Vignette for waterData—An R Package for Retrieval, Analysis, and Anomaly Calculation of Daily Hydrologic Time Series Data

Karen R. Ryberg and Aldo V. Vecchia

April 28, 2017

Contents

1 Introduction

U.S. Geological Survey (USGS) daily streamflow and continuous physical property data can be used as exogenous variables in trend models for water-quality analysis (Helsel and Hirsch, 2002). In addition, streamflow can be divided into multiple components, anomalies, representing variability over longer-term (annual or multi-year) and shorter-term (less than a year) time scales. Those components can be used as multiple exogenous variables in multiple regression models. The anomaly concept was first described by Vecchia (2003) and subsequently used and refined in numerous studies of surface-water quality (Alexander and Smith, 2006; Ryberg and Vecchia, 2006; Ryberg and others, 2010; Sullivan and others, 2009; Vecchia, 2005; Vecchia and others, 2008; Vecchia and others, 2009).

This package allows users to import USGS daily hydrologic time series data into R, plot and summarize the imported data, and calculate anomalies. There are four possible sets of streamflow anomalies. The following examples show users how to import, clean, summarize, and plot the

imported data and calculate and plot streamflow anomalies. The package **waterData** depends on the other R packages **lattice**, **latticeExtra**, and **XML**; however, the examples in this vignette also use the packages **xtable**, **maps**, and **mapdata**. Users may want to install those packages as well, or skip the examples using them.

2   Site Identification and Data Import

A major feature of this package is the direct import of USGS daily streamflow data from the USGS Daily Values Site Web Service, https://waterservices.usgs.gov/rest/DV-Service.html (U.S. Geological Survey, 2017d). To import, a USGS identification number (eight or more digits) that uniquely identifies a streamgage site is needed. (The USGS has many sites where water-quality and groundwater data are collected. Identification numbers for water-quality only sites or groundwater sites will not import any data into R using this package.) Users may search for surface-water sites and obtain their station identification numbers using the USGS Site Web Service, https://waterservices.usgs.gov/rest/Site-Service.html (U.S. Geological Survey, 2017e); using the National Water Information System: Mapper, https://maps.waterdata.usgs.gov/mapper/index.html (U.S. Geological Survey, 2017a); or using the National Water Information System: Web Interface to daily surface-water data, https://waterdata.usgs.gov/nwis/dv/?referred_module=sw (U.S. Geological Survey, 2017f).

Once a site has been selected (for example site 05054000, Red River of the North at Fargo, North Dakota), users may use the function *importDVs* to download water data. First, users must choose which continuously measured parameter to download. This particular gage measures gage height (stage), streamflow (discharge), specific conductance, pH, water temperature, turbidity, and dissolved oxygen. (Most USGS gages do not continuously measure all of these parameters.) For this example, we are interested in streamflow; therefore, the function argument code is set to "00060," the USGS parameter code for streamflow. Depending on the gage and parameter, users may select to download daily mean, maximum, minimum, or median. When working with daily streamflow, the daily mean generally is used and this is indicated by the statistics code "00003." A list of possible parameter codes is available at https://nwis.waterdata.usgs.gov/usa/nwis/pmcodes (U.S. Geological Survey, 2017b) and a list of statistics codes is available at https://nwis.waterdata.usgs.gov/nwis/help/?read_file=stat&format=table (U.S. Geological Survey, 2017c). Again, not all parameters are available at every gage and not every statistic is available for every parameter. Users also may indicate a start and end date, specified using the "sdate" and "edate" function arguments, for the desired data in the form YYYY-MM-DD, where YYYY represents year, MM month, and DD day. If no start date is specified, the start date used is "1851-01-01", and if no end date is specified, the current system date is used.

The following code and results show how to import data for the example site and view a portion of the data:

```
> # load waterData package, assuming it has already been installed on the system
> library(waterData)
> q05054000 <- importDVs("05054000", code="00060", stat = "00003",
+                        sdate = "2000-01-01", edate = "2010-12-31")
```

```
> # return first 6 rows for new data set to view a subset of the data
> head(q05054000)

     staid val       dates qualcode
1 05054000 470 2000-01-01        A
2 05054000 470 2000-01-02        A
3 05054000 470 2000-01-03        A
4 05054000 460 2000-01-04        A
5 05054000 460 2000-01-05        A
6 05054000 460 2000-01-06        A
```

The function checks to ensure that the station identification number is at least eight characters long and provides a warning message if it is not. If no data are returned, there may be an error in the date syntax, the station identification number may be invalid, or the USGS data service may be experiencing problems. Users may want the URL used to retrieve the data for troubleshooting and examining error messages from the data service. The uniform resource locator (URL) can be pasted into a browser to examine error messages or the data in the WaterML format (version 1.1; Valentine and Zaslavsky, 2009). The *tellMeURL* function will provide the URL.

```
> my.URL <- tellMeURL("05054000", code = "00060", stat = "00003",
+                     sdate = "2000-01-01", edate = "2010-12-31")
```

The object "my.URL" is now a character vector containing the following URL.
https://waterservices.usgs.gov/nwis/dv/?format=waterml,2.0&site=05054000&parameterCd=00060&statCd=00003&startDt=2000-01-01&endDt=2010-12-31

3   Plot

Users should then plot the data to check for potential problems. For example, some streamgage sites have large gaps in data collection and may not be appropriate for further use in hydrologic studies. Some sites may also have anomalous data values. The plot function uses *xyplot*, a lattice plot (Sarkar, 2008), and could accommodate data frames with data for multiple sites. The following code was used to produce figure 1. The streamflow data are plotted on an arithmetic scale rather than a log scale, commonly used for streamflow, because any zero or negative values would create errors when plotting on a logscale.

```
> data(exampleWaterData, package = "waterData")
> my.plot <- plotParam(badDataSet)
> print(my.plot)
```
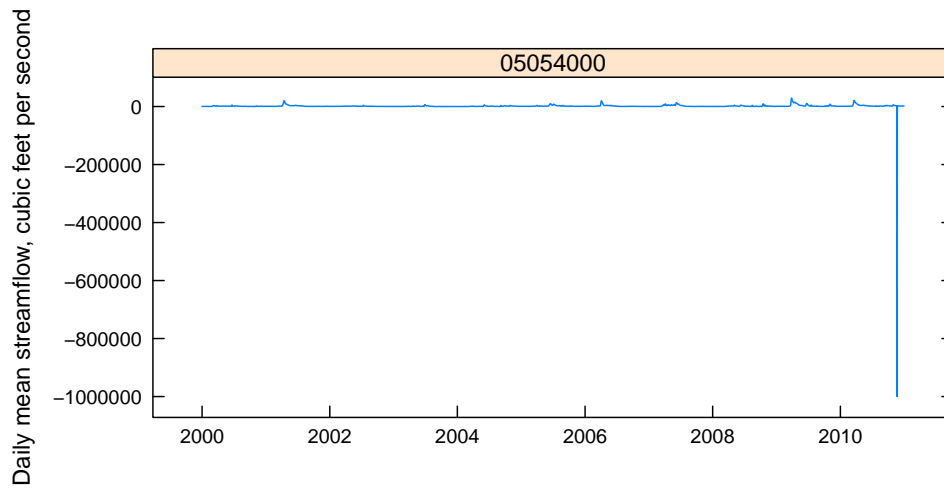

Figure 1: Mean daily streamflow for site 05054000 from January 1, 2000, through December 31, 2010.

4   Data Cleanup

Based on figure 1, this site contains an anomalous data value that needs to be explored further. The function *cleanUp* will identify negative and zero values as shown in the following example.

```
> cleanUp(badDataSet, task = "view")

[[1]]
        staid     val       dates qualcode
3979 05054000 -999999 2010-11-22  P Ice e

[[2]]
        staid val       dates qualcode
1365 05054000   0 2003-09-26        A
1366 05054000   0 2003-09-27        A
1367 05054000   0 2003-09-28        A
```

The streamflow value, column "val", is -999999.00 and has qualification codes of "P" (provisional data subject to revision), Ice (meaning the streamgage was affected by ice on this date), and "e" (value has been estimated). This value should be changed to "NA" to indicate missing data. Users are encouraged to use only those data with a qualification code of "A" (approved for publication–processing and review completed).

It may be necessary to remove negative values from the streamflow record or to replace values of 0 streamflow with 0.1 or 0.01. Streamflow anomalies are calculated using the base-10 logarithm

of streamflow; therefore negative and zero values will cause errors and warnings. To change the values, set the task argument to "fix" as in the following example. The results of the cleanup are shown in figure 2.

```
> q05054000Fix <- cleanUp(badDataSet, task = "fix", replace = 0.1)

> my.plot <- plotParam(q05054000Fix, code = "00060", stat = "00003", logscale = TRUE)
> print(my.plot)
```
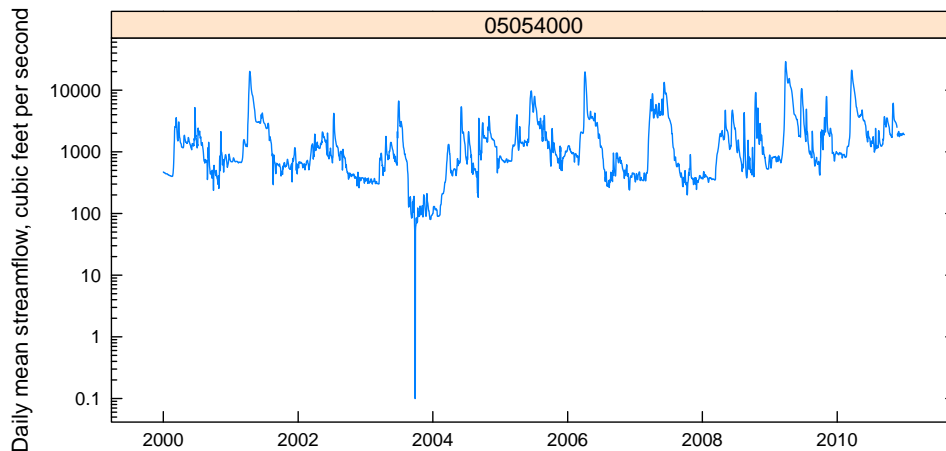

Figure 2: Daily streamflow with negative values changed to "NA" and 0 values replaced with 0.1.

Note that the default replacement value of 0.1 was explicitly used. This replacement does not fit well with the rest of the streamflow record, which is shown by figure 2. In this case, users may want to use a larger replacement value, such as 10 (see code below and figure 3). Alternatively, users could replace the zero values with NA and use the *fillMiss* function described next.

```
> q05054000Fix <- cleanUp(badDataSet, task = "fix", replace = 10)
```

```
> my.plot <- plotParam(q05054000Fix, code = "00060", stat = "00003", logscale = TRUE)
> print(my.plot)
```
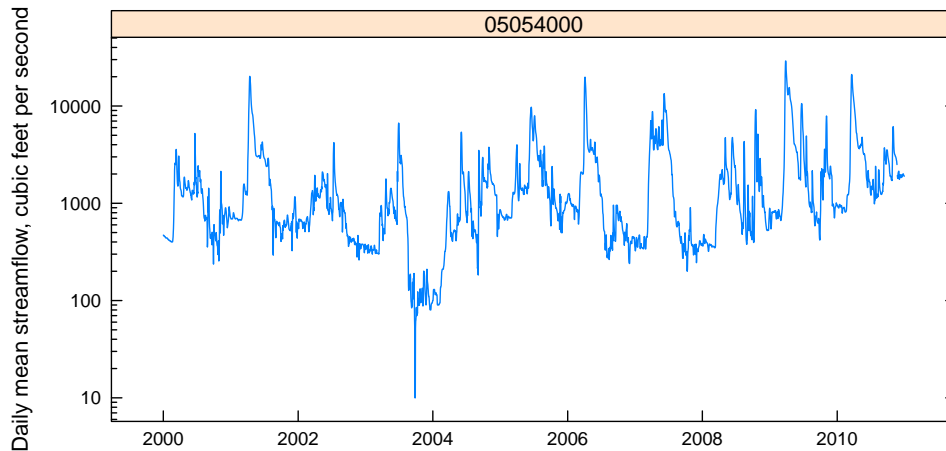

Figure 3: Daily streamflow with negative values changed to NA and 0 values replaced with 10.

The function *fillMiss* estimates values to replace NAs in the time series of hydrologic observations. The *fillMiss* function will check the percent of missing values and the size of the largest missing block of data. If there are large periods with missing values, the data may not be appropriate for analysis. If less than a user-specified percent (default is 40 percent) of the data are missing and the largest block is less than a user-specified number of days (default is 30), the data will be filled-in by using a structural time series, *StructTS* from the base **stats** package in **R** (R Development Team, 2012). The fitted structural time series is then smoothed via a state-space model, *tsSmooth* from the base **stats** package.

Many methods have been suggested for estimating streamflow; however, experiments showed that the functions in the base **stats** package worked well if the blocks of missing values were not long (less than 30 days). Users with larger blocks of missing values may want to explore other methods, including using nearby gages to estimate missing values at a streamgage. Additional methods for filling in missing hydrological data are summarized in Beauchamp (1989) and Elshorbagy and others (2000). An example follows for filling in missing values. Streamflow data for USGS streamgage 05054000, Red River of the North at Fargo, North Dakota, from January 1, 2000, to December 31, 2010, were downloaded and then 2,000 of the 4,018 daily values randomly were replaced with missing values. This provides a data set with slightly less than 50 percent of the values missing; however, the blocks of missing values generally are not very long. The data set is first summarized, showing the missing streamflow data.

```
> summary(misQ05054000)

     staid              val               dates            qualcode
 05054000:4018    Min.   :   46.0   Min.   :2000-01-01   A  :2705
                  1st Qu.:  522.0   1st Qu.:2002-10-01   A e:1313
                  Median :  939.5   Median :2005-07-01
```

```
            Mean    : 1797.8    Mean    :2005-07-01
            3rd Qu.: 1910.0    3rd Qu.:2008-03-31
            Max.    :27400.0   Max.    :2010-12-31
            NA's    :2000
```

Then, the function *fillMiss* is applied to the data. The process includes plotting the observed and estimated time series.

```
> my.newdata <- fillMiss(misQ05054000, block = 30, pmiss = 50, model = "trend",
+                        smooth = TRUE, log = "y")
```
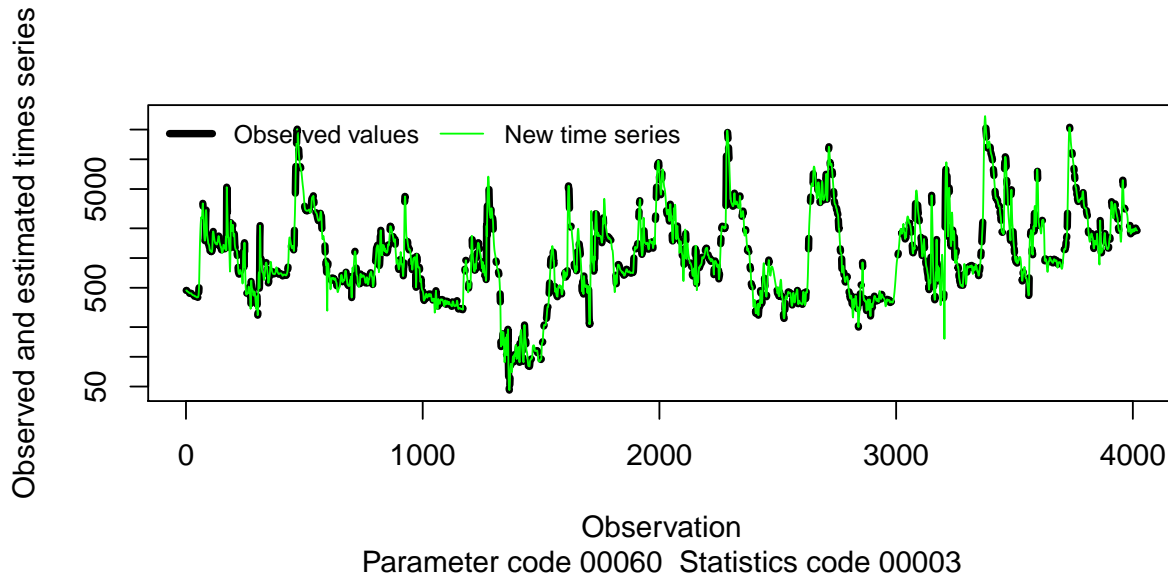

Figure 4: Plot of observed data, denoted by the black line in background, with data filled in for missing values, denoted by the green line in the foreground.

5   Data Summary

Summary statistics can be calculated for the daily streamflow series and a quantile-quantile plot may be used to check for approximate normality of the time series. For approximate normality, streamflow generally is transformed using the natural logarithm or base-10 logarithm. The following example applies the *summaryStats* function to the data and prints the results in an xtable, which can be used for LaTeX formatted tables.

```
> library(xtable)
> my.xtable <- xtable(summaryStats(q05054000Fix, staid = "05054000"),
+            cap = "Summary statistics for daily streamflow series.
+            Begin, the beginning date for the series; End, the ending
+            date for the series; n, the number of observations; NA, the
```

```
+               number of missing values; Neg, the number of negative values;
+               Min, the minimum value; Q1, the first quartile or 25th percentile;
+               Med, the median value; Mean, the mean value; Q3, the third
+               quartile or 75th percentile; Max, the maximum value; StdDev,
+               the standard deviation; IQR, the interquartile range.")
> print.xtable(my.xtable, size = c("scriptsize"))
```

	Begin	End	n	NA	Neg	Min	Q1	Med	Mean	Q3	Max	StdDev	IQR
05054000	2000-01-01	2010-12-31	4,018	1	0	10	523	923	1,797	1,900	29,100	2,647	1,377

Table 1: Summary statistics for daily streamflow series. Begin, the beginning date for the series; End, the ending date for the series; n, the number of observations; NA, the number of missing values; Neg, the number of negative values; Min, the minimum value; Q1, the first quartile or 25th percentile; Med, the median value; Mean, the mean value; Q3, the third quartile or 75th percentile; Max, the maximum value; StdDev, the standard deviation; IQR, the interquartile range.

One of the reasons for transforming hydrologic data with logarithms is to approximate normality, hence the need for replacing zero values with a small non-zero value. The effect of the logarithmic transformation is shown by contrasting figures 4 and 5.

```
> par(cex.lab = 0.9, las = 1, tcl = 0.5, xaxs = "r", yaxs = "r", cex.axis = 0.8)
> qqnorm(q05054000Fix$val)
> qqline(q05054000Fix$val, qtype = 9)
```
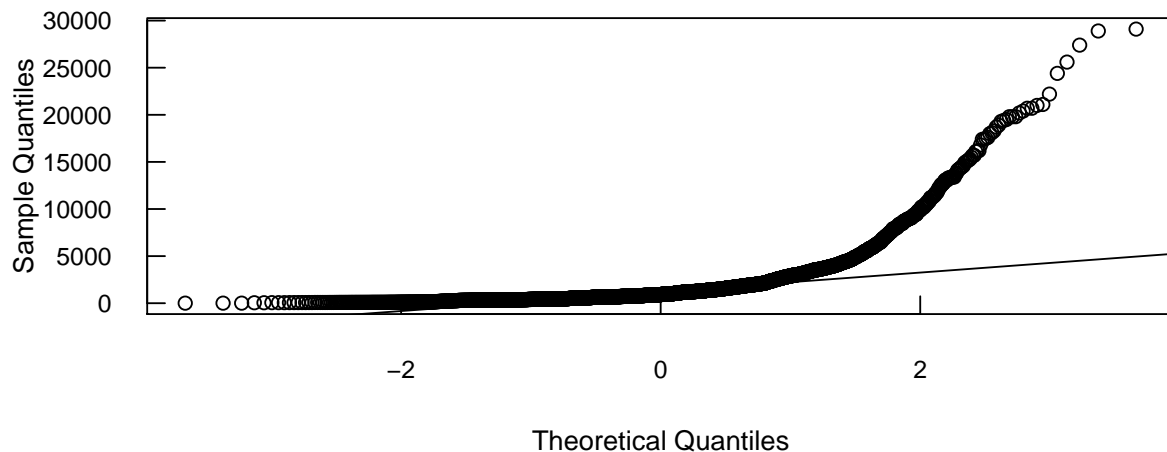
Normal Q–Q Plot

Figure 5: Quantile-quantile plot of streamflow to check for approximate normality. Streamflow is not log transformed and it is not approximated well by a normal distribution.

8

```
> par(cex.lab = 0.9, las = 1, tcl = 0.5, xaxs = "r", yaxs = "r", cex.axis = 0.8)
> qqnorm(log10(q05054000Fix$val))
> qqline(log10(q05054000Fix$val), qtype = 9)
```
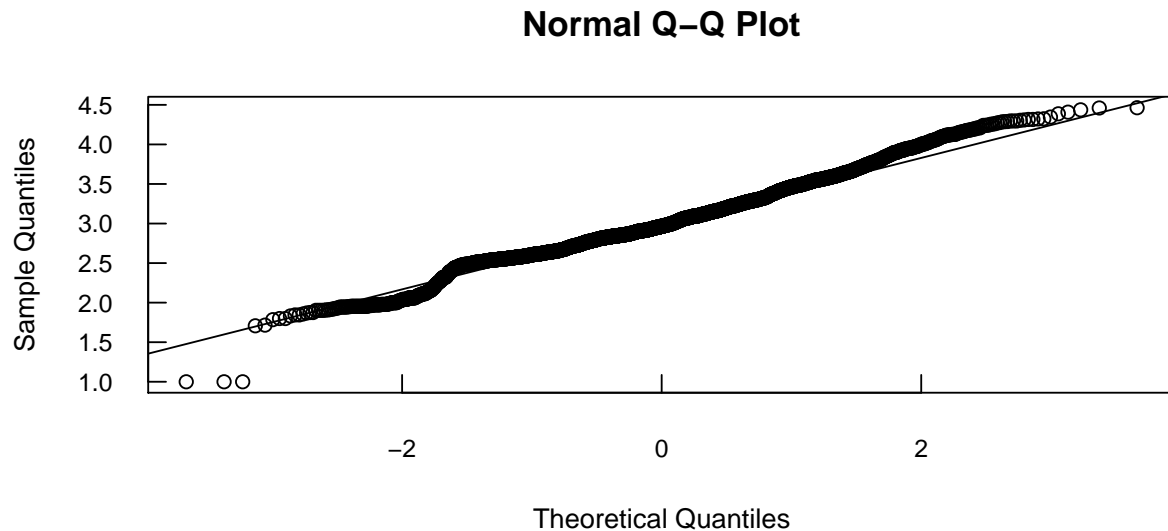

Figure 6: Quantile-quantile plot of the base-10 logarithm of streamflow to check for approximate normality. This shows that log-transformed streamflow is approximately normally distributed, with some low outliers.

6 Site Information

The user may also want additional information about streamgage sites, such as latitude and longitude. Additional information may be obtained from the USGS Site Information Service using the *siteInfo* function, as shown below. The *xtable* function can be applied to the results in an xtable, which can be used for LaTeX formatted tables.

```
> my.sites <- c("05054000", "05082500", "05061000", "05050000", "05058700",
+               "05267000", "06342500", "06478000", "06414000")
> my.siteInfo <- siteInfo(my.sites)
> xtable(my.siteInfo[order(my.siteInfo$staid),], cap = "Information for select
+       U.S. Geological Survey streamgage sites, sorted in downstream order.")
```

9

	staid	staname	lat	lng
4	05050000	BOIS DE SIOUX RIVER NEAR WHITE ROCK, SD	45.86	-96.57
1	05054000	RED RIVER OF THE NORTH AT FARGO, ND	46.86	-96.78
5	05058700	SHEYENNE RIVER AT LISBON, ND	46.45	-97.68
3	05061000	BUFFALO RIVER NEAR HAWLEY, MN	46.85	-96.33
2	05082500	RED RIVER OF THE NORTH AT GRAND FORKS, ND	47.93	-97.03
6	05267000	MISSISSIPPI RIVER NEAR ROYALTON, MN	45.83	-94.36
7	06342500	MISSOURI RIVER AT BISMARCK, ND	46.81	-100.82
9	06414000	RAPID CR AT RAPID CITY,SD	44.09	-103.24
8	06478000	JAMES R NEAR MITCHELL,SD	43.66	-97.92

Table 2: Information for select U.S. Geological Survey streamgage sites, sorted in downstream order.

The latitude and longitude information obtained using the *siteInfo* function may be used to plot a single site or multiple sites on a map.

```
> library(maps)
> library(mapdata)
> par(las = 1, tck = 0.02, mar = c(0, 0, 0, 0))
> map("state", region = c("minnesota", ".*dakota"))
> map("rivers", add = TRUE, col = 4)
> # label centered over gage site, jitter added to differentiate sites close
> # together
> mindif <- 0
> maxiterations <- 30
> iteration <- 1
> while (mindif < 0.085) {
+   y.offset <- as.numeric(my.siteInfo$lat) + runif(length(my.siteInfo$lat),
+                                                   0.12, 0.45)
+   mindif <- min(diff(unique(sort.int(round(as.numeric(y.offset), digits = 3)))))
+   iteration <- iteration + 1
+   if ( iteration >= maxiterations ) {
+     mindif <- 0.09
+     message("No ideal jitter found.  Some labels may conflict")
+   }
+ }
> points(my.siteInfo$lng, my.siteInfo$lat, pch = 19, col = "green")
>   text(xy.coords(my.siteInfo$lng, y.offset), my.siteInfo$staid, cex = 0.55)
> box()
> map.axes()
```
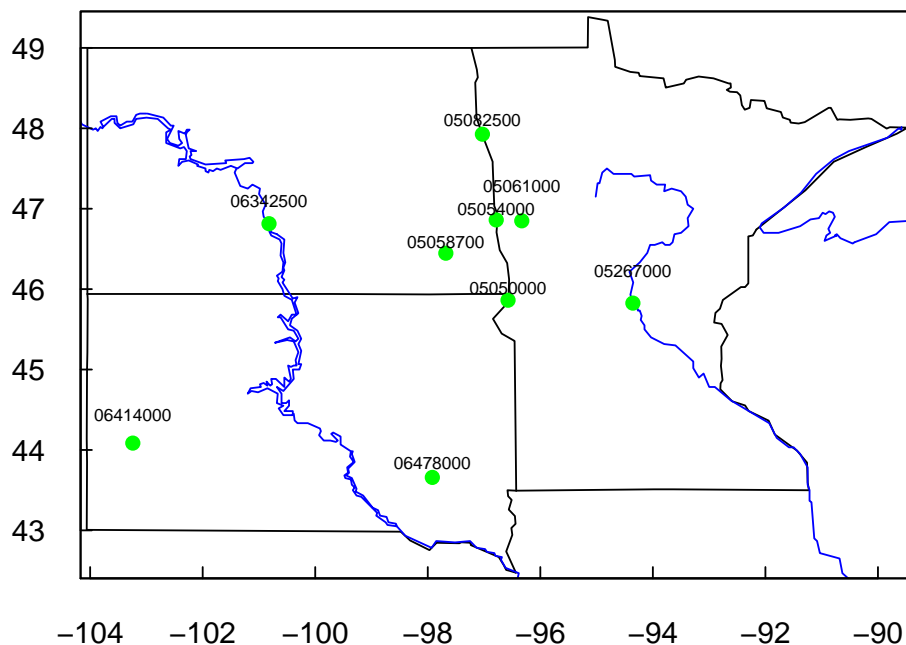

Figure 7: Map of selected U.S. Geological Survey streamgage sites.

11

One may also print the URL used by *siteInfo* to retrieve the site information.

```
> siteInfoURL <- tellMeSiteURL(c("05054000"))
```

The object "siteInfoURL" is now a character vector containing the following URL:
https://waterservices.usgs.gov/nwis/site?format=mapper&sites=05054000&siteOutput=expanded&
seriesCatalogOutput=true&outputDataTypeCd=all

7 Anomalies

Anomalies may be calculated spanning multiple time scales. For example, in an analysis of pesticide concentrations in urban streams (Ryberg and others, 2010), three streamflow anomalies were included in the time series model (SEAWAVE-Q; Sullivan and others, 2009) to help account for flow-related variability in pesticide concentrations. The calculation of those anomalies is described here to provide an example of a set of streamflow anomalies. Other time scales may be used. The time scales available in the package include a set of anomalies with the 1-year, 30-day, and 1-day time scales; a set with the 100-day, 10-day, and 1-day time scales; a set with the 30-day and 1-day time scales; and a set with the 10-year, 5-year, 1-year, one-quarter-year (seasonal), and 1-day time scales.

For example, using the 1-day, 10-day, and 100-day time scales, the anomalies can be computed using log-transformed daily flow. The first anomaly represents short-term (day-to-day) flow variability, and is defined using the following equation:

$$STFA(t) = X(t) - X_{10}(t) \tag{1}$$

where

$STFA(t)$ is the short-term flow anomaly (dimensionless) at time t;
$X(t)$ is log-transformed daily flow at time t; and
$X_{10}(t)$ is the average of log-transformed daily flow for 10 days up to and including time t.

Large positive values of STFA and associated increases in pesticide concentrations tend to occur near the beginning of a substantial rainfall-runoff event, whereas negative values of STFA and associated decreases in pesticide concentrations tend to occur after the event passes.

The second flow anomaly represents mid-term (10- to 100-day) flow variability and is defined using the following equation:

$$MTFA(t) = X_{10}(t) - X_{100}(t) \tag{2}$$

where

$MTFA(t)$ is the mid-term flow anomaly (dimensionless) at time t; and
$X_{100}(t)$ is the average of log-transformed daily flow for 100 days up to and including time t.

The third flow variable added to the model represents long-term (greater than 100-day) flow variability and is defined using the following equation:

$$LTFA(t) = X_{100}(t) - X*$$

(3)

where

$LTFA(t)$ is the long-term flow anomaly (dimensionless) at time t; and
$X*$ is the average of log-transformed daily flow for the specified trend assessment period.

Four different sets of anomalies may be calculated using the function *compAnom*. The "which" argument determines which set of anomalies is calculated. Setting "which" to 1 results in the 1-year, 30-day, and 1-day time scales; "which" equal to 2 results in a a set with the 100-day, 10-day, and 1-day time scales; "which" equal to 3 results in a a set with the 30-day and 1-day time scales; and "which" equal to 4 results in a set with the 10-year, 5-year, 1-year, one-quarter-year (seasonal), and 1-day time scales.

```
> anoms365.30.1 <- compAnom(q05054000, which = 1)
> anoms100.10.1 <- compAnom(q05054000, which = 2)
> anoms30.1 <- compAnom(q05054000, which = 3)
> anomsLT <- compAnom(q05054000, which = 4)
```

Plots of the first three possible sets of anomalies are shown in figures 8, 9, and 10. Ideally, the set of anomalies used for further applications would have similar standard deviations. Consideration should be given to gaps in the data (1 year of data is required to calculate the first 1-year anomaly, so gaps in daily streamflow will create larger gaps in the anomalies), size of the basin (the 100-, 10-, and 1-day set of anomalies may be better suited for smaller basins), and whether or not the streamgage is operated seasonally.
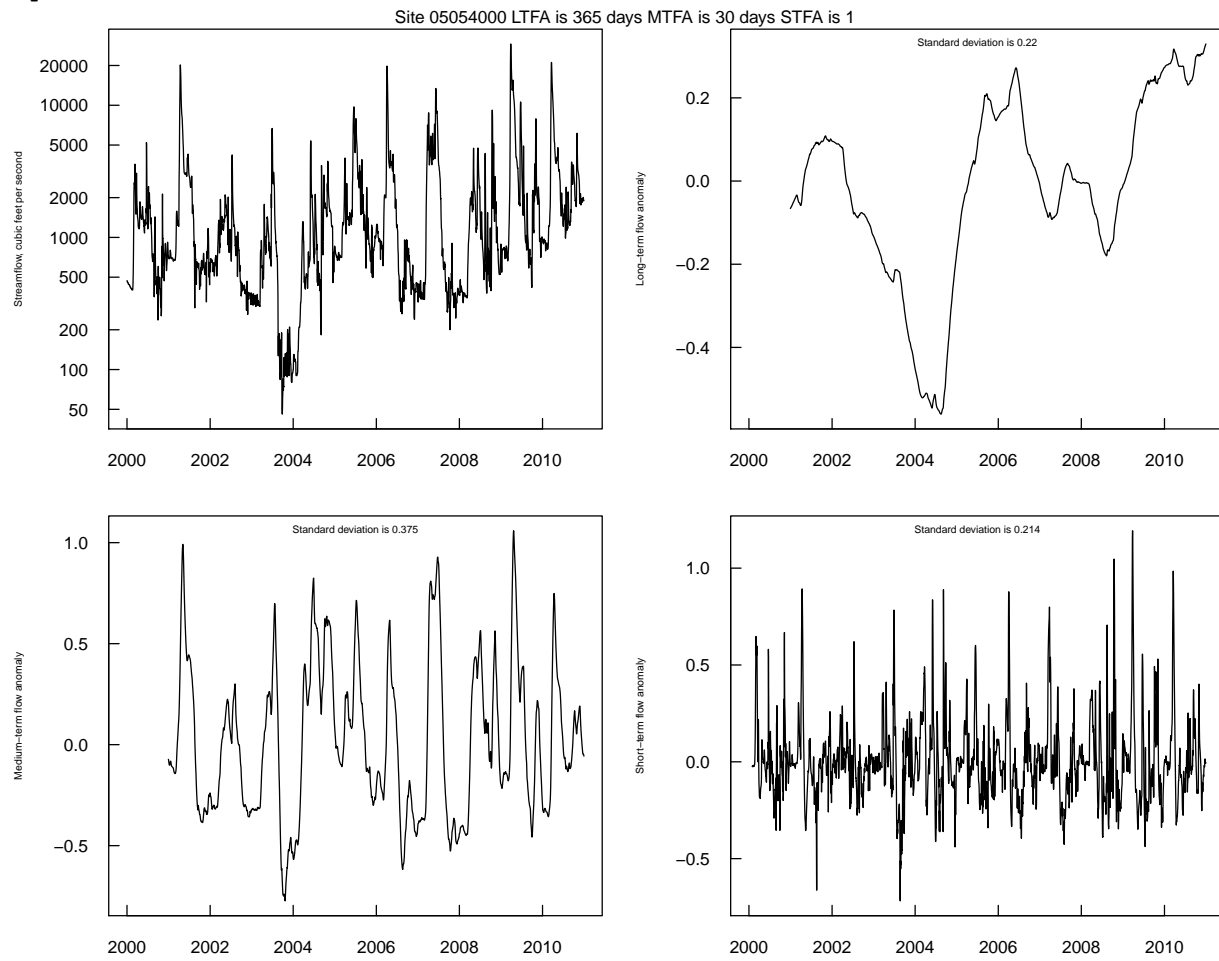
```
> plotAnoms(anoms365.30.1)
```


Figure 8: Streamflow and long-term, mid-term, and short-term anomalies computed using 365-day, 30-day, and 1-day time scales. These anomalies are best suited for sites with at least 10 years of data.
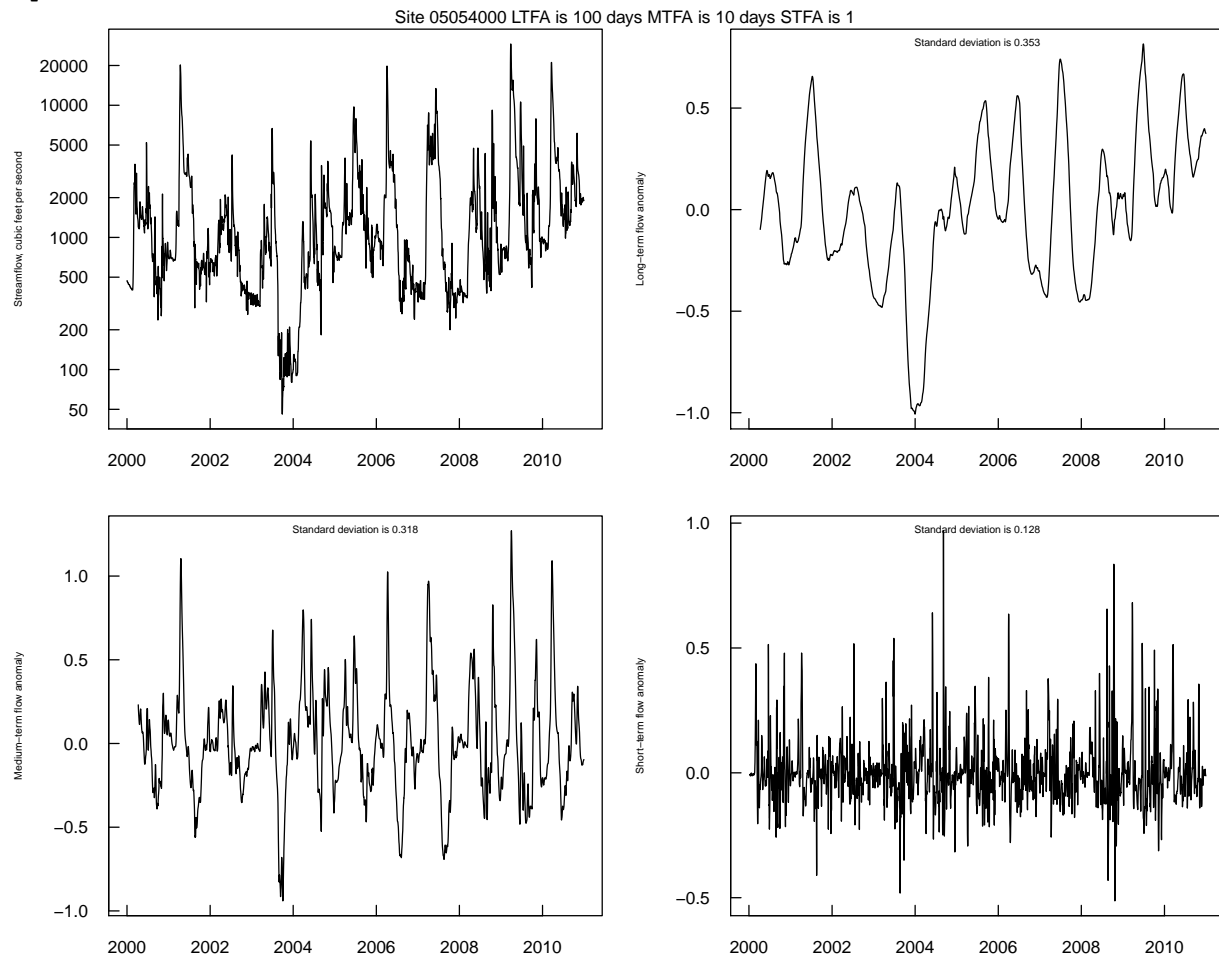
```
> plotAnoms(anoms100.10.1)
```


Figure 9: Streamflow and long-term, mid-term, and short-term anomalies computed using 100-day, 10-day, and 1-day time scales. These anomalies are best suited for sites with less than 10 years of data and for smaller drainage areas.
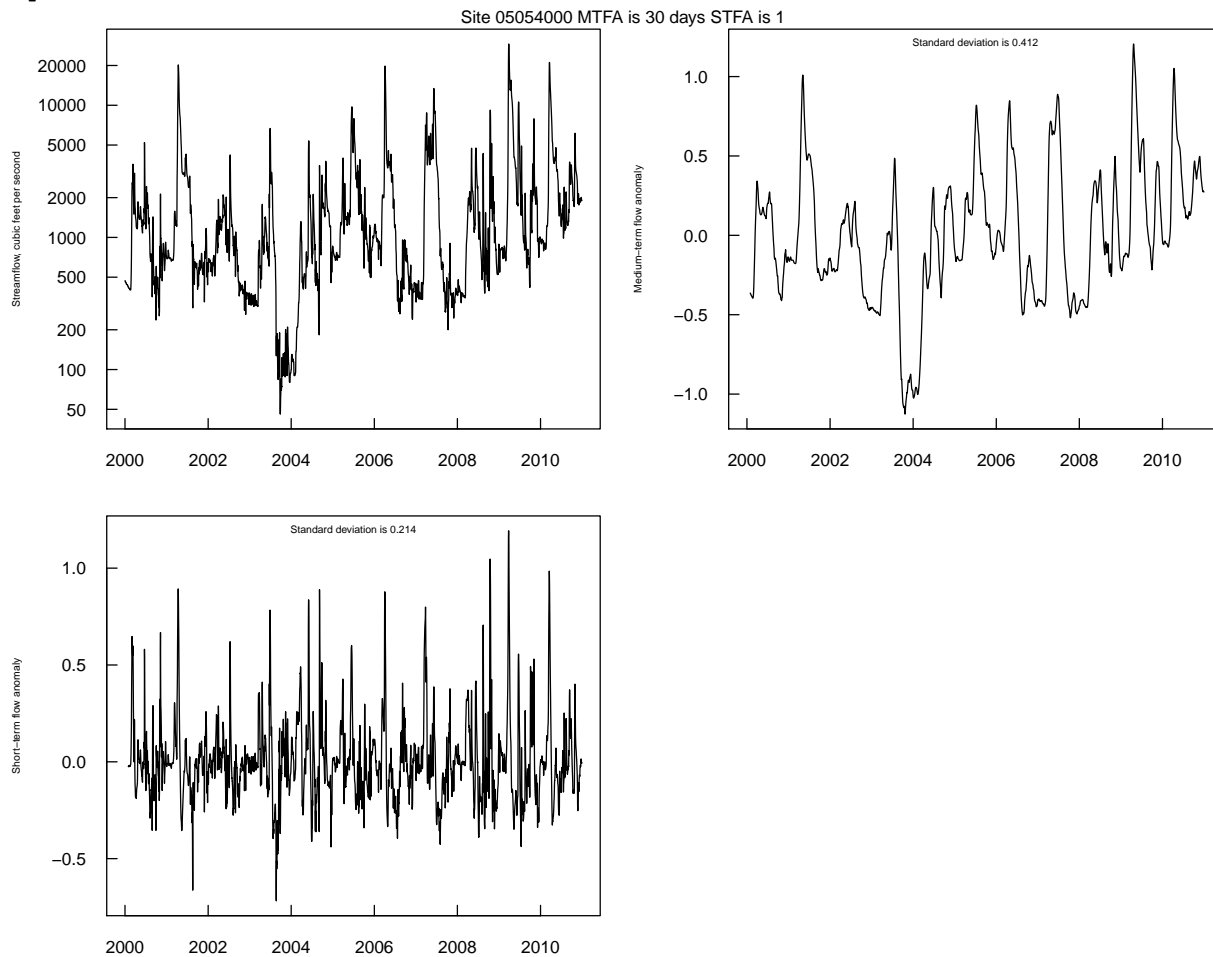
```
> plotAnoms(anoms30.1)
```

Figure 10: Streamflow, mid-term, and short-term anomalies calculated using 30-day and 1-day time scales. These anomalies are best suited for sites with seasonal operation.

The fourth possible set of anomalies requires long-term data sets with few, if any, missing values.

```
> # Import not run during vignette building because it may time out
> # over a slow connection, but this is the code to get the original data
> # q05054000LT<-importDVs("05054000", code="00060", stat="00003", sdate="1949-10-01",
> #                                                edate="2010-9-30")
> # data are available in example data set
> data(exampleWaterData)
> my.xtable <- xtable(summaryStats(q05054000LT, staid = "05054000"),
+                   cap = "Summary statistics for daily streamflow series.")
> print.xtable(my.xtable, size = c("footnotesize"))
```

	Begin	End	n	NA	Neg	Min	Q1	Med	Mean	Q3	Max	StdDev	IQR
05054000	1949-10-01	2010-09-30	22,280	0.0	0.0	0.1	218.0	463.5	1,014.6	996.5	29,100.0	1,873.1	778.2

Table 3: Summary statistics for daily streamflow series.

This series has some values of zero streamflow. Because the anomalies are based on logarithms of streamflow, the zero values are replaced with 0.1.

```
> q05054000LT <- cleanUp(q05054000LT, task = "fix")
> anomsLT <- compAnom(q05054000LT, which = 4)

> plotAnoms(anomsLT)
```
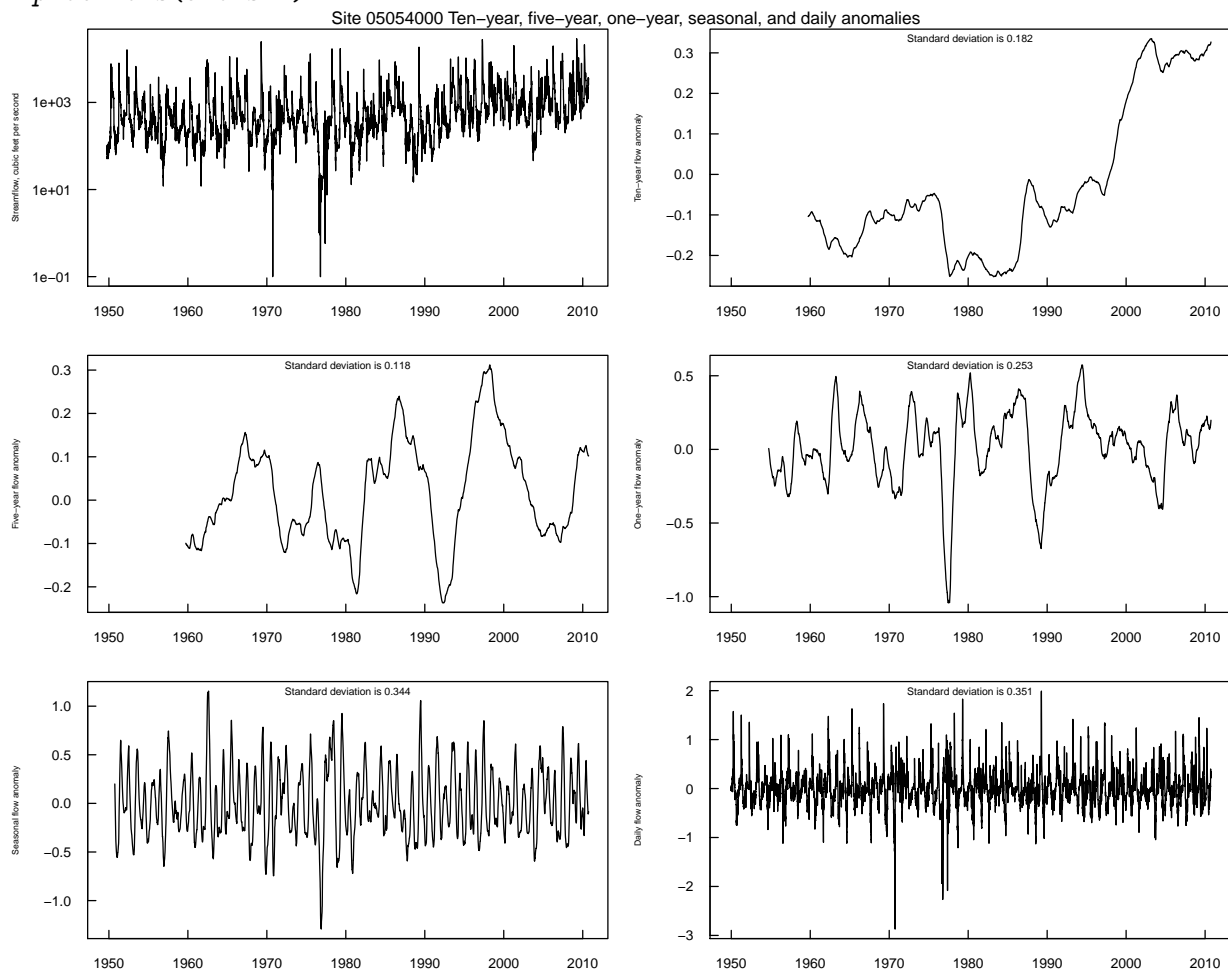

Figure 11: Streamflow and anomalies calculated using 10-year, 5-year, annual, and seasonal time scales. These anomalies are best suited for long-term streamgage sites and are not appropriate for sites with gaps in the data.

8 References

Alexander, R.B. and Smith, R.A., 2006, Trends in the nutrient enrichment of U.S. rivers during the late 20th century and their relation to changes in probably stream trophic conditions: Limnology and Oceanography, v. 51, no. 1, Part 2: Eutrophication of Freshwater and Marine Ecosystems, p. 639–654, accessed August 1, 2012, at http://www.jstor.org/stable/44996127.

Beauchamp, J.J., 1989, Comparison of regression and time-series methods for synthesizing missing streamflow records, Water Resources Bulletin, v. 25, no. 5, p. 961-975.

Elshorbagy, A.A., Panu, U.S., Simonovic, S.P., 2000, Group-based estimation of missing hydrological data—I. Approach and general methodology: Hydrological Sciences Journal, v. 45, no. 6, p. 849-866.

Helsel, D.R. and R. M. Hirsch, 2002, Statistical methods in water resources: U.S. Geological Survey Techniques of Water Resources Investigations, book 4, chap. A3, 522 p. (Also available at http://pubs.usgs.gov/twri/twri4a3/.)

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., 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/.

Sarkar, Deepayan, 2008, Lattice—multivariate data visualization with R: New York, Springer Science+Business Media, LLC, 268 p.

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/.

U.S. Geological Survey, 2017a, National Water Information System: Mapper, accessed January 3, 2017, at https://maps.waterdata.usgs.gov/mapper/index.html.

U.S. Geological Survey, 2017b, Parameter code definition, National Water Information System: Web Interface, accessed January 3, 2017, at http://nwis.waterdata.usgs.gov/usa/nwis/pmcodes.

U.S. Geological Survey, 2017c, Stat codes (stat_cd), National Water Information System: Web Interface, accessed January 3, 2017, at http://nwis.waterdata.usgs.gov/nwis/help/?read_file=stat&format=table.

U.S. Geological Survey, 2017d, USGS daily values site web service: REST Web Services, accessed January 3, 2017, at http://waterservices.usgs.gov/rest/DV-Service.html.

U.S. Geological Survey, 2017e, USGS site web service: REST Web Services, accessed January 3, 2017, at http://waterservices.usgs.gov/rest/Site-Service.html.

U.S. Geological Survey, 2017f, USGS surface-water daily data for the Nation: National Water Information System: Web Interface, accessed January 3, 2017, at http://waterdata.usgs.gov/nwis/dv/?referred_module=sw.

Valentine, David and Zaslavsky, Ilya, 2009, CUAHSI WATER ML 1.1 Specification Part 1— Introduction to WaterML Schema: San Diego Supercomputer Center, University of California at San Diego, 100 p., accessed May 21, 2012, at http://his.cuahsi.org/documents/WaterML_1_1_part1_v2.pdf.

Vecchia, A.V., 2003, Relation between climate variability and stream water quality in the Continental United States: Hydrological Science and Technology, v. 19, no. 1, p. 77–98.

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.