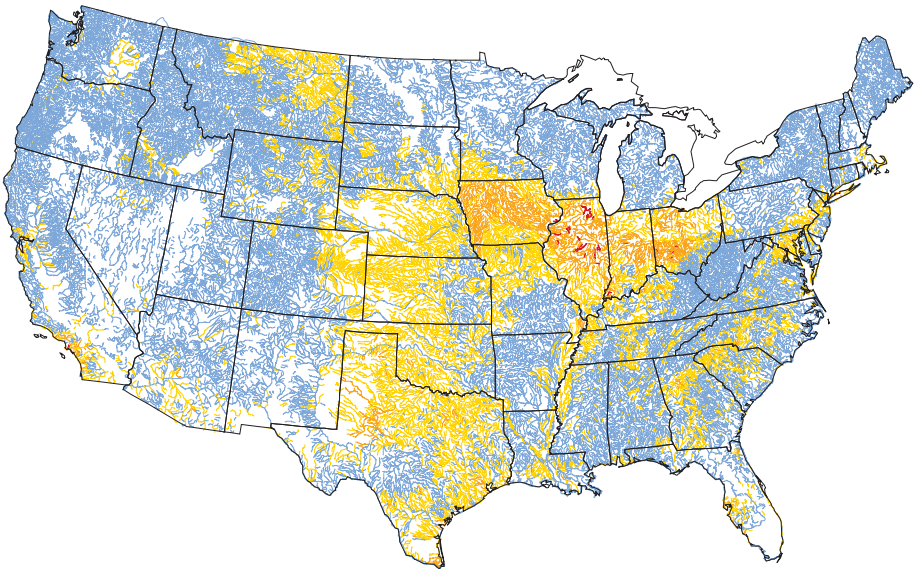
USGS
science for a changing world

National Water-Quality Assessment Program

Regression Model for Explaining and Predicting Concentrations of Dieldrin in Whole Fish from United States Streams



Scientific Investigations Report 2006–5020

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

Cover: Map shows the predicted dieldrin concentrations in whole fish (with 6.2 percent lipid content, the national average for sampled whole fish) from streams throughout the conterminous United States (see figure 7B).

Regression Model for Explaining and Predicting Concentrations of Dieldrin in Whole Fish from United States Streams

By Lisa H. Nowell, Charles G. Crawford, Naomi Nakagaki, Gail P. Thelin, and David M. Wolock

National Water-Quality Assessment Program

Scientific Investigations Report 2006–5020

Sacramento, California
2006

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

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

U.S. Geological Survey, Reston, Virginia: 2006
For Sale by U.S. Geological Survey, Information Services
Box 25286, Denver Federal Center
Denver, CO 80225-0286

For more information about the USGS and its products:
Telephone: 1-888-ASK-USGS

World Wide Web: http://www.usgs.gov

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

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

Suggested reference:
Nowell, L.H.; Crawford, C.G.; Nakagaki, Naomi; Thelin, G.P.; and Wolock, D.M., 2006, Regression model for explaining and predicting concentrations of dieldrin in whole fish from United States streams: U.S. Geological Survey Scientific Investigations Report 2006–5020, 30 p.

Contents

Abstract 1
Introduction 1
Background 1
Purpose and Scope 2
Acknowledgments 3
Methods 3
Dieldrin Data Used for Model Development 3
Sample Collection 3
Sample Work-Up and Analysis 3
Database Decisions for Model Development 4
Explanatory Variables 5
Fish Sample Characteristics 5
Past Sources of Aldrin and Dieldrin 8
Other Watershed Characteristics 9
Development of Regression Model 11
Regression Models 11
Transformation of Response and Explanatory Variables 12
Selection of Explanatory Variables 12
Analysis of Model Fit 13
Model Validation 13
Extrapolation to Unmonitored Streams 14
Data Sources for Unmonitored Streams Nationwide 14
Lipid Content 14
Uncertainty Analysis 15
Predicting the Potential Effects of Fish Consumption on Human Health and Wildlife 17
Potential Effects on Human Health 17
Potential Effects on Fish-Eating Wildlife 17
Results and Discussion 17
Regression Model for Dieldrin in Whole Fish 17
Analysis of Significant Explanatory Variables 19
Extrapolation to Unmonitored Streams 20
Significance to Human Health 20
Significance to Fish-Eating Wildlife 23
Uncertainty in Model Predictions 23
Summary and Conclusions 26
References 27

Figures

1. Map showing estimated historical use of aldrin plus dieldrin in agriculture (1966) in the conterminous United States 8
2. Map showing urban areas in the United States shown superimposed on the relative hazard of subterranean termite infestations 10
3. Graphs showing residual error developed from calibration data, and residual error developed from validation data 14
4. Graphs showing hypothetical regressions for which the response variable is log-transformed 16
5. Map showing measured dieldrin concentrations in fish at NAWQA sites, excluding sites with the highest and lowest 10 percent of lipid levels 18
6. Graph showing Effect of adding explanatory variables to the regression model for dieldrin in whole fish on the model's pseudo-R^2 and scale values 19
7. Map showing predicted dieldrin concentrations in whole fish for U.S. streams at (A) 4.2 percent lipid content (largemouth bass) and (B) 6.2 percent lipid content (all fish taxa) 21
8. Maps showing probability of streams having whole-fish samples that exceed (A) 5 µg/kg and (B) 25 µg/kg 23
9. Map showing probability of streams having whole-fish concentrations greater than 120 µg/kg 25

Tables

1. The 648 NAWQA stream sites used in model development, by major physiographic division and land use in the basin 4
2. Potential explanatory variables considered in developing the regression model for dieldrin in whole fish 6
3. Explanatory variables and coefficients for the regression model for dieldrin in whole fish 18

NOTE TO USGS USERS: , Use of hectare (ha) as an alternative name for square hectometer (hm^2) is restricted to the measurement of small land or water areas. Use of liter (L) as a special name for cubic decimeter (dm^3) is restricted to the measurement of liquids and gases. No prefix other than milli should be used with liter. Metric ton (t) as a name for megagram (Mg) should be restricted to commercial usage, and no prefixes should be used with it.

Temperature in degrees Celsius may be converted to degrees Fahrenheit (°F) as follows:
°F = (1.8×°C)+32.

Temperature in degrees Fahrenheit may be converted to degrees Celsius (°C) as follows:
°C = (°F-32)/1.8.

Abbreviations, Acronyms, and Initialisms

(Additional information given in parentheses)

AIC, Akaike Information Criterion
C_{est}, estimated dieldrin concentration
C_g, guideline concentration
DEM, digital elevation model
EPA, U.S. Environmental Protection Agency
FWS, Fish and Wildlife Service
GIRAS, Geographic Information Retrieval and Analysis System
HG, hydrologic groups
LULC, Land Use and Land Cover
NAWQA, National Water-Quality Assessment (USGS)
NCBP, National Contaminant Biomonitoring Program (FWS)
p, level of significance
P_{exc}, probability of exceedance
pR^2, pseudo-R^2
PI, prediction interval
PRISM, Parameter-elevation Regressions on Independent Slopes Model
R^2, coefficient of multiple determination
SD_{est}, standard error of the estimated logarithm of concentration
STATSGO, State Soil Geographic (database; National Resources Convervation Service)
SV, screening value
TIGER, Topologically Integrated Geographic Encoding and Referencing
TOPMODEL, topography model
USDA, U.S. Department of Agriculture
USGS, U.S. Geological Survey

Units of measurement

km, kilometer
lb/mi^2, pound per square mile
m, meter
mi, mile
µg/kg, microgram per kilogram
µg/L, microgram per liter

This page intentinally left blank.

Regression Model for Explaining and Predicting Concentrations of Dieldrin in Whole Fish from United States Streams

By Lisa H. Nowell, Charles G. Crawford, Naomi Nakagaki, Gail P. Thelin, and David M. Wolock

Abstract

A regression model was developed for estimating dieldrin concentrations in whole fish from United States streams using watershed characteristics, including estimates of past agricultural use intensity. Dieldrin concentrations were measured in composite whole-fish samples collected at about 650 sites nationwide during 1992–2001, as part of the U.S. Geological Survey's National Water-Quality Assessment Program. Important explanatory variables in the model include (1) fish lipid content; (2) the estimated historical agricultural use intensity of aldrin and dieldrin (which was calculated from regional application rates developed using U.S. Department of Agriculture data on pesticide use by farmers in 1966 and harvested acreage from the 1964 and 1969 Census of Agriculture); (3) a weighted termite-urban score (which represents past use for termite control and was developed from estimates of urban land use in the basin, combined with the national distribution of subterranean termite density); and (4) forested land in the basin (where past aldrin and dieldrin use was expected to be minimal). Other explanatory variables considered in developing the model included surrogates for past urban use (such as population density and percentage urban land use in the basin), taxon of the fish sampled, various watershed characteristics (including soil properties, hydrologic parameters, climatic variables, and agricultural management practices), and region of the country. Tobit regression methods were used for developing the model because about 60 percent of the dieldrin concentration values in the model-development data were reported as less than a detection threshold, resulting in censored data. The model explains 64 percent of the variability in dieldrin concentrations measured in whole fish. This model was used to predict dieldrin concentrations in whole fish in unmonitored streams throughout the United States using the U.S. Environmental Protection Agency's River Reach File 1.

Introduction

Background

The organochlorine pesticide dieldrin is still frequently detected in soil, sediment, biota, and air in the United States (Nowell and others, 1999; Schmitt and others, 1999; Wong and others, 2000), even though its uses were discontinued 15–30 years ago. Dieldrin residues in the environment may originate from past application of either aldrin, which degrades readily in the environment to dieldrin, or dieldrin itself. Aldrin and dieldrin (like DDT) are both organochlorine pesticides that were once used extensively, but were discontinued because of their persistence, tendency to bioaccumulate, carcinogenicity, and hazard to wildlife. Aldrin and dieldrin were applied in agriculture in the United States until the early 1970s, mainly on corn, but also on fruits and nuts, other vegetables, tobacco, and cotton. They also were used in nonagricultural applications—especially in termite control, which continued until at least the late 1980s. Dieldrin, like many organochlorine pesticides, is hydrophobic (low in water solubility, with a tendency to sorb to soil and sediment and to partition to organic matter) and resistant to degradation in the environment.

Three national studies have monitored dieldrin and other organochlorine pesticides in whole freshwater fish from rivers and streams across the United States. These are the Fish and Wildlife Service's (FWS) National Contaminant Biomonitoring Program (NCBP, 1969–1986), the U.S. Environmental Protection Agency's (EPA) National Study of Chemical Residues in Fish (1986–1987), and the U.S. Geological Survey's (USGS) National Water-Quality Assessment (NAWQA) Program (1992–2001). Evidence from these programs indicates that, nationally, organochlorine pesticide levels have declined since the use of these compounds was discontinued (Schmitt and others, 1999; Fuhrer and others, 1999), with the biggest declines occurring during the 1970s. Dieldrin residues in fish were highly variable during the 1970s, but were markedly lower after 1976 than before 1976; residues measured by the NCBP Program continued to decline during the 1980s (Schmitt and others, 1999; Nowell and others, 1999). NAWQA sampling data during 1992–2001 indicates the continued presence of dieldrin at low levels in fish and bed sediment from many streams in the United States (Wong and others, 2000; Nowell and Crawford, 2003)—dieldrin was detected in whole fish at over 50 percent of agricultural and urban streams, and in over 40 percent of streams in mixed land use areas (Nowell and Crawford, 2003). This suggests that dieldrin-contaminated soil is still being carried into receiving waters from both agricultural and urban land use settings.

Organochlorine pesticides have been shown to adversely affect the survival of various organisms (including aquatic invertebrates, fish, birds, and mammals) in laboratory tests; to disrupt the reproduction of fish and birds in the field; and to accumulate to high levels in fish-eating mammals (U.S. Environmental Protection Agency, 1975; Nowell and others, 1999). In addition, organochlorine pesticides (including dieldrin) induce the monooxygenase enzymes that hydroxylate testosterone (Haake and others, 1987), and they have been associated with endocrine and reproductive changes and immunosuppressive effects (Amdur and others, 1991). The biological significance of the low-level organochlorine residues detected in the environment today is more difficult to assess. Although the linkage between body burdens and effects in fish is not well understood, body burdens provide at least a direct indicator of exposure (Black and others, 2000). Organochlorine pesticide contamination in the field also has been associated with fish kills (Hunt and Linn, 1970; Madhun and Freed, 1990) and fish diseases (Myers and others, 1993). A few organochlorine pesticides (especially DDT and chlordane, and to a lesser extent dieldrin) are the subject of fish consumption advisories issued by the states for protection of human health from eating fish contaminated with bioaccumulative pollutants (U.S. Environmental Protection Agency, 2005).

Given the continued detection of low-level residues of organochlorine pesticides in fish, plus the uncertainty as to their potential effects on humans and wildlife from fish consumption, there is a need to ascertain the geographic extent and the degree of present-day exposure of aquatic populations to organochlorine pesticides. However, because of the high cost of monitoring and analysis, combined with decreasing interest in organochlorine pesticides as residues decline nationally, it is unlikely that substantial investment in monitoring will be made, except in areas where there is a strong probability of concern. An alternative to widespread direct monitoring is to predict concentrations using empirical regression models that relate measured concentrations in fish to watershed characteristics and estimates of past use.

A few prior studies have investigated the relationship between large-scale land use practices and the presence of organochlorine pesticides in sediment or biota (Truhlar and Reed, 1976; Munn and Gruber, 1997; Black and others, 2000). Black and others (2000) used logistic regression to develop predictive models for estimating the probability of detecting specific organochlorine pesticides in fish (sculpins) from streams in the Willamette and Puget Sound Basins of the Pacific Northwest. This regional-scale study predicted presence or absence, rather than specific concentration estimates. A number of studies have used linear regression or logistic regression models to predict concentrations of dissolved pesticides in stream water (Krueger and Tornqvist, 1996; Battaglin and Goolsby, 1997, 1998; Larson and Gilliom, 2001; Larson and others, 2004). Predictor variables in these models included pesticide use in the watershed, physical and chemical properties of the pesticides, and (or) various watershed characteristics. We are unaware of any studies that have developed models relating organochlorine pesticide concentrations in fish to land use and other watershed characteristics on a national scale.

Purpose and Scope

Assessment of risk to humans or wildlife from pesticides in the environment requires evaluation of both exposure and effects. The purpose of this report is to investigate the suitability and effectiveness of a regression model for predicting, and explaining variability in, dieldrin concentrations in whole fish from United States streams on a national scale. Specific objectives are to:

- Develop a regression model to estimate dieldrin concentrations in whole fish on the basis of national data on watershed characteristics, including estimates of historical pesticide use;
- Assess how much variability in pesticide concentrations in whole fish can be explained by factors in the regression model;
- Predict dieldrin concentrations in whole fish in streams across the United States, as a function of lipid (fat) content, from the regression model on the basis of watershed characteristics and estimates of historical pesticide use.

This model was developed using NAWQA data on dieldrin in whole fish at 648 stream sites sampled during 1992–2001. These stream sites are located in most parts of the country and represent a variety of land uses. Ultimately, the method described in this report will be applied to other organochlorine pesticides detected in fish by NAWQA.

In this report, the following terms are used: Concentrations measured at NAWQA sites are "observed concentrations." Application of the regression model using a given set of explanatory variables produces an "estimate" or "estimated concentration" for a site (in contrast to an observed, or measured, concentration). When the model is applied to an unmonitored site, the result is a "prediction" or "predicted concentration." Finally, when predictions are made for unmonitored sites over a wide geographic area, this builds a "spatial extrapolation."

Acknowledgments

The NAWQA data used in this study are the product of work by many people at the USGS who contributed to the study design, data collection, sample processing and analysis, and geographic information system (GIS) data analysis in the NAWQA Program. The authors are indebted to them all—especially to NAWQA study unit personnel, who developed the environmental framework and planned and conducted sampling for their study units, and provided basin-specific expertise for national land use classification and data analysis. The authors thank Larry R. Brown and David K. Mueller of the U.S. Geological Survey for their very helpful technical reviews of this report. The survreg procedure for the S language written by Terry M. Therneau of the Mayo Clinic was used to fit the regression model developed in this study. The authors are grateful to Michael Haverty (U.S. Dept. of Agriculture, Forest Service), Nan-Yao Su (University of Florida), and W. David Woodson (U.S. Geological Survey) for their assistance in obtaining termite distribution data and maps.

Methods

The methods used for this study included the use of appropriate data on dieldrin concentrations in whole fish in model development; estimating the past use of aldrin and dieldrin in agriculture; developing surrogates for past use of aldrin and dieldrin in urban settings; obtaining ancillary data on watershed characteristics from nationally available databases; developing a regression model to estimate dieldrin concentrations in whole fish; and using the model to predict dieldrin concentrations in whole fish in unmonitored streams across the United States.

Dieldrin Data Used for Model Development

Data on dieldrin in whole fish collected as part of the USGS's NAWQA Program were used for model development. The NAWQA Program collected samples from rivers and streams in 51 study areas (study units) across the United States. The study units are major hydrologic basins, which collectively encompass about 60 percent of the population and water use in the conterminous United States (Hirsch and others, 1988; Fuhrer and others, 1999).

The NAWQA program sampled whole fish (as well as bed sediment) at NAWQA sites to identify important hydrophobic contaminants, including dieldrin, and to determine their geographic distribution within the basin. Typically, fish sampling sites within a NAWQA study unit included the following types of sites: mixed land-use sites representing the large streams in the study unit, including major nodes in the drainage system; sites representing the principal environmental settings (land areas characterized by a unique combination of natural and anthropogenic factors, such as row crop cultivation on glacial-till soils); additional sites with known contamination (although point sources generally were avoided); and reference sites (generally forest or rangeland) where minimal occurrence of organochlorine compounds was expected.

Sample Collection

The NAWQA study design called for collecting at least one tissue sample (and one bed sediment sample) at each site. Whole-fish data were collected at 700 NAWQA sites. The distribution of sites does not quite cover the entire United States because clams were collected instead of fish at some sites (especially in the southeast), and appropriate fish and clam taxa were not available at some sites.

One fish taxon generally was sampled at each site, although multiple taxa were collected at about 10 percent of sites. Taxa were selected from a National Target Taxa List, and the same taxon was sampled at as many sites as possible within the study unit. Fish samples were composites of 5–8 individual whole fish, all from a single taxon (Crawford and Luoma, 1994).

Sample Work-Up and Analysis

Details on sample processing and analysis of organochlorine compounds in whole fish are published elsewhere (Leiker and others, 1995). Briefly, composite whole-fish samples were homogenized and a subsample was dried, Soxhlet-extracted with dichloromethane, and an aliquot was removed for lipid determination. The extract was cleaned up by gel permeation chromatography, solvent-exchanged to hexane, fractionated

using alumina/silica adsorption chromatography, then analyzed by gas chromatography with electron capture detection. Target analytes and their method detection limits in fish are listed at *http://ca.water.usgs.gov/pnsp/rep/bst/table1.html*.

Database Decisions for Model Development

Whole-fish data were collected at 700 NAWQA sites nationally. For model development, data for the Cook Inlet Basin (Alaska) and the Island of Oahu (Hawaii) study units, plus a few sites with watersheds extending into Canada or Mexico, were excluded because national ancillary databases are available only for the conterminous United States. Dieldrin data for the remaining 648 NAWQA sites were used in model development. A single fish sample from each site was used. At most sites, only one fish sample was collected. For sites where multiple fish samples were collected, preference was given to the fish sample(s) collected on the same date, or as close as possible in time, to bed sediment sampled at the same sites. (The availability of coincident sediment data was a priority because dieldrin concentrations in bed sediment were considered as potential explanatory variables in alternative regression models not discussed in this report.) If multiple fish samples were collected on the same date, preference was given to common carp samples (if available) or otherwise to the taxon with the highest lipid content.

In summary, the national data set used to develop the dieldrin fish model contains data for 648 sites, each with one composite fish sample collected during 1992–2001. The data represent 59 different fish taxa, of which the most common are common carp (29 percent of fish samples) and white sucker (26 percent). The 648 streams used in model development represent a variety of land use and environmental settings (*table 1*).

Table 1. The 648 NAWQA stream sites used in model development, by major physiographic division and land use in the basin.

[Major physiographic division was determined by clipping the national physiographic region coverage (Fenneman and Johnson, 1946) to the basin boundaries for each site. National land use classifications were based on the enhanced 1992 National Land Cover Data Set (unpublished data described in Nakagaki and Wolock, 2005) as follows: agricultural sites contained >50% agricultural land use and ≤5% urban land use in the watershed; urban sites contained >25% urban and ≤25% agricultural land use; undeveloped sites contained ≤25% agricultural and ≤5% urban land use; mixed sites were all other combinations of agricultural, urban, and undeveloped land. Land cover classifications used to designate sites as agricultural included pasture/hay, row crops, small grains, fallow, orchards/vineyards/other. Land cover classifications used to designate sites as urban included low intensity residential, high intensity residential, commercial/industrial/transportation, urban/recreational grasses, and forested residential. —, none; %, percent]

Physiographic division	Agricultural land use (sites)	Mixed land use (sites)	Undeveloped land use (sites)	Urban land use (sites)	Total sites
Appalachian Highland	16	82	57	28	183
Atlantic Plain	51	32	10	9	102
Interior Highlands	4	5	8	—	17
Interior Plains	79	49	21	9	158
Intermontane Plateaus	18	11	15	2	46
Laurentian Upland	1	1	10	—	12
Pacific Mountain System	8	9	20	10	47
Rocky Mountain System	—	16	63	4	83
Total sites	177	205	204	62	648

Explanatory Variables

The dieldrin fish model developed in this study is an empirical regression model. A large number of variables that could reasonably affect, or indicate an influence on, the use or transport of aldrin and (or) dieldrin were considered as potential explanatory variables for estimating dieldrin concentrations in fish. Statistical procedures were used to select from among the potential variables on the basis of their ability to reproduce the observed dieldrin concentrations in fish.

Table 2 lists the potential explanatory variables considered for inclusion in the regression model. Values of source and watershed variables for individual fish sampling sites were determined by applying basin boundaries for that site to various national ancillary data sets. Some potential explanatory variables were based on characteristics of the fish sample; most potential variables were based on watershed characteristics, including estimates of (or surrogates for) the use of aldrin and dieldrin in various applications. For ease of identification, potential explanatory variables are denoted in bold in the following section.

Fish Sample Characteristics

Three types of explanatory variables are based on characteristics of the fish sample.

- **Fish lipid content**: This is a measured parameter. Dieldrin concentrations in whole fish were significantly correlated with fish lipid content ($p \leq 0.0001$), although lipid content explained a relatively small percentage of the variance ($R^2 = 0.12$). This occurs because the sites sampled represented a wide variety of source conditions, ranging from reference streams (where pesticide use is expected to have been low) to streams in heavily agricultural or urbanized areas with histories of high pesticide use. Differences in source conditions (and therefore exposure) are likely to overwhelm differences in fish lipid content. Conversely, lipid content would be a confounding variable when investigating the relationship between dieldrin concentrations and source conditions. Therefore, fish lipid content was considered as a potential explanatory variable in the regression model. Because lipid content was retained as a variable in the final dieldrin model, prediction for unmonitored streams using the model therefore requires the assumption of a specified lipid content (see the next section "Extrapolation to Unmonitored Streams"). Alternatively, the model could have been developed for lipid-normalized dieldrin concentrations. However, the latter approach would have the following drawbacks: (1) it would reduce the variability explained by the model by ignoring the role of lipid content; and (2) the response variable then would be in units of micrograms of dieldrin per gram of lipid, whereas most fish data are expressed on a wet weight basis.
- **Taxon dummy variables**: Differences in contaminant residues between different species of fish have been widely reported in the literature. These have been attributed in part to differences in age, body size, lipid content, feeding habitat, and trophic state of the different species; but in many cases, there was no clear-cut or single explanation for the observed differences (Nowell and others, 1999). The NAWQA fish data set contained multiple species of fish. Although the model includes lipid content as an explanatory variable, this may not be sufficient to account for differences in contaminant levels among species. Therefore, a dummy variable was created for each of the eight most commonly sampled taxa (for example, common carp), and these eight dummy variables were considered as potential explanatory variables. Each taxon dummy variable has a value of either "1" (if the sample belongs to that taxon) or "0" (if it does not).
- **Time**: The elapsed time from 1966 (which represents peak historical agricultural use of aldrin and dieldrin) to the sampling year was considered as a potential explanatory variable. Because dieldrin trends in fish are declining nationally (Schmitt and others, 1999; Nowell and others, 1999), it is possible that samples collected earlier in the ten-year NAWQA sampling period (1992–2001) may contain higher levels than those collected later. Therefore, elapsed time was considered as a potential explanatory variable.

Table 2. Potential explanatory variables considered in developing the regression model for dieldrin in whole fish.

[Data sources are cited in text. km, kilometer; lb/mi^2, pound per square mile; m, meter]

Phase I—Source Variables

Past agricultural use—from combining county pesticide use data (1966) with mapped agricultural land from 1970s GIRAS land use and land cover data, at the 100-m resolution.
- Aldrin + dieldrin use intensity in agriculture, as amount applied per basin area (pounds per square mile).
- Aldrin + dieldrin use intensity in agriculture, as amount applied per area of agricultural land in basin (pounds per square mile).

Land use—from 1970s GIRAS land use and land cover data (polygons converted to 100-m resolution grid cells).
- Total agricultural land (percentage of basin area)—sum of cropland and pasture, orchards, groves, vineyards, nurseries, ornamental horticultural areas, confined feeding operations, other agricultural land.
- Cropland and pasture (percentage of basin area).
- Orchards, groves, vineyards, nurseries, ornamental horticultural areas (percentage of basin area).
- Total urban land (percentage of basin area)—sum of residential, commercial and services, industrial, transportation, communication, utilities, mixed urban and built-up land, other urban or built-up land.
- Residential land (percentage of basin area).
- Commercial-industrial land (percentage of basin area)—sum of commercial and services, industrial.
- Transportation land (percentage of basin area)—sum of transportation, communication, and utilities.
- Total forested land (percentage of basin area)—sum of deciduous, evergreen, and mixed forest land, and forested wetland.
- Total undeveloped land (percentage of basin area)—sum of deciduous, evergreen, and mixed forest land; forested wetland; nonforested wetland; herbaceous, shrub and brush, and mixed rangeland; herbaceous, shrub and brush, wet, and mixed tundra; dry salt flats; beaches; sandy areas other than beaches; bare exposed rock; bare ground; mixed barren land; streams and canals; lakes; reservoirs; perennial snowfields; glaciers.

Other surrogates for past urban use
- Mean 1990 population density (people per square kilometer)—from 1990 Census Population and Housing at the block group level processed at the 1-km resolution.
- Weighted termite-urban score, 1970s—urban land use (1970s GIRAS) weighted by termite density zone at the 100-m resolution.
- Weighted termite-urban score, adjusted to include new residential data from 1990s—urban land use (1970s–1990s GIRAS) weighted by termite density zone at the 100-m resolution.
- High termite-urban score, 1970s—urban land use (1970s GIRAS) in the very heavy plus moderate-to-heavy zones of termite growth at the 100-m resolution.
- High termite-urban score, adjusted to include new residential data from 1990s—urban land use (1970s–1990s GIRAS) in the very heavy plus moderate-to-heavy zones of termite growth at the 100-m resolution.

Characteristics of Sample
- Lipid content (percent)
- Elapsed time since 1966 (years)

Phase II—Taxon Dummy Variables

Taxon
- White sucker
- Common carp
- Largemouth bass
- Largescale sucker
- Longear sunfish
- Rock bass
- Brown trout
- Sculpins

Phase III—Watershed Characteristics

Physical watershed characteristics—delineated from sources ranging in scale from 1:24,000 to 1:100,000.
- Basin area (square kilometers).
- Mean basin slope (percent)—derived from National Elevation Data (NED) at 100-m resolution.
- Mean basin aspect (degrees)—derived from NED at 100-m resolution.
- Site elevation (meters)—derived from NED at 100-m resolution.
- Mean basin elevation (meters)—derived from NED at 100-m resolution.
- Road density (kilometers per square kilometer)—derived from Census Bureau 2000 Topologically Integrated Geographic Encoding and Referencing (TIGER) roads.

Table 2. Potential explanatory variables considered in developing the regression model for dieldrin in whole fish—Continued.

[Data sources are cited in text. km, kilometer; lb/mi^2, pound per square mile; m, meter]

Soil Properties—from the State Soil Geographic (STATSGO) data (polygons converted to 1-km resolution grid cells) unless noted otherwise.
- Mean available water capacity (inches per inch).
- Mean sand, silt, and clay composition (percent).
- Mean percent of soils classified as soil hydrologic groups A and B (indicator of well-drained soil).
- Mean percent of soils classified as soil hydrologic groups C, D, or C/D (indicator of poorly drained soil).
- Mean soil organic matter content (percent).
- Mean soil permeability (inchs per hour).
- Mean R factor from the Universal Soil Loss Equation, mean annual rainfall erosivity, 1971–2000 (from NOAA station data analyzed and interpolated to 1-km cells).
- Mean K factor for the uppermost soil horizon from the Universal Soil Loss Equation, soil erodability.

Hydrologic parameters
- Dunne overland flow—percentage of stream flow derived from runoff resulting from precipitation on saturated soil, generated from TOPMODEL (Wolock, 1993).
- Horton overland flow—percentage of stream flow derived from runoff resulting from precipitation exceeding soil infiltration rate, generated from TOPMODEL (Wolock, 1993).
- Soil contact time—Time required for precipitation to travel through subsurface flow paths before entering stream (years).
- Topographic wetness index, expressed as "ln(m)" (natural log of meters), derived from a 1-km resolution representation of the U.S. Geological Survey's Digital Elevation Model (U.S. Geological Survey, 1993).
- Mean annual runoff, 1951–1980 (millimeters).

Weather and Climate—estimated using the Parameter-elevation Regressions on Independent Slopes Model (PRISM) (Daly and others, 1997).
- Mean annual precipitation, 1961–1990 (millimeters).
- Mean annual temperature, 1961–1990 (degrees Celsius).
- Mean annual potential evapotranspiration—estimated using the PRISM temperature with the Hamon PET equation (Hamon, 1961).

Agricultural Management Practices—from 1992 Natural Resources Inventory (NRI) and agricultural land classifications from the National Land Cover Dataset 1992.
- Irrigation (percentage of agricultural land in watershed).
- Artificial drainage (percentage of agricultural land in watershed).
- Conservation tillage (percentage of agricultural land in watershed).

Phase IV—Regional Dummy Variables

USDA Farm Production Region—Each site was assigned to whichever region contained the largest percentage of the basin area.
Appalachians
Corn Belt
Delta
Lake States
Mountain States
Northeast
Northern Plains
Pacific States
Southern Plains
Southeast

Past Sources of Aldrin and Dieldrin

Several variables representing past sources of aldrin and dieldrin to the basin were considered. Because it is impossible to distinguish dieldrin residues that originated from application of aldrin from those originally applied as dieldrin, agricultural use estimates were calculated for the sum of aldrin plus dieldrin. Because quantitative estimates of urban pesticide use are not available, various urban-use surrogates were used as potential explanatory variables, as described below.

- **Agricultural use intensity**: The historical agricultural use intensity of aldrin plus dieldrin in the basin was estimated from regional application rates developed using U.S. Department of Agriculture (USDA) data on pesticide use by farmers in 1966, and harvested crop acreage from the 1964 and 1969 Census of Agriculture (Gail P. Thelin, U.S. Geological Survey, written commun., July 13, 2004). Briefly, regional pesticide-use-by-crop coefficients were developed for aldrin and dieldrin on 13 crops in each of the 10 USDA farm production regions using national and regional estimates of pesticide use in 1966 (Eichers and others, 1970) and total harvested crop acreage from the 1964 or 1969 (depending on the crop) Census of Agriculture (Economic Research Service, 1999). Next, these regional pesticide-use-by-crop coefficients were used to calculate use estimates for aldrin plus dieldrin in all counties in the conterminous United States that have agricultural land. The resulting estimated historical use of aldrin plus dieldrin in agriculture is shown in *figure 1* for the conterminous United States. Finally, historical agricultural estimates of aldrin and dieldrin use within the basin (watershed) were calculated for each sampling site by combining the basin boundaries for the site with county-based use estimates of pesticide use and 1970s Geographic Information Retrieval and Analysis System (GIRAS) land-use information for the basin (described further in the next paragraph). Two agricultural use intensity variables were calculated and considered in model development: one was computed per area of agricultural land in the basin and one per total basin area.

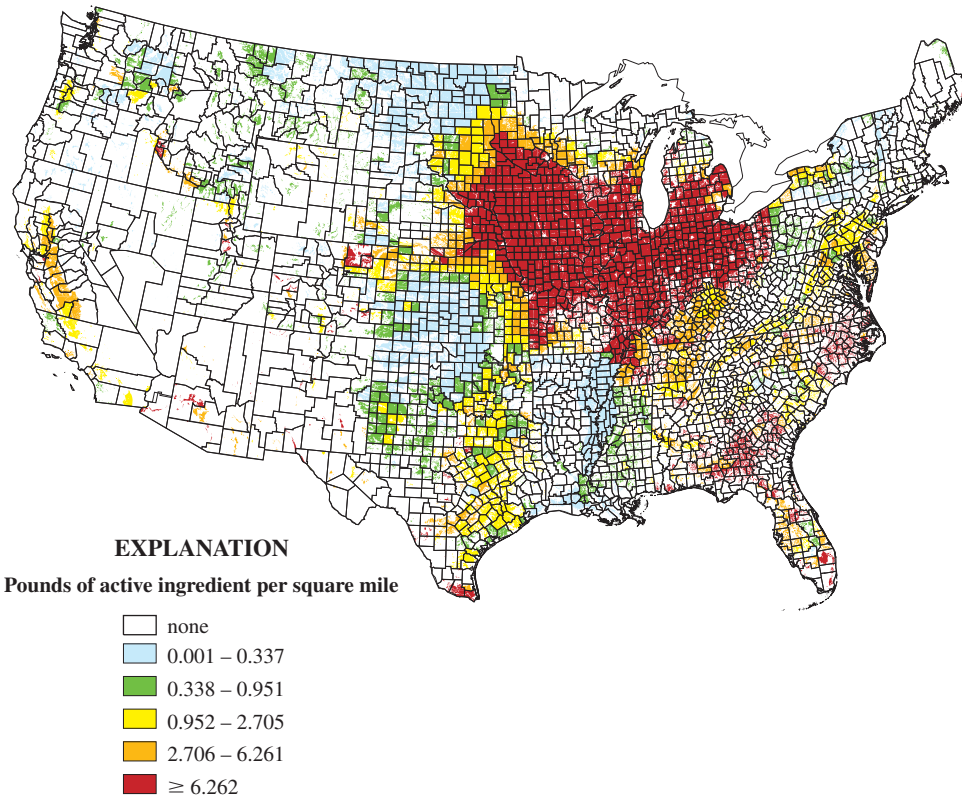


Figure 1. Estimated historical use of aldrin plus dieldrin in agriculture (1966) in the conterminous United States (from Gail P. Thelin, U.S. Geological Survey, unpub. data, 2005).

- Other agricultural source variables: The general agricultural source variables considered were the **percentage of cropland and pasture** in the basin, and the **percentage of total agricultural land** in the basin (including cropland, pasture, orchards, vineyards, and miscellaneous agriculture), both about 1970. These percentages were derived from the USGS's Land Use and Land Cover (LULC) data, which are based on aerial photography taken in the 1970s to mid-1980s (Fegeas and others, 1983; data from U.S. Geological Survey, 1998, as enhanced by Price and others, 2003). The LULC data are also known as Geographic Information Retrieval and Analysis System (GIRAS) data, which refers to the digital file storage format and software system used to develop the LULC data set (Mitchell and others, 1977). (In the remainder of this report, this data set will be referred to as the GIRAS data set.) Polygons of land use and land cover in the GIRAS data set were converted to grid cells at the 100-m resolution. (In this report, the term "grid" refers to a raster-based geodata set consisting of rows and columns of cells, each of which is referenced by its geographic x- and y-coordinates and has an associated data value.) The land use and land cover categories are based on a classification system developed by Anderson and others (1976).
- **Weighted termite-urban score**: Aldrin and dieldrin were used in subterranean termite control through the late 1980s (this use was voluntarily canceled by industry in 1987). Although EPA published national-scale estimates of pesticides used in subterranean termite control in 1980 and 1985 (U.S. Environmental Protection Agency, 1983; Esworthy, 1987), historical data on termiticide use at the county or even regional scale are not available. Therefore, a weighted termite-urban score was developed as a surrogate variable to represent the relative probability of past aldrin and dieldrin use in subterranean termite control. This score was calculated from the percentage of urban land use in the basin (1970s GIRAS data) and the relative hazard of subterranean termite infestations (Beal and others, 1994). Combined, these factors are expected to reflect the extent of past termiticide use, because termiticides were most likely applied to buildings in parts of the country where termites occurred. The national distributions of these two factors are shown in *figure 2*. The surrogate variable was calculated as follows: the urban land (as a percentage of basin area) located within each of four zones of termite density (none-to-slight, slight-to-moderate, moderate-to-heavy, and very heavy) was determined; the percentage of urban land within each zone was multiplied by a weighting factor for that zone (1 for none-to-slight, 2 for slight-to-moderate, 3 for moderate-to-heavy, and 4 for very heavy); and the four weighted percentages were summed to get a weighted termite-urban score. Alternative types of termite-urban scores were tried, including summing urban land in only the two highest zones of termite growth, and (or) calculating a termite-urban score using a version of the 1970s GIRAS land use data in which new residential areas were reclassified using 1990 census information (Hitt, 1994). However, the weighted termite-urban score that was based on the original 1970s GIRAS data had the strongest correlation with the response variable.
- Other urban source variables: Besides termite control, aldrin and dieldrin were used in other urban applications such as moth proofing, wood preservation, and control of disease vectors (such as mosquitoes) (U.S. Environmental Protection Agency, 1992a). Because quantitative use data are not available for such urban applications, two general urban-use surrogate variables were considered in model development: **the percentage of urban land use** in the basin and **population density** (people per square kilometer). The percentage of urban land use in the basin (which included residential, commercial, industrial, transportation, communication, and utilities) was determined from 1970s GIRAS data, as described above for agricultural land use variables. Population density data were derived from a 100-m resolution grid of the census block group boundaries and population counts (Price, 2003) derived from the 1990 Census of Population and Housing data files (U.S. Bureau of the Census, 1991).
- Minimal use variables: Additional source variables considered were the **percentage of total forested land** and the **percentage of total undeveloped land** in the basin, both about 1970, which each represent areas where past use of aldrin and dieldrin is expected to be minimal. Total forested land included deciduous, evergreen, and mixed forest land, and forested wetland. Total undeveloped land included total forested land, rangeland (herbaceous, shrub and brush, and mixed), water (streams, canals, lakes, reservoirs), nonforested wetland, tundra (herbaceous, shrub and brush, mixed, wet), barren land, beaches, other sandy areas, dry salt flats, bare rock, glaciers, and perennial snowfields. These land use percentages were determined from the 1970s GIRAS land use data.

Other Watershed Characteristics

Numerous watershed-based explanatory variables were considered, including soil characteristics, hydrologic parameters, climate variables, and agricultural management practices in the basin. Only those variables that were available for the entire conterminous United States were used. In addition, regional dummy variables were used to represent any regional differences not accounted for by the source- and watershed-based variables considered in the model.

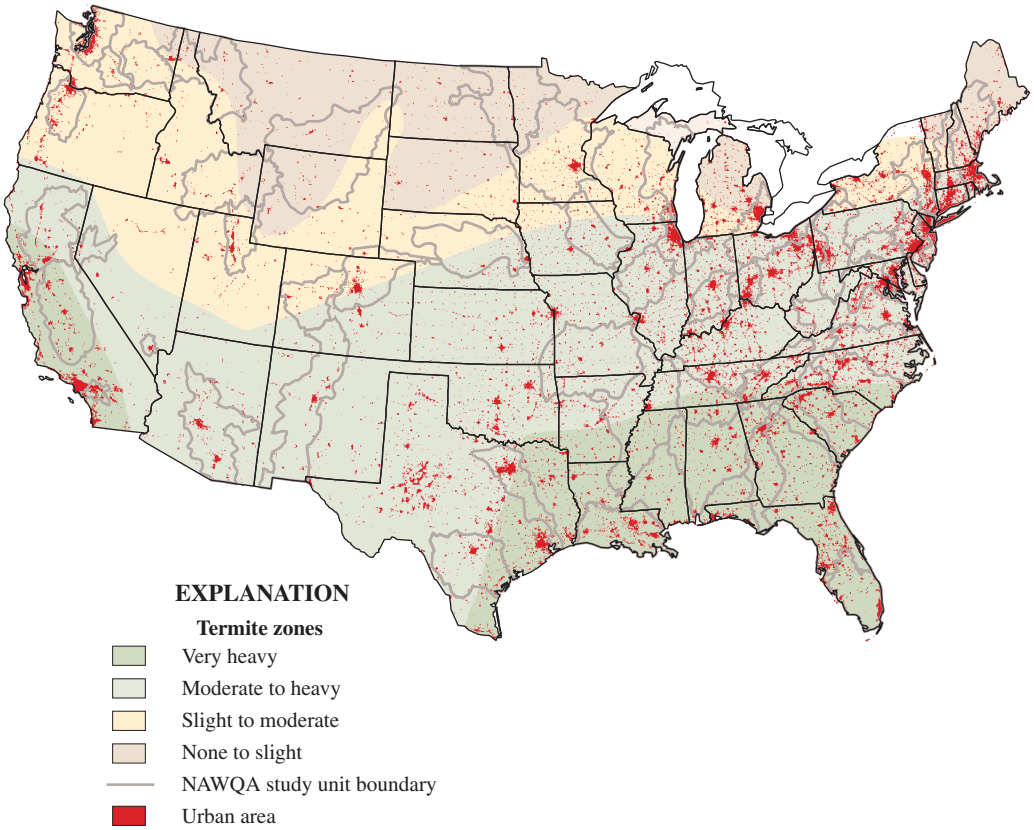


Figure 2. Urban areas in the United States (U.S. Geological Survey, 1986) shown superimposed on the relative hazard of subterranean termite infestations (Beal and others, 1994). [For urban areas to be visible on the map, the 100-m resolution urban areas were generalized to the 3-km resolution; only those 3-km grid cells consisting of at least 10 percent urban land use are colored on the map. km, kilometer; m, meter]

- Physical watershed characteristics included **watershed drainage area** (square kilometers), **site elevation** (meters), **mean basin elevation** (meters), **mean watershed slope** (fraction), **mean watershed aspect** (degrees), and **mean road density** (kilometer per square kilometer). Mean basin elevation, slope, and aspect were computed from USGS digital elevation model data (U.S. Geological Survey, 2000). Road density was derived from Census Bureau 2000 Topologically Integrated Geographic Encoding and Referencing (TIGER) roads (Geolytics, Inc., 2004).
- Soil characteristics included mean values within the basin of **soil organic matter content** (percent by weight); **clay**, **silt**, and **sand content** (percent) in the soil; **soil permeability** (inches per hour); and **available water capacity** (inches per inch). In addition, two soil type variables, **well-drained** and **poorly drained** (percent), were based on soil hydrologic groups (HG) A, B, C, and D (U.S. Department of Agriculture, 1999) by combining them into two groups (HGA+HGB and HGC+HGD) to differentiate the watersheds in terms of soil type. Higher values of HGA+HGB imply that soils in the watershed are generally well drained, with a relatively low potential for surface runoff, whereas higher values of HGC+HGD imply poorer drainage and generally more surface runoff. The previous soil characteristics were based on national 1-km resolution geodata sets produced by Wolock (1997) from data in the State Soil Geographic (STATSGO) database (National Resources Conservation Service, 1994). Two additional soil-related variables were the mean **R factor** (rainfall erosivity) and mean **K factor** (soil erodibility) values for the uppermost soil horizon from the Universal Soil Loss Equation. Mean R factors were estimated from mean annual (1971–2000) R factor values at the 1-km

grid cell resolution. The mean annual R factor values were estimated by the Illinois State Water Survey and interpolated to a 2.5 min resolution grid by the Spatial Climate Analysis Service at Oregon State University using the Parameter-elevation Regressions on Independent Slopes Model (PRISM) methodology (Greg Johnson, U.S. Department of Agriculture, written commun., May 12, 2003). K factor values were averaged for each State Soil Geographic (STATSGO) soil map unit for the uppermost soil horizon, and then spatially represented in a 1-km resolution grid (David M. Wolock, U.S. Geological Survey, written commun., May 9, 2003).

- Hydrologic parameters included **percent Dunne overland flow** (the percentage of total streamflow contributed by overland flow resulting from precipitation falling on saturated soil) and **percent Horton overland flow** (the percentage of total streamflow contributed by overland flow resulting from precipitation falling at a rate exceeding the infiltration rate of the soil), **topographic wetness index** (natural log of meters), **index of subsurface flow** contact time (days) and 1951–1980 **mean annual runoff** (inches). The Dunne and Horton overland flow parameters at the 5-km resolution (Wolock, 2003a,b) were generated using a national-scale application of the topography model, TOPMODEL rainfall-runoff model (Wolock, 1993). The topographic wetness index was derived from the U.S. Geological Survey's Digital Elevation Model (U.S. Geological Survey, 1993; Verdin and Greenlee, 1996) at the 1-km resolution using the single flow direction algorithm method developed by Wolock and McCabe (1995). Subsurface contact time was developed from the 100-m resolution Digital Elevation Model (U.S. Geological Survey, 1993) combined with soil characteristics (Wolock, 1997) using methods described in Wolock and others (1997). Mean annual runoff was created from Gebert and others (1987).
- Climatic variables include **mean annual temperature** (degrees Celsius) and **precipitation** (millimeters) (1961–1990), and **potential evapotranspiration** (millimeters). Mean annual temperature and precipitation are from climate data developed by the Spatial Climate Analysis Service at Oregon State University (Daly and others, 2002) using PRISM, which is an analytical model that uses point measurements and a Digital Elevation Model (DEM) to generate gridded estimates of monthly and annual precipitation, as well as other climatic parameters. Potential evapotranspiration was estimated using temperature data from PRISM with Hamon's equation for potential evapotranspiration (Hamon, 1961). All climatic data variables were processed as grid cells at the 1-km resolution.
- Agricultural management practice variables included the mean percentage of the basin under selected agricultural management practices. These variables were derived from a digital map of the basin overlaid with a national geodata set of agricultural practices integrated with mapped agricultural land. The selected agricultural practices consisted of **conservation tillage** (Wieczorek, 2004a), **irrigation** practices (Wieczorek, 2004b,c,d), and **artificial drainage** (Wieczorek, 2004e,f). The data sources used by Wieczorek in these publications were the 1992 National Resources Inventory (U.S. Department of Agriculture, 1995) and the U.S. Geological Survey's National Land Cover Dataset 1992 (Vogelman and others, 2001).
- Ten **regional dummy variables** were used, which correspond to the ten USDA farm production regions, as defined for 1971 agricultural pesticide use (Andrilenas, 1974). Each site was assigned to one primary region, defined as the region containing the largest percentage of the watershed area for that site. Each region is represented by a 1/0 dummy variable; for example, a site with most of its basin in the Corn Belt region will have a value of "1" assigned to the 'Corn Belt' dummy variable and a value of "0" assigned to the nine remaining regional dummy variables. These regional dummy variables were considered in case there were regional influences on pesticide use or transport that were not represented by the pesticide use estimates and surrogates and the watershed characteristics previously added to the model.

Development of Regression Model

The response variable in the regression model is the log-transformed concentration of dieldrin in one whole-fish composite sample from each NAWQA site. Because each sample is a composite of 5–8 individual fish from a given site, the measured concentration represents an approximate "average" for that site. Explanatory variables were tested in four phases: those representing (phase I) sources of the pesticide, fish lipid content, and time; (phase II) taxon of the fish sample; (phase III) characteristics of the watershed; and (phase IV) geographic region. Statistical methods were used for selecting and transforming variables (as needed) for inclusion in the model and in estimating coefficients of the regression equation.

Regression Models

Many of the dieldrin concentration values used as the response variable (62 percent of samples) were reported as less than a detection threshold (that is, dieldrin was not detected), resulting in censored data. With censored data, conventional least-squares regression methods will yield biased and inconsistent estimates (Judge and others, 1985, p. 780). Therefore, Tobit regression methods (also called maximum likelihood methods) were used for developing the model. These can be used to obtain parameter estimates of a censored linear model when the regression residual errors are independent, identically and normally distributed, and with a mean of zero and finite variance (Maddala, 1983). Maximum likelihood methods

implemented in the survreg procedure (Therneau, 1999) in the statistical analysis program S-PLUS (Insightful Corporation, 1999) were used to estimate the parameters of the regression models.

Measures of goodness of fit used in conventional least squares analysis—such as the standard deviation of the residual error, or the coefficient of multiple determination (R^2)—cannot be computed for the Tobit regression model. The standard deviation of the residual error is alternatively referred to as the scale parameter in maximum likelihood estimation. Estimates of the scale parameter from the maximum likelihood procedure 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 the degree of censoring. In this report, biased estimates of the standard deviation of the residual error are referred to as "scale" in tables and figures. As an alternative to a conventional R^2, a pseudo-R^2 (pR^2) suitable for use in Tobit regression model was calculated using the method of Laitila (1993). As with conventional R^2, this pR^2 ranges from 0 to 1 and is an estimate of the proportion of the variation in the response variable explained by the regression model ("0" indicates no variation is explained; "1" indicates all variation is explained).

Transformation of Response and Explanatory Variables

The maximum likelihood methods used for estimating the parameters of the models require several assumptions. The relation between the variables must be linear in the parameters, and the residual error must be identically and normally distributed. Departures from these assumptions can result in estimates of model coefficients with considerable error. One means of addressing departures from model assumptions is through transformations of the response and (or) explanatory variables. Various transformations were considered to minimize departures from the assumptions of the maximum likelihood methods used. For the response variable, a logarithmic transformation was used. For explanatory variables, the logarithmic, square-root, and square transformations of each variable were considered, as well as the untransformed value. (For variables that have zero as a possible value, 1 was added to variable values before applying the logarithmic transformation.)

Because the response variable is a logarithmic transformation, concentrations predicted by the model (after retransformation) are the median concentrations expected for sites that have a given set of explanatory values, rather than mean concentrations. For example, for 10 basins with identical explanatory variables, the model will predict the same dieldrin concentration for all 10 basins. In actuality, the true concentrations are expected to be less than the predicted value at half of the 10 basins and greater than the predicted value at the other half. Thus, the "median" response refers to the distribution of concentrations among sites. In this study, predicted concentrations were not adjusted for transformation bias (Bradu and Mundlak, 1970; Duan, 1983) because estimates of site medians were considered appropriate for the study objectives.

Selection of Explanatory Variables

Potential explanatory variables were considered and added to the model in four phases, as described below.

- Phase I: Source variables, lipid content, time—Estimates of historical pesticide use in agriculture, various surrogates for agricultural and urban pesticide use, and surrogates for minimal use were added to represent past pesticide use in the basin. In addition, fish lipid content (a measured parameter) was included in phase I because of its significant correlation with dieldrin residues in fish ($p \leq 0.0001$). Elapsed time was also considered during this phase. These variables were considered first because they were expected to be the most important in terms of explaining the variance in concentrations among sites.
- Phase II: Taxon—Dummy variables representing the most commonly sampled fish taxa were considered as explanatory variables, in case lipid content (added to the model during phase I) did not completely account for differences in contaminant bioaccumulation among fish taxa.
- Phase III: Watershed variables—Additional watershed characteristics (soil properties, hydrologic parameters, climate variables, and agricultural management practices) were considered. These variables have potential to affect contaminant fate and transport and, therefore, may contribute to differences in contaminant concentrations measured in fish from different watersheds, once differences between sources and fish taxa have been taken into account.
- Phase IV: Region—Dummy variables representing USDA farm production regions were considered in case there were regional differences not accounted for by source estimates and watershed characteristics.

During each phase, a stepwise procedure that was based on the Akaike Information Criterion (AIC) was applied (Venables and Ripley, 1999, p. 186–188). This procedure is similar to stepwise regression, except that the AIC was used to select variables for inclusion in the model. AIC is a function of the maximized likelihood and the number of parameters in the model; it balances model goodness of fit with the number of parameters needed to achieve that fit. AIC attempts to quantify the concept of model parsimony, in that simpler models are favored over complex ones unless a complex one substantially improves the fit. During each phase of model development, a stepwise Akaike-based procedure (stepAIC) was used that computed the AIC for multiple (alternative) models and selected the one that best described the response variable data. Candidate variables not selected by the stepAIC procedure

were eliminated, and the stepAIC procedure was rerun until no further variables were eliminated from the model. This process, described below in more detail for phase I, was repeated during each of the four phases of model development.

To select variables during phase I: (1) The stepAIC procedure was run to identify candidate phase I variables that best describe the response variable data. (2) For each variable identified as important by the stepAIC procedure, that variable and three alternate transformations (the logarithm, square, and square root of the variable) were each plotted against the response variable and examined to see if the relation was approximately linear, and if the variance was relatively constant across the range of values of the variable. The stepAIC procedure also was run on all four versions of the variable (untransformed and the three transformations) to determine whether transformation(s) improved the fit. For each untransformed explanatory variable that was selected by the stepAIC procedure, the following was performed: (a) if the relationship between the response variable and that untransformed explanatory variable showed non-normality or nonconstant variance, but the problem was corrected by one or more transformations of that explanatory variable, then that transformation was substituted for the original, untransformed variable in the model; (b) if the relationship between the response variable and that untransformed explanatory variable was highly skewed, showed non-normality or nonconstant variance, and these problems were not improved by transformation, then that variable was dropped from consideration in the model. (3) The stepAIC procedure was run again on the revised list of candidate variables. (4) Any variables that were not selected by the stepAIC procedure were eliminated and the stepAIC procedure was run again. (5) Step 4 was repeated until no further phase I variables were eliminated by the stepAIC procedure. (6) Any variables that were selected by the stepAIC procedure, but had p-values greater than 0.05, were dropped from the model. Prior to this step, some variables with p-values greater than 0.05 may have been selected, and some variables with p-values less than 0.05 excluded, by the stepAIC procedure because this procedure relies on AIC (not on p-values) to select variables. During this step, any variables with p-values greater than 0.05 were dropped to simplify the model and to limit the overall number of variables in the model. (7) Occasionally the stepAIC procedure selected two or more redundant variables, such as two transformations of the same variable, or two variables that represented the same or overlapping ancillary data (such as total forested land and total undeveloped land, or two termite-urban score variables calculated in different ways); in such cases, the variable with the strongest correlation was selected, and the redundant variable(s) were eliminated. At this point, phase I ended and phase II began.

Beginning with phase II, each phase started with the best model from the preceding phase, and then the new set of variables was added and tested. The seven steps listed above were repeated during each of phases II through IV (except that dummy variables in phases II and IV were not transformed). During the later stages of model development, sometimes variables selected during previous phases were no longer retained (selected) by the stepAIC procedure; that is, addition of new explanatory variables to the model caused some previously selected variables to be eliminated from the model.

After phase IV, a subsampling technique was used to reduce the number of variables and to eliminate overfitting. The AIC-based selection procedure was run on a random subset of 50 percent of the dieldrin data, and a record was made of any variables dropped from the model (that is, no longer selected by stepAIC). This random subsampling procedure was repeated 50 times, and any variables that were routinely dropped (that is, dropped in at least 20 percent of replicate runs) were eliminated from the model. This subsampling approach was repeated with the abbreviated model, and again variables were eliminated if they were dropped in 20 percent or more of the replicates. The subsampling approach was repeated until no variables were routinely dropped. The subsampling routines reduced the final dieldrin model from 13 to 6 variables.

Analysis of Model Fit

Condition indices and variance decomposition proportions were used to detect colinearity among variables. One variable was eliminated from the dieldrin model because of colinearity. Diagnostics for censored regression (Escobar and Meeker, 1992) available in the survreg procedure in the S-PLUS program (Insightful Corporation, 1999) were used to identify influential observations and to aid in variable selection. These diagnostics included case-weight perturbations (a measure of the effect that dropping an observation has on the inference), response perturbations (a measure of the effect that errors in the observed response variable has on the inference), shape parameter perturbations (a measure of the effect that unequal variances of the response variable have on the inference), deviance residuals (a residual transformed to a log-likelihood scale—the sum of the squared deviance residuals totals to the deviance), and leverage (a measure of how much influence an observation has on the overall regression). For individual observations identified as high values for any of these diagnostics, the regression was rerun excluding this observation to determine the effect of the observation on the regression; however, in all cases, the Akaike selection procedure retained (that is, selected) all variables in the regression model, there was no change in the sign of the coefficients, and the pR^2 and scale values did not change markedly. Therefore, all observations were retained for model development.

Model Validation

The final model was cross-validated as follows: The NAWQA fish data used in model development were randomly divided into two subsets, a calibration data set (containing 70 percent of the data) and a validation data set (30 percent of the data). The model was first fit to the calibration data subset,

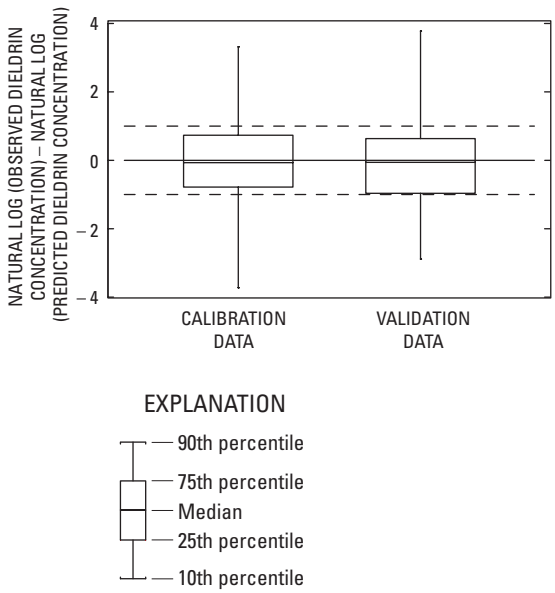


Figure 3. (A) Residual error developed from calibration data. Dashed lines are reference lines at y values of 1 and –1. Dieldrin concentrations are in units of micrograms per kilogram, wet weight.

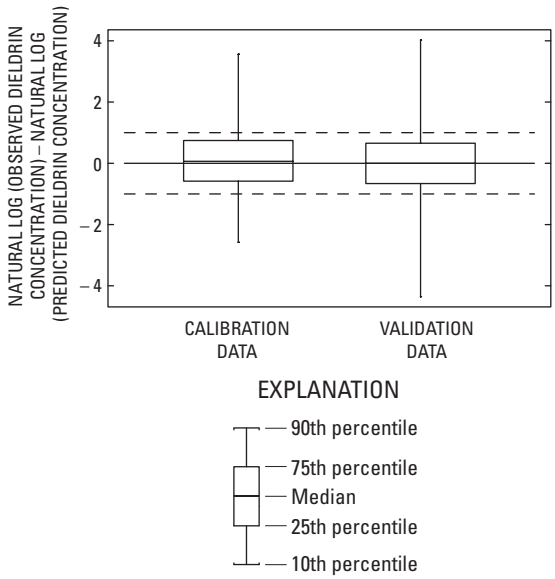


Figure 3. (B) Continued. Residual error developed from validation data.

and predicted values and residuals were computed for both the calibration data and validation data subsets. Then the process was reversed, and the model was fit to the validation data subset. There was little difference in the predicted and residual errors obtained from the two fitted models that were based on the different data subsets (*fig. 3)*.

Extrapolation to Unmonitored Streams

The regression model was used to predict the dieldrin concentration in whole fish in all streams in the conterminous United States, using national data sources to estimate values of model variables for unmonitored streams. Unlike the other explanatory variables, lipid content is not available from a nationwide data set; therefore, the extrapolation must assume a specified lipid content value that is appropriate to the objectives of the extrapolation.

Data Sources for Unmonitored Streams Nationwide

The U.S. streams were defined by the EPA River Reach File 1 (Nolan and others, 2003), which includes over 600,000 mi 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. This step required delineation of the entire watershed area draining to the downstream end of each stream reach and then computing area-weighted average values for each explanatory variable. The regression model then was used to predict the concentration of dieldrin in whole fish for the stream at the outlet of each watershed.

Lipid Content

Nationwide extrapolations were made at two different values for fish lipid content: (a) 4.2 percent and (b) 6.2 percent. These correspond to the average lipid content of (a) whole-body largemouth bass and (b) all whole fish that were sampled by the NCBP (1969–1986, from U.S. Geological Survey, 2004) and the NAWQA Program (1992–2001) combined, respectively. The 4.2 percent value is intended to represent the average lipid content of whole largemouth bass (a nationally distributed sport fish that is popular with recreational fishers), and the 6.2 percent value is intended to represent the average lipid content of all whole fish, in United States streams.

Uncertainty Analysis

A regression model provides an estimate of the response variable conditional for a given a set of values of the explanatory variables. If the form of the regression model reasonably describes the relation between the response and explanatory variables and the sample data are representative of the population of all data, then the model can be used to predict values of the response variable for additional sets of explanatory variable values not included among the sample data. In this study, each prediction represents the median, or "mid-point" estimate of the response variable (logarithm of dieldrin concentration in whole fish) for all unmonitored sites with an identical set of explanatory variable values. Thus, for a group of sites with identical explanatory variable values, there is a 50 percent probability that the measured concentration at any site exceeds the predicted concentration for that site; the actual value is expected to be lower than the model prediction at 50 percent of these sites and higher than the predicted concentration at 50 percent of these sites. Moreover, because the model was developed using measured concentrations in composite whole-fish samples, the model predictions also represent composite whole-fish samples. Finally, because the response variable is the logarithm of dieldrin concentration, the antilogarithm of the estimate from the regression model must be computed to obtain the median estimate of dieldrin concentration.

If the assumptions of the maximum likelihood method used to fit the model parameters pertaining to the residual error distribution are valid, then uncertainty in the estimated response variable values can be assessed by calculating prediction intervals. Upper and lower prediction limits are constructed around the estimated concentration from the model error (scale). Specifically, a prediction interval (*PI*) is computed as follows:

$$PI = \left[\log(C_{est}) \pm SD_{est} \times t\left(\frac{\alpha}{2}\right)\right] \tag{1}$$

where $\log(C_{est})$ is the logarithm of the estimated dieldrin concentration, $t(\alpha/2)$ is the point on student's t-distribution with a probability of exceedance of $\alpha/2$, and SD_{est} is the standard error of the estimated logarithm of concentration.

This means of assessing uncertainty is only approximate because it doesn't fully take into account the effect of censoring of the observed concentration data. A prediction interval represents the likelihood that the actual unknown concentration for an individual unmonitored stream falls within a specified interval of the predicted concentration. For a 90th percentile prediction interval, for example, only 10 percent of measured concentrations in fish samples (with the appropriate lipid content) collected from a given stream would be expected to fall outside that interval—5 percent would fall below the lower limit, and 5 percent above the upper limit. (That is, there is only a 5 percent probability that the measured concentration at any site exceeds the upper 90th percentile prediction interval limit and a 95 percent probability that it exceeds the lower prediction interval limit.) The wider the prediction interval, the greater the uncertainty in the estimated concentration.

Sometimes it is useful to be able to determine the probability that an estimated concentration exceeds a certain concentration, such as a guideline for protection of human health or wildlife. This probability is a function of the estimated concentration for a site and the overall model error, and it can be approximated by algebraically rearranging the formula for computing prediction interval limits (equation 1). The probability of exceedance (P_{exc}) is thus equal to

$$P_{exc} = \frac{\left[\log(C_g) - \log(C_{est})\right]}{SD_{est}} \tag{2}$$

where C_g is the guideline concentration. This concept is illustrated in *figure 4*. *Figure 4A* shows a hypothetical regression line that represents the estimated values of the response variable for the explanatory variable shown on the x-axis. For four selected values of the explanatory variable, f*igure 4A* also shows the error distribution about the regression line and the 90th percentile prediction interval limits. 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, as shown. The probability that the estimated response value will exceed the guideline is thus represented by the area under the error distribution curve above the point where the guideline intersects the curve. In the example shown in *figure 4*, the exceedance probabilities for the four selected values of the explanatory variable are (from left to right) 0.02, 2.3, 31, and 84 percent. (*Figure 4* is a 2-dimensional representation of a 3-dimensional figure. In three dimensions, the error distribution curve would be perpendicular to the paper and the prediction interval limits would lie on the curve.) *Figure 4B* shows the same regression and error relation after antilogarithms have been computed. Once this has been done, the error distribution around the regression line is no longer symmetric, nor does it have constant variance. Because the error distribution is no longer symmetric, the magnitude of the uncertainty is greater above the estimated concentration than it is below it. Likewise, because the error distribution no longer has constant variance, the magnitude of the uncertainty about the estimated dieldrin concentrations is much greater for larger concentrations than for smaller ones.

In this report, for both human health and wildlife guidelines, the probability that an unmonitored stream would exceed a given guideline was calculated and the national distribution of probabilities was mapped for unmonitored streams across the United States. This entailed assuming a default lipid content appropriate for each type of guideline, as discussed further in the next section.

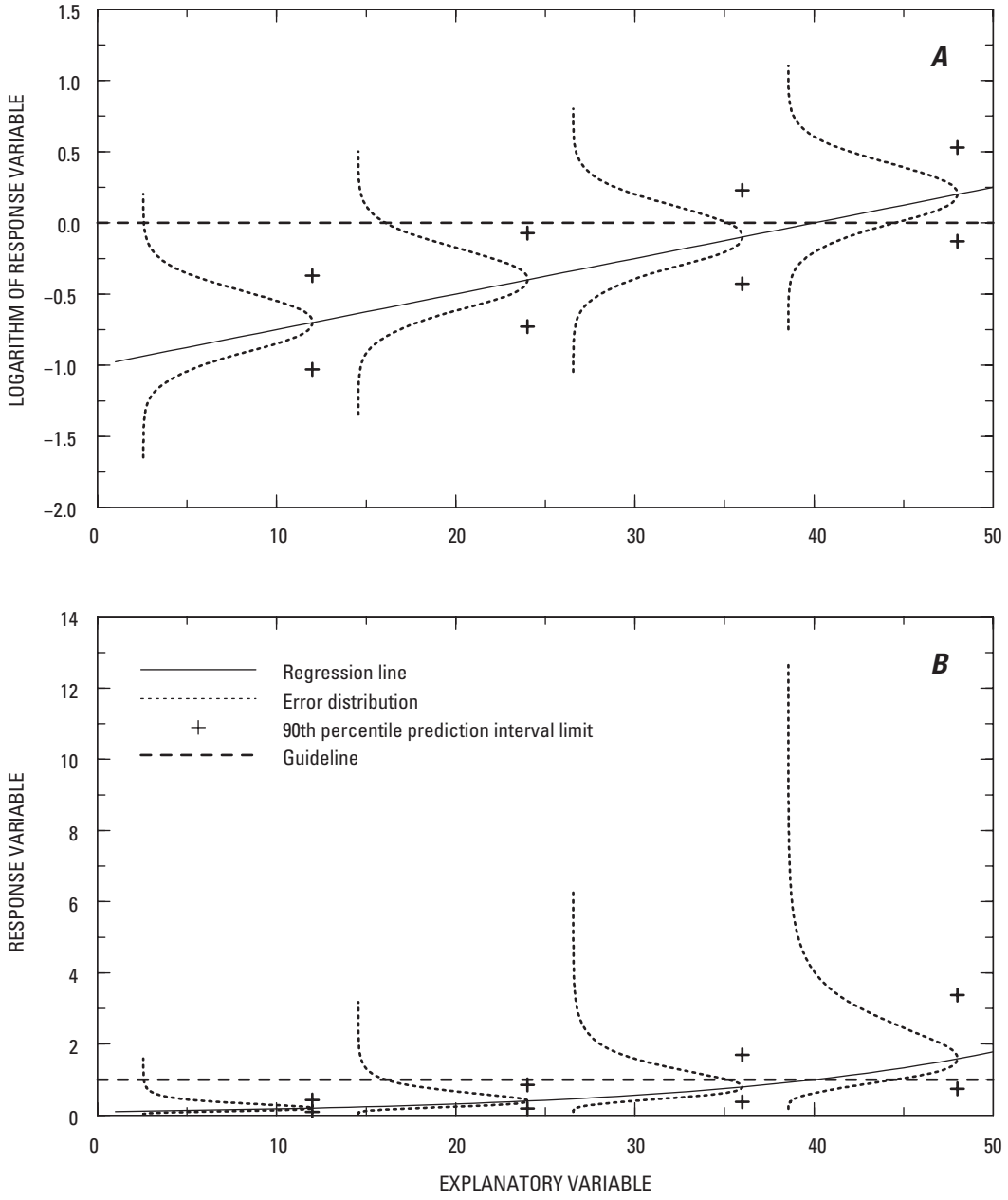


Figure 4. (*A*) A hypothetical regression for which the response variable is log-transformed, showing the regression line; the error distribution and the upper and lower 90th percentile prediction interval limits for four selected estimated values of the response variable; and a hypothetical guideline value. (*B*) Same as A, after taking the antilogarithm of the response variable.

Predicting the Potential Effects of Fish Consumption on Human Health and Wildlife

Potential Effects on Human Health

Potential effects on human health from fish consumption were assessed by (1) assuming a lipid content of 4.2 percent, and then (2) comparing predicted dieldrin concentrations in whole fish for unmonitored streams across the United States with the EPA recommended screening values for dieldrin in edible fish (U.S. Environmental Protection Agency, 2000). As noted above, the lipid content of 4.2 percent was chosen because it is the average lipid content of whole-body largemouth bass sampled by the NCBP and NAWQA Programs. This choice of taxon was considered appropriate from a human health perspective because largemouth bass is a common sport fish with a national distribution. This choice of tissue type (whole-body fish) is not optimal from a human health perspective because people tend to consume only the edible portion (fillets)—however, it is necessary because the model was developed from whole-body fish data, and it cannot be used to predict concentrations in fish fillets. Therefore, comparison with human-health guidelines (EPA screening values) is considered a screening-level assessment (discussed further in the section on "Significance to Human Health").

The EPA screening value for a contaminant in edible fish tissue is defined as the concentration of that contaminant in edible fish or shellfish that is of public health concern and that is used as a threshold value against which tissue residue levels of that contaminant in edible fish and shellfish can be compared (U.S. Environmental Protection Agency, 2000). For potential carcinogens such as dieldrin, EPA uses a conservative (protective) model of carcinogenesis that assumes there is no safe level of a carcinogen; instead, EPA calculates a screening value that corresponds to a maximum acceptable cancer risk (such as 1 in 100,000). If the maximum acceptable cancer risk decreases, the allowable dieldrin concentration also will decrease accordingly. The EPA screening values (SV) can be adjusted for different acceptable risk levels, as for example:

$$SV_{(1:10,000)} = 10 \times SV_{(1:100,000)} \tag{3}$$

where $SV_{(1:10,000)}$ is the screening value at a maximum acceptable cancer risk of 1 in 10,000, and $SV_{(1:100,000)}$ is the screening value at a cancer risk of 1 in 100,000.

For dieldrin, the screening value is 25 µg/kg for a cancer risk level of 1 in 10,000, and 2.5 µg/kg for a cancer risk level of 1 in 100,000. Both of these screening values (at the 1:10,000 and 1:100,000 cancer risk levels) are used in the extrapolation for human health (see "Significance to Human Health").

Potential Effects on Fish-Eating Wildlife

Potential effects on fish-eating wildlife were assessed by (1) assuming a default lipid content of 6.2 percent and (2) comparing predicted dieldrin concentrations in whole fish for unmonitored streams across the U.S. with the New York guideline for protection of fish-eating wildlife (Newell and others, 1987). As noted above, the default lipid content of 6.2 percent was chosen because it is the average lipid content of all whole fish sampled by the NCBP and NAWQA Programs. This assumption is appropriate because wildlife will eat a variety of fish taxa and they tend to consume the whole fish. The New York wildlife guideline of 120 µg/kg is intended to protect fish-eating birds and mammals from adverse effects other than cancer, such as mortality, reproductive impairment, and organ damage (Newell and others, 1987).

Dieldrin is a potential carcinogen (U.S. Environmental Protection Agency, 1992b), and the 120-µg/kg guideline does not address the potential incidence of cancer in wildlife populations from dieldrin exposure. Although Newell and others (1987) also determined a cancer risk guideline for dieldrin, this guideline assumes a maximum acceptable cancer risk of 1 in 100—that is, it assumes that a 1:100 cancer risk would result in virtually no reduction in a population from contaminant-induced cancer, an assumption that could not be fully justified at the time (Newell and others, 1987). Because it is not clear what effect this level of contamination and cancer risk would have on wildlife populations, the New York cancer-risk guideline was not used as a benchmark in this report.

Results and Discussion

The model development data set contains data for 59 different fish taxa at 648 stream sites. Site locations and measured dieldrin concentrations in whole fish are shown in *figure 5*. However, sites with the highest and lowest 10 percent of lipid levels were excluded from *figure 5* so that the range of lipid levels shown (2.3–10.4 percent, for n = 514 sites) more closely approximates the lipid levels used in the nationwide extrapolations.

Regression Model for Dieldrin in Whole Fish

The explanatory variables included in the dieldrin fish model, as well as their transformations and coefficients, are summarized in *table 3* .

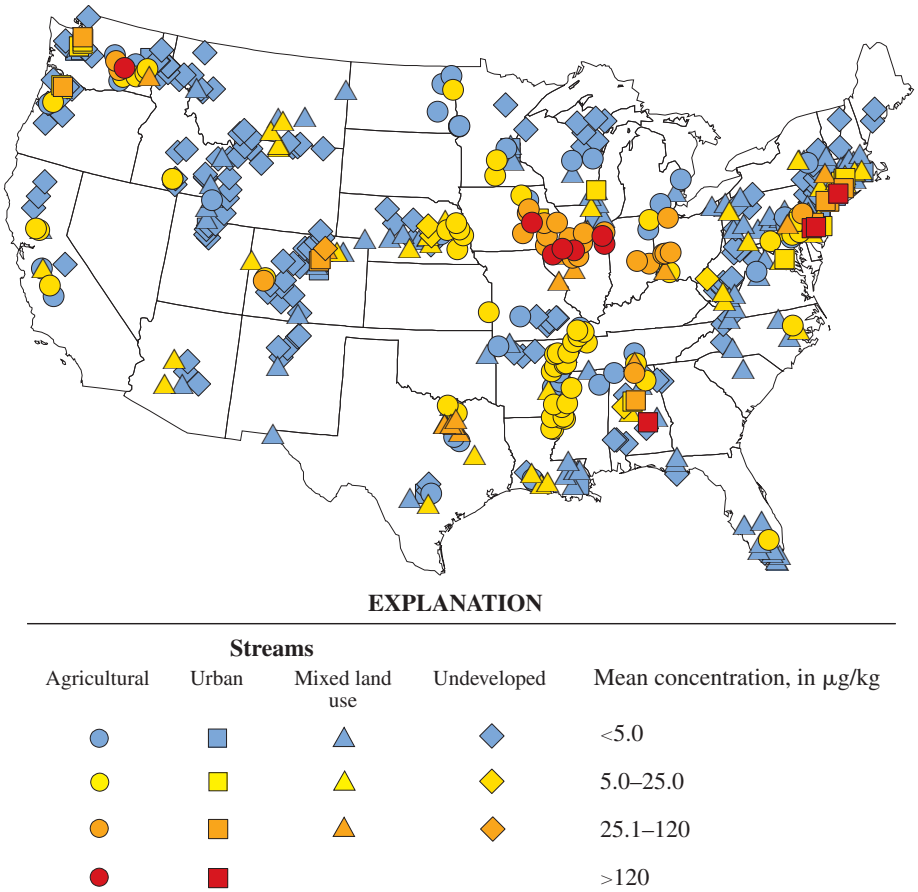


Figure 5. Measured dieldrin concentrations in fish at NAWQA sites, excluding sites with the highest and lowest 10 percent of lipid levels. The 514 sites plotted have 2.3–10.4 percent lipid content. µg/kg, microgram per kilogram.

Table 3. Explanatory variables and coefficients for the regression model for dieldrin in whole fish.

[—, none; lb/mi^2, pound per square mile]

Explanatory variable	Transformation	Coefficient	p-value
Fish lipid content (%)	Logarithmic	1.0542	1.32E–21
Aldrin–dieldrin use in agriculture, per basin area (lb/mi^2)	—	0.0733	7.91E–26
Forest land (% of basin)	Square root	–0.0829	5.18E–16
Urban-termite score	Logarithmic(x + 1)	0.5406	2.91E–27
Organic matter content of basin soils (average %)	Logarithmic	–0.4838	2.94E–10
Southern Plains region	Binary group variable (1/0)	0.2714	0.0146
(Intercept)	—	–0.1674	0.132

Analysis of Significant Explanatory Variables

The first four and most important explanatory variables in the model have reasonable relations to dieldrin concentrations in fish, in terms of their relative importance and the direction of their influence. The measured lipid content is positively related to dieldrin concentration in fish, which is consistent with the tendency of organochlorine pesticides to accumulate in fat tissue. Coefficients are positive for agricultural use of aldrin and dieldrin and the urban-termite score (representing historical use in agriculture and termite control, respectively), and negative for the variable representing forested land in the basin (where use would have been minimal). Together these four variables explain 58 percent of the variability in dieldrin concentrations measured in fish (*fig. 6*).

The addition of two more variables, soil organic matter and the Southern Plains regional variable, made a modest improvement in model performance. The final model explains 64 percent of the variability in dieldrin concentrations in whole fish, and the scale parameter (standard deviation of the residual error) for the final regression is 0.485 (*fig. 6*). The last two variables are significantly, though weakly, correlated with the response variable. Because these variables explain relatively little of the variability, possible reasons for their influence on the response variable are speculative. The soil organic matter variable has a negative coefficient, and multiple explanations are possible. High values for soil organic matter may be associated with either or both (1) reduced transport of contaminated soil to streams and (or) (2) a higher organic carbon content in stream sediment, which would tend to have a higher fraction of dieldrin sorbed to sediment and a correspondingly lower fraction in water, which could in turn decrease uptake by fish. The first possible explanation is supported by the observation that, in data from the STATSGO database for NAWQA fish sampling sites, mean soil organic matter is positively correlated with mean soil permeability. In support of the second possible explanation, the logarithm of the dieldrin concentration measured in fish is also negatively correlated ($p < 0.05$) with the logarithm of the measured organic carbon content of streambed sediment collected at the same sites.

The Southern Plains regional variable has a positive coefficient, indicating that without this variable, the regression model would tend to underestimate dieldrin concentrations in fish in the Southern Plains region (Texas and Oklahoma). The reason for this underestimation is not known. However, in this region, aldrin and dieldrin were used in agriculture (especially on cotton), as well as in urban applications (the eastern parts of Texas and Oklahoma are in the zone of heaviest termite density). Of the NAWQA sites sampled in the Southern Plains region, the highest dieldrin residues in fish were from the Trinity River Basin study unit in eastern Texas—especially at sites in large basins with mixed (agricultural and urban) land use—whereas dieldrin was low or not detected at strictly agricultural sites in this study unit. Although not conclusive,

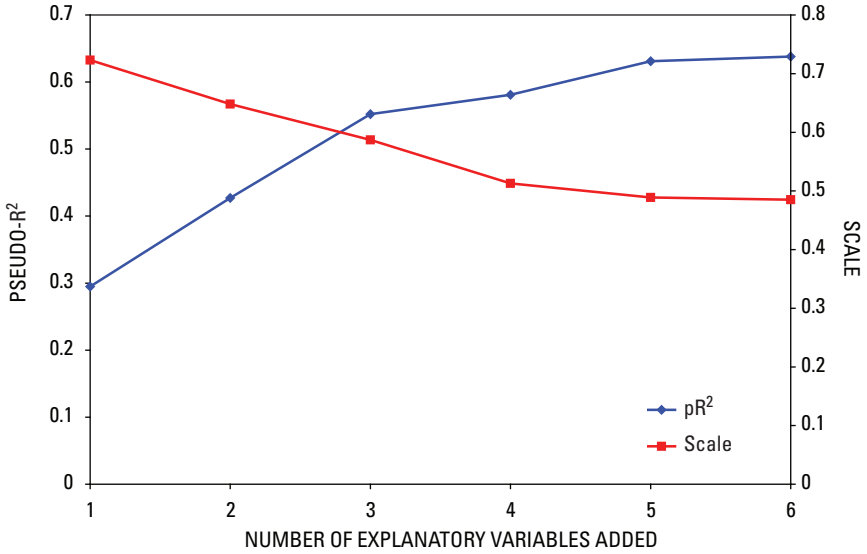


Figure 6. Effect of adding explanatory variables to the regression model for dieldrin in whole fish on the model's pseudo-R^2 and scale values.

this suggests that urban use (rather than agricultural use) of aldrin and dieldrin in this area may have been underestimated by the use terms in the regression model. In fact, most urban sites in the Trinity River Basin are located in the moderate-to-heavy zone of termite density, but very near (just outside) the boundary of the very-heavy termite zone (*fig. 2*); so, it is possible that the weighted termite-urban score may under-represent past termiticide use in these urban areas. Because this regional variable explains very little of the variability (about 1 percent), the apparent regional bias is slight and the explanation for it is speculative.

Extrapolation to Unmonitored Streams

The regression model developed here can be used for spatial extrapolation to streams that have not yet been assessed by direct measurements, because the model is based on certain characteristics of each hydrologic system, such as estimated historical pesticide use and soil characteristics within the basin, for which data or estimates are available for the entire nation. This approach to extrapolation is fundamental to extending the targeted local and regional studies of NAWQA to a comprehensive national assessment (Gilliom and others, 2006).

The results of spatial extrapolation to all United States streams are shown in *figure 7*, which maps the dieldrin concentration in composite whole fish predicted by the model, for whole fish with a lipid content of (A) 4.2 percent and (B) 6.2 percent. As noted previously, lipid content is not available from a nationwide data set, as are the other explanatory variables, and therefore default lipid values have been assumed that are appropriate to the objectives of the extrapolations. These maps show the spatial distribution of the best or most likely concentrations of dieldrin in composite whole-fish samples (for each of the specified lipid contents).

In *figure 7*, unmonitored streams are color coded to correspond to different concentration ranges for the predicted dieldrin concentration in whole fish (at the lipid content specified in the figure). A single set of color-coded dieldrin concentration ranges is used in both maps, and the boundary concentrations defining these ranges correspond to specific dieldrin guidelines for either human health or wildlife. The highest concentration category (>120 µg/kg) is defined by the New York guideline for protection of fish-eating wildlife (Newell and others, 1987). The lowest category consists of streams predicted to have dieldrin whole-fish residues below the reporting limit (<5 µg/kg), and the two middle categories are defined in relation to EPA screening values for consumption of edible fish or shellfish (U.S. Environmental Protection Agency, 2000). Specifically, the 5–25 µg/kg category is defined at the lower end by the reporting limit (5 µg/kg), which is associated with a cancer risk of 2 in 100,000 (or 1:50,000). (Note that the EPA recommended screening value at a 1:100,000 cancer risk is 2.5 µg/kg, which is lower than NAWQA was able to reliably quantitate in fish.) For the 25–120 µg/kg category, all sites have predicted values greater than the EPA screening value at a 1:10,000 cancer risk (25 µg/kg). These predictions are discussed further below in relation to potential significance to human health and wildlife.

Because the response variable is a logarithmic transformation, concentrations predicted by the model (after retransformation) are *median* concentrations expected for a given set of explanatory values. This means that, for sites with the identical predicted dieldrin concentration, the actual concentrations are expected to be greater than the predicted value (for a given lipid content) at half of the sites, and lower than predicted at half of the sites. In *figure 7*, therefore, actual concentrations (for composite whole-fish samples) at half the sites are expected to exceed the predicted values. Uncertainty in model predictions is discussed below in the context of human health and wildlife guidelines.

Because fish lipid content (a measured parameter) is a variable in the model and has a positive coefficient, predicted dieldrin concentrations will increase as fish lipid content increases. This is illustrated by comparing predictions for 6.2 percent lipid (*fig. 7B*) with predictions for 4.2 percent lipid (*fig. 7A*). The extrapolation at 6.2 percent lipid (*fig. 7B*) has fewer blue streams, and more yellow, orange, and red streams than the extrapolation at 4.2 percent lipid (*fig. 7A*). Whole fish with a lipid content greater than 6.2 percent would be predicted to have still higher dieldrin concentrations than those shown in *figure 7B*, and whole fish with a lipid content less than 4.2 percent would be expected to have lower concentrations than those shown in *figure 7A*. As a point of reference, the average whole-fish lipid content in NAWQA and NCBP fish samples for some additional species are lake trout, 15 percent; channel catfish, 7.5 percent; common carp, 6.5 percent; smallmouth buffalo, 6.5 percent; white sucker, 5.8 percent; yellow perch, 4.3 percent; sculpins, 3.8 percent; and bluegill, 3.1 percent.

Significance to Human Health

NAWQA fish data were not collected with the intention of assessing potential effects on human health from consumption of contaminated fish. Instead, NAWQA's primary fish sampling objective was to assess the occurrence and geographic distribution of hydrophobic contaminants in streams, and the fish sampling methods were designed accordingly. Two aspects of the NAWQA study design detract from the

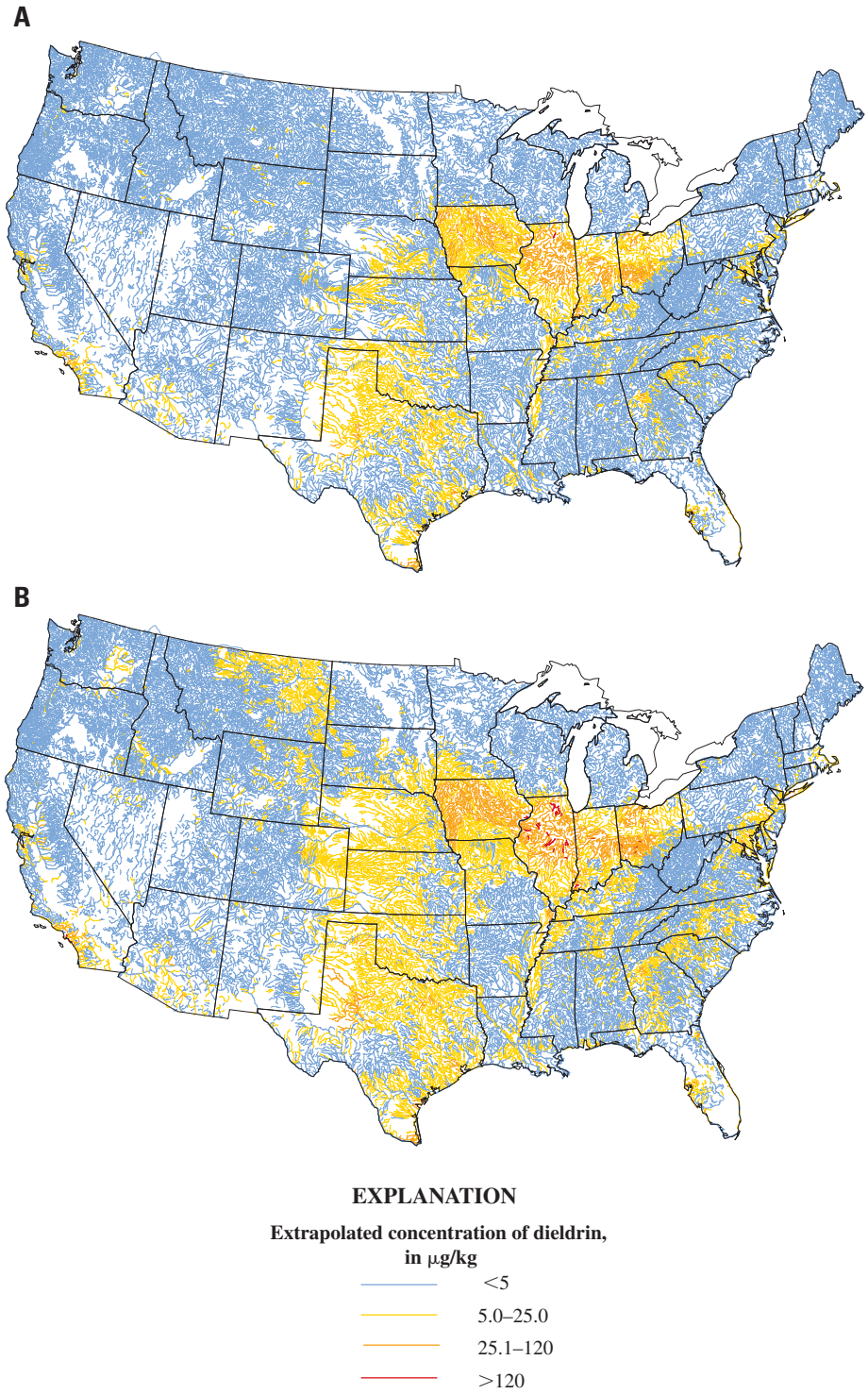


Figure 7. Predicted dieldrin concentrations in whole fish for U.S. streams at (*A*) 4.2 percent lipid content (largemouth bass) and (*B*) 6.2 percent lipid content (all fish taxa).

usefulness of NAWQA fish data in the assessment of human exposure and health risk. First, NAWQA analyzed whole-body residues, whereas people generally consume only the edible portion (fillet or muscle tissue), and second, many of the fish taxa collected are bottom feeders that are not commonly consumed by people. Nonetheless, careful data analysis will allow some inference with respect to potential human health effects.

It is not necessarily straightforward to predict how residues of organochlorine pesticides will compare in a whole-body bottom feeder versus a game-fish fillet, because two factors (fish taxon and fish tissue type) have changed. In its National Study of Chemical Residues in Fish, EPA measured bioaccumulative chemicals in whole-body bottom-feeders and game-fish fillets, and found that the relative magnitude of residues in each depended on the specific chemical being measured (U.S. Environmental Protection Agency, 1992b). Of the organochlorine pesticide chemicals measured, some (including dieldrin, oxychlordane, and DDE) were present in roughly the same average concentrations in game-fish fillets as in whole-body bottom-feeders, whereas others (including chlordane, nonachlor, and heptachlor epoxide) had higher average concentrations in whole-body bottom feeders than in game-fish fillets.

These findings suggest that, with caveats, NAWQA whole-fish data can be compared with EPA screening values for edible fish consumption in a *screening-level* assessment—such as in deciding whether and where to conduct future dieldrin monitoring in game-fish fillets. For example, in streams where dieldrin residues in whole fish are found (or predicted) to be high, there exists some potential for concentrations in edible game fish to be high as well. Therefore, if such streams are frequented by sport or subsistence fishermen, this suggests that additional sampling and analysis of game-fish fillets from these streams may be warranted. Conversely, where dieldrin residues in whole fish are low or undetected, concentrations in edible fish tissue are also likely to be low—allowing for differences that are due to species and (or) lipid content.

To assess relevance to human health, the regression model was used to predict dieldrin concentrations in whole fish, assuming a 4.2 percent lipid content (to represent the average lipid content of whole-body largemouth bass), in unmonitored streams. As noted previously, the 4.2 percent lipid value pertains to *whole-body* largemouth bass, whereas the EPA screening values for fish consumption apply to the *edible* portion of the fish, which tends to have a lower lipid content. Data from EPA's National Study for Chemical Residues in Fish indicate that lipid levels in bass fillets average 0.84 percent, or about one fifth the content of whole-body largemouth bass sampled by the NCBP and NAWQA programs (U.S. Environmental Protection Agency, 1992b). Therefore, the comparison shown in *figure 7A* represents a worst-case assessment *at the 4.2 percent lipid level* from the perspective of human health. For fish with a higher lipid level, such as lake trout (with an average whole-body lipid content of 15 percent), *whole-body* dieldrin concentrations would likely be much higher than the whole-body concentrations calculated at the 4.2 percent lipid level (and shown in *fig. 7A*). We cannot make specific predictions about dieldrin concentrations in *edible* fish tissues for largemouth bass, lake trout, or any other taxon because the data required to make such predictions were not collected.

In *figure 7A*, dieldrin concentrations in whole fish from the dark-blue streams are predicted to be less than the reporting limit (5 µg/kg) for the method NAWQA used to analyze for dieldrin in fish. In edible fish tissue, a concentration less than 5 µg/kg would correspond to a cancer risk level of less than 1 in 50,000. Because dieldrin concentrations in edible fish are expected to be even lower than those in whole fish, the dark blue streams are likely to have median dieldrin concentrations in edible fish tissue that are below the EPA screening value at 1:50,000 cancer risk.

In *figure 7A*, streams with predicted dieldrin concentrations greater than 25 µg/kg (red and orange shaded streams) are mostly concentrated in the Corn Belt, with scattered patches in Texas, the southeast, urbanized parts of the northeast, and California. If measured in edible fish tissue, dieldrin concentrations greater than 25 µg/kg would exceed the EPA recommended screening value for 1 in 10,000 cancer risk, and concentrations in the range of 5–25 µg/kg (yellow shaded streams) would exceed the EPA screening value for 1 in 50,000 cancer risk. Exactly what constitutes an acceptable maximum level of cancer risk is not clearly defined by EPA. Many state and EPA programs use a cancer risk level of between 1 in 100,000 and 1 in 1,000,000. Certainly, cancer risks exceeding 1:10,000 are higher than the maximum acceptable cancer risk levels used by most state and EPA programs (Nowell and Resek, 1994). Again, this comparison must be qualified because predicted dieldrin concentrations in whole fish probably overestimate concentrations in edible fish. In streams where whole-fish dieldrin concentrations are predicted by the model to exceed EPA screening values at 1:10,000 (red and orange streams) or 1:50,000 (yellow streams), it is not clear that concentrations in edible fish also would exceed screening values. However, for such streams that are actively fished, the extrapolation results can be used to guide future monitoring and assessment priorities for measuring dieldrin in the edible tissues of game fish.

Significance to Fish-Eating Wildlife

The extrapolation shown in *figure 7B* assumes a fish lipid content of 6.2 percent lipid (to represent the average lipid content of all whole fish, regardless of species). Because wildlife consume many fish species and typically eat the entire fish, it is appropriate to directly compare model predictions at 6.2 percent lipid with whole-fish guidelines for protection of wildlife.

The New York wildlife guideline used in this analysis (120 µg/kg dieldrin) is intended to protect fish-eating birds and mammals from adverse effects other than cancer, as noted previously. Relatively few streams are predicted to have median concentrations that exceed this dieldrin guideline (mapped in red in *fig. 7B*) and these few are in the Corn Belt, where past use of aldrin and dieldrin in agriculture was highest.

Uncertainty in Model Predictions

In the preceding model predictions, it is important to consider uncertainty, especially when comparing predicted concentrations with human-health or aquatic-life guidelines. In figures *7A* and *B*, measured fish samples (at the appropriate lipid content) will have a 50 percent probability of exceeding the predicted concentration. For example, even for dark blue streams in *figure 7A*, where the predicted median concentration is <5 µg/kg, this means that 50 percent of streams are expected to have whole-fish concentrations greater than this predicted concentration, and some may in fact be greater than EPA screening values of 2.5 or even 25 µg/kg, which are associated with cancer risk levels of 1:100,000 and 1:10,000, respectively.

The maps in *figures 8A* and *8B* and *9* help put this uncertainty into perspective by showing the probability of streams having whole-fish concentrations that exceed a

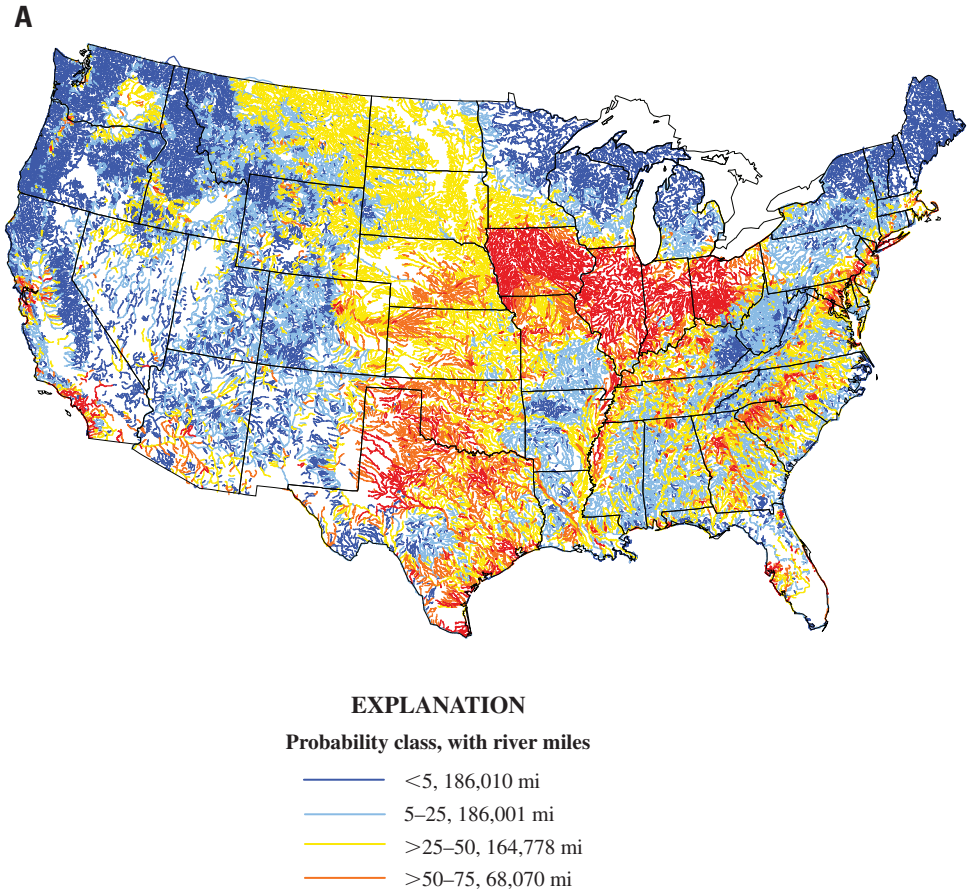


Figure 8. Probability of streams having whole-fish samples that exceed (A) 5 µg/kg (which is the reporting level for dieldrin in whole fish and also corresponds to an EPA screening value at a cancer risk of 1:50,000) and (B) 25 µg/kg (which corresponds to the EPA screening value at a cancer risk of 1:10,000). In both maps, whole-fish samples are assumed to have 4.2 percent lipid content (the mean value for largemouth bass sampled by the NCBP and NAWQA programs combined). The EPA screening values apply to edible fish tissue. mi, mile; µg/kg, microgram per kilogram.

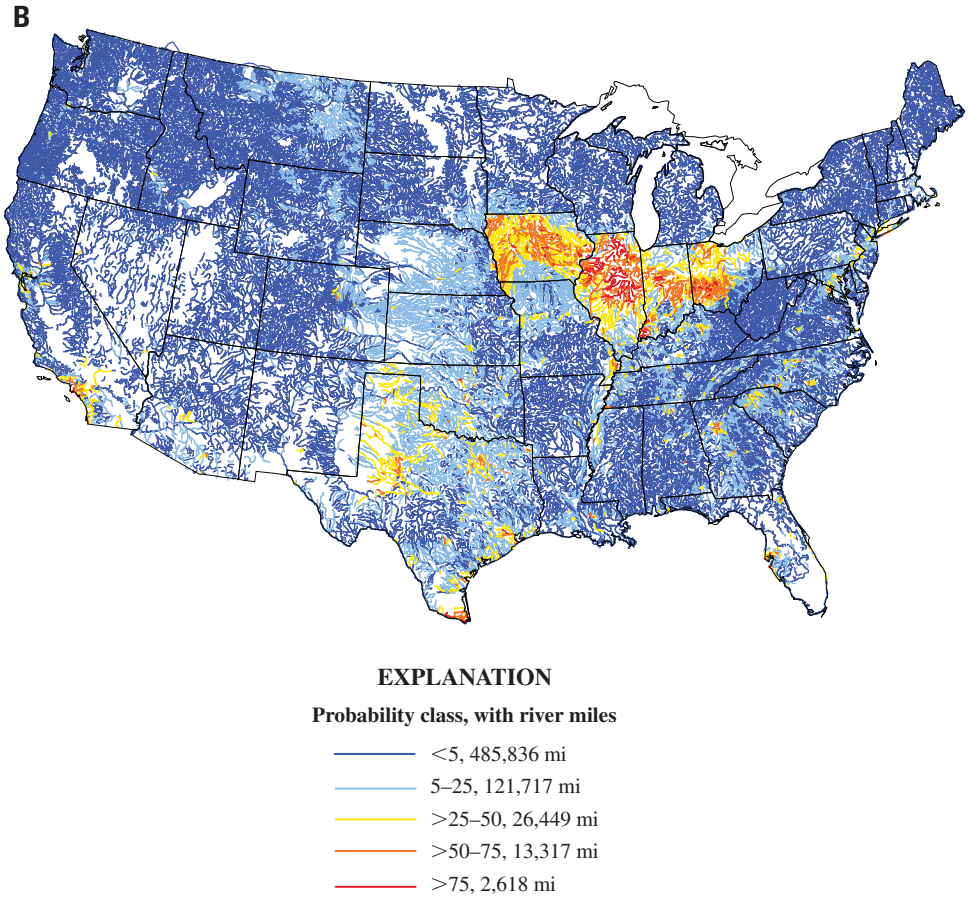


Figure 8. Continued.

specific guideline concentration. The maps in *figure 8A* and *8B* assume a lipid content of 4.2 percent and relate to human health guidelines, whereas the map in *figure 9* assumes a lipid content of 6.2 percent and relates to potential effects on fish-eating wildlife.

The maps in *figure 8* show the probability of streams having whole-fish concentrations (at 4.2 percent lipid content) that exceed 5 µg/kg (*fig. 8A*) and 25 µg/kg for dieldrin (*fig. 8B*). For dark blue streams in *figure 8A*, whole-fish samples (with 4.2 percent lipid) have less than a 5 percent probability of exceeding 5 µg/kg; in other words, fewer than 1 out of 20 dark blue streams are expected to have whole-fish dieldrin concentrations greater than 5 µg/kg. In edible fish, a 5-µg/kg concentration would correspond to a maximum acceptable cancer risk of 1 in 50,000. Dark blue streams account for only 29 percent of total stream miles in the United States. The remaining 71 percent of total stream miles have at least a 5 percent probability of whole-fish concentrations exceeding 5 µg/kg. Of these streams, whole-fish concentrations are expected to exceed 5 µg/kg in more than 1 out of 20 light blue streams (>5 percent probability), more than 1 out of 4 yellow streams (>25 percent probability), more than half of orange streams (>50 percent probability), and more than 3 out of 4 red streams (>75 percent probability). Orange and red streams, which have at least a 50 percent probability of whole-fish concentrations exceeding 5 µg/kg, make up 17 percent of total river miles and are located mostly in the Corn Belt, parts of Texas and California, and scattered urban areas.

Far fewer streams are likely to have whole-fish samples (at 4.2 percent lipid) with dieldrin concentrations exceeding 25 µg/kg, as shown in *figure 8B*. In edible fish tissue, a concentration of 25 µg/kg would correspond to the EPA screening value at a cancer risk of 1:10,000. Here, most streams (75 percent of stream miles) are shaded dark blue and have less than a 5 percent probability of whole-fish concentrations exceeding 25 µg/kg. Of the remaining streams, more than 1 out of 20 light blue streams (>5 percent probability), more than 1 out of 4 yellow streams (>25 percent probability), more than half of orange streams (>50 percent probability), and more than 3 out of 4 red streams (>75 percent probability) are expected to have whole-fish concentrations that exceed 25 µg/kg. Orange and red streams account for only about 2 percent of total stream miles, and they are located mostly in the Corn Belt. Again, these maps show the probability of *whole-fish* samples exceeding the predicted concentrations—edible fish tissue is likely to have lower lipid levels and lower dieldrin concentrations than in the whole fish.

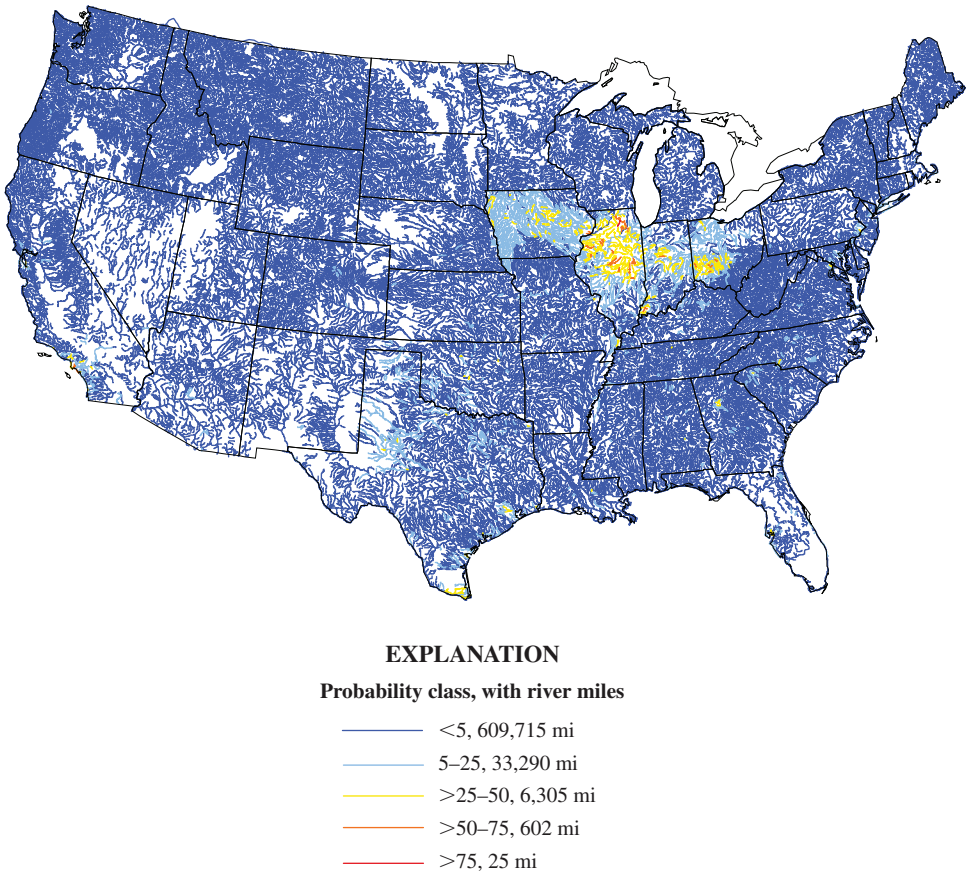


Figure 9. Probability of streams having whole-fish concentrations greater than 120 µg/kg, which is the New York criterion for protection of fish-eating wildlife (noncancer effects) from Newell and others (1987). Whole-fish samples are assumed to have 6.2 percent lipid content (the mean value for all whole fish sampled by the NCBP and NAWQA programs combined). mi, mile; µg/kg, microgram per kilogram.

Figure 9 relates to potential effects on fish-eating wildlife. This map shows the probability of streams that have whole-fish concentrations (at 6.2 percent lipid content) greater than the New York wildlife guideline of 120 µg/kg. Most streams (about 94 percent of total stream miles) are shaded dark blue and have less than a 5 percent probability of whole-fish samples (at 6.2 percent lipid) exceeding 120 µg/kg—in other words, fewer than 1 out of 20 dark blue streams are expected to exceed the wildlife guideline. More than 1 out of 20 light blue streams (> 5 percent probability), more than 1 out of 4 yellow streams (>25 percent probability), more than half of orange streams (>50 percent probability), and 3 out of 4 red streams (>75 percent probability) are expected to exceed the New York wildlife guideline. Orange and red streams—for the most part located in the Corn Belt—account for less than 1 percent of total stream miles. Fish with a lipid content greater than 6.2 percent would likely have higher dieldrin concentrations than those predicted by the model; conversely, fish with a lower lipid content would likely have lower dieldrin concentrations than predicted by the model.

Although dieldrin residues in fish have declined substantially since the 1960s (Schmitt and others, 1999; Nowell and others, 1999), the recent NAWQA data and results of these extrapolations suggest that detectable residues are present in some streams across the country at levels sufficient to be of potential human health or wildlife concern. A number of fish consumption advisories for dieldrin have been rescinded in recent years, consistent with the national declining trends. As of 2004, there were still 22 active fish consumption advisories for dieldrin affecting seven states within the United States (U.S. Environmental Protection Agency, 2005). Models such as the one described here for dieldrin may assist in planning future monitoring efforts to support the setting or rescinding of fish consumption advisories.

Summary and Conclusions

A regression model was developed for estimation of dieldrin concentrations in whole fish in United States streams, using nationally available data on watershed characteristics and estimates of (or surrogates for) the past uses of aldrin and dieldrin. The model also included lipid content (a measured parameter) as an explanatory variable. The model accounts for 64 percent of the variability in dieldrin concentrations in whole fish from the 648 streams used for model development. Four variables were the most important and together accounted for 58 percent of the variability. Three of these variables have positive coefficients: lipid content, the estimated agricultural use intensity of aldrin and dieldrin per basin area, and a weighted termite-urban score (which is a surrogate for past use of aldrin and dieldrin in termite control). The fourth important variable, total forested land (where past use was probably minimal), has a negative coefficient.

The regression model performed well when cross-validation procedures were conducted, in which the NAWQA fish data were randomly divided into two subsets (calibration and validation data sets); the model was fit to both data sets, and there was little difference between predicted and residual errors obtained from the two fitted models that were based on the different data subsets.

The regression model was used to predict the median dieldrin concentrations in whole fish from unmonitored streams across the United States. Because lipid content is an explanatory variable in the model, these predictions were made by assuming a specific lipid content value. Two nationwide spatial extrapolations were done—at lipid content values of 4.2 and 6.2 percent lipid—to assess potential effects on human health and wildlife, respectively, from consumption of dieldrin-contaminated fish. These lipid values represent the average lipid content of whole-body largemouth bass (4.2 percent) and all species of whole fish (6.2 percent) sampled by the NCBP and NAWQA programs, combined.

The extrapolation at 4.2 percent lipid was used to predict dieldrin concentrations in whole-body largemouth bass, which is a nationally distributed sport fish popular with recreational fishers, in unmonitored streams across the U.S. For fish with a higher lipid content (such as lake trout, which has an average lipid content of 15 percent), whole-fish dieldrin concentrations are likely to be much higher than those predicted for a 4.2 percent lipid content. On the other hand, people generally consume only the edible fish tissue (fillets), which is expected to have a lower lipid content and lower dieldrin concentrations than in the whole fish. The model should not be used to predict specific dieldrin concentrations in edible fish tissue. However, it can be a useful screening tool in identifying streams where dieldrin contamination may be of potential human health concern. For streams that are actively fished and where predicted whole-body dieldrin concentrations are greater than EPA screening values for human consumption of edible fish, it may be worthwhile to monitor dieldrin residues in game-fish fillets from these streams. A substantial fraction of streams have at least a 5 percent probability (1 in 20) of whole-fish concentrations exceeding EPA screening values for edible fish consumption that are based on potential carcinogenic effects of dieldrin—for 71 percent of total stream miles, 1 in 20 streams are expected to have whole-fish concentrations greater than 5 µg/kg (which has a cancer risk of 1:50,000); and for about 25 percent of total stream miles, 1 in 20 streams are expected to have whole-fish concentrations greater than 25 µg/kg (which has a cancer risk of 1:10,000). The streams with the highest probability of exceeding EPA screening values are located primarily in the Corn Belt, where aldrin and dieldrin were used in the past on corn, and in scattered areas across the United States, especially in the urbanized northeast, the southeast, Texas, and California.

The extrapolation at 6.2 percent lipid was used to predict dieldrin concentrations in whole fish from unmonitored streams across the United States, to represent fish that may be consumed by wildlife. Predicted concentrations were compared with the fish flesh criterion for protection of fish-eating wildlife (protective of noncancer effects) from the State of New York. Relatively few streams—located predominantly in the Corn Belt—were predicted to have median dieldrin concentrations that exceeded the New York wildlife guideline of 120 µg/kg. For most unmonitored streams in the United States (94 percent of total stream miles), fewer than 1 out of 20 streams are expected to exceed the wildlife guideline for dieldrin.

References

Aitkin, Murry, 1981, A note on the regression of censored data: Technometrics, v. 23, no. 2, p. 161–163.

Amdur, M.O.; Doull, John; Klaassen, C.D.; eds., 1991, Casarett and Doull's toxicology—the basic science of poisons: Pergamon Press, New York, p. 314, 505–506, 565–622. [Pergamon is an imprint of Elsevier Publishers]

Anderson, J.R., Hardy, E.E., Roach, J.T., and Witmer, R.E., 1976, A land use and land cover classification system for use with remote sensor data: U.S. Geological Survey Professional Paper 964, 28 p. [also available at *http://uyypubs.er.usgs.gov/pubs/pp/pp964]*

Andrilenas, P.A., 1974, Farmers' use of pesticides in 1971—quantities: U.S. Department of Agriculture, Economic Research Service, Agricultural Economic Report no. 252, 56 p.

Beal, R.H., Mauldin, J.K., and Jones, S.C., 1994, Subterranean termites: their prevention and control in buildings: U.S. Department of Agriculture, Home and Garden Bulletin no. 64.

Battaglin, W.A., and Goolsby, D.A., 1997, Statistical modeling of agricultural chemical occurrence in midwestern rivers: Journal of Hydrology, v. 196, p. 1–25.

Battaglin, W.A., and Goolsby, D.A., 1998, Regression models of herbicide concentrations in outflow from reservoirs in the midwestern USA, 1992–1993: Journal of the American Water Resources Association, v. 34, no. 6, p. 1369–1390.

Black, R.W., Haggland, A.L., Voss, F.V., 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: Environmental Toxicology and Chemistry, v. 19, no. 4, p. 1044–1054.

Bradu, D. and Mundlak, Y., 1970, Estimation in lognormal linear models: Journal of the American Statistical Association, v. 65, no. 329, p. 198–211.

Crawford, J.K., and Luoma, S.N., 1994, Guidelines for studies of contaminants in biological tissues for the National Water-Quality Assessment Program: U.S. Geological Survey Open-File Report 92-494, 69 p.

Daly, Christopher; Gibson, W.P.; Taylor, G.H.; Johnson, G.L.; and Pasteris, Phillip; 2002, A knowledge-based approach to the statistical mapping of climate: Climate Research, v. 22, p. 99–113. [also available at *http://www.ocs.orst.edu/prism/docs/index.phtml*]

Daly, Christopher; Neilson, R.P.; and D.L. Phillips; 1994, A statistical-topographic model for mapping climatological precipitation over mountainous terrain: Journal of Applied Meteorology, v. 33, p. 140–158.

Duan, Naihua, 1983, Smearing estimate—a nonparametric retransformation method: Journal of the American Statistical Association, v. 78, no. 383, p. 605–610.

Economic Research Service, 1999, Historical census of agriculture data, 1949–1974, records of the Economic Research Service (record Group 354): Washington, D.C., data compiled from Section 1, chap. 3–4 of the Census of Agriculture, v. 1, Custom order CD-ROM, National Archives.

Eichers, Theodore; Andrilenas, Paul; Blake, Helen; Jenkins, Robert; and Fox, Austin; 1970, Quantities of pesticides used by farmers in 1966: U.S. Department of Agriculture, Economic Research Service, Agricultural Economic Report no. 179, 61 p.

Escobar, L.A., and Meeker, W.Q., Jr., 1992, Assessing influence in regression analysis with censored data: Biometrics: v. 48, p. 507–528.

Esworthy, R.F., 1987, Incremental benefit analysis—restricted use of all pesticides registered for subterranean termite control: U.S. Environmental Protection Agency, Benefits and Use Division, Economic Analysis Branch.

Fegeas, R.G., Claire, R.W., Guptill, S.C., Anderson, K.E., and Hallam, C.A., 1983, Land use and land cover digital data: U.S. Geological Survey Circular 895-E, 21 p.

Fenneman, N.M., and Johnson, D.W., 1946, Physical divisions of the United States, accessed May 1, 2003, at *http://water.usgs.gov/lookup/getspatial?physio*

Fuhrer, G.J., Gilliom, R.J., Hamilton, P.A., Morace, J.L., Nowell, L.H., Rinella, J.F., Stoner, J.D., and Wentz, D.A., 1999, The quality of our nation's waters: nutrients and pesticides: U.S. Geological Survey Circular 1225, 82 p.

Gebert, W.A., Graczyk, D.J., and Krug, W.R., 1987, Average annual runoff in the United States: U.S. Geological Survey Hydrologic Atlas 710. [also available at *http://water.usgs.gov/lookup/getspatial?runoff*]

Geolytics, Inc., 2004, Census CD 2000 and Street CD 2000 [CDROM]. [also available at http://www.geolytics.com].

Gilliom, R.J.; Barbash, J.E.; Crawford, C.G.; Hamilton, P.A.; Martin, J.D.; Nakagaki, Naomi; Nowell, L.H.; Scott, J.C.; Stackelberg, P.E.; Thelin, G.P.; and Wolock, D.M., 2006, The quality of our nations's water—pesticides in the nation's streams and ground water, 1992–2001: U.S. Geological Survey Circular 1291. [also available at *http://pubs.usgs.gov/circ/2005/1291/*]

Haake, J., Kelley, M., Keys, B., and Safe, S., 1987, The effects of organochlorine pesticides as inducers of testosterone and benzo[a]pyrene hydroxylases: General Pharmacology, v. 18, no. 2, p. 165–169.

Hamon, W.R., 1961, Estimating potential evapotranspiration: Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers, v. 87, p. 107–120.

Hirsch, R.M., Alley, W.M., and Wilber, W.G., 1988, Concepts for a National Water-Quality Assessment Program: U.S. Geological Survey Circular 1021, 42 p.

Hitt, K.J., 1994, Refining 1970's land-use data with 1990 population data to indicate new residential development. U.S. Geological Survey Water-Resources Investigation Report 94-4250, 14 p.

Hunt, E.G., and Linn, J.D., 1970, Fish kills by pesticides, *in* Gillett, J.W., ed., The biological impact of pesticides in the environment: Corvallis, Oreg., Oregon State University Press, p. 97–103.

Insightful Corporation, 1999, S-PLUS 2000 user's guide: Seattle, Wash., Insightful Corporation, Data Analysis Products Division, 558 p.

Judge, G.G.; Griffiths, W.E.; Hill, R.C.; Lütkepohl, Helmut; and Lee, T.C.; 1985, The theory and practice of econometrics (2d ed.): New York, John Wiley and Sons, 1019 p.

Krueger, J., and Tornqvist, L., 1996, Multiple regression analysis and prediction of pesticide occurrence in subsurface drainage and streamflow related to pesticide intrinsic properties and quantities applied, *in* Del Re., A.A.M.; Capri, Ettore; Evans, S.P.; and Trevisan, Marco; eds., The environmental fate of xenobiotics: Proceedings of the Tenth Symposium on Pesticide Chemistry, September 30 to October 2, 1996, Castelnuevo Fogliana, Piacenza, Italia, La Goliardica Pavese, Pavia, Italy, p. 681–685.

Laitila, Thomas, 1993, A pseudo-R-square measure for limited and qualitative dependent variable models: Journal of Econometrics, v. 56, p. 341–356.

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

Larson, S.J., and Gilliom, R.J., 2001, Regression models for estimating herbicide concentrations in U.S. streams from watershed characteristics: Journal of the American Water Resources Association, v. 37, no. 5, p. 1349–1367.

Leiker, T.J., Madsen, J.E., Deacon, J.R., and Foreman, W.F., 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: U.S. Geological Survey Open-File Report 94-710, 42 p.

Maddala, G.S., 1983, Limited-dependent and qualitative variables in econometrics: Cambridge, U.K., Cambridge University Press, 401 p.

Madhun, Y.A., and Freed, V.H., 1990, Impact of pesticides on the environment, *in* Cheng, H.H., ed., Pesticides in the soil environment—processes, impacts, and modeling: Madison, Wis., Soil Science Society of America, p. 429–466.

Mitchell, W.W., Guptill, S.C., Anderson, K.E., Fegeas, R.G., and Hallam, C.A., 1977, GIRAS—a geographic information system analysis system for handling land use and land cover data: U.S. Geological Survey Professional Paper 1059, 16 p.

Munn, M.D., and Gruber, S.J., 1997, The relationship between land use and organochlorine compounds in streambed sediment and fish in the Central Columbia Plateau, Washington and Idaho, area: Environmental Toxicology and Chemistry, v. 16, no. 9, p. 1877–1887.

Myers, M.S., Stehr, C.M., Olson, O.P., Johnson, L.L., McCain, B.B., Chan, S-L., and 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): Seattle, Wash., National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Northwest Fisheries Center.

Nakagaki, Naomi; and Wolock D.M.; 2005, Estimation of agricultural pesticide use in drainage basins using land cover maps and county pesticide data: U.S. Geological Survey Open File Report 2005-1188 [also available at *http://pubs.water.usgs.gov/ofr20051188/*]

National Resources Conservation Service, 1994, State soil geographic (STATSGO) database—data use information: U.S. Department of Agriculture, National Soil Survey Center, Miscellaneous Publication no. 1492, 110 p.

Newell, A.J., Johnson, D.W., and Allen, L.K., 1987, Niagara River Biota Contamination Project: fish flesh criteria for piscivorous wildlife: New York State Department of Environmental Conservation, Division of Fish and Wildlife, Bureau of Environmental Protection, Technical Report 87-3. [also available at *http://www.dec.state.ny.us/website/dfwmr/habitat/niagarabiotacontamproj.pdf*]

Nolan, J.V., Brakebill, J.W., Alexander, R.B., and Schwarz, G.E., 2003, Enhanced River Reach File 2.0, Version 2.0, November 10, 2003: U.S. Geological Survey Open-File Report 02-4 [also available at *http://water.usgs.gov/lookup/getspatial?erf1_2*]

Nowell, L.H., Capel, P.D., and Dileanis, P.D., 1999, Pesticides in stream sediment and aquatic biota—distribution, trends, and governing factors: Boca Raton, Fla., CRC Press, Pesticides in the Hydrologic System series, v. 4, 1040 p.

Nowell, L.H., and Crawford, C.G., 2003, Organochlorine pesticides and PCBs in bed sediment and whole fish from United States rivers and streams: Summary statistics; Results from Cycle I of the National Water-Quality Assessment Program (NAWQA), 1992–2001, accessed Oct. 26, 2005, at *http://ca.water.usgs.gov/pnsp/oc_doc.html*

Nowell, L.H., and Resek, E.A., 1994, National standards and guidelines for pesticides in water, sediment, and aquatic organisms: application to water-quality assessments in Ware, G.W., ed., Reviews of Environmental Contamination and Toxicology, v. 140, p. 1–164.

Price, C.V., 2003, 1990 Population density by block group for the conterminous United States, digital map, accessed Jan. 24, 2005, at *http://water.usgs.gov/lookup/getspatial?uspopd90x10g*

Price, C.V.; Nakagaki, Naomi; Hitt, K.J; and Clawges, R.M.; 2003, Mining GIRAS: Improving on a national treasure of land use data: Proceedings of the ESRI 2003 International User Conference. [also available at *http://www.esri.com/uc*]

Schmitt, C.J., Zajicek, J.L., May, T.W., and Cowman, D.F., 1999, Organochlorine residues and elemental; contaminants in U.S. freshwater fish, 1976–1986: National Contaminant Biomonitoring Program: Reviews of Environmental Contamination and Toxicology v. 162, p. 43–104.

Therneau, T.M., 1999, A package of survival routines for S: Rochester, Minn., Mayo Clinic Department of Health Sciences Research, 83 p. [also available at *http://www.may.edu/hsr/people/therneau.html*]

Truhlar, J.F., and Reed, L.A., 1976, Occurrence of pesticide residues in four streams draining different land use areas in Pennsylvania, 1969–71: Pesticides Monitoring Journal v. 10, no. 3, p. 101–110.

U.S. Bureau of the Census, 1991, Census of population and housing, 1990: Public Law (P.L.) 94-171: U.S. Bureau of the Census: Washington, DC, digital tabular data on CD-ROM. [also available at *http://www.census.gov*]

U.S. Department of Agriculture, 1995, 1992 National Resources Inventory: Fort Worth, Tex., Soil Conservation Service National Cartography and GIS Center, tabular digital data (on CD-ROM).

U.S. Department of Agriculture, 1999, National soil survey handbook: U.S. Department of Agriculture, Natural Resources Conservation Service.

U.S. Environmental Protection Agency, 1975, DDT: a review of scientific and economic aspects of the decision to ban its use as a pesticide: U.S. Environmental Protection Agency EPA-540/1-75-022.

U.S. Environmental Protection Agency, 1983, Analysis of the risks and benefits of seven chemicals used for subterranean termite control: U.S. Environmental Protection Agency EPA-540/9-83-005, variously paged.

U.S. Environmental Protection Agency, 1992a, National study of chemical residues in fish, volume I: U.S. Environmental Protection Agency, Office of Science and Technology, EPA-823-R-92-008a, variously paged.

U.S. Environmental Protection Agency, 1992b, National study of chemical residues in fish, volume II: U.S. Environmental Protection Agency, Office of Science and Technology, EPA-823-R-92-008b, variously paged.

U.S. Environmental Protection Agency, 2000, Guidance for assessing chemical contaminant data for use in fish advisories, volume 1, Fish sampling and analysis, 3rd ed.: U.S. Environmental Protection Agency, Office of Water, EPA 823-B-00-007, accessed Jan. 1, 2004, at *http://www.epa.gov/OST/fishadvice/volume1/*

U.S. Environmental Protection Agency, 2005, The national listing of fish advisories database, release of 2004 data: U.S. Environmental Protection Agency, Office of Water, accessed Nov. 9, 2005, at *http://www.epa.gov/waterscience/fish/advisories/index.html*

U.S. Geological Survey, 1986, Land use and land cover digital data from 1:250,000- and 1:100,000-scale maps: Data User Guide 4: U.S. Geological Survey, 25 p.

U.S. Geological Survey, 1993, Digital elevation models: Data users guide 5: U.S. Geological Survey, 48 p.

U.S. Geological Survey, 1998, Land use and land cover digital data from 1:250,000- and 1:100,000-scale maps [digital files in GIRAS format], accessed Oct. 27, 2005, at http://edc.usgs.gov/products/landcover/lulc.html

U.S. Geological Survey, 2000, U.S. GeoData digital Elevation models: U.S. Geological Survey Fact Sheet FS-040-00, 2 p. [also available at *http://mac.usgs.gov/mac/isb/pubs/factsheets/fs04000.pdf*]

U.S. Geological Survey, 2004, National Contaminant Biomonitoring Program (NCBP): NCBP Fish Database, accessed July 29, 2004, at *http://www.cerc.usgs.gov/data/ncbp/fish.htm*

Venables, W.N., and Ripley, B.D., 1999, Modern applied statistics with S-Plus, 3rd ed.: New York, Springer, 501 p.

Verdin, K.L., and Greenlee, S.K., 1996, Development of continental scale digital elevation models and extraction of hydrographic features, *in* Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, New Mexico, January 21–26, 1996, Proceedings: Santa Barbara, California, National Center for Geographic Information and Analysis.

Vogelmann, J.E., Howard, S.M., Yang, L., Larson, C.R., Wylie, B.K., and Van Driel, N., 2001, Completion of the 1990's national land cover dataset for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources: Photogrammetric Engineering and Remote Sensing, v. 67, p. 650–662.

Wieczorek, M.E., 2004a, Surface drainage, main or laterals on agricultural land in the conterminous United States: U.S. Geological Survey Open-File Report 2004-1189, ver. 1.1, U.S. Geological Survey digital map data, accessed October 27, 2005, at *http://water.usgs.gov/lookup/getspatial?ofr041189329*

Wieczorek, M.E., 2004b, Irrigation source on agricultural land in the conterminous United States: U.S. Geological Survey Open-File Report 2004-1189, ver. 1.1, U.S. Geological Survey digital map data, accessed October 27, 2005, at *http://water.usgs.gov/lookup/getspatial?ofr041189it01*.

Wieczorek, M.E., 2004c, Pressure irrigation source on agricultural land in the conterminous United States: U.S. Geological Survey Open-File Report 2004-1189, ver. 1.1, U.S. Geological Survey digital map data, accessed October 27, 2005, at *http://water.usgs.gov/lookup/getspatial?ofr041189it02*

Wieczorek, M.E., 2004d, Gravity and pressure irrigation source on agricultural land in the conterminous United States: U.S. Geological Survey Open-File Report 2004-1189, ver. 1.1, U.S. Geological Survey digital map data, accessed October 27, 2005, at *http://water.usgs.gov/lookup/getspatial?ofr041189it03*

Wieczorek, M.E., 2004e, Subsurface drains on agricultural land in the conterminous United States: U.S. Geological Survey Open-File Report 2004-1189, v. 1.1, U.S. Geological Survey digital map data, accessed October 27, 2005, at *http://water.usgs.gov/lookup/getspatial?ofr041189606*

Wieczorek, M.E., 2004f, Surface drainage, field ditches on agricultural land in the conterminous United States: U.S. Geological Survey Open-File Report 2004-1189, ver. 1.1, U.S. Geological Survey digital map data, accessed October 27, 2005, at *http://water.usgs.gov/lookup/getspatial?ofr041189607*

Wolock, D.W., 1993, Simulating the variable-source-area concept of streamflow generation with the watershed model TOPMODEL: U.S. Geological Survey Water Resources-Investigations Report 93-4124, 33 p.

Wolock, D.M., 1997, STATSGO soil characteristics for the conterminous United States: U.S. Geological Survey Open-File Report 97-656, digital map, accessed Oct. 27, 2005, at *http://water.usgs.gov/lookup/getspatial?muid*

Wolock, D.M., 2003a, Saturation overland flow estimated by TOPMODEL for the conterminous United States: U.S. Geological Survey Open-File Report 03-264, Metadata report and data available at *http://water.usgs.gov/lookup/getspatial?satof*

Wolock, D.M., 2003b, Infiltration-excess overland flow estimated from TOPMODEL for the conterminous United States: U.S. Geological Survey Open-File Report 03-310, Metadata report and data available at *http://water.usgs.gov/lookup/getspatial?ieof*

Wolock, D.M., Fan, J., and Lawrence, G.B., 1997, Effects of basin size on low-flow stream chemistry and subsurface contact time in the Neversinic River watershed, New York: Hydrological Processes, v. 11, p. 1273–1286.

Wolock, D.M., and McCabe, G.J., 1995, Comparison of single and multiple flow direction algorithms for computing topographic parameter in TOPMODEL: Water Resources Research, May 1995, v. 31, no. 5, p. 1315–1324.

Wong, C.S., Capel, P.D., and Nowell, L.H., 2000, Organochlorine pesticides and PCBs in stream sediment and aquatic biota: initial results from the National Water-Quality Assessment Program, 1992–1995: U.S. Geological Survey Water-Resources Investigations Report 00-4053, 88 p. [also available at *http://ca.water.usgs.gov/pnsp/rep/wrir004053/*]

Inside back cover.