Vignette for seawaveQ—An R Package for Analyzing Trends in Pesticide Concentrations in Streams with a Seasonal Wave (seawave) and Adjustment for Streamflow (Q) and Other Ancillary Variables

Karen R. Ryberg and Aldo V. Vecchia

June 7, 2013

Contents

1 Introduction

This R package, **seawaveQ**, is designed for fitting a parametric regression model for assessing variability and trends in pesticide concentration in streams and was developed by Vecchia and others (2008), and subsequently refined and referred to as the "seawave-Q" model in several trend analyses (Ryberg and others, 2010; Sullivan and others, 2009; Vecchia and others, 2009). In these publications, "seawave-Q" stands for seasonal wave (seawave) with adjustment for streamflow (Q). 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). This R package provides a standardized methodology for fitting the seawave-Q model and makes the trend analysis method widely available for use by others. In addition, several enhancements to the seawave-Q model have been included such as: procedures for preparing and summarizing input data; flexibility to include other explanatory variables besides streamflow; graphical methods for assessing model fit; and additional model outputs. A flow chart showing how the various function

in the package work together is shown in figure 1 of the U.S. Geological Survey Open-File Report documenting this package (Ryberg and Vecchia, 2013).

The statistical methodology for the seawave-Q model is described in Vecchia and others (2008) and in the U.S. Geological Survey Open-File Report documenting this package (Ryberg and Vecchia, 2013). Users new to this model should read both of those documents before applying the model to their own data. An important part of the model and the output shown below is the seasonal wave. The seasonal wave is a periodic (with a period of 1 year) solution to a differential equation (Vecchia and others, 2008) that has a pulse input function, a seasonal shift that determines the time at which the seasonal wave reaches its maximum, and a model half-life.

2 Input Data

The seawaveQ model needs two types of input data. The first is the the water-quality sample data including dates, the concentration data, and qualification codes, indicating which values are censored (less than a long-term method detection level). The second type of data is the continuous ancillary data used in the model, such as streamflow anomalies (Ryberg and Vecchia, 2012). These ancillary data are also used to produce a continuous estimate of pesticide concentration. Examples of the necessary format of these two data sets are provided and documented in the package. The following code shows how to access the example data.

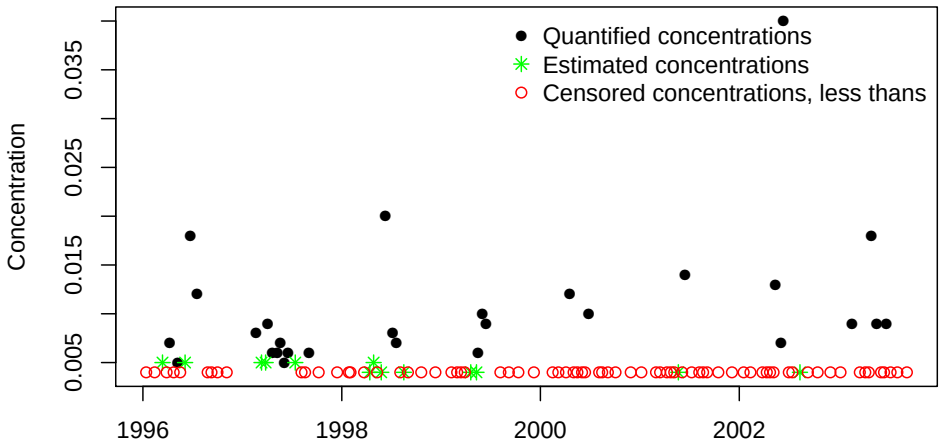
```
> # load waterData package, assuming it has already been installed on the system
> library(seawaveQ)
> # load example data that comes with the package
> data(swData)
> # show first few rows of water-quality data for Missouri River at Omaha, Nebraska
> head(qwMoRivOmaha)
```

```
     staid      dates times R04035 P04035 R04037 P04037 R04041 P04041 R39415 P39415 R46342 P463
1 06610000 1996-01-13  1130      <  0.004      _  0.024      <  0.008      _  0.006      <  0.0
2 06610000 1996-02-13  1200      <  0.004      E  0.005      <  0.008      _  0.200      <  0.0
3 06610000 1996-03-13  1000      E  0.005      E  0.004      <  0.008      _  0.026      <  0.0
4 06610000 1996-03-28  1030      <  0.004      E  0.005      _  0.009      _  0.026      <  0.0
5 06610000 1996-04-09  1100      _  0.007      E  0.006      _  0.014      _  0.075      E  0.0
6 06610000 1996-04-23  1000      <  0.004      <  0.004      _  0.012      _  0.040      <  0.0
```

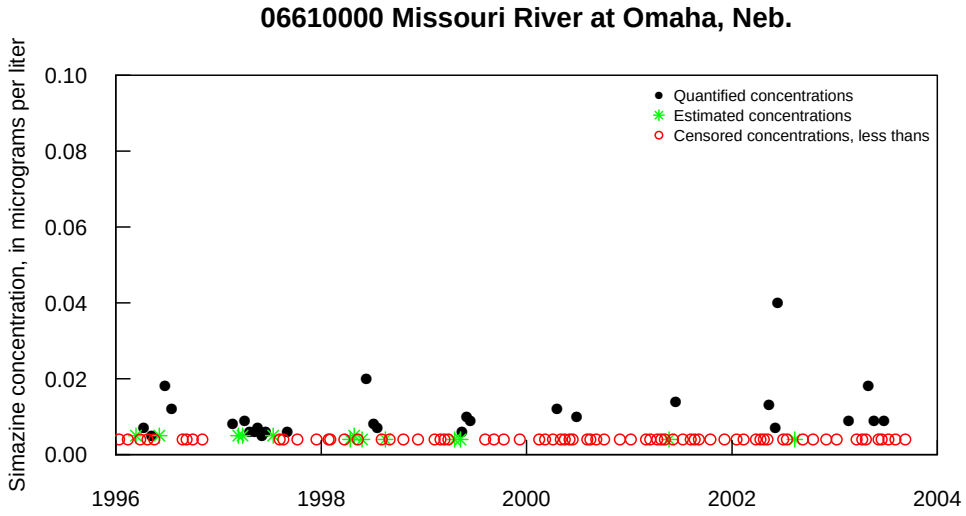
```
> # get a description of the data including definitions of the columns
> # by viewing the help documentation
> ?qwMoRivOmaha
```

Optionally, functions have been provided to plot concentration data. These functions produce scatter plots and box plots that indicate or take into account the censored, less than, values. The functions are *cenScatPlot* and *rosBoxPlot* and examples of their use follow. The box plots are generated using the function *ros*, regression on order statistics, in the R package **NADA** (Lee, 2012). It is an implementation of a regression on order statistics designed for multiply-censored analytical-chemistry data (Helsel, 2005).

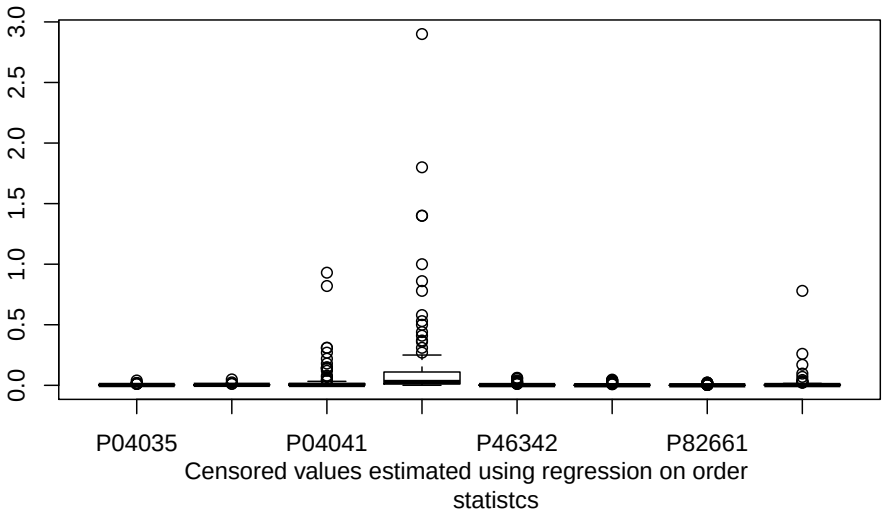
```
> # scatter plot showing unqualified, estimated, and censored  values
> cenScatPlot(qwMoRivOmaha, pname="04035")
```



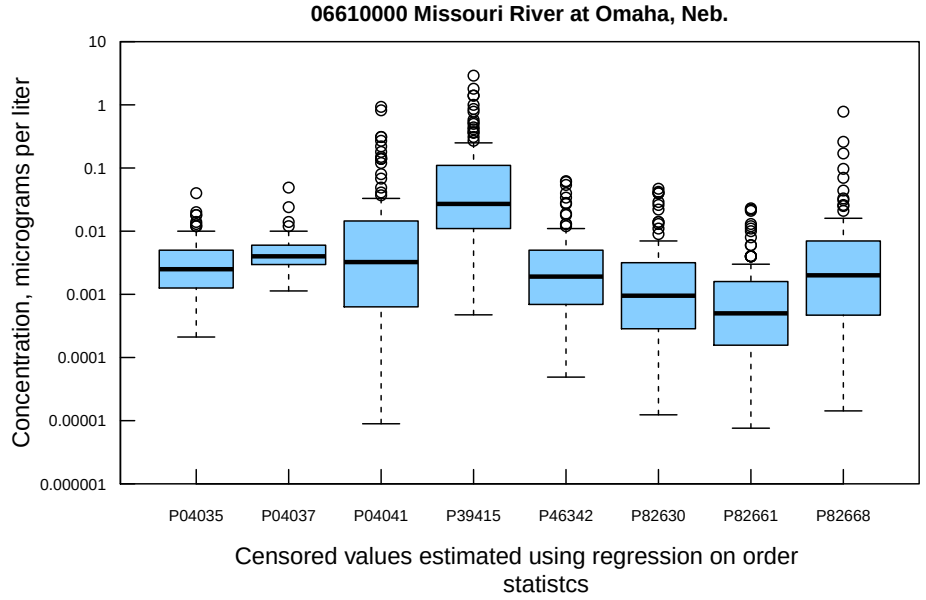
```
> # scatter plot with many additional plotting arguments
> # these options provide a plot closer to the plotting standards
> #  of the U.S. Geological Survey however, these plots may not
> # meet all U.S. Geological Survey publication requirements
> par(las=1, tcl=0.5)
> cenScatPlot(qwMoRivOmaha, pname="04035",
+                       site="06610000 Missouri River at Omaha, Neb.",
+                       ylabel="Simazine concentration, in micrograms per liter",
+                       legcex=0.7, qwcols=c("R", "P"), ylim=c(0,0.1), yaxs="i",
+                       cex.lab=0.9, cex.axis=0.9, xlim=c(as.Date("1996-01-01"),
+                       as.Date("2004-01-01")), xaxs="i", xaxt="n")
> axdates <- c("1996-01-01", "1998-01-01", "2000-01-01",
+                       "2002-01-01", "2004-01-01")
> axis(1, as.Date(axdates),
+                       labels=c("1996", "1998", "2000", "2002", "2004"), cex.axis=0.9)
```



```
> # simple box plots of water-quality concentrations
> rosBoxPlot(qwMoRivOmaha, qwcols=c("R", "P"))
```



```
> # same boxplot function with many additional plotting arguments
> rosBoxPlot(qwMoRivOmaha, site="06610000 Missouri River at Omaha, Neb.",
+                      log="y", yaxt="n", ylim=c(0.000001, 10), qwcols=c("R", "P"),
+                      ylab=c("Concentration, micrograms per liter"), col="skyblue1",
+                      cex.axis=0.7, cex.sub=0.8,
+                      par(tcl=0.5, las=1, yaxs="i", mgp=c(3,0.5,0), mar=c(5,5,2,2),
+                      cex.main=0.9))
> axis(2, at=c(0.000001, 0.00001, 0.0001, 0.001, 0.01, 0.1, 1, 10),
+ labels=c("0.000001", "0.00001", "0.0001", "0.001", "0.01",
+ "0.1", "1", "10"), cex.axis=0.7)
```



The second data set needed is the one containing the continuous ancillary data for building the model that describes pesticide concentrations.

```
> data(swData)
> # show last few rows of water-quality data for Missouri River at Omaha, Nebraska
> tail(cqwMoRivOmaha)
```

	staid	dates	dflow	flowa30	flowa1	dsed	seda30	seda1
2917	06610000	2003-09-25	29200	-0.09433312	-0.00527612	176	-0.1546322	-0.10766231
2918	06610000	2003-09-26	28800	-0.09268448	-0.01291513	177	-0.1566163	-0.10321762
2919	06610000	2003-09-27	28700	-0.09102975	-0.01608045	181	-0.1579888	-0.09213983
2920	06610000	2003-09-28	28700	-0.08926147	-0.01784872	184	-0.1588294	-0.08416002
2921	06610000	2003-09-29	28600	-0.08754373	-0.02108233	189	-0.1586754	-0.07267005
2922	06610000	2003-09-30	28700	-0.08577546	-0.02133474	201	-0.1569162	-0.04769499

```
> # get a description of the data including definitions of the columns
> # by viewing the help documentation
```

```
> ?cqwMoRivOmaha
```

In this case, the continuous ancillary data includes daily streamflow (dflow) and daily sediment concentration (dsed), as well as the 30-day and 1-day streamflow (flowa30 and flowa1) and sediment (seda30 and seda1) anomalies. (The anomalies were calculated using the **waterData** package for R.)

In order to build a model using one or more of these ancillary variables as explanatory variables for pesticide concentration, the continuous ancillary variables need to be associated with the water-quality samples. The function *combineData* will combine water-quality sample data and continuous (daily) ancillary variables and drop unnecessary columns. One needs to specify the water-quality sample data, the continuous ancillary data, and the columns representing the station identifier, the data, the qualification code columns, and the concentration columns as shown in the following code.

```
> data(swData)
> MoRivOmaha<-combineData(qwdat=qwMoRivOmaha, cqwdat=cqwMoRivOmaha,
+ qwcols=c("staid", "dates", "R", "P"))
> # view combined data set
> head(MoRivOmaha)
```

```
     staid      dates R04035 P04035 R04037 P04037 R04041 P04041 R39415 P39415 R46342 P46342 R82
1 06610000 1996-01-13      <  0.004      _  0.024      <  0.008      _  0.006      <  0.003
2 06610000 1996-02-13      <  0.004      E  0.005      <  0.008      _  0.200      <  0.003
3 06610000 1996-03-13      E  0.005      E  0.004      <  0.008      _  0.026      <  0.003
4 06610000 1996-03-28      <  0.004      E  0.005      _  0.009      _  0.026      <  0.003
5 06610000 1996-04-09      _  0.007      E  0.006      _  0.014      _  0.075      E  0.003
6 06610000 1996-04-23      <  0.004      <  0.004      _  0.012      _  0.040      <  0.003
       flowa30       flowa1 dsed      seda30       seda1
1 -0.111771936 -0.041600453  255  0.04313266 -0.14439969
2 -0.155914620  0.075222364  312  0.02706313 -0.04071575
3 -0.043752697 -0.008021798  236 -0.02856792 -0.10632729
4 -0.004315925  0.066689687  609  0.01503934  0.26177074
5  0.073100169  0.063475721  528  0.13734452  0.07748219
6  0.126711034 -0.003283307  368  0.23175763 -0.17371703
```

3 Fitting the seawaveQ Model

One can now fit the seawaveQ model using the data explored and combined in the previous code examples. The follow code fits three different seawaveQ models (with differing continuous ancillary variables) for two pesticides in the data set. The pesticides are 04035, simazine, and 04041, cyanazine. See the help documentation for further information about the function arguments shown and additional arguments.

```
> data(swData)
> # associate continuous water-quality data with each sample
> # combineData does this for you
> modMoRivOmaha<-combineData(qwdat=qwMoRivOmaha, cqwdat=cqwMoRivOmaha)
> # then fit model(s)
> myfit1 <- fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,
+ tanm="myfit1", pnames=c("04035", "04041"), yrstart=1995,
+ yrend=2003, tndbeg=1995, tndend=2003, iwcav=c("flowa30", "flowa1"),
+ dcol="dates", qwcols=c("R","P"))
> myfit2 <- fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,
+ tanm="myfit2", pnames=c("04035", "04041"), yrstart=1995,
+ yrend=2003, tndbeg=1995, tndend=2003, iwcav=c("seda30", "seda1"),
+ dcol="dates", qwcols=c("R","P"))
> myfit3 <- fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,
+ tanm="myfit3", pnames=c("04035", "04041"), yrstart=1995,
+ yrend=2003, tndbeg=1995, tndend=2003, iwcav=c("flowa30", "flowa1",
+ "seda30", "seda1"), dcol="dates", qwcols=c("R","P"))
```

4 Model Output

The model fitting process finds the best pulse input function and model half-life for the concentration data and uses survival regression to fit a regression model. Three types of output are provided: 1) a list, the first element being a data frame with information about the model and its parameters, the second element being the survival regression summary; 2) text files showing 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 data frame results for the three models for simazine and cyanazine are shown below.

```
> # get the first element of the list for each model/constituent combination
> # the data frame with information about each model/constituent combination
> myfit1[[1]]
  pname mclass jmod hlife   cmaxt     scl    loglik     cint   cwave     ctnd cflowa30 cflowa1
1 04035      1    3     1 0.48087 0.27199 -32.79163 -2.49349 0.55437 -0.02793  0.05386 2.10427
2 04041      1    4     3 0.48087 0.42405 -43.05521 -2.26965 2.17692 -0.24887 -0.02829 2.98901
  pvaltnd
1 0.18598
2 0.00000
> myfit2[[1]]
  pname mclass jmod hlife   cmaxt     scl    loglik     cint   cwave     ctnd cseda30  cseda1
1 04035      1    5     3 0.48087 0.24677 -24.32398 -2.68378 0.79650 -0.02405 0.25361 0.59703 0
2 04041      1    3     2 0.48087 0.40826 -39.64286 -2.12752 1.58343 -0.22011 1.13733 0.63650 0
```

```
  pvaltnd
1 0.12907
2 0.00000
```

```
> myfit3[[1]]
```

```
  pname mclass jmod hlife   cmaxt     scl    loglik     cint   cwave     ctnd cflowa30  cflowa1
1 04035      1    5     3 0.48087 0.24621 -23.83644 -2.67838 0.76533 -0.01817  0.17914 -0.36126
2 04041      1    3     2 0.48087 0.38981 -37.63645 -2.07403 1.71880 -0.24112 -0.61646  1.58972
  seflowa30 seflowa1 seseda30 seseda1 pvaltnd
1   0.30546  0.64210  0.20464 0.15225 0.34346
2   0.49387  1.11121  0.30870 0.26845 0.00000
```

```
> # get the second element of the list for each model/constituent combination
> # the survival regression summary for each model/constituent combination
> myfit1[[2]]
```

```
[[1]]

Call:
survreg(formula = Surv(time = clogtmp, time2 = indcen, type = "left") ~
    xxxmat - 1, dist = "gaussian")
                Value Std. Error       z        p
xxxmatintcpt  -2.4935     0.0491 -50.782 0.00e+00
xxxmatwavest   0.5544     0.1117   4.962 6.97e-07
xxxmattndlin  -0.0279     0.0211  -1.323 1.86e-01
xxxmatflowa30  0.0539     0.3049   0.177 8.60e-01
xxxmatflowa1   2.1043     0.4689   4.488 7.20e-06
Log(scale)    -1.3020     0.1205 -10.804 3.31e-27

Scale= 0.272

Gaussian distribution
Loglik(model)= -32.8   Loglik(intercept only)= -60.9
        Chisq= 56.16 on 4 degrees of freedom, p= 1.9e-11
Number of Newton-Raphson Iterations: 5
n= 115


[[2]]

Call:
survreg(formula = Surv(time = clogtmp, time2 = indcen, type = "left") ~
    xxxmat - 1, dist = "gaussian")
                Value Std. Error        z         p
xxxmatintcpt  -2.2697     0.0853 -26.5946 7.85e-156
xxxmatwavest   2.1769     0.2594   8.3929  4.74e-17
```

```
xxxmattndlin  -0.2489     0.0368  -6.7576  1.40e-11
xxxmatflowa30 -0.0283     0.5138  -0.0551  9.56e-01
xxxmatflowa1   2.9890     0.7801   3.8317  1.27e-04
Log(scale)    -0.8579     0.1036  -8.2828  1.20e-16

Scale= 0.424

Gaussian distribution
Loglik(model)= -43.1   Loglik(intercept only)= -103.7
        Chisq= 121.33 on 4 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 6
n= 115
```

> myfit2[[2]]

```
[[1]]

Call:
survreg(formula = Surv(time = clogtmp, time2 = indcen, type = "left") ~
    xxxmat - 1, dist = "gaussian")
              Value Std. Error      z        p
xxxmatintcpt -2.6838     0.0685 -39.17 0.00e+00
xxxmatwavest  0.7965     0.1867   4.27 1.99e-05
xxxmattndlin -0.0241     0.0158  -1.52 1.29e-01
xxxmatseda30  0.2536     0.1729   1.47 1.43e-01
xxxmatseda1   0.5970     0.1025   5.82 5.78e-09
Log(scale)   -1.3993     0.1184 -11.82 3.11e-32

Scale= 0.247

Gaussian distribution
Loglik(model)= -24.3   Loglik(intercept only)= -60.9
        Chisq= 73.09 on 4 degrees of freedom, p= 5e-15
Number of Newton-Raphson Iterations: 6
n= 115


[[2]]

Call:
survreg(formula = Surv(time = clogtmp, time2 = indcen, type = "left") ~
    xxxmat - 1, dist = "gaussian")
              Value Std. Error      z         p
xxxmatintcpt -2.128     0.0751 -28.31 2.31e-176
xxxmatwavest  1.583     0.1894   8.36  6.18e-17
xxxmattndlin -0.220     0.0293  -7.51  5.96e-14
```

```
xxxmatseda30  1.137     0.2874   3.96  7.60e-05
xxxmatseda1   0.636     0.1813   3.51  4.46e-04
Log(scale)   -0.896     0.1027  -8.73  2.65e-18

Scale= 0.408

Gaussian distribution
Loglik(model)= -39.6   Loglik(intercept only)= -103.7
        Chisq= 128.16 on 4 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 6
n= 115
```

> myfit3[[2]]

```
[[1]]

Call:
survreg(formula = Surv(time = clogtmp, time2 = indcen, type = "left") ~
    xxxmat - 1, dist = "gaussian")
               Value Std. Error       z        p
xxxmatintcpt  -2.6784     0.0672 -39.830 0.00e+00
xxxmatwavest   0.7653     0.1844   4.150 3.33e-05
xxxmattndlin  -0.0182     0.0192  -0.947 3.43e-01
xxxmatflowa30  0.1791     0.3055   0.586 5.58e-01
xxxmatflowa1  -0.3613     0.6421  -0.563 5.74e-01
xxxmatseda30   0.2273     0.2046   1.111 2.67e-01
xxxmatseda1    0.6665     0.1522   4.378 1.20e-05
Log(scale)    -1.4016     0.1188 -11.802 3.79e-32

Scale= 0.246

Gaussian distribution
Loglik(model)= -23.8   Loglik(intercept only)= -60.9
        Chisq= 74.07 on 6 degrees of freedom, p= 6e-14
Number of Newton-Raphson Iterations: 6
n= 115


[[2]]

Call:
survreg(formula = Surv(time = clogtmp, time2 = indcen, type = "left") ~
    xxxmat - 1, dist = "gaussian")
              Value Std. Error      z         p
xxxmatintcpt  -2.074     0.0784 -26.47 2.27e-154
xxxmatwavest   1.719     0.1999   8.60  8.10e-18
```

```
xxxmattndlin  -0.241     0.0336  -7.18  7.02e-13
xxxmatflowa30 -0.616     0.4939  -1.25  2.12e-01
xxxmatflowa1   1.590     1.1112   1.43  1.53e-01
xxxmatseda30   1.039     0.3087   3.37  7.65e-04
xxxmatseda1    0.290     0.2684   1.08  2.80e-01
Log(scale)    -0.942     0.1033  -9.12  7.53e-20

Scale= 0.39

Gaussian distribution
Loglik(model)= -37.6   Loglik(intercept only)= -103.7
        Chisq= 132.17 on 6 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 6
n= 115

>
```

The first item of the list, the data frame, contains information about each model including the pesticide analyzed; the model class (an option in seawaveQ that is not currently implemented but that will provide additional model options in the future); the choice of model or pulse input function, an integer 1 through 14; the model half-life in months, an integer, 1 to 4 months; the decimal season of highest concentration; the scale factor from the survreg.object; the log-likelihood for the model; the coefficient for the model intercept; the coefficient for the seasonal wave; the coefficient for the trend component of the model; 0 or more values representing coefficients for the continuous ancillary variables; the standard error for the intercept; the standard error for the seasonal wave; the standard error for the trend; and 0 or more columns representing standard errors for the continuous ancillary variables.

The second item of the list is provided so that users could extract the attributes of the survival regression summary programmatically (rather than viewing them in the text file) and create their own summaries or plots of the results.

```
> attributes(myfit1[[2]][[1]])

$names
 [1] "call"         "df"           "loglik"       "iter"         "idf"          "scale"
[10] "correlation"  "parms"        "n"            "chi"          "robust"


$class
[1] "summary.survreg"

> myfit1[[2]][[1]]$n

[1] 115

> myfit1[[2]][[1]]$table

                    Value Std. Error           z            p
xxxmatintcpt  -2.49349164 0.04910211 -50.7817616 0.000000e+00
```

```
xxxmatwavest   0.55436953 0.11171616   4.9623039 6.966189e-07
xxxmattndlin  -0.02793350 0.02112089  -1.3225534 1.859839e-01
xxxmatflowa30  0.05385545 0.30491533   0.1766243 8.598035e-01
xxxmatflowa1   2.10426938 0.46888353   4.4878296 7.195245e-06
Log(scale)    -1.30197275 0.12051232 -10.8036489 3.307909e-27
```

```
>
```

The text file for the first of the function calls above is inserted here as an example. Users may run the model fitting code themselves and view the resulting text files for all three models. The results for all three are too long to include in this vignette.

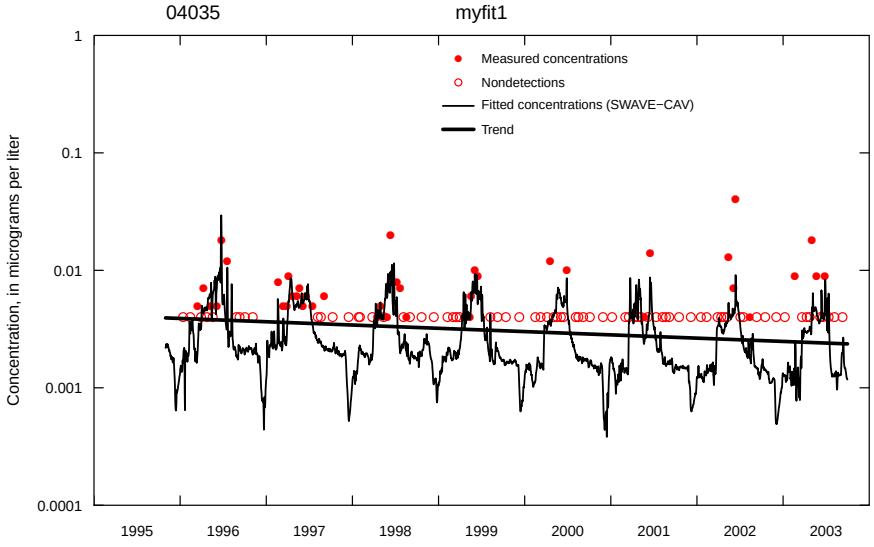
```
Friday 07 Jun 2013 16:03:19 PM CDT
R version 3.0.1 (2013-05-16)
seawaveQ version 0.7.0
x86_64-apple-darwin10.8.0 (64-bit)

Final model survreg results for 04035
Call:
survreg(formula = Surv(time = clogtmp, time2 = indcen, type = "left") ~
    xxxmat - 1, dist = "gaussian")
                Value Std. Error       z        p
xxxmatintcpt  -2.4935     0.0491 -50.782 0.00e+00
xxxmatwavest   0.5544     0.1117   4.962 6.97e-07
xxxmattndlin  -0.0279     0.0211  -1.323 1.86e-01
xxxmatflowa30  0.0539     0.3049   0.177 8.60e-01
xxxmatflowa1   2.1043     0.4689   4.488 7.20e-06
Log(scale)    -1.3020     0.1205 -10.804 3.31e-27

Scale= 0.272

Gaussian distribution
Loglik(model)= -32.8   Loglik(intercept only)= -60.9
Chisq= 56.16 on 4 degrees of freedom, p= 1.9e-11
Number of Newton-Raphson Iterations: 5
n= 115

AIC (Akaike's An Information Criterion) is:  77.58325
BIC (Bayesian Information Criterion) is:  94.05285
Model class is 1
Pulse input function is 3
Half life is 1
Seasonal value of the maximum concentration is 0.4808743.
```

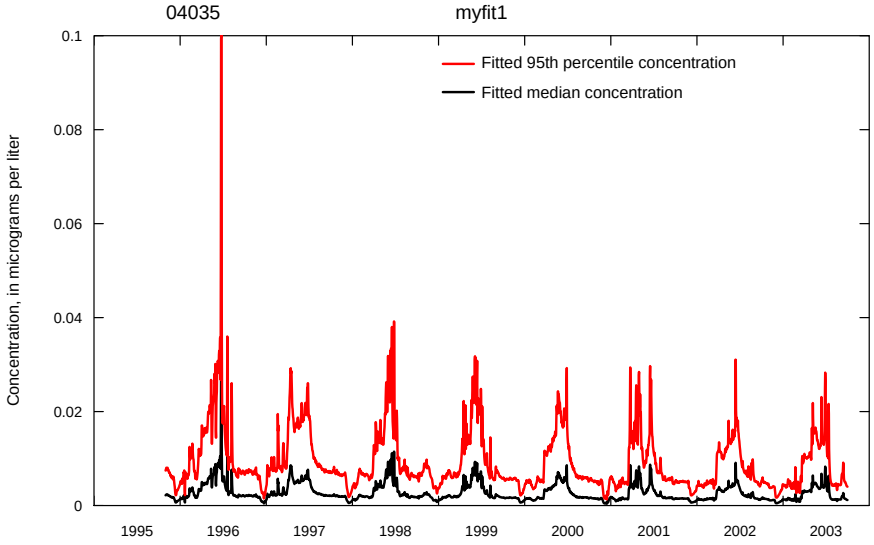
```
Friday 07 Jun 2013 16:03:20 PM CDT
R version 3.0.1 (2013-05-16)
seawaveQ version 0.7.0
x86_64-apple-darwin10.8.0 (64-bit)

Final model survreg results for 04041
Call:
survreg(formula = Surv(time = clogtmp, time2 = indcen, type = "left") ~
    xxxmat - 1, dist = "gaussian")
               Value Std. Error        z         p
xxxmatintcpt  -2.2697     0.0853 -26.5946 7.85e-156
xxxmatwavest   2.1769     0.2594   8.3929  4.74e-17
xxxmattndlin  -0.2489     0.0368  -6.7576  1.40e-11
xxxmatflowa30 -0.0283     0.5138  -0.0551  9.56e-01
xxxmatflowa1   2.9890     0.7801   3.8317  1.27e-04
Log(scale)    -0.8579     0.1036  -8.2828  1.20e-16

Scale= 0.424

Gaussian distribution
Loglik(model)= -43.1   Loglik(intercept only)= -103.7
Chisq= 121.33 on 4 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 6
n= 115

AIC (Akaike's An Information Criterion) is:  98.11042
BIC (Bayesian Information Criterion) is:  114.58
Model class is 1
Pulse input function is 4
Half life is 3
Seasonal value of the maximum concentration is 0.4808743.


Friday 07 Jun 2013 16:04:57 PM CDT
R version 3.0.1 (2013-05-16)
seawaveQ version 0.7.0
x86_64-apple-darwin10.8.0 (64-bit)

Final model survreg results for 04035
Call:
survreg(formula = Surv(time = clogtmp, time2 = indcen, type = "left") ~
    xxxmat - 1, dist = "gaussian")
               Value Std. Error       z        p
```

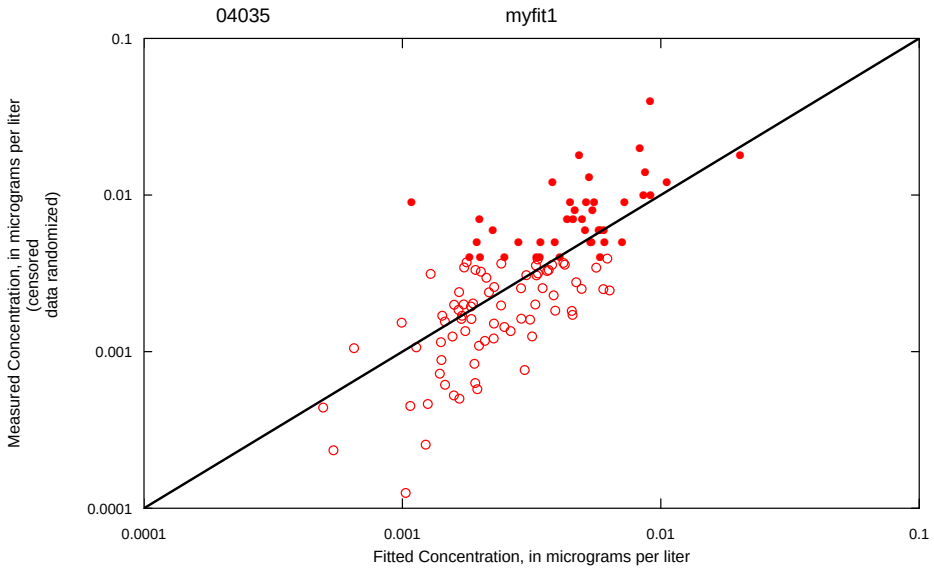
```
xxxmatintcpt  -2.4935     0.0491 -50.782 0.00e+00
xxxmatwavest   0.5544     0.1117   4.962 6.97e-07
xxxmattndlin  -0.0279     0.0211  -1.323 1.86e-01
xxxmatflowa30  0.0539     0.3049   0.177 8.60e-01
xxxmatflowa1   2.1043     0.4689   4.488 7.20e-06
Log(scale)    -1.3020     0.1205 -10.804 3.31e-27

Scale= 0.272

Gaussian distribution
Loglik(model)= -32.8   Loglik(intercept only)= -60.9
Chisq= 56.16 on 4 degrees of freedom, p= 1.9e-11
Number of Newton-Raphson Iterations: 5
n= 115

AIC (Akaike's An Information Criterion) is:  77.58325
BIC (Bayesian Information Criterion) is:  94.05285
Model class is 1
Pulse input function is 3
Half life is 1
Seasonal value of the maximum concentration is 0.4808743.


Friday 07 Jun 2013 16:04:57 PM CDT
R version 3.0.1 (2013-05-16)
seawaveQ version 0.7.0
x86_64-apple-darwin10.8.0 (64-bit)

Final model survreg results for 04041
Call:
survreg(formula = Surv(time = clogtmp, time2 = indcen, type = "left") ~
    xxxmat - 1, dist = "gaussian")
               Value Std. Error        z         p
xxxmatintcpt  -2.2697     0.0853 -26.5946 7.85e-156
xxxmatwavest   2.1769     0.2594   8.3929  4.74e-17
xxxmattndlin  -0.2489     0.0368  -6.7576  1.40e-11
xxxmatflowa30 -0.0283     0.5138  -0.0551  9.56e-01
xxxmatflowa1   2.9890     0.7801   3.8317  1.27e-04
Log(scale)    -0.8579     0.1036  -8.2828  1.20e-16

Scale= 0.424

Gaussian distribution
Loglik(model)= -43.1   Loglik(intercept only)= -103.7
Chisq= 121.33 on 4 degrees of freedom, p= 0
```

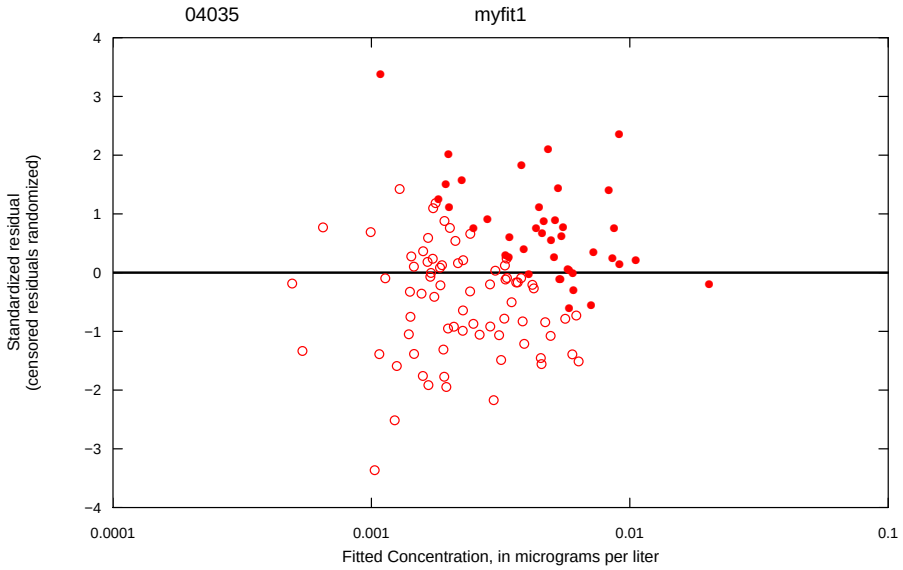
```
Number of Newton-Raphson Iterations: 6
n= 115

AIC (Akaike's An Information Criterion) is:  98.11042
BIC (Bayesian Information Criterion) is:  114.58
Model class is 1
Pulse input function is 4
Half life is 3
Seasonal value of the maximum concentration is 0.4808743.
```

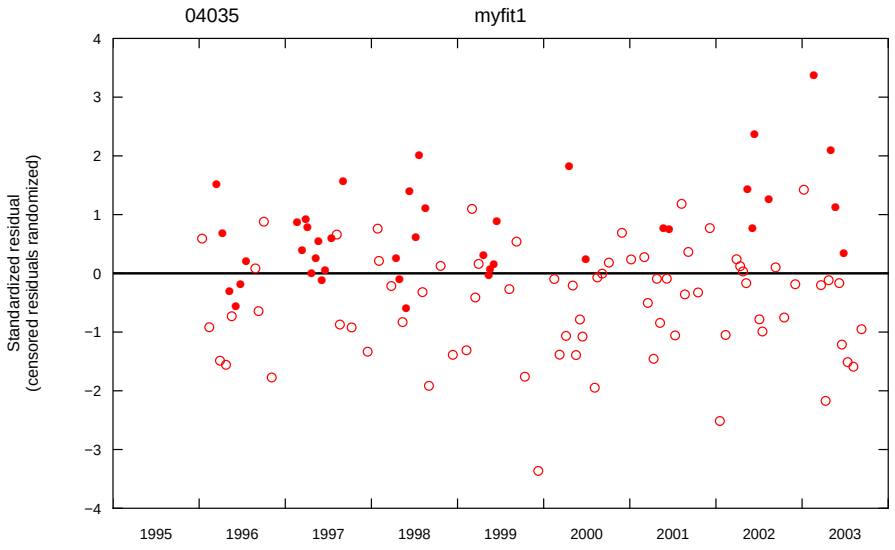
The plots written to a pdf file for the first pesticide, 04035, simazine, in the first model, myfit1, are included below. As with the text results, the plots for all three models and all pesticides are too numerous to include here. Users are encouraged to run the code themselves and examine all of the plots.

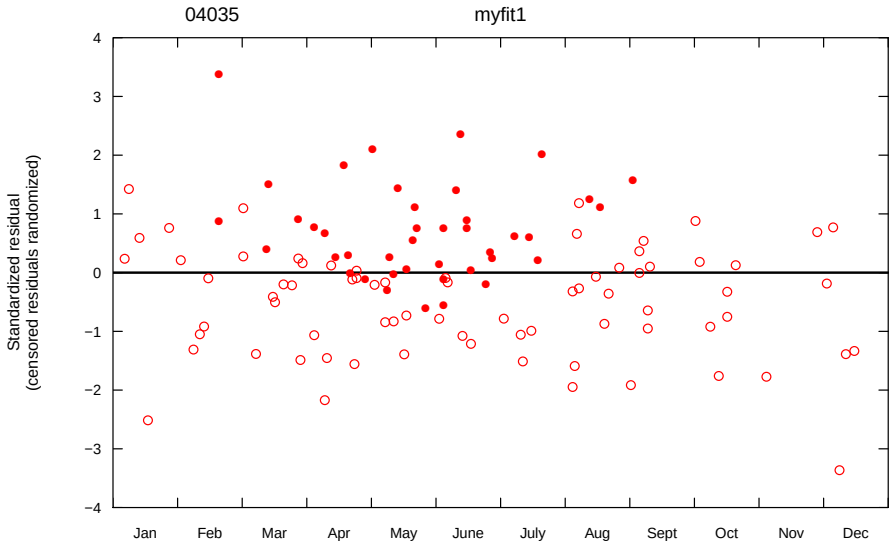

04035
myfit1
Measured concentrations
Nondetections
Fitted concentrations (SWAVE−CAV)
Trend
Concentration, in micrograms per liter
1
0.1
0.01
0.001
0.0001
1995
1996
1997
1998
1999
2000
2001
2002
2003


04035
myfit1
Fitted 95th percentile concentration
Fitted median concentration
Concentration, in micrograms per liter
0.1
0.08
0.06
0.04
0.02
0
1995
1996
1997
1998
1999
2000
2001
2002
2003


04035
myfit1
Measured Concentration, in micrograms per liter
(censored
data randomized)
0.1
0.01
0.001
0.0001
0.0001
0.001
0.01
0.1
Fitted Concentration, in micrograms per liter


04035
myfit1
Standardized residual
(censored residuals randomized)
4
3
2
1
0
−1
−2
−3
−4
0.0001
0.001
0.01
0.1
Fitted Concentration, in micrograms per liter


04035
myfit1
Standardized residual
(censored residuals randomized)
4
3
2
1
0
−1
−2
−3
−4
1995
1996
1997
1998
1999
2000
2001
2002
2003


04035
myfit1
Standardized residual
(censored residuals randomized)
4
3
2
1
0
−1
−2
−3
−4
Jan
Feb
Mar
Apr
May
June
July
Aug
Sept
Oct
Nov
Dec

The plotting position used for representing censored values in the model plots (produced by the internal function *seawaveQPlots* further described in the package help documentation) is an important consideration for interpreting model fit. Plotting values obtained by using the censoring limit, or something smaller such as one-half of the censoring limit, produce plots that are difficult to interpret if there are a large number of censored values. Therefore, to make the plots more representative of diagnostic plots used for standard (non-censored) regression, a method for substituting randomized residuals in place of censored residuals was used . If a log-transformed concentration is censored at a particular limit, $logC < L$, then the residual for that concentration is censored as well, $logC - fitted(logC) < L - fitted(logC) = rescen$). In that case, a randomized residual was generated from a conditional normal distribution:

```
resran  <-  scl * qnorm(runif(1) * pnorm(rescen / scl))
```

where *scl* is the scale parameter from the survival regression model, *pnorm* is the R function for computing cumulative normal probabilities, *runif* is the R function for generating a random variable from the uniform distribution, and *qnorm* is the R function for computing quantiles of the normal distribution. Under the assumption that the model residuals are uncorrelated, normally distributed random variables with mean zero and standard deviation scl, the randomized residuals generated in this manner are an unbiased sample of the true (but unknown) residuals for the censored data. This is an application of the probability integral transform (Mood, 1974) to generate random variables from continuous distributions. The plotting position used a censored concentration is $fitted(logC) + resran$. Note that each time a new model fit is performed, a new set of randomized residuals is generated and thus the plotting positions for censored values can change.

5 References

Helsel, D.R., 2005, Nondetects and data analysis: New York, John Wiley and Sons.

Lee, Lopaka, 2012, Nondetects and data analysis for environmental data: R package version 1.5-4, http://CRAN.R-project.org/package=NADA.

Mood, A.M., Graybill, F.A., and Boes, D.C., 1974, Introduction to the theory of statistics (3d ed.): New York, McGraw-Hill, Inc., 564 p.

Ryberg, K.R. and Vecchia, A.V., 2006, Water-quality trend analysis and sampling design for the Devils Lake Basin, North Dakota, January 1965 through September 2003: U.S. Geological Survey Scientific Investigations Report 2006–5238, 64 p., accessed May 1, 2012, at http://pubs.usgs.gov/sir/2006/5238/.

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., accessed March 1, 2013, at http://pubs.usgs.gov/of/2012/1168/.

Ryberg, K.R. and Vecchia, A.V., 2013, seawaveQ—An R Package for Analyzing Trends in Pesticide 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-XXXX, XX p.

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., accessed May 1, 2012, 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., accessed May 1, 2012, at http://pubs.usgs.gov/sir/2009/5132/.

Vecchia, A.V., 2005, Water-quality trend analysis and sampling design for streams in the Red River of the North Basin, Minnesota, North Dakota, and South Dakota, 1970–2001: U.S. Survey Scientific Investigations Report 2005–5224, 54 p., accessed May 1, 2012, at http://pubs.usgs.gov/sir/2005/5224/.

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. 9,096-9,102, accessed May 1, 2012, at http://water.usgs.gov/nawqa/pubs/es902122j.pdf.

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, pp. 1,308-1,324, accessed May 1, 2012, at http://dx.doi.org/10.1111/j.1752-1688.2008.00225.x.