science for a changing world

National Water-Quality Assessment 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

Open-File Report 2013–1255

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

By Karen R. Ryberg and Aldo V. Vecchia

National Water-Quality Assessment Program

Open-File Report 2013–1255

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

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2013

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment, visit http://www.usgs.gov or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit http://www.usgs.gov/pubprod

To order this and other USGS information products, visit http://store.usgs.gov

Suggested citation:
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, *http://dx.doi.org/10.3133/ofr20131255.*

2331-1258 (online)

Contents

Figures

Tables

Conversion Factors

Inch/Pound to SI

Multiply	By	To obtain
	Area	
square mile (mi^2)	259.0	hectare (ha)
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)

SI to Inch/Pound

	Area	
square kilometer (km^2)	247.1	acre
square kilometer (km^2)	0.3861	square mile (mi^2)
	Flow rate	
cubic meter per second (m^3/s)	35.31	cubic foot per second (ft^3/s)

Acknowledgments

The authors thank the time and effort of the U.S. Geological Survey personnel that reviewed this report, tested the R package and vignette, and contributed suggestions to improve the package. The U.S. Geological Survey testers and reviewers were Bill Damschen, North Dakota Water Science Center and Patrick Phillips, New York Science Center.

The authors thank Evan Newell, Department of Ecology, State of Washington, who also reviewed the report and tested the R package and vignette.

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

By Karen R. Ryberg and Aldo V. Vecchia

Abstract

The **seawaveQ** R package fits a parametric regression model (seawaveQ) to pesticide concentration data from streamwater samples to assess variability and 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, as well as providing additional utility functions for plotting pesticide and other chemical concentration data.

Introduction

Vecchia and others (2008) first introduced the concept of using a periodic solution to a conceptual storage equation with pulse input function to model seasonality in pesticide concentrations in streamwater. The resulting seasonal model (referred to in later publications as the “seawave” 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” (described later in this report) 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. The methods compared included the seasonal Kendall (SEAKEN) test for nonflow-adjusted concentrations [a modified version of the nonparametric seasonal Kendall test 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 approach 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 seawaveQ model. Thus, the seawaveQ model was selected as the statistical tool for analyzing pesticide trends for cornbelt streams (Sullivan and others, 2009), analyzing pesticide trends for urban streams (Ryberg and others, 2010), and for use in ongoing pesticide assessments. In response to requests from outside agencies for a software package to allow them to apply the seawaveQ model to their datasets, the R package, **seawaveQ**, was developed.

The model was developed to “handle a number of difficulties often found in pesticide data, such as strong seasonality in response to use patterns, high numbers of concentrations below laboratory reporting levels (RLs), complex relations between streamflow and concentration, and intermittent or changing sampling frequencies (both inter-annually and intra-annually)” (Vecchia and others, 2008). 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 (Vecchia and others, 2008).

The R package **seawaveQ** was developed to provide a standardized methodology for fitting the seawaveQ model and to make the trend analysis method widely available for public use. In addition, several enhancements to the seawaveQ model have been included, 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 body 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.

Description Of The SeawaveQ Package

The **seawaveQ** package is a collection of functions written for R (*http://www.r-project.org/; R Core Team, 2012a*), 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 seawaveQ model to pesticide concentration data. 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 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. Table 1 lists the functions in **seawaveQ** and provides a brief description of each.

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 PDF documents that contain examples of R code and results of running the code, as well as descriptive text (R Core Team, 2012c). Vignettes can be used as tutorials for the package and the vignette for **seawaveQ** is included in this document 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, 2013). Additional information on the installation and administration of R and packages that extend it are available in the manual *R Installation and*

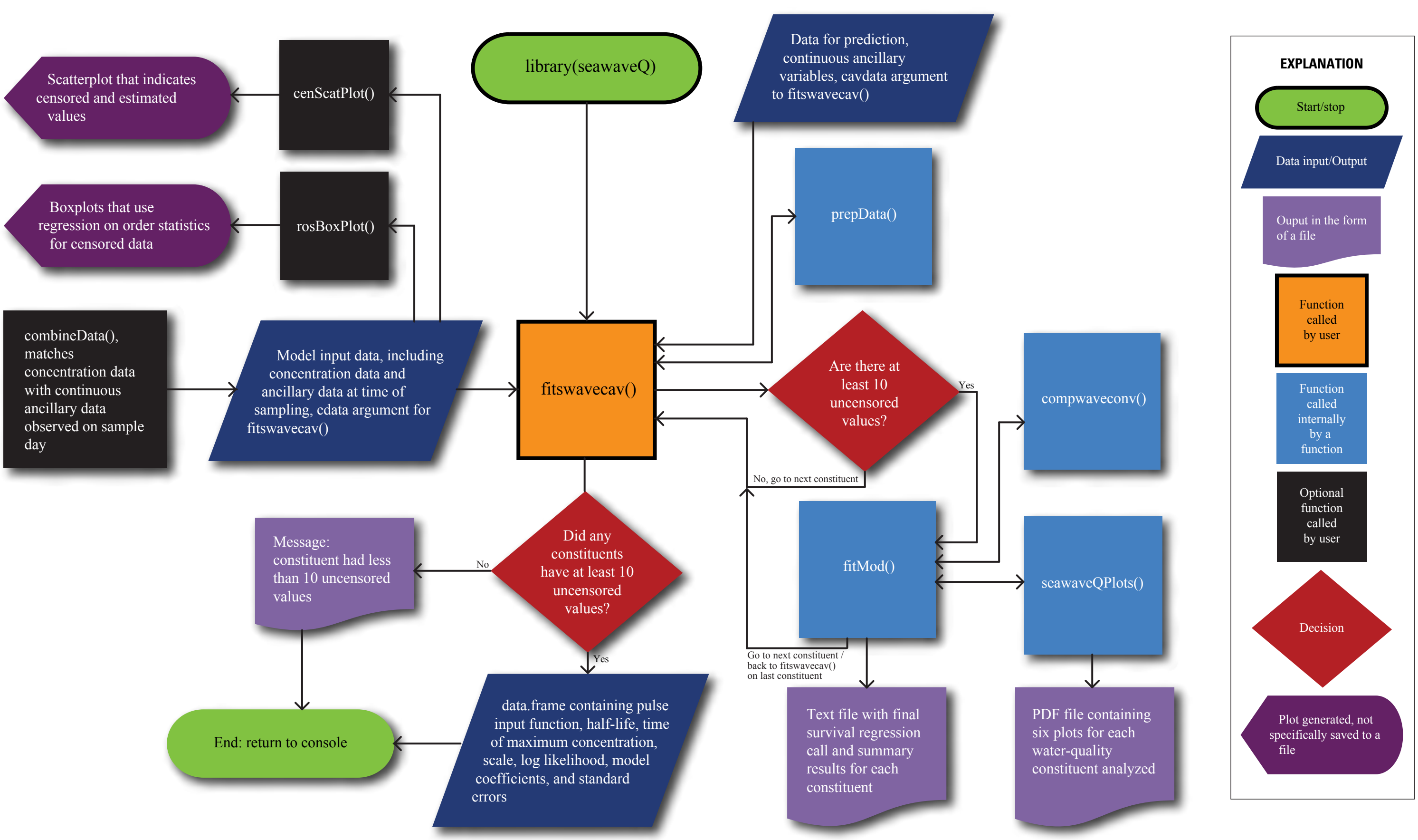

Figure 1. Flowchart showing the input, functions, and output that are part of **seawaveQ**.

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

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 expalantory 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 the seawaveQ 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 seawaveQ 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 prodcue plots.	required	user supplied	a pdf file, a text file, and a list.
prepData	An internal function (usually called from within fitswavecav, but can be invoked directly) that prepares the data for subsequent fitting of the seawaveQ model.	internal	objects passed to it by the fitswavecav function	a list.
rosBoxPlot	Creates boxplots of water-quality data that include censored values.	optional	user supplied	a boxplot.
seawaveQPlots	An internal function (usually called from within fitMod but can be invoked directly) that generates plots based on the data and the fitted seawaveQ model.	internal	objects passed to it by the fitMod function	a pdf file.

Administration (*http://streaming.stat.iastate.edu/CRAN/doc/manuals/r-release/R-admin.pdf*, R Core Team, 2012b).

Input Data

In addition to the functions, **seawaveQ** provides some sample datasets that are used to illustrate the format of the 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 necessary data format. It is assumed that the unit of concentration is 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 they 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 titleed “R Data Import/Export” (R Core Team, 2013a).

Statistical Methodology

The seawaveQ 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,

$$Log\ C(t) = \beta_0 + \beta_1\ W(t) + \beta_2\ LTFA(t) + \beta_3\ MTFA(t) + \beta_4\ STFA(t) + \beta_5\ t + \eta(t) \tag{1}$$

where

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

The seasonal wave is a dimensionless, periodic function of time with an annual cycle, similar to a mixture of sine and

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

[USGS, U.S. Geological Survey]

Data set	Description, see appendix 2 for more detail
cqwMoRivOmaha	Continuously monitored streamflow and sediment for USGS gage 06610000, Missouri River at Omaha, Nebraska.
IllRivValleyCty	Pesticide concentration data for USGS gage 05586100, Illinois River at Valley City, Illinois.
qwMoRivOmaha	Pesticide concentration data for USGS gage 06610000, Missouri River at Omaha, Nebraska.
The following data sets are provided to show how to invoke the internal functions directly. Most users will not need to use these	
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 is passed internally to subfunctions of fitswavecav.
examplecentmp	An example of a logical vector that is passed internally to subfunctions of fitswavecav. It 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. It 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 represent water-quality concentration data.
examplestpars	An example of data that is 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 a to the seawaveQ-Plots function.

cosine functions often 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):

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

where

$\lambda(.)$ is a pulse input function with $\lambda(.) > 0$ during specified application season(s) and $\lambda(.)=0$ 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, Appendix 2), 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 it from the chemical half-life of pesticides. Thus, 56 (14x4) choices for the wave function are available. As described in Sullivan and others (2009, appendix 2), 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 shows the 56 different seasonal waves with two different seasonal shifts, specified as cmaxt=0.3 and cmaxt=0.6.

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 (LTFA), mid-term (MTFA), and short-term (STFA) 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, LTFA represented annual streamflow variability, MTFA represented monthly variability within years, and STFA represented daily variability within months. In Ryberg and others (2010), which focused on urban pesticides for small basins, LTFA represented streamflow variability for 100-day intervals, MTFA represented variability for 10-day intervals within 100-day intervals, and STFA 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 seawaveQ model (and can be computed in the **waterData** package).

The seawaveQ model (equation 1) is fitted to pesticide data using maximum likelihood methods for censored data, as described in Sullivan and others (2009), using this package.

The vignette, appendix 1, guides the user through an example that uses anomalies based on streamflow and anomalies based on sediment.

Summary

The seawaveQ model fits a parametric regression model to concentration data from streamwater samples to assess variability and trends. The model incorporates the strong seasonality (through a pulse input function and a model half-life) and the high degree of censoring (through survival regression) 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 to standardize the methodology for trend analysis, to document the code, and to provide help and tutorial information.

In addition to the functionality of the original seawaveQ model, several enhancements have been included such as procedures for preparing and summarizing input data; flexibility to include explanatory variables besides streamflow; graphical methods for assessing model fit; and procedures for managing model output.

The package is available in the free, public Comprehensive R Archive Network, *http://cran.r-project.org/* (R Core Team, 2013b) The appendixes of this document provide an example of how to use the R package and document the functions.

Disclaimer

References Cited

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.

R Core Team, 2012a, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing, [ISBN 3-900051-07-0]. (Also available at *http://www.R-project.org/*.)

R Core Team, 2012b, R Installation and Administration, Version 2.15.2, 2012-10-26: accessed March 4, 2013, at *http://streaming.stat.iastate.edu/CRAN/doc/manuals/r-release/R-admin.pdf*.

R Core Team, 2012c, Writing R extensions, version 2.15.2, 2012-10-26: accessed March 4, 2013, at *http://cran.r-project.org/doc/manuals/R-exts.pdf*.

R Core Team, 2013a, R data import/export: accessed July 9, 2013, at *http://cran.r-project.org/doc/manuals/R-data.pdf*.

R Core Team, 2013b, The comprehensive R archive network; accessed March 4, 2013, at *http://cran.r-project.org/*.

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. (Also available at *http://pubs.usgs.gov/of/2012/1168/*.)

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. (Also available at *http://pubs.usgs.gov/sir/2010/5139/*.)

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. (Also available at *http://pubs.usgs.gov/sir/2009/5132/*.)

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, p. 9096–9102.

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.

Venables, W.N., Smith, D.M., and the R Core Team, 2012, An introduction to R, version 2.15.2, 2012-10-26: accessed March 3, 2013, at *http://cran.r-project.org/doc/manuals/R-intro.pdf*.

Appendixes 1–3

Appendix 1. Vignette

Vignettes are the established R community method for providing examples of how to use the package.

The pdf file can be accessed at *http://pubs.usgs.gov/of/2013/1255/Downloads/Appendix_1.pdf*

Appendix 2. R Documentation

The official R documentation for this package can be accessed at *http://pubs.usgs.gov/of/2013/1255/Downloads/Appendix_2.pdf*

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 it from the chemical half-life of pesticides. Thus, 56 (14x4) 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 and 0.6. 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”)

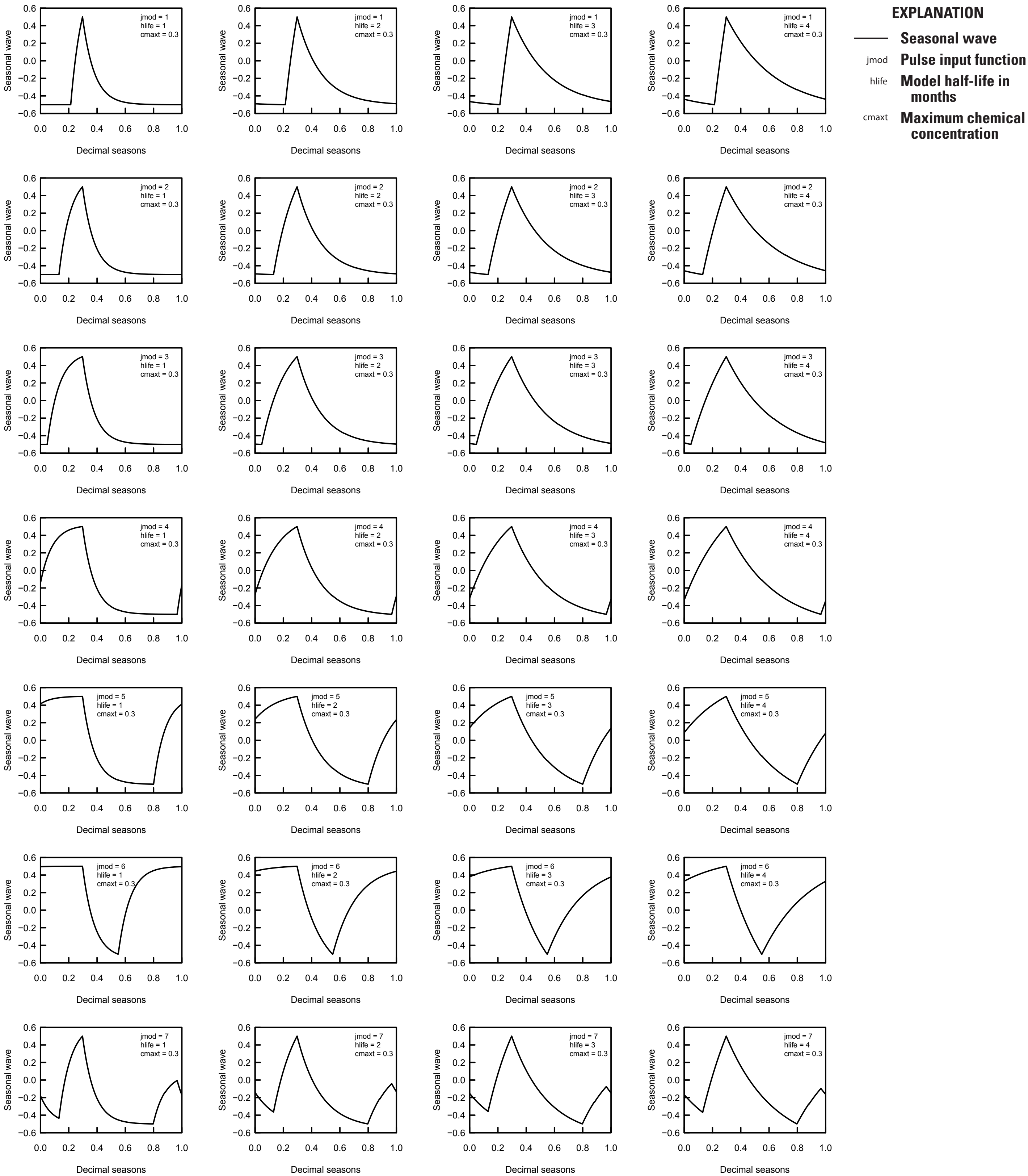

Figure 3–1. Visualization of seasonal waves with seasonal shift of the decimal season of maximum chemical concentration (cmaxt) equal to 0.3. jmod is the pulse input function, hlife is model half-life in months, and decimal season is the fraction of the calendar year represented by a specific month and day.

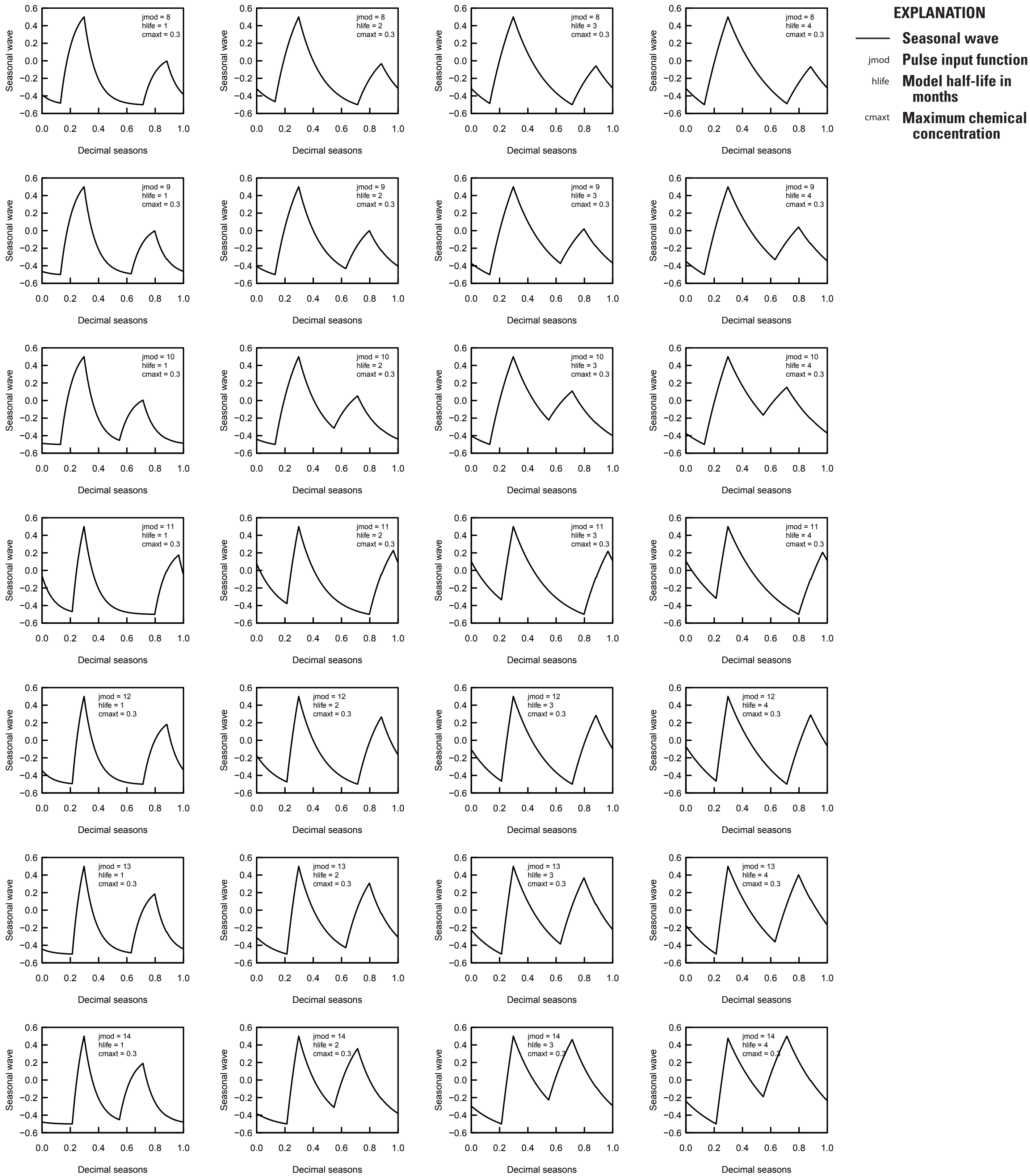

Figure 3–1. Visualization of seasonal waves with seasonal shift of the decimal season of maximum chemical concentration (cmaxt) equal to 0.3. jmod is the pulse input function, hlife is model half-life in months, and decimal season is the fraction of the calendar year represented by a specific month and day.—Continued

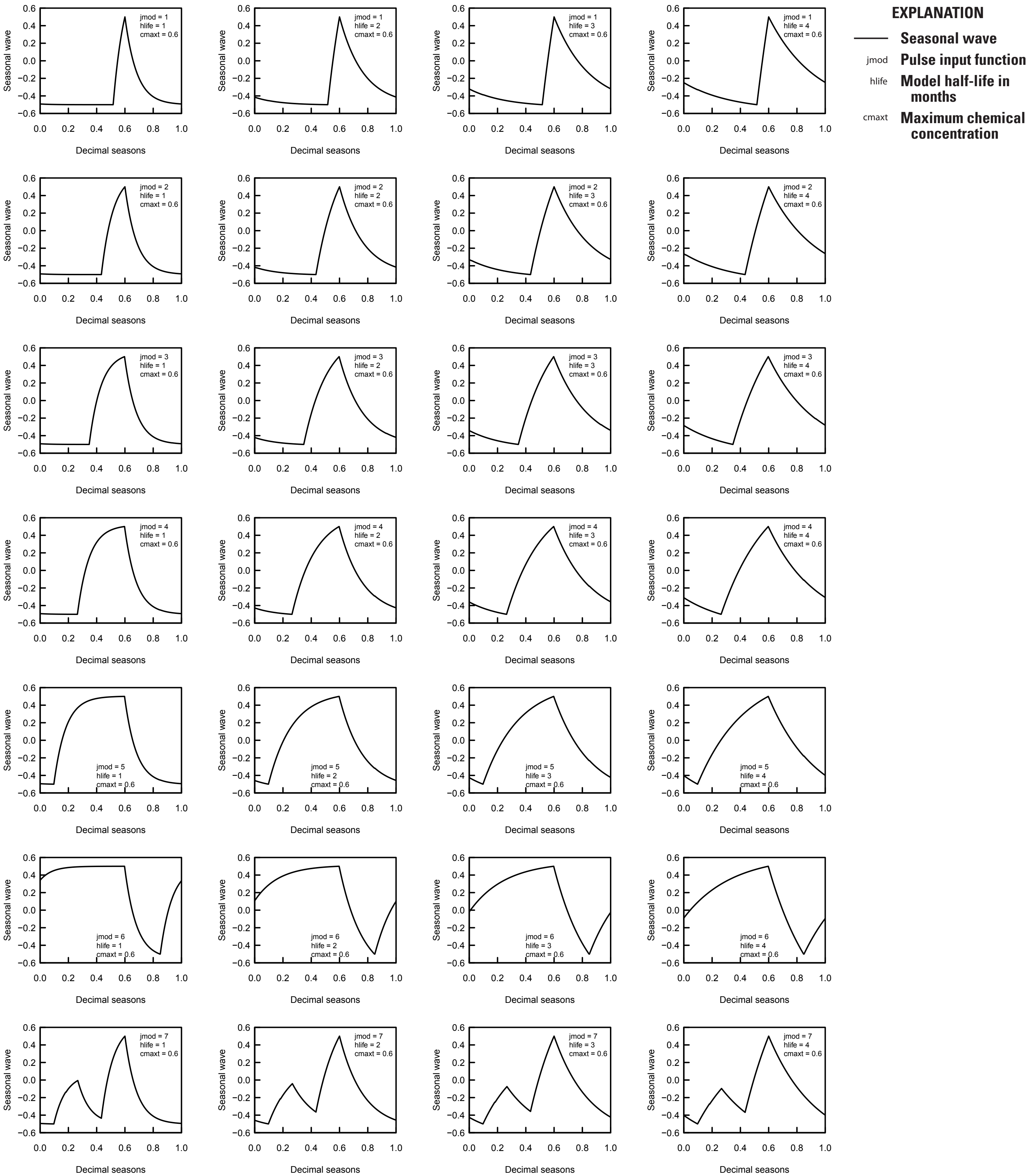

Figure 3–2. Visualization of seasonal waves with seasonal shift of the decimal season of maximum chemical concentration (cmaxt) equal to 0.6. jmod is the pulse input function, hlife is model half-life in months, and decimal season is is the fraction of the calendar year represented by a specific month and day.

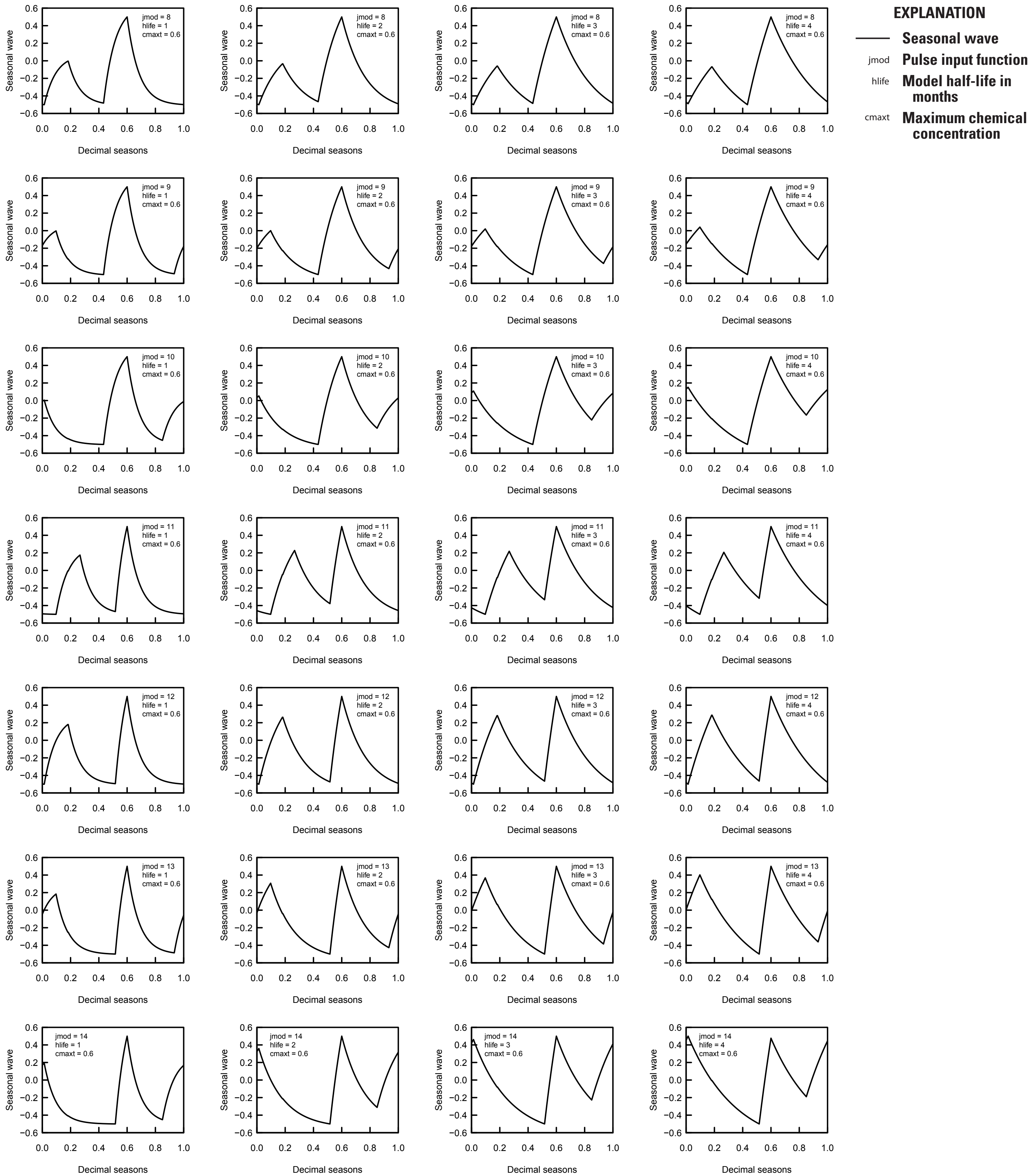

Figure 3–2. Visualization of seasonal waves with seasonal shift of the decimal season of maximum chemical concentration (cmaxt) equal to 0.6. jmod is the pulse input function, hlife is model half-life in months, and decimal season is is the fraction of the calendar year represented by a specific month and day.—Continued

Publishing support provided by:
Rolla Publishing Service Center

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

Or visit the North Dakota Water Science Center Web site at:
http://nd.water.usgs.gov/

Ryberg and Vecchia—**seawaveQ—An R Package Providing a Model and Utilities for Analyzing Trends in Chemical Concentrations**—Open-File Report 2013–1255

2331-1258 (online)
http://dx.doi.org/10.3133/ofr20131255