USGS
science for a changing world

National Water Quality Program

Model Methodology for Estimating Pesticide Concentration Extremes Based on Sparse Monitoring Data

Scientific Investigations Report 2017–5159

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

Cover. Photograph showing grassed waterways that carry runoff from crop fields preventing erosion (photograph by Lynn Betts, U.S. Department of Agriculture, 1999).

Model Methodology for Estimating Pesticide Concentration Extremes Based on Sparse Monitoring Data

By Aldo V. Vecchia

National Water Quality Program

Scientific Investigations Report 2017–5159

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

U.S. Department of the Interior
RYAN K. ZINKE, Secretary

U.S. Geological Survey
William H. Werkheiser, Acting Director
exercising the authority of the Director

U.S. Geological Survey, Reston, Virginia: 2018

Suggested citation:
Vecchia, A.V., 2018, Model methodology for estimating pesticide concentration extremes based on sparse monitoring data: U.S. Geological Survey Scientific Investigations Report 2017–5159, 47 p., https://doi.org/10.3133/sir20175159.

ISSN 2328-0328 (online)

Foreword

Sustaining the quality of the Nation's water resources and the health of our diverse ecosystems depends on the availability of sound water-resources data and information to develop effective, science-based policies. Effective management of water resources also brings more certainty and efficiency to important economic sectors. Taken together, these actions lead to immediate and long-term economic, social, and environmental benefits that make a difference to the lives of the almost 400 million people projected to live in the United States by 2050.

In 1991, Congress established the National Water-Quality Assessment (NAWQA) to address where, when, why, and how the Nation's water quality has changed, or is likely to change in the future, in response to human activities and natural factors. Since then, NAWQA has been a leading source of scientific data and knowledge used by national, regional, state, and local agencies to develop science-based policies and management strategies to improve and protect water resources used for drinking water, recreation, irrigation, energy development, and ecosystem needs (https://water.usgs.gov/nawqa/applications/). Plans for the third decade of NAWQA (2013–23) address priority water-quality issues and science needs identified by NAWQA stakeholders, such as the Advisory Committee on Water Information and the National Research Council, and are designed to meet increasing challenges related to population growth, increasing needs for clean water, and changing land-use and weather patterns.

Understanding the occurrence and distribution of pesticides in the Nation's waters is one priority water-quality issue. Pesticides in drinking water and aquatic ecosystems have the potential to adversely impact humans and aquatic life. The evaluation of pesticide exposure has traditionally required high-frequency (for example, daily) sampling. However, because of the prohibitive cost of daily sampling, most monitoring sites have sparse (weekly or less frequent) sampling. This report provides a modeling methodology for using sparse pesticide monitoring data to generate synthetic time series of daily concentrations that reproduce the statistical properties of actual daily time series data. The synthetic data can be used in future studies to better understand pesticide exposure risk and uncertainty in our Nation's rivers and streams.

The purpose of this publication is to provide insight and information to meet water-resource needs and to foster increased citizen awareness and involvement in the protection and restoration of our Nation's waters. The information in this report is intended primarily for those interested or involved in resource management and protection, conservation, regulation, and policymaking at the regional and national levels.

Dr. Donald W. Cline
Associate Director for Water
U.S. Geological Survey

Acknowledgments

James Hetrick, Rochelle Bohaty, Charles Peck, Sarah Hafner, Christine Hartless, and Dana Spatz with the U.S. Environmental Protection Agency and Matthew Bischof with the Washington State Department of Agriculture provided valuable assistance throughout the process of devel-oping and testing the methods described in this report.

Contents

Figures

Tables

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
	Area	
square mile (mi^2)	2.590	square kilometer (km^2)
	Flow rate	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)

Supplemental Information

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

Abbreviations

AMDC	annual maximum daily concentration
ASF	annual sampling frequency
CR	censoring rate
CTS	correlation time scale
maxLT–MDL	maximum long-term method detection level
MTFA	mid-term flow anomaly
NAWQA	National Water-Quality Assessment
NOBS	total number of pesticide observations
NOBSG	number of pesticide observations in a generic year
NUCG	number of uncensored pesticide observations in a generic year
NWQN	National Water Quality Network
PUC	percent of pesticide observations that are uncensored
qlow50	median concentration of low-level detections
RL	record length
SEAWAVE–Q	seasonal wave with streamflow adjustment
SEAWAVE–QEX	seasonal wave with streamflow adjustment with extended capability
SSD	seasonal standard deviation
STFA	short-term flow anomaly
USGS	U.S. Geological Survey

Model Methodology for Estimating Pesticide Concentration Extremes Based on Sparse Monitoring Data

By Aldo V. Vecchia

Abstract

This report describes a new methodology for using sparse (weekly or less frequent observations) and potentially highly censored pesticide monitoring data to simulate daily pesticide concentrations and associated quantities used for acute and chronic exposure assessments, such as the annual maximum daily concentration. The new methodology is based on a statistical model that expresses log-transformed daily pesticide concentration in terms of a seasonal wave, flow-related variability, long-term trend, and serially correlated errors. Methods are described for estimating the model parameters, generating conditional simulations of daily pesticide concentration given sparse (weekly or less frequent) and potentially highly censored observations, and estimating concentration extremes based on the conditional simulations. The model can be applied to datasets with as few as 3 years of record, as few as 30 total observations, and as few as 10 uncensored observations. The model was applied to atrazine, carbaryl, chlorpyrifos, and fipronil data for U.S. Geological Survey pesticide sampling sites with sufficient data for applying the model. A total of 112 sites were analyzed for atrazine, 38 for carbaryl, 34 for chlorpyrifos, and 33 for fipronil. The results are summarized in this report; and, R functions, described in this report and provided in an accompanying model archive, can be used to fit the model parameters and generate conditional simulations of daily concentrations for use in investigations involving pesticide exposure risk and uncertainty.

Introduction

Potential human and ecological exposure to pesticides in streams and rivers commonly is evaluated by comparing measured or predicted pesticide concentrations or concentration statistics to acute and chronic water-quality benchmarks for human health and aquatic life. An accurate assessment of water-quality conditions, therefore, is dependent on characterization of the highest pesticide concentrations that may have occurred. Predicting when the highest concentrations of a pesticide may occur in a stream or river is difficult because of temporal and spatial complexity of pesticide use, pesticide transport, and hydrology. Likewise, systematic sampling, such as monthly or weekly sampling, may not accurately characterize high concentrations, especially in small, flashy streams. Exposure estimates calculated from water samples collected as frequently as every 4 days were determined to be biased low in comparison to estimates calculated from more frequent sampling (Lerch and others, 2011). Simulation analysis was used by Crawford (2004) to indicate that, for small streams, a sampling frequency of 10 times per month produced estimates of the time-weighted 99th percentile concentrations that were within 50 percent of the true concentrations most of the time. The problems with characterizing acute pesticide exposure in streams and rivers based on existing pesticide monitoring data have been described and discussed during Federal Insecticide, Fungicide, and Rodenticide Act Science Advisory Panel meetings regarding atrazine (U.S. Environmental Protection Agency, 2010a, 2010b, 2011, 2012).

To accurately characterize extreme pesticide concentrations in streams, daily sampling may be necessary during active pesticide runoff periods. The cost of sample collection and analysis prohibits this high sampling frequency for most monitoring programs. Sampling frequencies of every 4 days, 10 times per month, or more frequently also are uncommon among programs monitoring pesticide concentrations in streams. Sampling frequencies of weekly to monthly are more common and are referred to in this report as sparse monitoring data.

Approaches to estimate extreme pesticide concentrations from temporally sparse data include the use of sampling bias factors developed from high-frequency monitoring sites (U.S. Environmental Protection Agency, 2012). In this approach, concentration extremes computed from various subsamples of the high-frequency monitoring data, such as monthly or weekly sampling, are compared to the actual concentration extremes to compute bias factors. The resulting bias factors are related to available data on pesticide use, soil characteristic, and other properties of the upstream basins in order to estimate bias factors for sites with sparse sampling. However, the estimated bias factors may be highly uncertain, especially for sites that are not well represented by the small subset of sites with high-frequency sampling.

A new model methodology was developed by the U.S. Geological Survey (USGS) to address the need for using

sparse and potentially highly censored pesticide monitoring data to estimate pesticide concentration extremes, such as the annual maximum daily concentration. The new methodology is based on a statistical model that expresses log-transformed daily pesticide concentration in terms of a seasonal wave, flow-related variability, long-term trend, and serially correlated errors. The seasonal wave models seasonality in pesticide concentration because of site-specific timing and duration of the pesticide application season. Flow-related variability is modeled using two variables that are called mid-term and short-term flow anomalies, and the variables are computed using antecedent daily discharge. The model errors are assumed to have seasonal standard deviation that can increase with increasing pesticide concentration and serial correlation that is modeled using an exponential correlation function. The statistical model can be used to simulate daily concentrations that are equal to (for days with uncensored observations) or less than (for days with censored observations) the monitoring data and that consist of model-generated values for days with no observations. The simulated daily concentrations reproduce the statistical time series characteristics of actual daily pesticide concentrations, such as serial correlation, seasonal means and variances, and flow-related variability. The simulated daily concentrations are called conditional simulations, because the simulated concentrations are conditioned on the observed monitoring data. The conditional simulations can be used to estimate exposure metrics, such as the annual maximum daily or 7-day moving average concentration, and to evaluate bias factors.

Pesticide concentration data from the USGS National Water Quality Network (NWQN) were used to develop this methodology (Vecchia and Williams-Sether, 2017). The USGS National Water-Quality Assessment (NAWQA) project maintains the NWQN with the goal of understanding the nation's water quality, including the occurrence of pesticides in natural waters (Rowe and others, 2013). A primary consideration in developing the methodology was to maximize the number of pesticide sampling sites that could be analyzed with the model, so that whenever possible, monitoring data could be used to estimate daily concentrations and associated exposure metrics. Given the short historical records, sparse sampling frequencies, and often highly censored concentration data available for many of the sites and pesticides, the methodology needed to be as simple and robust as possible while still providing unbiased and informative estimates of concentration extremes.

The starting point for the new methodology is a regression model for analyzing pesticide concentration trends developed by Vecchia and others (2008) and referred to in later applications as the seasonal wave with streamflow adjustment (SEAWAVE–Q) model. The SEAWAVE–Q model has been used in a number of studies to analyze long-term trends in annual median pesticide concentrations for agricultural, urban, and mixed land-use streams (Sullivan and others, 2009; Vecchia and others, 2009; Ryberg and others, 2010; Ryberg and Gilliom, 2015). For simulating daily pesticide concentrations, several enhancements to the SEAWAVE–Q model were required. These enhancements include algorithms for estimating serial correlation and nonconstant variance of the model errors and generating conditional simulations of daily concentrations. The enhanced model is referred to as seasonal wave with streamflow adjustment and extended capability (SEAWAVE–QEX), where the "EX" stands for extended capability to produce simulated daily concentrations.

Purpose and Scope

This report describes the SEAWAVE–QEX model methodology and associated R (R core team, 2016) functions for estimating the model parameters and generating conditional simulations of daily pesticide concentration. Model test results for selected atrazine, carbaryl, chlorpyifos, and fipronil datasets demonstrate the robust properties of the SEAWAVE–QEX model for sparse and potentially highly censored datasets. The effects of serial correlation, degree of censoring, and sampling frequency on the simulated daily concentrations and associated concentration extremes are evaluated. Model application results are summarized for atrazine, carbaryl, chlorpyrifos, and fipronil datasets for USGS pesticide sampling sites with sufficient data for applying the model. The conditional simulations of daily pesticide concentrations produced by the model can be used in investigations involving pesticide exposure risk and uncertainty.

The SEAWAVE–QEX model is intended for use with sparse (weekly or less frequent) monitoring data, and the model is not a replacement for other methods when high-frequency monitoring data are available. Simulated daily concentrations from the model are intended for use in regional to national scale assessments of pesticide exposure risk and uncertainty. Model-simulated concentrations for individual sites may be highly uncertain, and should not be used for site-specific evaluation of concentration extremes without careful consideration of model assumptions and uncertainty.

Model Methodology

The SEAWAVE–Q model is a regression model that was used in previous investigations for analyzing seasonality, flow-related variability, and trends in pesticide concentrations. The SEAWAVE–QEX model is an extension to the SEAWAVE–Q model to include seasonal variance and serial correlation in the model errors. The following sections describe the underlying principles and equations used for the SEAWAVE–Q and SEAWAVE–QEX models.

SEAWAVE–Q

The SEAWAVE–Q model was introduced by Vecchia and others (2008) and has been used as the basis for a number of pesticide trend analysis studies (Sullivan and others, 2009; Vecchia and others, 2009; Ryberg and others, 2010; Johnson and others, 2011; Kalkhoff and others, 2012; Ryberg and others, 2014). The model is expressed as

$$\log\{C(t)\} = \beta_0 + \beta_1 W(t) + \beta_2 A_{MT}(t) + \beta_3 A_{ST}(t) + \beta_4 (t - t_m) + \varepsilon(t) \quad (1)$$

where

log	is the base-10 logarithm;
$C(t)$	is concentration, in micrograms per liter;
$\beta_0, \beta_1, \ldots, \beta_4$	are regression coefficients;
$W(t)$	is the seasonal wave described later in this section (dimensionless);
$A_{MT}(t)$	is the mid-term flow anomaly described later in this section (dimensionless);
$A_{ST}(t)$	is the short-term flow anomaly described later in this section (dimensionless);
t	is the time in decimal years;
t_m	is the midpoint of the time interval being analyzed; and
$\varepsilon(t)$	is the model error (dimensionless).

The seasonal wave (W) models seasonality in log-transformed pesticide concentration resulting from site-specific timing and duration of the pesticide application season. It is computed using a conceptual storage model in which the pesticide is being applied in the upstream basin during one or more distinct application seasons (Vecchia and others, 2008). For the sake of model simplicity, the number of distinct application seasons is assumed to be, at most, two seasons. This conceptual storage model is called a pulse input model. For this report, two classes of seasonal wave models were used, one with a single application season and the other with two distinct application seasons,

$$\begin{gathered} W(t) \in WaveCL1 = \{SW1(m,h,s); \\ m=1,2, \ldots ,6;\ h=1, 2, 3, 4;\ s=0, .5, 1, \ldots ,11.5\} \\ W(t) \in WaveCL2 = \{SW2(m,h,s); \\ m=1,2,\ldots,12;\ h=1, 2;\ s=0, .5, 1, \ldots ,11.5\} \end{gathered} \quad (2)$$

where

*WaveCL*1	is wave class 1,
SW1	is a seasonal wave from class 1,
m	denotes the pulse input model,
h	is the decay rate (decimal month),
s	is the phase shift (decimal month),
*WaveCL*2	is wave class 2, and
SW2	is a seasonal wave from class 2.

In equation 2, there are 576 seasonal waves in each class (24 choices for [m, h] times 24 choices for the phase shift), or a total of 1,152 seasonal waves. For notational convenience, the decimal year is divided into 12 equal-length decimal months instead of calendar months. So, for example, a phase shift of 0.5 decimal month corresponds to 1/24 decimal year, or approximately 15.5 (365/24) days.

Examples of the seasonal waves for *WaveCLS1* are shown in figure 1 for the six choices of pulse input model (m) consisting of single application seasons of durations 1, 2, 3, 4, 6, or 9 months, and for h=1 (left plots) and h=4 (right plots). In each case the phase shift was selected so that the beginning of the application season corresponded with the end of the first decimal month. Each wave ranges from -0.5 to +0.5, increases during the application season, and decreases exponentially (at a rate determined by h) after the end of the application season. Increasing h has the effect of lessening the rate of increase during the beginning of the application season and lessening the rate of decrease after the end of the application season. In previous investigations, h was referred to as the "half-life," but is an empirical coefficient and should not be confused with the chemical half-life of the pesticide. The seasonal waves are analogous to sine and cosine functions, which are often used to model seasonality in hydrologic and water-quality time series data. However, in the case of pesticides, the seasonal waves generally work much better than sine and cosine functions for representing seasonality. Examples of the seasonal waves for *WaveCL*2 are shown in figure 2 for the 12 choices of pulse input model and for h=1. Each pulse input model consists of two application seasons of durations 2, 3, or 4 months separated by nonapplication seasons of durations 2, 3, or 4 months. For pulse input models m=1 through m=6 (left plots), the application rate is the same for both seasons, and for models m=7 through m=12 (right plots), the application rate is higher for the first season. The higher h-values (3 and 4) are not included for *WaveCL*2 because in those cases, the double peak becomes less distinct and the waves are not easily distinguishable from waves with a single application season.

The seasonal wave generally is the most important term in the SEAWAVE–Q model (eq. 1) because the seasonal wave usually explains the most variability in concentrations compared to the other variables in the regression model. However, flow-related variability also is an important consideration. The two flow-anomaly terms in equation 1 are dimensionless variables calculated from a daily streamflow record assumed to be available from a streamflow-gaging station at or near the site being modeled. The mid-term flow anomaly (MTFA) is computed using log-transformed daily streamflow for 30 days up to and including the current time as follows:

$$A_{MT}(t) = \frac{1}{30}\sum_{j=0}^{29}\{\log Q(\tilde{t} - j) - M\} \quad (3)$$

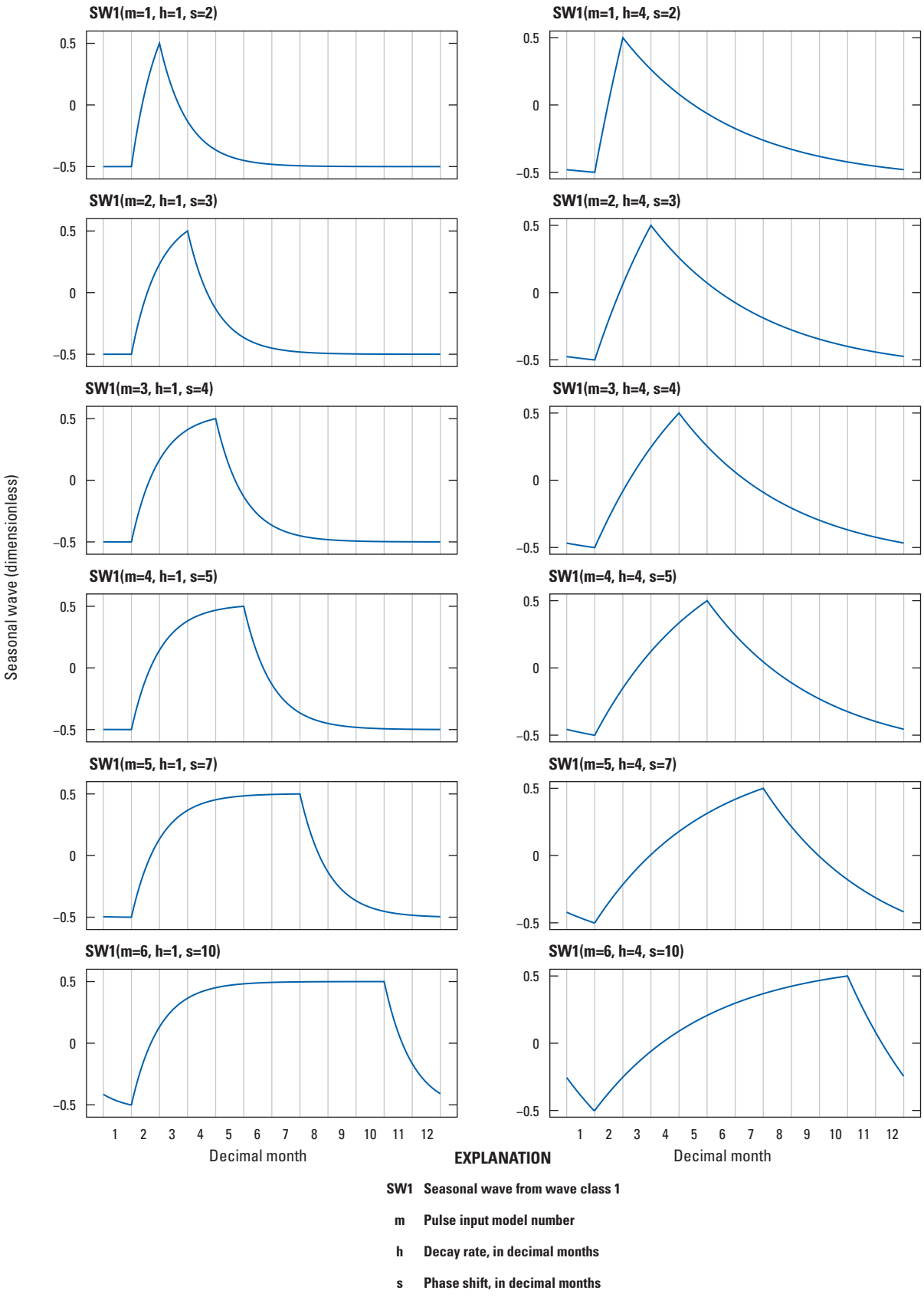

Figure 1. Examples of seasonal waves for wave class 1, consisting of pulse input models with one application season of various durations.

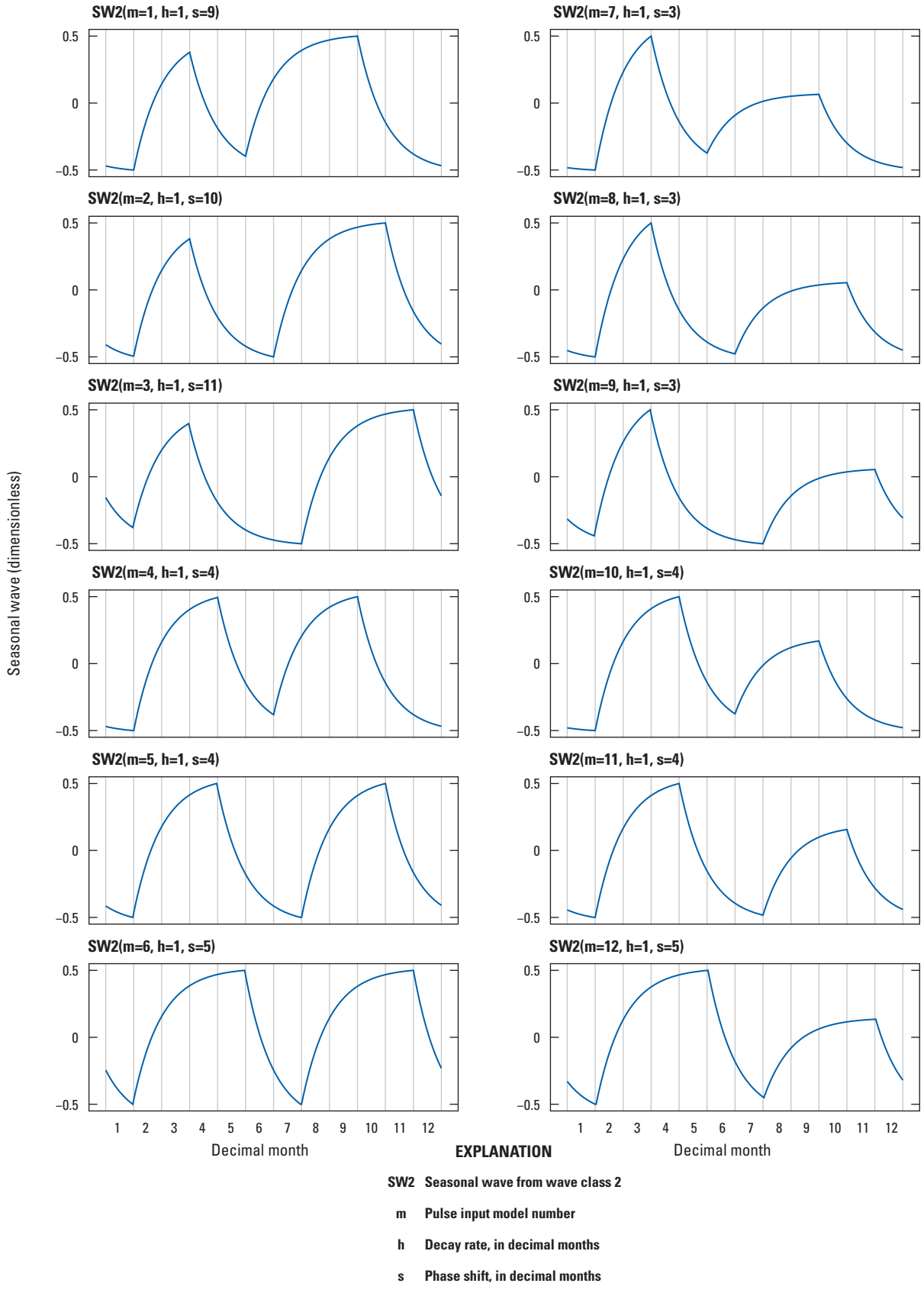

Figure 2. Examples of seasonal waves for wave class 2, consisting of pulse input models with two application seasons of various durations.

where

$A_{MT}(t)$ is the MTFA;
j is the time lag, in days;
log is the base-10 logarithm;
Q is daily mean streamflow, in cubic feet per second;
$\tilde{t}$ is the integer day associated with the current decimal time; and
M is the mean of log-transformed daily flow for the period being analyzed.

The short-term flow anomaly (STFA) is computed by subtracting the mean and the MTFA from log-transformed flow,

$$A_{ST}(t) = \log Q(t) - M - A_{MT}(t) \quad (4)$$

where

$A_{ST}(t)$ is the STFA.

An example of the flow anomalies for Sope Creek near Marietta, Georgia (USGS station number 02335870; U.S. Geological Survey, 2017) is shown in figure 3. MTFA captures seasonal variability relative to the long-term mean (M), and STFA captures higher-frequency variability relative to the long-term mean plus MTFA. As is typically the case for small, flashy streams such as Sope Creek, STFA tends to be highly positively skewed compared to MTFA. Including both anomalies in the model generally explains more variability in pesticide concentrations than including only log-transformed flow. Often the coefficient for MTFA will be negative, indicating that dilution is the primary process at the seasonal scale, and the coefficient for STFA will be positive, indicating that short-term increases in the hydrograph (such as from a rainfall-runoff event) cause concentrations to increase. If daily streamflow is not available, surrogate variables computed using estimated precipitation from the watershed may be considered in place of streamflow (Johnson and others, 2011). The fifth term on the right-hand side of the SEAWAVE–Q model (eq. 1) is a linear trend term for modeling gradual increases or decreases in concentration on a longer (interannual) time scale.

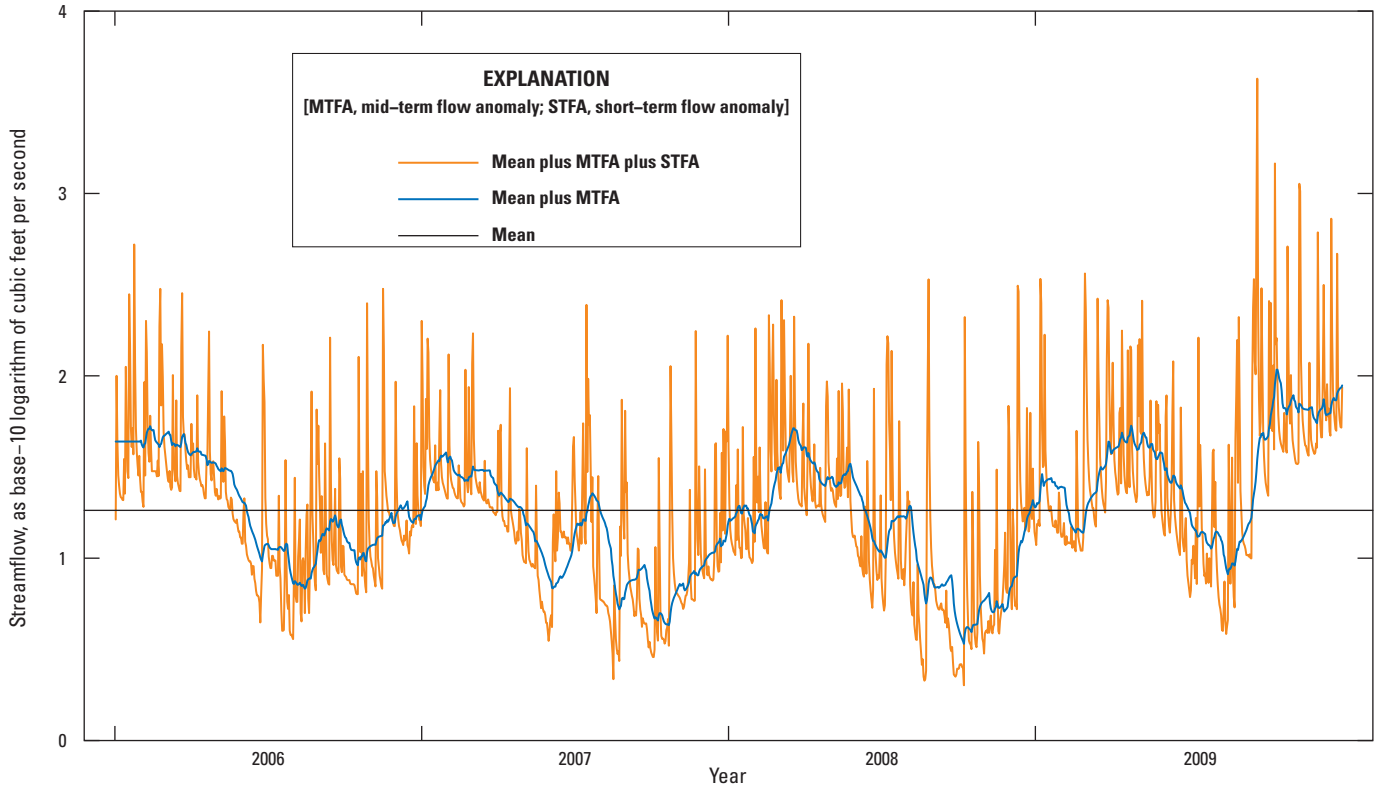

Figure 3. Daily streamflow for Sope Creek near Marietta, Georgia (U.S. Geological Survey station number 02335870) for 2006–09, showing mid-term flow anomaly and short-term flow anomaly.

SEAWAVE–QEX

Analyzing long-term trends was the primary objective in previous investigations using the SEAWAVE–Q model. For analyzing trends with sparse monitoring data, the error term in equation 1 generally can be assumed to consist of independent normal random variables with constant variance. However, for the objective of this study, namely simulating daily concentrations, the properties of the error term needed to be more carefully considered. In particular, nonconstant (seasonal) variance as well as serial correlation between errors for closely spaced observations are important considerations for simulating realistic daily concentrations.

The SEAWAVE–QEX model consists of the SEAWAVE–Q model (eq. 1) modified to include seasonal variance and serial correlation in the model errors,

$$\log\{C(t)\} = \beta_0 + \beta_1 W(t) + \beta_2 A_{MT}(t) + \beta_3 A_{ST}(t) + \beta_4 (t - t_m) + SSD(t)Z(t) \quad (5)$$

$$SSD(t) = \sigma[1 + \alpha W(t)]^{1/2} \quad (6)$$

where

log	is the base-10 logarithm;
$C(t)$	is concentration, in micrograms per liter;
$\beta_0, \beta_1, \ldots, \beta_4$	are regression coefficients;
$W(t)$	is a seasonal wave (eq. 2);
$A_{MT}(t)$	is the mid-term flow anomaly (eq. 3);
$A_{ST}(t)$	is the short-term flow anomaly (eq. 4);
t	is the time, in decimal years;
t_m	is the midpoint of the time interval being analyzed;
$SSD(t)$	is the seasonal standard deviation (dimensionless);
$\sigma>0$	is a seasonal standard deviation parameter;
$0\leq\alpha<2$	is a seasonal standard deviation parameter; and
$Z(t)$	is the normalized model error.

The normalized model errors are assumed to have a normal distribution with mean zero, variance one, and serial correlation function

$$Corr(k) = EV[Z(\tilde{t})Z(\tilde{t}+k)] = e^{-|k|/CTS} \quad (7)$$

where

$Corr(k)$	is the serial correlation function;
k	is the time lag, in days, between observations;
$EV[.]$	denotes expected value;
Z	is the normalized model error (eq. 5);
$\tilde{t}$	is the integer day associated with decimal time t;
e	is Euler's constant; and
$CTS>0$	is the correlation time scale (days).

The correlation function (eq. 7) is an exponential correlation function that is parameterized in terms of the correlation time scale (*CTS*), which is the time lag for which the correlation equals e^{-1}, or approximately 0.37. The correlation between observations separated by 1 day, or lag-1 correlation, equals $e^{-1/CTS}$. For example, a 15-day *CTS* is equivalent to a lag-1 correlation of $e^{-1/15} = 0.936$. The rationale for using an exponential correlation function is discussed later in the "Examples of SEAWAVE–QEX Model Results" section.

The methodology for estimating the SEAWAVE–QEX model parameters, generating conditional simulations of daily pesticide concentration, and using the conditional simulations for estimating pesticide concentration extremes is summarized in figure 4. The first step in analyzing a pesticide dataset for a particular site is to prepare the data for analysis and determine if the data are sufficient for applying the SEAWAVE–QEX model (fig. 4). This step will be explained in detail in the "Data Preparation and Screening" section of this report. The remaining steps of the methodology are described in nontechnical terms in the remainder of this section, and examples are provided in the "Examples of SEAWAVE–QX Model Results" section to illustrate model output and interpretation. Details of the methodology and self-contained R code for completing the computations are described in the appendix.

Regression Analysis and Selection of the Best-Fit Seasonal Wave

The second step of the analysis is to estimate the regression coefficients (β_0, β_1, ..., β_4 in equation 5) and select the best-fit seasonal wave (fig. 4). For this step, the seasonal standard deviation (SSD) is assumed to be constant (α=0 in equation 6) and correlation is ignored (*CTS*=0 in equation 7, equivalent to uncorrelated observations). Censored regression (survreg function in the survival library in R) is used to estimate the regression coefficients for each of the 1,152 choices for the seasonal wave (eq. 2). The seasonal wave with the largest value of the likelihood function is selected as the best-fit seasonal wave. Fitted values for log-transformed concentration are computed from equation 5 with the estimated regression coefficients in place of the true coefficients and the normalized errors equal to zero as follows:

$$FV(t_i) = b_0 + b_1 W^*(t_i) + b_2 A_{MT}(t_i) + b_3 A_{ST}(t_i) + b_4 (t_i - t_m) \quad (8)$$

where

$FV(t_i)$	is the fitted value of log-transformed concentration for the ith observation,
t_i	is the observation time for the ith observation,
$b_0, b_1, \ldots, b_4$	are the estimated regression coefficients, and
$W^*(t_i)$	is the best-fit seasonal wave.

The fitted values represent variability in log-transformed concentration that is explained by the seasonal wave, the flow anomalies, and the trend. The regression residuals are

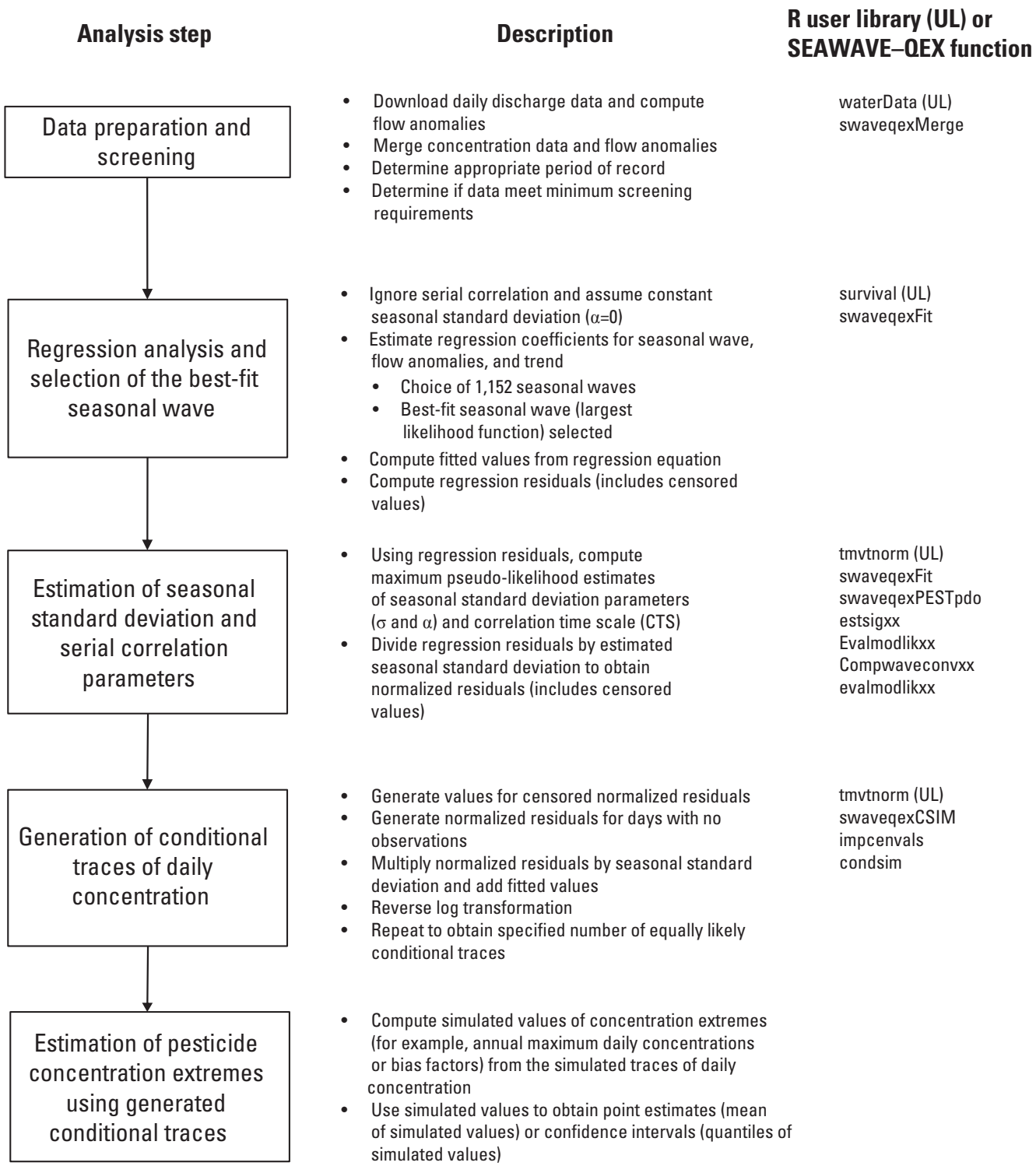

Figure 4. SEAWAVE–QEX modeling methodology.

obtained by subtracting the fitted values from the observations as follows:

$$RR(t_i) = \begin{Bmatrix} \log\{C(t_i)\} - FV(t_i), \\ \text{if the } i\text{th observation is not censored} \\ \log\{L_i\} - FV(t_i), \\ \text{if the } i\text{th observation is censored} \end{Bmatrix} \quad (9)$$

where

$RR(t_i)$ is the regression residual for the ith observation, and

L_i is the censoring limit for the ith observation if the observation is censored ($C(t_i)<L_i$).

Note that regression residuals for censored observations also are censored values. The regression residuals represent variability in log-transformed concentration remaining after removing variability explained by the seasonal wave, the flow anomalies, and the trend. For isolating the seasonal variability, it is useful to compute adjusted concentrations by removing the flow-related variability and the trend as follows:

$$AC(t_i) = \begin{Bmatrix} \log\{C(t_i)\} - b_2 A_{MT}(t_i) - b_3 A_{ST}(t_i) - b_4(t_i - t_m), \\ \text{if } i\text{th observation is not censored;} \\ \log\{L_i\} - b_2 A_{MT}(t_i) - b_3 A_{ST}(t_i) - b_4(t_i - t_m), \\ \text{if } i\text{th observation is censored} \end{Bmatrix} \quad (10)$$

where

$AC(t_i)$ is the adjusted concentration for the ith observation.

Estimation of Seasonal Standard Deviation and Serial Correlation Parameters

The third step in the analysis is to estimate the SSD parameters (σ and α) in equation 6 and the correlation time scale (*CTS*) in equation 7 (fig. 4). These parameters are estimated using the regression residuals (eq. 9) along with an estimation method known as maximum pseudo-likelihood estimation (see appendix). The normalized residuals are computed by dividing the regression residuals by the estimated SSD as follows:

$$NR(t_i) = RR(t_i) / SSD^*;\ SSD^* = \sigma^*[1 + \alpha^* W^*(t_i)]^{1/2} \quad (11)$$

where

$NR(t_i)$ is the normalized residual for the ith observation,

SSD^* is the estimated SSD, and

σ^*, α^* are the estimated values of the SSD parameters σ and α.

The normalized residuals, which are estimates of the normalized model errors (Z in eq. 5), are useful for model diagnostic purposes.

Generation of Conditional Traces of Daily Concentration

The fourth step in the analysis is to generate conditional traces of daily concentration given the parameter estimates, fitted values, and normalized residuals from the previous steps (fig. 4). First, a conditional trace of the normalized residuals is generated as described in the appendix. The generated values of the normalized residuals for days with uncensored observations are equal to the observed values. However, normalized residuals for days with censored observations and normalized residuals for days with no observations differ for each trace and are randomly generated in such a way that the statistical time series characteristics of the data are maintained. After generating a conditional trace of the normalized residuals, the normalized residuals are multiplied by the estimated SSD and added to the fitted values from the regression model to obtain generated values for log-transformed concentration. Note that the fitted values from the regression model can be computed for each day of record provided there are no missing values for the flow anomalies. If some days are missing flow anomalies, the generated concentrations for those days also will be missing values. After generating a conditional trace for log-transformed concentration, the log transformation is reversed to obtain a conditional trace of untransformed concentration. The generated daily concentrations from the conditional trace are equal to the observed concentrations for days with uncensored observations; are less than the censoring limit for days with censored observations; and for days with no observations, the generated concentrations should be indistinguishable (statistically) from values that could have occurred if samples had been collected daily. The trace is referred to as "conditional" trace because it is conditioned on the observed concentrations. The process for generating a conditional trace is repeated to obtain a specified number (N) of randomly generated conditional traces, where guidelines for specifying N are discussed in the next step.

Estimation of Concentration Extremes using Generated Conditional Traces

The fifth step in the analysis is to use the conditional traces of daily concentration from the previous step to estimate pesticide concentration extremes (fig. 4). This step depends on the particular objectives of the analysis. For example, one objective might be to obtain an estimated value or confidence interval for the annual maximum daily concentration (AMDC). For this application, it may be appropriate to generate N=100 conditional traces, where each trace consists of n years of generated daily values and n is the record length (RL). For each year of record, there are N=100 simulated

values for the AMDC. The mean of the 100 simulated values is the estimated value for the AMDC and the 10th and 90th percentiles of the simulated values provide an approximate 80 percent confidence interval for the AMDC. If a higher confidence level is required, generating more than 100 traces may be necessary. For example, to obtain a 95 percent confidence interval, N=250 traces may be required. Generally, to obtain a confidence level of P percent, N(1–P/100) should be at least 20 to obtain a reasonable confidence interval. The same technique described for AMDC can be used to estimate other exposure metrics, such as the annual maximum 7-day or 30-day moving average concentration. Another objective may be to obtain estimates of bias factors (U.S. Environmental Protection Agency, 2012). For estimating bias factors, N=50 conditional traces may be sufficient provided a constant bias factor for a particular site is used, rather than bias factors that vary by year. In the former case there would be 50 times n generated values of the bias factors (for all n simulation years combined), or at least 150 values if the RL (n) is at least 3 years.

Examples of SEAWAVE–QEX Model Results

Several examples are presented in this section to illustrate the SEAWAVE–QEX modeling methodology and to help reinforce the definitions, terminology, and interpretation of model output. Details of the data sources, sample collection and preservation methods, laboratory analytical methods, and other data considerations are described later in the "SEAWAVE–QEX Model Applications" section and are not important for purposes of this section.

Atrazine

The first example is for atrazine concentration for Little Buck Creek near Indianapolis, Indiana (USGS station number 03353637) for 1993–2002 (fig. 5). This site has a drainage area of 44 square kilometers (km^2) and mixed urban and agricultural (mostly corn and soybeans) land use, and atrazine

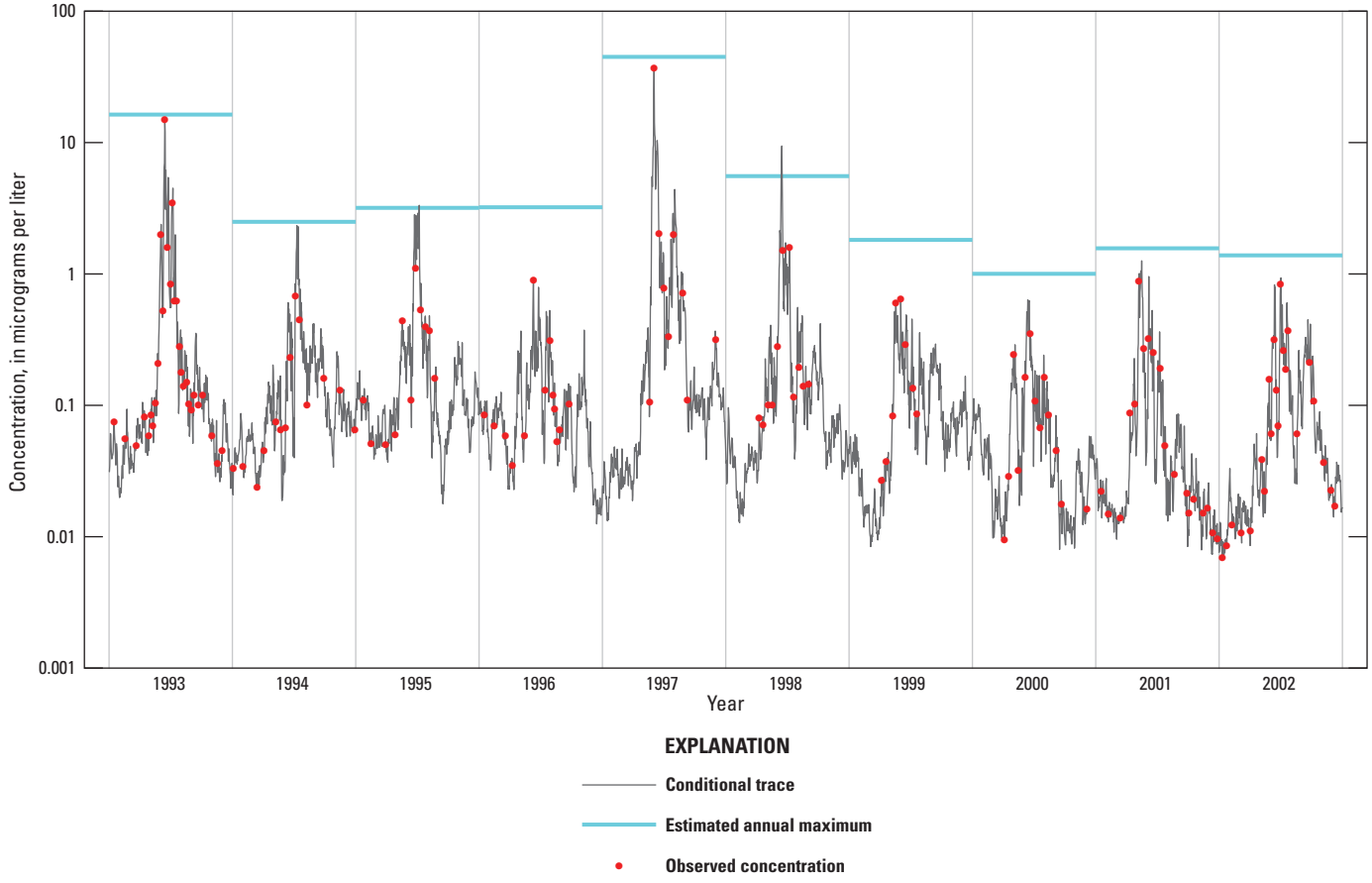

Figure 5. Observed atrazine concentrations, simulated conditional trace of daily concentrations, and estimated annual maximum daily concentrations for Little Buck Creek near Indianapolis, Indiana (U.S. Geological Survey station number 03353637) for 1993–2002.

is an agricultural herbicide used mostly on corn (Ryberg and others, 2010). The red points in figure 5 indicate observed concentrations. There are no censored values for this dataset. The dataset has 148 observations, or an average annual sampling frequency (ASF) of about 15 observations per year. However, ASF was variable, ranging from 28 samples in 1993 to 8 samples in 1999. The grey line in figure 5 shows a single conditional trace (out of 100) generated from the SEAWAVE–QEX model, which is equal to the observed atrazine concentration values during observation days and hypothetical randomly generated values for the remaining days. The horizontal blue lines show the estimated annual maximum concentration for each year, which is the average of the annual maxima for all 100 conditional traces.

The best-fit seasonal wave (*W**), estimated linear regression coefficients (b_0, b_1, ..., b_4), estimated SSD parameters (σ* and α*), and estimated CTS (*CTS**) for this dataset were as follows:

*W**=*SW1*(*m*=3, *h*=2, *s*=6);

b_0=-0.88, b_1=1.35, b_2=-0.30, b_3=0.16, b_4=-0.06; and

σ*=0.379, α*=1.11, *CTS**=15.1.

The regression coefficients for the seasonal wave (1.35), MTFA (-0.30), STFA (0.16), and trend (-0.06) were significant (*p*-values less than 0.05; note that the *p*-values are approximate because they are based on assumption of constant variance and no serial correlation). The downward trend probably resulted from increasing urban land use (and decreasing corn acreage) during the analysis period (Ryberg and others, 2010). The best-fit seasonal wave (b_0+b_1W*) is shown in figure 6. For this example, *W** was from wave class 1 and the pulse input model consisted of a 3-month application season (*m*=3), decay rate (*h*) of 2 months, and phase shift (*s*) of 6 months. The application season extends from early April through early July. The points in figure 6 show adjusted concentrations (eq. 10) to highlight the seasonal variability along with the fitted seasonal wave. Observations are shown for all years combined. Also shown are lines at plus and minus 2 times the estimated SSD (b_0+b_1W*+/-2 σ* $(1+\alpha^*W^*)^{1/2}$). For these data, α*=1.11, thus the seasonal standard deviation was considerably higher during the application season than during other times of the year.

The normalized residuals (eq. 11) for this example are shown in figure 7. No obvious problems with lack-of-fit (nonconstant mean or variance, skewness, outliers, and so on) are indicated, with the possible exception of three residuals equal to 3, which may be outliers. In the estimation algorithms described in the appendix, to reduce the effects of potential outliers on the parameter estimates, the normalized residuals were truncated at plus or minus 3. In this case, three of the residuals were truncated at plus 3. Further examination of the simulation results indicated that only one of the truncated residuals, during June 1997, may have adversely affected the estimated annual maximum concentration for that year. Further examination of the data for that observation may be warranted. Note that serial correlation may produce apparent clustering in the normalized residuals, and such clustering should not adversely affect the results. In this case, clustering is apparent during May through August, during which concentrations (and sampling frequencies) were highest. The fitted exponential correlation function (eq. 7) for this example is shown in figure 8. The estimated CTS for this example was about 15 days (*CTS**=15.1). The points in figure 8 show empirical estimates of the correlation function at selected lags and were computed directly from the normalized residuals using methods for estimating the empirical variogram (Cressie, 1991, p. 74) and converting the empirical variogram to the empirical correlogram. The exponential correlation function provided a reasonable approximation to the empirical correlogram for this example. The exponential correlation function has been found generally to be a reasonable approximation to the empirical correlogram for datasets (such as this one) that have sampling frequencies sufficient for computing the empirical correlogram for short time lags. The exponential correlation function has two features that have considerable influence on the conditional simulations—the assumption that the correlation function approaches one as the time lag approaches zero and the assumption of an exponential decay rate as the time lag increases. Measurement error (differences between actual and measured concentrations because of factors such as sample collection and preservation techniques or laboratory analytical techniques) could cause a discontinuity in the correlation function at lag zero. For USGS pesticide data analyzed for this report, measurement error generally can be considered to be negligible in relation to natural variability.

Carbaryl

The next example is for carbaryl concentration for Kisco River below Mount Kisco, New York (USGS station number 01374987) for 2000–2008 (fig. 9). This site has a drainage area of 49 km^2 and mixed urban and agricultural land use, and carbaryl is an insecticide with many urban and agricultural uses (Ryberg and others, 2010). The dataset includes 186 observations with ASF ranging from 28 in 2002 to 15 in 2008. This example is typical of the high censoring rate (CR) for many pesticides. About 71 percent (134) of the observations were censored. The censored values in figure 9 and in subsequent figures are shown as open circles; and, for plotting purposes, the values are equal to the randomly generated value from the conditional trace. The best-fit seasonal wave model, estimated linear regression coefficients, estimated SSD parameters, and estimated CTS for this site were as follows:

*W**=*SW1*(*m*=4, *h*=3, *s*=6.5);

b_0=-2.42, b_1=0.906, b_2=0.17, b_3=0.75, b_4=0.04; and

σ*=0.542, α*=0, *CTS**=4.7.

The regression coefficients for the seasonal wave (0.906) and STFA (0.75) were significant. The best-fit seasonal wave for this example (fig. 10) was from wave class 1 and consisted of a 4-month application season from late March through late July (*m*=4 and *s*=6.5) and a decay rate of 3 months (*h*=3). The estimated SSD was constant (α* = 0) and the estimated CTS was about 5 days (*CTS**=4.7). Note that a lower bound of zero was assumed for the estimated value of α (α*≥0) and for this example the estimate was equal to the lower bound.

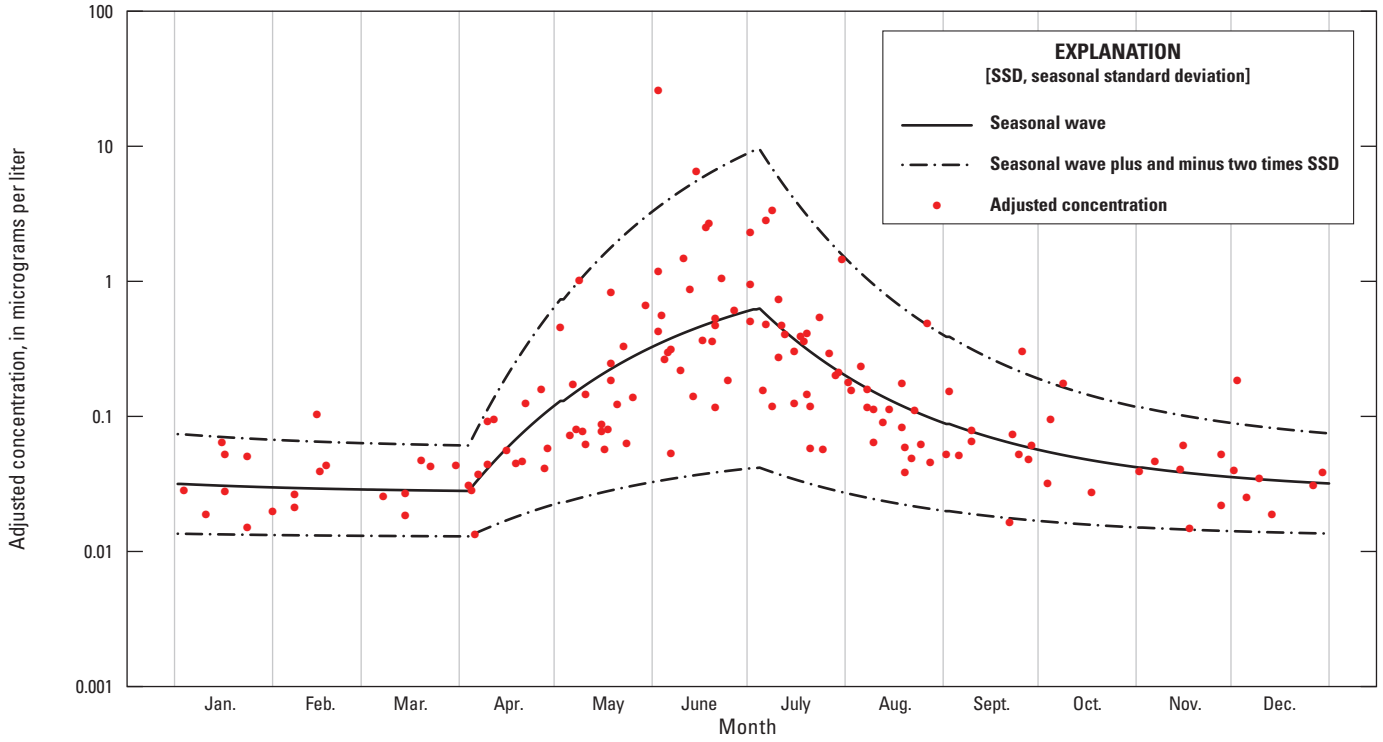

Figure 6. Adjusted (minus trend and flow-related variability) atrazine concentrations, fitted seasonal wave, and fitted seasonal standard deviation for Little Buck Creek near Indianapolis, Indiana (U.S. Geological Survey station number 03353637) for 1993–2002.

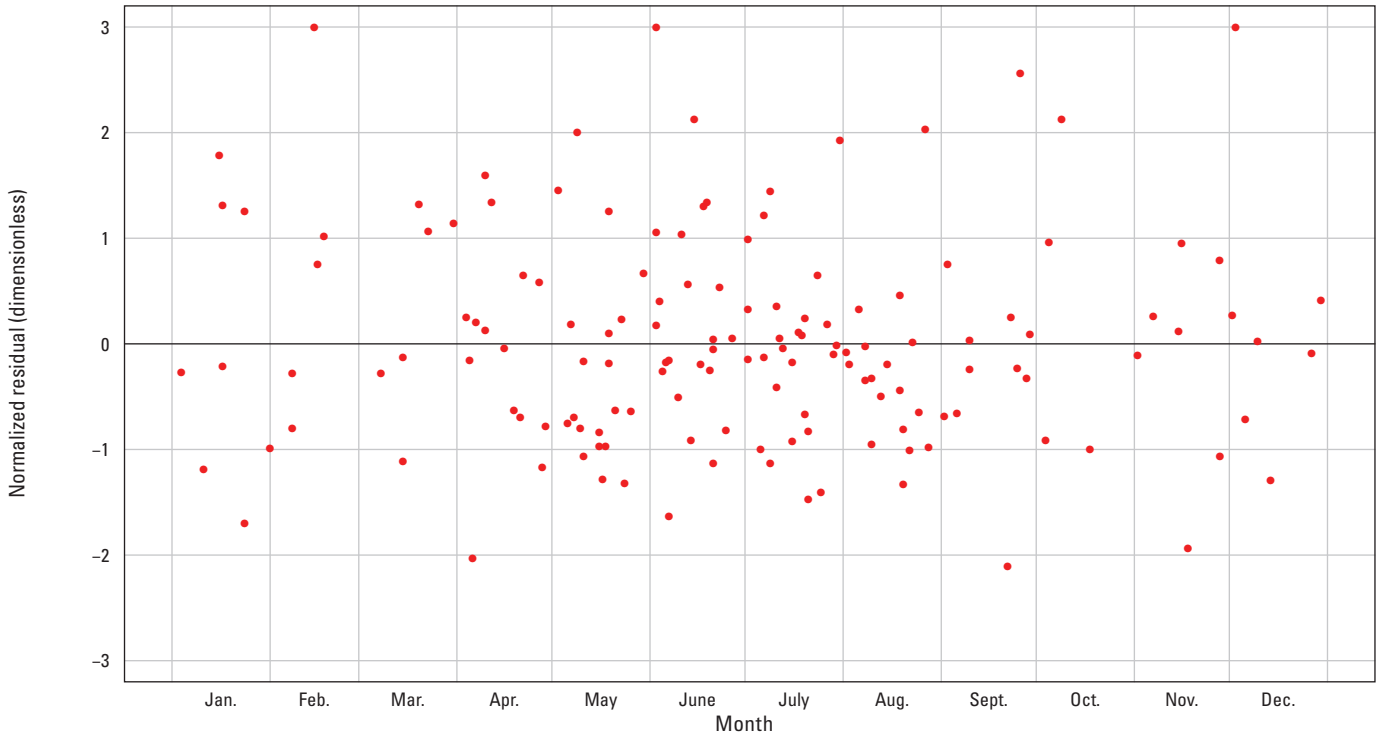

Figure 7. Normalized residuals from SEAWAVE–QEX model for atrazine concentration for Little Buck Creek near Indianapolis, Indiana (U.S. Geological Survey station number 03353637) for 1993–2002.

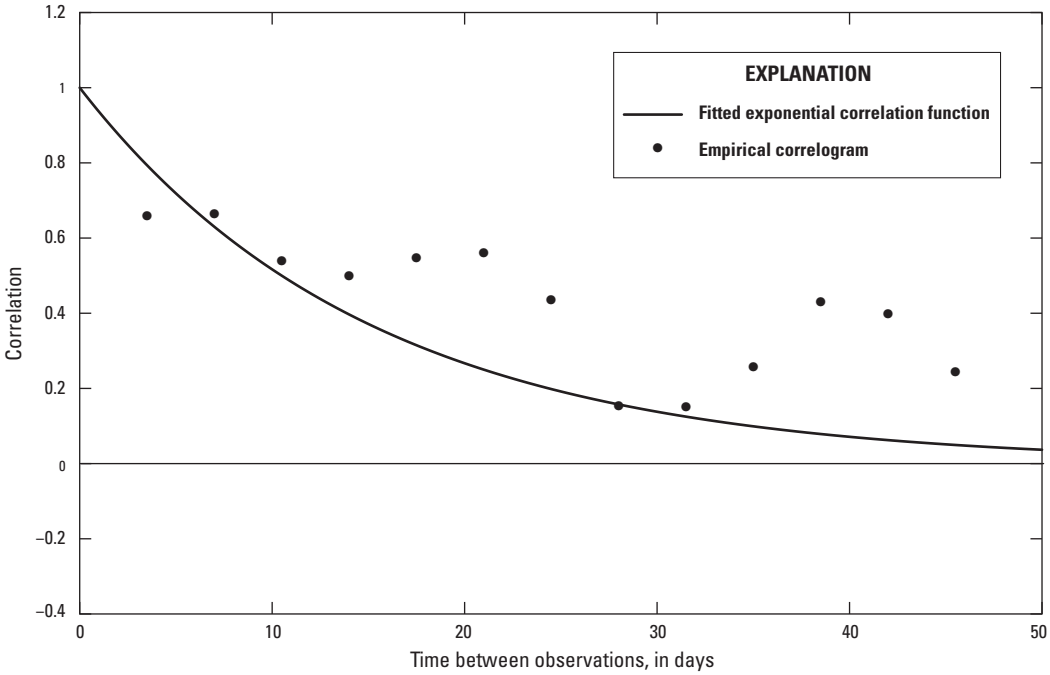

Figure 8. Empirical correlogram and fitted exponential correlation function for normalized residuals from SEAWAVE–QEX model for atrazine concentration for Little Buck Creek near Indianapolis, Indiana (U.S. Geological Survey station number 03353637) for 1993–2002.

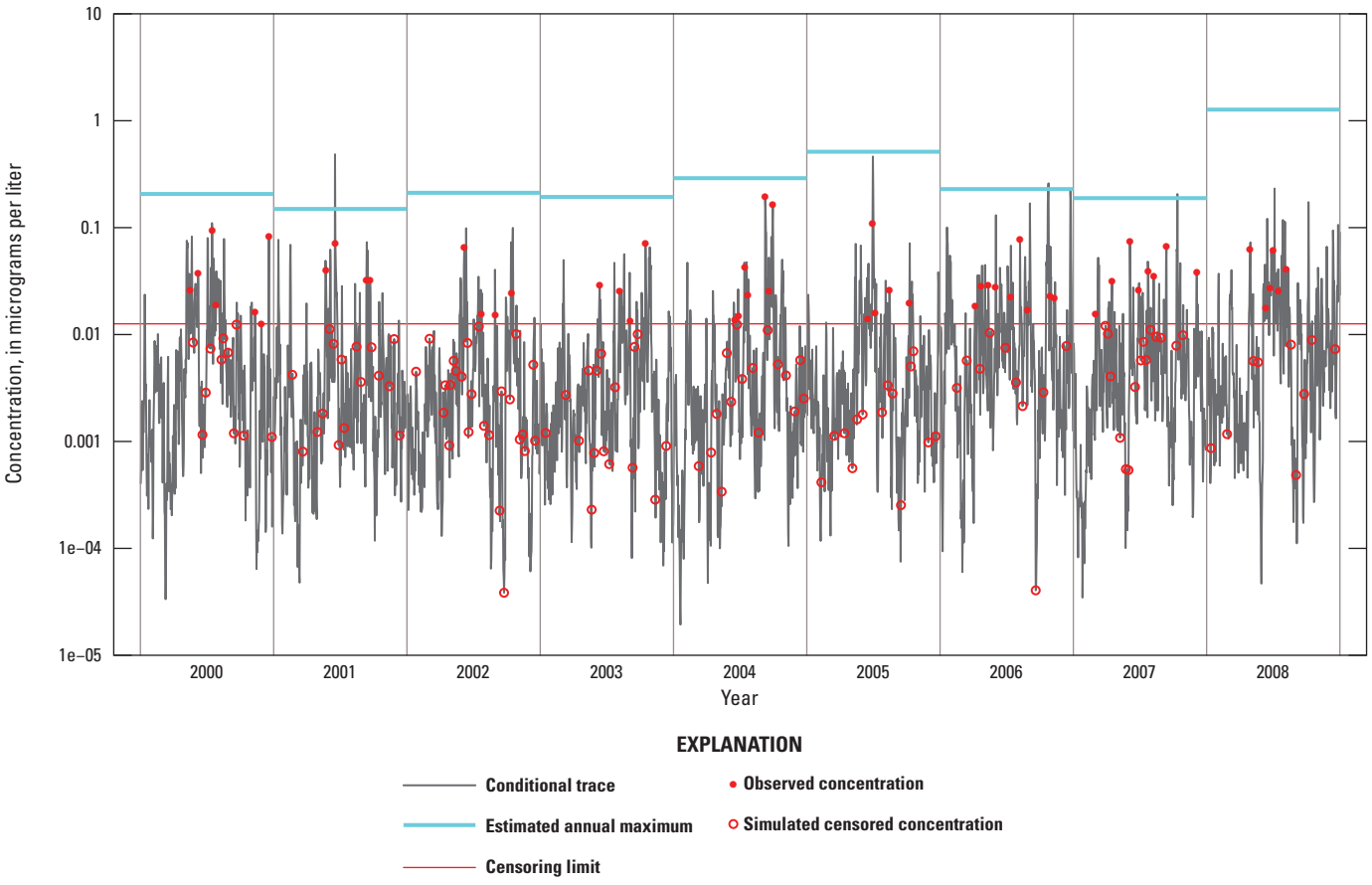

Figure 9. Observed carbaryl concentrations, simulated conditional trace of daily concentrations, and estimated annual maximum daily concentrations for Kisco River near Mount Kisco, New York (U.S. Geological Survey station number 01374987) for 2000–2008.

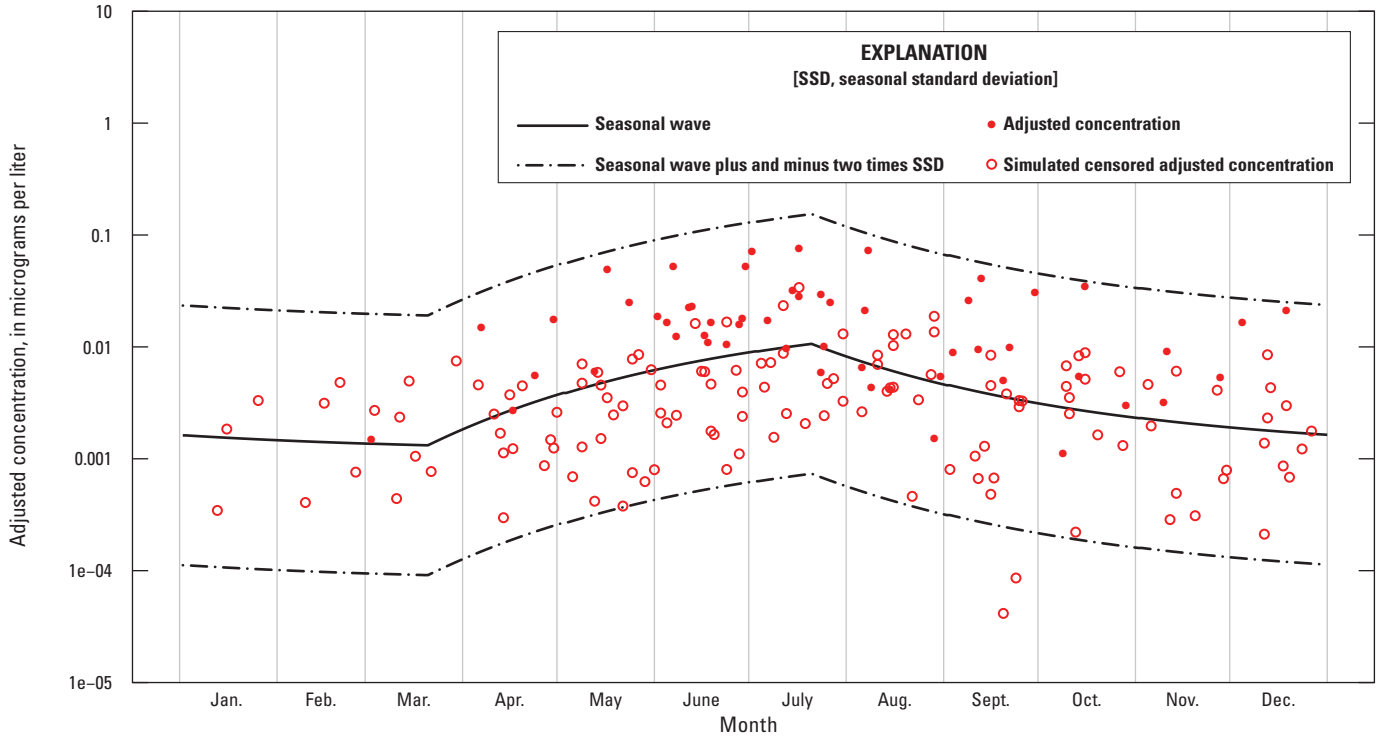

Figure 10. Adjusted (minus trend and flow-related variability) carbaryl concentrations, fitted seasonal wave, and fitted seasonal standard deviation for Kisco River near Mount Kisco, New York (U.S. Geological Survey station number 01374987) for 2000–2008.

Chlorpyrifos

The next example is for chlorpyrifos concentration for Sope Creek near Marietta, Georgia (USGS station number 02335870), for 1993–2002 (fig. 11). This site has a drainage area of about 80 km^2 and urban land use of about 70 percent (Ryberg and others, 2010). During 1993–2002, chlorpyrifos was a widely used insecticide in a variety of urban and agricultural settings, but residential use was phased out starting in about 1997 and essentially eliminated by 2002 (Ryberg and others, 2010). The dataset includes 118 observations with ASF of 32 in 1993, 23 in 2002, and from 0 to 12 in the other years. About 70 percent (83) of the values were censored. The best-fit wave model, estimated linear regression coefficients, estimated SSD parameters, and estimated CTS for this example were as follows:

$W^*=SW2(m=4, h=2, s=5.5)$;

$b_0 = -2.78$, $b_1 = 0.80$, $b_2 = 0.92$, $b_3 = 0.61$, $b_4 = -0.06$; and

$\sigma^* = 0.374$, $\alpha^* = 0$, $CTS^* = 3.5$.

The regression coefficients for the seasonal wave (0.80), MTFA (0.92), STFA (0.61), and trend (-0.06) all were significant. The downtrend probably resulted from the aforementioned phasing out of residential uses. The best-fit seasonal wave for this example (fig. 12) was from wave class 2 and consisted of two 3-month application seasons, from mid-October through mid-January and from mid-March through mid-June (m=4 and s=5.5), and a decay rate of 2 months (h=2). The estimated SSD was constant (α^*=0) and the estimated CTS was 3.5 days. Note that a lower bound for the estimated CTS of 3.5 days ($CTS^* \geq 3.5$) was assumed because sampling frequencies are rarely sufficient to discern correlation structure at such small time lags. For this example, CTS^* was equal to the lower bound.

Fipronil

The final example is for fipronil concentration for Sope Creek near Marietta, Georgia (U.S. Geological Survey station number 02335870) for 2003–12 (fig. 13). Fipronil is an insecticide that was first registered for use in 1996 and quickly gained widespread use by 2002 as a replacement for other insecticides, such as diazinon and chlorpyrifos, that were being phased out for residential use (Ryberg and others, 2010). The dataset includes 127 observations with ASF ranging from 22 in 2008 to 6 in several other years. The dataset includes 68 censored observations (about 54 percent). The best-fit seasonal wave model, estimated linear regression coefficients, estimated SSD parameters, and estimated CTS for this site were as follows:

$W^*=SW2(m=5, h=1, s=7)$;

b_0=-2.15, b_1=0.17, b_2=0.14, b_3=0.29, b_4=0.01; and

σ^*=0.201, α^*=0.218, CTS^*=5.9.

The regression coefficients for the seasonal wave (0.17), MTFA (0.14), and STFA (0.29) all were significant. The best-fit seasonal wave (fig. 14) was from wave class 2 and consisted

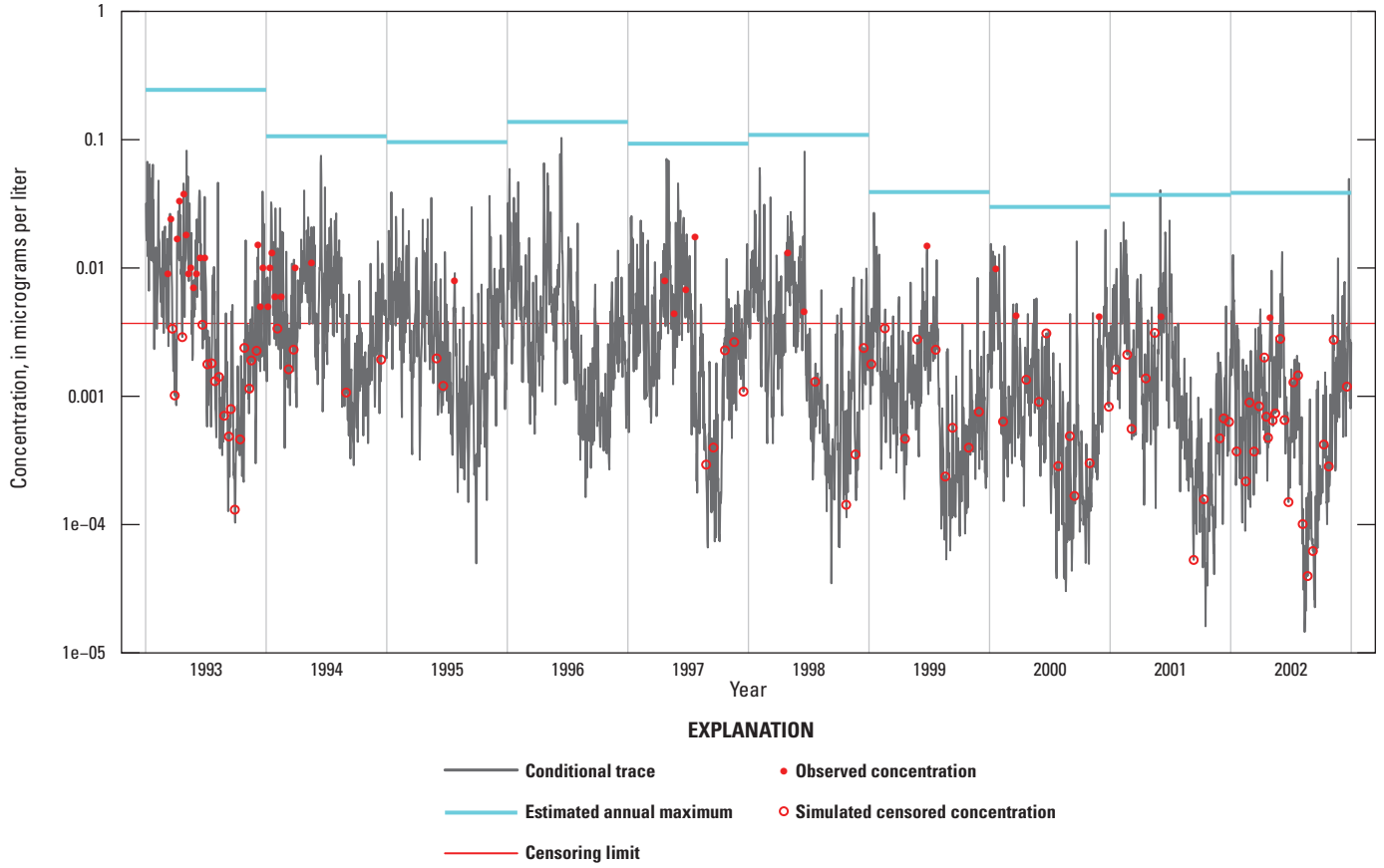

Figure 11. Observed chlorpyrifos concentrations, simulated conditional trace of daily concentrations, and estimated annual maximum daily concentrations for Sope Creek near Marietta, Georgia (U.S. Geological Survey station number 02335870) for 1993–2002.

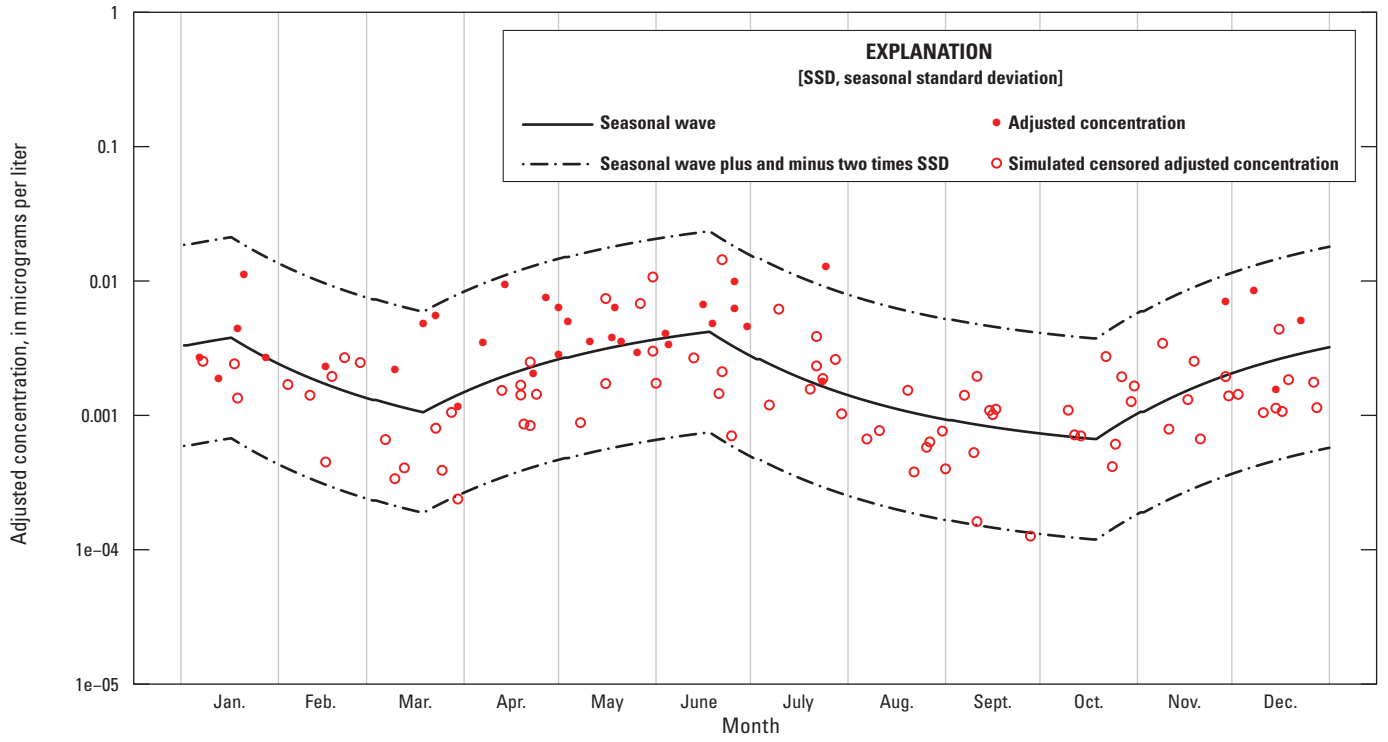

Figure 12. Adjusted (minus trend and flow-related variability) chlorpyrifos concentrations, fitted seasonal wave, and fitted seasonal standard deviation for Sope Creek near Marietta, Georgia (U.S. Geological Survey station number 02335870) for 1993–2002.

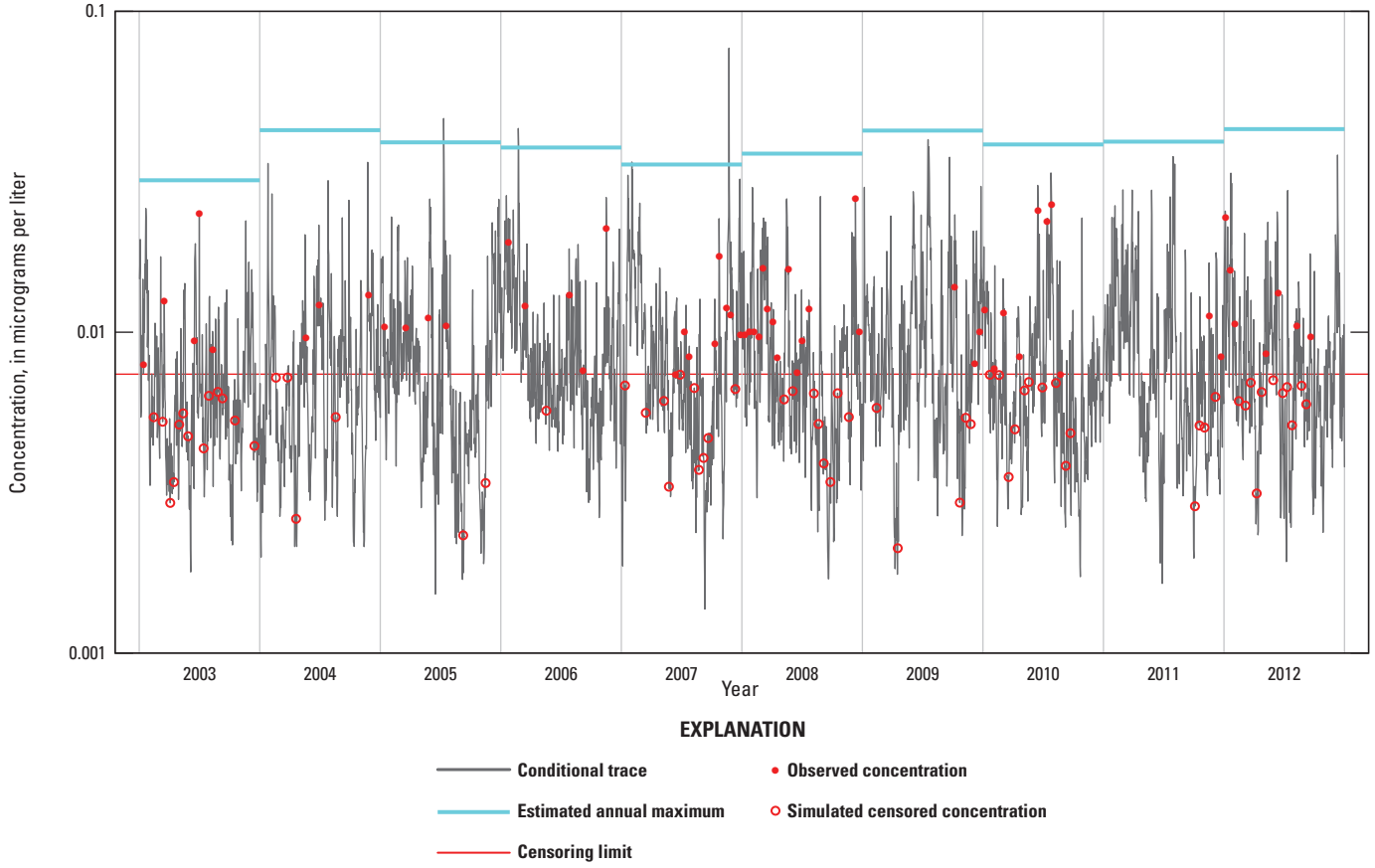

Figure 13. Observed fipronil concentrations, simulated conditional trace of daily concentrations, and estimated annual maximum daily concentrations for Sope Creek near Marietta, Georgia (U.S. Geological Survey station number 02335870) for 2003–12.

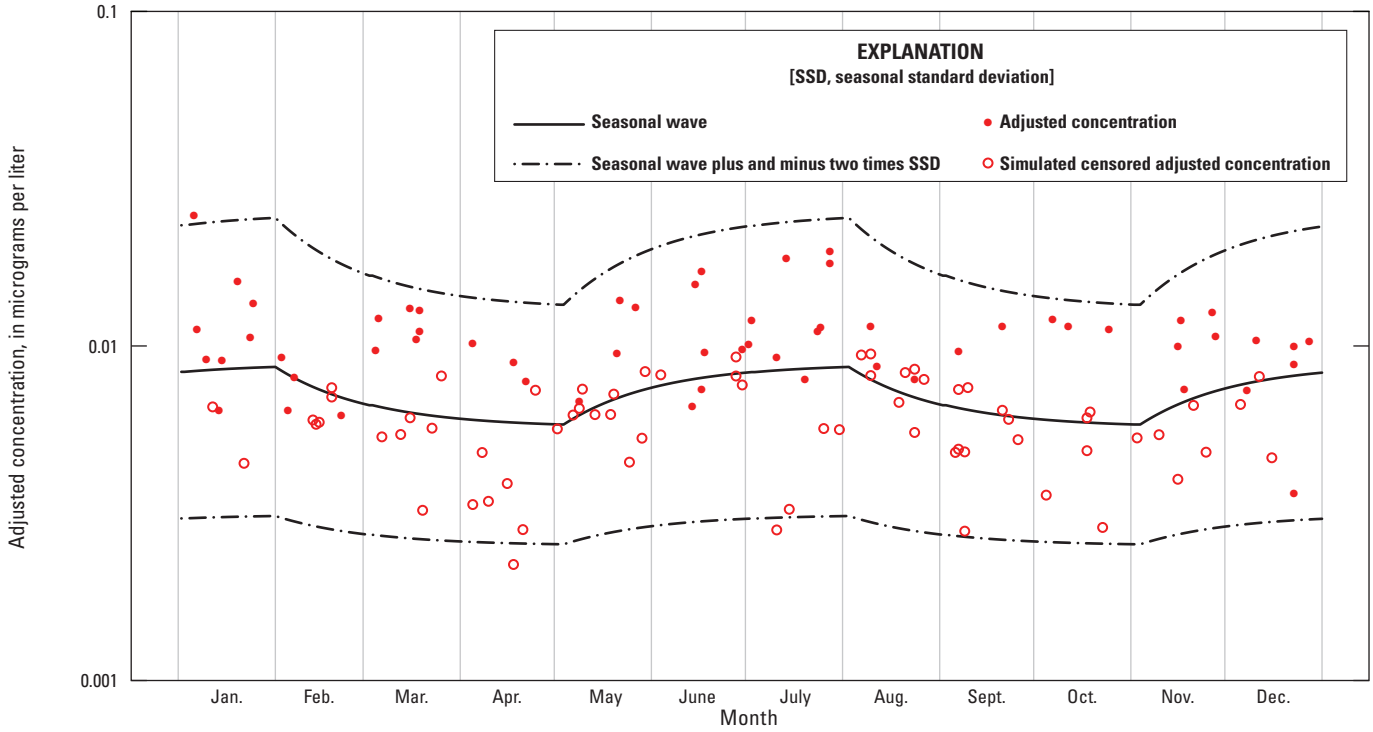

Figure 14. Adjusted (minus trend and flow-related variability) fipronil concentrations, fitted seasonal wave, and fitted seasonal standard deviation for Sope Creek near Marietta, Georgia (U.S. Geological Survey station number 02335870) for 2003–12.

of two 3-month application seasons, from early November through early February and from early May through early August (m=5 and s=7), and a decay rate of 1 month (h=1). The seasonal wave for fipronil was similar to the seasonal wave for chlorpyrifos (fig. 12), but with somewhat different timing of the application seasons and faster decay rate. The SSD had a small degree of seasonality (α^*=0.24), and the estimated CTS was about 6 days (CTS^*=5.9).

Model Testing

Simulation experiments were used to test the SEAWAVE–QEX model and determine limits for the RL, ASF, and CR for applying the model. The objective of the SEAWAVE–QEX model is to provide a robust tool that can be used to generate conditional simulations of daily concentration and estimate associated statistics, such as the AMDC, for as many pesticide monitoring sites as possible and for pesticides (such as carbaryl, chlorpyrifos, and fipronil) that may have short records, sparse sampling frequencies, or a high percentage of censored data. Important considerations for verifying any statistical model include (1) verification of model assumptions and (2) verification of model results. The first consideration will be discussed later in the "Model Limitations" section of this report. The simulation experiments described in this section were designed with the second consideration in mind, namely, given that the model assumptions are satisfied, does the model provide reasonable estimates of concentration extremes? How do properties of the estimates (for example, bias and uncertainty) depend on RL, ASF, and CR?

For the simulation experiments, the atrazine, carbaryl, and chlorpyrifos examples described in the "Examples of SEAWAVE–QEX Model Results" section were used to randomly generate time series of "known" daily pesticide concentrations assuming the true model was given by the fitted model for the respective examples, except for CTS, which was varied. The "known" daily pesticide concentrations (which have no censored values or days with missing observations) are referred to as model-generated traces to distinguish them from the conditional traces described previously. For each model-generated trace, the RL, ASF, CTS, and CR were varied to represent a reasonable range of values that may be expected. In particular, all combinations of two RLs (RL=3 and 6 years), three ASFs (ASF=12, 24, and 48 samples per year), three CTSs (CTS=7, 15, and 30 days), and two CRs (CR=30 and 70 percent) were considered. For each pesticide and each combination of RL, ASF, CTS, and CR, 100 model-generated traces were generated from the SEAWAVE–QEX model. The flow anomalies for each model-generated trace were computed using consecutive years of daily streamflow from the period of record used for the examples, where the beginning year was selected at random for each trace. For example, for the atrazine model (where the period of record used for the example was 1993–2002) and RL=3, the starting year was selected at random from eight possibilities (1993–2000). The lowest ASF (ASF=12) represents approximate monthly sampling. For that case, the data used for fitting the model were assumed to consist of 12 observations per year, where the first observation was selected at random from days 1 to 30 and the remaining 11 observations were selected every 30 days thereafter. For ASF=24, representing approximate bimonthly sampling, the first observation was selected at random from days 1 to 15 and the remaining 23 observations were selected every 15 days thereafter. For ASF=48, representing approximate weekly sampling, 24 additional observations were added to the ASF=24 design each year (about midway between adjacent observations for the ASF=24 design, resulting in alternating 7- or 8-day spacing between observations). After selecting the observations used for fitting the model, the known concentrations for the selected observations were artificially censored at either the lower (CR=30) or higher (CR=70) level. For example, with RL=3, ASF=12, and CR=70, there were 36 observations for all 3 years, 11 of which (about 30 percent) were uncensored and 25 of which (about 70 percent) were censored.

For each model-generated trace, the SEAWAVE–QEX model methodology described previously (fig. 4 and associated discussion) was applied to fit the model, generate N=100 conditional traces, and compute estimates of the AMDC for each simulation year. The following statistics were used to evaluate the estimated AMDC:

- Bias—the geometric mean of the estimated AMDCs minus the geometric mean of the known AMDCs, as a percent of the geometric mean of the known AMDCs.

- 80-percent error bounds—(P10, P90), where P10 and P90 are the 10th and 90th percentiles of the estimation errors. The estimation errors are the differences between the estimated and known values of AMDC, expressed as a percent of the known value. For example, if the 80-percent error bounds are (-10, 50), most (80 percent) of the estimated AMDCs are within -10 and +50 percent of the known value.

Bias was categorized as low (between -10 and 25 percent), moderate (between 25 and 50 percent), or high (greater than 50 percent). No cases had substantial downward (negative) bias (less than -10 percent)—in most cases there was upward (positive) bias. Thus, the estimated values of AMDC generally, on average, were higher than the known values.

Results of the simulation experiments for all three models (atrazine, carbaryl, and chlorpyrifos) indicated that, as expected, both bias and uncertainty were lower for RL=6 years compared to RL=3 years. Therefore, only the results for RL=3 years, given in table 1, are discussed.

For the lower CR (CR=30), bias was low for all three models and for all combinations of ASF and CTS (table 1). However, estimation errors (uncertainty) depended on ASF and CTS. For CR=30 and ASF=12 (30 days between observations), estimation uncertainty was high relative to ASF=24

Table 1. SEAWAVE–QEX model testing results based on simulated data from the atrazine, carbaryl, and chlorpyrifos models with record length 3 years.

Censoring rate, in percent	Annual sampling frequency, in samples per year	Correlation time scale, in days	Atrazine model			Carbaryl model			Chlorpyrifos model		
			[1]Bias	[2]P10	[2]P90	[1]Bias	[2]P10	[2]P90	[1]Bias	[2]P10	[2]P90
30	12	7	-7	-61	112	19	-66	264	16	-49	141
30	12	15	10	-48	129	13	-49	157	24	-36	142
30	12	30	19	-32	107	22	-52	192	16	-31	90
30	24	7	-5	-50	86	23	-36	137	19	-30	84
30	24	15	5	-33	73	12	-35	92	6	-30	54
30	24	30	9	-27	59	16	-31	86	10	-21	59
30	48	7	3	-33	53	15	-31	83	10	-26	57
30	48	15	2	-26	39	2	-35	53	2	-21	31
30	48	30	6	-16	36	2	-23	36	8	-15	40
70	12	7	15	-56	227	21	-66	272	24	-47	140
70	12	15	33	-54	237	21	-60	224	42	-41	210
70	12	30	48	-33	266	46	-43	269	40	-42	190
70	24	7	32	-35	166	48	-33	268	19	-36	136
70	24	15	31	-21	129	20	-41	151	7	-31	70
70	24	30	18	-20	78	6	-41	97	3	-37	52
70	48	7	31	-13	104	45	-12	153	13	-20	70
70	48	15	18	-13	71	20	-24	88	9	-22	54
70	48	30	16	-9	45	12	-23	67	2	-23	36

[1]Difference between geometric means of estimated and known annual maximum daily concentrations, as a percent of the known value.

[2]Tenth (P10) and 90th (P90) percentiles of the estimation errors, where the estimation error is the difference between the estimated and known annual maximum daily concentration, as a percent of the known value.

(15 days between observations) or 48 (approximately 7 days between observations). For ASF=12, spacing between samples was equal to (for CTS=30), twice (for CTS=15), or about 4 times (for CTS=7) the CTS. Thus, correlation between neighboring observations was low and the conditional simulations of daily concentrations for days between neighboring observations (which are used to estimate AMDC) were uncertain. Uncertainty generally decreased as spacing between observations became small in relation to CTS. For example, for CR=30, CTS=15, and ASF=48 (spacing between samples equal to about one-half of the CTS), most of the estimated values for AMDC for carbaryl were between -35 and 53 percent of the known values compared to -49 and 157 percent for CR=30, CTS=15, and ASF=12 (spacing between samples equal to twice the CTS). In addition to higher spacing between observations (relative to CTS), higher SSD also causes higher estimation uncertainty. The maximum value of the SSD ($\sigma[1+0.5\alpha]^{1/2}$, which occurs when the seasonal wave equals its maximum value of 0.5) was about 0.49 for atrazine, 0.54 for carbaryl, and 0.37 for chlorpyrifos. Thus, estimation uncertainty generally was highest for the carbaryl model and lowest for the chlorpyrifos model. For example, for CR=30, ASF=24, and CTS=15, most of the estimated values of AMDC for the carbaryl model were between -35 and 92 percent of the known values compared to -30 and 54 percent for the chlorpyrifos model. The dependence of estimation uncertainty on the SSD was diminished as the observations became closely spaced with respect to CTS. For example, for CR=30, ASF=48, and CTS=30, the 80-percent error bounds for the carbaryl model (-23 to 36 percent) were similar to the error bounds for the chlorpyrifos model (-15 to 40 percent) and the atrazine model (-16 to 36 percent).

As expected, increasing the CR increased bias of the estimated AMDCs (table 1). For CR=70, the estimated AMDC had moderate upward bias (between 25 and 50 percent) for many cases. However, none of the cases had high upward bias (greater than 50 percent). The cases with moderate upward bias tended to occur when the spacing between observations was high in relation to CTS. Increasing the CR also increased estimation uncertainty, especially when the spacing between observations was high in relation to CTS. For example, for CR=70, ASF=24, and CTS=15, most of the estimated values for AMDC for the carbaryl model were between -41 and 151 percent of the known values compared to -35 and

92 percent for CR=30, ASF=24, and CTS=15. The effect of high censoring rates on estimation uncertainty diminished as the spacing between observations became small in relation to CTS. For example, for CR=70, ASF=48, and CTS=15, most of the estimated values for AMDC for the carbaryl model were between -24 and 88 percent of the known values compared to -35 and 53 percent for CR=30, ASF=48, and CTS=15.

Overall, based on the model testing, if the model assumptions are satisfied, the SEAWAVE–QEX model should produce reasonable estimates of AMDCs with as few as 3 years of data, ASF as low as 12 samples per year, and CRs as high as 70 percent. The estimates should be relatively unbiased (bias between -10 and 25 percent) for low CRs (less than 30 percent), but can have moderate upward bias (between about 25 and 50 percent) for higher CRs. The estimates can have high uncertainty, especially when the CR is high and the spacing between observations is large in relation to the CTS. However, provided estimation uncertainty is quantified (for example, using confidence intervals instead of point estimates), the estimates should be useful for evaluating pesticide exposure risk and uncertainty.

Model Assumptions and Limitations

The simulation experiments described in the "Model Testing" section provide evidence that the SEAWAVE–QEX modeling methodology can be a useful and robust tool for estimating concentration extremes based on sparse monitoring data. For the models tested, the estimated AMDCs produced by the model were relatively unbiased and had reasonable uncertainty. Note, however, that the model testing assumed that in each case, the model assumptions were satisfied because the "known" concentrations were generated from the model. In the examples used for selecting the models that were tested, the period of record was long enough and the sampling frequent enough to provide reasonable assurance that the model assumptions were met. When applying the model in practice, the normalized residuals (eq. 11) should be examined in an effort to verify the following assumptions:

- The normalized residuals are approximately normally distributed with mean equal to zero and variance equal to one. Plots of the normalized residuals with time of year (for example, fig. 7) should be examined to identify any obvious seasonality remaining in the mean or standard deviation, as well as any obvious nonnormality (skewness, outliers, and so forth).

- The normalized residuals have exponential correlation function. The empirical correlogram (for example, fig. 8) should be examined for obvious discontinuity at the origin or nonexponential decay as the time lag increases.

The shorter the sampling record, the sparser the sampling frequency, or the more highly censored the data, the more difficult the assumptions are to verify; therefore, more caution should be taken when using the model results. Model verification can be especially difficult when a large percentage of the data (more than 50 percent) are censored, in which case a large percentage of the normalized residuals are generated values from a conditional trace and not observed residuals. The generated residuals, by definition, satisfy the model assumptions.

In addition to verifying model assumptions, the following general limitations of the methodology should be noted:

- Alternative, more data intensive, methods may be available for estimating concentration extremes for certain pesticides, such as atrazine, that have been extensively monitored through various sampling programs (Mosquin and others, 2012). The SEAWAVE–QEX model is proposed as an alternative—not a replacement—for other methods, with the primary goal of increasing the geographic extent, hydrologic conditions, and pesticide use characteristics of sites that can be analyzed. Comparing SEAWAVE–QEX to other approaches is beyond the scope of this report.

- Measurement error is assumed to be negligible. The SEAWAVE–QEX model was designed to handle sparse sampling and censoring, but the model was not designed to handle measurement error (differences between the actual concentration and the observed concentration because of sample collection and preservation techniques or laboratory analytical methods). For the model, measurement error is assumed to be small in relation to the standard deviation of the actual concentrations (σ in eq. 6) and, for practical purposes, can be ignored. The conditional traces of daily concentrations from the model are "exact interpolators" in that the simulated daily concentrations coincide with the observed concentrations on days when samples were collected (unless the observation is censored, in which case the simulated concentration is less than the censoring limit).

- The generated conditional simulations of daily pesticide concentrations from the SEAWAVE–QEX model should not be interpreted as actual daily concentrations. The simulations are designed to mimic the statistical properties of actual data, not to predict actual concentrations. Estimated concentration extremes derived from the conditional simulations for individual sites should not be reported without careful consideration of model assumptions and estimation uncertainty.

Data Preparation and Screening

The simulation experiments described in the "Model Testing" section were used in conjunction with practical considerations involving actual monitoring datasets, such as variable ASF, variable spacing between samples, and gaps in the record, to develop a stepwise screening procedure for determining which datasets are appropriate for SEAWAVE–QEX model applications (fig. 4, data preparation and screening). The following data preparation and screening procedure steps were used for this report. The R function swaveqexMerge (see appendix) can be used for completing the data preparation and screening.

Step 1. Assemble daily streamflow data and compute flow anomalies.—Daily streamflow data from a streamflow-gaging station, at or near the pesticide sampling site, is used to compute the flow anomalies. Generally, the streamflow-gaging station should have upstream drainage area within 5 percent of the drainage area upstream from the pesticide sampling site, and no major diversions or other flow modifications should be between the sampling site and the streamflow-gaging station. The daily streamflow record should consist of complete calendar years starting on January 1 of the beginning year of the pesticide observations and ending on December 31 of the ending year of the pesticide observations, but excluding leap days. Each year, therefore, should have 365 values. Leap days are excluded because the SEAWAVE–QEX model uses a fixed period of 365 days per year for all of the computational algorithms. Missing daily streamflow values are allowed at this step, and every day should have a streamflow record (with placeholders for missing values inserted for days with missing record). The MTFA (eq. 3) and STFA (eq. 4) are computed using the daily streamflow record. If part of the streamflow record is missing, the long-term mean required for computing the MTFA (M in eq. 3) is computed using the nonmissing streamflow record, but the MTFA and STFA for a given day are missing if there are any missing streamflow values for the 30-day period up to and including the given day.

Step 2. Merge pesticide data and daily streamflow data and determine appropriate period of record.—Pesticide observations with missing streamflow anomalies should be removed. These observations cannot be used by the model and removing them, rather than substituting missing values, is required because the algorithms used for fitting the model do not allow missing values for the observed concentrations. Let NOBSG denote the number of pesticide observations and NUCG the number of uncensored observations in a generic year. If, after removing observations with missing flow anomalies, there are fewer than 3 individual years of record that meet the minimum requirements that NOBSG≥6 and NUCG≥2, the dataset is not appropriate for analysis using the SEAWAVE–QEX model. Otherwise, continue with the data screening. If the beginning or ending year of the pesticide record do not meet the minimum requirements (NOBSG≥6 and NUCG≥2), increase the beginning year or decrease the ending year, as required, so that both the beginning and ending year of record meet the minimum requirements. Note that years in the middle of the record that do not meet NOBSG≥6 and NUCG≥2 should be included; however, if there is a long gap (3 or more consecutive years) that fails to meet the minimum requirements, consider adjusting the beginning and ending year, if possible, to exclude the gap and still maintain at least 3 years that meet the minimum requirements.

Step 3. Thin the pesticide data to maintain at least 3 days between consecutive observations.— Most pesticide monitoring datasets, such as those described in the "SEAWAVE–QEX Model Applications" section of this report, have sustained high-frequency sampling (more than two samples per week) for, at most, 1 or 2 years, and even then only for short periods such as 1–3 months. Including all of those high-frequency samples can result in undue influence of those years for fitting the SEAWAVE–QEX model, especially with regard to selecting the best-fit seasonal wave and estimating the linear regression coefficients. The seasonal wave is meant to model "typical" seasonal behavior of pesticide concentrations during a span of several years; not fit the observations in any single year as closely as possible. Therefore, thinning to at least 3 days between samples (at most two samples per week) is advised.

Step 4. Determine if the pesticide data are sufficient for SEAWAVE–QEX model application.—The fourth and final data preparation step is to verify that the pesticide observations remaining after the first three steps satisfy the following minimum data requirements:

- at least 3 individual years with 6 or more observations, 30 percent or more of which are uncensored;
- at least 30 observations for all years combined; and
- at least 10 uncensored observations for all years combined.

These steps should cull most of the sites for which the data are too sparse for analysis, while at the same time keeping as many sites as possible for use in investigations using the simulated daily concentrations from the SEAWAVE–QEX model. The minimum data requirements given in step 4 were motivated by the simulation experiments described in the "Model Testing" section of this report, which indicated that the model results should be reasonable provided there are at least 3 years of pesticide data, ASF consists of 12 or more (equally spaced) observations per year, and at least 30 percent of the observations are uncensored (or, equivalently, at most 70 percent are censored). Those data constraints were relaxed somewhat to accommodate common issues with actual datasets, such as nonuniform spacing between samples, nonuniform ASF, and gaps in the record. As with any model, results should be examined in each case to help confirm that the simulated daily concentrations are reasonable.

SEAWAVE–QEX Model Applications

For each of the four pesticides considered in this report (atrazine, carbaryl, chlorpyrifos, and fipronil), available concentration data from USGS NWQN were screened to determine datasets that could be analyzed using the SEAWAVE–QEX model. Sample collection and preservation methods, laboratory analytical techniques, quality control procedures, and data preparation methods are described in this section and are similar to previous studies (Martin, 2009; Ryberg and others, 2010). The concentration data used in this report are provided in a separate data release (Vecchia and Williams-Sether, 2018), and the discharge data are available in the USGS National Water Information System (U.S. Geological Survey, 2017). Self-contained R functions described in the appendix of this report can be used to verify the results, to complete additional analyses with the data provided, or to analyze user-provided datasets.

Pesticide Data-Processing Steps

Pesticide concentration data for this report are based on flow-weighted, depth- and width-integrated water samples collected and preserved using standard USGS methods and analyzed by the USGS National Water Quality Laboratory using gas chromatography/mass spectrometry (see Ryberg and others, 2010, and references therein, for details). Methods developed by Martin (2009) and modified as described in appendixes 1–4 of Ryberg and others (2010) were used to adjust raw (laboratory reported) concentrations for possible recovery bias and to determine appropriate censoring limits to use for nondetections. The data adjustment process consisted of the following steps (see appendix 1 in Ryberg and others, 2010).

Step 1. Identification of laboratory reporting levels.—"Routine" reporting levels (those excluding raised reporting levels caused by matrix interference or other analytical difficulties) were identified for each pesticide. The laboratory reporting level is the "less-than" concentration value reported for samples in which the pesticide was not detected. The reporting level from the USGS laboratory changed through time and was not generally interpretable as a method detection limit (MDL).

Step 2. Reassignment of censoring limit for routine nondetections.—Detailed analyses of quality assurance/quality control data were used to determine the maximum long-term method detection level (maxLT–MDL) for each pesticide analyte. The detection level may have changed through time, but generally did not indicate any consistent patterns (increasing or decreasing) or large changes for the pesticides during the period of record (1992–2012) used for this report. The maxLT–MDL is the maximum of the MDLs. The maxLT–MDL was determined to be 0.004 micrograms per liter (µg/L) for atrazine, 0.03 µg/L for carbaryl, 0.003 µg/L for chlorpyrifos, and 0.01 µg/L for fipronil (table 1 in Martin, 2009). These values are the default censoring limits assigned to routine nondetections for each site. The default censoring limit may be lowered as described in step 5.

Step 3. Rounding concentrations to a consistent level of precision.—The precision (number of decimal places reported for samples with detected and quantified pesticide concentration) changed though time. Concentrations were rounded to a consistent level of precision depending on various concentration ranges (table 4 in Martin, 2009).

Step 4. Adjustment of concentrations for temporal changes in recovery.—Ideally, the gas chromatography/mass spectrometry analytical method should have 100 percent recovery (quantified concentration equal, on average, to actual concentration). However, because of instrument drift between calibrations, complexity of analytical techniques, and other factors, the actual recovery may vary above or below 100 percent through time. Quality control data from "spiked" samples (samples with known pesticide concentration) were used to model changes in recovery through time, and concentration values for samples with detected and quantified concentrations were adjusted to represent 100 percent recovery. Recovery adjusted concentrations were re-rounded as described in step 3. The recovery adjustment was not applied to the censoring limit (maxLT–MDL) used for nondetections.

Step 5. Reassignment of censoring limit for selected sites to the median of low-level detections.—The default censoring limit assigned for routine nondetections ("<maxLT–MDL") was lowered for some pesticide-site combinations to reflect site-specific differences in detection sensitivity. Some pesticide-site combinations can have frequent occurrences of low-level detections, which are concentration values for which the pesticide was detected and quantified at a value less than maxLT–MDL. For sites with 10 or more low-level detections, the censoring limit was lowered to the median concentration of the low-level detections, denoted as qlow50. For such cases, the concentration value for routine nondetections for that site is recoded to "<qlow50" rather than "<maxLT–MDL". Finally, all quantified concentrations that are less than the censoring limit (either maxLT–MDL or qlow50, depending on the site) are recoded as censored values.

Overview of SEAWAVE–QEX Model Results

A broad overview of the SEAWAVE–QEX parameter estimation results for each pesticide is provided in this section. The parameter estimates for each pesticide and site (best-fit seasonal wave, estimated regression coefficients and approximate *p*-values, estimated values of the SSD parameters, and CTS) are provided in the model archive (see appendix). R functions described in the appendix and provided in the model archive can be used to verify the parameter estimates, generate diagnostic plots (similar to figs. 5–8), and generate conditional traces of daily pesticide concentration for use in investigations of pesticide exposure risk and uncertainty.

To facilitate comparisons of the relative magnitudes of the regression coefficients among sites and pesticides, the estimated regression coefficients (eq. 8) were multiplied by scaling factors:

$$\text{CSWAVE} = 0.5\, b_1;\ \text{CMTFA} = \text{SDMTFA}\, b_2;\ \text{CSTFA} = \text{SDSTFA}\, b_3 \quad (12)$$

where

b_1, b_2, b_3 are the estimated regression coefficients for the seasonal wave, MTFA, and STFA;
CSWAVE is the scaled coefficient for the seasonal wave;
CMTFA is the scaled coefficient for the MTFA;
SDMTFA is the standard deviation of the MTFA;
CSTFA is the scaled coefficient for the STFA; and
SDSTFA is the standard deviation of the STFA.

Loosely interpreted, the higher the magnitude (absolute value) of a scaled coefficient, the more variability of the observed pesticide concentrations is explained by that variable and the more important that variable is in the model. To simplify the notation and discussion of results in this section, SSD is used to denote the estimated seasonal standard deviation ($SSD^*=\sigma^*(1+\alpha^* W^*)^{1/2}$, eq. 11) and CTS is used to denote the estimated correlation time scale (CTS^*). For $\alpha^*>0$ (nonconstant SSD), the minimum and maximum values of SSD (which occur when $W^*= -0.5$ and $W^*=0.5$, respectively) differ. For $\alpha^*=0$, and minimum and maximum values of SSD both equal the same constant value (σ^*). The maximum value of the estimated seasonal standard deviation ($\sigma^*(1+0.5\alpha^*)^{1/2}$) is denoted by maxSSD.

Atrazine

For atrazine, 112 USGS sampling sites were identified for the SEAWAVE–QEX model application (table 2). Atrazine is one of the most widely used and commonly detected pesticides in streams throughout the Nation (Stone and others, 2014). Consequently, many sites had long RLs and low CRs. A total of 900 site years were analyzed, with 99 sites having RL≥5 years and 45 sites having RL=10 years. Many sites had more than 10 years of record available, but a maximum RL of 10 years was used. When selecting the period of record for sites with more than 10 years of data, or for sites with gaps in the record, preference was given to including the most recent years. The total number of observations (NOBS) generally was large, and the percent of the observation that were uncensored (PUC) was high: 85 sites had NOBS≥60 and 100 sites had PUC≥80.

The scaled coefficients for the seasonal wave, MTFA, and STFA (CSWAVE, CMTFA, and CSTFA in eq. 12) for the atrazine sites are shown in figure 15. Sites are ordered with respect to increasing USGS station number, and the site numbers are given in table 2. Most (88) sites had CSWAVE≥0.4 and all but one of the coefficients was statistically significant (approximate *p*-value less than 0.05) (fig. 15*A*). Only 24 sites had CSWAVE<0.4, including 11 sites among sites 93–112 (these sites are in the Western United States; table 2). The values for CMTFA (fig. 15*B*) generally were smaller in magnitude compared with CSWAVE. However, CMTFA was significant for many (51) sites and the significant coefficients were evenly distributed between positive and negative values. The values for CSTFA (fig. 15*C*) were similar in magnitude to CMTFA and most (74) were significant. However, unlike CMTFA, most (68) of the significant values for CSTFA were positive.

The estimated SSDs and CTSs for the atrazine sites are shown in figure 16. The SSD (fig. 16*A*) generally was small in relation to CSWAVE (fig. 15*A*): most (93) sites had maxSSD<0.4 whereas most (88) sites had CSWAVE≥0.4. The CTS (fig. 16*B*) was variable: 48 sites had CTS≤10 days and 64 sites had CTS>10 days. There was no readily apparent correlation between the values of CTS and SSD or between the values of CTS and CSWAVE. Generally, sites with lower maxSSD, higher CTS, or both, should have lower uncertainty in estimated pesticide concentration extremes computed using conditional simulations from the SEAWAVE–QEX model.

Carbaryl

For carbaryl, 38 USGS sampling sites were identified for SEAWAVE–QEX model application (table 3). Compared to atrazine, carbaryl concentrations tended to be much more highly censored. Consequently, there were only 38 sites that met the minimum data requirements for carbaryl. A total of 308 site years were analyzed, with 36 sites having RL≥5 years and 13 sites having RL≥10 years. When selecting the period of record for sites with gaps in the record, preference was given to including the most recent years. NOBS generally was large (34 sites had NOBS≥60) and PUC was moderate to low (only 9 sites had PUC>40).

The scaled coefficients for the seasonal wave, MTFA, and STFA for the carbaryl sites are shown in figure 17. Sites are ordered with respect to increasing USGS station number and the site number is given in table 3 (note that the site numbers for carbaryl differ from the site numbers for atrazine). Most (28) carbaryl sites had CSWAVE≥0.4 and all but 3 of the coefficients were significant (fig. 17*A*). The values for CMTFA (fig. 17*B*) generally were smaller in magnitude compared with CSWAVE, and only nine of the coefficients were significant. The values for CSTFA (fig. 17*C*) generally were smaller in magnitude compared with CSWAVE, but unlike CMTFA, most (27) of the coefficients were significant and 26 of the significant coefficients were positive.

The estimated SSDs and CTSs for the carbaryl sites are shown in figure 18. The SSD (fig. 18*A*) generally was similar in magnitude to CSWAVE (fig. 17*A*): most (36 sites) had maxSSD≥0.4 and most (28) sites had CSWAVE≥0.4. This similarity between the values of maxSSD and CSWAVE for carbaryl is in contrast to atrazine, for which maxSSD generally was smaller than CSWAVE. Also, unlike atrazine, which had variable CTS (fig. 16*B*), most (34) of the carbaryl sites had CTS<10. Because of the larger values of maxSSD and lower values of CTS for carbaryl (compared to atrazine), estimated

Table 2. U.S. Geological Survey water-quality sampling sites, period of record, and number of observations used for application of SEAWAVE–QEX model for atrazine.

[USGS, U.S. Geological Survey; km², square kilometers; NOBS, number of observations; NUC, number of uncensored observations; PUC, percent uncensored observations]

Site number	USGS station number	Site name	Drainage area, in km²	Starting year	Ending year	Record length, in years	NOBS	NUC	PUC
1	01184000	Connecticut River at Thompsonville, Connecticut	25,012	1997	2006	10	107	65	61
2	01349150	Canajoharie Creek near Canajoharie, New York	155	1998	2007	10	145	144	99
3	01356190	Lisha Kill northwest of Niskayuna, New York	49	2001	2010	10	98	39	40
4	01357500	Mohawk River at Cohoes, New York	8,986	1997	2006	10	134	131	98
5	01374987	Kisco River below Mount Kisco, New York	46	2000	2008	9	186	66	35
6	01463500	Delaware River at Trenton, New Jersey	17,574	2003	2012	10	99	97	98
7	01464907	Little Neshaminy Creek at Valley Road near Neshaminy, Pennsylvania	71	1999	2004	6	77	76	99
8	01472157	French Creek near Phoenixville, Pennsylvania	158	1999	2004	6	58	57	98
9	01474500	Schuylkill River at Philadelphia, Pennsylvania	4,889	1999	2004	6	68	67	99
10	01493112	Chesterville Branch near Crumpton, Maryland	18	1999	2002	4	51	51	100
11	01493500	Morgan Creek near Kennedyville, Maryland	33	2002	2004	3	52	51	98
12	01555400	East Mahantango Creek at Klingerstown, Pennsylvania	116	1997	2000	4	48	48	100
13	01578310	Susquehanna River at Conowingo, Maryland	70,161	1996	2004	9	65	65	100
14	01621050	Muddy Creek at Mount Clinton, Virginia	43	1997	2001	5	66	66	100
15	01654000	Accotink Creek near Annandale, Virginia	62	1997	2001	5	63	50	79
16	02087580	Swift Creek near Apex, North Carolina	56	2002	2011	10	116	94	81
17	02089500	Neuse River at Kinston, North Carolina	7,020	2003	2012	10	142	138	97
18	02091500	Contentnea Creek at Hookerton, North Carolina	1,898	1997	2006	10	119	110	92
19	02169570	Gills Creek at Columbia, South Carolina	164	2001	2006	6	59	58	98
20	02174250	Cow Castle Creek near Bowman, South Carolina	64	1999	2008	10	105	89	85
21	02306774	Rocky Creek at State Highway 587 at Citrus Park, Florida	49	2002	2004	3	49	38	78
22	02318500	Withlacoochee River at U.S. 84 near Quitman, Georgia	3,872	1999	2008	10	100	99	99
23	02335870	Sope Creek near Marietta, Georgia	86	2003	2012	10	127	121	95
24	02338000	Chattahoochee River near Whitesburg, Georgia	6,252	2003	2012	10	144	141	98
25	02350080	Lime Creek near Cobb, Georgia	162	2001	2007	7	89	77	87
26	02359170	Apalachicola River near Sumatra, Florida	49,771	2008	2012	5	52	52	100
27	0242354750	Cahaba Valley Creek at Cross Creek Road at Pelham, Alabama	71	2003	2012	10	99	99	100
28	02444490	Bogue Chitto Creek near Memphis, Alabama	138	1999	2004	6	73	73	100
29	03086000	Ohio River at Sewickley, Pennsylvania	50,496	2001	2007	7	88	83	94
30	03216600	Ohio River at Greenup Dam near Greenup, Kentucky	159,235	1998	2007	10	132	129	98
31	03267900	Mad River at Saint Paris Pike at Eagle City, Ohio	798	1999	2004	6	103	100	97
32	03303280	Ohio River at Cannelton Dam, at Cannelton, Indiana	249,355	2003	2012	10	125	125	100

Table 2. U.S. Geological Survey water-quality sampling sites, period of record, and number of observations used for application of SEAWAVE–QEX model for atrazine. —Continued

[USGS, U.S. Geological Survey; km², square kilometers; NOBS, number of observations; NUC, number of uncensored observations; PUC, percent uncensored observations]

Site number	USGS station number	Site name	Drainage area, in km²	Starting year	Ending year	Record length, in years	NOBS	NUC	PUC
33	03353637	Little Buck Creek near Indianapolis, Indiana	51	1995	2004	10	127	127	100
34	03357330	Big Walnut Creek near Roachdale, Indiana	374	2002	2011	10	54	54	100
35	[1]03374100	White River at Hazleton, Indiana	29,069	2003	2012	10	136	135	99
36	[2]03378500	Wabash River at New Harmony, Indiana	75,585	2003	2012	10	127	127	100
37	03466208	Big Limestone Creek near Limestone, Tennessee	203.8	1996	2004	9	95	95	100
38	03467609	Nolichucky River near Lowland, Tennessee	4,374	1996	2004	9	92	89	97
39	03575100	Flint River at Brownsboro, Alabama	967	1999	2007	9	105	105	100
40	04072150	Duck Creek near Howard, Wisconsin	279	1997	2001	5	57	57	100
41	04161820	Clinton River at Sterling Heights, Michigan	811	2002	2006	5	46	46	100
42	04178000	St. Joseph River near Newville, Indiana	1,537	1996	2004	9	81	81	100
43	04186500	Auglaize River near Fort Jennings, Ohio	861	2002	2008	7	74	73	99
44	04193500	Maumee River at Waterville, Ohio	16,274	1998	2007	10	113	112	99
45	04234000	Fall Creek near Ithaca, New York	327	1997	1999	3	37	37	100
46	04264331	St. Lawrence River at Cornwall Ontario near Massena, New York	10,366	2008	2012	5	57	57	100
47	05288705	Shingle Creek at Queen Avenue in Minneapolis, Minnesota	103	1997	2006	10	104	87	84
48	05320270	Little Cobb River near Beauford, Minnesota	331	1998	2007	10	83	82	99
49	05330000	Minnesota River near Jordan, Minnesota	41,787	1996	1998	3	37	37	100
50	05331580	Mississippi River below Lock and Dam 2 at Hastings, Minnesota	95,179	1996	2004	9	88	87	99
51	05412500	Turkey River at Garber, Iowa	4,023	2004	2012	9	100	100	100
52	05420500	Mississippi River at Clinton, Iowa	220,066	2003	2012	10	116	116	100
53	05420680	Wapsipinicon River near Tripoli, Iowa	901	1996	2004	9	56	56	100
54	05422000	Wapsipinicon River near De Witt, Iowa	41	2004	2012	9	99	99	100
55	05449500	Iowa River near Rowan, Iowa	1,105	1996	2004	9	87	87	100
56	05451210	South Fork Iowa River northeast of New Providence, Iowa	583	2003	2012	10	125	125	100
57	05464220	Wolf Creek near Dysart, Iowa	773	1996	1999	4	51	51	100
58	05465500	Iowa River at Wapello, Iowa	32,341	2003	2012	10	136	136	100
59	05474000	Skunk River at Augusta, Iowa	11,165	2004	2012	9	98	98	100
60	05490500	Des Moines River at Keosauqua, Iowa	36,310	2004	2012	9	121	121	100
61	05525500	Sugar Creek at Milford, Illinois	1,159	1999	2004	6	64	64	100
62	05531500	Salt Creek at Western Springs, Illinois	306	1999	2007	9	99	99	100
63	05532500	Des Plaines River at Riverside, Illinois	1,656	1999	2004	6	46	46	100
64	05572000	Sangamon River at Monticello, Illinois	1,447	1997	2006	10	122	122	100
65	05586100	Illinois River at Valley City, Illinois	68,925	2003	2012	10	102	102	100
66	05587455	Mississippi River below Grafton, Illinois	444,074	2004	2012	9	105	105	100
67	06295000	Yellowstone River at Forsyth, Montana	102,211	1999	2004	6	66	48	73
68	06329500	Yellowstone River near Sidney, Montana	177,171	2001	2007	7	54	35	65

Table 2. U.S. Geological Survey water-quality sampling sites, period of record, and number of observations used for application of SEAWAVE–QEX model for atrazine. —Continued

[USGS, U.S. Geological Survey; km², square kilometers; NOBS, number of observations; NUC, number of uncensored observations; PUC, percent uncensored observations]

Site number	USGS station number	Site name	Drainage area, in km²	Starting year	Ending year	Record length, in years	NOBS	NUC	PUC
69	06485500	Big Sioux River at Akron, Iowa	17,815	2004	2010	7	79	79	100
70	06607500	Little Sioux River near Turin, Iowa	9,179	2004	2012	9	104	104	100
71	06609500	Boyer River at Logan, Iowa	2,259	2004	2012	9	100	100	100
72	06610000	Missouri River at Omaha, Nebraska	780,965	2003	2012	10	129	129	100
73	06713500	Cherry Creek at Denver, Colorado	1,057	2001	2009	9	111	107	96
74	06714000	South Platte River at Denver, Colorado	10,001	1994	2000	7	59	47	80
75	06753990	Lonetree Creek near Greeley, Colorado	1,481	2001	2004	4	44	44	100
76	06754000	South Platte River near Kersey, Colorado	25,021	1997	2006	10	108	108	100
77	06800000	Maple Creek near Nickerson, Nebraska	962	1997	2006	10	171	171	100
78	06800500	Elkhorn River at Waterloo, Nebraska	17,446	2002	2004	3	41	41	100
79	06805500	Platte River at Louisville, Nebraska	12	2003	2012	10	114	114	100
80	06810000	Nishnabotna River above Hamburg, Iowa	7,279	2004	2012	9	104	103	99
81	06817700	Nodaway River near Graham, Missouri	3,933	1996	2001	6	38	38	100
82	06934500	Missouri River at Hermann, Missouri	1,277,918	2003	2012	10	132	132	100
83	07022000	Mississippi River at Thebes, Illinois	1,769,764	2003	2012	10	129	129	100
84	07053250	Yocum Creek near Oak Grove, Arkansas	138	2006	2012	7	58	25	43
85	07144100	Little Arkansas River near Sedgwick, Kansas	3,242	1995	2004	10	48	48	100
86	07288650	Bogue Phalia near Leland, Mississippi	1,439	1999	2008	10	101	99	98
87	07288955	Yazoo River below Steele Bayou near Long Lake, Mississippi	34,742	2003	2012	10	134	134	100
88	07369500	Tensas River at Tendal, Louisiana	2,514	1996	1999	4	62	62	100
89	07374000	Mississippi River at Baton Rouge, Louisiana	2,880,682	2004	2012	9	103	103	100
90	07374525	Mississippi River at Belle Chasse, Louisiana	2,881,390	2009	2012	4	46	46	100
91	07381590	Wax Lake Outlet at Calumet, Louisiana	6,297	2009	2012	4	48	48	100
92	07381600	Lower Atchafalaya River at Morgan City, Louisiana	3,130,494	2006	2012	7	90	90	100
93	08057200	White Rock Creek at Greenville Avenue Dallas, Texas	190	2002	2011	10	121	120	99
94	08057410	Trinity River below Dallas, Texas	16,253	2003	2012	10	143	143	100
95	08064100	Chambers Creek near Rice, Texas	2,091	2001	2004	4	52	52	100
96	08116650	Brazos River near Rosharon, Texas	103,578	2008	2012	5	56	56	100
97	08178800	Salado Creek at Loop 13, San Antonio, Texas	489	2002	2008	7	72	72	100
98	08181800	San Antonio River near Elmendorf, Texas	4,529	1997	2004	8	77	77	100
99	09163500	Colorado River near Colorado-Utah State Line	46,122	1997	2004	8	61	41	67
100	10168000	Little Cottonwood Creek at Jordan River near Salt Lake City	105	1999	2007	9	92	64	70
101	11074000	Santa Ana River below Prado Dam, California	5,680	2004	2012	9	125	68	54
102	11274538	Orestimba Creek at River Road near Crows Landing, California	465	1997	2004	8	130	59	45
103	11303500	San Joaquin River near Vernalis, California	35,855	2003	2012	10	166	80	48
104	12505450	Granger Drain at Granger, Washington	150	1999	2004	6	105	104	99

Table 2. U.S. Geological Survey water-quality sampling sites, period of record, and number of observations used for application of SEAWAVE–QEX model for atrazine. —Continued

[USGS, U.S. Geological Survey; km², square kilometers; NOBS, number of observations; NUC, number of uncensored observations; PUC, percent uncensored observations]

Site number	USGS station number	Site name	Drainage area, in km²	Starting year	Ending year	Record length, in years	NOBS	NUC	PUC
105	12510500	Yakima River at Kiona, Washington	14,060	1999	2008	10	91	78	86
106	13092747	Rock Creek above Highway 30/93 crossing at Twin Falls, Idaho	632	1997	2005	9	127	98	77
107	13154500	Snake River at King Hill, Idaho	62,321	1998	2007	10	115	87	76
108	13351000	Palouse River at Hooper, Washington	6,219	1997	2003	7	93	42	45
109	14201300	Zollner Creek near Mount Angel, Oregon	41	1999	2008	10	120	120	100
110	14206950	Fanno Creek at Durham, Oregon	82	2001	2011	11	142	102	72
111	14211720	Willamette River at Portland, Oregon	28,922	2003	2012	10	155	136	88
112	14246900	Columbia River at Beaver Army Terminal near Quincy, Oregon	619,784	2003	2012	10	100	59	59

[1]Streamflow data are for White River at Petersburg, Indiana (U.S. Geological Survey station number 03374000).

[2]Streamflow data are for Wabash River at Mt. Carmel, Illinois (U.S. Geological Survey station number 03377500).

carbaryl concentration extremes computed using conditional simulations from the SEAWAVE–QEX model generally should have high uncertainty compared to atrazine.

Chlorpyrifos

For chlorpyrifos, 34 USGS sampling sites were identified for SEAWAVE–QEX model application (table 4). A total of 292 site years were analyzed, with 30 sites having RL≥5 and 14 sites having RL≥10. When selecting the period of record for sites with gaps in the record, preference was given to including the most recent years. NOBS was variable (28 sites had NOBS≥60 and 5 sites had NOBS<40) and PUC was low (only 9 sites had PUC≥40 and 17 sites had PUC<25).

The coefficients for the seasonal wave, MTFA, and STFA for the chlorpyrifos sites are shown in figure 19. Sites are ordered with respect to increasing USGS station number and the site number is given in table 4. CSWAVE was variable (fig. 19*A*): 18 sites had CSWAVE≥0.4, 16 sites had CSWAVE<0.4, and all but 4 of the coefficients were significant. The values for CMTFA (fig. 19*B*) generally were smaller in magnitude compared with CSWAVE; however, 12 coefficients were significant and most (11) of the significant coefficients were positive. This result is in contrast to atrazine and carbaryl (figs. 15*B* and 17*B*), for which the values of CMTFA were evenly split between positive and negative values. However, similar to atrazine and carbaryl, the values of CSTFA for chlorpyrifos (fig. 19*C*) were mostly positive and many (15) were significant.

The estimated SSDs for chlorpyrifos (fig. 20*A*) were variable: 20 sites had maxSSD≥0.4 and 14 sites had maxSSD<0.4. The estimated values of CTS (fig. 20*B*) also were variable: 18 sites had CTS≤10 and 16 sites had CTS>10. The combination of variable SSD and variable CTS indicates that uncertainty of the estimated chlorpyrifos concentration extremes computed using conditional simulations from the SEAWAVE–QEX model should have variable uncertainty.

Fipronil

For fipronil, 33 USGS sampling sites were identified for SEAWAVE–QEX model application (table 5). A total of 219 site years were analyzed, with 23 sites having RL≥5 years and 4 sites having RL≥10 years. When selecting the period of record for sites with gaps in the record, preference was given to including the most recent years. NOBS was variable (22 sites had NOBS≥60 and 5 sites had NOBS<40) and PUC was variable (16 sites had PUC>40 and 10 sites had PUC<25).

The coefficients for the seasonal wave, MTFA, and STFA for the fipronil sites are shown in figure 21. Sites are ordered with respect to increasing USGS station number and the site number is given in table 5. CSWAVE (fig. 21*A*) generally was small (29 sites had CSWAVE≤0.4 and only 4 sites had CSWAVE>0.4) and all but 7 of the coefficients were significant. The values for CMTFA (fig. 21*B*) generally were comparable in magnitude compared with CSWAVE, 23 coefficients were significant, and (like atrazine and carbaryl) the significant coefficients were equally split between positive and negative values. Similar to the other pesticides, many (16) of the values of CSTFA for fipronil (fig. 21*C*) were significant and most (15) of the significant coefficients were positive. However, the values of CSTFA for fipronil generally were small in comparison to the other pesticides.

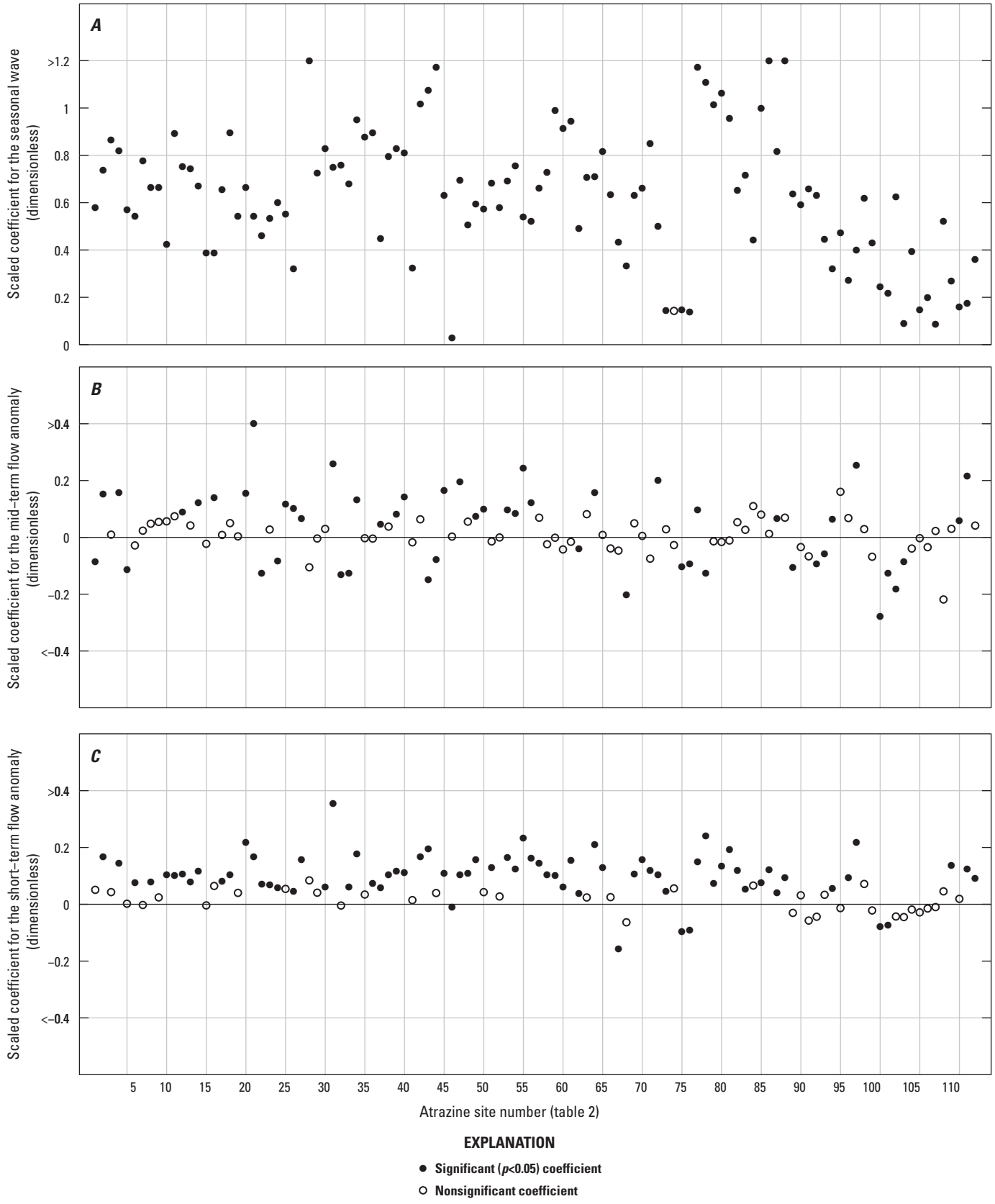

Figure 15. Estimated regression coefficients from SEAWAVE–QEX model results for atrazine. *A*, scaled coefficient for seasonal wave; *B*, scaled coefficient for mid-term flow anomaly; *C*, scaled coefficient for short-term flow anomaly.

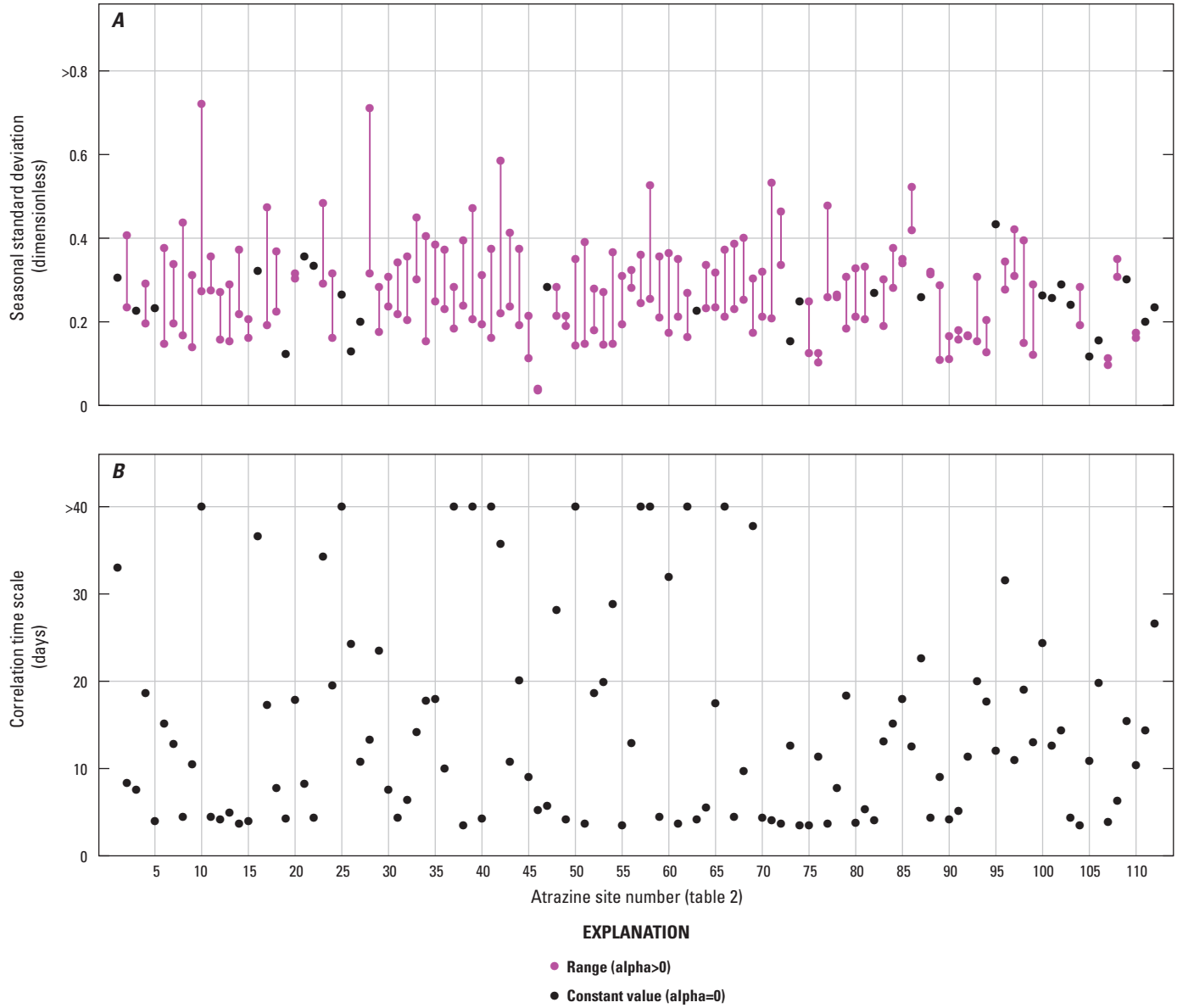

Figure 16. Estimated seasonal standard deviations and correlation time scales from SEAWAVE–QEX model results for atrazine. *A*, seasonal standard deviation; *B*, correlation time scale.

Table 3. U.S. Geological Survey water-quality sampling sites, period of record, and number of observations used for application of SEAWAVE–QEX model for carbaryl.

[USGS, U.S. Geological Survey; km², square kilometers; NOBS, number of observations; NUC, number of uncensored observations; PUC, percent uncensored observations]

Site number	USGS station number	Site name	Drainage area, in km²	Starting year	Ending year	Record length, in years	NOBS	NUC	PUC
1	01356190	Lisha Kill northwest of Niskayuna, New York	49	2002	2008	7	70	23	33
2	01374987	Kisco River below Mount Kisco, New York	46	2000	2008	9	186	54	29
3	01464907	Little Neshaminy Creek at Valley Road near Neshaminy, Pennsylvania	71	1999	2004	6	77	28	36
4	01474500	Schuylkill River at Philadelphia, Pennsylvania	4,889	1999	2004	6	68	21	31
5	01654000	Accotink Creek near Annandale, Virginia	62	1994	2000	7	79	39	49
6	02087580	Swift Creek near Apex, North Carolina	56	2002	2009	8	92	41	45
7	02089500	Neuse River at Kinston, North Carolina	7,020	2003	2012	10	128	42	33
8	02091500	Contentnea Creek at Hookerton, North Carolina	1,897	1997	2008	12	142	35	25
9	02169570	Gills Creek at Columbia, South Carolina	164	1996	2005	10	86	19	22
10	02318500	Withlacoochee River at U.S. 84, near Quitman, Georgia	3,872	1995	2004	10	102	15	15
11	02335870	Sope Creek near Marietta, Georgia	86	2002	2012	11	150	42	28
12	02338000	Chattahoochee River near Whitesburg, Georgia	6,252	2001	2010	10	141	61	43
13	0242354750	Cahaba Valley Creek at Cross Creek Road at Pelham, Alabama	71	2001	2010	10	98	23	23
14	03216600	Ohio River at Greenup Dam near Greenup, Kentucky	159,235	1997	2002	6	80	16	20
15	03353637	Little Buck Creek near Indianapolis, Indiana	51	1995	2004	10	127	34	27
16	04161820	Clinton River at Sterling Heights, Michigan	811	2003	2006	4	38	16	42
17	05288705	Shingle Creek at Queen Avenue in Minneapolis, Minnesota	103	2003	2010	8	76	19	25
18	05531500	Salt Creek at Western Springs, Illinois	306	1999	2007	9	99	30	30
19	05532500	Des Plaines River at Riverside, Illinois	1,656	1999	2004	6	53	17	32
20	06713500	Cherry Creek at Denver, Colorado.	1,057	2001	2011	11	132	52	39
21	06714000	South Platte River at Denver, Colorado.	10,001	1994	2000	7	59	37	63
22	06754000	South Platte River near Kersey, Colorado	25,021	1997	2005	9	104	35	34
23	07031692	Fletcher Creek at Sycamore View Road at Memphis	80	1997	2004	8	52	26	50
24	08057200	White Rock Creek at Greenville Avenue, Dallas, Texas	190	2002	2011	10	121	49	40
25	08057410	Trinity River below Dallas, Texas	16,253	2003	2012	10	143	35	24
26	08178800	Salado Creek at Loop 13, San Antonio, Texas	489	2002	2008	7	72	16	22
27	08181800	San Antonio River near Elmendorf, Texas	4,529	1997	2004	8	77	16	21
28	10168000	Little Cottonwood Creek at Jordan River near Salt Lake City	105	1999	2009	11	109	34	31
29	11074000	Santa Ana River below Prado Dam, California	5,680	2003	2012	10	133	23	17
30	11274538	Orestimba Creek at River Road near Crows Landing, California	465	1992	2000	9	104	24	23
31	11303500	San Joaquin River near Vernalis, California	35,855	1993	1999	7	85	22	26
32	11447360	Arcade Creek near Del Paso Heights, California	98	2001	2008	8	82	39	48
33	12128000	Thornton Creek near Seattle, Washington	31	2001	2007	7	74	20	27
34	12505450	Granger Drain at Granger, Washington	150	2002	2004	3	83	23	28
35	12510500	Yakima River at Kiona, Washington	14,060	1999	2004	6	69	11	16
36	14201300	Zollner Creek near Mount Angel, Oregon	41	1999	2005	7	94	21	22
37	14206950	Fanno Creek at Durham, Oregon	82	2001	2005	5	73	35	48
38	14211720	Willamette River at Portland, Oregon	28,922	2006	2011	6	104	22	21

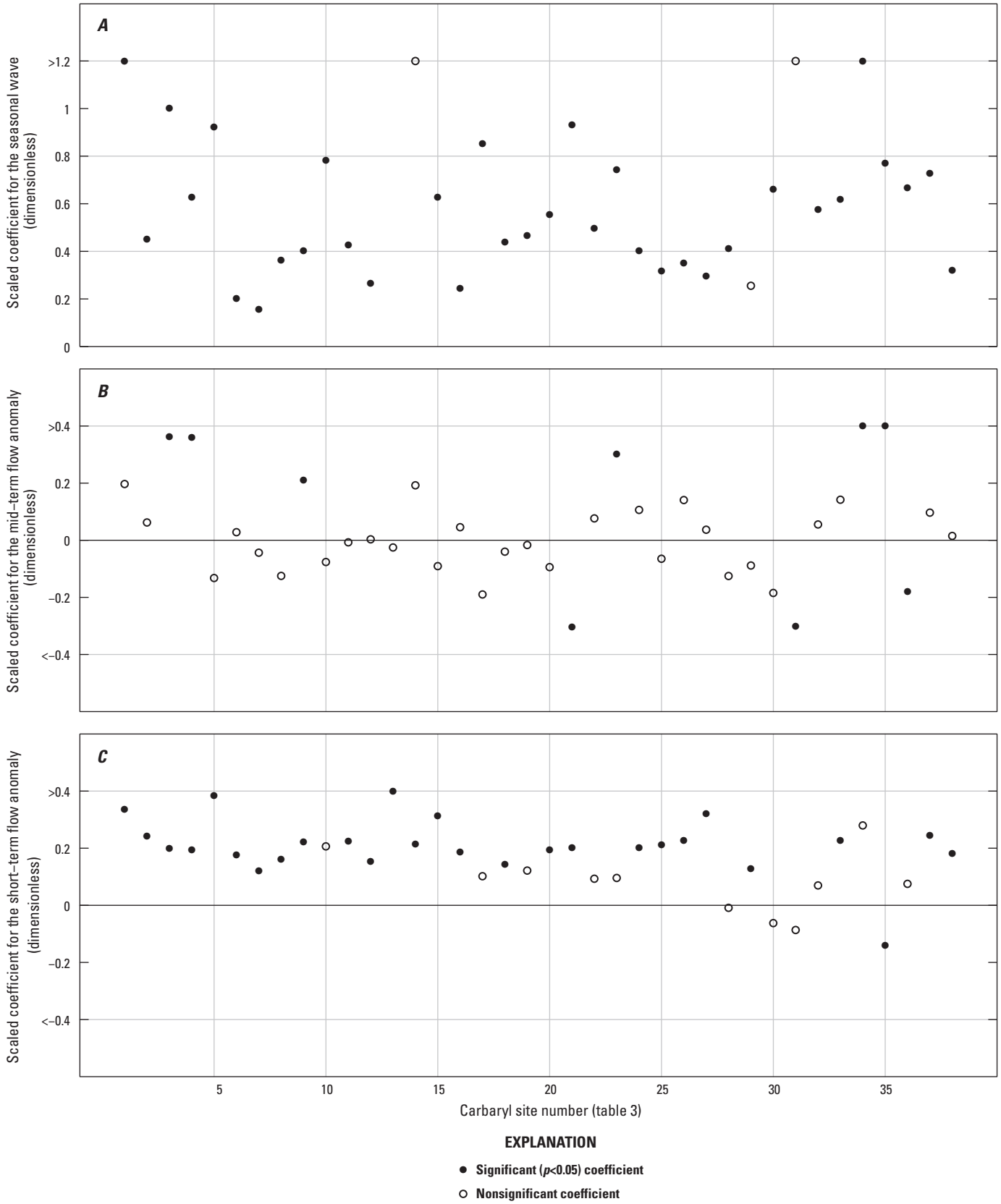

Figure 17. Estimated regression coefficients from SEAWAVE–QEX model results for carbaryl. *A*, scaled coefficient for seasonal wave; *B*, scaled coefficient for mid-term flow anomaly; *C*, scaled coefficient for short-term flow anomaly.

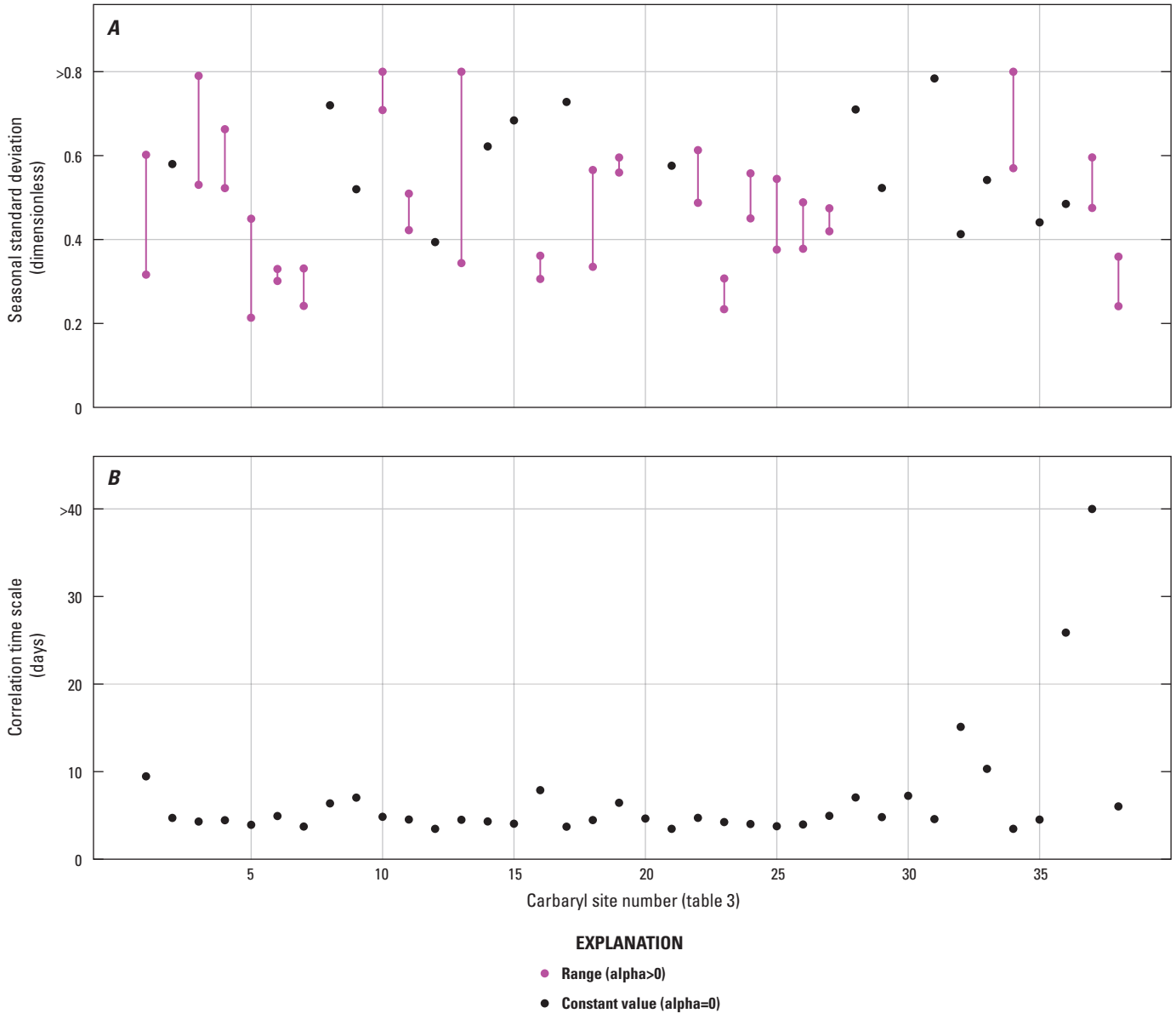

Figure 18. Estimated seasonal standard deviations and correlation time scales from SEAWAVE–QEX model results for carbaryl. *A*, seasonal standard deviation; *B*, correlation time scale.

The estimated SSDs for fipronil (fig. 22*A*) generally were small in comparison to the other pesticides: 29 sites had maxSSD≤0.4 and only 4 sites has maxSSD>0.4. Like atrazine and carbaryl, the estimated values of CTS for fipronil (fig. 22*B*) were variable: 20 sites had CTS<10 and 14 sites had CTS>10. However, unlike the other pesticides, for fipronil there was a statistically significant positive correlation between the values of SSD and CTS—the Kendall's rank correlation (r) between σ* and *CTS**was r=0.353 (p=0.004). Consequently, all four of the sites with large estimated SSD (maxSSD>0.4) also had high estimated CTS (CTS>20). This positive correlation between SSD and CTS, combined with the generally small values of maxSSD, indicate that uncertainty of the estimated fipronil concentration extremes computed using conditional simulations from the SEAWAVE–QEX model generally should have low uncertainty compared to the other pesticides.

Table 4. U.S. Geological Survey water-quality sampling sites, period of record, and number of observations used for application of SEAWAVE–QEX model for chlorpyrifos.

[USGS, U.S. Geological Survey; km², square kilometers; NOBS, number of observations; NUC, number of uncensored observations; PUC, percent uncensored observations]

Site number	USGS station number	Site name	Drainage area, in km²	Starting year	Ending year	Record length, in years	NOBS	NUC	PUC
1	01403900	Bound Brook at Middlesex, New Jersey	132	1996	1998	3	37	18	49
2	01555400	East Mahantango Creek at Klingerstown, Pennsylvania	116	1993	2000	8	91	13	14
3	01654000	Accotink Creek near Annandale, Virgina	62	1994	2000	7	79	31	39
4	02091500	Contentnea Creek at Hookerton, North Carolina	1,897	1997	2008	12	142	18	13
5	02174250	Cow Castle Creek near Bowman, South Carolina	64	1999	2008	10	105	53	50
6	02335870	Sope Creek near Marietta, Georgia	86	1993	2000	8	84	33	39
7	0242354750	Cahaba Valley Creek at Cross Creek Road at Pelham, Alabama	71	1999	2005	7	92	17	18
8	02424000	Cahaba River at Centreville, Alabama	2,659	1999	2000	2	35	14	40
9	02444490	Bogue Chitto Creek near Memphis, Alabama	138	1999	2003	5	66	16	24
10	03303280	Ohio River at Cannelton Dam at Cannelton, Indiana	249,355	1996	2000	5	63	13	21
11	03353637	Little Buck Creek near Indianapolis, Indiana	51	1992	2001	10	150	65	43
12	[1]03374100	White River at Hazleton, Indiana	29,069	1992	1996	5	119	34	29
13	[2]03378500	Wabash River at New Harmony, Indiana	75,585	1997	2004	8	109	18	17
14	04186500	Auglaize River near Fort Jennings, Ohio	861	1996	2005	10	78	13	17
15	04193500	Maumee River at Waterville, Ohio	16,273	1996	2007	12	142	29	20
16	05320270	Little Cobb River near Beauford, Minnesota	331	2005	2007	3	38	12	32
17	05572000	Sangamon River at Monticello, Illinois	1,447	1997	2008	12	145	14	10
18	06609500	Boyer River at Logan, Iowa	2,259	2004	2012	9	100	12	12
19	06713500	Cherry Creek at Denver, Colorado.	1,057	1993	1994	2	31	12	39
20	06800000	Maple Creek near Nickerson, Nebraska	962	1997	2012	16	239	52	22
21	06805500	Platte River at Louisville, Nebraska	12	1992	2000	9	97	13	13
22	07031692	Fletcher Creek at Sycamore View Road at Memphis	80	1997	2004	8	52	26	50
23	08057200	White Rock Creek at Greenville Avenue, Dallas, Texas	190	1997	2007	11	123	37	30
24	08057410	Trinity River below Dallas, Texas	16,253	1995	1999	5	36	11	31
25	11273500	Merced River at River Road Bridge near Newman, California	3,287	2002	2009	8	80	27	34
26	11274538	Orestimba Creek at River Road near Crows Landing, California	465	1997	2006	10	145	88	61
27	11303500	San Joaquin River near Vernalis, California	35,855	1997	2012	16	273	132	48
28	11447360	Arcade Creek near Del Paso Heights, California	98	1997	2008	12	110	56	51
29	11447650	Sacramento River at Freeport, California	59,572	2004	2011	8	115	28	24
30	12505450	Granger Drain at Granger, Washington	150	1999	2004	6	105	19	18
31	12510500	Yakima River at Kiona, Washington	14,059	1999	2008	10	91	19	21
32	14201300	Zollner Creek near Mount Angel, Oregon	41	1997	2008	12	140	106	76
33	14206950	Fanno Creek at Durham, Oregon	82	2001	2006	6	82	15	18
34	14211720	Willamette River at Portland, Oregon	28,921	1996	2012	17	249	54	22

[1]Streamflow data are for White River at Petersburg, Indiana (U.S. Geological Survey station number 03374000).

[2]Streamflow data are for Wabash River at Mt. Carmel, Illinois (U.S. Geological Survey station number 03377500).

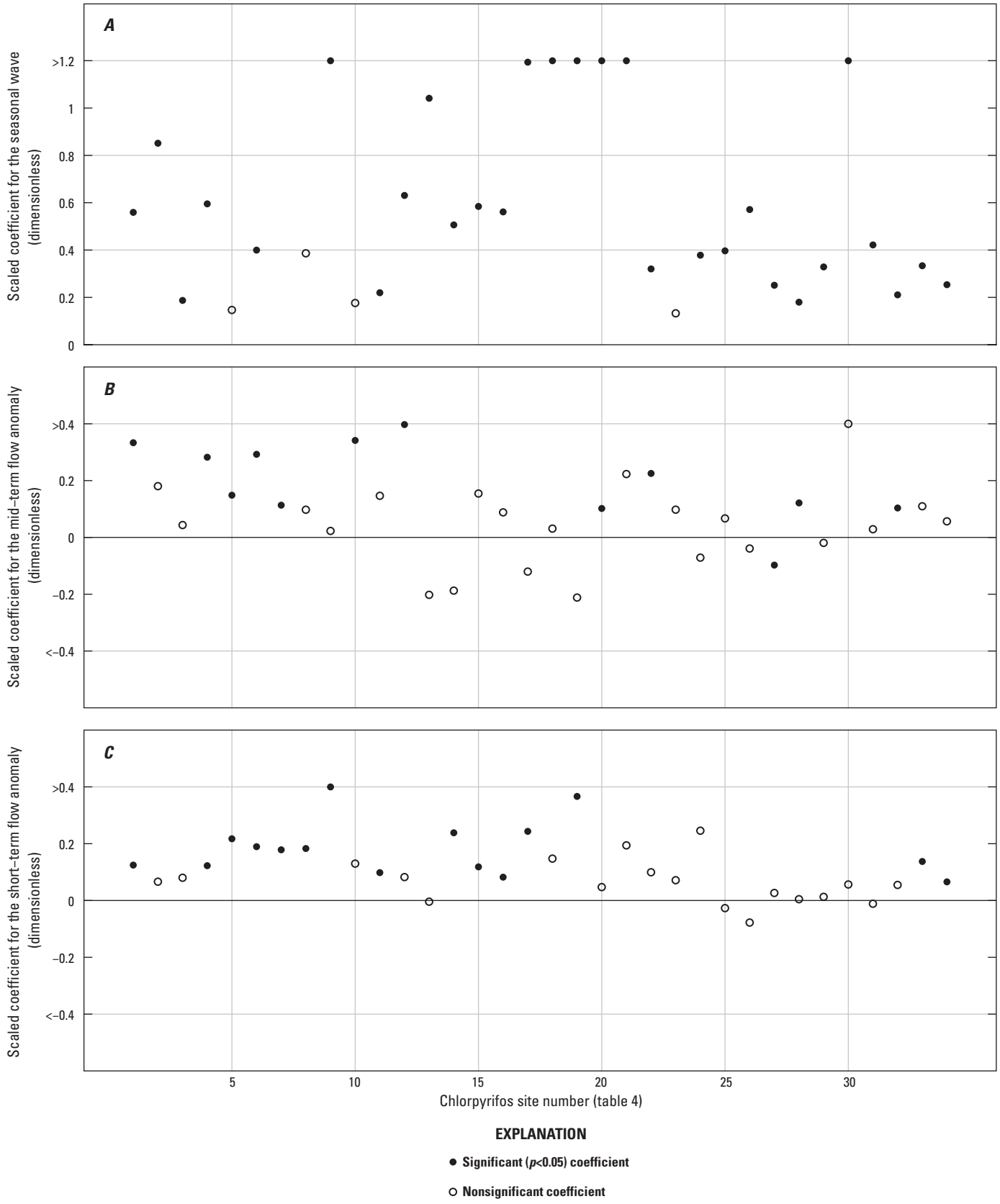

Figure 19. Estimated regression coefficients from SEAWAVE–QEX model results for chlorpyrifos. *A*, scaled coefficient for seasonal wave; *B*, scaled coefficient for mid-term flow anomaly; *C*, scaled coefficient for short-term flow anomaly.

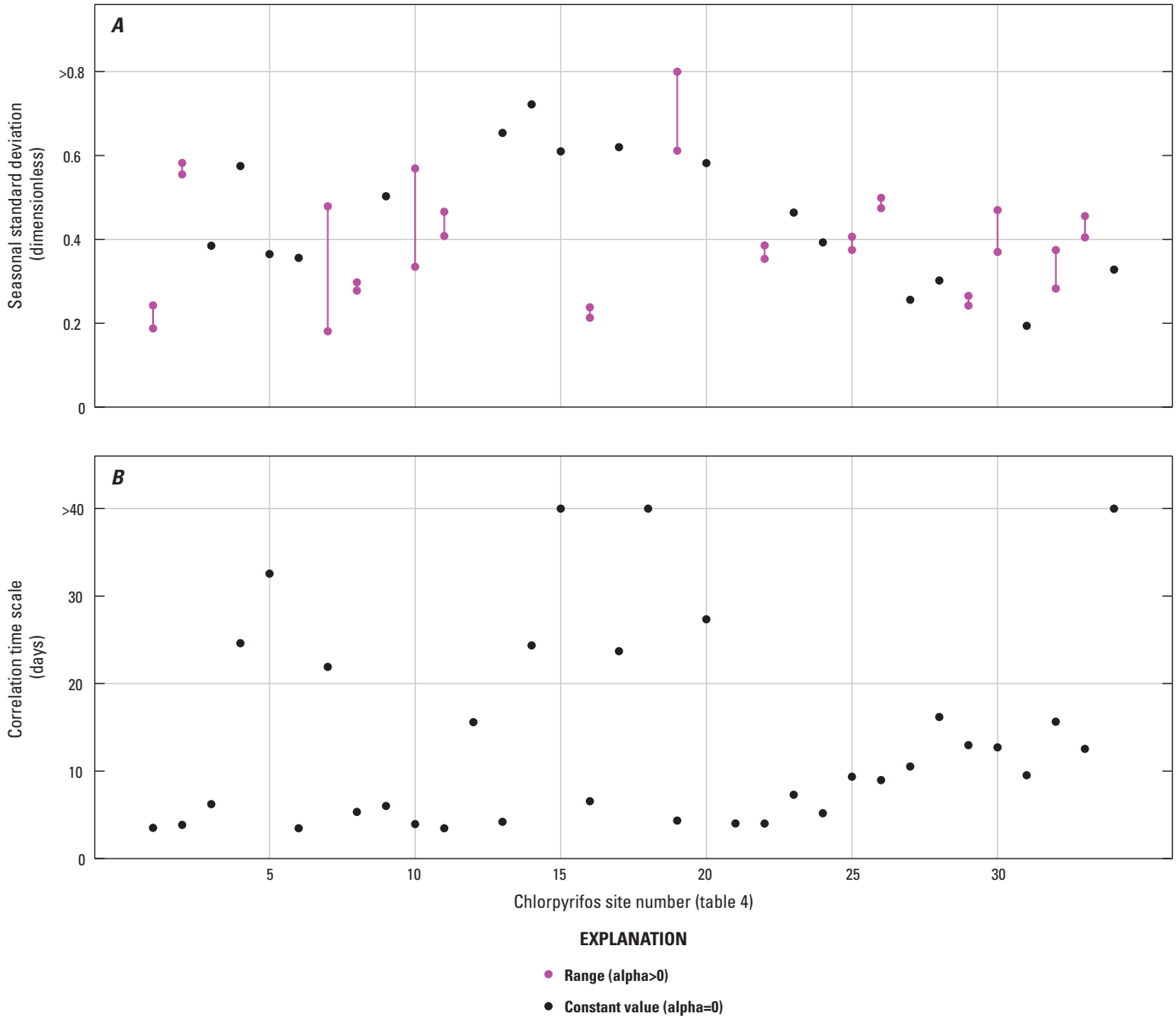

Figure 20. Estimated seasonal standard deviations and correlation time scales from SEAWAVE–QEX model results for chlorpyrifos. *A*, seasonal standard deviation; *B*, correlation time scale.

Table 5. U.S. Geological Survey fipronil water-quality sampling sites, period of record, and number of observations used for application of SEAWAVE–QEX model.

[USGS, U.S. Geological Survey; km², square kilometers; NOBS, number of observations; NUC, number of uncensored observations; PUC, percent uncensored observations]

Site number	USGS station number	Site name	Drainage area, in km²	Starting year	Ending year	Record length, in years	NOBS	NUC	PUC
1	01374987	Kisco River below Mount Kisco, New York	46	2005	2007	3	61	11	18
2	01403900	Bound Brook at Middlesex, New Jersey	132	2004	2011	8	42	19	45
3	02087580	Swift Creek near Apex, North Carolina	56	2003	2011	9	103	54	52
4	02089500	Neuse River at Kinston, North Carolina	7,020	2005	2012	8	117	57	49
5	02091500	Contentnea Creek at Hookerton, North Carolina	1,897	2005	2008	4	33	20	61
6	02169570	Gills Creek at Columbia, South Carolina	164	2005	2010	6	41	17	41
7	02318500	Withlacoochee River at U.S. 84, near Quitman, Georgia	3,872	2005	2012	8	36	18	50
8	02335870	Sope Creek near Marietta, Georgia	86	2003	2012	10	127	59	46
9	02338000	Chattahoochee River near Whitesburg, Georgia	6,252	2003	2012	10	144	69	48
10	0242354750	Cahaba Valley Creek at Cross Creek Road at Pelham, Alabama	71	2005	2012	8	83	44	53
11	03303280	Ohio River at Cannelton Dam at Cannelton, Indiana	249,355	2008	2011	4	48	14	29
12	04186500	Auglaize River near Fort Jennings, Ohio	861	2005	2008	4	51	27	53
13	04193500	Maumee River at Waterville, Ohio	16,273	2003	2006	4	39	18	46
14	05465500	Iowa River at Wapello, Iowa	32,341	2006	2012	7	106	17	16
15	05490500	Des Moines River at Keosauqua, Iowa	36,310	2008	2011	4	60	13	22
16	05531500	Salt Creek at Western Springs, Illinois	306	2003	2011	9	70	34	49
17	05572000	Sangamon River at Monticello, Illinois	1447	2005	2012	8	75	16	21
18	05586100	Illinois River at Valley City, Illinois	68,924	2005	2012	8	78	35	45
19	05587455	Mississippi River below Grafton, Illinois	444,073	2007	2011	5	60	12	20
20	06485500	Big Sioux River at Akron, Iowa	17,814	2004	2009	6	70	13	19
21	06805500	Platte River at Louisville, Nebraska	12	2003	2009	7	75	18	24
22	06934500	Missouri River at Hermann, Missouri	1,277,917	2003	2009	7	93	14	15
23	07022000	Mississippi River at Thebes, Illinois	1,769,764	2007	2009	3	38	10	26
24	07288650	Bogue Phalia near Leland, Mississippi	1,439	2003	2008	6	73	14	19
25	07288955	Yazoo River below Steele Bayou near Long Lake, Mississippi	34,742	2005	2009	5	56	10	18
26	07381590	Wax Lake Outlet at Calumet, Louisiana	6,296	2007	2009	3	35	11	31
27	08057200	White Rock Creek at Greenville Avenue, Dallas, Texas	190	2003	2011	9	97	60	62
28	08057410	Trinity River below Dallas, Texas	16,253	2003	2012	10	143	65	45
29	08178800	Salado Creek at Loop 13, San Antonio, Texas	490	2005	2012	8	67	18	27
30	11074000	Santa Ana River below Prado Dam, California	5,680	2005	2012	8	117	44	38
31	11447360	Arcade Creek near Del Paso Heights, California	98	2003	2012	10	73	62	85
33	14211720	Willamette River at Portland, Oregon	28,921	2006	2008	3	46	12	26

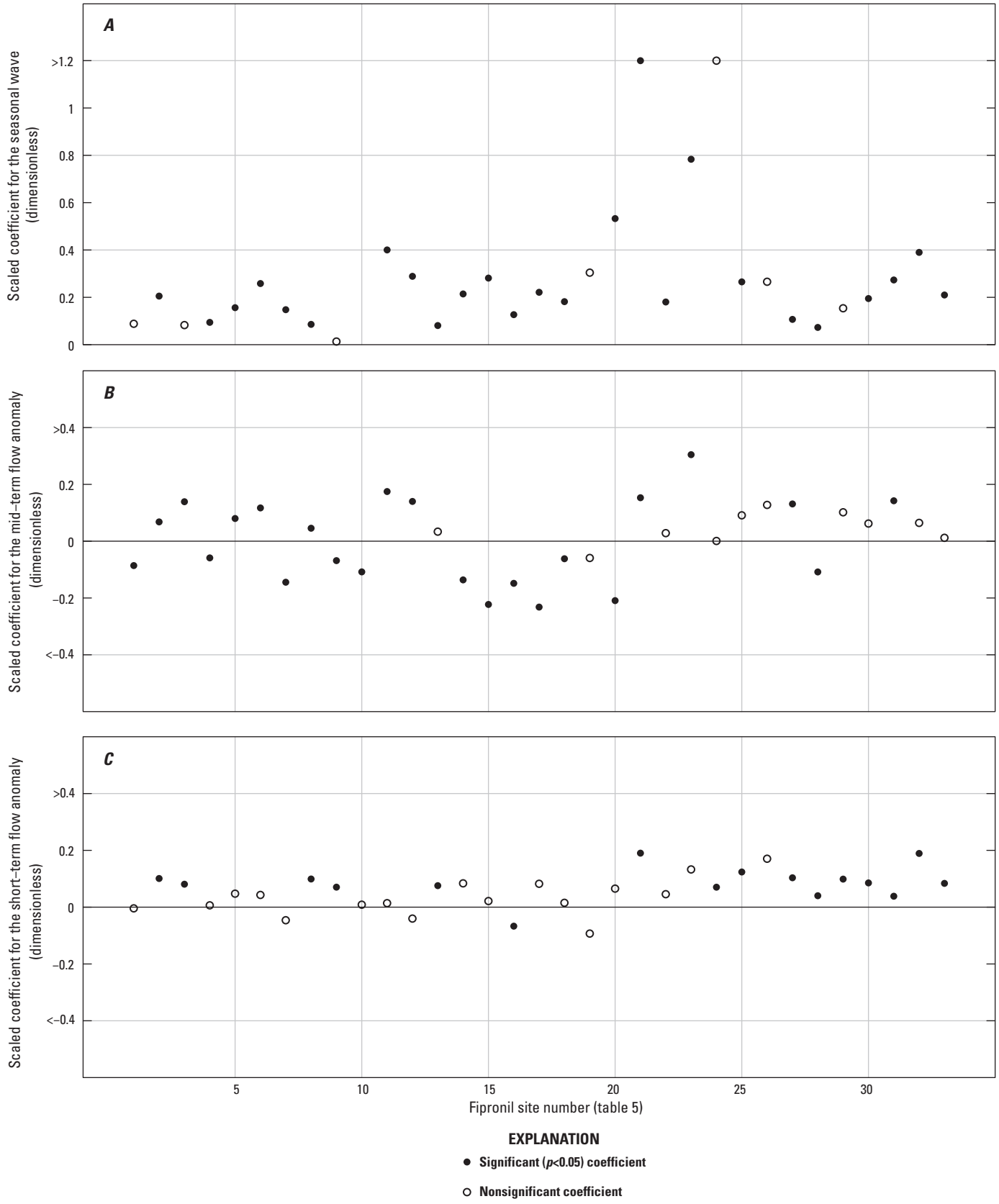

Figure 21. Estimated regression coefficients from SEAWAVE–QEX model results for fipronil. *A*, scaled coefficient for seasonal wave; *B,* scaled coefficient for mid-term flow anomaly; *C*, scaled coefficient for short-term flow anomaly.

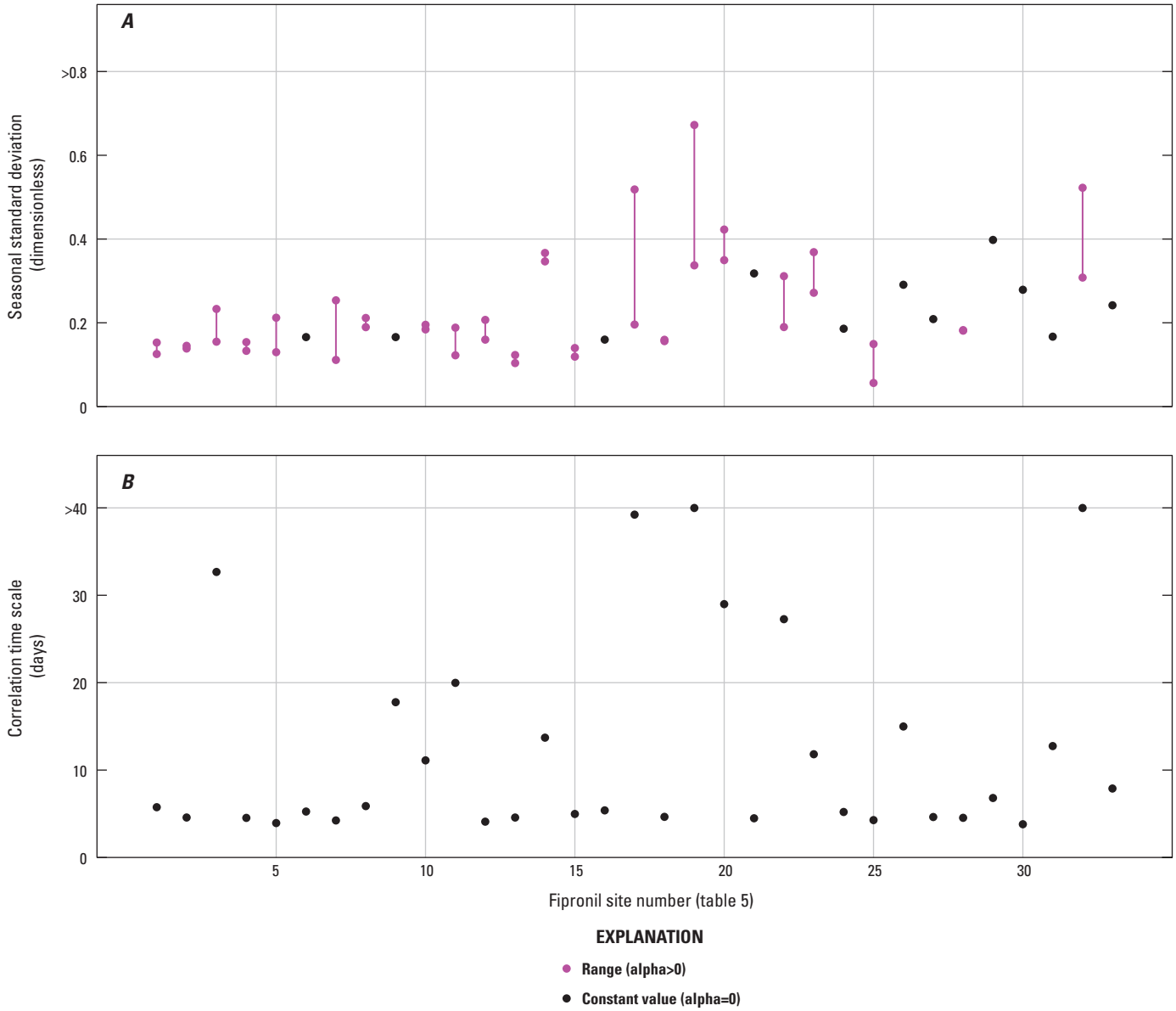

Figure 22. Estimated seasonal standard deviations and correlation time scales from SEAWAVE–QEX model results for fipronil. *A*, seasonal standard deviation; *B*, correlation time scale.

Summary and Conclusions

To accurately characterize extreme pesticide concentrations in streams, daily sampling may be necessary during active pesticide runoff periods. However, the cost of sample collection and analysis prohibits this high sampling frequency for most monitoring programs. Sampling frequencies of every 4 days, 10 times per month, or more frequently also are uncommon among programs monitoring pesticide concentrations in streams. Sampling frequencies of weekly to monthly are more common and are referred to in this report as sparse monitoring data.

A new model methodology was developed for using sparse and potentially highly censored pesticide monitoring data to estimate pesticide concentration extremes, such as the annual maximum daily concentration. The new methodology is based on a statistical model, called SEAWAVE–QEX (seasonal wave with streamflow adjustment and extended capability to produce simulated daily concentrations). The SEAWAVE–QEX model expresses log-transformed daily pesticide concentration in terms of a seasonal wave, flow-related variability, long-term trend, and serially correlated errors. The seasonal wave models seasonality in pesticide concentration because of site-specific timing and duration of the pesticide application season. Flow-related variability is modeled using two variables that are called mid-term and short-term flow anomalies, and the variables are computed using antecedent daily discharge. The model errors are assumed to have seasonal standard deviation that can increase with increasing pesticide concentration and serial correlation that is modeled using an exponential correlation function. The SEAWAVE–QEX model can be used to simulate daily concentrations that are equal to (for days with uncensored observations) or less than (for days with censored observations) the monitoring data and that consist of randomly generated values for days with no observations. The simulated daily concentrations reproduce the statistical time series characteristics of actual daily pesticide concentrations, such as serial correlation, seasonal means and variances, and flow-related variability. The simulated daily concentrations are called conditional simulations, because the simulated concentrations are conditioned on the observed monitoring data. The conditional simulations can be used to estimate concentration extremes, such as the annual maximum daily concentration, or to estimate bias factors.

This report describes the SEAWAVE–QEX modeling methodology, model testing, and data requirements. Self-contained R functions for fitting the model parameters and generating conditional simulations of daily concentrations are provided in an accompanying model archive (appendix). The model can be applied to datasets with as few as 3 years of record, as few as 30 total observations, and as few as 10 uncensored observations. Model testing indicated that, provided the model assumptions are verified, estimated annual maximum daily concentrations produced by the model should be relatively unbiased (bias between -10 and 25 percent) for low censoring rates (less than 30 percent), but can have moderate upward bias (between about 25 and 50 percent) for higher censoring rates. The estimates can have high uncertainty, especially when the censoring rate is high and the spacing between observations is large in relation to the correlation time scale of the model errors. However, provided estimation uncertainty is quantified (for example, using confidence intervals instead of point estimates), the estimates should be useful for evaluating pesticide exposure risk and uncertainty. The model was applied to atrazine, carbaryl, chlorpyrifos, and fipronil data from the U.S. Geological Survey (USGS) National Water Quality Network (NWQN). A total of 112 sites were analyzed for atrazine, 38 for carbaryl, 34 for chlorpyrifos, and 33 for fipronil. Uncertainty in the estimated concentration extremes computed using conditional simulations from the SEAWAVE–QEX model increases as the standard deviation of the model error increases or the correlation time scale of the model error decreases. Based on the parameter estimation results for the sites analyzed, uncertainty generally should be highest for carbaryl, lowest for fipronil, and mixed for atrazine and chlorpyrifos.

References Cited

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, accessed October 24, 2017, at http://onlinelibrary.wiley.com/doi/10.1111/j.1752-1688.2004.tb01045.x/abstract.

Cressie, N.C., 1991, Statistics for spatial data: New York, Wiley, 900 p.

Johnson, H.M., Domagalski, J.L., and Saleh, D.K., 2011, Trends in pesticide concentrations in streams of the western United States, 1993–2005: Journal of the American Water Resources Association, v. 47, no. 2, p. 265–286, accessed October 24, 2017, at http://onlinelibrary.wiley.com/doi/10.1111/j.1752-1688.2010.00507.x/full.

Kalkhoff, S.J., Vecchia, A.V., Capel, P.D., and Meyer, M.J., 2012, Eleven-year trend in acetanilide pesticide degradates in the Iowa River, Iowa: Journal of Environmental Quality, v. 41, no. 5, p. 1566–1579, accessed October 24, 2017, at https://dl.sciencesocieties.org/publications/jeq/abstracts/41/5/1566.

Lerch, R.N., Sadler, E.J., Sudduth, K.A., Baffaut, C., and Kitchen, N.R., 2011, Herbicide transport in Goodwater Creek experimental watershed—I. Long-term research on atrazine: Journal of the American Water Resources Association, v. 47, no. 2, p. 209–223, accessed October 24, 2017, at http://onlinelibrary.wiley.com/doi/10.1111/j.1752-1688.2010.00503.x/abstract.

Martin, J.D., 2009, Sources and preparation of data for assessing trends in concentrations of pesticides in streams of the United States, 1992–2006: U.S. Geological Survey Scientific Investigations Report 2009–5062, 41 p, accessed October 24, 2017, at https://pubs.usgs.gov/sir/2009/5062.

Mosquin, P., Whitmore, R.W., and Chen, W., 2012, Estimation of upper centile concentrations using historical atrazine monitoring data from community water systems: Journal of Environmental Quality, v. 41, no. 3, p. 834–844, accessed October 24, 2017, at https://www.ncbi.nlm.nih.gov/pubmed/22565265.

R Core Team, 2016. R—A language and environment for statistical computing: R Foundation for Statistical Computing, Vienna, Austria, accessed July 31, 2017, at https://www.R-project.org.

Rowe, G.L., Jr.; Kenneth, Belitz; Demas, C.R.; Essaid, H.I.; Gilliom, R.J.; Hamilton, P.A.; Hoos, A.B.; Lee, C.J.; Munn, M.D.; and Wolock, D.W., 2013, Design of Cycle 3 of the National Water-Quality Assessment Program, 2013–23—Part 2—Science plan for improved water-quality information and management: U.S. Geological Survey Open-File Report 2013–1160, 110 p., accessed October 24, 2017, at https://pubs.usgs.gov/of/2013/1160/.

Ryberg, K.R., and Gilliom, R.J., 2015, Trends in pesticide concentrations and use for major rivers of the United States: Science of the Total Environment, v. 538, p. 431–444, accessed October 24, 2017, at https://www.ncbi.nlm.nih.gov/pubmed/26318227.

Ryberg, K.R., Vecchia, A.V., Gilliom, R.J., and Martin, J.D., 2014, Pesticide trends in major rivers of the United States, 1992–2010: U.S. Geological Survey Scientific Investigations Report 2014–5135, 63 p., accessed October 24, 2017, at https://doi.org/10.3133/sir20145135.

Ryberg, K.R., Vecchia, A.V., Martin, J.D., and Gilliom, R.J., 2010, Trends in pesticide concentrations in urban streams in the United States, 1992–2008: U.S. Geological Survey Scientific Investigations Report 2010–5139, 42 p.

Stone, W.W., Gilliom, R.J., and Ryberg, K.R., 2014, Pesticides in U.S. streams and rivers—Occurrence and trends during 1992–2011: Environmental Science and Technology, v. 48, n. 19, p. 11025–11030, accessed October 24, 2017, at http://pubs.acs.org/doi/abs/10.1021/es5025367.

Sullivan, D.J., Vecchia, A.V., Lorenz, D.L., Gilliom, R.J., and Martin, J.D, 2009, Trends in pesticide concentrations in corn-belt streams, 1996–2006: U.S. Geological Survey Scientific Investigations Report 2009–5132, 75 p., accessed October 24, 2017, at https://pubs.usgs.gov/sir/2009/5132/.

U.S. Environmental Protection Agency, 2010a, Reevaluation of human health effects of atrazine—Review of experimental animal and in vitro studies and drinking water monitoring frequency: Federal Insecticide, Fungicide and Rodenticide Act Scientific Advisory Panel, Docket Number EPA–HQ–OPP–2010–0125–0008, April 26–29, 2010, accessed July 18, 2017, at https://archive.epa.gov/scipoly/sap/meetings/web/pdf/042610transcript.pdf.

U.S. Environmental Protection Agency, 2010b, Re-evaluation of human health effects of atrazine—Review of non-cancer effects and drinking water monitoring frequency: Federal Insecticide, Fungicide and Rodenticide Act Scientific Advisory Panel, Docket Number EPA–HQ–OPP–2010–0481, September 14–17, 2010, accessed July 18, 2017, at https://www.regulations.gov/docket?D=EPA-HQ-OPP-2010-0481.

U.S. Environmental Protection Agency, 2011, Re-evaluation of human health effects of atrazine—Review of cancer epidemiology, non-cancer experimental animal and in vitro studies and drinking water and monitoring frequency: Federal Insecticide, Fungicide and Rodenticide Act Scientific Advisory Panel, Docket Number EPA–HQ–OPP–2011–0399, accessed July 18, 2017, at http://atrazinefacts.com/docs/EPA%20Issue%20Paper%20for%20July%202011%20SAP-EPA-HQ-OPP-2011-0399-0013.pdf.

U.S. Environmental Protection Agency 2012, Problem formulation for the reassessment of ecological risks from the use of atrazine: Federal Insecticide, Fungicide and Rodenticide Act Scientific Advisory Panel, Docket Number EPA–HQ–OPP–2012–0230, September 11–14, 2012, accessed July 18, 2017, at https://www.epa.gov/sites/production/files/2015-06/documents/061212minutes.pdf.

U.S. Geological Survey, 2017, U.S. Geological Survey water data for the Nation: National Water Information System database, accessed December 14, 2017, at https://doi.org/10.5066/F7P55KJN.

Vecchia, A.V., and Williams-Sether, T., 2018, Data files to support SEAWAVE–QEX model for simulating concentrations of selected pesticides in the continental United States, 1992–2012: U.S. Geological Survey data release, https://doi.org/10.5066/F7NV9H50.

Vecchia, A.V., Martin, J.D., and Gilliom, R.J., 2008, Modeling variability and trends in pesticide concentrations in streams: Journal of the American Water Resources Association, v. 44, no. 5, p. 1308–1324.

Vecchia, A.V., Gilliom, R.J., Sullivan, D.J., Lorenz, D.L., and Martin, J.D., 2009, Trends in concentrations and use of agricultural herbicides for Corn Belt rivers, 1996–2006: Environmental Science and Technology, v. 43, no. 24, p. 9096–9102.

Appendix

Appendix. Description of R Functions and Model Archive for Running SEAWAVE–QEX.

R functions are provided for merging daily discharge and pesticide concentration data, preparing the data for analysis, fitting the seasonal wave with streamflow adjustment and extended capability (SEAWAVE–QEX) model parameters, and generating conditional simulations of daily pesticide concentrations. In many cases, the R functions complete straightforward tasks described in the main body of this report and require little explanation. Some tasks, such as maximum likelihood estimation of the seasonal variance and serial correlation parameters and conditional simulation of daily concentrations, have additional explanation. The functions are listed and described in the following section, and the code for creating the functions and instructions for running the model are provided in the "Model Archive" section of this appendix.

SEAWAVE–QEX Functions

Function: **swaveqexMerge**

Purpose: Merges daily discharge and pesticide concentration data, completes data screening steps, produces rough data plots, and creates object for input to swaveqexFit.

Required R libraries: waterData

Usage

> qexfitinput <– swaveqexMerge(cdatin, qwstnum, ddstnum, yrbeg, yrend, getdd="WD")

cdatin is a data frame with the pesticide concentration data

- the first column should be the station number (character)
- the second column should be the date, in "yyyy-mm-dd" format (character)
- the concentration value should be in a column named "final_value" (numeric)
- the remark should be in a column named "final_remark" (character)

qwstnum is the station number from cdatin to analyze (character).

ddstnum is the station number for daily discharge (character, usually the same as qwstnum). If getdd is omitted or gettdd="WD", the waterData package is used to download daily discharge for the specified U.S. Geological Survey station number (ddstnum). If getdd="File", the daily discharge data are assumed to be in a tab-delimited text file called dd_ddstnum.txt in the current working directory. The first column of the text file for daily discharge should be the date (in yyyy-mm-dd format) and the second column should be the discharge value. The text file should not have a header and there should be no missing values.

yrbeg and yrend are the beginning and ending calendar years for analysis (numeric). If unknown, the entire period of record can be analyzed by setting yrbeg=0 and yrend=3000.

Examples

qexfitinput <- swavqexMerge(SWqexAtrazineData,"03353637","03353637",1993,2002)

Prepares the atrazine data for Little Buck Creek near Indianapolis, Indiana (USGS station number 03353637 in the SWqexAtrazineData dataframe) for 1993–2002.

qexfitinput <- swavqexMerge(SWqexAtrazineData,"03353637","03353637",1993,2002,getwd="File")

looks for discharge data in a file called dd_03353637.txt in the current working directory.

Output

Rough data plots (sent to the default plot device). These plots can be used to adjust yrbeg and yrend and see if data are sufficient for analysis.

An object (list) named qexfitinput or any other user-specified name, for input to swaveqexFit

Function: **swaveqexFit**

Purpose: Uses input object prepared by swaveqexMerge to estimate the model parameters, produce diagnostic plots, and generate conditional simulations of daily concentration.

Required R libraries: tmvtnorm, survival

Other functions required (described later): swaveqexPESTpdo, swaveqexCSIM

Usage

```
> qexfitout <- swaveqexFit(qexfitinput,outfolder,ncs=50)
```

qexfitinput is an object (list) produced by swaveqexMerge.

outfolder is a character name for the folder to save the diagnostic plots and conditional simulations. The folder needs to be created ahead of time. For example, "outatrazine\\" will save the results in a folder called outatrazine in the default working directory.

ncs is the number of conditional simulations to generate (default is 50, maximum is 250).

Output

A Portable Document Format (PDF) file called "PlotsXXX.pdf", where XXX is the qwstnum used to produce qexfitinput. This file contains diagnostic plots similar to figures 5–8 of this report.

A tab-delimited text file called "CSIMSXXX.txt" (where XXX is the qwstnum) with daily output information for the period of record, including columns with the observed concentrations, daily discharges, and ncs conditional simulations. See the "Model Archive" section of this appendix for description of the CSIMXXX.txt file.

A list with three elements, saved as qexfitout or any other user specified name, with the parameter estimates and other information as follows:

qexfitout[[1]] is the station number for pesticide concentration data (character)

qexfitout[[2]] is a vector of length 25 with the output names (character)

qexfitout[[3]] is a vector of length 25 with the output values (numeric)

The output names and descriptions are as follows:

yrbeg	beginning year of record
yrend	ending year of record
rlen	record length
nobs	number of observations
nucen	number of uncensored observations
prucen	proportion of uncensored observations
int	regression intercept
cswave	regression coefficient for seasonal wave
pswave	approximate *p*-value for cswave
cmtfa	regression coefficient for mid-term flow anomaly
pmtfa	approximate *p*-value for cmtfa
cstfa	regression coefficient for short-term flow anomaly
pstfa	approximate *p*-value for cstfa
ctnd	regression coefficient for trend term
ptnd	approximate *p*-value for ctnd
wmcls	wave model class (1 or 2)
wmodno	pulse input model number (1 through 6)
hlife	modeled "half-life" (1 through 4)

wshft	phase shift
sigma	estimated error standard deviation
alph	estimated value of alpha
cts	estimated correlation time scale
n2LLIK	negative 2 times the log-likelihood value
sdmtfa	standard deviation of the mid-term flow anomaly
sdstfa	standard deviation of the short-term flow anomaly

Function: **swaveqexPESTpdo**

Purpose: Selects the best wave model and computes estimates of the regression coefficients and maximum pseudo-likelihood estimates of the seasonal standard deviation and serial correlation parameters.

Required R libraries: tmvtnorm, survival

Other functions used (described later): estsigxx, evalmodlikxx, compwaveconvxx

This function is called internally from swaveqexFit. User does not need to call this function.

Additional details: With highly censored data, exact maximum likelihood estimation is intractable. An alternative method, based on the pseudo-likelihood function is used. This method has been determined to be comparable (in terms of bias and efficiency) to exact maximum likelihood while being much simpler to compute (Besag, 1977; Zeger and Brookmeyer, 1986).

Function: **estsigxx**

Purpose: Finds iterative solution for sigma to maximize the pseudo-likelihood given values for alpha and cts. User does not need to call this function.

Function: **evalmodlikxx**

Purpose: Computes value of negative 2 times the log-pseudo-likelihood. User does not need to call this function.

Function: **compwaveconvxx**

Purpose: Computes the seasonal wave given the model class, model number, model half-life, and phase shift. Function used internally.

Function: **swaveqexCSIM**

Purpose: Computes conditional simulations of daily pesticide concentration given estimated model parameters and other information passed from swaveqexFit. Input and output are processed within swaveqexFit.

Required R libraries: tmvtnorm

Other functions required (described later): impcenvals, condsim

Function: **impcenvals**

Purpose: Imputes values for censored normalized residuals. Input and output are processed within swaveqexCSIM. Function used internally.

Required R libraries: tmvtnorm

Additional details: The imputed values for each block of consecutive censored residuals are generated at random from a truncated conditional multivariate normal distribution for the censored residuals given the closest uncensored values before and after the block. This process relies on the assumption of an exponential correlation function, for which the residuals have a first-order Markov dependence structure and, thus, only the closest uncensored values are required.

Function: **condsim**

Purpose: Computes a conditional trace for the normalized residuals given the uncensored residuals and the imputed censored residuals. Input and output are processed within swaveqexCSIM and are not important for user.

Required R libraries: tmvtnorm

Additional details: The values for each block of days in between the observed/imputed values are generated at random from a conditional multivariate normal distribution given the closest values before and after the block. This process relies on the assumption of an exponential correlation function, for which the residuals have a first-order Markov dependence structure and, thus, only the closest values before and after the block are required.

Model Archive

The following files are available for download at https://doi.org/10.3133/sir20175159.

- swaveqexFunctions_V1.R

  Text file with the code required to create the SEAWAVE–QEX functions.

- swaveQEX.Rdata

  R workspace containing the following dataframes:

  - *SWqexAtrazineData, SWqexCarbarylData, SWqexChlorpyrifosData, SWqexFipronilData*

    The atrazine, carbaryl, chlorpyrifos, and fipronil dataframes used for the applications in this report.

  - *SWqexAtrazineSites, SWqexCarbarylSites, SWqexChlorpyrifosSites, SWqexFipronilSites*

    The site lists and other information for each pesticide (see tables 2–5 of this report).

  - *SWqexAtrazinePest, SWqexCarbarylPest, SWqexChlorpyrifosPest, SWqexFipronilPest*

    The SEAWAVE–QEX parameter estimates for each pesticide/site using the period of record (yrbeg, yrend) that is specified in the site list files. See description of output for the swaveqexFit function for the variable names.

Instructions for Running SEAWAVE–QEX

A recent version of R (v.3.3.0 or later) is required, and installing Rstudio is recommended. The user libraries waterData, tmvtnorm, and survival also need to be installed.

Step 1. Open the swaveQEX.Rdata workspace containing the dataframes described previously.

Step 2. Create the SEAWAVE–QEX functions in your user environment using the source command:

```
> source("swaveqexFunctions.txt")
```

Step 3. Before running the model, attach the following required libraries:

```
> library("survival")
> library("tmvtnorm")
> library("waterData")
```

SEAWAVE–QEX should now be fully functional. Start by verifying some of the results from the report, provided in the dataframes. For example, reproduce the results for the example model for carbaryl (figs. 9 and 10 and related discussion). This site is the second site in the carbaryl site list (USGS station number 01374987; the second row of the dataframe SWqexCarbarylSites; also the second row in table 3 of the report). The data first need to be prepared using swaveqexMerge:

```
> Kisco <- swaveqexMerge(SWqexCarbarylData,"01374987","01374987",2000,2008)
```

This command will create an object named Kisco in your workspace that is ready for swaveqexFit. Before running swaveqexFit, create a folder in the default directory called CarbarylOutput or some other name for storing the output files. Also, make sure that no plotting devices are open. Then, run swaveqexFit:

```
> KiscoPest <- swaveqexFit(Kisco,"CarbarylOutput\\",ncs=100)
```

This command will create the object KiscoPest in your workspace with the SEAWAVE–QEX parameter estimates. To make the results easier to view, a dataframe can be created with the first column consisting of the parameter names and the second column consisting of the parameter estimates:

```
> KiscoPest <- data.frame(KiscoPest[[2]],KiscoPest[[3]])
```

The parameter estimates should be identical to the values in the second row of the SWqexCarbarylPest dataframe.

The two files named "Plots01374987.pdf" and "CSIM01374987.txt" also should be in the CarbarylOutput folder. Look at the plots in the PDF file. The first plot should look similar to figure 9. The points labeled as "observed concentrations" (the uncensored observations) should be identical to figure 9 because those points do not change depending on the conditional simulation. However, the conditional trace (including the points labeled "simulated censored concentrations") are randomly generated and will differ for each plot. The estimated annual maximum concentrations (which are the average of ncs=100 values) may differ slightly for each plot.

The second plot should look like figure 10 of the report. Note that the curves showing the fitted seasonal wave and +/- two seasonal standard deviations (and the solid points) should be identical to figure 10. However, the open points (corresponding to simulated censored observations) are from the conditional trace and will be different for each plot.

Now look at the CSIM01374987.txt file. The file should look similar to the following:

date	year	jday	qobs	cobs	crem	estreg	estcmu	csim1	csim2	csim3
2000-01-01	2000	1	24.0			1.11	2.14	0.20	0.10	2.38
2000-01-02	2000	2	24.0			1.11	3.48	0.22	0.41	1.05
2000-01-03	2000	3	29.0			1.28	3.11	0.58	1.08	1.90
2000-01-04	2000	4	32.0			1.37	3.23	0.91	2.88	1.85
2000-01-05	2000	5	61.0			2.22	9.26	1.34	1.47	1.27

This is a tab-delimited text file with a row for each day of the period analyzed and the following columns:

date (column 1): YYYY–MM–DD format

year (column 2): the calendar year

jday (column 3): integer day

qobs (column 4): observed daily discharge (cubic foot per second)

cobs (column 5): observed concentrations (micrograms per liter, blanks for missing values)

crem (column 6): concentration remark (< for censored value, blank for missing or uncensored values)

The remaining columns (7 through ncs+8) contain transformed model generated concentrations (TC):

TC = Round(1000C,2), where C is concentration, in micrograms per liter. TC is obtained by multiplying the model generated concentration by 1,000 and rounding to 2 decimal places. To obtain concentration, in micrograms per liter, out to 5 decimal places, divide TC by 1,000.

Note that missing values of TC are coded as numeric value -9

estreg (column 7): Fitted value of TC from the regression model

estmu (column 8): Mean of the ncs (for example, ncs=100) conditional traces

csim1, csim2, csim3, … (columns 9 through ncs+8): conditional traces of TC.

References Cited

Besag, Julian, 1977, Efficiency of Pseudoliklihood estimation for simple Gaussian fields: Biometrika, v. 64, no. 3 p. 616–618.

Zeger, S.L., and Brookmeyer, Ron, 1986, Regression analysis with censored autocorrelated data: Journal of the American Statistical Association, v. 81, no. 395, p. 722–729.

For more information about this publication, contact
Director, USGS Dakota Water Science Center, North Dakota Office
821 East Interstate Avenue
Bismarck, ND 58503
(701) 250–7400

For additional information visit https://nd.water.usgs.gov

Publishing support provided by the
Rolla Publishing Service Center

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20175159