Subdaily Time-Step Models

Dave Lorenz

December 3, 2015

Abstract

This example illustrates how to set up and build a rating curve model for a sub-daily time step. Sub-daily time steps can be used for small streams that have peak durations less than one day and have sampling schemes that measure the mean concentration over part or all of the peak flow. This example build a simple model and assumes that the user is familiar with the general steps in build a rating curve model using the functions in **rloadest**.

Contents

1 Introduction **2**

2 Determine the Time Step **4**

3 Build the Model **5**

4 Load Estimation **10**

1 Introduction

This example builds on and extends the analysis of Landers and others (2007) and Joiner and others (2014) of whole-water ammonia plus organic nitrogen in the Brushy Fork Creek at Beaver Road near Loganville, Georgia, USGS streamgage 02207400. This example covers the period beginning in water year 2005 through water year 2013.

```
> # Load the necessary packages and the data
> library(rloadest)
> library(dataRetrieval)
> # Get the QW data
> Brushy <- "02207400"
> # Parameter code 00665 is whole-water phosphorus
> # Parameter code 71123 is mean streamflow for peak flow
> # Parameter code 00060 is the daily mean streamflow
> # Parameter code 00061 is the measured streamflow
> # 71123 will be used as the flow for peak flows and 00061
> # for base flow or 00060 if 00061 is missing.
> BrushyQW <- importNWISqw("02207400",
+   params=c("00625", "72123", "00060", "00061"),
+   begin.date="2004-10-02", end.date="2013-09-30")
> # Convert the separate columns of dates and times to a single column
> # Uses functions from smwrBase
> BrushyQW <- transform(BrushyQW,
+   StartDateTime = setTZ(sample_dt + as.timeDay(sample_tm), tzone_cd, force.stz=TRUE),
+   EndDateTime = setTZ(sample_end_dt + as.timeDay(sample_end_tm), tzone_cd, force.stz=TRUE))
> # A few rows of data:
> head(BrushyQW)

    site_no  sample_dt sample_tm tzone_cd medium_cd sample_end_dt sample_end_tm Discharge_cfs
13 02207400 2004-10-05     09:35      EST        WS          <NA>                          NA
16 02207400 2004-10-05     09:36      EST        WS          <NA>                          NA
17 02207400 2004-12-21     11:10      EST        WS          <NA>                          NA
20 02207400 2004-12-21     11:11      EST        WS          <NA>                          NA
75 02207400 2005-01-13     21:33      EST        WS    2005-01-14         06:26           NA
65 02207400 2005-02-28     02:33      EST        WS    2005-02-28         08:04           NA
   InstDischarge_cfs Kjeldahl_WW.N.00625 DischargeMeanStorm_cfs       StartDateTime
13               7.8               E0.43                      8 2004-10-05 09:35:00
16               7.8                  NA                     NA 2004-10-05 09:36:00
17              12.0               E0.31                     12 2004-12-21 11:10:00
20              12.0                  NA                     NA 2004-12-21 11:11:00
75                NA                0.64                     49 2005-01-13 21:33:00
65                NA                0.56                     42 2005-02-28 02:33:00
           EndDateTime
13                <NA>
16                <NA>
17                <NA>
20                <NA>
75 2005-01-14 06:26:00
65 2005-02-28 08:04:00
```

```
> # Subset to remove the missing values in Kjeldahl_WW.N.00625
> # Note that this works only because we are interested in Kjeldahl_WW.N.00625
> BrushyQW <- subset(BrushyQW, !is.na(Kjeldahl_WW.N.00625))
```

2 Determine the Time Step

There is no set method to determine the time step. The approach should determine a typical peak duration. For new sites, with little streamflow record, the user must examine the history of peaks and determine a sampling strategy that can characterize the peak flow. Landers and others (2007) has a brief description of the process. For the Brushy Fork Creek data, with a fairly long history of composite sample collection designed to characterize the peak flow volume, the median duration of the composite sample can be used. The median duration, from the code shown following this paragraph is 352 minutes—the closest even divisor into 24 hours is 6 hours. The code following this paragraph also computes the sample date and time and flow for the regression model.

```
> # Compute the median peak duration.
> with(BrushyQW, median(EndDateTime - StartDateTime, na.rm=T))

Time difference of 352 mins

> # Compute the sample date and time
> BrushyQW <- transform(BrushyQW,
+   dateTime=ifelse(is.na(EndDateTime), StartDateTime,
+     StartDateTime + (EndDateTime - StartDateTime)/2))
> # Need to convert to POSIXct. Note that the original data were
> # recorded in standard time only, so the correct time zone is
> # "America/Jamaica," which preserves the correct time offset.
> BrushyQW <- transform(BrushyQW, dateTime=as.POSIXct(dateTime,
+   origin="1970-01-01", tz="America/Jamaica"))
> # Now the flow, coalesce is in smwrBase
> BrushyQW <- transform(BrushyQW,
+   Flow=ifelse(!is.na(EndDateTime), DischargeMeanStorm_cfs,
+   coalesce(InstDischarge_cfs, Discharge_cfs)))
```

3 Build the Model

Surrogate data including turbidity and specific conductance are available for this site, but only as daily values. If those unit or instantaneous values were available, then it would be best to examine models that included surrogate data. The model must be built using streamflow and time. The code following this paragraph builds the 6-hour time step model, using the best-model selection process.

```
> # Find the "best" predefined model
> Brushy.lreg <- selBestModel("Kjeldahl_WW.N.00625", BrushyQW, flow="Flow",
+   dates="dateTime", time.step="6 hour", station=Brushy)
> print(Brushy.lreg)

*** Load Estimation ***

Station: 02207400
Constituent: Kjeldahl_WW.N.00625

           Number of Observations: 89
Number of Uncensored Observations: 59
           Center of Decimal Time: 2009.234
                  Center of ln(Q): 2.4272
                 Period of record: 2004-10-05 09:35:00 to 2013-09-10 10:15:00

Model Evaluation Criteria Based on AMLE Results
-----------------------------------------------

  model   AIC  SPCC  AICc
1     1 178.2 185.6 178.4
2     2 180.1 190.0 180.5
3     3 179.1 189.1 179.6
4     4 180.0 192.4 180.7
5     5 181.0 193.5 181.7
6     6 181.8 196.8 182.8
7     7 181.4 196.3 182.4
8     8 183.2 200.6 184.6
9     9 184.9 204.9 186.7
Model # 1 selected

Selected Load Model:
--------------------

Kjeldahl_WW.N.00625 ~ model(1)

Model coefficients:
            Estimate Std. Error z-score p-value
(Intercept)    2.534    0.09485   26.72       0
lnQ            1.338    0.07526   17.77       0

AMLE Regression Statistics
```

```
Residual variance: 0.5673
Generalized R-squared: 76.57 percent
G-squared: 129.1 on 1 degrees of freedom
P-value: <0.0001
Prob. Plot Corr. Coeff. (PPCC):
  r = 0.9556
  p-value = 0.0011
Serial Correlation of Residuals: 0.0604


Comparison of Observed and Estimated Loads
------------------------------------------
      Summary Stats: Loads in kg/d
---------------------------------------------
     Min  25%  50%  75% 90% 95% Max
Est 0.31 6.79 28.7 84.0 151 226 563
Obs 0.12 4.77 17.3 67.8 167 227 983

Bias Diagnostics
----------------
 Bp: 0.1471 percent
PLR: 1.001
  E: 0.4428
```

Model 1 was selected—including only flow. The residual variance is quite large at 0.5673, but the printed diagnostics in general indicate an acceptable model. The diagnostic plots in figures 1-3 also indicate an acceptable model: the fit is reasonably linear and appear to have uniform scatter, there are no issues with serial correlation, and the standardized residuals fall very near the 1:1 line.

```
> # Set up for graph in vignette
> setSweave("graph01", 6, 6)
> plot(Brushy.lreg, which = 1, set.up=FALSE)
> dev.off()

null device
          1
```

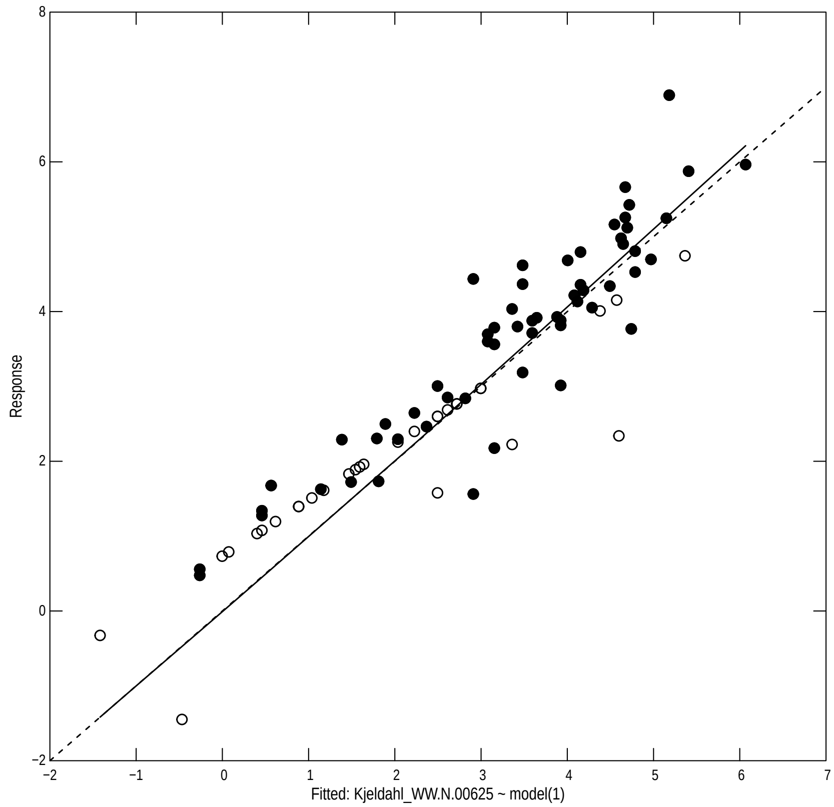

Figure 1. The overall fit.

```
> # Set up for graph in vignette
> setSweave("graph02", 6, 6)
> plot(Brushy.lreg, which = 4, set.up=FALSE)
> dev.off()

null device
          1
```

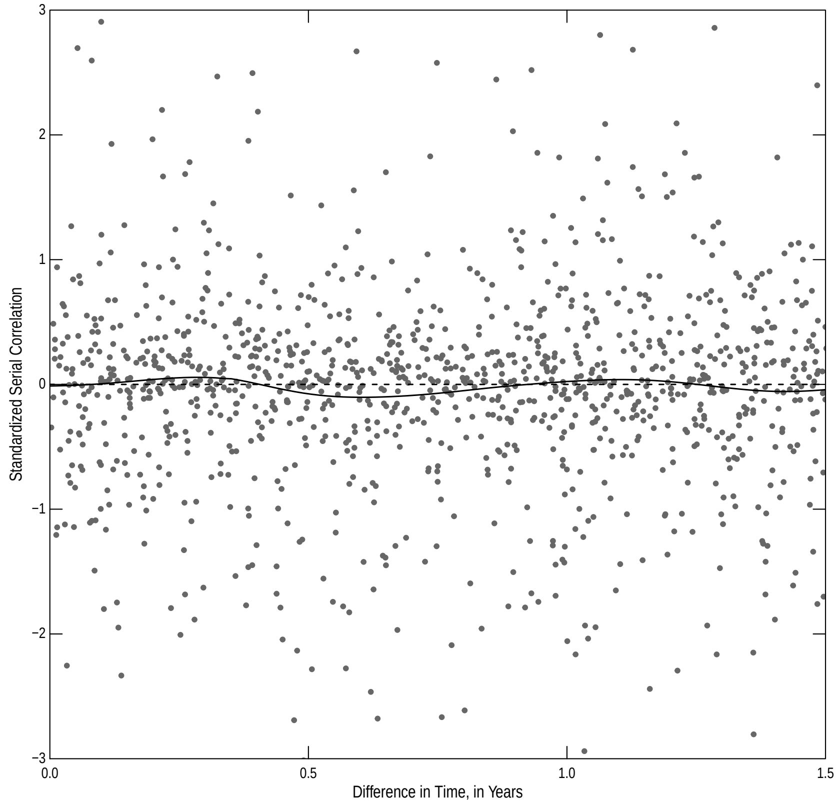

Figure 2. The correlogram.

```
> # Set up for graph in vignette
> setSweave("graph03", 6, 6)
> plot(Brushy.lreg, which = 5, set.up=FALSE)
> dev.off()

null device
          1
```

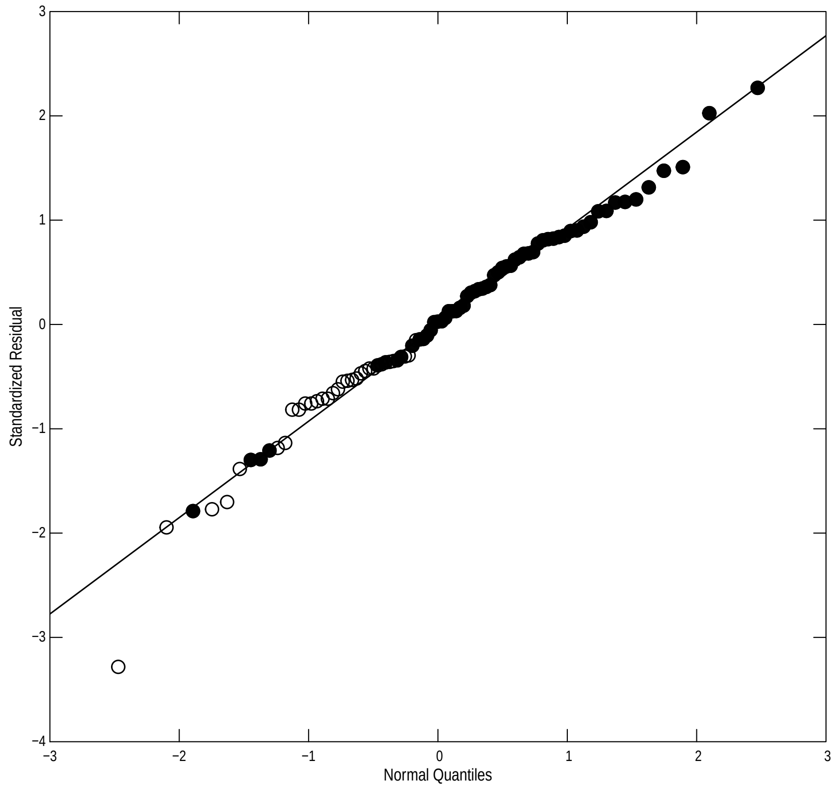

Figure 3. The q-normal plot.

As a final check on the simple model, compute the jackknife estimates of the parameters. The code following this paragraph demonstrates the use of the `jackStats` function. The relative bias is small, not indicating a problem with the model.

```
> # Jackknife statistics
> jackStats(Brushy.lreg)


jackknife estimates:

Coefficients:
            Estimate       Bias Std. Error Rel. Bias
(Intercept)    2.534 -0.0012435    0.11311  0.010994
lnQ            1.338  0.0002289    0.08365  0.002736
```

4 Load Estimation

Load estimation can easily be done using unit-value data, rather than the user aggregating data by the time step. The code following this paragraph demonstrates a simple load estimation for the month of May, 2013. The data are retrieved, making sure the time zone is set to the local time. The unit-value data are aggregated into 6-hour time steps, taking all of the unit values within the time period; a complete record is not required, but each 6-hour period must have at least one unit value. The Brushy Fork Creek data do not have any unit values for 5:15 through 6:00 on May 22, as indicated in the code below. The user must check for gaps in the record; `predLoad` will fail if there are gaps greater than the time step.

```
> # Get the data for May, 2013
> BrushyQ <- readNWISuv(Brushy, "00060", startDate="2013-05-01",
+   endDate="2013-05-31", tz="America/New_York")
> # Rename the Flow column, must be done manually as renameNWISColumns
> # appends _Inst
> names(BrushyQ)[5] <- "Flow"
> # Show the gap in the record:
> BrushyQ[2030:2040,]

     agency_cd  site_no            dateTime            tz_cd Flow X_00060_00011_cd
2030      USGS 02207400 2013-05-22 03:15:00 America/New_York   15                A
2031      USGS 02207400 2013-05-22 03:30:00 America/New_York   15                A
2032      USGS 02207400 2013-05-22 03:45:00 America/New_York   15                A
2033      USGS 02207400 2013-05-22 04:00:00 America/New_York   15                A
2034      USGS 02207400 2013-05-22 04:15:00 America/New_York   15                A
2035      USGS 02207400 2013-05-22 04:30:00 America/New_York   15                A
2036      USGS 02207400 2013-05-22 04:45:00 America/New_York   15                A
2037      USGS 02207400 2013-05-22 05:00:00 America/New_York   15                A
2038      USGS 02207400 2013-05-22 06:15:00 America/New_York   15                A
2039      USGS 02207400 2013-05-22 06:30:00 America/New_York   15                A
2040      USGS 02207400 2013-05-22 06:45:00 America/New_York   15                A

> # predict the load
> predLoad(Brushy.lreg, BrushyQ, by="month")

    Period Ndays    Flux  Std.Err      SEP      L95      U95
1 May 2013    31 44.7511 5.692819 15.65127 21.70634 82.20623
```

References

[1] Joiner, J.K., Aulenbach, B.T., and Landers, M.N., 2014, Watershed characteristics and water-quality trends and loads in 12 watersheds in Gwinnett County, Georgia: U.S. Geological Survey Scientific Investigations Report 2014-5141, 79 p., http://dx.doi.org/10.3133/sir20145141.

[2] Landers, M.N., Ankcorn, P.D., and McFadden, K.W., 2007, Watershed effects on streamflow quantity and quality in six watersheds of Gwinnett Count, Georgia: U.S. Geological Survey Scientific Investigations Report 2007-5132, 62 p., Web-only publication at http://pubs.usgs.gov/sir/2007/5132/