National Water Quality Program

seawaveQ—An R Package Providing a Model and Utilities for Analyzing Trends in Chemical Concentrations in Streams with a Seasonal Wave (seawave) and Adjustment for Streamflow (Q) and Other Ancillary Variables, Version 2.0.0

Open-File Report 2020–1082

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

seawaveQ—An R Package Providing a Model and Utilities for Analyzing Trends in Chemical Concentrations in Streams with a Seasonal Wave (seawave) and Adjustment for Streamflow (Q) and Other Ancillary Variables, Version 2.0.0

By Karen R. Ryberg and Benjamin C. York

National Water Quality Program

Open-File Report 2020–1082

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

U.S. Department of the Interior
DAVID BERNHARDT, Secretary

U.S. Geological Survey
James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia: 2020

Suggested citation:
Ryberg, K.R., and York, B.C., 2020, seawaveQ—An R package providing a model and utilities for analyzing trends in chemical concentrations in streams with a seasonal wave (seawave) and adjustment for streamflow (Q) and other ancillary variables, version 2.0.0: U.S. Geological Survey Open-File Report 2020–1082, 25 p., https://doi.org/10.3133/ofr20201082.

ISSN 2331-1258 (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.

Federal, state, and local agencies have invested billions of dollars to reduce the amount of pollution entering rivers and streams that millions of Americans rely on for drinking water, recreation, and irrigation. Tracking changes in the quality of these waterways throughout multiple decades is crucial for evaluating the effectiveness of pollution control efforts and protecting the Nation's water resources into the future. This report provides the methodology to assess pesticide trends by documenting a method developed to address challenges in water-quality trend analysis specific to pesticide data. These challenges include seasonality that differs in nature from other water-quality constituents such as nutrients, complex relations between streamflow and concentration, a high percentage of concentrations less than laboratory reporting levels (nondetections), and intermittent or changing sampling frequencies. All NAWQA reports are available online at https://water.usgs.gov/nawqa/bib/.

We hope this publication will provide you with insights and information to meet your water-resource needs and will 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

Contents

Figures

Tables

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
	Flow rate	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)

International System of Units to U.S. customary units

Multiply	By	To obtain
	Area	
square kilometer (km^2)	0.3861	square mile (mi^2)
	Mass	
kilogram (kg)	2.205	pound avoirdupois (lb)

Supplemental Information

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

Abbreviations

AIC	Akaike’s information criterion
BIC	Bayesian information criterion
PDF	portable document format
R^2	coefficient of determination
RCS	restricted cubic splines
USGS	U.S. Geological Survey

seawaveQ—An R Package Providing a Model and Utilities for Analyzing Trends in Chemical Concentrations in Streams with a Seasonal Wave (seawave) and Adjustment for Streamflow (Q) and Other Ancillary Variables, Version 2.0.0

By Karen R. Ryberg and Benjamin C. York

Abstract

The **seawaveQ** R package provides functionality and help to fit a parametric regression model, SEAWAVE-Q, to pesticide concentration data from stream-water samples to assess trends. The model incorporates the strong seasonality and high degree of censoring common in pesticide data, and users can incorporate numerous ancillary variables such as streamflow anomalies. The model is fitted to pesticide data using maximum likelihood methods for censored data and is robust in terms of pesticide, stream location, and degree of censoring of the concentration data. This R package standardizes this methodology for trend analysis, documents the code, and provides help and tutorial information.

In previous investigations, the SEAWAVE-Q model assumed a linear trend across the period analyzed. For short trend periods, this assumption of a linear trend is adequate. However, as the period of record analyzed becomes longer, the assumption of linearity is problematic because of changes in pesticide regulation and use, some of which can be abrupt. In this update to the model, a restricted cubic spline option was added for long trend periods. This option allows for more flexibility in the time component of the model. Bootstrap functionality is included to determine statistical significance. Model results with the new restricted cubic spline option are compared to the linear trend option for two pesticide-site combinations.

Introduction

The concept of using a periodic solution to a conceptual storage equation with pulse input function to model seasonality in pesticide concentrations in stream water was first introduced by Vecchia and others (2008). The resulting seasonal model was shown to provide much better representation of seasonality in pesticide data than the standard sine and cosine functions used to model seasonality in other water-quality constituents such as nutrients and major ions. Vecchia and others (2008) also demonstrated the usefulness of using “streamflow anomalies” as predictor variables for pesticide concentration. Sullivan and others (2009) compared several methods for analyzing trends in pesticide concentrations for 31 sites and 11 pesticides in the Corn Belt of the United States. The methods compared included the seasonal Kendall, SEAKEN, test for nonflow-adjusted concentrations (a modified version the method described by Hirsch and Slack [1984]); a parametric regression model incorporating the seasonal model from Vecchia and others (2008) and referred to in Sullivan and others (2009) as the “SEAWAVE” model; and the seawave model with the addition of streamflow anomalies, called “SEAWAVE-Q.” The best model, in terms of maximizing the number of sites and pesticides that could be assessed and accounting for variable streamflow conditions when comparing trends for multiple sites and pesticides, was determined to be the SEAWAVE-Q model. Thus, the SEAWAVE-Q model was selected as the statistical tool for analyzing pesticide trends for corn-belt streams (Sullivan and others, 2009), analyzing pesticide trends for urban streams (Ryberg and others, 2010), analyzing pesticide trends in major river (Ryberg and others, 2014; Ryberg and Gilliom, 2015), and analyzing pesticide trends at sites across the Nation (Oelsner and others, 2017).

The SEAWAVE-Q model was specifically developed to address challenges in water-quality trend analysis specific to pesticide data, including seasonality that differs in nature from other water-quality constituents, such as nutrients; complex relations between streamflow and concentration; a high percentage of concentrations less than laboratory reporting levels (nondetections); and intermittent or changing sampling frequencies (Vecchia and others, 2008). Pesticide concentrations in surface water have a strong seasonal signal depending on the time of application. Pesticides may be applied during one or more general periods. Some pesticides used on food and feed crops are applied in the spring after crops have sprouted (postemergent), whereas other pesticides may be preemergent and applied in the spring or fall, or both. The seasonal hydrologic cycle has less effect on pesticide concentrations than on other more naturally occurring or chemically stable water-quality constituents, such as nutrients or major ions. However, streamflow-related variability does have an effect on pesticide concentrations and is modeled using streamflow anomalies (Vecchia and others, 2008; Ryberg and Vecchia, 2012), which are a way of separating streamflow variability into short (daily to seasonal) to long (1 year or more) time scales. Although nutrients may also have censored concentrations (that is, less than laboratory reporting levels), pesticides tend to have a much higher percentage of censored concentrations. SEAWAVE-Q uses maximum likelihood regression and survival analysis (Therneau and Grambsch, 2000; Therneau, 2015) and, therefore, is robust for highly censored constituents. Generally, only 10 noncensored values are needed for a trend period (if a sufficient number of censored samples and streamflow data are available; Vecchia and others, 2008; Sullivan and others, 2009; Ryberg and others, 2010; Ryberg and others, 2014; Ryberg and Gilliom, 2015; Oelsner and others, 2017). As funding and objectives change, water-quality sampling frequency may change; thus, SEAWAVE-Q has been tested and applied to simulated and observed data with changing sampling frequencies and gaps in the data (Vecchia and others, 2008).

In the interest of standardizing the model code and in response to requests from outside agencies for a software package to allow them to apply the SEAWAVE-Q model to their datasets, the R package, **seawaveQ**, was developed (Ryberg and Vecchia, 2013). The R package was named "**seawaveQ**" rather than "SEAWAVE-Q" because R package names cannot contain hyphens and generally start with lowercase letters. In addition to the original model functionality, enhancements were included in terms of plots and model output, as well as utility functions for working with chemical concentration data. These enhancements and utilities include procedures for preparing and summarizing input data, added flexibility to include other explanatory variables besides streamflow, graphical methods for assessing model fit, and plotting routines that may be used for pesticide and other chemical concentration data. The new version of **seawaveQ** documented in this report, 2.0.0, includes an option to model the trend with restricted cubic splines (RCS) and the ability to calculate pesticide loads.

The main text of this report provides a brief overview of the **seawaveQ** package, including the statistical methods used. A complete example using **seawaveQ** for trend analysis is provided in the vignette, or tutorial, in appendix 1. Additional detail for each function, including the arguments and returned values, is provided in appendix 2. Visual examples of the seasonal wave part of the model are provided in appendix 3 to help users understand how the seasonal wave fits into the model. Model comparisons of the linear trend model and variations of the RCS model are presented in appendix 4.

Description of the seawaveQ Package

The **seawaveQ** package is a collection of functions written for R (http://www.r-project.org/; R Core Team, 2018c), an open source language and a general environment for statistical computing and graphics that runs on a variety of operating systems including Linux, Mac OS, UNIX, and Windows.

The main purpose of this package is to fit the SEAWAVE-Q model to pesticide concentration data that are assumed to be expressed in micrograms per liter. The main function, fitswavecav, internally calls other functions to prepare the data, fit the model, and plot model results and regression diagnostics. The internal functions need not be called by the user, but they are documented in the R help. In addition, optional functions are provided to plot water-quality data and to combine the water-quality data with ancillary data. A flowchart showing how the input data, main function, optional functions, internal functions, and output work together to create the **seawaveQ** package is shown in figure 1 The functions in the seawaveQ package and a brief description of each function are listed in table 1.

More details on **seawaveQ** functions and tutorial examples of preparing datasets for analysis are available in appendix 1 (vignette) and appendix 2 (help documentation). Vignettes are portable document format (PDF) files that contain examples of R code and results of running the code, as well as descriptive text (R Core Team, 2018e). Vignettes can be used as tutorials for the package, and the vignette for **seawaveQ** is included in this report to familiarize users with the functions in **seawaveQ**. In addition to the vignette, the package has detailed help documentation files for each function (appendix 2). After installing the package, the help documentation may be accessed in the same manner as the help for other R functions. Help features within R are further described in the manual "An Introduction to R" (Venables and others, 2018). Additional information on the installation and administration of R and packages that extend R are available in the manual "R Installation and Administration" (https://cran.r-project.org/doc/manuals/r-release/R-admin.html, R Core Team, 2018b).

In addition to the functions, the **seawaveQ** package provides some sample datasets that are used to illustrate the format of the input datasets used by the package functions, and the datasets are used in the vignette. The sample datasets are listed in table 2. To use **seawaveQ**, an analyst is required to have concentration data and continuous ancillary data. The sample datasets IllRivValleyCty and qwMoRivOmaha provide examples of the input data format. The unit of concentration is assumed to be micrograms per liter. Continuous ancillary data can be dimensionless streamflow or sediment anomalies, as in the example dataset cqwMoRivOmaha, computed using the R package **waterData** (Ryberg and Vecchia, 2012), or the ancillary data can be anomalies based on other continuously monitored data, such as specific conductance or temperature. See Vecchia and others (2008) for more information on anomalies. Users that need additional help getting their own data into R should consult the R manual titled "R Data Import/Export" (R Core Team, 2018a).

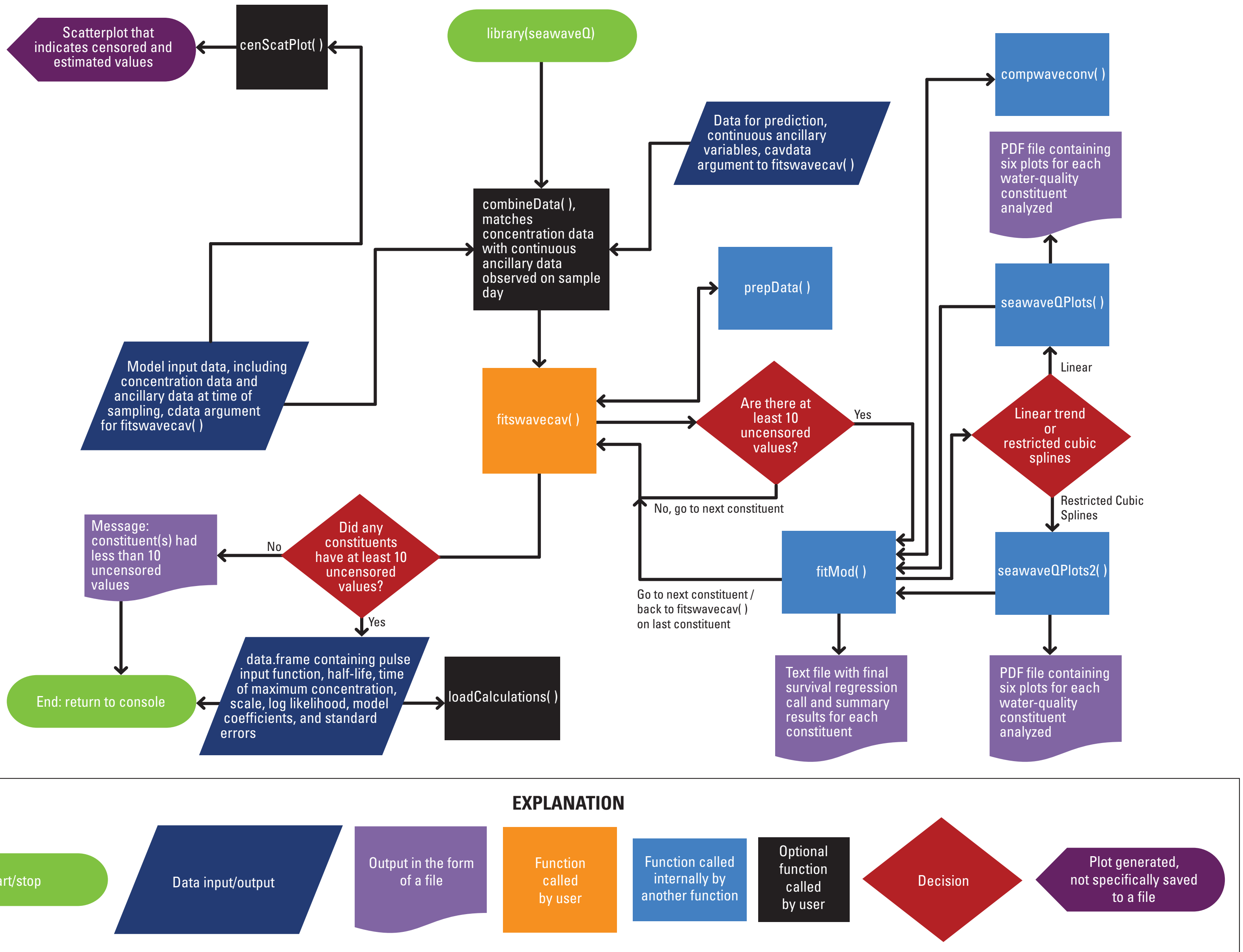

Figure 1. Flowchart showing the input, functions, and output that are part of the seawaveQ package.

Table 1. List of functions in the seawaveQ package and short description of each function.

[PDF, portable document format]

Function	Description	Use	Input, see appendix 2 for more detail	Output, see appendix 2 for more detail
cenScatPlot	Creates scatterplots of water-quality data and indicates which values are censored or estimated.	Optional	User supplied	A scatterplot.
combineData	Combines water-quality sample data with other explanatory variables.	Optional	User supplied	A data frame.
compwaveconv	An internal function (usually called from within fitMod and seawaveQPlots but can be invoked directly) that computes the seasonal wave of the SEAWAVE-Q model.	Internal	Objects passed to it by the fitMod function	A numeric vector.
fitMod	An internal function (usually called from within fitMod but can be invoked directly) that fits the SEAWAVE-Q model.	Internal	Objects passed to it by the fitswavecav function	A PDF file, a text file, and a list.
fitswavecav	The main (wrapper) function that calls other functions to prepare the data, fit the model, and produce plots.	Required	User supplied	A PDF file, a text file, and a list.
pesticideTrendCalcs	Internal function to summarize the trend results.	Internal	Objects passed to it by the fitswavecav function	A data frame.
prepData	An internal function (usually called from within fitswavecav, but can be invoked directly) that prepares the data for subsequent fitting of the SEAWAVE-Q model.	Internal	Objects passed to it by the fitswavecav function	A list.
seawaveQPlots	An internal function (usually called from within fitMod but can be invoked directly) that generates plots based on the data and the fitted SEAWAVE-Q linear trend model.	Internal	Objects passed to it by the fitMod function	A PDF file.
seawaveQPlots2	An internal function (usually called from within fitMod but can be invoked directly) that generates plots based on the data and the fitted SEAWAVE-Q model with restricted cubic splines.	Internal	Objects passed to it by the fitMod function	A PDF file.

Table 2. List of sample datasets in the seawaveQ package and short description of each.

[USGS, U.S. Geological Survey]

Dataset	Description, see appendix 2 for more detail
cqwMoRivOmaha	Continuously monitored streamflow and sediment for USGS streamgage 06610000, Missouri River at Omaha, Nebraska, data provided in 2011, by Patrick Phillips, USGS, New York Water Science Center.
IllRivValleyCty	Pesticide concentration data for USGS streamgage 05586100, Illinois River at Valley City, Illinois, data provided in 2011, by Patrick Phillips, USGS, New York Water Science Center.
qwMoRivOmaha	Pesticide concentration data for USGS streamgage 06610000, Missouri River at Omaha, Nebraska, data provided in 2011, by Patrick Phillips, USGS, New York Water Science Center.
The following datasets are provided to demonstrate how to invoke the internal functions directly. Most users will not need to use these datasets.	
examplecavdat	An example of the continuous ancillary data that is passed internally to subfunctions of fitswavecav.
examplecavmat	An example of the continuous ancillary matrix that is passed internally to subfunctions of fitswavecav.
examplecdatsub	An example of the water-quality data that are passed internally to subfunctions of fitswavecav.
examplecentmp	An example of a logical vector that is passed internally to subfunctions of fitswavecav. The vector indicates which water-quality concentrations are censored.
exampleclog	An example of a numeric vector that is used internally by fitMod and passed to its subfunction seawaveQPlots. The vector represents the base-10 logarithm of water-quality concentrations.
exampleqwcols	An example of the character vector used internally to indicate which columns represent qualification codes and which columns represent water-quality concentration data.
examplestpars	An example of data that are passed internally to subfunctions of fitswavecav.
exampletndlin	An example of a numeric vector.
exampletndlinpr	An example of a numeric vector.
exampletseas	An example of a numeric vector.
exampletseaspr	An example of a numeric vector.
exampletyr	An example of the numeric vector used to pass decimal dates for the sample data to the seawaveQPlots function.
exampletyrpr	An example of the numeric vector used to pass decimal dates for the continuous ancillary variables to the seawaveQPlots function.

Statistical Methodology of Original Model

The SEAWAVE-Q model is a parametric regression model specifically designed for analyzing seasonal- and flow-related variability and trends in pesticide concentrations. The model is expressed as follows,

$$\log C(t) = \beta_0 + \beta_1 W(t) + \beta_2 LTFA(t) + \beta_3 MTFA(t) + \beta_4 STFA(t) + \beta_5 t + \eta(t) \quad (1)$$

where

- log denotes the base-10 logarithm;
- C is pesticide concentration, in micrograms per liter;
- t is decimal time, in years, with respect to an arbitrary time origin;
- $\beta_0, \beta_1, \ldots,$ are regression coefficients; and
- W is a seasonal wave representing periodic (seasonal) variability in concentration;
- $LTFA$ is a long-term dimensionless flow anomalies computed from daily streamflow;
- $MTFA$ is a mid-term dimensionless flow anomalies computed from daily streamflow,
- $STFA$ is a short-term dimensionless flow anomalies computed from daily streamflow; and
- $\eta(t)$ is the model error.

The seasonal wave is a dimensionless, periodic function of time with an annual cycle, like a mixture of sine and cosine functions commonly used to model seasonality in concentration data. However, the seasonal wave is better suited for modeling seasonal behavior of pesticide data than a mixture of sines and cosines. The seasonal wave is a periodic (with a period of 1 year) solution to the following differential equation (Vecchia and others, 2008):

$$\frac{d}{dt} W(t) = \lambda(t + s^*) - \varphi W(t) \quad (2)$$

where

- $\frac{d}{dt}$ is the derivative with respect to time;
- $\lambda(.)$ is a pulse input function with $\lambda(.)$ greater than zero during specified application season(s), and $\lambda(.)$ equals zero otherwise;
- s^* is a seasonal shift that determines the time at which W reaches its maximum; and
- φ is a decay rate corresponding with an approximate half-life of $12/\varphi$ months.

As in Sullivan and others (2009), the pulse input function is selected from a menu of 14 choices with either 1 or 2 distinct application seasons (when pesticides may be transported to the stream) of lengths from 1 to 6 months and the half-life is selected from 4 choices (1, 2, 3, or 4 months). The half-life is referred to as a model half-life when discussing model results to distinguish the model half-life from the chemical half-life of pesticides. Thus, 56 (14 multiplied by 4) choices for the wave function are available. As described in Sullivan and others (2009), the observed concentration data were used to select the best wave function and to estimate the seasonal shift (s^*) through a combination of graphical and maximum likelihood techniques. Appendix 3 presents the 56 different seasonal waves with two different seasonal shifts.

The dimensionless flow anomalies in equation 1 are orthogonal (uncorrelated) time series computed using daily flow records from a streamgage at the site of the pesticide sampling. The anomalies represent different scales of streamflow variability—long-term, mid-term, and short-term—and the specific scales selected can vary depending on the site or pesticide being analyzed. In Sullivan and others (2009), which focused on agricultural pesticides and included many large basins, long-term dimensionless flow anomaly represented annual streamflow variability, mid-term dimensionless flow anomaly represented monthly variability within years, and short-term dimensionless flow anomaly represented daily variability within months. In Ryberg and others (2010), which focused on urban pesticides for small basins, long-term dimensionless flow anomaly represented streamflow variability for 100-day intervals, mid-term dimensionless flow anomaly represented variability for 10-day intervals within 100-day intervals, and short-term dimensionless flow anomaly represented daily variability within 10-day intervals. The anomalies can be computed using the R package **waterData** (Ryberg and Vecchia, 2012), which has several options for the time scales used. In addition, dimensionless anomalies similar to the streamflow anomalies but calculated from other continuously monitored parameters such as temperature, specific conductance, or turbidity can be calculated and used as ancillary variables in the SEAWAVE-Q model (and can be computed in the **waterData** package).

The SEAWAVE-Q model (eq. 1) is fitted to pesticide data using maximum likelihood methods for censored data, as described in Sullivan and others (2009). The maximum likelihood estimates are computed assuming the model errors are independent and normally distributed. The vignette, appendix 1, guides the user through an example that uses anomalies based on streamflow and anomalies based on sediment.

Addition of Restricted Cubic Splines Option

The original SEAWAVE-Q model assumes a linear trend during a given trend period. Linearity is a common, useful assumption in trend analysis, but becomes problematic as trend periods become longer because of the potential for abrupt, or gradual, nonlinear changes. Examples include cancellation of some or all of the uses of a particular pesticide, phaseouts of a pesticide, or other changes in usage. As additional data are collected, trends for longer periods may be calculated, but a strict linear trend may be inappropriate for many site-pesticide combinations. Thus, RCS (transformations of the time variable) were incorporated into the SEAWAVE-Q model to allow flexibility with time.

Background on Restricted Cubic Splines

When using RCS to model the time variable in equation 1, the range of values of the time variable is split up, with "knots" defining the end of one segment and the start of the next (fig. 2). Separate curves are then fit to each segment (Harrell, 2010, 2016). Overall, the splines are defined so that the resulting fitted curve is smooth and continuous, passing through each knot. Other studies have indicated that 3 to 5 knots typically are sufficient for most models (Korn and Graubard, 1999). For small sample sizes (less than 30, a sample size insufficient for a pesticide trend model), 3 knots can be used; 4 knots are sufficient for most models and represent a compromise between overfitting and model flexibility (Harrell, 2010; Croxford, 2016). If the sample size is large (greater than 100, which is often the case for pesticide trend modeling) or the relation changes quickly during the trend period, 5 to 7 knots could be used, although 5 knots are usually sufficient (Stone, 1986; Croxford, 2016).

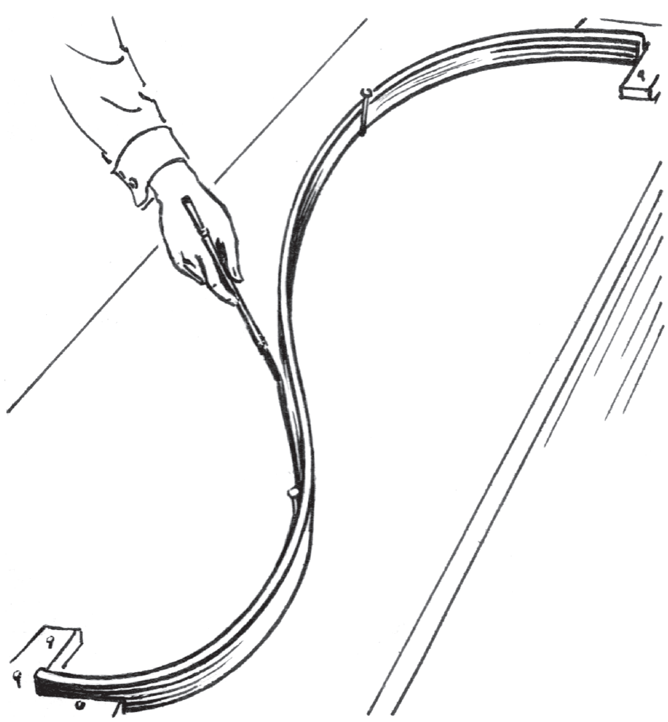

Figure 2. A spline used for drafting. The knots force the spline to pass through specific points. In the restricted cubic spline case, the ends, from the start to the first knot and the last knot to the end, are restricted to straight lines. Public domain image, Spline_(PSF).png, https://commons.wikimedia.org/w/index.php?title=File:Spline_(PSF).png&oldid=265423511, accessed July 30, 2018.

The use of RCS increases the number of parameters in equation 1, where the number of parameters is one less than the number of knots. For example, with 4 knots $\beta_5\, t$ (the time component of the model) is now described as the sum of 3 cubic splines, with coefficients β_5, β_6, and β_7. The outer line segments, from the minimum t to the first knot and from the fourth knot to the maximum t, are restricted to straight lines.

Comparison of Different Types of SEAWAVE-Q Models

Previously, several ways were available to model chemical trends using SEAWAVE-Q. The addition of RCS adds more options. The modeling options are summarized in table 3. See the definition of equation 1 for definitions of the model components other than S_k, which represents the new cubic spline components.

Sample Size Considerations

None of the variations of SEAWAVE (table 3) is a simple model. The model complexity increases with flow anomalies and with the addition of RCS. As the number of parameters in the model increases, so must the sample size. However, model complexity and sample size are not the only considerations. The degree of censoring of concentrations also affects the model's ability to converge to a solution. Ryberg and others (2014) determined that for a simple survival regression model for pesticide use, the combination of a limited number of observations (10) and censored regression resulted in several pesticide use trend models not converging. Therefore, SEAWAVE is not designed for small sample sizes. The minimum sample size is not an easy calculation, however, because the sample size depends on model complexity, the degree of censoring in the data, and whether the samples are representative of the range of pesticide concentrations with time.

In the original publication of the model, Vecchia and others (2008) used Monte Carlo simulation to show that trends were unbiased with accurate p-values for as few as 10 uncensored concentrations during a 3-year period, assuming a sampling frequency of 15 samples per year. In a study of pesticides in the U.S. Corn Belt, Vecchia and others (2009) used samples sizes from 57 to 103 during a 7-year period with a percentage of the samples censored ranging from 0 to 76 percent. Ryberg and others (2010) did not require 15 samples per year but did require that the samples be representative of the 9-year trend period studied and at least 10 uncensored values. See Ryberg and others (2010) for text and tables summarizing the number of samples and degree of censoring. Ryberg and others (2014) considered the minimum sampling criteria for a particular site during a 10-year trend period to be (1) at least 10 uncensored values, (2) at least 5 years of samples, (3) 6 or more samples in at least 2 of the first 5 years of the period, and (4) 6 or more samples in at least 2 of the last 5 years of the period. Oelsner and others (2017)

Table 3. Summary of modeling options and trend expression.

[β_0, β_1, …, are regression coefficients; *W* is a seasonal wave representing periodic (seasonal) variability in concentration; *t* is decimal time, in years, with respect to an arbitrary time origin; *LTFA* is a long-term dimensionless flow anomalies computed from daily streamflow; *MTFA* is a mid-term dimensionless flow anomalies computed from daily streamflow, *STFA* is a short-term dimensionless flow anomalies computed from daily streamflow; S_k, represents the new cubic spline components; and *k* is the number of knots]

Model terminology	Seasonal wave	Flow-related variability	Trend	Comments
SEAWAVE-Q	$\beta_0 + \beta_1 W(t)$	$\beta_2 LTFA(t) + \beta_3 MTFA(t) + \beta_4 STFA(t)$	$\beta_5 t$	This is the model described in equation 1.
SEAWAVE-Q/NoFlow	$\beta_0 + \beta_1 W(t)$	None	$\beta_2 t$	This is simpler than equation 1, so the coefficient for the time trend is renumbered as 2.
SEAWAVE-Q/ConFlow	$\beta_0 + \beta_1 W(t)$	$\beta_2 \log Q(t)$	$\beta_3 t$	This is simpler than equation 1, with streamflow used rather than a set of anomalies.
SEAWAVE-Q/RCS-k	$\beta_0 + \beta_1 W(t)$	$\beta_2 LTFA(t) + \beta_3 MTFA(t) + \beta_4 STFA(t)$	$\beta_5 S_1(t) + \beta_6 S_2(t) + \ldots + \beta_{4+k-1} S_k(t)$	With the use of restricted cubic splines, the time trend in equation 1 ($\beta_5 t$) is now described as the sum of *k*-1 cubic splines and the number of knots is *k*.

reported 10- and 20-year trend periods and for both periods, required at least 10 uncensored concentrations. For both trend periods, 50 percent of the trend period needed to have at least three quarterly samples per year, with a flexible definition of quarters. See Oelsner and others (2017) for more details on the quarterly sampling and how gaps in sampling were handled.

Trend Calculation for Model with Restricted Cubic Splines

As listed in table 3, the trend component of a model using RCS is more complex than with previous implementations of the model. Because the model now has curvilinear components, the trend can be expressed as a change in concentration from the beginning of the curvilinear trend line to the end of the curvilinear trend line (see “Model Output” section for more details). In some cases, where the addition of RCS has little change to the model, this trend may be the same as that of the linear model; however, sometimes the trend can be dramatically different because RCS can respond to abrupt or gradual changes in use or regulation that cannot be modeled linearly. In addition, the determination of statistical significance is different for the model with RCS. The following section describes how statistical significance is determined for SEAWAVE-Q/RCS-k.

Bootstrapping for Statistical Significance with Restricted Cubic Splines

To determine if a trend in pesticide concentration calculated using RCS differs, to a statistically significant degree, from a pattern produced by chance alone, bootstrap analysis functionality was added. Bootstrapping creates a distribution by resampling the sample that contains *n* observations. By resampling with replacement, the analyst can create, from a single sample, a bootstrap distribution that resembles the sampling distribution (Matthiopoulos, 2011). Because each sample is shuffled, the single sample becomes many new samples with the same number of elements as the original sample (Önöz and Bayazit, 2012). Normal bootstrapping eliminates all trends and produces trend-free data, whereas block bootstrapping preserves any autocorrelation detected in the sample if the sample is serially dependent (Önöz and Bayazit, 2012). To ensure that block bootstrapping preserves autocorrelation, blocks are often overlapped and are wrapped around from the end to the start of the series (Canty, 2002).

The bootstrap analysis in **seawaveQ** uses a block bootstrap and an attained significance level similar to the method used in Hirsch and Ryberg (2012). However, Hirsch and Ryberg (2012) used a random length block with block length geometrically distributed with a mean length of 20 years to replicate hydrologic conditions that might have climatic persistence of variable length. For **seawaveQ**, the goal was not to randomize the seasonal patterns in concentration trends, because then the significance of the seasonal wave in the model would be tested rather than the trend. Therefore, the block units are entire years. For each bootstrap replicate, the resampled time series of pesticide concentrations is regressed against the variables in the SEAWAVE-Q model. The significance calculation is based on an attained level of significance method described in Hirsch and Ryberg (2012). For iteration *j*, m_j is the estimate of the trend with time (the coefficient for *t*, eq. 1). The m_j values form a distribution of possible trend values across the bootstrap resampled dataset. The test statistic, *M*, is the estimated trend in time (difference in modeled concentration from beginning to end of trend period) calculated using the observations from the original data (data in their original order). The null hypothesis is that the expected value of *M* is zero (no trend with time). The probability value (p-value) is the fraction of the iterations in which $|m_j|$ is greater than or equal to $|M|$. If few of the trends from the randomized data (the m_j values) are larger in absolute value than the absolute value of original trend, the trend with time is statistically significant, depending on the analyst's choice of significance level. An attained p-value is reported so that a user may choose to compare the p-value to any significance level. See appendix 2 for the function arguments the produce bootstrap results.

Model Output

Model output takes several forms, which are shown in figure 1 and appendix 1 and are described in appendix 2. The following are three types of model output: (1) a list, with the first element being a data frame with information about the model and its parameters, the second element being the survival regression summary, the third element being the observed concentration (censored and uncensored), the fourth element being the concentrations predicted by the model, and the fifth element being the summary statistics for the predicted concentrations; (2) text files that provide a summary of the survival regression results, like the second element of the list but with additional measures of model quality and information about the R session; and (3) a PDF file of plots showing the model, trend, and diagnostic plots. The format of these outputs has not been changed from the previous version of **seawaveQ**, but additions have been made to the data frame within output 1 and to the text file.

List Output

A sixth element has been added to the list output that provides further trend metrics. In the original SEAWAVE-Q model (expressed as SEAWAVE-Q, SEAWAVE-Q/NoFlow, or SEAWAVE-Q/ConFlow), the trend is the coefficient of the trend component. The trend also can be expressed as a percentage in net change during the period of record or as a net change in concentration in the original units (micrograms per liter) concentration value during the period of record. These calculations were performed for a national trend assessment to make the pesticide trends results equivalent to other trend reporting results. Equations 26–31 in Oelsner and others (2017) were used to bias correct the results (eq. 26), to calculate a net change in percentage (eq. 27) and a net change in micrograms per liter (eq. 28), to calculate confidence intervals (eqs. 29 and 30), and to calculate a likelihood (eq. 31) for a likelihood-based approach to reporting trend results (Hirsch and others, 2015). See appendix 2 and Oelsner and others (2017) for additional details and see U.S. Geological Survey (2017) for an example of trends expressed in this manner in a national trend study.

In the linear trend model, an internal function called pesticideTrendCalcs (see appendix 2) produces the data frame that is returned as this new sixth element. This internal function returns the base concentration (the concentration at the beginning of the trend line) and trends and confidence intervals for concentration trends expressed as a percentage during the period of record and for concentration trends expressed in original units (micrograms per liter) during the period of record. See appendix 2 and Oelsner and others (2017) for column definitions and formulas. The output also contains a concentration trend likelihood value. The likelihood value is based on the two-sided p-value associated with the significance level of the trend and is determined as follows: likelihood = (1 - (p-value / 2)), where p-value is the p-value for the trend coefficient (Oelsner and others, 2017; Hirsch and others, 2015).

For the RCS model (SEAWAVE-Q/RCS-k), the data frame returned as this new sixth element contains columns for the modeled concentration at the beginning of the trend period and at the end of the trend period, a trend as a percentage during the period of record, a trend in original units calculated based on the percentage of increase (decrease) from the beginning to the end of the modeled trend, a p-value based on the bootstrapping routine, and a trend likelihood. See appendix 2 for additional details.

Text File Output

The model output in the text file format "text file with final survival regression call and summary results for each constituent" in figure 1, was modified to include the addition of a generalized coefficient of determination (R^2) as an additional criterion for comparing potential models. The generalized R^2 based on the likelihood-ratio test (Allison, 1995, p. 247–249). In model selection, the goal is to maximize the generalized R^2. The other model selection criterion, including Bayesian information criterion (BIC, also known as Schwarz's criterion; Schwarz, 1978) and Akaike's information criterion (AIC; Akaike, 1974), that existed in the previous version of the software were retained. The BIC takes into account the goodness of fit of the model and applies a penalty for increasing the number of parameters in a model. The BIC approximates the Bayes factor, which is a "summary of the evidence provided by the data in favor of one scientific theory, represented by a statistical model, as opposed to another" (Kass and Raftery 1995, p. 777). The AIC is an asymptotically efficient criterion and minimizes prediction error (Aho and others, 2014). In model selection using BIC or AIC, the goal is to minimize BIC and AIC. The formulas for BIC and AIC are similar, but the two criteria have different philosophies and properties (Aho and others, 2014; Vrieze, 2012); and, when the number of observations is greater than or equal to 8, the penalty for each additional parameter is larger for BIC than for AIC. Therefore, BIC can be useful when a user has many potential explanatory variables, such as when adding RCS, but wants a parsimonious model. However, in a simulation study, Vrieze (2012) indicated that AIC has less risk than BIC of selecting a bad model. Users may consult comparison studies (such as Aho and others, 2014) to determine which criterion to give the most weight in model selection.

Portable Document Format (PDF) File Output

The plots are unchanged for this version of **seawaveQ**. Some internal functionality was changed for generating the plots for models with RCS (see fig. 1 and appendix 1); however, this change does not affect how users interact with the program or the output.

Model Comparisons Using Example Sites and Pesticides

The **seawaveQ** R package was used to model trends for two pesticide-site combinations using eight cases. The cases used for the model comparisons were (1) no flow and linear trend (SEAWAVE/NoFlow), (2) daily flow and linear trend (SEAWAVE/ConFlow), (3) flow anomalies and linear trend (SEAWAVE-Q; eq. 1), and (4–8) flow anomalies and RCS with 3, 4, 5, 6, and 7 knots (SEAWAVE-Q/RCS-k). Data from 1992 to 2012 for each of the sites were used in the model comparisons and are available in a U.S. Geological Survey (USGS) data release (Ryberg and others, 2017).

Selected Sites and Pesticides

The first pesticide/site combination is alachlor for White River at Hazelton, Indiana, USGS streamgage 03374100. This site has a drainage area of 29,275 square kilometers, is 58-percent agricultural land, and is 9-percent developed land (Falcone, 2017). Alachlor is a selective herbicide for control of broadleaf weeds and grasses (Ryberg and others, 2014). According to Sullivan and others (2009), alachlor use has steadily declined since 1994 because of two main factors—the introduction of acetochlor (which widely replaced alachlor for corn) and the introduction of glyphosate-resistant soybeans (reducing the need for alachlor for soybeans). Alachlor has a low sorption to soil/sediment value, is moderately soluble in water, is volatile from water, is not persistent in aerobic soil, and is very persistent in water (Ryberg and others, 2014).

The second pesticide/site is deethylatrazine for Sope Creek near Marietta, Georgia, USGS streamgage 02335870. This site has a drainage area of 80 square kilometers is 0-percent agricultural land, and is 78-percent developed land (Falcone, 2017). Atrazine is one of the most widely used agricultural pesticides in the United States and deethylatrazine is one of atrazine’s degradation products. Deethylatrazine has a low sorption to soil/sediment value, is soluble in water, is moderately volatile from water, and is persistent in aerobic soil (Ryberg and others, 2014).

Example Output

The model output plots for each of the seven cases for the two pesticide-site combinations are presented in this section of the report. Each site has the following seven plots in figure 3: (1) no flow (SEAWAVE-Q/NoFlow); (2) with daily flow (SEAWAVE-Q/ConFlow); (3) flow anomalies (SEAWAVE-Q); and (4–7) with flow anomalies and RCS with 4, 5, 6, and 7 knots (SEAWAVE-Q/RCS-k, where k is 4, 5, 6, and 7). Additional output for these models is provided appendix 4.

The seven models compared for the White River site (USGS streamgage 03374100) are on the left-hand-side of figure 3. The top of the first page of figure 3 shows a model fitted with no streamflow and a linear trend. The next plot down shows the model fitted with daily streamflow and a linear trend. The third plot down has streamflow anomalies in the model and a linear trend. The fourth plot depicts the results of a model with streamflow anomalies and an RCS model with 4 knots. Page two of figure 3 continues on the left-hand-side with model results for the RCS with 5, 6, and 7 knots, respectively moving from top to bottom. The right-hand-side of figure 3 shows the same set of plots for the Sope Creek site (USGS streamgage 02335870).

In developing the **seawaveQ** functionality for RCS models, hundreds of model comparisons were done for numerous pesticide-site combinations (not shown). Results indicated that in most cases the RCS models represented variation in the pesticide concentrations better than the linear trend models. In addition, 4 knots were sufficient, as determined by others (Harrell, 2010; Croxford, 2016). In some instances, users may want to incorporate more knots; however, caution should be exercised because additional knots may result in multicollinearity, or a relation, between the time and streamflow components for the model (a result with undesirable consequences for regression analysis, see Helsel and others [2020] for more details on multicollinearity). Users of **seawaveQ** are encouraged to do their own comparisons to determine whether a linear or an RCS trend model is appropriate for the pesticide, length of trend period, and number of samples to be analyzed.

Load Calculation

A function was added to calculate pesticide loads, *loadCalculations*(). Daily pesticide concentration estimates provided in the output of *fitswavecav*() are used by this function and corrected for retransformation bias (the concentration model is built on the base-10 logarithm of concentration;

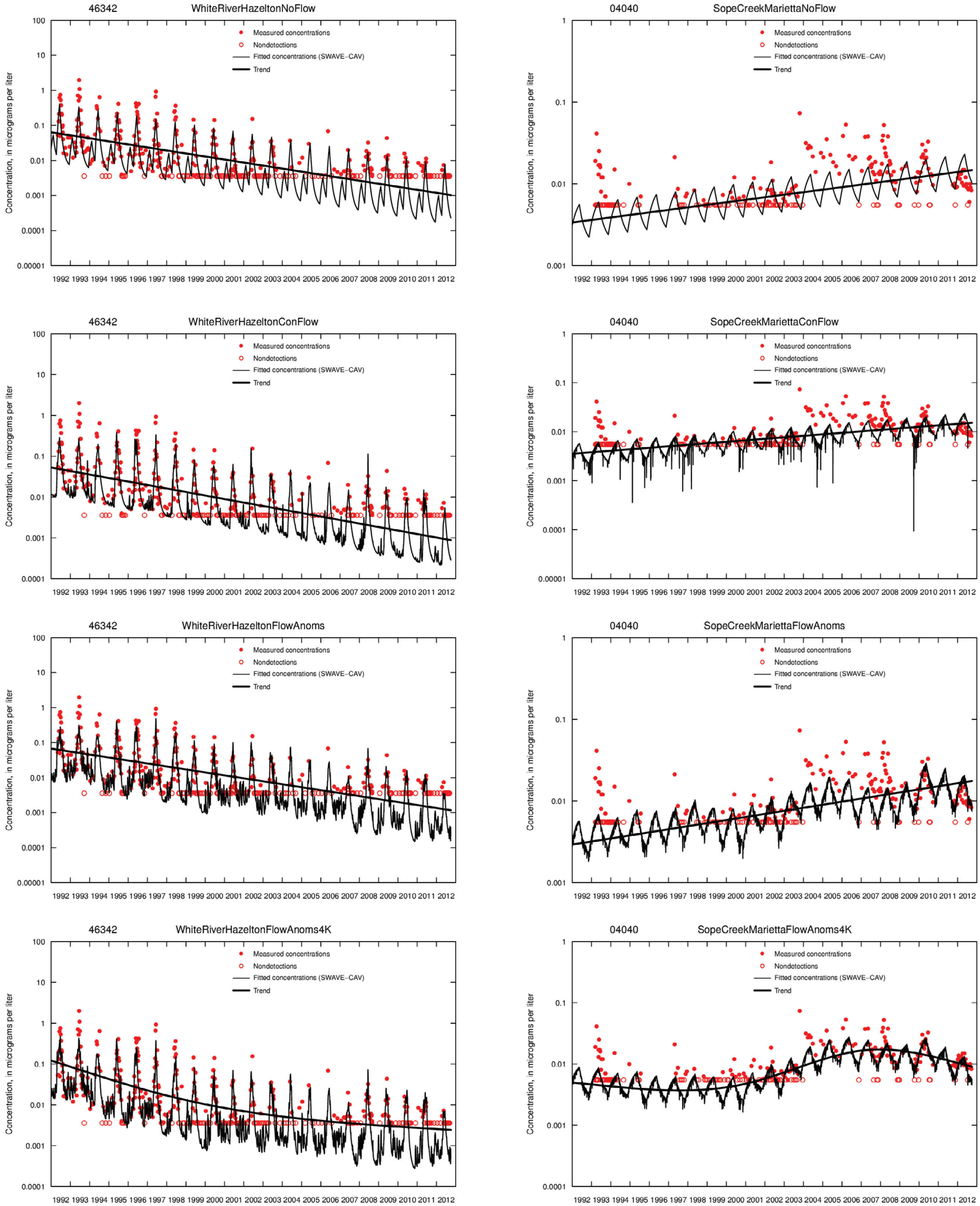

Figure 3. Model comparisons for alachlor concentrations, parameter code 46342, at White River near Hazleton, Indiana (USGS streamgage 03374100) and deethylatrazine (a degradation product of atrazine) concentrations, parameter code 04040, at Sope Creek near Marietta, Georgia (USGS streamgage 02335870). Measured concentrations from Ryberg and others (2017). This figure represents part of the R package output and users may need to modify the figure further to meet U.S. Geological Survey, or other publisher, publication standards.

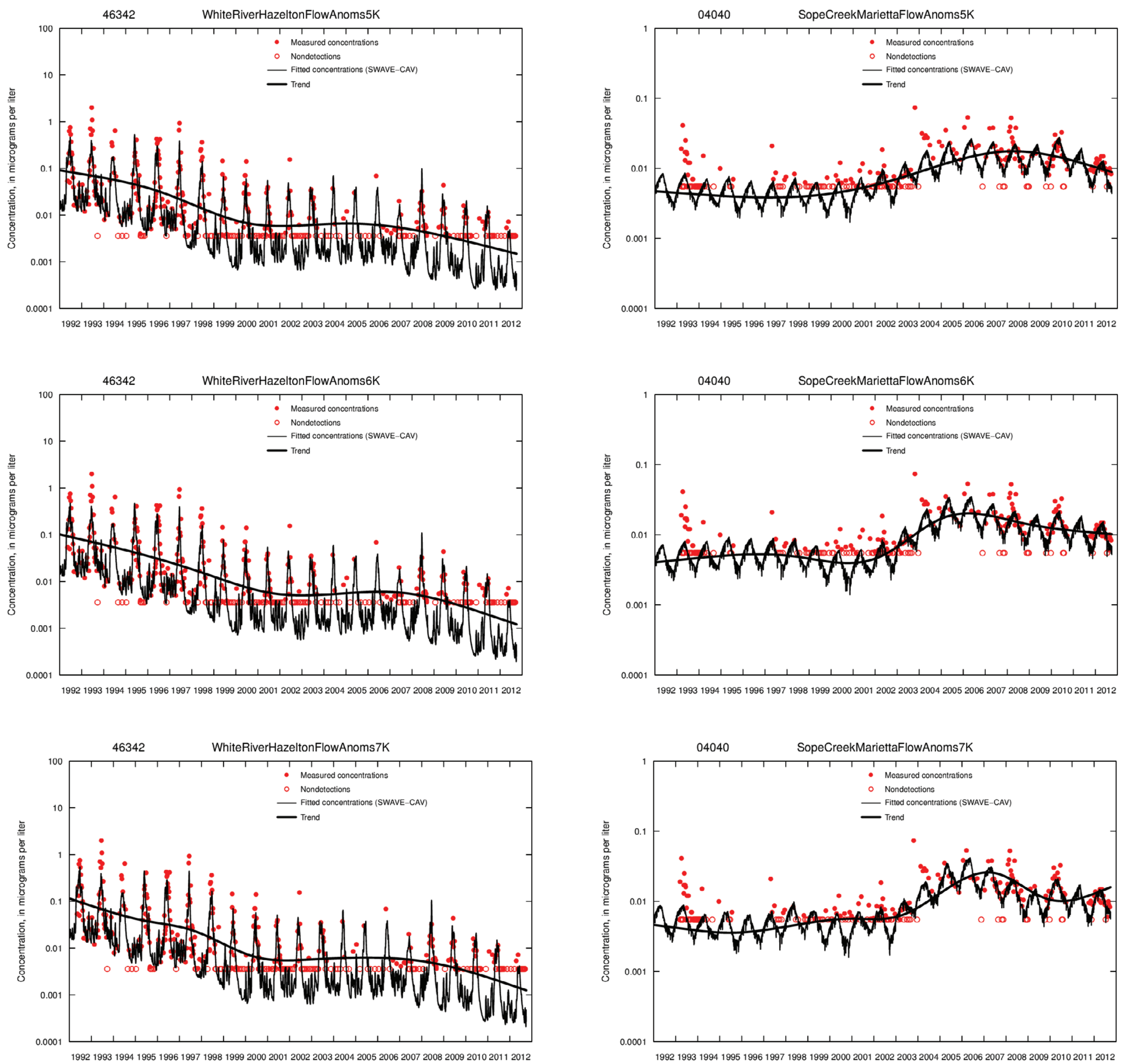

Figure 3. Model comparisons for alachlor concentrations, parameter code 46342, at White River near Hazleton, Indiana (USGS streamgage 03374100) and deethylatrazine (a degradation product of atrazine) concentrations, parameter code 04040, at Sope Creek near Marietta, Georgia (USGS streamgage 02335870). Measured concentrations from Ryberg and others (2017). This figure represents part of the R package output and users may need to modify the figure further to meet U.S. Geological Survey, or other publisher, publication standards.—Continued

therefore, a bias correction is required when transforming back to the original units) and then used to calculate daily loads. The bias correction is given by (see Oelsner and others, 2017):

$$e^{\frac{(\log(10)*scl)^2}{2}} \quad (3)$$

where

- e is the exponentiation function,
- log(10) is the natural logarithm of the number 10, and
- scl is the scale parameter from the **seawaveQ** output.

This correction, based on the quasi-maximum likelihood estimator (Cohn and others, 1989) that was developed for natural logarithms, was adjusted for the SEAWAVE-Q model, which models the base-10 logarithm of the concentration. To calculate loads, the bias corrected concentrations were multiplied by daily streamflow and a constant, 0.892998605, which converts the load units (micrograms per liter multiplied by cubic feet per second) to kilograms per day. Daily loads can be summed to annual values by the user.

Summary

The **seawaveQ** R package fits a parametric regression model, SEAWAVE-Q, to concentration data from stream-water samples to assess variability and trends. The model incorporates the strong seasonality and the high degree of censoring common in pesticide data, and users can incorporate numerous ancillary variables, such as streamflow anomalies. The model is fitted to pesticide data using maximum likelihood methods for censored data and is robust in terms of pesticide, stream location, and degree of censoring of the concentration data. The model was incorporated into an R package in 2013 to standardize the method, to formally document the code, and to provide help and tutorial information. This updated model retains the past functionality with the addition of being able to use restricted cubic splines (RCS) for a more flexible trend definition.

The **seawaveQ** R package fits a parametric regression model SEAWAVE-Q to pesticide concentration data from stream-water samples to assess variability and trends. For this study, RCS were incorporated into the SEAWAVE-Q model as an option for modeling nonlinear trends. To determine if a trend in pesticide concentration calculated using RCS differs, to a statistically significant degree, from a pattern produced by chance alone, bootstrap analysis functionality was added.

The **seawaveQ** R package was used to model seven cases for comparison for two pesticide-site combinations. The cases used for the model comparisons were (1) no flow (SEAWAVE/NoFlow), (2) with daily flow (SEAWAVE-ConFlow with a linear trend), (3) flow anomalies (SEAWAVE-Q with a linear trend), and (4–7) with flow anomalies and RCS with 4, 5, 6, and 7 knots (SEAWAVE-Q/RCS-k).

Additional functionality was added to the package to calculate pesticide loads. The package is available in the free, public Comprehensive R Archive Network, http://cran.r-project.org/ (R Core Team, 2018d). The appendixes of this document provide an example of how to use the R package and document the functions.

Disclaimer

Acknowledgments

The authors thank Charlie Crawford and Skip Vecchia for reviewing this report. The authors thank Charlie Crawford and Parker Norton for reviewing the associated software package.

References Cited

Aho, K., Derryberry, D., and Peterson, T., 2014, Model selection for ecologists—The worldviews of AIC and BIC: Ecology, v. 95, no. 3, p. 631–636, https://doi.org/10.1890/13-1452.1.

Akaike, H., 1974, A new look at the statistical model identification: IEEE Transactions on Automatic Control, v. 19, no. 6, p. 716–723, https://doi.org/10.1109/TAC.1974.1100705.

Allison, P.D., 1995, Survival analysis using the SAS system—A practical guide: Cary, N.C., SAS Institute Inc., 304 p.

Canty, A.J., 2002, Resampling methods in R—The boot package: R News, v. 2, no. 3, p. 2–7.

Cohn, T.A., DeLong, L.L., Gilroy, E.J., Hirsch, R.M., and Wells, D.K., 1989, Estimating constituent loads: Water Resources Research, v. 25, no. 5, p. 937–942, https://doi.org/10.1029/WR025i005p00937.

Croxford, R., 2016, Restricted cubic spline regression—A brief introduction: SAS Institute Inc., Paper 5621.

Falcone, J.A., 2017, Watershed characteristics for study sites of the Surface Water Trends project, National Water Quality Program: U.S. Geological Survey data release, https://doi.org/10.5066/F7TX3CKP.

Harrell, F.E., Jr., 2010, Regression modeling strategies—With applications to linear models, logistic regression, and survival analysis: New York, Springer-Verlag, 568 p.

Harrell, F.E., Jr., 2016, rms—Regression modeling strategies; R package version 5.0–1: accessed August 14, 2018, at https://cran.r-project.org/web/packages/rms/index.html.

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., and Gilroy, E.J., 2020, Statistical methods in water resources: U.S. Geological Survey Techniques and Methods, book 4, chap. A3, 458 p., https://doi.org/10.3133/tm4a3.

Hirsch, R.M., and Ryberg, K.R., 2012, Has the magnitude of floods across the USA changed with global CO_2 levels?: Hydrological Sciences Journal, v. 57, no. 1, p. 1–9, accessed September 20, 2018, at https://doi.org/10.1080/02626667.2011.621895.

Hirsch, R.M., and Slack, J.R., 1984, A nonparametric trend test for seasonal data with serial dependence: Water Resources Research, v. 20, no. 6, p. 727–732, https://doi.org/10.1029/WR020i006p00727.

Hirsch, R.M., Archfield, S.A., and De Cicco, L.A., 2015, A bootstrap method for estimating uncertainty of water quality trends: Environmental Modelling & Software, v. 73, p. 148–166, https://doi.org/10.1016/j.envsoft.2015.07.017.

Kass, R.E., and Raftery, A.E., 1995, Bayes Factors: Journal of the American Statistical Association, v. 90, no. 430, p. 773–795, https://doi.org/10.1080/01621459.1995.10476572.

Korn, E.L., and Graubard, B.I., 1999, Appendix C—Restricted cubic regression splines, *in* Korn, E., ed., L, and Graubard, B.I., Analysis of health surveys: New York, John Wiley & Sons, p. 345–346, https://doi.org/10.1002/9781118032619.

Matthiopoulos, J., 2011, How to be a Quantitative Ecologist—The 'A to R' of Green Mathematics and Statistics: West Sussex, John Wiley & Sons, Ltd, 467 p., https://doi.org/10.1002/9781119991595.

Oelsner, G.P., Sprague, L.A., Murphy, J.C., Zuellig, R.E., Johnson, H.M., Ryberg, K.R., Falcone, J.A., Stets, E.G., Vecchia, A.V., Riskin, M.L., De Cicco, L.A., Mills, T.J., and Farmer, W.H., 2017, Water-quality trends in the Nation's rivers and streams, 1972–2012—Data preparation, statistical methods, and trend results (ver. 2.0, October 2017): U.S. Geological Survey Scientific Investigations Report 2017–5006, 136 p., https://doi.org/10.3133/sir20175006.

Önöz, B., and Bayazit, M., 2012, Block bootstrap for Mann–Kendall trend test of serially dependent data: Hydrological Processes, v. 26, no. 23, p. 3552–3560, accessed September 20, 2018, at https://doi.org/10.1002/hyp.8438.

R Core Team, 2018a, R data import/export, version 3.5.1, 2018–07–02: accessed August 2, 2018, at http://cran.r-project.org/doc/manuals/R-data.pdf.

R Core Team, 2018b, R installation and administration, version 3.5.1, 2018–07–02: accessed August 2, 2018, at https://cran.r-project.org/doc/manuals/r-release/R-admin.pdf.

R Core Team, 2018c, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing, [ISBN 3–900051–07–0], accessed August 2, 2018, at http://www.R-project.org/.

R Core Team, 2018d, The comprehensive R archive network: accessed August 2, 2018, at http://cran.r-project.org/.

R Core Team, 2018e, Writing R extensions, version 3.5.1, 2018–07–02: accessed August 2, 2018, at http://cran.r-project.org/doc/manuals/R-exts.pdf.

Ryberg, K.R., and Gilliom, R.J., 2015, Trends in pesticide concentrations and use for major rivers of the United States: The Science of the Total Environment, v. 538, p. 431–444, https://doi.org/10.1016/j.scitotenv.2015.06.095.

Ryberg, K.R., and Vecchia, A.V., 2012, waterData—An R package for retrieval, analysis, and anomaly calculation of daily hydrologic time series data, version 1.0: U.S. Geological Survey Open-File Report 2012–1168, 8 p., https://pubs.usgs.gov/of/2012/1168/.

Ryberg, K.R., and Vecchia, A.V., 2013, seawaveQ—an R package providing a model and utilities for analyzing trends in chemical concentrations in streams with a seasonal wave (seawave) and adjustment for streamflow (Q) and other ancillary variables: U.S. Geological Survey Open-File Report 2013–1255, 13 p., with 3 appendixes, https://dx.doi.org/10.3133/ofr20131255.

Ryberg, K.R., Murphy, J.C., Falcone, J.A., Riskin, M.L., Wieben, C.M., and Vecchia, A.V., 2017, Pesticide concentration and streamflow datasets used to evaluate pesticide trends in the Nation's rivers and streams, 1992–2012: U.S. Geological Survey data release, https://dx.doi.org/10.5066/F7BC3WPC.

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., 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, 101 p., https://doi.org/10.3133/sir20105139.

Schwarz, G., 1978, Estimating the dimension of a model: Annals of Statistics, v. 6, no. 2, p. 461–464, https://doi.org/10.1214/aos/1176344136.

Stone, C.J., 1986, Comment—Generalized additive models: Statistical Science, v. 1, no. 3, p. 312–314, https://doi.org/10.1214/ss/1177013607.

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., https://doi.org/10.3133/sir20095132.

Therneau, T.M., 2015, Survival—A package for survival analysis in S, R package version 2.38, accessed August 14, 2018, at https://CRAN.R-project.org/package=survival.

Therneau, T.M., and Grambsch, P.M., 2000, Modeling survival data—Extending the Cox model: New York, Springer-Verlag, 350 p., https://doi.org/10.1007/978-1-4757-3294-8.

U.S. Geological Survey, 2017, Water-quality changes in the Nation's streams and rivers, accessed June 21, 2018, at https://nawqatrends.wim.usgs.gov/swtrends/.

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 & Technology, v. 43, no. 24, p. 9096–9102, accessed September 20, 2018, at https://doi.org/10.1021/es902122j.

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, accessed September 20, 2018, at https://doi.org/10.1111/j.1752-1688.2008.00225.x.

Venables, W.N., Smith, D.M., and the R Core Team, 2018, An introduction to R, version 3.5.1, 2018–07–02, accessed August 2, 2018, at http://cran.r-project.org/doc/manuals/R-intro.pdf.

Vrieze, S.I., 2012, Model selection and psychological theory—A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC): Psychological Methods, v. 17, no. 2, p. 228–243, https://doi.org/10.1037/a0027127.

Appendixes 1–4

Appendix 1. Vignette

Vignettes are the established R community method for providing examples of how to use the package. The portable document format (PDF) file can be accessed at https://doi.org/10.3133/ofr20201082.

Appendix 2. R Documentation

The official R documentation for this package can be accessed at https://doi.org/10.3133/ofr20201082.

Appendix 3. Visualizations of the Seasonal Wave

The pulse input function, jmod, is selected from a menu of 14 choices with either 1 or 2 distinct application seasons (when pesticides may be transported to the stream) of lengths from 1 to 6 months and the half-life (hlife) is selected from 4 choices (1, 2, 3, or 4 months). The half-life is referred to as a model half-life when discussing model results to distinguish the model half-life from the chemical half-life of pesticides. Thus, 56 (14 multiplied by 4) choices for the wave function are available. The observed concentration data are used to select the best wave function and to estimate the seasonal shift, the decimal season of maximum chemical concentration, cmaxt, through a combination of graphical and maximum likelihood techniques. The following plots show the 56 different seasonal waves with 2 different seasonal shifts, 0.3 (fig. 3.1) and 0.6 (fig. 3.2). The plotted seasonal waves were generated by setting the arguments of the function compwaveconv to 0.3 and 0.6 for cmaxt, 1 through 14 for jmod, 1 to 4 for hlife, and generating a plot for each combination of options, for example:

```
swave <- compwaveconv(cmaxt = 0.3, jmod = 1, hlife = 1, mclass = 1)
plot(seq(0, 1, 1 / 360), swave, typ = “l,” xaxs = “i,” yaxs = “i,”
ylim = c(-0.6, 0.6), cex.axis = 0.6, cex.lab = 0.6,
ylab = “Seasonal wave,” xlab = “Decimal seasons”)
```

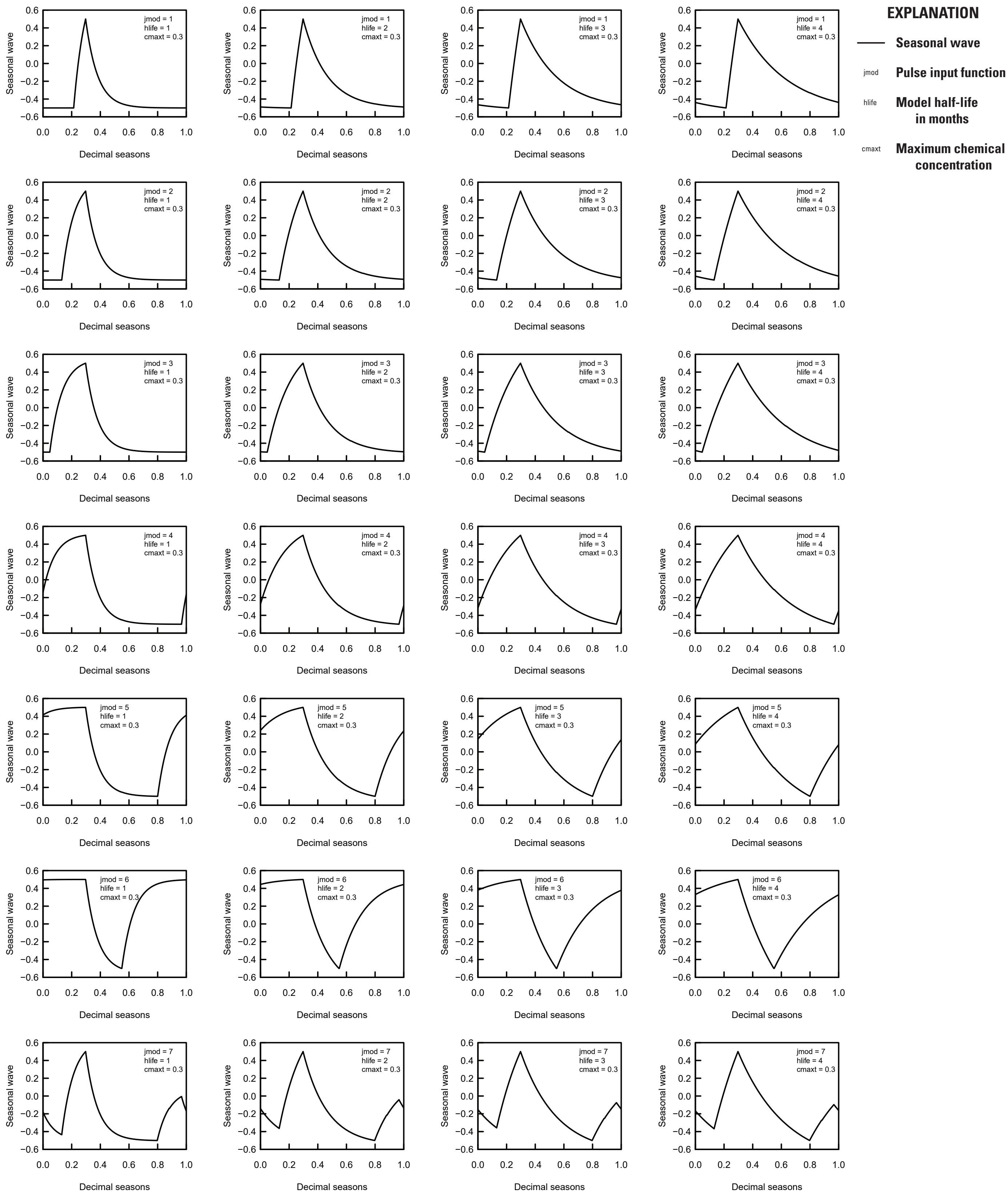

Figure 3.1. Visualization of seasonal waves with seasonal shift of the decimal season of maximum chemical concentration (cmaxt) equal to 0.3 (from Ryberg and Vecchia, 2013). jmod is the pulse input function, hlife is model half-life in months, and decimal season is the fraction of the calendar year represented by a specific month and day.

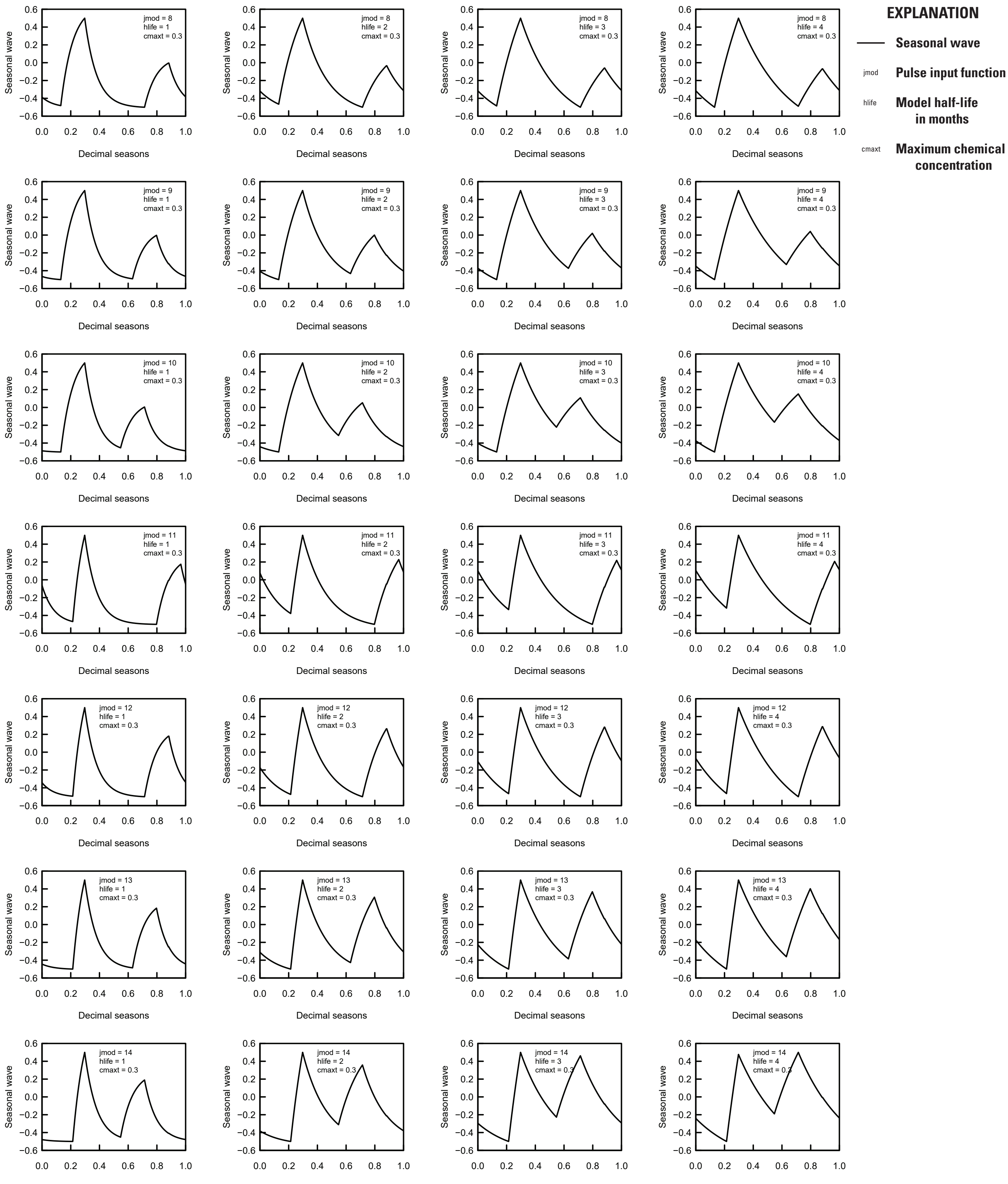

Figure 3.1. Visualization of seasonal waves with seasonal shift of the decimal season of maximum chemical concentration (cmaxt) equal to 0.3 (from Ryberg and Vecchia, 2013). jmod is the pulse input function, hlife is model half-life in months, and decimal season is the fraction of the calendar year represented by a specific month and day.—Continued

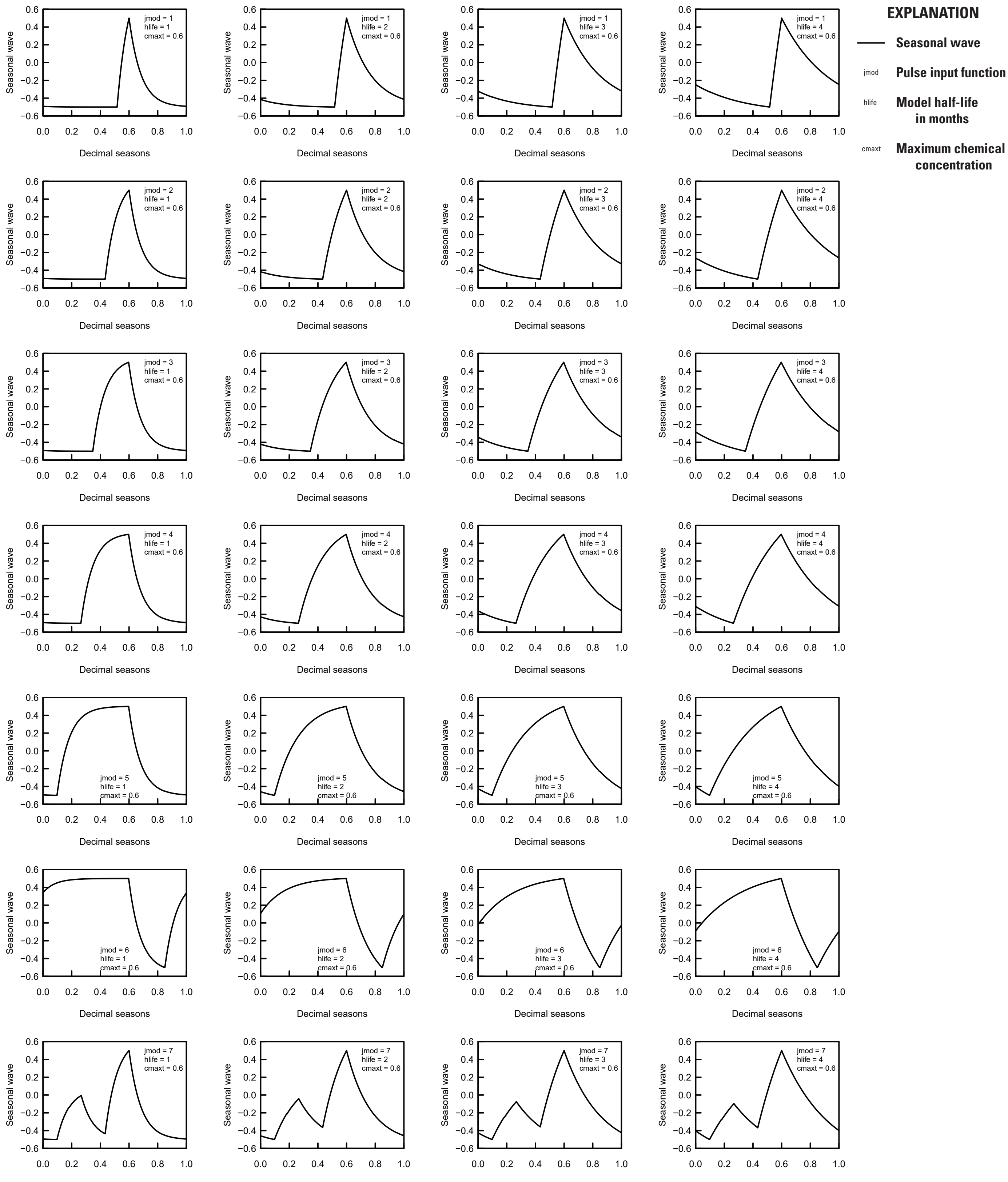

Figure 3.2. Visualization of seasonal waves with seasonal shift of the decimal season of maximum chemical concentration (cmaxt) equal to 0.6 (from Ryberg and Vecchia, 2013). jmod is the pulse input function, hlife is model half-life in months, and decimal season is the fraction of the calendar year represented by a specific month and day.

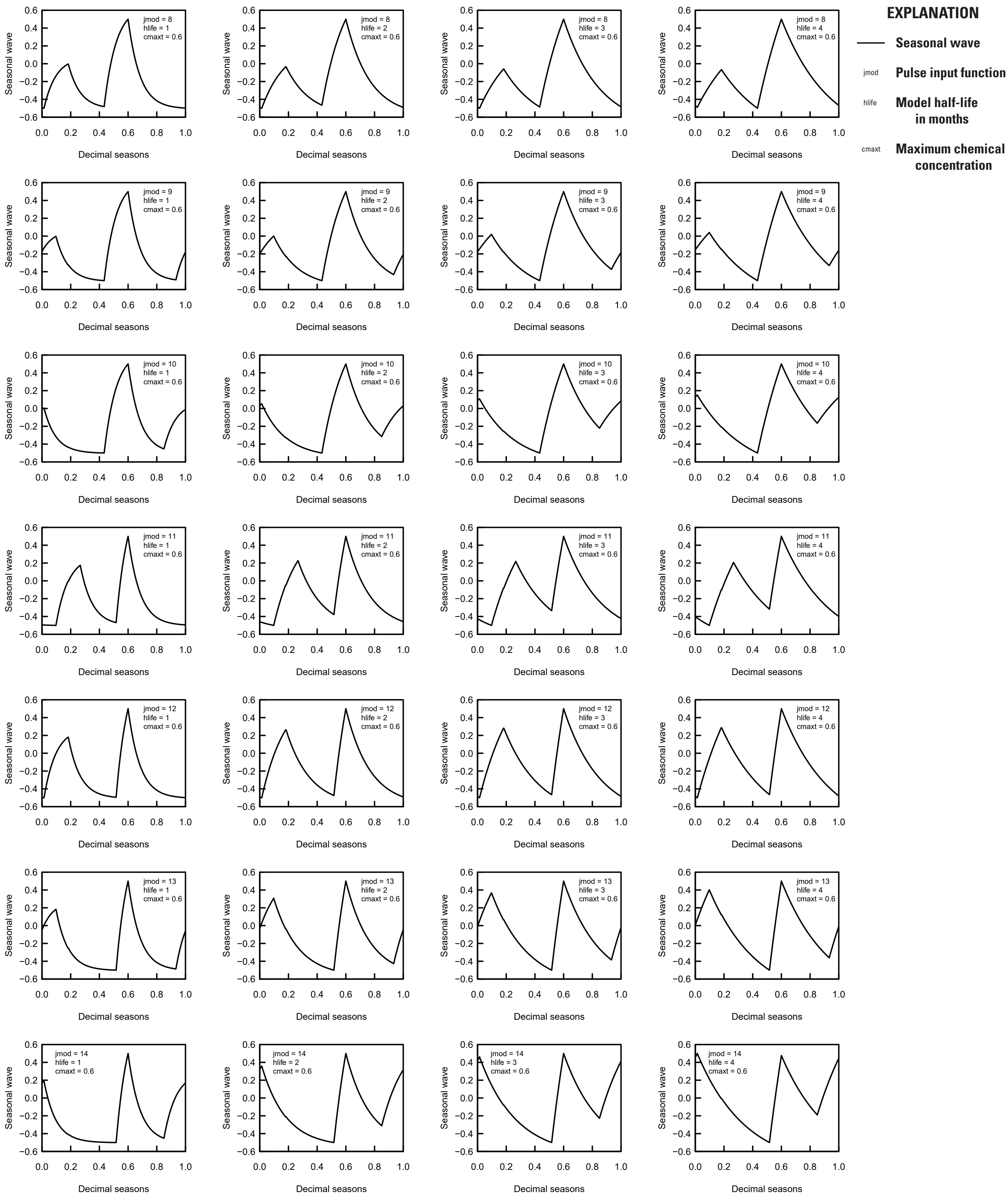

Figure 3.2. Visualization of seasonal waves with seasonal shift of the decimal season of maximum chemical concentration (cmaxt) equal to 0.6 (from Ryberg and Vecchia, 2013). jmod is the pulse input function, hlife is model half-life in months, and decimal season is the fraction of the calendar year represented by a specific month and day.—Continued

References Cited

Ryberg, K.R., and Vecchia, A.V., 2013, seawaveQ—an R package providing a model and utilities for analyzing trends in chemical concentrations in streams with a seasonal wave (seawave) and adjustment for streamflow (Q) and other ancillary variables: U.S. Geological Survey Open-File Report 2013–1255, 13 p., with 3 appendixes, https://doi.org/10.3133/ofr20131255.

Appendix 4. Model Comparisons

An R markdown document presenting the code used to create figure 3, as well as additional model output, can be accessed at https://doi.org/10.3133/ofr20201082.

For more information concerning this publication, contact:

Director, USGS Dakota Water Science Center
821 East Interstate Avenue
Bismarck, North Dakota 58503
(701) 250–7400

Or visit the Dakota Water Science Center Web site at:
https://www.usgs.gov/centers/dakota-water

Publishing support provided by:
Rolla Publishing Service Center

Ryberg and York—**seawaveQ—An R Package Providing a Model and Utilities for Analyzing Trends in Chemical Concentrations**—Open-File Report 2020–1082

ISSN 2331-1258 (online)
https://doi.org/10.3133/ofr20201082