean of zero [38]. For total chlordane and total *p,p′*-DDX models (in which the response variable represents summed concentrations of four and three components, respectively), the method was modified to include interval censoring. This entails computing two values for the response variable—a lower bound (with concentrations summed assuming censored values equal to zero) and an upper bound (with concentrations summed assuming censored values equal to the reporting level). For example, for a sample containing 20 μg/kg *p,p′*-DDE, 10 μg/kg *p,p′*-DDD, and <5 μg/kg *p,p′*-DDT, the interval censoring approach computes total *p,p′*-DDX as between 30 and 35 μg/kg, not simply as the censored sum, <35 μg/kg. When all three components are censored, the summed concentration (response variable) was considered left-censored and set equal to the sum of the reporting levels for all components (for example, total *p,p′*-DDX would be <15 μg/kg for a sample containing <5 μg/kg each of *p,p′*-DDE, *p,p′*-DDD, and *p,p′*-DDT). Maximum likelihood methods implemented in the *survreg* procedure ([39]; http://mayoresearch.mayo.edu/mayo/research/biostat/splusfunctions.cfm) in the statistical analysis program S-PLUS® 7 (TIBCO Software) [40] were used to estimate the parameters of the regression models.

In all models, the response variable is the log-transformed concentration of the pesticide (or group) in whole fish (wet wt). Many potential explanatory variables were transformed (either logarithmic or square root) to minimize deviations from the assumptions of the maximum likelihood method—that the relationship between the variables is linear in the parameters and the residual error is identically and normally distributed.

Measures of goodness of fit used in conventional least-squares analysis—such as the multiple coefficient of determination (R^2) and the standard deviation of the residual error—cannot be computed exactly for the Tobit regression model. Instead, a pseudo-R^2 (pR^2) suitable for use in the Tobit regression model was calculated using the method of Laitila [41]. As with conventional R^2, pR^2 ranges from zero to one (also expressed as 0 to 100%) and is an estimate of the proportion of the variation in the response variable explained by the regression model. The scale parameter, a low-biased estimate of the standard deviation of the residual error, is also used. In maximum likelihood estimation, scale provides only asymptotically unbiased estimates of the standard deviation of the residual error when estimated from sample data [42], with bias a function of sample size and degree of censoring.

Models were developed using a stepwise procedure similar to stepwise regression, except that the Akaike information criterion (AIC) [43] was used to select variables for inclusion in the model. The AIC balances model goodness of fit with the number of parameters needed to achieve that fit. Simpler models are favored over complex ones unless a complex model substantially improves the fit. Models were developed in the following four phases.

Phase I. Source variables, *fish lipid content*, and *time*, which were expected to be the most important in terms of explaining the variance in concentrations among sites. For sediment-based models, sediment variables also were considered.

Phase II. Taxon dummy variables, which could be important if lipid content did not adequately account for differences in bioaccumulation among fish taxa.

Phase III. Watershed variables, which have the potential to affect contaminant fate and transport. These may contribute to differences in contaminant concentrations measured in fish from different watersheds, once differences between sources and fish sample characteristics (lipid content and taxon) are considered.

Phase IV. Regional dummy variables, which represent any regional differences not accounted for by source estimates and watershed characteristics.

During each phase, new variables were added, and the stepwise AIC-based (stepAIC) procedure was applied to select explanatory variables that produce the model that best describes the response variable data. Occasionally, the stepAIC procedure selected two or more variables that were redundant (i.e., derived from the same ancillary data set), highly correlated with one another, or dependent (as indicated by high condition index and variance decomposition proportions). In this case, the variable with the lowest p value was retained, and the other redundant, correlated, or dependent variable(s) were eliminated from the model. At the end of the stepwise procedure, a subsampling technique was used to reduce the number of variables and to eliminate overfitting. A maximum of six variables was selected in each model, as a practical upper limit. Regression diagnostics available in the *survreg* procedure in the S-PLUS program [40] were used to identify influential observations and to aid in variable selection. These diagnostics include leverage, deviance residuals, and several diagnostics developed by Escobar and Meeker [44] for censored regression (case-weight, response, and shape parameter perturbations).

The final models were cross-validated by dividing model-development data randomly into two data subsets: a calibration data set (containing 70% of the data) and a validation data set (30% of the data). First, each model was fit to the calibration data subset, and the predicted values and residuals were computed for both the calibration data and the validation data subsets. Then, the process was reversed, and the model was

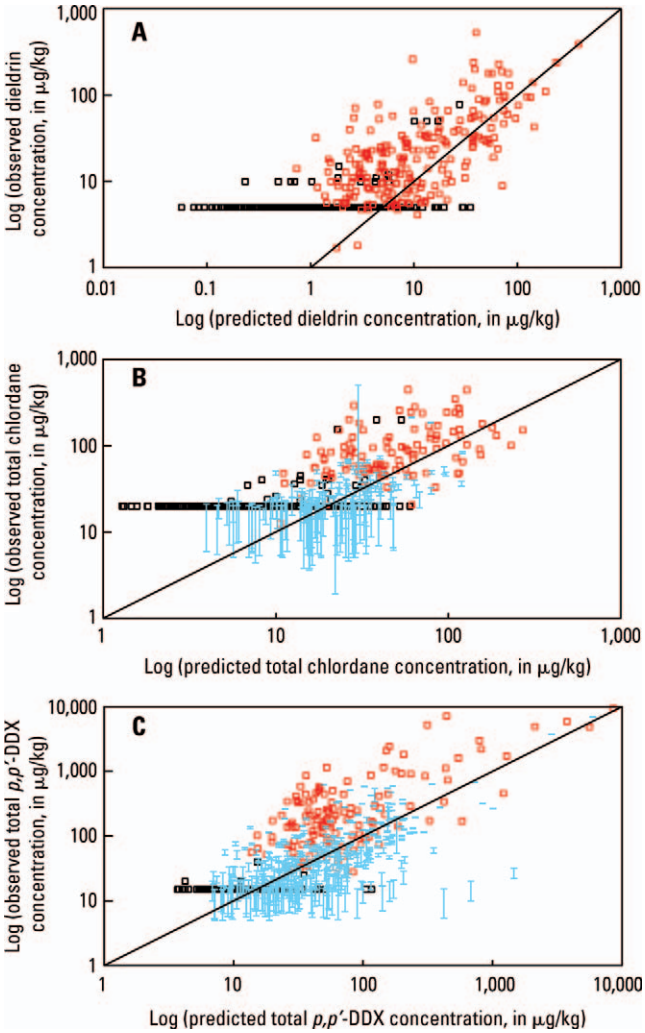

Fig. 3. Comparison of observed and predicted pesticide concentrations for sites in the model-development data set. Predicted concentrations are calculated using measured fish lipid values. The 1:1 line is shown. (**A**) Dieldrin model: Red squares, detected concentrations. Black squares, nondetections plotted as the reporting level. (**B**) Total chlordane model and (**C**) total *p,p′*-DDX model: Red squares, the sum of concentrations when all components are detected. Black squares, samples with no components detected, plotted as the sum of component reporting levels. Blue line segments, range of possible concentrations (from lower bound to upper bound; see text) when one or more, but not all, components are detected. DDD = dichlorodiphenyldichloroethane; DDE = dichlorodiphenyldichloroethylene; DDX = sum of DDT, DDD, and DDE.

fit to the validation data subset. In all cases, there was little difference in the predicted and residual errors obtained from the two fitted models based on the different data subsets; comparison of boxplots showed comparable median and interquartile range values (e.g., Fig. 3 in Nowell et al. [26]).

Extrapolation to all streams

Regression models for dieldrin, total chlordane, and total *p,p′*-DDX were used to estimate concentrations in whole fish for all streams in the conterminous United States. The network of streams in the conterminous United States was defined by the U.S. Environmental Protection Agency (U.S. EPA) River Reach File 1 ([45]; http://water.usgs.gov/lookup/getspatial?erf1_2). The River Reach File 1 (RF1) streams include over 600,000 miles of streams and more than 60,000 watersheds at a scale of approximately 1:500,000. For every watershed, values of the explanatory variables required by the regression models were computed from nationwide data sources using geospatial analysis tools. Because all models contain the variable *fish lipid content*, all extrapolations must assume a specific lipid content value—in this case, the value used is 6.2%, which is the average lipid content of the composite whole-fish samples collected by the NAWQA Program from 1992 to 2001 [26] and the U.S. Fish and Wildlife Service's National Contaminant Biomonitoring Program from 1970 to 1986 (http://www.cerc.usgs.gov/data/ncbp/fish.htm). The regression model then was used to estimate the concentration of the pesticide in whole fish for the stream at the outlet of each watershed.

Because the response variable is a logarithmic transformation and the regression residuals were approximately normally distributed, concentrations predicted by the model (after retransformation) approximate the median concentrations expected for sites that have a given set of explanatory values, rather than mean concentrations. For example, for 10 watersheds with identical explanatory variables, the model will predict the same pesticide concentration for all 10 watersheds. In actuality, the true concentrations are expected to be less than the predicted value at half of the 10 watersheds and greater than the predicted value at the other half. Thus, the median response refers to the distribution of concentrations among sites. Predicted concentrations were not adjusted for transformation bias [46,47] because estimates of site medians were considered appropriate for the present study objectives.

Probability of exceeding wildlife benchmarks

The models can be used to estimate the probability that the actual concentration at a site exceeds a certain concentration, such as a benchmark for protection of wildlife. This probability is a function of the predicted concentration for a site and the model error (standard error). The standard errors were estimated from the maximum likelihood scale parameter, using the adjustment suggested by Aitkin [42]. Specifically, a prediction interval (PI) is computed as follows:

$$\mathrm{PI} = \left[\log(C_{\mathrm{est}}) \pm \mathrm{SE}_{\mathrm{est}} \times t\left(\frac{\alpha}{2}\right)\right] \quad (1)$$

where $\log(C_{\mathrm{est}})$ is the logarithm of the estimated pesticide concentration, $t(\alpha/2)$ is the point on Student's t distribution with a probability of exceedance of one-half of α, and $\mathrm{SE}_{\mathrm{est}}$ is the standard error of the estimated logarithm of concentration. This means of assessing uncertainty in the logarithm of the predicted pesticide concentration (response variable) is approximate because it does not fully account for the effect of censoring of the observed concentration data.

The probability that the actual concentration at a site will exceed a certain concentration (P_{exc}) is approximated by algebraically rearranging the formula for computing prediction interval limits (Eqn. 1), as follows:

$$P_{\mathrm{exc}} = \frac{[\log(C_{\mathrm{b}} - \log(C_{\mathrm{est}})]}{\mathrm{SE}_{\mathrm{est}}} \quad (2)$$

where C_{b} is the benchmark concentration. When the regression residual error is normally distributed, the error distribution about an estimated value of the response 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. Therefore, the magnitude of the uncertainty is greater

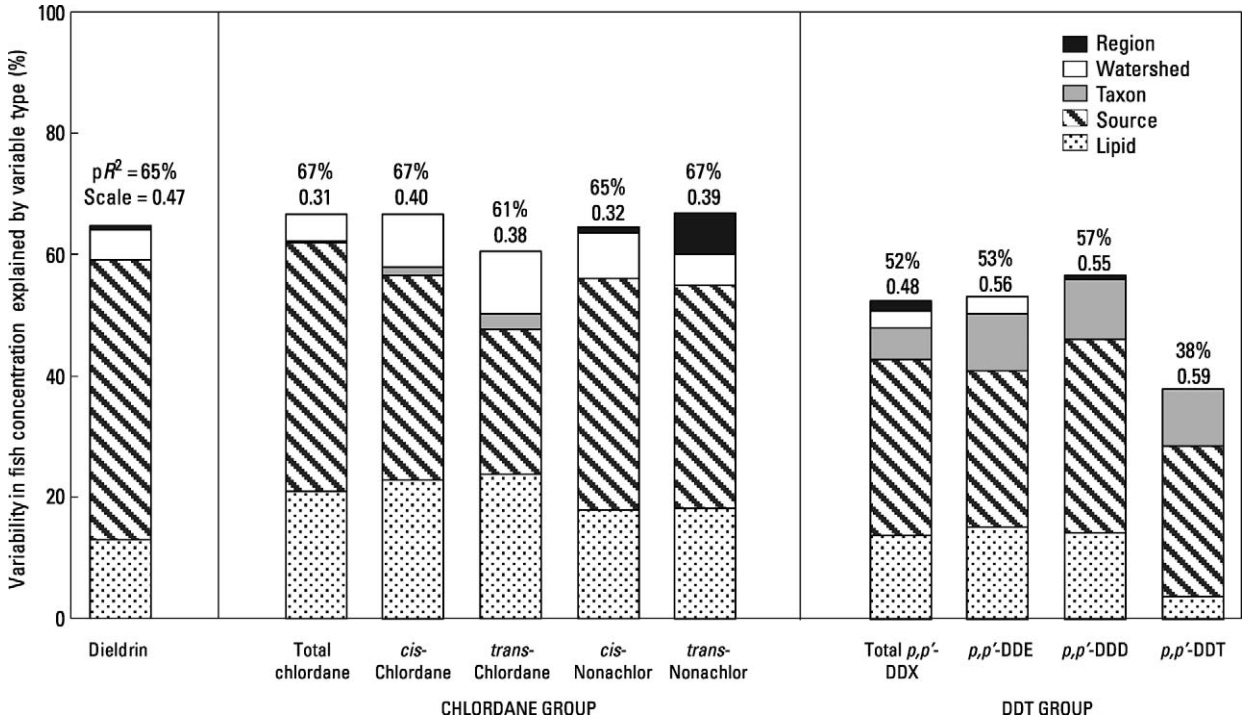

Fig. 4. Variability in pesticide concentrations explained by different types of explanatory variables in fish regression models for dieldrin, total chlordane and its four components, and total *p,p′*-DDX and its three components. Pseudo-R^2 (pR^2) and scale were calculated using the methods of Laitila [41] and Aitkin [42], respectively. DDD = dichlorodiphenyldichloroethane; DDE = dichlorodiphenyldichloroethylene; DDX = sum of DDT, DDD, and DDE; R^2, coefficient of multiple determination.

above the estimated concentration than it is below it, and it is greater for higher concentrations than for lower concentrations (Fig. 4 in Nowell et al. [26]).

The regression models were used to estimate the probability that predicted pesticide concentrations would exceed applicable benchmarks for protection of fish-eating wildlife, and the national distribution of probabilities was mapped for streams across the conterminous United States. This analysis assumes 6.2% fish lipid content (to represent the mean whole-body lipid content of all fish taxa) because wildlife consumes various fish species and typically eats the entire fish. Predicted organochlorine pesticide concentrations were compared with a range of systematically derived wildlife benchmarks available from the literature, following Gilliom et al. [2]. The ranges of benchmark values are 81 to 120 μg/kg for dieldrin, 300 to 4,200 μg/kg for total chlordane, and 6 to 200 μg/kg for total DDT. The present study compares predicted concentrations with both the high and low threshold values (called high benchmark and low benchmark) for a given pesticide—except that for total *p,p′*-DDX the reporting level (15 μg/kg) was substituted for the low benchmark (6 μg/kg). Because the low benchmark was below the reporting level, the probability of exceeding this benchmark could not be reliably estimated using the model. Technically, the total DDT benchmarks apply to the sum of *p,p′* plus *o,p′* isomers of DDT, DDD, and DDE. Therefore, predictions for total *p,p′*-DDX (*p,p′* isomers only) from the present study may slightly underestimate benchmark exceedances. However, use of total DDT benchmarks is reasonable because total *p,p′*-DDX constitutes 99% (on average) of total DDT concentrations in NAWQA fish samples.

RESULTS AND DISCUSSION

Regression models

Regression models for dieldrin, total chlordane, and total *p,p′*-DDX explained 52 to 67% of the variability in concentrations in whole fish (Table 1). For these three models, Figure 3 shows the observed versus predicted pesticide concentrations for sites in the model-development data set. For dieldrin (Fig. 3A), censored samples (nondetections) are plotted at the reporting level. For total chlordane (Fig. 3B) and total *p,p′*-DDX (Fig. 3C), the concentration in partially censored samples—which have detections for one or more, but not all, components of the sum—is unknown but falls somewhere between a lower bound (which assumes zero concentrations for nondetected components of the sum) and an upper bound (which assumes that nondetected components have concentrations equal to the reporting level). This range is indicated by the blue line segments in Figure 3B and C. Left-censored samples (in which all components are nondetected) have actual concentrations somewhere between zero and the value plotted in Figure 3B and C, which is the sum of the reporting levels for all components.

Important explanatory variables

Fish lipid content and source variables representing past agricultural and urban sources of pesticides were the most important explanatory variables in all models, together accounting for over three-quarters of the total variability explained by each model (Table 1 and Fig. 4). One or more taxon, watershed, and/or regional variables were selected in all models, but these tend to explain relatively little of the variability in the response variables. Therefore, possible reasons for their influence are speculative and are not discussed below in detail.

Models for individual components of total chlordane and total *p,p′*-DDX often have similar source and lipid variables but poorer model performance (lower or comparable pR^2 and higher scale) compared with models for total chlordane and total *p,p′*-DDX. Although not discussed here in detail, com-

Table 1. Regression models for dieldrin, total chlordane, and total *p,p′*-DDX in whole fish: Explanatory variables, model performance, and percentage of censored data in the model-development data set[a]

Pesticide compound or group	Explanatory variables[b]				Model performance				Censored data (%)	
							Attributed to source variables only			
	Source variables	Fish taxon variables	Watershed variables	Regional variables	Overall pseudo-R^2 (%)[c]	Scale	Pseudo-R^2 (%)[c]	Variability explained (%)	Left censored	Interval censored
Dieldrin[d]	log(Lipid), Aldrin+dieldrin use, log(Termite+1), √Forest	—	log(OM)	SPla	65	0.47	59	91	64	—
Total chlordane[e]	log(Lipid), Cropland, log(Termite+1)	Carp	log(Drainage), log(Irrigation +1)	—	67	0.31	62	93	53	28
p,p′-DDX[f]	log(Lipid), DDT+TDE use, log(Popdensity+1)	Carp	√Rfact	SPla	52	0.48	43	82	20	59

[a] Footnotes contain regression model equations. Data sources for variables are in *Supporting Information* Table S2 (http://dx.doi.org/10.1897/08-508.S2), and references are in *Supporting Information* S4 (http://dx.doi.org/10.1897/08-508.S4).

[b] Abbreviations (complete variable names are in italics): Aldrin+dieldrin use = *agricultural use intensity* of aldrin plus dieldrin (lbs/mi^2); DDT+TDE use = *agricultural use intensity* of DDT plus TDE (lbs/mi^2); Cropland = *cropland and pasture* (percent of basin land); Drainage = *basin drainage area* (km^2); Forest = *forested land* (percent of basin land); Irrigation = *irrigated land* (percent of basin); log = logarithm; Lipid = *fish lipid content* (percent); OM = mean *soil organic matter* content (percent by weight); Popdensity = *population density* (people per km^2); Rfact = mean *R factor*; SPla = *Southern Plains* region; TDE = tetrachlorodiphenylethane; Termite = *termite-urban score*; μg/kg = microgram per kilogram wet weight; (—) represents none. One was added to some variables prior to transformation to avoid taking the logarithm of zero.

[c] R^2 = multiple coefficient of determination.

[d] Model: log(dieldrin concentration, μg/kg) = −0.2050 + 1.0740 × log(Lipid) + 0.09111 × (Aldrin+dieldrin use) + 0.5246 × log(Termite+1) − 0.0773 × √Forest − 0.4827 × log(OM) + 0.2628 × (So Plain).

[e] Model: log(total chlordane concentration, μg/kg) = −0.134709 + 0.65579 × log(Lipid) + 0.00419 × (Cropland) + 0.71783 × log(Termite+1) + 0.15066 × (Carp) + 0.09490 × log(Drainage) − 0.19322 × log(Irrigation+1). Total chlordane is the sum of *cis*-chlordane, *trans*-chlordane, *cis*-nonachlor, and *trans*-nonachlor.

[f] Model: log(total *p,p′*-DDX concentration, μg/kg) = 0.77912 + 0.71958 × log(Lipid) + 0.00698 × (DDT+TDE use) + 0.31874 × log(Popdensity+1) + 0.38140 × (Carp) − 0.03187 × √Rfact + 0.42418 × (SPla). *p,p′*-DDX is the sum of *p,p′* isomers of DDT, DDD, and DDE. (DDD = dichlorodiphenyldichloroethane; DDE = dichlorodiphenyldichloroethylene.)

ponent models are mentioned when relevant to explanations of total chlordane and total *p,p′*-DDX models.

Dieldrin model. The dieldrin model explained 65% of the variability in dieldrin concentrations in fish (Table 1). Most was accounted for by *agricultural use intensity* of aldrin plus dieldrin (which by itself explained 25% of variability), fish lipid content (13%), *termite-urban score* (12%), and *forested land* (9%). These source variables are consistent with the past use history of aldrin and dieldrin, which were used in agriculture until the early 1970s and for termite control until the late 1980s [3,6–8]. *Forested land* is a surrogate variable representing minimal use and is the only source variable with a negative regression coefficient. Two secondary variables, *soil organic matter* and the *Southern Plains* region, explain an additional 5% and 1%, respectively. All variables selected are identical to those in a preliminary published model [26], although the regression coefficients changed slightly in value because some corrections were made in the model data development set after the preliminary model was developed.

Total chlordane model. Of the 67% of total chlordane variability in fish explained by this model (Table 1), most was accounted for by the *termite-urban score* (34% of variability), *fish lipid content* (21%), and *cropland and pasture* (7%). Past chlordane use for termite control in the 1980s was higher (*Supporting Information* Table S1; http://dx.doi.org/10.1897/08-508.S1) and more recent than its use in agriculture, which was discontinued in the early 1970s [3,6–8]. The surrogate variable *cropland and pasture* was selected over the estimated *agricultural use intensity* of chlordane, perhaps because uncertainty in the chlordane use estimate was high (*Supporting Information* S3; http://dx.doi.org/10.1897/08-508.S3). Of the three secondary variables in the model, *drainage basin area* accounts for 5% of the variability, and *carp* and *irrigation* each explain less than 1%.

Total p,p′*-DDX model.* This model explained 52% of the variability in total *p,p′*-DDX concentrations in fish (Table 1). Most was accounted for by *agricultural use intensity* of DDT and TDE (19% of the variability), *fish lipid content* (14%), and *population density* (10%). Selection of these source terms is consistent with the past use of DDT in agriculture and for disease vector control in urban areas [8]. The three secondary variables selected were *carp* (5% of the variability), *R factor* (3%), and *Southern Plains* (2%). *Carp*'s coefficient is positive, so its presence in the model results in higher predictions for carp compared with other fish taxa. The *R factor* (rainfall erosivity)—selected in both total *p,p′*-DDX and *p,p′*-DDE models—was the only watershed variable in any DDT-group models.

The DDT-group models are notable because more of the variability is explained by taxon variables and less by watershed variables than in all of the other models (Fig. 4). In addition, the pR^2 is lower for DDT-group models compared with other models, and the scale is higher (Fig. 4). These differences suggest that the explanatory variables selected in DDT-group models do not represent the source and/or fate properties of the modeled pesticide compound(s) as effectively

as the dieldrin and chlordane-group models do. It is possible that taxon variables are important because lipid content may not sufficiently account for species differences in total *p,p′*-DDX bioaccumulation, which may be affected by trophic level (e.g., [48]). Alternatively, each individual taxon has its own national distribution (Fig. 1), so these terms may represent a geographic correction. Note that the total *p,p′*-DDX model has greater explanatory capability (higher pR^2) and greater predictive capability (lower scale) than the model for the parent compound, *p,p′*-DDT (Fig. 4). The total *p,p′*-DDX model benefits from the interval censoring approach because there is less left-censored data (20%) than in the *p,p′*-DDT model (75%).

Sediment-based models. When variables representing pesticides in sediment were considered as potential explanatory variables in fish models for dieldrin, total chlordane, and total *p,p′*-DDX, one or more of these sediment-based variables were selected. Dieldrin and total chlordane fish models each contain one sediment variable (*nondetections*), whereas the total *p,p′*-DDX model contains three (*detection <10 μg/kg, ≥10 μg/kg*, and *sediment organic carbon*). *Fish lipid content* and most source variables were retained in sediment-based models, but *forested land* was dropped from the dieldrin model and *population density* was dropped from the total *p,p′*-DDX model when sediment-based variables were present. There also were changes in secondary (taxon, watershed, and regional) variables. With the incorporation of sediment-based variables into the fish models, model performance improved slightly for total *p,p′*-DDX—the pR^2 (now 59%) showed a 7% increase and the scale a 6% decrease—but not substantially for dieldrin (pR^2 decreased by 1% and scale by 2%) or total chlordane (pR^2 increased by 4% and scale decreased by 1%).

National extrapolation and significance to wildlife

Because models for total chlordane and total *p,p′*-DDX both contain the taxon variable *carp*, national extrapolations were done separately for carp and noncarp fish taxa. Figure 5 shows both pesticide concentrations in composite whole fish (6.2% lipid) predicted by the model (Fig. 5A, C, and E) and the probability of concentrations exceeding wildlife benchmarks (Fig. 5B, D, and F) for these pesticides: dieldrin in all taxa (Fig. 5A and B), total chlordane in noncarp taxa (Fig. 5C and D), and total *p,p′*-DDX in noncarp taxa (Fig. 5E and F). Because the variable *carp* has a positive coefficient in models for both total chlordane and total *p,p′*-DDX, predicted concentrations in carp (not shown) are somewhat higher than those shown in Figure 5. Streams in Figure 5A, C, and E are color-coded to correspond to pesticide concentrations in whole fish predicted by the model. In each case, the lowest category is defined by the reporting level in the present study, and the upper categories are based on the 50th and 90th percentiles of predicted concentrations (by stream reach). In Figure 5B, D, and F, streams are color-coded according to the probability of exceeding two wildlife benchmarks from the literature, the high benchmark and low benchmark—except that in Figure 5F the probability of exceeding the reporting level (15 μg/kg) is used instead of the low benchmark (6 μg/kg), which is below the reporting level. In model applications, the choice of benchmark typically is made in the context of study objectives. For example, to identify streams with the highest probability of adverse effects, it may be appropriate to compare predicted concentrations with high benchmark values. However, low benchmark values might be more appropriate for use as part of a sensitive screening approach to identify watersheds with endangered species concerns. Because the present study takes a screening approach, the probability maps in Figure 5 emphasize comparison with low benchmark values.

Dieldrin. The dieldrin model predicts (assuming 6.2% fish lipid) that 72% of RF1 stream miles will have concentrations in whole fish that are less than the 5 μg/kg reporting level (Fig. 5A). The highest concentrations are predicted for streams in the Corn Belt (where aldrin was heavily used on corn, Fig. 2A), Texas, Oklahoma, and scattered urban areas (where aldrin, dieldrin, or both were likely used for termite control).

When predicted concentrations are compared to wildlife benchmarks (still assuming 6.2% fish lipid), most U.S. stream miles (92%) have less than a 5% probability of exceeding the low benchmark for dieldrin (81 μg/kg), and 7% of stream miles have a 5 to 50% probability of exceeding this same benchmark (Fig. 5B). Although less than 1% of stream miles have more than a 50% probability of exceeding the same low benchmark, this 1% totals almost 2,400 stream miles and represents a substantial portion of streams in the central Corn Belt. Red streams are the most likely to show adverse effects on wildlife, because these streams have more than a 50% probability of exceeding the high benchmark (120 μg/kg).

Total chlordane. Whole-fish concentrations of total chlordane (at 6.2% fish lipid) are predicted to be less than the 20 μg/kg reporting level in 86% of U.S. stream miles for noncarp taxa (Fig. 5C); this value is 78% of stream miles for carp, which have slightly higher predicted concentrations (not shown) than those in noncarp taxa (Fig. 5C). The highest predicted concentrations occur in the Corn Belt (where chlordane was applied to corn, Fig. 2B); in Texas, Louisiana, and the Southeast region (in the zone of highest termite density, Fig. 2D); and in urban areas across the country, including eastern states from Virginia to Massachusetts (Fig. 2D).

Almost all (99%) stream miles have less than a 5% probability of exceeding the low benchmark for chlordane (300 μg/kg), in either carp (not shown) or noncarp fish taxa (Fig. 5D). The few streams with more than a 5% probability of exceeding this benchmark tend to be located in scattered urban areas within the two highest termite zones (Fig. 2D). No streams have even a 5% probability of exceeding the high benchmark (4,200 μg/kg).

Total p,p′-DDX. In contrast to the two previous models, whole-fish concentrations of total *p,p′*-DDX are predicted to be less than the 15 μg/kg reporting level in only 18 to 30% of RF1 stream miles (for carp and noncarp taxa, respectively), assuming 6.2% fish lipid. Predicted concentrations for carp (not shown) are slightly higher than those for noncarp taxa (Fig. 5E). The highest predicted concentrations are in the Southeast, Mississippi Delta states, Texas, and Oklahoma (where DDT was heavily used on cotton; Fig. 2C and *Supporting Information* Table S1; http://dx.doi.org/10.1897/08-508.S1); in some agricultural areas (such as in California, where DDT was used on a variety of crops); and in scattered urban areas nationwide. Figure 5E shows a distinct boundary along state lines separating Texas and Oklahoma from the adjacent states, which is an artifact of the *Southern Plains* variable in the model. It is reasonable to expect high predicted concentrations in the Southern Plains states because of past use in the region on cotton but not as pronounced of a demarcation across state lines as was predicted by the model.

For total DDT, available wildlife benchmarks span a wide concentration range (6–200 μg/kg), so there is considerable uncertainty in the threshold concentration for adverse effects

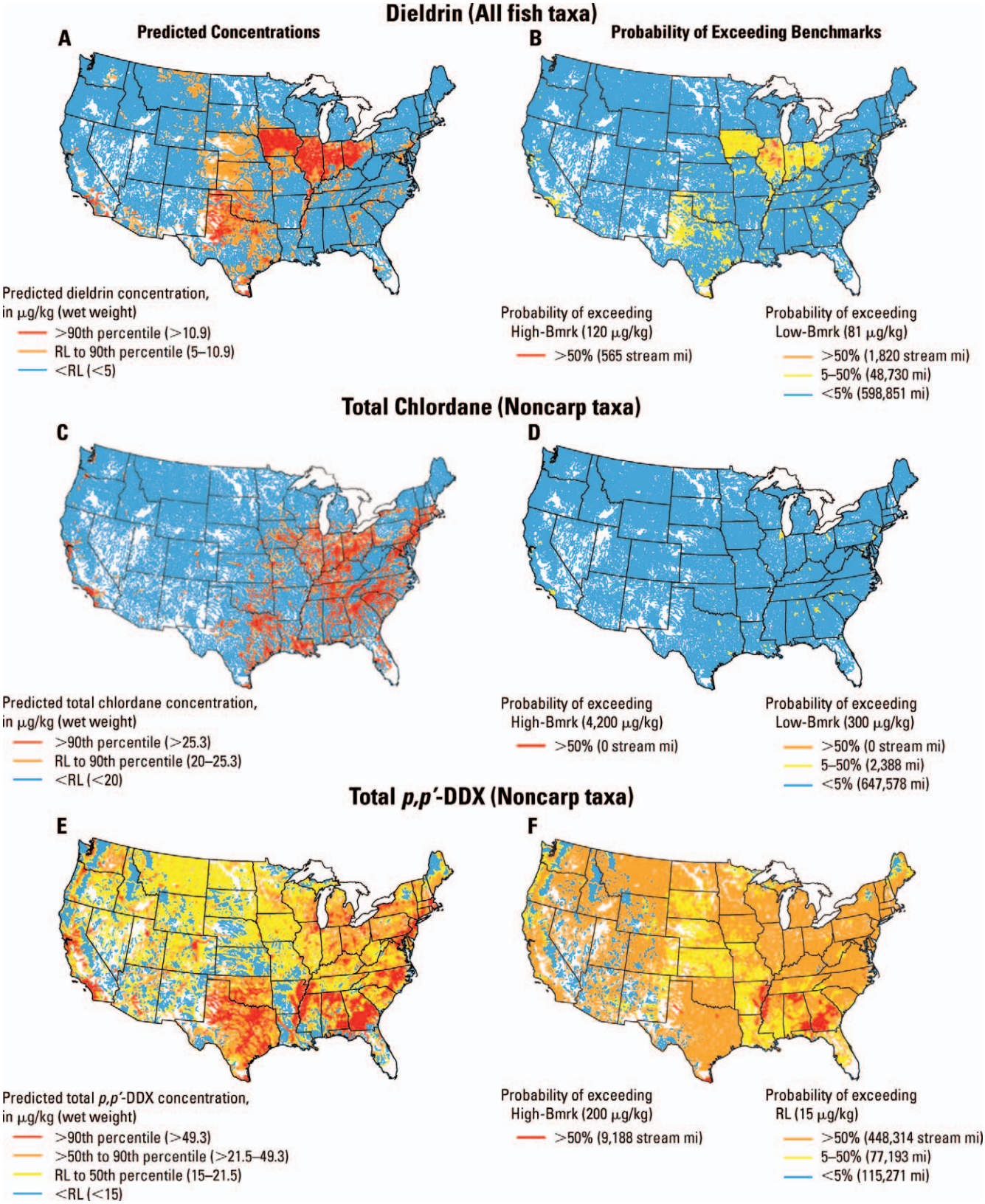

Fig. 5. Predicted concentrations of pesticides in whole fish (at 6.2% lipid) from U.S. streams and probability that concentrations exceed wildlife benchmarks. (**A**) Predicted concentrations and (**B**) probability of exceeding benchmarks for dieldrin in all fish taxa; (**C**) predicted concentrations and (**D**) probability of exceeding benchmarks for total chlordane in noncarp taxa; and (**E**) predicted concentrations and (**F**) probability of exceeding benchmarks for total *p,p′*-DDX in noncarp taxa. In panels A, C, and E, percentiles refer to predicted concentrations by stream reach. In panels B, D, and F, the number of stream miles is shown in parentheses for each probability category. Panel F shows the probability of exceeding the reporting level instead of the low benchmark, which is below the reporting level. DDD = dichlorodiphenyldichloroethane; DDE = dichlorodiphenyldichloroethylene; DDX = sum of DDT, DDD, and DDE; High-Bmrk = high benchmark; Low-Bmrk = low benchmark; mi = miles; RL = reporting level.

on fish-eating wildlife. Over 70% of stream miles across the United States have more than a 50% probability of exceeding the reporting level—70% for noncarp taxa (Fig. 5F) and 82% for carp (not shown)—and, therefore, of also exceeding the low benchmark (6 μg/kg), which is less than the reporting level (15 μg/kg). Red streams are the most likely to show adverse effects, because these have more than a 50% probability of exceeding the high benchmark for total DDT (200 μg/kg). Although these streams make up only 1 to 4% (for noncarp taxa and carp, respectively) of the total stream miles in the United States, they include a substantial proportion of stream miles in the Southeast and Mississippi Delta states.

Comparatively, total *p,p′*-DDX in whole-fish tissue is a much more widespread, continuing concern than dieldrin or total chlordane, and the potential concern for dieldrin is more widespread than that for chlordane. At least 70% of U.S. stream miles (~500,000 stream miles) are predicted to have more than a 50% probability of *p,p′*-DDX concentrations above the low benchmark, compared with less than 1% of stream miles (2,400 stream miles) for dieldrin and 0% for total chlordane. Those streams with the greatest probability of exceeding benchmarks are the highest priority for further investigation.

Fish with lipid content greater than 6.2% would likely have higher pesticide concentrations, and therefore also a higher probability of exceeding wildlife benchmarks, than shown in Figure 5. The reverse is true for fish with lipid content lower than 6.2%. In making predictions with the regression models, the lipid content assumption should be appropriate for the objectives of extrapolation. For example, one might assume a lipid content appropriate for an individual fish taxon of interest, such as largemouth bass (mean whole-body lipid content of 4.2%) or channel catfish (mean of 7.5%). At these lipid values, the percentage of stream reaches with predicted dieldrin concentrations greater than 10.9 μg/kg would change from 10% of stream reaches assuming 6.2% fish lipid content (shown as red streams in Fig. 5A) to 6% of stream reaches at 4.2% lipid and to 12 to 13% of stream reaches at 7.5% lipid. Predicted concentrations of dieldrin at two lipid values are shown in Figure 7 of Nowell et al. [26].

The empirical regression models developed in the present study have several limitations. There is considerable uncertainty in the models—depending on the pesticide, approximately 30 to 50% of variability in pesticide concentrations in fish was not explained by the model. The models can be applied to make predictions only where ancillary data on watershed characteristics are available (i.e., the conterminous United States only) and where watersheds are delineated. Also, the models apply to whole fish; this limits their applicability to human exposure assessment, which typically is concerned with contaminants in edible fish tissue. Moreover, because the models are based on data collected during the 1990s, future applications of the models may overestimate predicted pesticide concentrations in fish, given that residues in fish are declining nationally (albeit slowly) over time. The variable *time* (elapsed time between 1966 and the sampling date) was not selected as an explanatory variable in any of the final regression models, suggesting that time was not important over the 10-year sampling period (1992–2001). Although the *time* variable was initially selected by the stepAIC procedure in the total *p,p′*-DDX model, it was subsequently dropped in favor of more influential variables during model development. As expected, *time* had a negative coefficient in this model when initially selected, indicating declining residues over time.

CONCLUSIONS

Nationally available estimates of past pesticide use and watershed characteristics are useful explanatory variables for regression modeling of legacy organochlorine pesticide concentrations in whole fish from U.S. streams. The most important variables in explaining variability were the fish lipid content (measured) and source terms representing past agricultural and urban uses in the basin. Particularly important source terms are *agricultural use intensity* estimated from USDA data on pesticide use by farmers in 1966 (in dieldrin and total *p,p′*-DDX models); the *termite-urban score*, a surrogate for past termiticide use (in dieldrin and total chlordane models); and *population density*, a surrogate for general urban use (in the *p,p′*-DDX model). The resulting empirical models explained approximately 50 to 70% of the variability in concentrations of dieldrin, total chlordane, and total *p,p′*-DDX in whole fish collected by the NAWQA Program from 648 streams from 1992 to 2001. This degree of explanatory power is considerable, given that these models are based on data aggregated for multiple species of fish; the pesticides being modeled were not used in the United States for 5 to 30 years prior to sample collection; and only those watershed characteristics with nationally available data could be used as explanatory variables. Predicted concentrations were nearly always within an order of magnitude of the measured concentrations for the model-development streams. Results from cross-validation tests (in which one subset of the model-development data was used to calibrate, and another subset to validate, the model) were encouraging, but validation of the regression models using independent data sets has not yet been done.

National extrapolation to U.S. streams was performed by applying the regression models to streams in U.S. EPA's RF1 River Reach File, assuming a fish lipid content of 6.2% (to represent average whole-body lipid content of all fish taxa). Predicted concentrations are higher for total DDT than those for dieldrin or total chlordane, and the probability of exceeding wildlife benchmarks is also highest for total DDT. The vast majority of stream miles have less than a 5% probability of exceeding the most sensitive benchmarks for dieldrin (92% of stream miles) and total chlordane (over 99%). For total DDT, the wide range of available wildlife benchmarks (6–200 μg/kg) results in considerable uncertainty as to which streams have potential for adverse effects on aquatic life. Taking a sensitive screening approach, 70 to 82% of stream miles (depending on the fish taxon) have more than a 50% probability of exceeding the most sensitive benchmark.

Future application of these regression models may overestimate pesticide concentrations in fish, given that environmental residues continue to decline over time, decades after the uses of these pesticides were discontinued. Nonetheless, these models provide a cost-effective predictive tool to supplement existing monitoring data, assess the need for further monitoring, and guide the design and location of future sampling.

SUPPORTING INFORMATION

Table S1. For organochlorine pesticides in the present study, pesticide use information and occurrence data in whole fish from the U.S. Geological Survey's National Water Quality Assessment Program, 1992–2001.

Found at DOI: 10.1897/08-508.S1 (13 KB PDF).

Table S2. Potential explanatory variables considered in de-

veloping regression models for organochlorine pesticides in whole fish.

Found at DOI: 10.1897/08-508.S2 (57 KB PDF).

Supporting Information S3. Methods for determining key source variables: agricultural use intensity and termite-urban score.

Found at DOI: 10.1897/08-508.S3 (17 KB PDF).

Supporting Information S4. References cited in Tables S1 and S2 and Supporting Information S3.

Found at DOI: 10.1897/08-508.S4 (18 KB PDF).

Acknowledgement—This research was conducted as part of the U.S. Geological Survey (USGS) National Water Quality Assessment (NAWQA) Program. The authors thank the many individuals in the USGS whose work contributed to the present study, especially NAWQA Study Unit teams. We also thank Larry Brown, Norman Spahr, and David Mueller for their advice and manuscript review and two anonymous reviewers. The *survreg* procedure for the S language written by Terry M. Therneau of the Mayo Clinic was used to fit the regression models developed in the present study. The authors thank Michael Haverty, Nan-Yao Su, and W. David Woodson for assistance in obtaining termite distribution maps.

REFERENCES

1. Schmitt CJ, Zajicek JL, May TW, Cowman DF. 1999. Organochlorine residues and elemental contaminants in U.S. freshwater fish, 1976–1986: National Contaminant Biomonitoring Program. *Rev Environ Contam Toxicol* 162:43–104.
2. Gilliom RJ, Barbash JE, Crawford CG, Hamilton PA, Martin JD, Nakagaki N, Nowell LH, Scott JC, Stackelberg PE, Thelin GP, Wolock DM. 2006. The quality of our nation's waters—Pesticides in the nation's streams and ground water, 1992–2001. Circular 1291. U.S. Geological Survey, Reston, VA.
3. U.S. Environmental Protection Agency. 1990. Suspended, cancelled, and restricted pesticides. EN-342. 20T-4002. Office of Pesticides and Toxic Substances, Washington, DC.
4. Hickey JP, Batterman SA, Chernyak SM. 2006. Trends of chlorinated organic contaminants in Great Lakes trout and walleye from 1970 to 1998. *Arch Environ Contam Toxicol* 50:97–110.
5. Van Metre PC, Mahler BJ. 2005. Trends in hydrophobic organic contaminants in urban and reference lake sediments across the United States, 1970–2001. *Environ Sci Technol* 39:5567–5574.
6. Nowell LH, Capel PD, Dileanis PD. 1999. *Pesticides in Stream Sediment and Aquatic Biota—Distribution, Trends, and Governing Factors.* Pesticides in the Hydrologic System Series, Vol 4. CRC, Boca Raton, FL, USA.
7. U.S. Environmental Protection Agency. 1992. National study of chemical residues in fish, Vol II. EPA-823-R-92-008b. Office of Science and Technology, Washington, DC.
8. U.S. Environmental Protection Agency. 2000. Great Lakes binational toxics strategy: The Level I pesticides in the binational strategy. Section 3.0, Review of domestic use status by pesticide. Final Report. Great Lakes National Program Office, Chicago, IL.
9. Van Metre PC. 2006. Trends in total DDT and chlordane in lake sediments. In Gilliom RG, Barbash JE, Crawford CG, Hamilton PA, Martin JD, Nakagaki N, Nowell LH, Scott JC, Stackelberg PE, Thelin GP, Wolock DM. The quality of our nation's waters—Pesticides in the nation's streams and ground water, 1992–2001. Circular 1291. U.S. Geological Survey, Reston, VA.
10. Schmitt CJ, Zajicek JL, Peterman PL. 1990. National Contaminant Biomonitoring Program: Residues of organochlorine chemicals in freshwater fishes of the United States, 1976–1984. *Arch Environ Contam Toxicol* 19:748–782.
11. Li X, Yang L, Jans U, Melcer ME, Zhang P. 2007. Lack of enantioselective microbial degradation of chlordane in Long Island Sound sediment. *Environ Sci Technol* 41:1635–1640.
12. U.S. Environmental Protection Agency. 1975. DDT: A review of scientific and economic aspects of the decision to ban its use as a pesticide. EPA-540/1-75-022. Washington, DC.
13. Haake J, Kelley M, Keys B, Safe S. 1987. The effects of organochlorine pesticides as inducers of testosterone and benzo[*a*]pyrene hydroxylases. *Gen Pharmacol* 18:165–169.
14. Myers MS, Stehr CM, Olson OP, Johnson LL, McCain BB, Chan S-L, Varanasi U. 1993. National benthic surveillance project: Pacific coast, fish histopathology and relationships between toxicopathic lesions and exposure to chemical contaminants for Cycles I to V (1984–88). NMFS NWFSC-6. NOAA Technical Memorandum. Northwest Fisheries Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA.
15. Thomas MJ, Thomas JA. 2001. Toxic responses of the reproductive system. In Klaassen CD, ed, *Casarett and Doull's Toxicology—The Basic Science of Poisons*, 6th ed. McGraw-Hill, New York, NY, USA, p 701.
16. Burns-Naas LA, Meade BJ, Munson AE. 2001. Toxic responses of the immune system. In Klaassen CD, ed, *Casarett and Doull's Toxicology—The Basic Science of Poisons*, 6th ed. McGraw-Hill, New York, NY, USA, pp 441–442.
17. Weseloh DV, Hughes KD, Ewins PJ, Best D, Kubiak T, Shieldcastle MC. 2002. Herring gulls and great black-backed gulls as indicators of contaminants in bald eagles in Lake Ontario, Canada. *Environ Toxicol Chem* 21:1015–1025.
18. Beckvar N, Dillon TM, Read LB. 2005. Approaches for linking whole-body fish tissue residues of mercury or DDT to biological effects thresholds. *Environ Toxicol Chem* 24:2094–2105.
19. Truhlar JF, Reed LA. 1976. Occurrence of pesticide residues in four streams draining different land use areas in Pennsylvania, 1969–71. *Pestic Monit J* 10:101–110.
20. Munn MD, Gruber SJ. 1997. The relationship between land use and organochlorine compounds in streambed sediment and fish in the Central Columbia Plateau, Washington and Idaho, area. *Environ Toxicol Chem* 16:1877–1887.
21. Black RW, Haggland AL, Voss FV. 2000. Predicting the probability of detecting organochlorine pesticides and polychlorinated biphenyls in stream systems on the basis of land use in the Pacific Northwest, USA. *Environ Toxicol Chem* 19:1044–1054.
22. Krueger J, Tornqvist L. 1996. Multiple regression analysis and prediction of pesticide occurrence in subsurface drainage and streamflow related to pesticide intrinsic properties and quantities applied. *Proceedings*, 10th Symposium on Pesticide Chemistry, Castelnuevo Fogliana, Piacenza, Italia, La Goliardica Pavese, Pavia, Italy, September 30–October 2, pp 681–685.
23. Battaglin WA, Goolsby DA. 1997. Statistical modeling of agricultural chemical occurrence in midwestern rivers. *J Hydrology* 196:1–25.
24. Battaglin WA, Goolsby DA. 1998. Regression models of herbicide concentrations in outflow from reservoirs in the midwestern USA, 1992–1993. *J Am Water Resour Assoc* 34:1369–1390.
25. Larson SJ, Crawford CG, Gilliom RJ. 2004. Development and application of watershed regressions for pesticides (WARP) for estimating atrazine concentration distributions in streams. Water-Resources Investigations Report 03-4047. U.S. Geological Survey, Reston, VA.
26. Nowell LH, Crawford CG, Nakagaki N, Thelin GP, Wolock DW. 2006. Regression model for explaining and predicting concentrations of dieldrin in whole fish from United States streams. Scientific Investigations Report 2006-5020. U.S. Geological Survey, Reston, VA.
27. Crawford JK, Luoma SN. 1994. Guidelines for studies of contaminants in biological tissues for the National Water-Quality Assessment Program. Open-File Report 92-494. U.S. Geological Survey, Reston, VA.
28. Leiker TJ, Madsen JE, Deacon JR, Foreman WF. 1995. Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory—Determination of chlorinated pesticides in aquatic tissue by capillary-column gas chromatography with electron-capture detection. Open-File Report 94-710W. U.S. Geological Survey, Reston, VA.
29. Shelton LR, Capel PD. 1994. Guidelines for collecting and processing samples of stream bed sediment for analysis of trace elements and organic contaminants for the National Water-Quality Assessment Program. Open-File Report 94-458. U.S. Geological Survey, Reston, VA.
30. Foreman WT, Conner BF, Furlong ET, Vaught DG, Merten LM. 1995. Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory—Determination of organochlorine pesticides and polychlorinated biphenyls in bottom sediment by dual capillary-column gas chromatography with electron-capture detection. Open-File Report 95-140. U.S. Geological Survey, Reston, VA.
31. Eichers T, Andrilenas P, Blake H, Jenkins R, Fox A. 1970. Quan-

tities of pesticides used by farmers in 1966. Agricultural Economic Report 179. Economic Research Service, U.S. Department of Agriculture, Washington, DC.
32. Andrilenas PA. 1974. Farmers' use of pesticides in 1971—Quantities. Agricultural Economic Report 252. Economic Research Service, U.S. Department of Agriculture, Washington, DC.
33. Economic Research Service. 1999. Historical census of agriculture data, 1949–1974, records of the Economic Research Service (record Group 354). Section 1, Chapter 3-4, Census of Agriculture, Vol 1. National Archives, Washington, DC, USA.
34. U.S. Environmental Protection Agency. 1983. Analysis of the risks and benefits of seven chemicals used for subterranean termite control. EPA-540/9-83-005. Washington, DC.
35. Beal RH, Mauldin JK, Jones SC. 1994. Subterranean termites: Their prevention and control in buildings. Home and Garden Bulletin 64. U.S. Department of Agriculture, Washington, DC.
36. U.S. Bureau of the Census. 1991. Census of population and housing, 1990. Public Law 94-171. Washington, DC.
37. Judge GG, Griffiths WE, Hill RC, Lütkepohl H, Lee TC. 1985. *The Theory and Practice of Econometrics*, 2nd ed. John Wiley, New York, NY, USA.
38. Maddala GS. 1983. *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge University, Cambridge, UK.
39. Therneau TM. 1999. A package of survival routines for S. Mayo Clinic Department of Health Sciences Research, Rochester, MN, USA.
40. Insightful. 2005. S-Plus® 7 for Windows Release 6. Seattle, WA, USA.
41. Laitila T. 1993. A pseudo-R-square measure for limited and qualitative dependent variable models. *J Econometrics* 56:341–356.
42. Aitkin M. 1981. A note on the regression of censored data. *Technometrics* 23:161–163.
43. Venables WN, Ripley BD. 1999. *Modern Applied Statistics with S-Plus*, 3rd ed. Springer, New York, NY, USA.
44. Escobar LA, Meeker WQ Jr. 1992. Assessing influence in regression analysis with censored data. *Biometrics* 48:507–528.
45. Nolan JV, Brakebill JW, Alexander RB, Schwarz GE. 2003. Enhanced river reach file 2.0, version 2.0, November 10, 2003. Open-File Report 02-4. U.S. Geological Survey, Reston, VA.
46. Bradu D, Mundlak Y. 1970. Estimation in lognormal linear models. *J Am Stat Assoc* 65:198–211.
47. Duan N. 1983. Smearing estimate—A nonparametric retransformation method. *J Am Stat Assoc* 78:605–610.
48. Rasmussen JB, Rowan DJ, Lean DRS, Carey JH. 1990. Food chain structure in Ontario Lakes determines PCB levels in lake trout (*Salvelinus namaycush*) and other pelagic fish. *Can J Fish Aquat Sci* 47:2030–2038.