

Watershed Regressions for Pesticides (WARP) for Predicting Annual Maximum and Annual Maximum Moving-Average Concentrations of Atrazine in Streams

Open-File Report 2008-1186

Watershed Regressions for Pesticides (WARP) for Predicting Annual Maximum and Annual Maximum Moving-Average Concentrations of Atrazine in Streams

By Wesley W. Stone, Robert J. Gilliom, and Charles G. Crawford

Open-File Report 2008–1186

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

U.S. Department of the Interior
DIRK KEMPTHORNE, Secretary

U.S. Geological Survey
Mark D. Myers, Director

U.S. Geological Survey, Reston, Virginia: 2008

For product and ordering information:
World Wide Web: <http://www.usgs.gov/pubprod>
Telephone: 1-888-ASK-USGS

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

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

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

Suggested citation:
Stone, W.W., Gilliom, R.J., and Crawford, C.G. 2008, Watershed Regressions for Pesticides (WARP) for Predicting Annual Maximum and Maximum Moving-Average Concentrations of Atrazine in Streams: U.S. Geological Survey Open-File Report 08-1186, 19 p.

Foreword

The U.S. Geological Survey (USGS) is committed to providing the Nation with credible scientific information that helps to enhance and protect the overall quality of life and that facilitates effective management of water, biological, energy, and mineral resources (<http://www.usgs.gov/>). Information on the Nation's water resources is critical to ensuring long-term availability of water that is safe for drinking and recreation and is suitable for industry, irrigation, and fish and wildlife. Population growth and increasing demands for water make the availability of that water, now measured in terms of quantity and quality, even more essential to the long-term sustainability of our communities and ecosystems.

The USGS implemented the National Water-Quality Assessment (NAWQA) Program in 1991 to support national, regional, State, and local information needs and decisions related to water-quality management and policy (<http://water.usgs.gov/nawqa>). The NAWQA Program is designed to answer: What is the condition of our Nation's streams and ground water? How are conditions changing over time? How do natural features and human activities affect the quality of streams and ground water, and where are those effects most pronounced? By combining information on water chemistry, physical characteristics, stream habitat, and aquatic life, the NAWQA Program aims to provide science-based insights for current and emerging water issues and priorities. From 1991-2001, the NAWQA Program completed interdisciplinary assessments and established a baseline understanding of water-quality conditions in 51 of the Nation's river basins and aquifers, referred to as Study Units (<http://water.usgs.gov/nawqa/studyu.html>).

Multiple national and regional assessments are ongoing in the second decade (2001–2012) of the NAWQA Program as 42 of the 51 Study Units are reassessed. These assessments extend the findings in the Study Units by determining status and trends at sites that have been consistently monitored for more than a decade, and filling critical gaps in characterizing the quality of surface water and ground water. For example, increased emphasis has been placed on assessing the quality of source water and finished water associated with many of the Nation's largest community water systems. During the second decade, NAWQA is addressing five national priority topics that build an understanding of how natural features and human activities affect water quality, and establish links between sources of contaminants, the transport of those contaminants through the hydrologic system, and the potential effects of contaminants on humans and aquatic ecosystems. Included are topics on the fate of agricultural chemicals, effects of urbanization on stream ecosystems, bioaccumulation of mercury in stream ecosystems, effects of nutrient enrichment on aquatic ecosystems, and transport of contaminants to public-supply wells. These topical studies are conducted in those Study Units most affected by these issues; they comprise a set of multi-Study-Unit designs for systematic national assessment. In addition, national syntheses of information on pesticides, volatile organic compounds (VOCs), nutrients, selected trace elements, and aquatic ecology are continuing.

The USGS aims to disseminate credible, timely, and relevant science information to address practical and effective water-resource management and strategies that protect and restore water quality. We hope this NAWQA publication will provide you with insights and information to meet your needs, and will foster increased citizen awareness and involvement in the protection and restoration of our Nation's waters.

The USGS recognizes that a national assessment by a single program cannot address all water-resource issues of interest. External coordination at all levels is critical for cost-effective management, regulation, and conservation of our Nation's water resources. The NAWQA Program, therefore, depends on advice and information from other agencies—Federal, State, regional, interstate, Tribal, and local—as well as nongovernmental organizations, industry, academia, and other stakeholder groups. Your assistance and suggestions are greatly appreciated.

Matthew C. Larsen
Acting Associate Director for Water

Contents

Foreword	iii
Abstract.....	1
Introduction.....	1
Purpose and Scope	2
Sources of Model Development and Validation Data.....	2
Statistical Methods.....	2
Calculation of Moving-Average Concentrations	2
Accuracy of Concentration Statistics	2
Regression Methods	3
Analysis of Model Fit.....	3
Estimation of Prediction Intervals	4
Accuracy of Concentration Statistics for Model Development Sites	4
Atrazine Models	4
Model Performance.....	4
Model Development Sites	4
Model Validation Sites	8
Uncertainty in Model Predictions	11
Model Limitations.....	15
Summary and Conclusions.....	16
References Cited.....	17
Appendix 1.....	19

Figures

1–4 Graphs showing:

1. Differences between observed (high-frequency sampling) and simulated (22 samples per year) atrazine concentration statistics for the four sites studied by the Water Quality Laboratory of Heidelberg College.5
2. Atrazine concentration statistics from concentrations observed at model development sites in relation to values of the same statistics predicted by the WARP models.7
3. Residual errors for atrazine concentration statistics predicted by WARP models in relation to watershed area for the 112 model development sites.9
4. Residual errors for atrazine concentration statistics predicted by the WARP models in relation to geographic region for the 112 model development sites.10
5. Map showing U.S. Department of Agriculture Farm Resource Regions (U.S. Department of Agriculture, 2000) and regions used for evaluation of the WARP models.....11

6–9 Graphs showing:

6. Map showing Atrazine concentration statistics from concentrations observed at model validation sites in relation to values of the same statistics predicted by the WARP models.....12
7. Residual errors for atrazine concentration statistics predicted by the WARP models using data from model development and validation sites.13
8. Residual errors for atrazine concentration statistics predicted by the WARP models in relation to the type of validation site.14
9. Potential prediction errors among concentration statistics predicted using the atrazine WARP models.15

Tables

1. Summary of results from the Monte Carlo analysis evaluating the accuracy of concentration statistics estimated using a reduced sampling frequency design for four WQL sites.....6
2. Summary of statistics and coefficients for the four atrazine WARP models.6
3. Summary of atrazine WARP model performance for model development and model validation sites.8

Conversion Factors

Multiply	By	To obtain
kilometer (km)	0.6214	mile (mi)
square kilometer (km ²)	247.1	acre
square kilometer (km ²)	0.3861	square mile (mi ²)
liter (L)	0.2642	gallon (gal)
kilogram (kg)	2.205	pound avoirdupois (lb)
microgram per liter (µg/L)	1	part per billion (ppb)

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

$$^{\circ}\text{F}=(1.8\times^{\circ}\text{C})+32$$

Abbreviations and Acronyms

µg/L	microgram per liter
L	liter
ARP	Acetochlor Registration Partnership
CI	confidence interval
DF	Dunne overland flow
K	soil erodibility factor (K-factor)
pR ²	pseudo R-squared (R-squared value for tobit regression)
NASQAN	National Stream-Quality Accounting Network
NAWQA	National Water-Quality Assessment
PI	prediction interval
R ²	coefficient of multiple determination
R	rainfall erosivity factor (R-factor)
UI	annual agricultural atrazine use in the watershed (kg)/ watershed area (km ²)
USGS	U.S. Geological Survey
USLE	Universal Soil Loss Equation
WA	watershed area (km ²)
WARP	Watershed Regressions for Pesticides
WQL	Water Quality Laboratory of Heidelberg College

Watershed Regressions for Pesticides (WARP) for Predicting Annual Maximum and Annual Maximum Moving-Average Concentrations of Atrazine in Streams

By Wesley W. Stone, Robert J. Gilliom, and Charles G. Crawford

Abstract

Regression models were developed for predicting annual maximum and selected annual maximum moving-average concentrations of atrazine in streams using the Watershed Regressions for Pesticides (WARP) methodology developed by the National Water-Quality Assessment Program (NAWQA) of the U.S. Geological Survey (USGS). The current effort builds on the original WARP models, which were based on the annual mean and selected percentiles of the annual frequency distribution of atrazine concentrations. Estimates of annual maximum and annual maximum moving-average concentrations for selected durations are needed to characterize the levels of atrazine and other pesticides for comparison to specific water-quality benchmarks for evaluation of potential concerns regarding human health or aquatic life.

Separate regression models were derived for the annual maximum and annual maximum 21-day, 60-day, and 90-day moving-average concentrations. Development of the regression models used the same explanatory variables, transformations, model development data, model validation data, and regression methods as those used in the original development of WARP. The models accounted for 72 to 75 percent of the variability in the concentration statistics among the 112 sampling sites used for model development. Predicted concentration statistics from the four models were within a factor of 10 of the observed concentration statistics for most of the model development and validation sites.

Overall, performance of the models for the development and validation sites supports the application of the WARP models for predicting annual maximum and selected annual maximum moving-average atrazine concentration in streams and provides a framework to interpret the predictions in terms of uncertainty. For streams with inadequate direct measurements of atrazine concentrations, the WARP model predictions for the annual maximum and the annual maximum moving-average atrazine concentrations can be used to characterize the probable levels of atrazine for comparison to specific water-quality benchmarks. Sites with a high probability of exceeding a benchmark for human health or aquatic life can be prioritized for monitoring.

Introduction

Pesticide concentrations in streams vary widely across the United States. Each pesticide has a unique pattern of occurrence because factors such as pesticide use, application practices and timing, climate, and watershed characteristics vary geographically (Gilliom and others, 2006b). Because of the geographic and temporal complexity of pesticide occurrence and concentrations, adequate monitoring of pesticide concentrations in the numerous streams in the United States that are potentially affected by pesticides is prohibitively expensive, particularly for sampling frequencies that are high enough to reliably estimate the concentration statistics needed for risk assessments. The lack of adequate direct measurements of pesticide concentrations leads to the need for tools that can be used to predict pesticide concentrations for unmonitored streams. Although such estimates cannot replace direct measurements when reliability requirements are high, they are useful for the initial screening-level steps of risk assessment and for efficiently guiding intensive monitoring.

Larson and others (2004) developed and applied a method of predicting annual frequency distributions of atrazine concentrations in streams using watershed characteristics. The method, known as Watershed Regressions for Pesticides (WARP), originated from the work of Larson and Gilliom (2001) and is based on empirical relations among pesticide concentrations in samples collected at monitoring sites and selected watershed characteristics available nationally, such as pesticide use, and soil and hydrologic characteristics. The models developed in these studies make it possible to predict specific concentration percentiles and to characterize the annual frequency distribution of concentrations of selected herbicides for unmonitored streams.

Since the development and application of WARP by Larson and others (2004), interest in predicting concentrations of atrazine and other pesticides in unmonitored streams has grown to include annual maximum and certain annual maximum moving-average pesticide concentrations in streams. For example, during the initial screening-level steps of risk assessment, water-quality benchmarks for evaluating potential concerns for human health or aquatic life are compared to

2 Watershed Regressions for Pesticides (WARP) for Predicting Annual Concentrations of Atrazine in Streams

estimated environmental concentrations, which for atrazine are represented by annual maximum and moving-average atrazine concentrations for specified durations.

Purpose and Scope

This report describes WARP models that were developed to predict annual maximum atrazine concentrations and annual maximum moving-average atrazine concentrations for selected durations. The WARP models are based on the most complete year of observations for each of 112 model development sites. This year is referred to as the “WARP development period.” The moving-average durations of 21, 60, and 90 days were selected to correspond to exposure durations used to estimate environmental concentrations for comparison to specific atrazine water-quality benchmarks for human health or aquatic life (U.S. Environmental Protection Agency, 2003).

Sources of Model Development and Validation Data

The 112 model development sites and water-quality data used by Larson and others (2004) also were used in this study. Data from these 112 sites were collected for the U.S. Geological Survey (USGS) National Water-Quality Assessment (NAWQA) and National Stream-Quality Accounting Network (NASQAN) programs. Model validation data from 26 sites collected by the Water Quality Laboratory (WQL) of Heidelberg College in Tiffin, Ohio, and the Acetochlor Registration Partnership (ARP) used by Larson and others (2004) also were used in this study. The model development and validation site selection procedure is discussed in detail in Larson and others (2004). The explanatory variables — atrazine use intensity, R-factor, K-factor, watershed area, and Dunne overland flow — used in the WARP models developed by Larson and others (2004) also were used in this study, as well as the transformations of the response and explanatory variables. The difference between the models developed in this study and those from Larson and others (2004) is the concentration statistics used as the response variables. This study used the annual maximum atrazine concentration and annual maximum moving-average atrazine concentrations (described in detail below), rather than annual frequency distributions of atrazine concentrations used by Larson and others (2004).

Statistical Methods

The statistical methods used were (1) calculation of annual maximum moving-average concentrations from the sample data, (2) assessment of the accuracy of estimated annual maximum and annual maximum moving-average concentrations for model development sites, and (3) construction of the specific regression models. Each is described below.

Calculation of Moving-Average Concentrations

The annual maximum moving-average concentrations of atrazine are the highest moving-average concentrations for the selected three durations— 21 days, 60 days, and 90 days— for each site during the year used for model development or validation. Computation of the moving-average concentrations of atrazine generally followed the methods described in Gilliom and others (2006a), which are provided in Appendix 1. Hourly atrazine concentrations were estimated for the entire period of record (multiple years) for each site through linear interpolation of actual observations. The hourly concentration estimates were averaged to obtain an estimated daily concentration. The hourly estimates facilitated computations for days with multiple samples, but were not used for other purposes. Moving-average concentrations for the selected durations (21, 60, and 90 days) were computed for each day; for example, the 21-day moving average for a particular day includes the average of that day and the 20 previous days. The estimated moving-average concentrations then were truncated to the 1-year WARP development period for each site, and the maximum moving-average concentrations were determined for each of the durations.

In some cases, insufficient observations were available for a particular site to calculate moving-average concentrations (21-, 60-, or 90-day durations) for the beginning of the 1-year WARP development period. For example, the WARP development period for a particular site begins April 1, 1997, and no observations were available prior to that date. In these situations, the actual observations for the 1-year WARP development period (in this case the data prior to April 1, 1998) were used as surrogate observations for the year before the WARP development period.

Nondetections, referred to as “censored observations” for statistical analysis, were assigned a value of zero for the process of linear interpolation and computation of moving-average concentrations (Appendix 1). Any annual maximum or annual maximum moving-average concentration estimated as less than 0.001 $\mu\text{g/L}$ was considered censored at the 0.001 $\mu\text{g/L}$ level for development of the models.

Accuracy of Concentration Statistics

The accuracy of the annual maximum and annual maximum moving-average concentrations computed from data for model development sites is a function of the number of samples collected and the timing of sample collection. The model-development site sampling frequency averaged 25 samples per year (NAWQA sampling design), which likely underestimates the annual maximum concentration in a stream (Crawford, 2004). To characterize the effects of sampling design on estimates of the annual maximum and annual maximum moving-average concentrations, four WQL sites were used in a Monte Carlo analysis following the approach used by Crawford (2004).

The four WQL sites used in the Monte Carlo analysis individually have 10 years of data (1989–1998) collected at higher frequencies than the model development data. Typically, samples were collected at the WQL sites at a frequency of 16 to 20 samples per month from mid-April to mid-August and 2 to 4 samples per month for the remainder of the year (Crawford 2004). The annual sample count for the four WQL sites ranged from 61 to 199, with an average of 97 samples per site collected each year. In addition, the WQL collected multiple samples during selected days to characterize runoff events.

For each site and year, a subset of the WQL data was selected on the basis of a reduced sampling frequency similar to that of the model development data. The reduced sampling frequency included one sample every 7 days for May through August, one sample every 14 days for April and September, one sample for October through December, and one sample for January through March (22 samples per year). The simulation for each site and year began by randomly selecting (all random selections in this analysis are without replacement) a sample from each day, creating a subset of WQL data. The random selection of a single sample for each day reduced the influence of runoff-event characterization on the final selection of samples, which was based on the reduced sampling frequency. From this subset of WQL data, one sample was selected randomly from the October through December series and one from the January through March series. The April through September samples in the analysis represented fixed time intervals. The initial sample for the first sampling period (April, September, and May) of a particular year was selected randomly. Subsequent samples were randomly selected within fixed time intervals of 6 to 8 days for May through August and 12 to 16 days for April and for September. The simulation process was repeated 1,000 times for each site and year. Replicate analysis showed the results based on 1,000 simulations to be equivalent. For each site and year, annual maximum and annual maximum moving-average concentrations were computed from the simulated subset of samples from the Monte Carlo analysis and compared to the concentration statistics computed from all the observations. The concentration statistics generated from the subsets of samples from the Monte Carlo analysis are referred to as “simulated concentration statistics” in subsequent discussions.

Regression Methods

Some of the annual maximum and annual maximum moving-average concentrations computed for sites included in the regression analysis were less than the censoring threshold. In the context of an explanatory model, a censored observation is one in which the value of the response variable could not be quantified below a specified level. Conventional least-squares methods for estimating parameters of the explanatory model, using either the entire sample or the subsample of complete observations, yield biased and inconsistent estimates (Judge

and others, 1985). Therefore, tobit regression methods (Judge and others, 1985; Tobin, 1958), as used in Larson and others (2004), were used in model development.

Measures of goodness of fit, such as the standard deviation of the residual error (commonly referred to as the root mean square error in conventional regression analysis) or the coefficient of multiple determination (R^2), used for conventional least-squares regression analysis, cannot be computed for the tobit regression model. The standard deviation of 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 degree of censoring. In this report, biased estimates of the standard deviation of residual error are referred to as “scale” in figures and tables. Several pseudo R^2 (pR^2) measures suitable for use with the tobit regression model have been proposed in the literature as alternatives to R^2 . For this study, pR^2 was calculated using the method of Laitila (1993). As with conventional R^2 , the 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 explained; 1 indicates all variation is explained).

Analysis of Model Fit

In this study, the WARP explanatory variables and transformations developed by Larson and others (2004) were not altered. Analysis of model goodness of fit focused on evaluating model performance with annual maximum concentrations and annual maximum moving-average concentrations instead of annual frequency distributions of atrazine concentrations as the response variables in the WARP models. Box and whisker plots (Tukey, 1977) were used to qualitatively assess model performance. These plots, also known as boxplots, summarize a group of data by showing a measure of central tendency (the median), the variation (interquartile range), the range (shown by the whiskers, which extend to only 1.5 times the interquartile range when extreme values are present), and extreme values (shown by individual points). Boxplots were used for displaying the distribution of model residuals and comparing residuals among groups of data (for example, different regions of the country).

Comparisons between predicted concentration statistics and concentration statistics computed from observations are made frequently in the discussion of model performance. Terms used for these comparisons are defined here for clarity. Concentration statistics generated by the WARP models are referred to as “predicted concentration statistics,” and concentration statistics computed from observations are referred to, for convenience, as “observed concentration statistics.” Predicted concentration statistics within a factor of 10, or “order of magnitude,” of the observed concentration statistics

4 Watershed Regressions for Pesticides (WARP) for Predicting Annual Concentrations of Atrazine in Streams

are between one-tenth and 10 times the observed concentration statistic. For example, for an observed concentration statistic of 3 µg/L, predicted concentration statistics from 0.3 to 30 µg/L are within a factor of 10 of the observed concentration statistic; predicted concentration statistics from 0.6 to 15 µg/L are within a factor of 5.

Estimation of Prediction Intervals

Prediction intervals were approximated using normal theory and the t-distribution. That is, methods for ordinary least square regression were used, which are only approximate when applied to censored data. The standard errors were estimated from the maximum likelihood scale parameter, using the adjustment suggested by Aitkin (1981).

Accuracy of Concentration Statistics for Model Development Sites

The Monte Carlo analysis indicated that the simulated annual maximum atrazine concentrations determined from the reduced sampling frequency (similar to model development sites) were biased low in comparison to annual maximum concentrations computed from much more frequently collected data from the four WQL sites in Ohio (fig. 1). The simulated annual maximum concentrations were within a factor of 2 of the annual maximum concentrations derived from observed concentrations for 51 to 66 percent of the simulations, and within a factor of 5 for 94 to 97 percent of the simulations (table 1). Quantification of the degree of underestimation derived by this Monte Carlo analysis should be tempered by the knowledge that the annual maximum concentrations derived from observations at each of the WQL sites are likely underestimates of the true annual maximum concentrations because of the limitations of any sampling frequency.

In contrast to findings for the annual maximum, the comparison of simulated annual maximum moving-average concentrations from the reduced sampling frequency (similar to model development sites) to those derived from the observed concentrations for the WQL sites did not show a bias for any of the three durations (fig. 1). The simulated annual maximum 21-day, 60-day, and 90-day moving-average concentrations were within a factor of 2 of the observed concentration statistics for more than 85 percent of the simulations (table 1). This limited analysis indicates that the sampling frequency for the model development sites was adequate for yielding unbiased estimates of annual maximum 21-day, 60-day, and 90-day moving-average concentrations.

Atrazine Models

The models for the annual maximum and the annual maximum 21-day, 60-day, and 90-day moving-average concentrations of atrazine have the same form as reported by Larson and others (2004):

$$\log_{10}(\text{concentration}) = f[(UI)^{1/4}, \log_{10}(R), K, (WA)^{1/2}, DF] \quad (1)$$

where

UI is the atrazine use intensity, annual agricultural atrazine use in the watershed (kg) divided by watershed area (km²);

R is the rainfall erosivity factor (R-factor) from the Universal Soil Loss Equation (USLE);

K is the soil erodibility factor (K-factor) from USLE;

WA is the watershed area (km²); and

DF is the Dunne overland flow, the percentage of total streamflow derived from surface runoff caused by precipitation on saturated soil.

Statistics for all four models are given in table 2. Regression coefficients and statistics were based on model fit using tobit regression. Four of the five variables were significant ($p < 0.05$) in all models; watershed area, which is included in the models for annual maximum and annual maximum 21-day moving-average concentrations, was the exception. Values of pR^2 ranged from 0.72 for the annual maximum concentration model to 0.75 for the annual maximum 60- and 90-day moving-average concentration models, meaning that the models accounted for 72 to 75 percent of the variability in the concentration statistics among the 112 sites used for model development. Concentration statistics predicted by these models represent the median expected concentration for all sites with the same values for the explanatory variables.

Model Performance

Model performance was evaluated by assessment of (1) goodness of fit and residual errors for the model development sites, (2) residual errors for the model validation sites, and (3) uncertainty in model predictions. Each is described below.

Model Development Sites

Plots of the goodness of fit of the four regression models are shown in figure 2. For all four models, 92 percent or more of the predicted concentration statistics were within a factor of 10, and 75 percent or more were within a factor of 5 of the observed concentration statistics at the development sites (table 3).

Residual errors were calculated by subtracting the predicted concentration statistics from the observed concentration statistics; therefore, residual errors less than zero indicate over-prediction of the concentration statistic, and residual errors greater than zero indicate under-prediction. A residual error of zero indicates exact agreement; residuals of -1 and +1 indicate that the prediction for a given site was 10 times and

one-tenth of the observed concentration statistic, respectively. Boxplots of the residual errors from the models show the performance of the models with respect to geographic region and watershed area for the model development sites. Residual errors grouped by watershed area and geographic region are shown in figures 3 and 4, respectively.

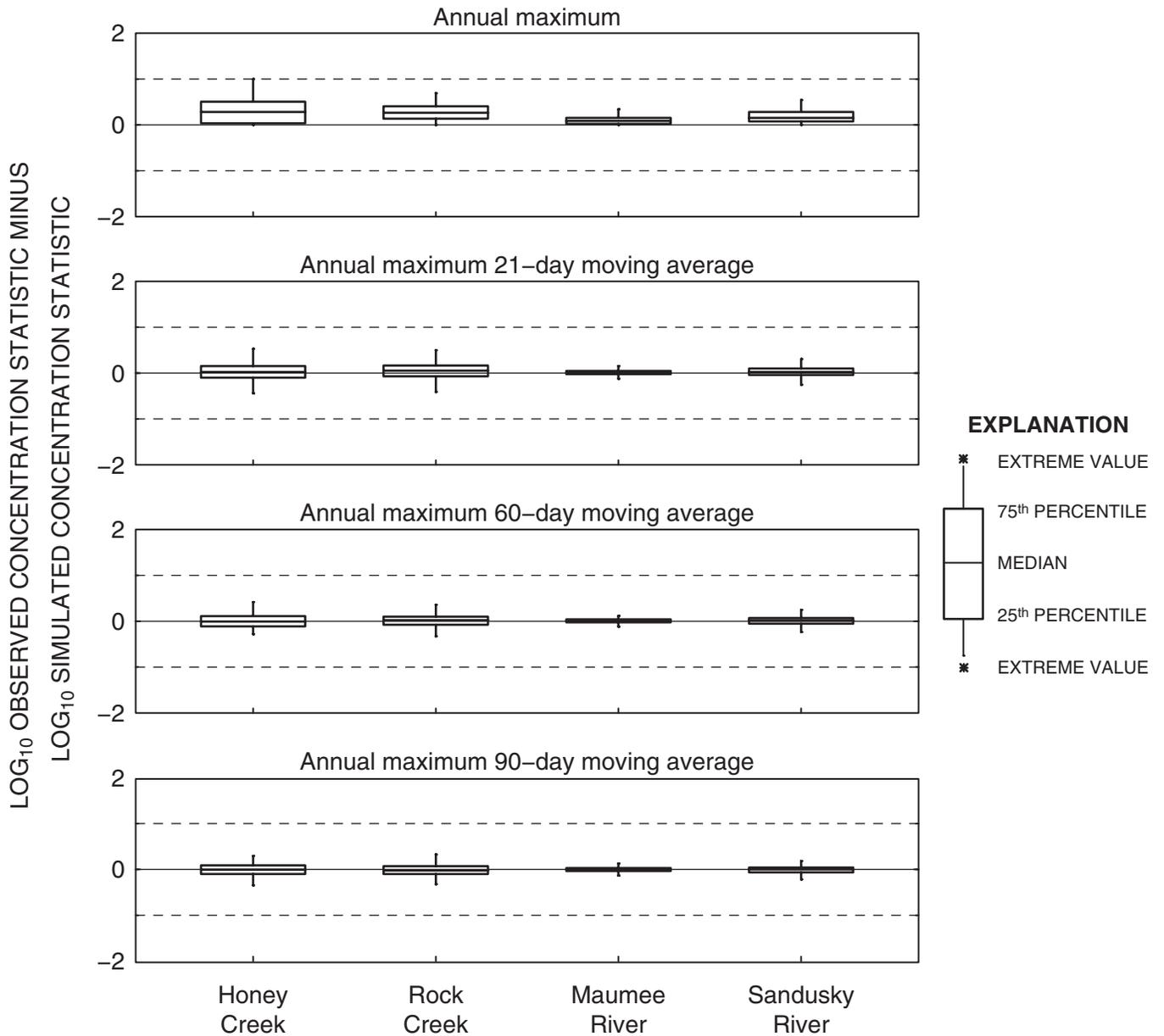


Figure 1. Differences between observed (high-frequency sampling) and simulated (22 samples per year) atrazine concentration statistics for the four sites studied by the Water Quality Laboratory of Heidelberg College. Simulated annual maximum and annual maximum moving-average concentration statistics were generated from Monte Carlo simulations in which 22 samples were selected from each WQL site for each year (10 years of data). Each WQL site has from 61 to 199 observations per year. The analysis was replicated 1,000 times for each site and year; each boxplot represents 10,000 simulated values. Outliers are not shown.

6 Watershed Regressions for Pesticides (WARP) for Predicting Annual Concentrations of Atrazine in Streams

Table 1. Summary of results from the Monte Carlo analysis evaluating the accuracy of concentration statistics estimated using a reduced sampling frequency design for four WQL sites.

[Percentage of simulated concentration statistics (22 samples per year) within a factor of 10, 5, and 2 of the observed concentration statistics (high-frequency sampling) are shown for each of the four sites. Simulated concentration statistics “within a factor of x” are between (1/x) of the (observed concentration statistic) and (x) times the (observed concentration statistic); WQL, Water Quality Laboratory of Heidelberg College; Max, annual maximum concentration; 21-day, annual maximum 21-day moving-average concentration; 60-day, annual maximum 60-day moving-average concentration; 90-day, annual maximum 90-day moving-average concentration]

Duration	Percentage of simulated concentration statistics within a factor of 10, 5, or 2											
	Honey Creek (Ohio)			Rock Creek (Ohio)			Maumee River (Ohio)			Sandusky River (Ohio)		
	10	5	2	10	5	2	10	5	2	10	5	2
Max	100	94	51	100	97	60	100	97	66	100	97	64
21-day	100	100	85	100	100	87	100	100	87	100	100	87
60-day	100	100	95	100	100	97	100	100	97	100	100	97
90-day	100	100	97	100	100	99	100	100	99	100	100	99

Table 2. Summary of statistics and coefficients for the four atrazine WARP models.

[WARP, Watershed Regression for Pesticides; Pseudo R-square, R-squared value for tobit regression; Scale, tobit regression analogue of the root mean squared error obtained from ordinary least squares regression; $UI^{1/4}$, fourth root of atrazine use intensity; $\log_{10}(R)$, \log_{10} of R-factor; K, K-factor; $(WA)^{1/2}$, square root of watershed area; DF, Dunne overland flow; <, less than; Max, annual maximum concentration from the WARP model; 21-day, annual maximum 21-day moving-average concentration from the WARP model; 60-day, annual maximum 60-day moving-average concentration from the WARP model; 90-day, annual maximum 90-day moving-average concentration from the WARP model]

Model	Regression coefficients (p -value)						Pseudo R-square	Scale	Percentage of censored observations
	Intercept	$UI^{1/4}$	$\log_{10}(R)$	K	$(WA)^{1/2}$	DF			
Max	-3.98 (<0.001)	0.90 (<0.001)	0.95 (<0.001)	3.14 (<0.001)	0.00016 (0.510)	-0.14 (0.004)	0.72	0.58	1
21-day	-4.09 (<0.001)	0.86 (<0.001)	0.90 (<0.001)	3.17 (<0.001)	0.00039 (0.085)	-0.12 (0.006)	0.74	0.54	1
60-day	-4.29 (<0.001)	0.82 (<0.001)	0.90 (<0.001)	3.33 (<0.001)	0.00050 (0.023)	-0.12 (0.006)	0.75	0.51	2
90-day	-4.34 (<0.001)	0.80 (<0.001)	0.90 (<0.001)	3.25 (<0.001)	0.00052 (0.015)	-0.11 (0.006)	0.75	0.50	2

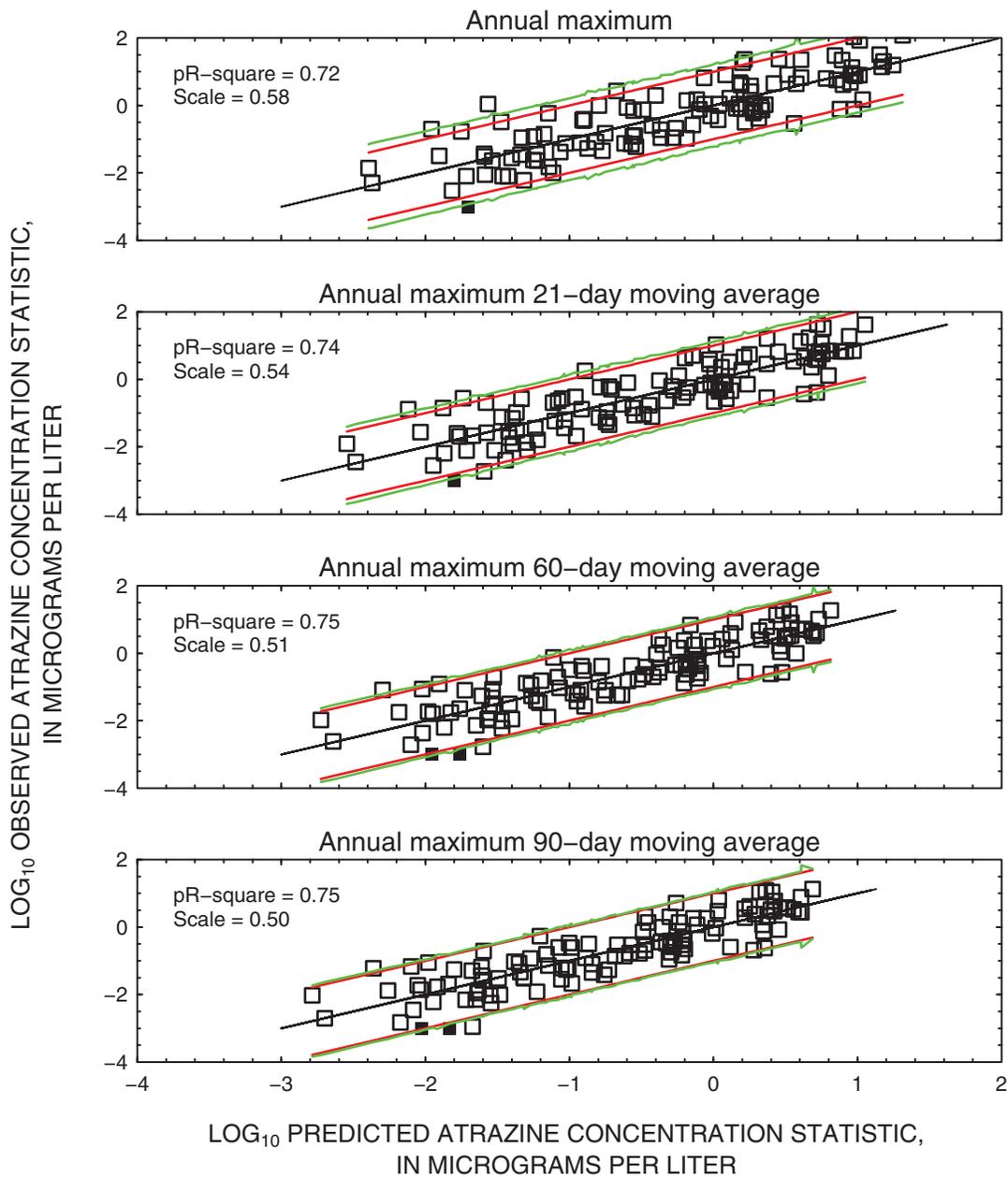


Figure 2. Atrazine concentration statistics from concentrations observed at model development sites in relation to values of the same statistics predicted by the WARP models. The black line is a 1:1 line, indicating exact agreement of the observed and predicted concentration statistics. The red lines are plus and minus one log unit from the 1:1 line; the area between a red and black line represents predicted concentration statistics within a factor of 10 of the observed concentration statistics. The green lines represent the 95-percent prediction interval. Filled symbols indicate censored observed concentration statistics. pR-square is the pseudo R-squared (R-square value used for tobit regression).

8 Watershed Regressions for Pesticides (WARP) for Predicting Annual Concentrations of Atrazine in Streams

Table 3. Summary of atrazine WARP model performance for model development and model validation sites.

[WARP, Watershed Regression for Pesticides; Percentage of predicted concentration statistics within a factor of 10, 5, and 2 of the observed concentration statistics are shown for each of the four models. Predicted concentration statistics “within a factor of x” are between $(1/x) \times (\text{observed concentration statistic})$ and $x \times (\text{observed concentration statistic})$; N, number of sites; Max, annual maximum concentration from the WARP model; 21-day, annual maximum 21-day moving-average concentration from the WARP model; 60-day, annual maximum 60-day moving-average concentration from the WARP model; 90-day, annual maximum 90-day moving-average concentration from the WARP model]

Model	Percentage of predicted concentration statistics within a factor of 10, 5, or 2					
	Model development sites (N=112)			Model validation sites (N=26)		
	10	5	2	10	5	2
Max	92	75	34	96	73	46
21-day	92	82	35	96	85	58
60-day	95	83	43	96	85	62
90-day	97	84	43	96	85	62

Predictions of atrazine concentration statistics were not biased with respect to watershed area, as indicated by figure 3. For this plot, watershed areas of the 112 model development sites were divided into quintiles (that is, the first quintile contains the smallest 20 percent of watershed areas, the second quintile contains the second smallest 20 percent, and so on). Residual errors were similar for all five groups, which cover more than five orders of magnitude in watershed area (17 to 2,965,000 km²).

Predictions of atrazine concentration statistics had no substantial regional bias, as indicated by figure 4. All four models had a slight tendency to under-predict concentration statistics for Region 4 (Northern Great Plains and Prairie Gateway; fig. 5). However, most of the predicted concentration statistics for this region were within a factor of 10 of the observed concentration statistics for all four models. The regional groupings used for this plot were based on the U.S. Department of Agriculture Farm Resource Regions (U.S. Department of Agriculture, 2000), which categorized agricultural regions of the conterminous United States on the basis of climate, topography, soil types, and dominant agricultural activities. The nine Farm Resource Regions were consolidated into five regions (fig. 5) so that each region would have sufficient sites for the computation of statistics summarizing model fit.

Model Validation Sites

The four models were applied to concentration data from the 26 validation sites, and plots of the predictions for these sites are shown in figure 6. A summary of the model performance for the validation sites is shown in table 3. Predicted concentration statistics from the four models were within a factor of 10 of the observed concentration statistics for 96

percent of the sites, and within a factor of 5 of the observed concentration statistics for 73 to 85 percent of the sites.

Larson and others (2004) found that WARP predictions were biased low for most of the validation sites, and this under-prediction was attributed to a combination of higher sampling frequency and the geographic location of the WQL validation sites. A similar under-prediction bias was evident in the WARP models of this study when the models were applied to concentrations from the validation sites, which were the same validation sites and data used by Larson and others (2004) (fig. 7).

The under-prediction bias for the model validation sites is further illustrated in figure 8, which compares the residual errors for USGS and ARP sites to the WQL sites. WARP predictions for WQL sites were biased low, whereas the predictions were not biased for USGS and ARP sites (fig. 8). An average of 21 samples per year were collected at USGS and ARP validation sites, whereas an average of 65 samples per year were collected at WQL sites. The under-prediction bias for the annual maximum concentrations was expected due to differences in sampling frequencies among the model development sites and the more intensively sampled WQL model validation sites. An average of 25 samples per year were collected at model development sites (NAWQA sampling design), compared to the 65 samples per year average for the WQL sites. At the model development sites, samples collected were not likely to contain the highest concentrations of atrazine during the year; therefore, the computed annual maximum concentrations used in model development would be biased low, leading to under-prediction. The sample design effect of underestimating the annual maximum concentrations for the sampling frequencies similar to that of the model development sites is shown in figure 1.

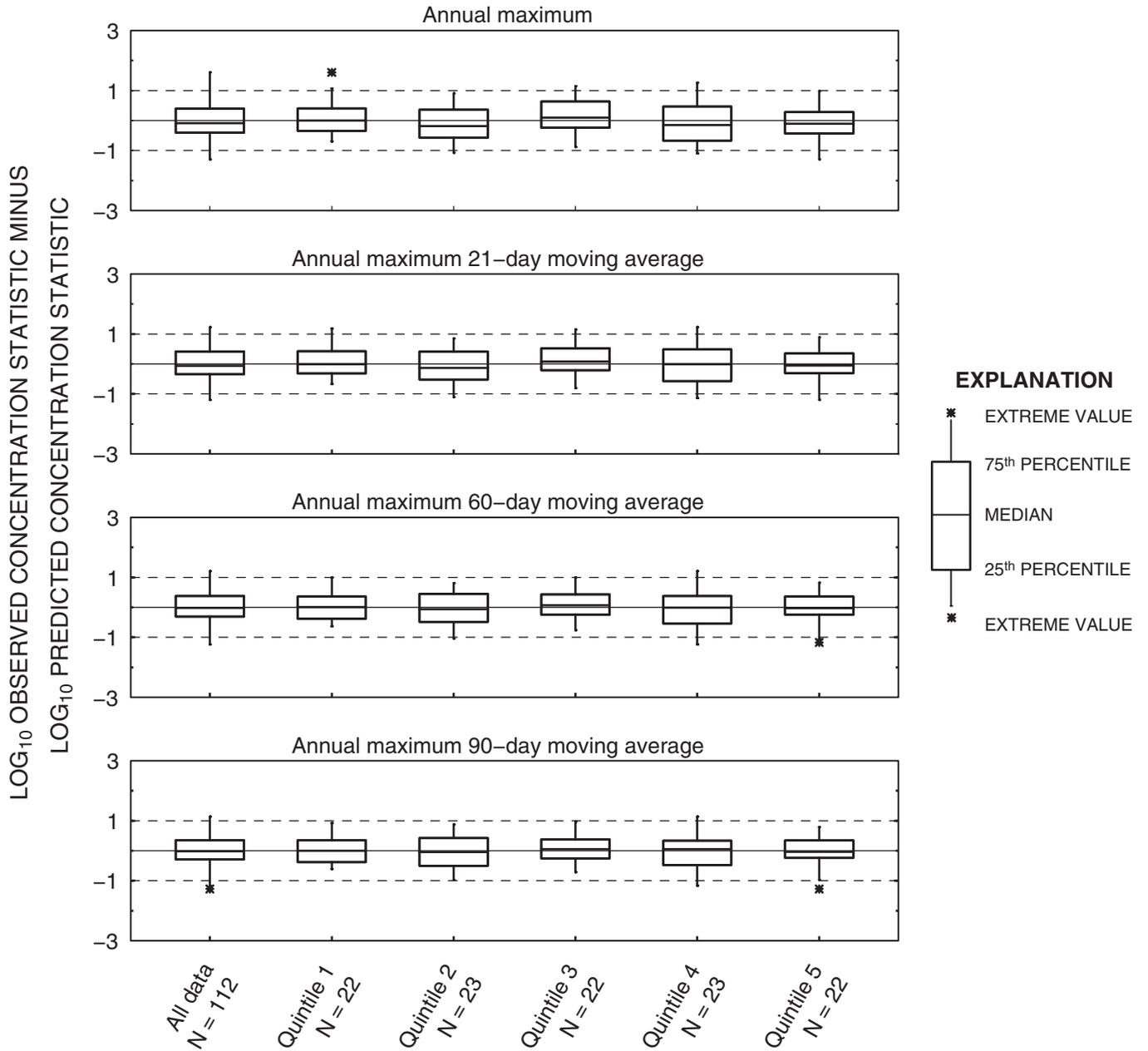


Figure 3. Residual errors for atrazine concentration statistics predicted by WARP models in relation to watershed area for the 112 model development sites. Sites are grouped by quintiles of watershed area. First boxplot in each group shows the residual errors for all 112 sites. Remaining boxplots show the residual errors for sites grouped into five classes based on quintiles of watershed area. Residual error is $[\log_{10}(\text{observed concentration statistic}) - \log_{10}(\text{predicted concentration statistic})]$. N is the number of sites.

10 Watershed Regressions for Pesticides (WARP) for Predicting Annual Concentrations of Atrazine in Streams

In contrast to the annual maximum concentrations, the Monte Carlo analysis did not indicate bias in the estimates of maximum moving-average concentrations as a result of sampling frequency differences between WQL sites and model development sites (fig. 1). The geographic location of the validation sites, however, may have contributed to the model under-prediction bias, which was primarily evident for WQL

sites. The nine WQL sites are in central and northern Ohio. The model under-prediction bias for these specific validation sites may be the result of local or regional variables affecting pesticides in runoff that are not included in the nationally based models. Field runoff in the eastern Corn Belt, which includes central and northern Ohio, is somewhat higher than in other regions of the United States (Capel and Larson,

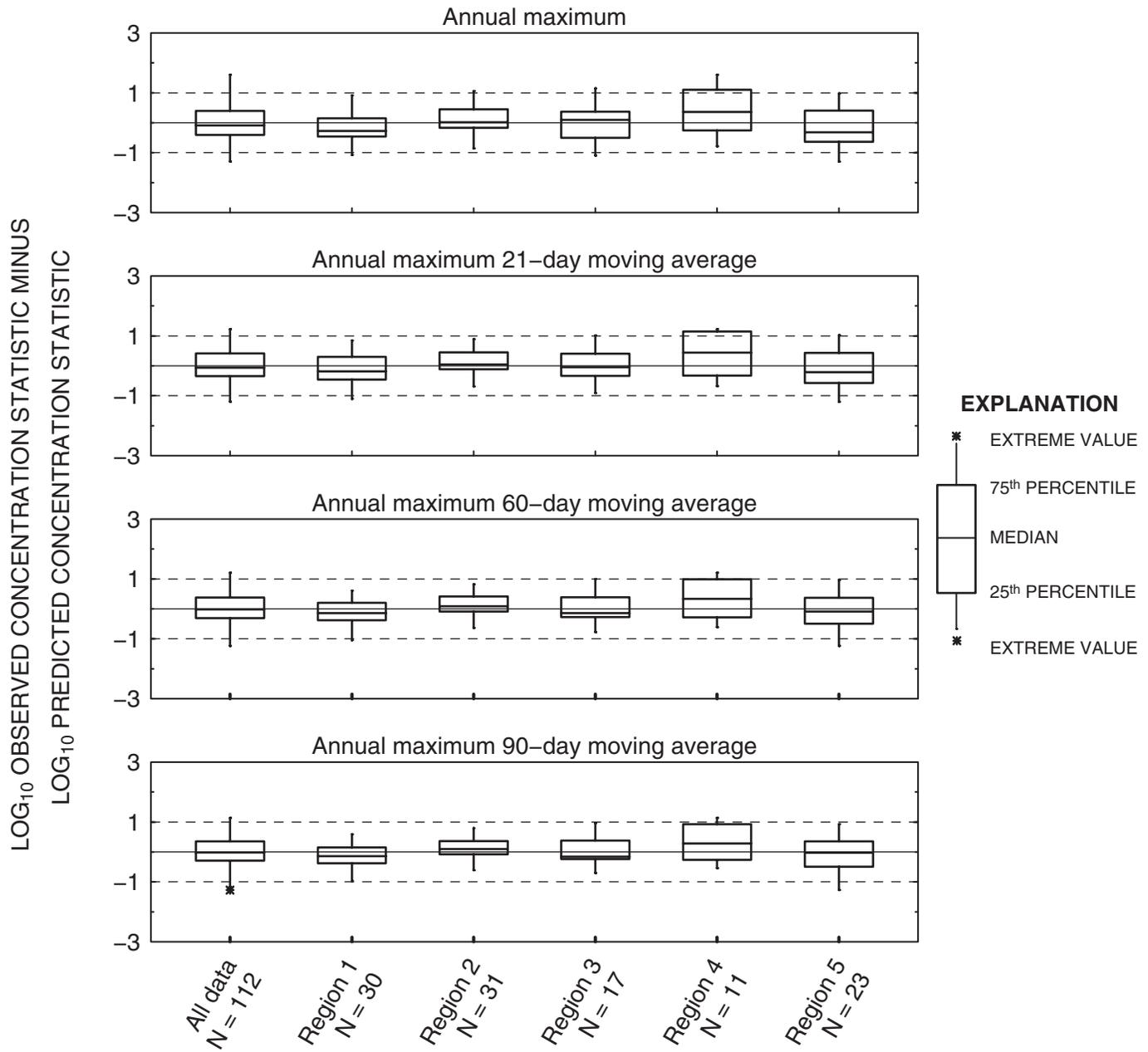


Figure 4. Residual errors for atrazine concentration statistics predicted by the WARP models in relation to geographic region for the 112 model development sites. First boxplot in each group shows the residual errors for all 112 sites. Remaining boxplots show the residual errors for sites grouped by region. Regions are shown in figure 5. Residual error is $[\log_{10}(\text{observed concentration statistic}) - \log_{10}(\text{predicted concentration statistic})]$. N is the number of sites.

2001). This may have contributed to the under-prediction bias for these validation sites by models developed from widely geographically distributed sampling sites. Predictions for the development sites in northwestern Ohio also showed a pattern of under-prediction similar to that observed with the validation sites.

Uncertainty in Model Predictions

Uncertainty in the prediction of a concentration statistic can be expressed in terms of a prediction interval (PI) for a specified confidence level—the confidence level used in this study is 95 percent. Conceptually, each predicted concentration statistic is the median estimate of the particular concentration statistic (annual maximum and annual maximum moving averages) for all the stream sites that have the same combination of values for the explanatory variables. The PI is the range

of values for a concentration statistic within which 95 percent of the actual concentration-statistic values are expected to occur for all stream sites with the same values of explanatory variables. In addition, the PI can be interpreted as the range within which the actual concentration statistic for an individual site and year is expected to fall 95 percent of the time.

The PIs for the four models are shown as green lines in figure 2. Concentration statistics are expressed as logarithms in these plots, resulting in symmetrical intervals for the PI (the high and low bounds of the intervals are the same distance from the predicted value). However, expressing the concentration statistics as logarithms obscures the fact that the intervals are skewed – the upper part of the PI interval covers a wider range of values than the lower part. Comparison of the PIs (green lines) to the exact agreement between predicted concentration statistics and observed concentration statistics (black line) and to the predicted concentration statistics within

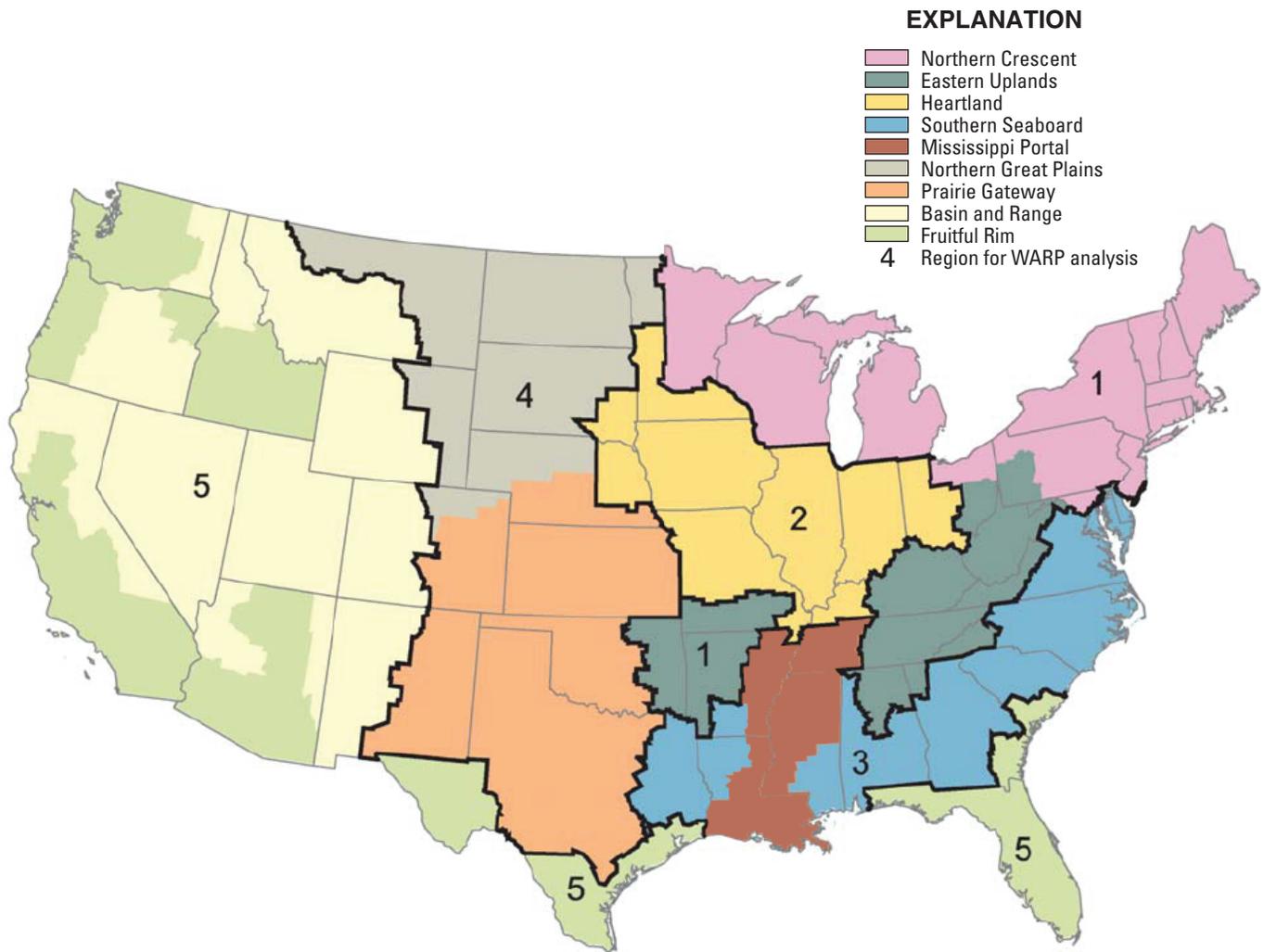
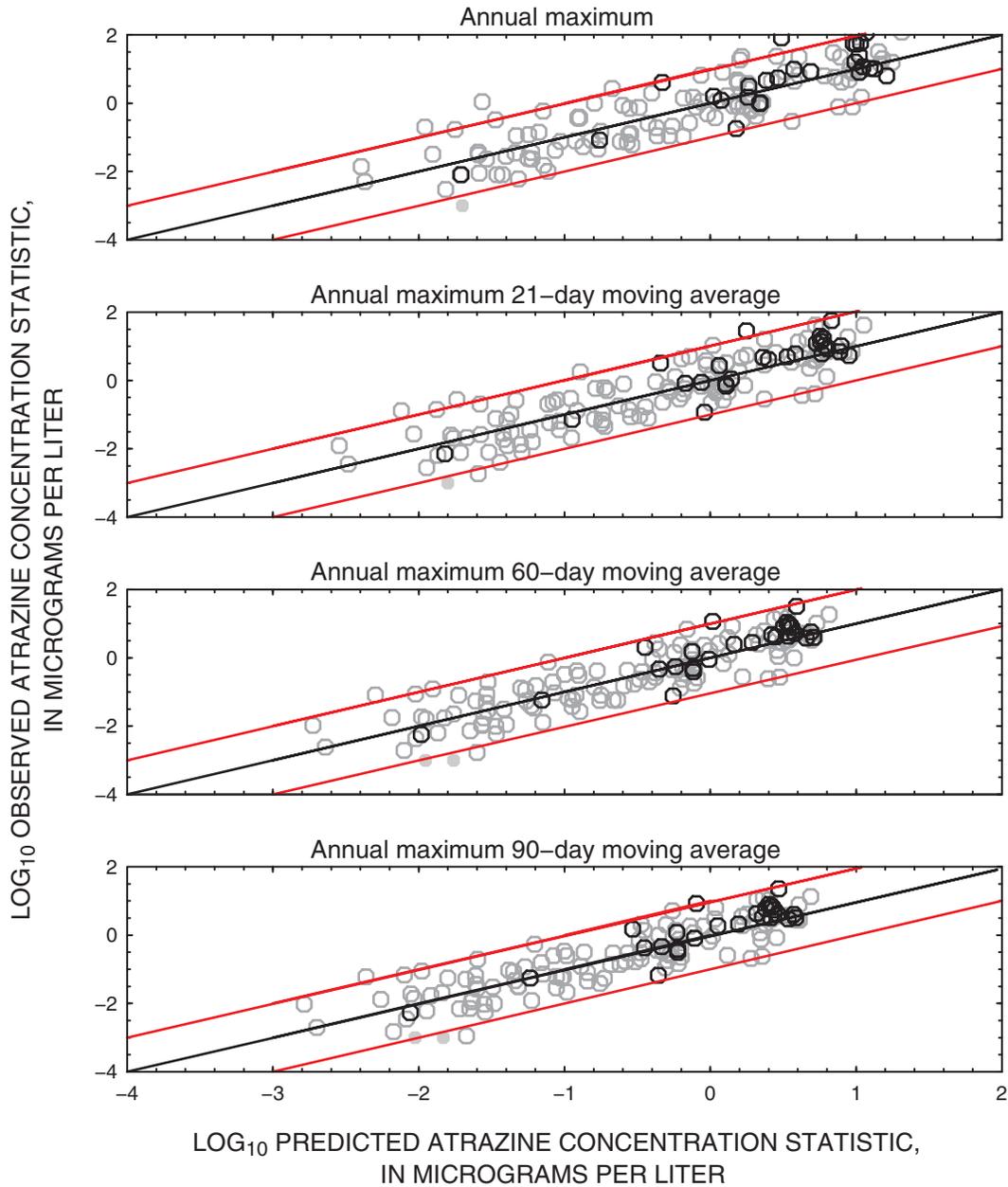


Figure 5. U.S. Department of Agriculture Farm Resource Regions (U.S. Department of Agriculture, 2000) and regions used for evaluation of the WARP models.



EXPLANATION

- Model development sites
- Model validation sites
- Censored observations for model defelopment sites

Figure 6. Atrazine concentration statistics from concentrations observed at model validation sites in relation to values of the same statistics predicted by the WARP models. Values for the model development sites are shown for comparison. The black line is a 1:1 line, indicating exact agreement of the observed concentration statistics and predicted concentration statistics. Red lines correspond to predictions of one-tenth (upper line) and 10 times (lower line) the observed concentration statistics. Filled symbols indicate censored observed concentration statistics.

a factor of 10 of the observed concentration statistics (red lines) for each of the models shows that the PIs were largest for the annual maximum model and smallest (PIs near or equal to a factor of 10) for the annual maximum 90-day moving-average model (fig. 2). The PI is a function of the fit of the model and the amount of variability explained by the model. The annual maximum 90-day moving-average model had a better model fit and explained more variability (scale = 0.50; $pR^2 = 0.75$) than the annual maximum model (scale = 0.58; $pR^2 = 0.72$).

The levels of uncertainty among the four models are compared in figure 9. The size of the PIs is represented as the ratio of the upper boundary of the prediction interval to the predicted atrazine concentration statistic (this is the same as the ratio of the predicted concentration statistic to the lower boundary of the interval). The extreme values (shown as asterisks) are for the same four sites for each of the models, which have one or more explanatory variables that are relatively extreme in value when compared to values for the rest of the sites (fig. 9). The highest ratios, shown by the asterisks well

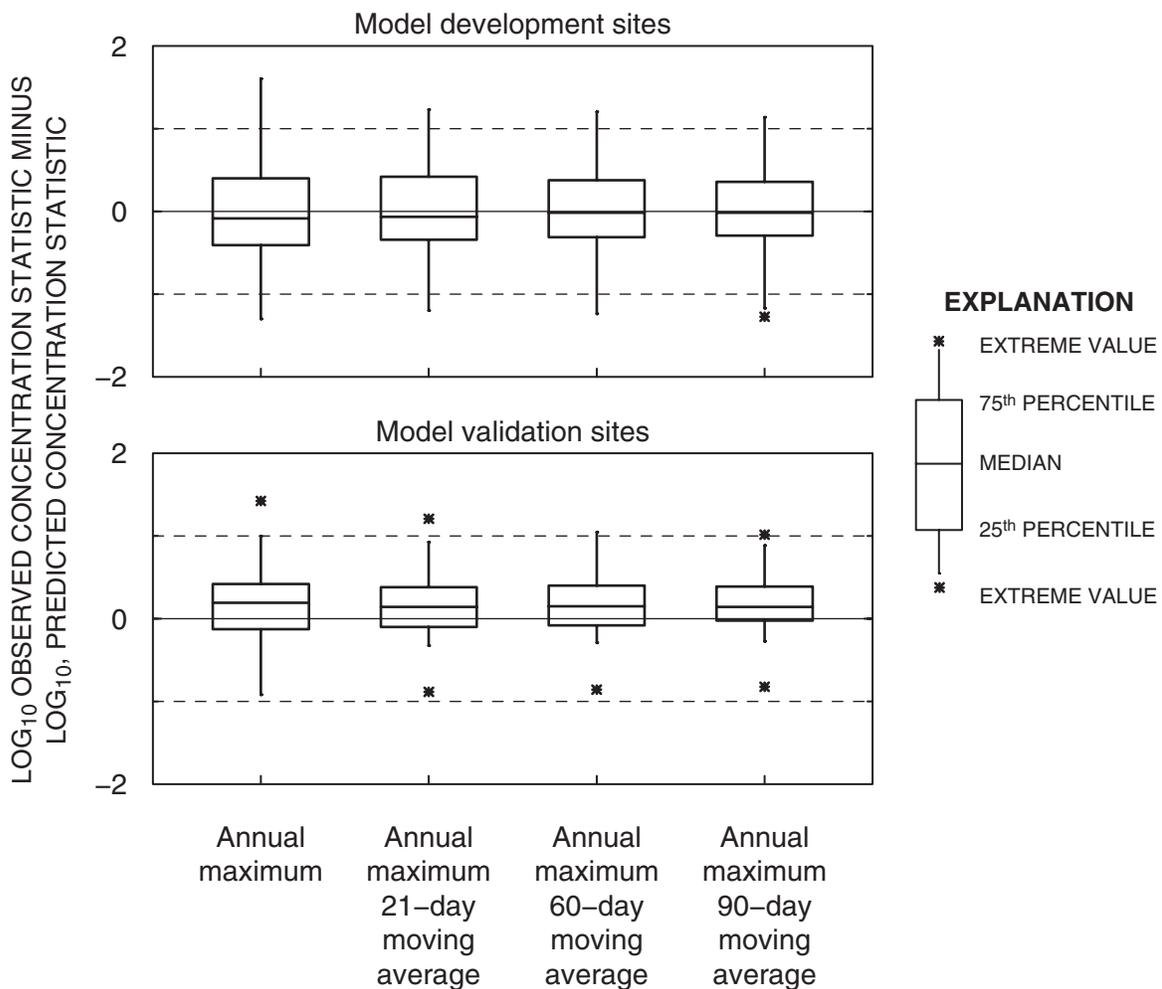


Figure 7. Residual errors for atrazine concentration statistics predicted by the WARP models using data from model development and validation sites. First boxplot in each group shows the residual errors for the annual maximum atrazine concentration model. Remaining boxplots show the residual errors for the annual maximum moving-average concentration statistic models grouped by duration. Residual error is $[\log_{10}(\text{observed concentration statistic}) - \log_{10}(\text{predicted concentration statistic})]$.

above the others, are for the Mississippi River at St. Francisville, La, which has a watershed area far greater than the rest of the sites. PIs for the annual maximum moving-average models (21-, 60-, and 90-day) extend to a factor of 10 to 14

above and below the predicted concentration statistic for most sites. PIs for the annual maximum model extend to a factor of 16 to 18 above and below the predicted annual maximum concentration for most sites.

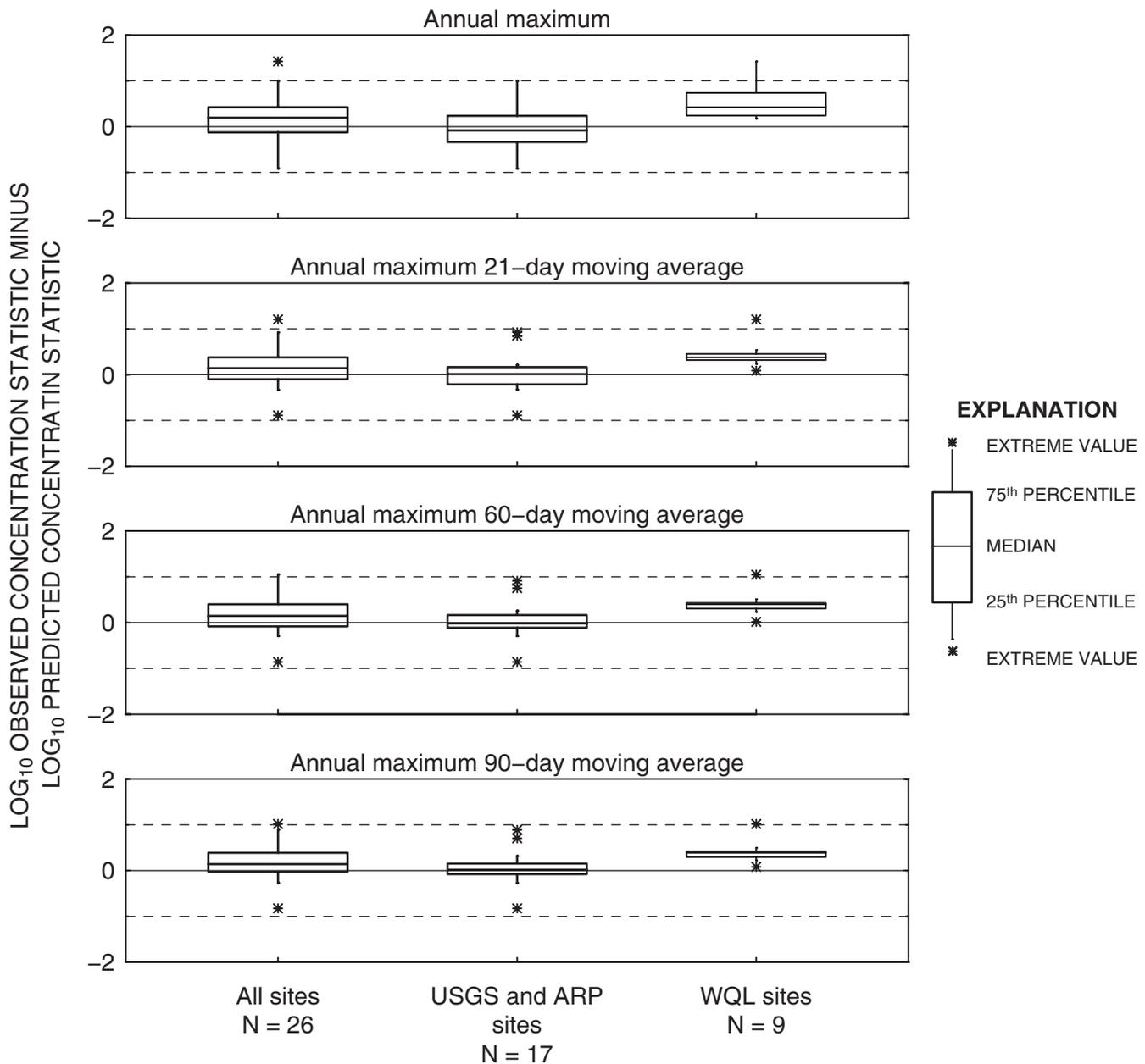


Figure 8. Residual errors for atrazine concentration statistics predicted by the WARP models in relation to the type of validation site. First boxplot in each group shows the residual errors for all 26 sites. Second boxplot in each group shows the residual errors for the USGS and ARP sites. Third boxplot in each group shows the residual errors for the WQL sites. Residual error is $[\log_{10}(\text{observed concentration statistic}) - \log_{10}(\text{predicted concentration statistic})]$. ARP, Acetochlor Registration Partnership. WQL, Water Quality Laboratory of Heidelberg College. N, number of sites.

Model Limitations

Use of the regression models for predicting atrazine concentration statistics, and the WARP methodology in general, are subject to the following limitations.

1. The sampling frequencies of the model development sites were not sufficient to reliably characterize the highest concentrations during a year. Thus, application of the models to predict the annual maximum concentration is expected to under-predict the actual annual maximums.
2. The regression models are designed for prediction of atrazine concentration statistics for streams of the conterminous United States. Although the 112 sites used for model development represent a wide variety of environmental settings and a large range of watershed areas, it is likely that some watersheds have one or more characteristics outside the ranges of the watershed parameters used to develop the models. Application of the models to streams draining such watersheds would result in increased uncertainty in predicted concentrations.
3. The models were developed using concentration data from streams. Application of the models to lakes or reservoirs would likely result in biased predictions.
4. The atrazine-use data used in the models are estimates for applications to agricultural land only. Substantial non-agricultural use of atrazine in a watershed could result in under-prediction of atrazine concentrations in a stream, if such use cannot be estimated. In addition, the atrazine-use data are based on periodic summaries of data on land use, agricultural crops, and pesticide use with a gap of 3 to 5 years between updates. Substantial changes in farming practices or pesticide use in a watershed between updates of the data could result in reduced accuracy of atrazine concentration predictions. This would be more likely for streams draining small agricultural watersheds and watersheds undergoing changes in land use (for example, urban areas) than for large rivers.

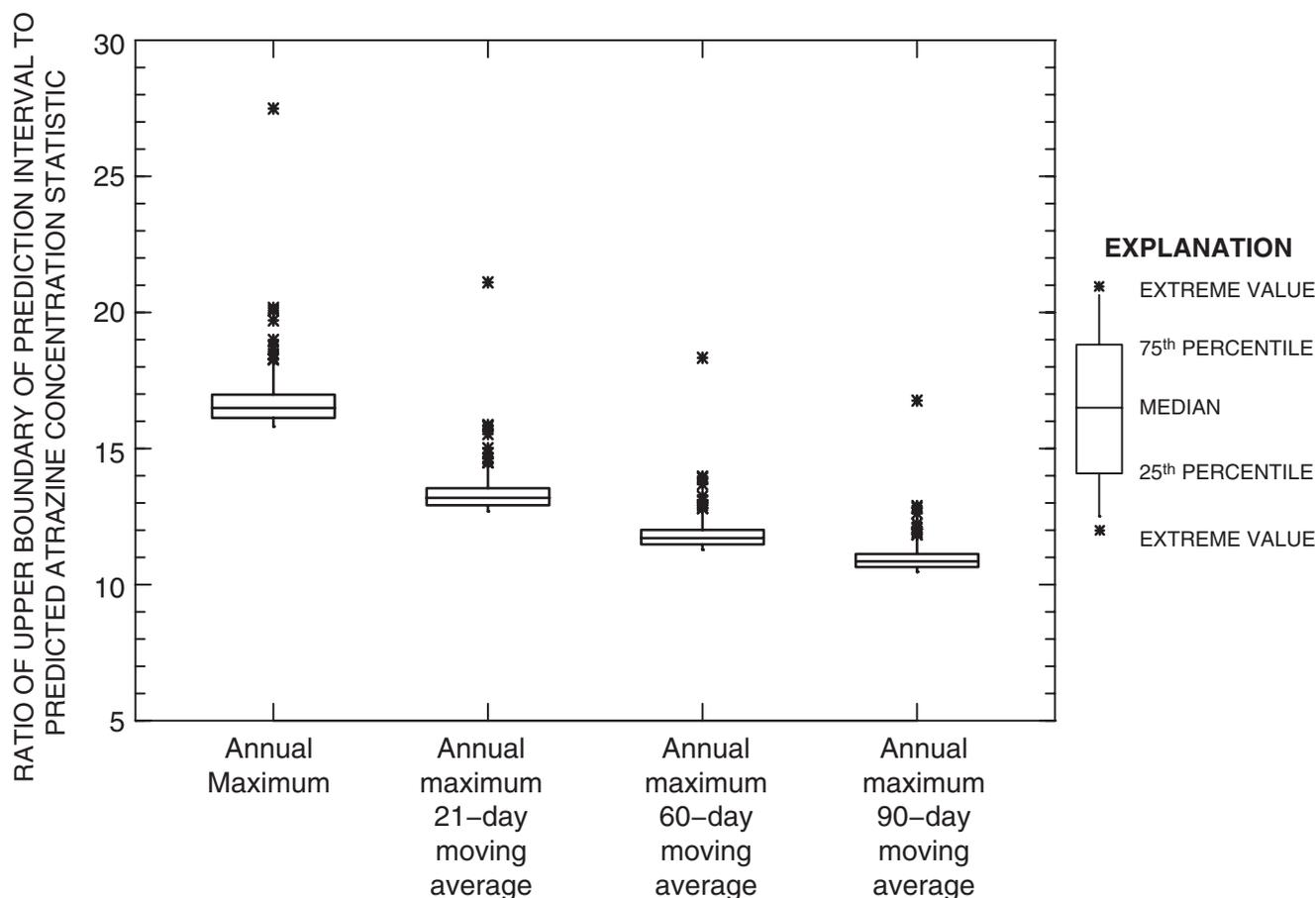


Figure 9. Potential prediction errors among concentration statistics predicted using the atrazine WARP models. Potential prediction error is represented by the ratio of the upper boundary of the 95-percent prediction interval to the predicted atrazine concentration statistic. Each boxplot shows the ratios among the 112 model development sites.

Summary and Conclusions

Regression models were developed for predicting annual maximum and annual maximum moving-average atrazine concentrations in streams, using nationally available data on watershed characteristics and atrazine use. Annual maximum and annual maximum moving-average concentration predictions generated by these models can be used to characterize the levels of atrazine for comparison to specific water-quality benchmarks for evaluation of potential concerns regarding human health and aquatic life.

Separate models were developed for the annual maximum concentration and the annual maximum 21-day, 60-day, and 90-day moving-average concentrations. The models accounted for 72 to 75 percent of the variability in concentration statistics among the 112 nationwide sites used for model development. Uncertainty in predicted concentration statistics was expressed in terms of prediction intervals. For the four models, 95-percent prediction intervals extend to a factor of 10 to 18 above and below the predicted concentration statistic in most cases.

Results for the 26 model validation sites show that concentration statistics were predicted within a factor of 10 of the observed concentration statistics in nearly all cases. All four models showed a bias toward under-prediction for the validation sites when considered together, but most of the under-prediction occurred for the nine WQL validation sites in central and northern Ohio. The tendency of the annual maximum model to under-predict concentrations was expected because Monte Carlo analysis showed that sampling frequencies similar to those for the model development sites would yield estimates of annual maximum concentrations that were

consistently biased low. However, results of the Monte Carlo analysis did not indicate bias for annual maximum moving-average concentrations. Geographic location of the WQL validation sites is likely a primary cause of model under-prediction of the annual maximum moving-average concentrations for these sites. Annual maximum moving-average concentration predictions for these WQL sites in central and northern Ohio were consistently biased low when compared to the predicted concentrations for other validation sites. The model under-prediction for these specific validation sites may be the result of local or regional variables that affect pesticides in runoff that are not included in the nationally based models.

Adequate monitoring of pesticide concentrations in the numerous streams in the United States is prohibitively expensive, particularly for sampling frequencies that are high enough to reliably estimate the concentrations statistics needed for risk assessments. The WARP models are tools for predicting pesticide concentrations in unmonitored or inadequately monitored streams. Overall, performance of the models for the development and validation sites supports the application of the WARP models for predicting annual maximum and selected annual maximum moving-average atrazine concentrations in streams and provides a framework to interpret the predictions in terms of uncertainty. For streams with inadequate direct measurements of atrazine, the WARP model predictions for the annual maximum and the annual maximum moving-average atrazine concentrations can be used to characterize the probable levels of atrazine for comparison to specific water-quality benchmarks. Sites with a high probability of exceeding a benchmark for human health or aquatic life could be prioritized for monitoring.

References Cited

- Aitken, M., 1981, A note on the regression analysis of censored data: *Technometrics*, v. 23, no. 2, p. 161–163.
- Capel, P.D., and Larson, S.J., 2001, Effect of scale on the behavior of atrazine in surface waters: *Environmental Science and Technology*, v. 35, no. 4, p. 648–657.
- Crawford, C.G., 2004, Sampling strategies for estimating acute and chronic exposures of pesticides in streams: *Journal of the American Water Resources Association*, v. 40, no. 2, p. 485–502.
- Gilliom and others, 2006a, The Quality of Our Nation's Waters—Pesticides in the Nation's Streams and Ground Water, 1992–2001: U.S. Geological Survey Circular 1291, Supplemental technical information for figure 6–5, accessed March 6, 2008, http://water.usgs.gov/nawqa/pnsp/pubs/circ1291/show_description.php?chapter=6&figure=5
- Gilliom and others, 2006b, The Quality of Our Nation's Waters—Pesticides in the Nation's Streams and Ground Water, 1992–2001: U.S. Geological Survey Circular 1291, 172 p.
- Judge, C.G., Griffiths, W.E., Carter-Hill, R., Lutkepohl, H., and Lee, T.C., 1985, *The theory and practice of econometrics* (2nded.): New York, John Wiley and Sons, 1019 p.
- Laitila, T., 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. 1349–1368.
- Tobin, J., 1958, Estimation of relationships for limited dependent variables: *Econometrics*, v. 26, p. 24–36.
- Tukey, J.W., 1977, *Exploratory data analysis*: Reading, Mass., Addison-Wesley, 506 p.
- U.S. Department of Agriculture, 2000, Farm resource regions: U.S. Department of Agriculture, Economic Research Service, Agricultural Information Bulletin Number 760 [also available at <http://www.ers.usda.gov/publications/aib760/aib-760.pdf>]
- U.S. Environmental Protection Agency, 2003, Interim reregistration eligibility decision for atrazine, case no. 0062: U.S. Environmental Protection Agency, Office of Prevention, Pesticides and Toxic Substances, Office of Pesticide Programs, Special Review and Reregistration Division, accessed March 26, 2008, at http://www.epa.gov/oppsrrd1/REDs/atrazine_combined_docs.pdf

Appendix 1

The following is an excerpt from Gilliom and others (2006a).

Moving-average concentrations for each of the 62 pesticide compounds at each of the 186 stream-water sites were estimated as follows:

1. A 1-year period with the most samples and analytes was selected for each site.
2. The time of sampling was rounded to the nearest hour. A few samples rounded to the same hour. The times for these samples were manually edited to round to sequential hours.
3. A program was written to estimate concentrations for every hour in the selected 1-year period. The hourly estimates were calculated by interpolation of the time series of concentrations measured in water samples. Nondetections of pesticides were assigned a zero concentration for interpolation of concentrations.
4. The program to estimate hourly concentrations required a concentration on the first and last days of the 1-year period. The measured concentration of the last sample in the time series was assigned to the first day, first hour of the 1-year period. Similarly, the measured concentration of the first sample in the time series was assigned to the last day, last hour of the 1-year period.
5. The 1-day average concentrations were calculated for each day in the 1-year period by calculating the mean of the 24 hourly concentrations for the day.
6. The 4-day moving-average concentrations were calculated for each day in the 1-year period by calculating the mean of four, 1-day average concentrations (the selected day and the three previous days). 21-day and 60-day moving-average concentrations were calculated similarly to the 4-day moving average.
7. Moving-average concentrations at the beginning of the 1-year period are calculated by extending the 1-day average concentration of the first day forward in time. For example, assume the first four 1-day average concentrations are: 1 on day 1, 2 on day 2, 3 on day 3, and 4 on day 4. The 4-day moving-average concentration for day 1 is 1.00 ($[1+1+1+1]/4$), for day 2 is 1.25 ($[1+1+1+2]/4$), for day 3 is 1.75 ($[1+1+2+3]/4$), and for day 4 is 2.50 ($[1+2+3+4]/4$).

Gilliom and others, 2006a, *The Quality of Our Nation's Waters—Pesticides in the Nation's Streams and Ground Water, 1992-2001*: U.S. Geological Survey Circular 1291, Supplemental technical information for figure 6-5, accessed March 6, 2008, http://water.usgs.gov/nawqa/pnsp/pubs/circ1291/show_description.php?chapter=6&figure=5

