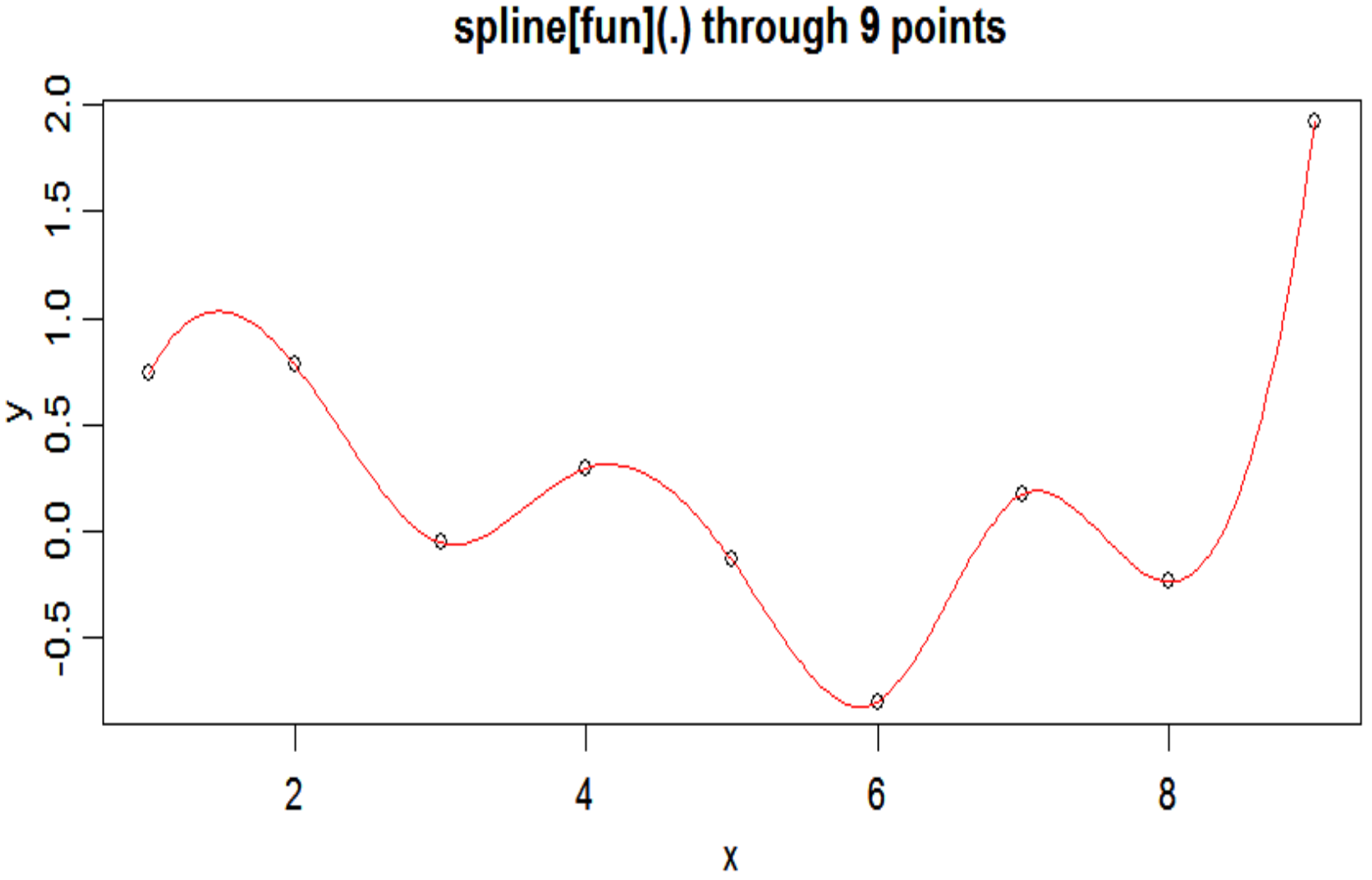
Using Generalized Additive Models for Trend Analysis

Elgin Perry
eperry@chesapeake.net
12 / 18/ 2013

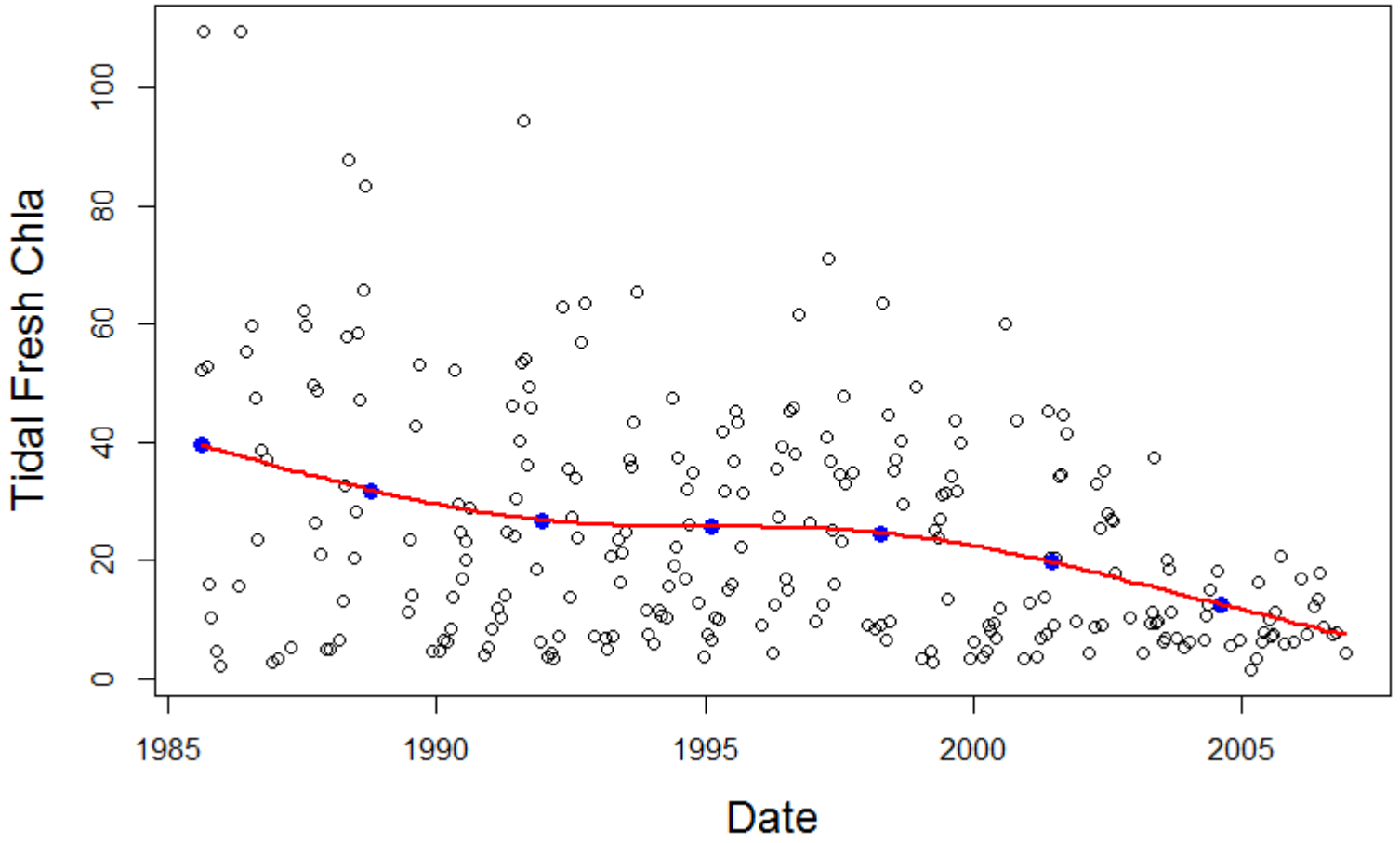
Road Map:

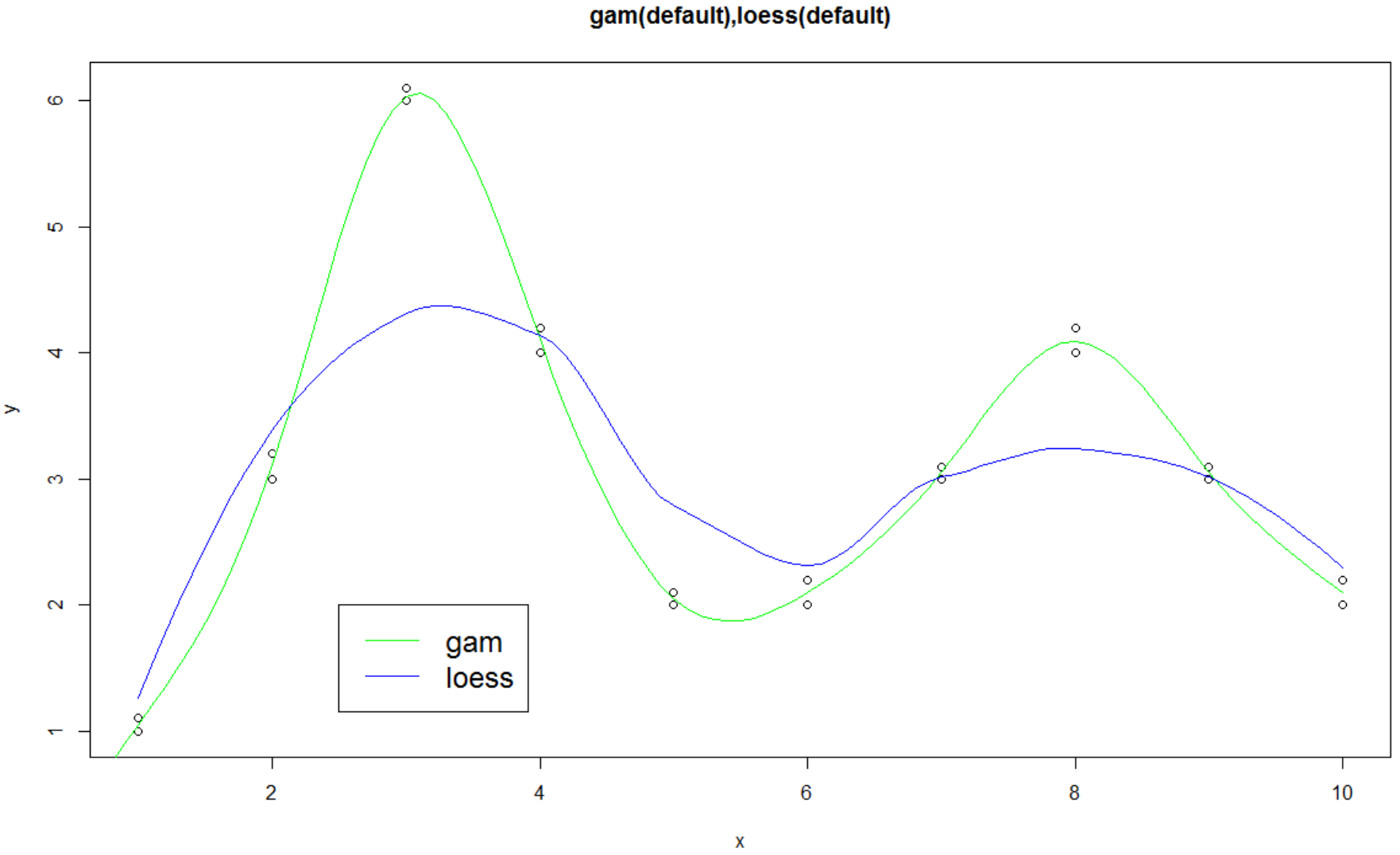
1. Heuristic development of GAM
2. Comparison of GAM and LOESS
3. Quick review of GAM tools
4. Trend example using GAM

Illustration of using a cubic spline interpolator to produce a smooth curve passing through 9 points. Each circle indicates a transition point.

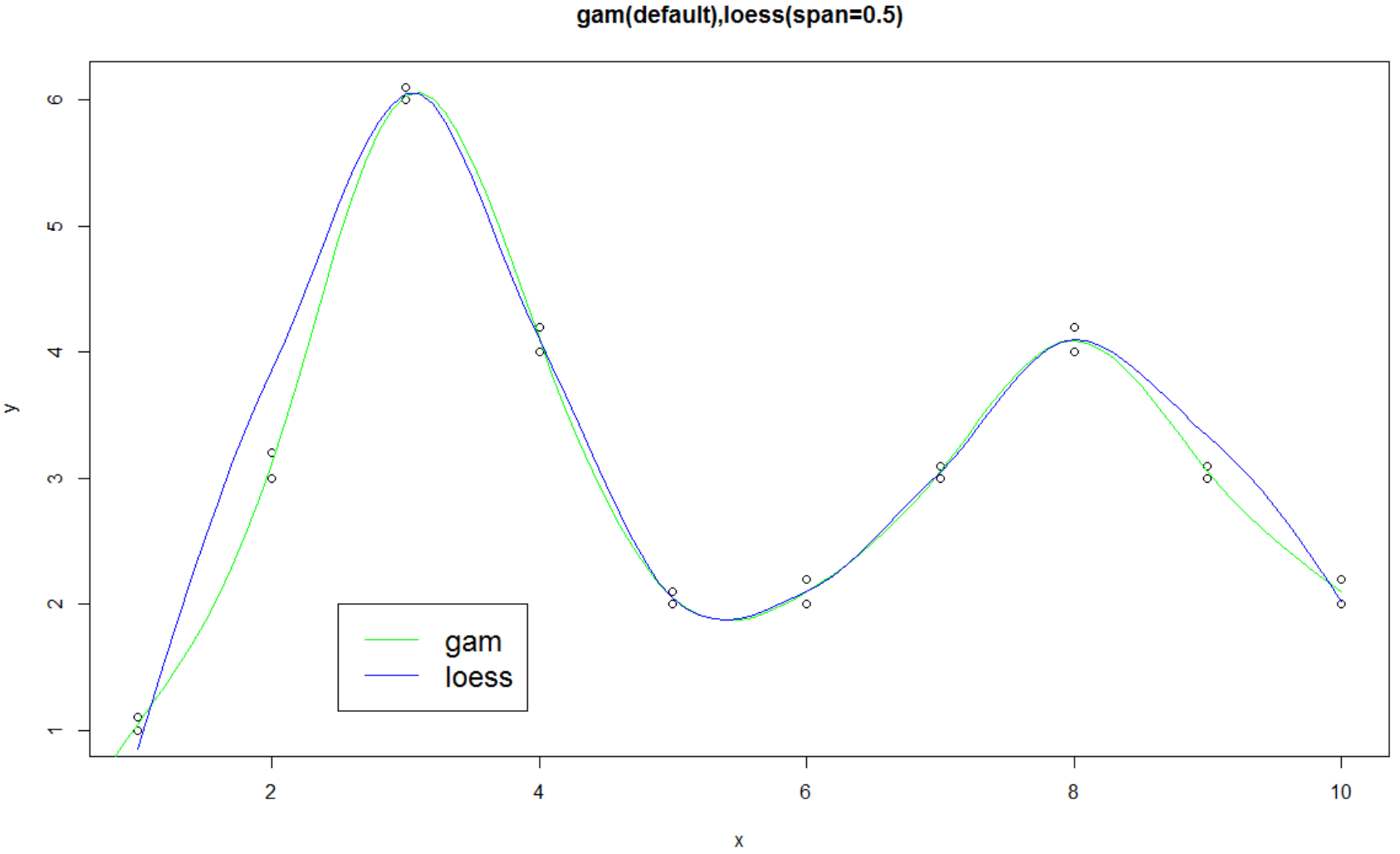


Use knots (blue dots) to define number of intervals for cubic polynomials and fit using least squares.





GAM and LOESS fits to same data using default parameters.



With a slight tweek to the loess, it can be made to produce predictions that are very similar to the gam.

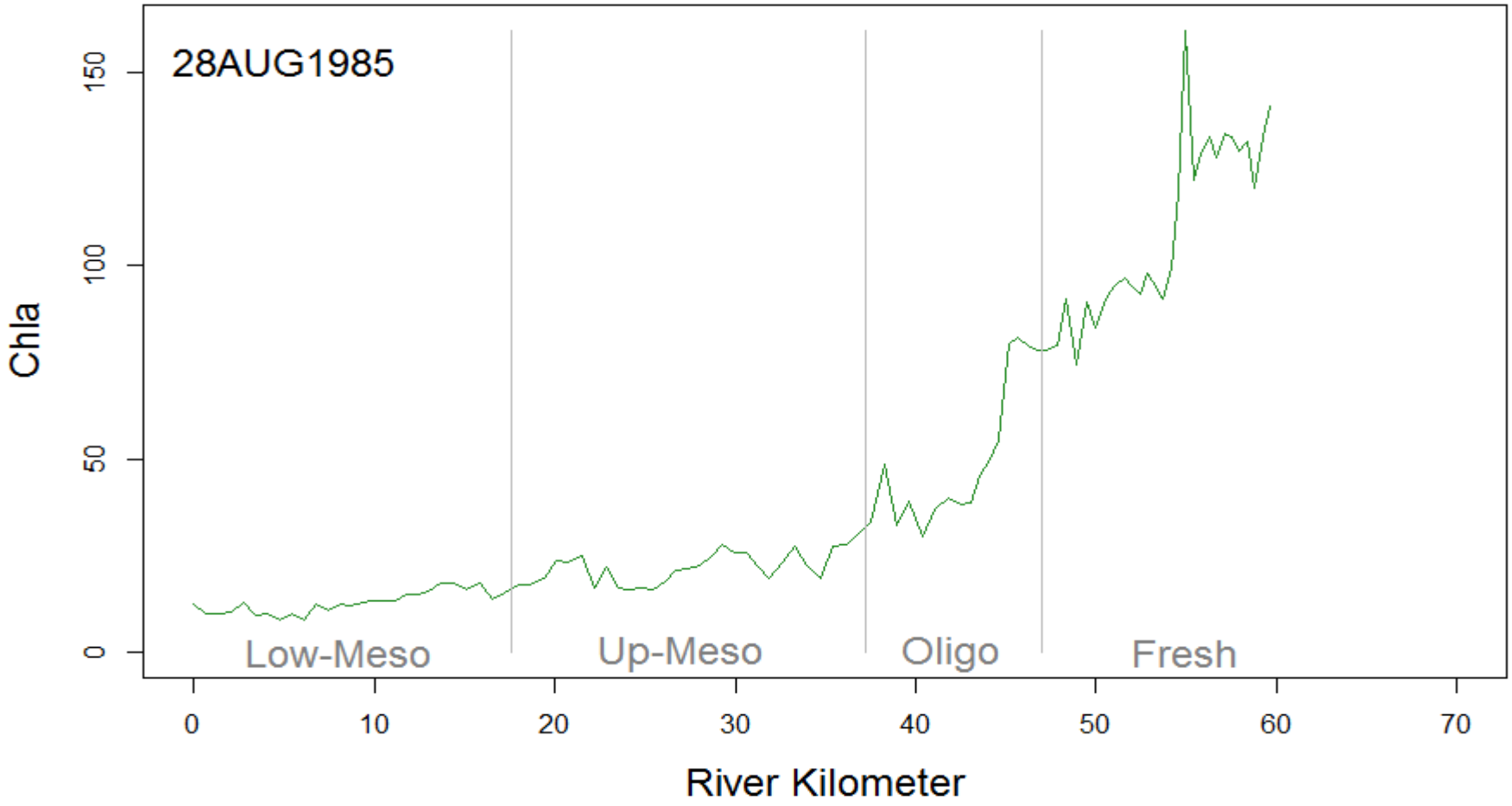
Geek Slide

```
library(mgcv)    #Wood's gam package


tngam1 <- gam(LOG_TN ~ YEAR,  data=tn)

tngam2 <- gam(LOG_TN ~ s(YEAR),  data=tn)

tngam3 <- gam(LOG_TN ~ s(YEAR) + as.factor(month),  data=tn)

tngam4 <- gam(LOG_TN ~ s(YEAR) + (doy,bs="cc"),  data=tn)

tngam5 <- gam(LOG_TN ~ s(YEAR) + s(doy,bs="cc") +
s(LOG_SRF),data=tn)

tngam6 <- gam(LOG_TN ~ tp(YEAR,LOG_SRF) + s(doy,bs="cc"),data=tn)

ptngam6 <- predict(tngam6,newdata=ptn, se.fit=TRUE)
```

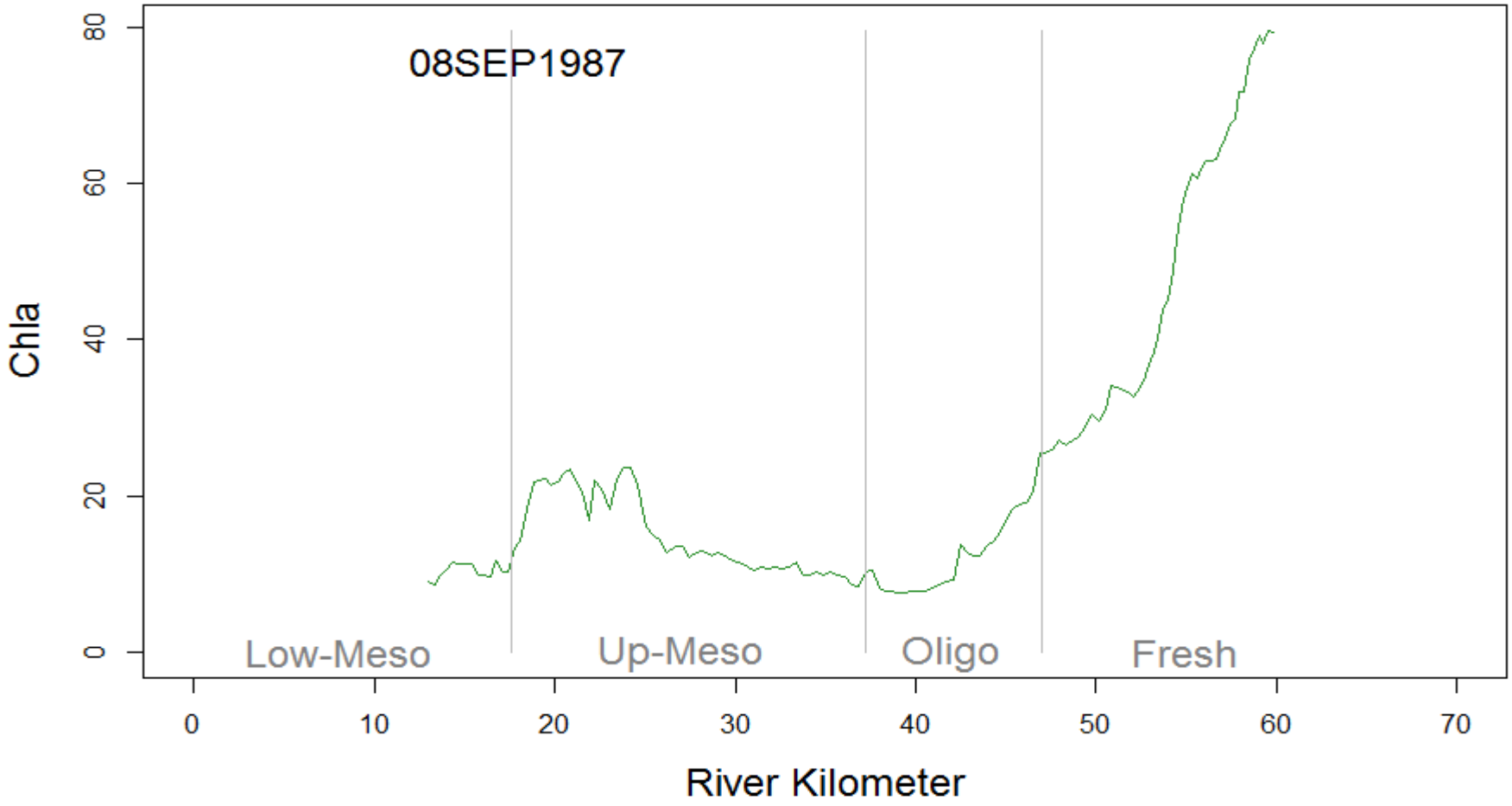
Steps

1. Look at response data
2. Look at data relative to forcing functions
3. Develop quantitative measures of forcing functions
4. GAM development.

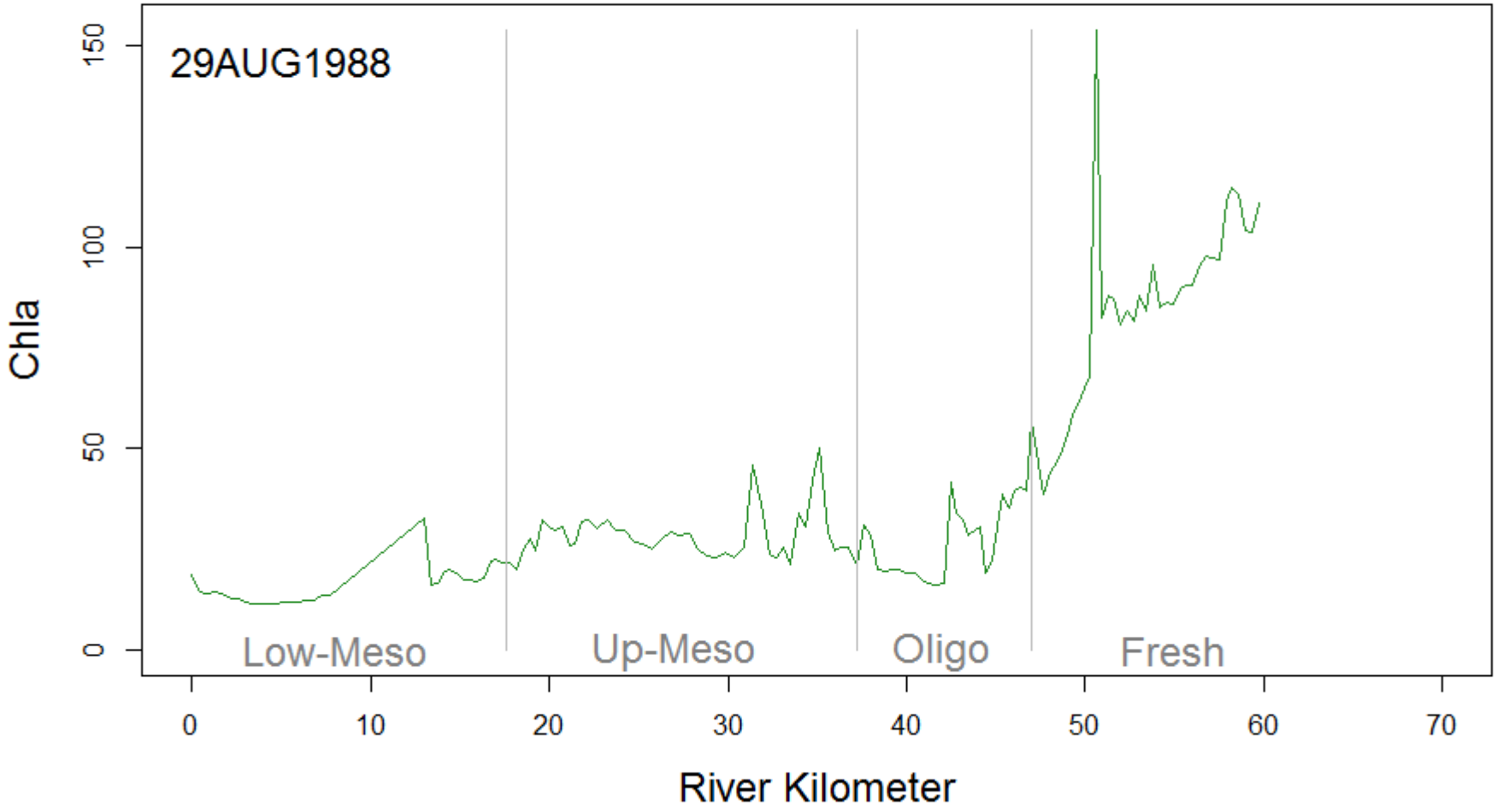
Response data (388 dates)



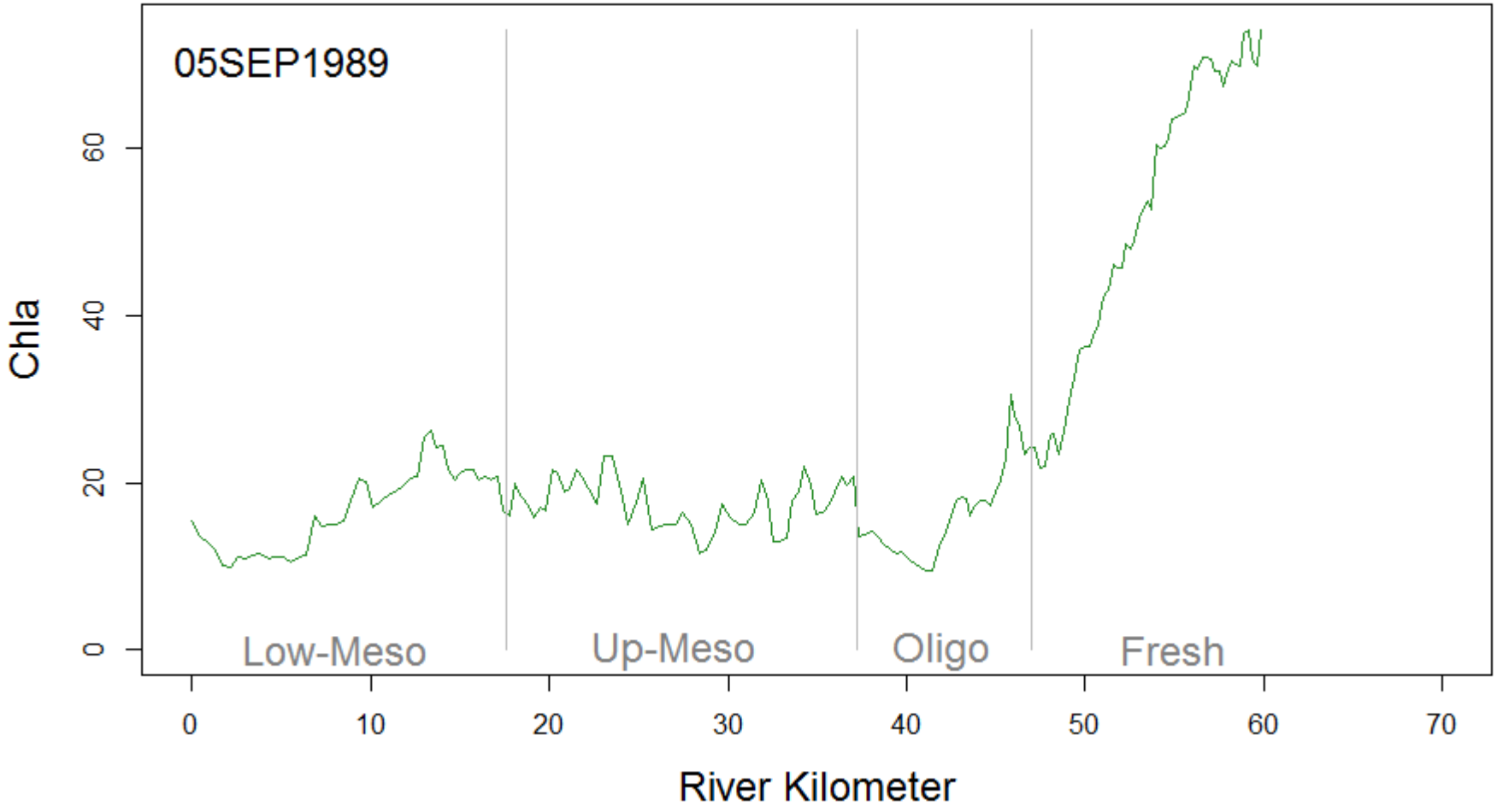
Late Summer


08SEP1987
Chla
0
20
40
60
80
Low-Meso
Up-Meso
Oligo
Fresh
0
10
20
30
40
50
60
70
River Kilometer

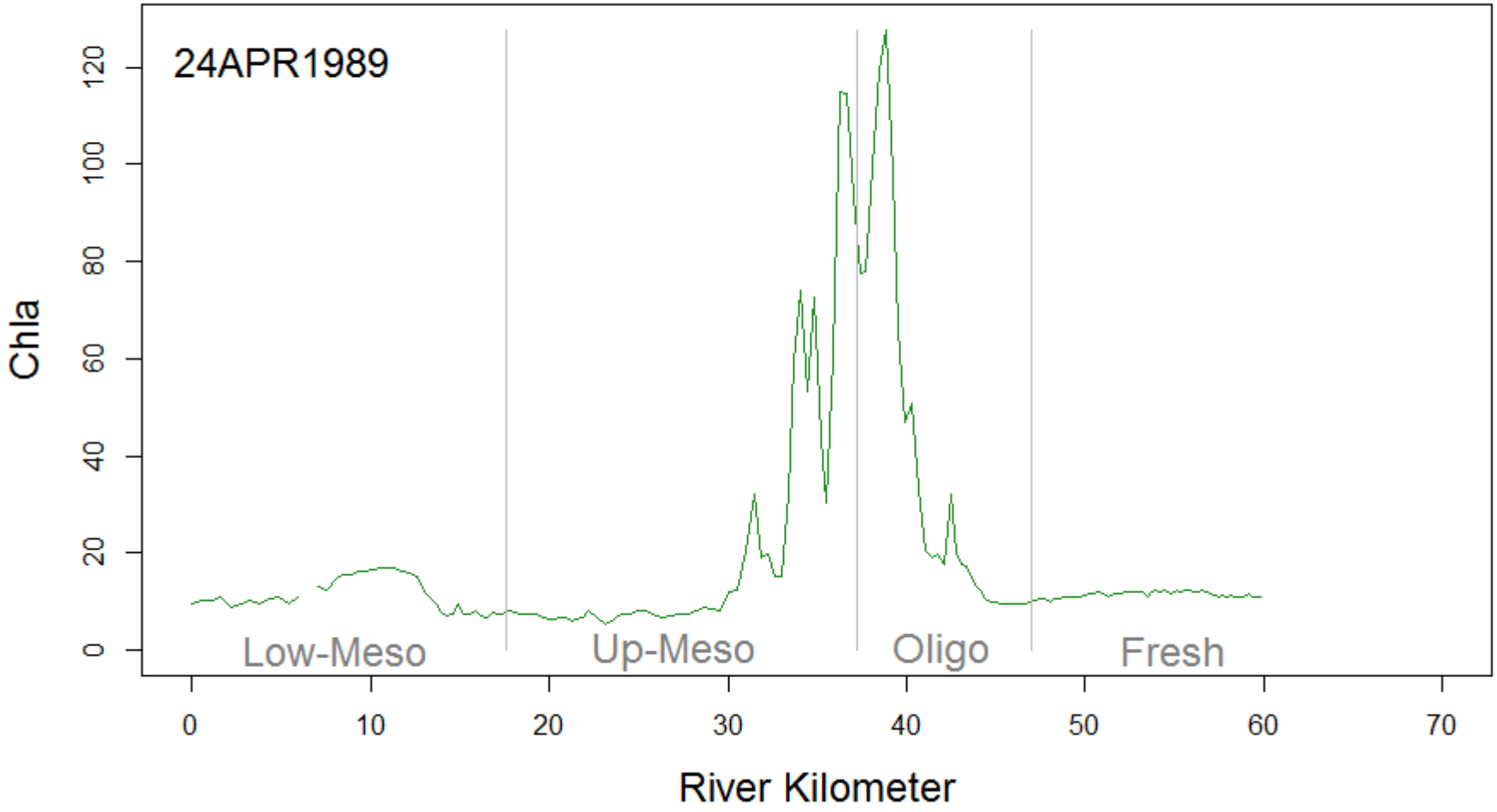
Late Summer


29AUG1988
Chla
0
50
100
150
Low-Meso
Up-Meso
Oligo
Fresh
0
10
20
30
40
50
60
70
River Kilometer

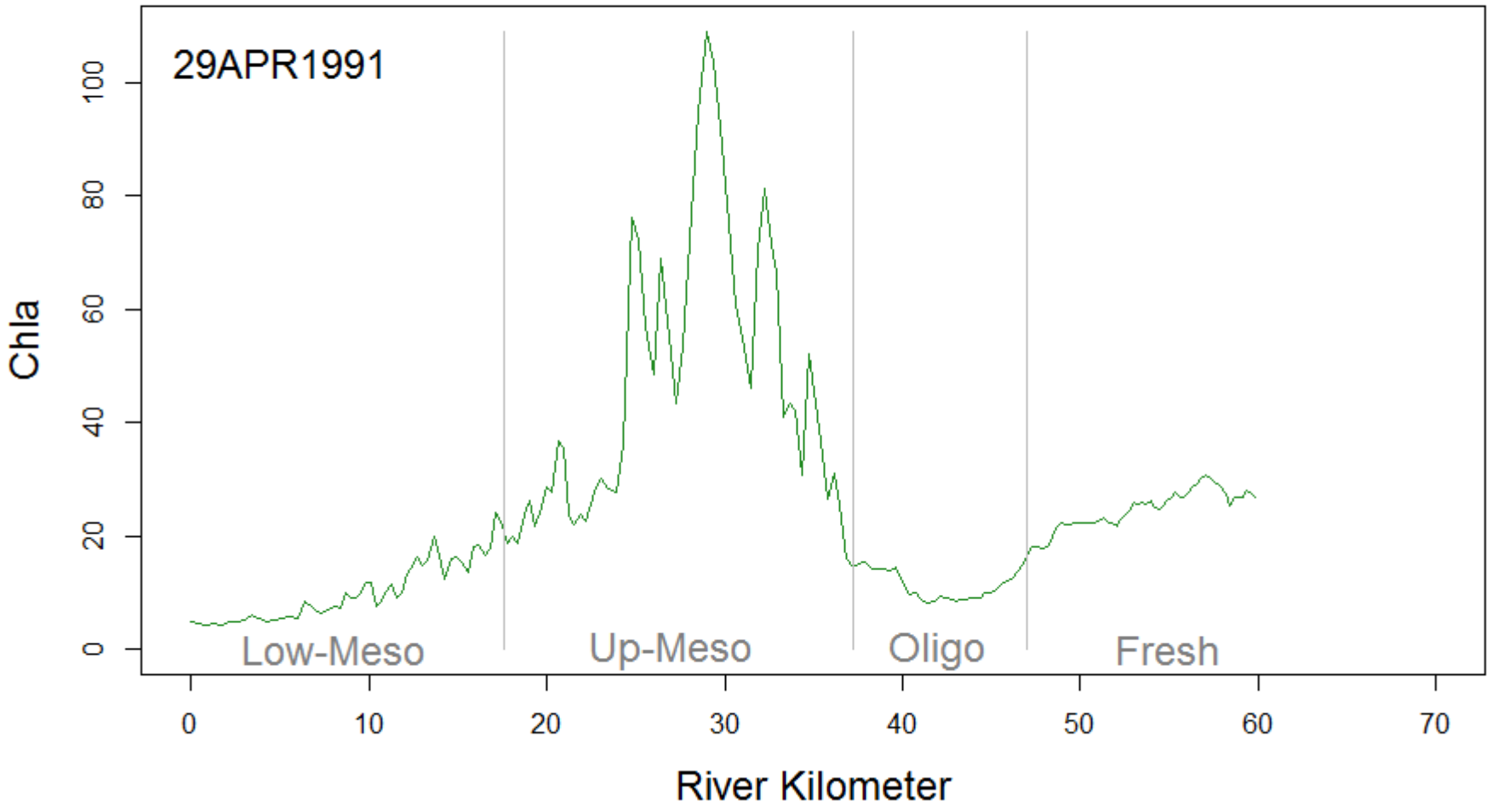
Late Summer


05SEP1989
Chla
0
20
40
60
Low-Meso
Up-Meso
Oligo
Fresh
0
10
20
30
40
50
60
70
River Kilometer

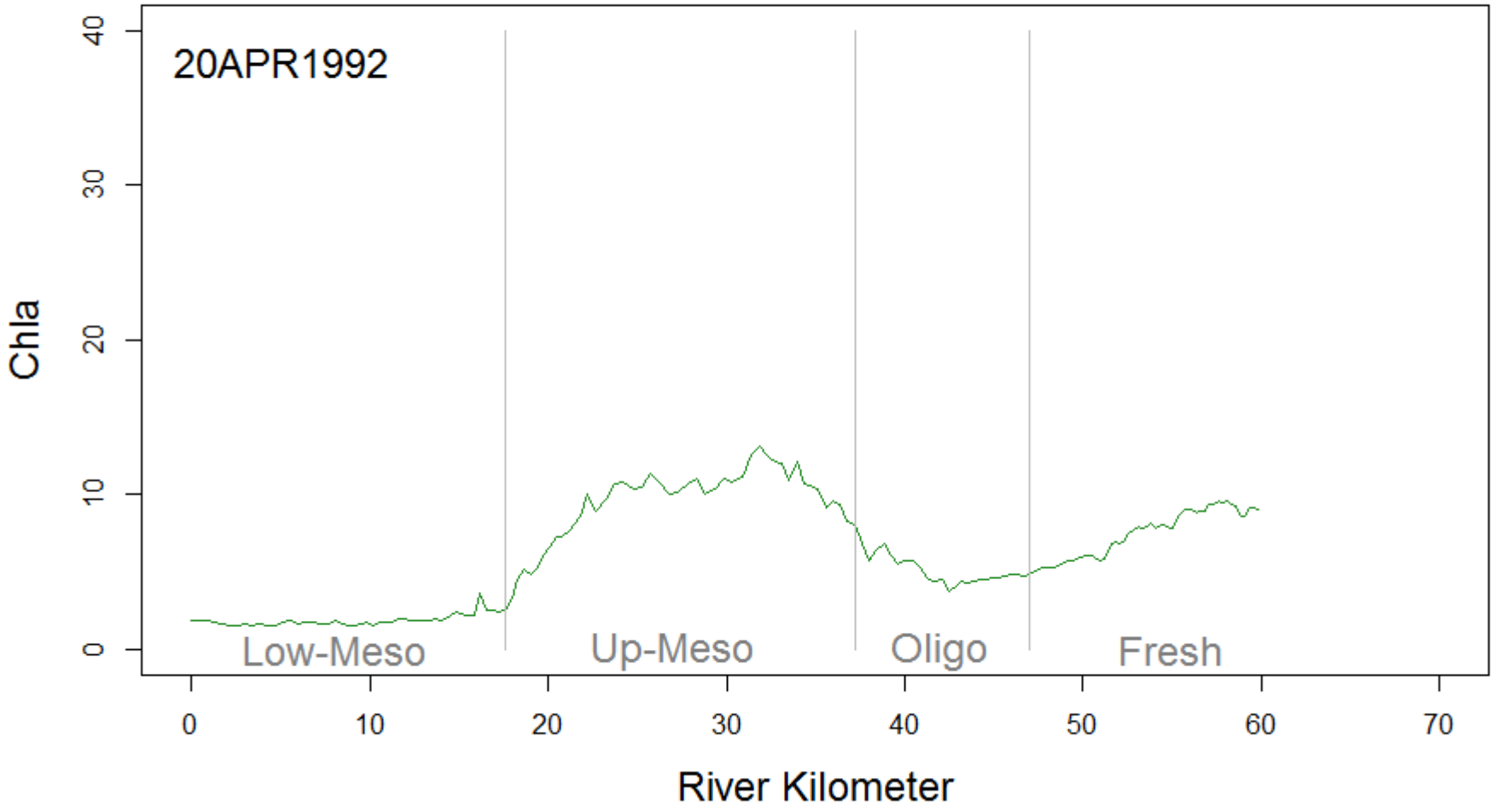
Spring Bloom



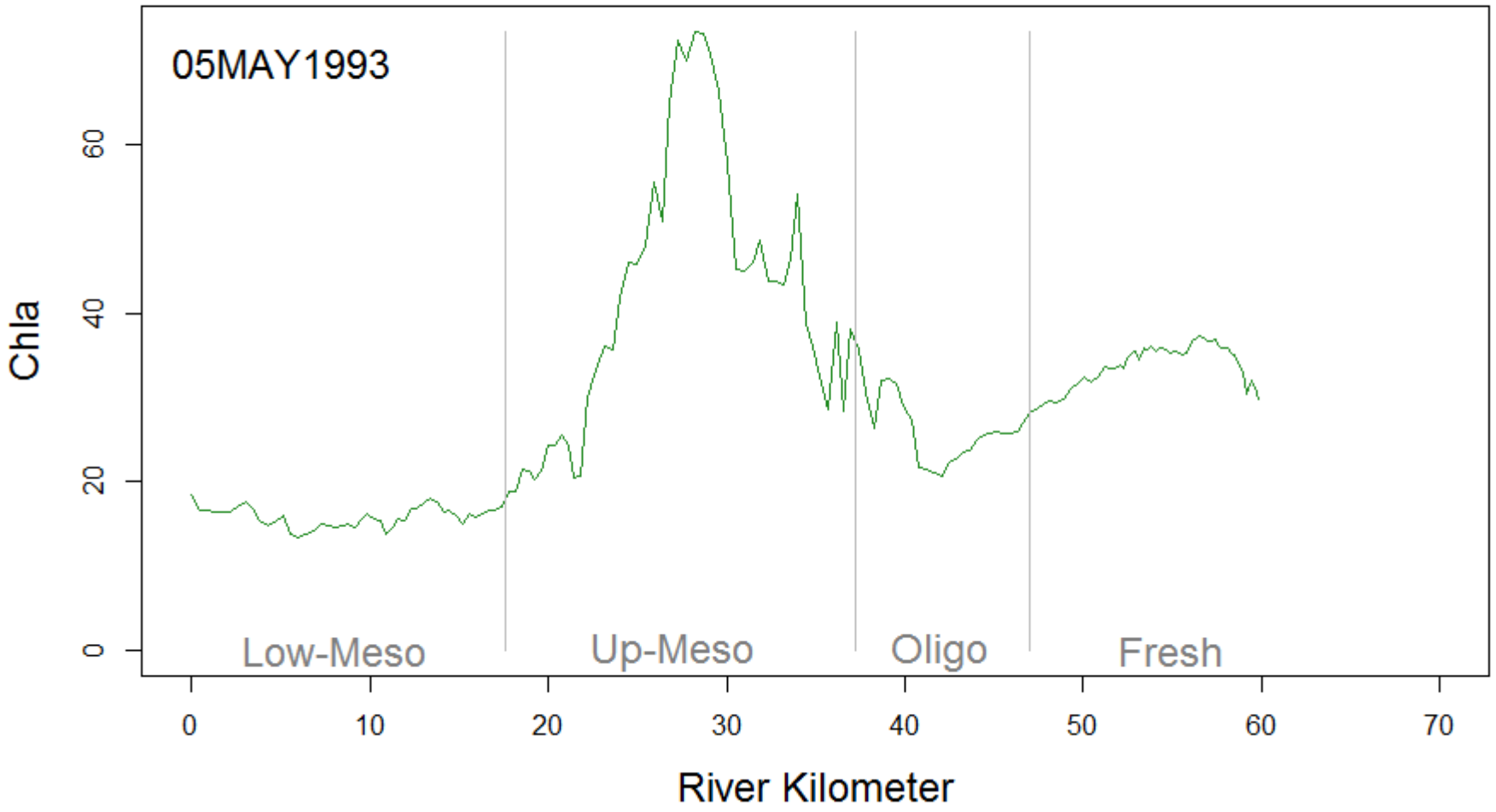
Spring Bloom



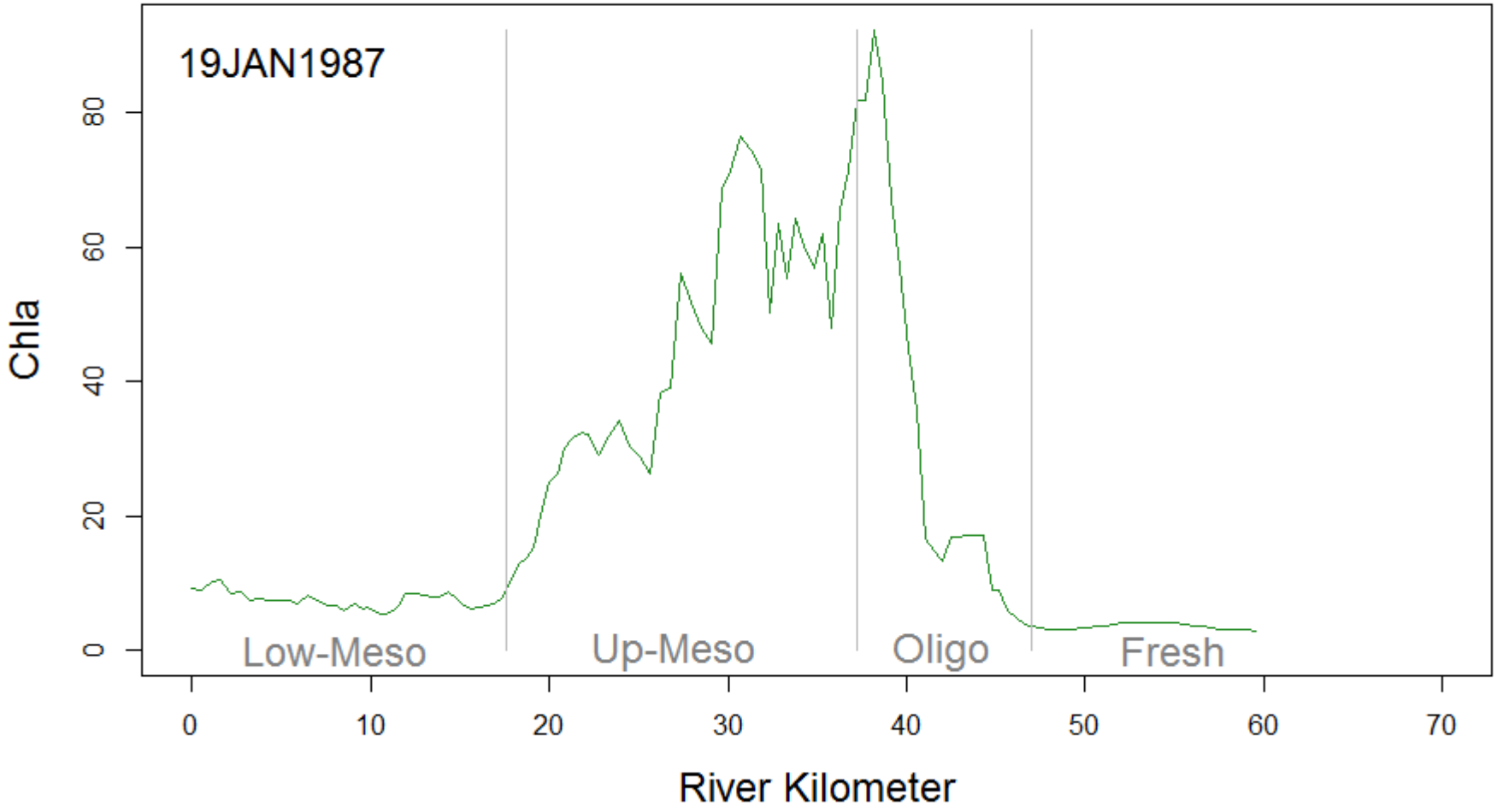
Spring Bloom



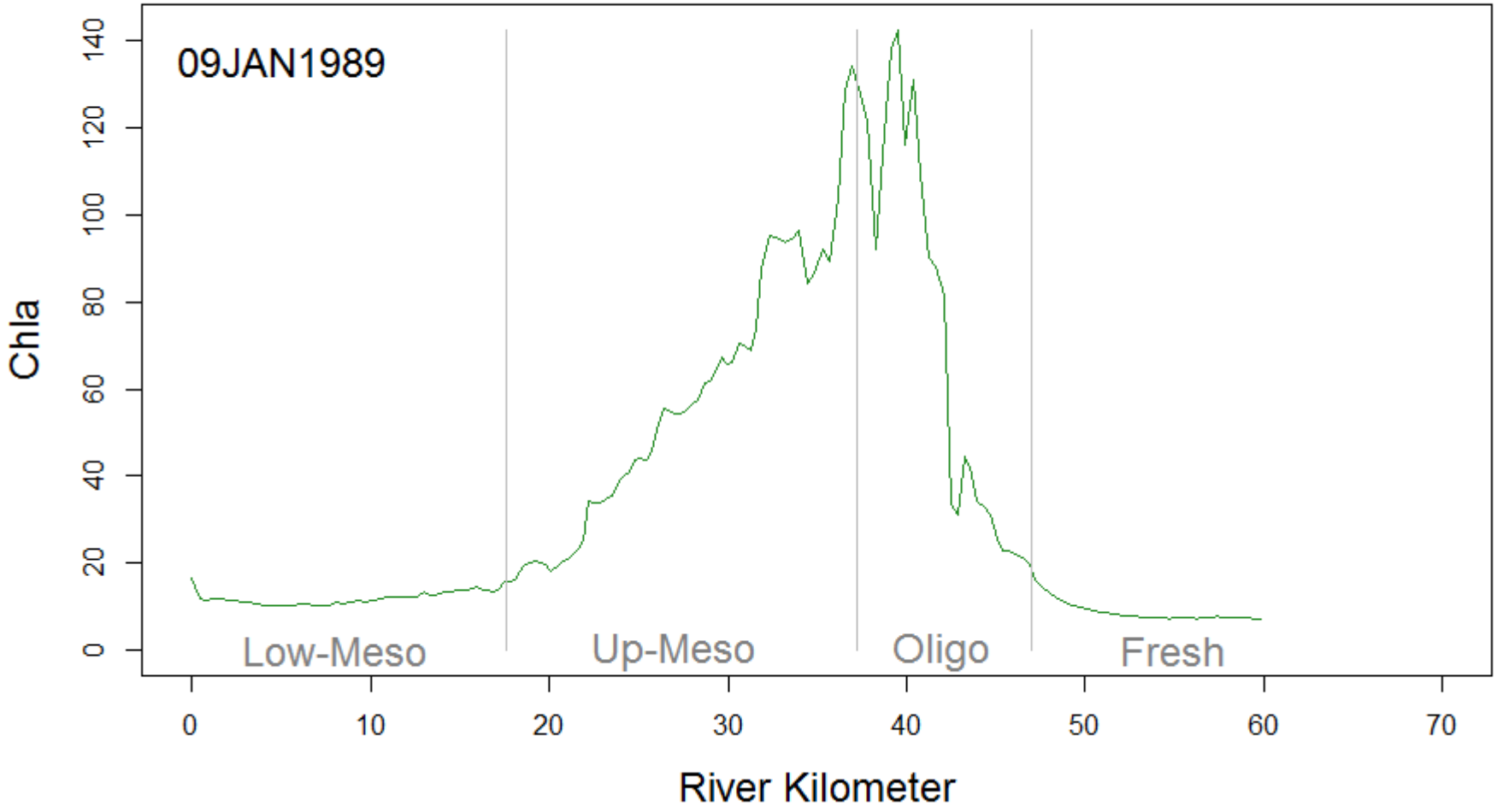
Spring Bloom

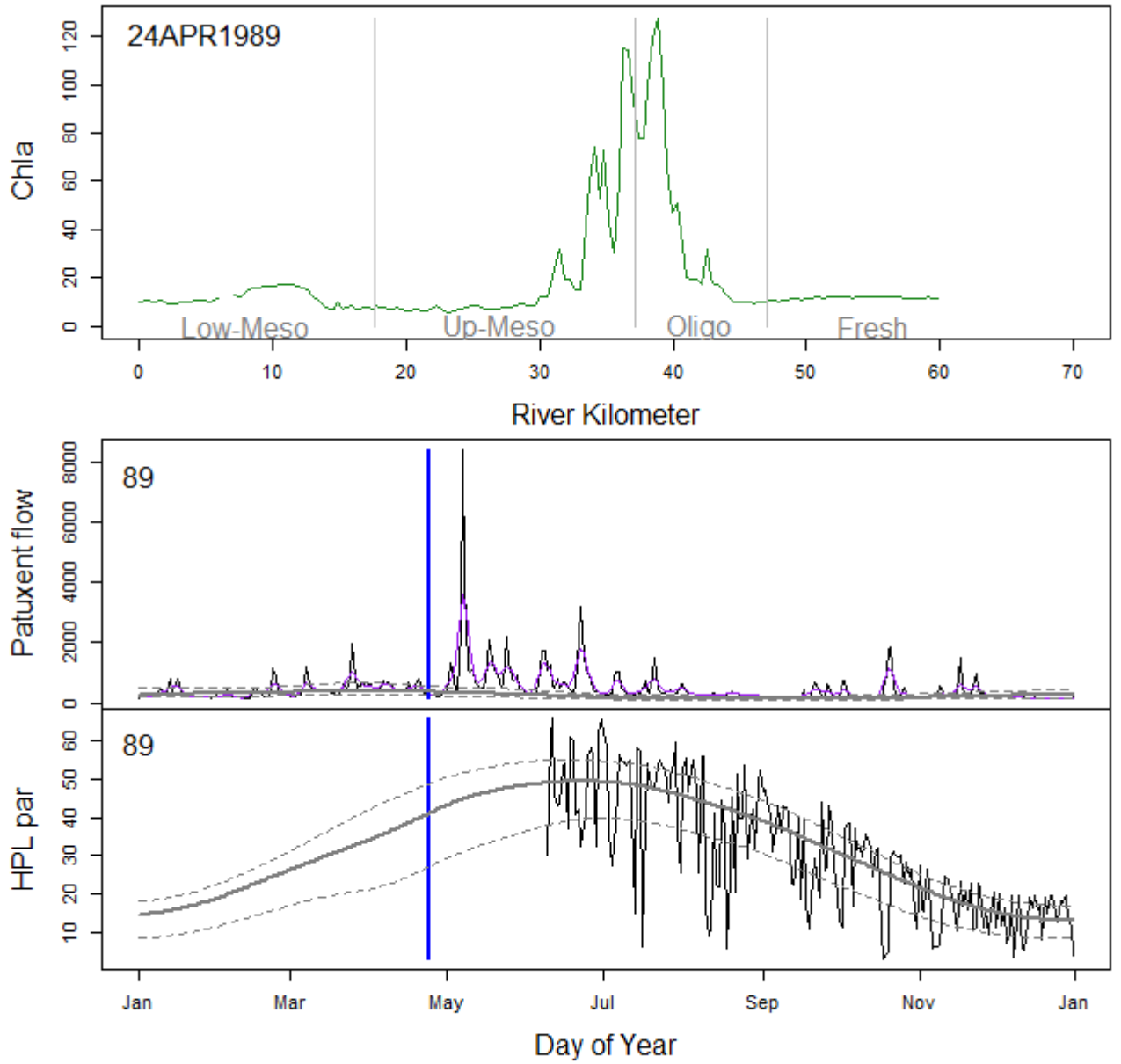


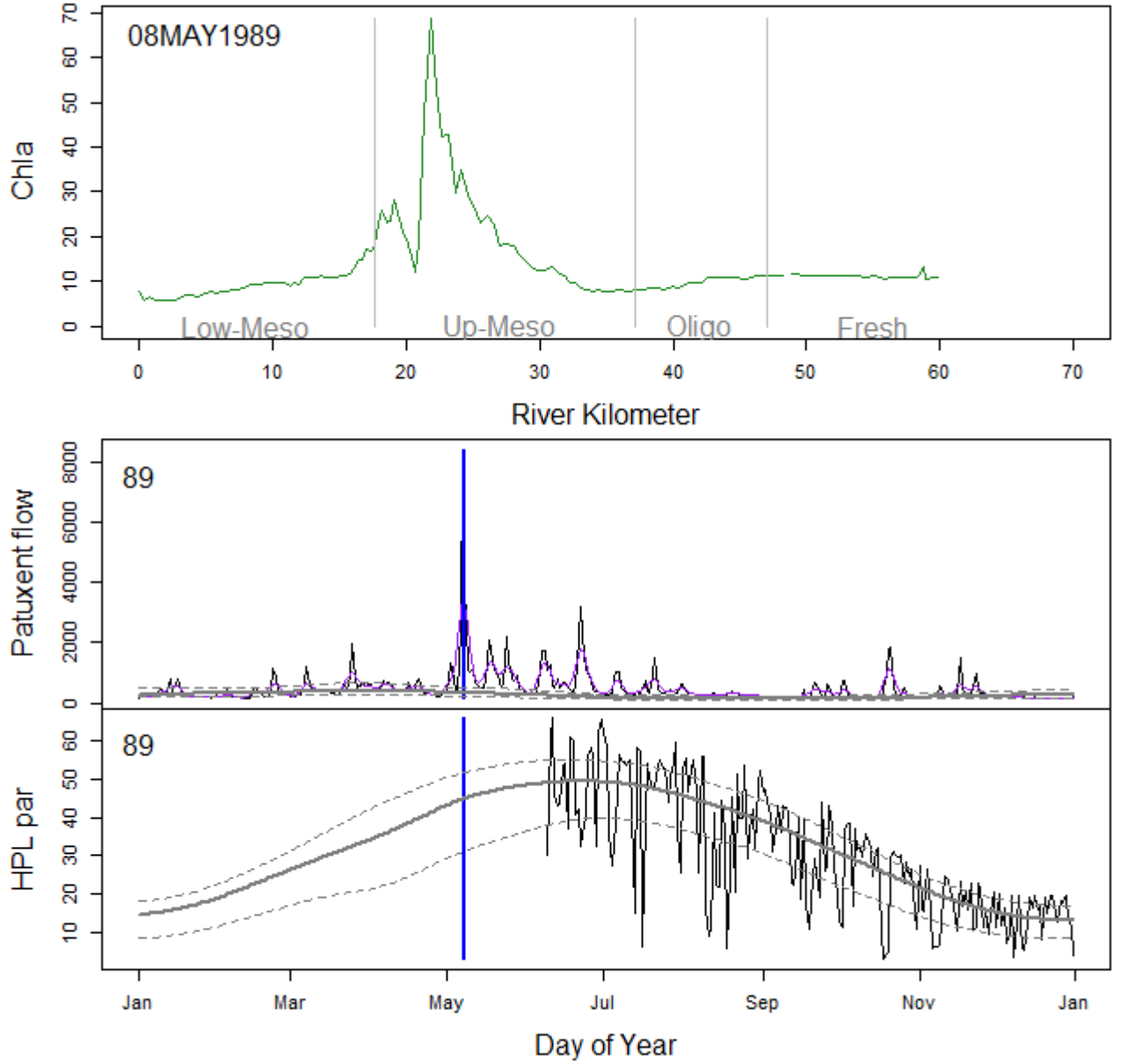
Mid-late Winter

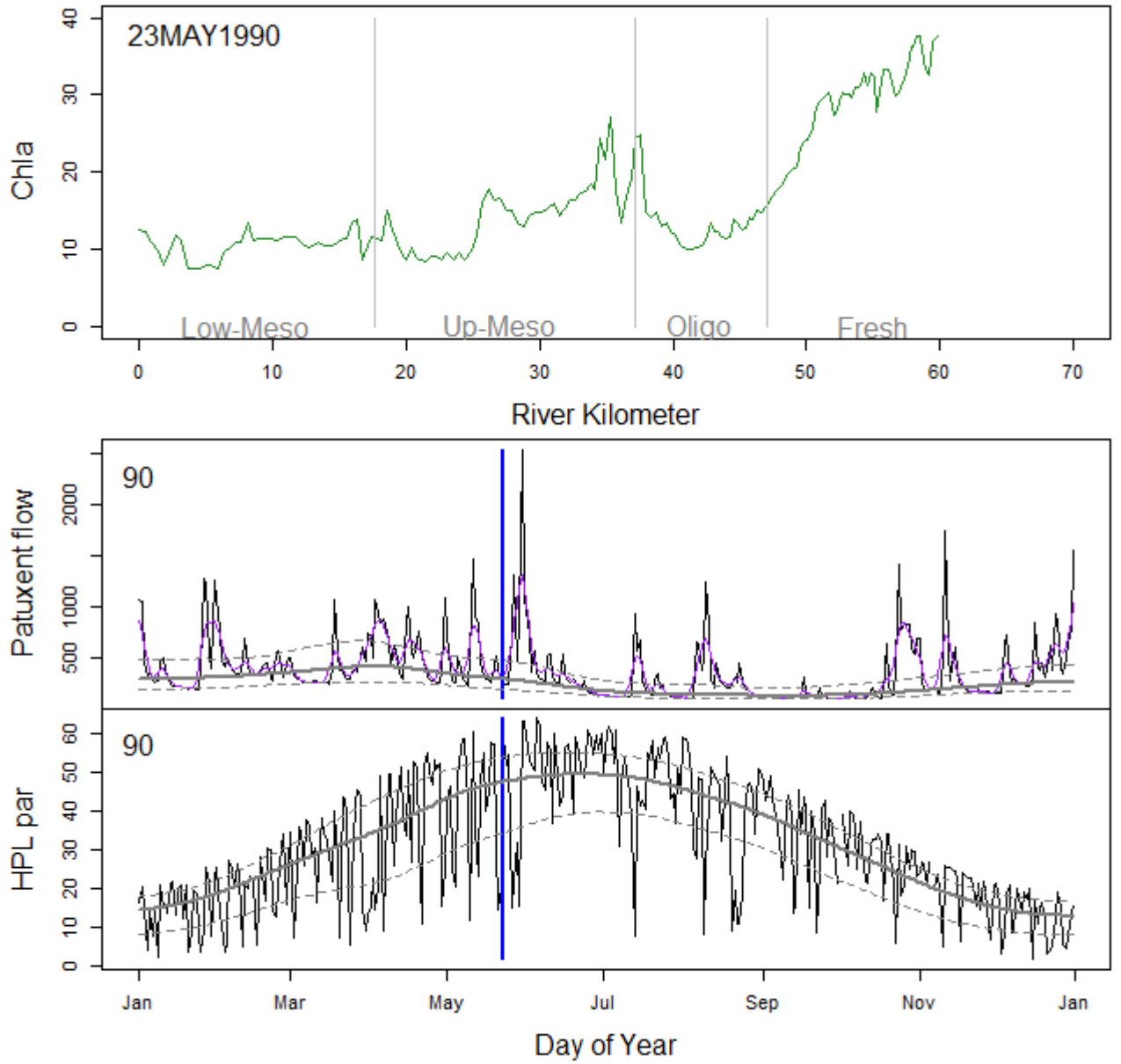


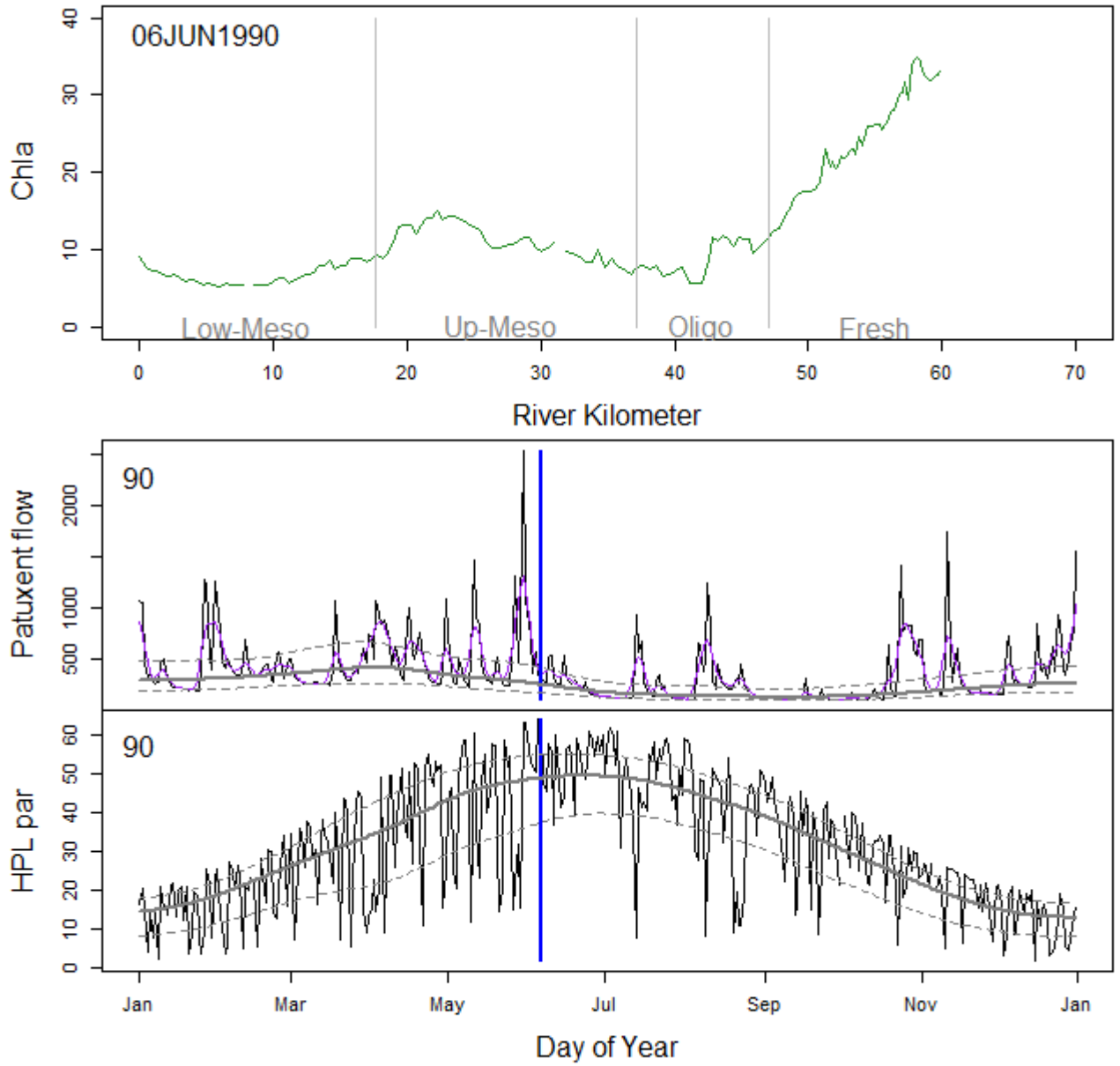
Mid-late Winter




24APR1989
Chla
Low-Meso
Up-Meso
Oligo
Fresh
River Kilometer
89
Patuxent flow
89
HPL par
Jan
Mar
May
Jul
Sep
Nov
Jan
Day of Year


08MAY1989
Chla
Low-Meso
Up-Meso
Oligo
Fresh
River Kilometer
89
Patuxent flow
89
HPL par
Jan
Mar
May
Jul
Sep
Nov
Jan
Day of Year


23MAY1990
Chla
Low-Meso
Up-Meso
Oligo
Fresh
River Kilometer
Patuxent flow
90
HPL par
90
Jan
Mar
May
Jul
Sep
Nov
Jan
Day of Year


06JUN1990
Chla
Low-Meso
Up-Meso
Oligo
Fresh
River Kilometer
Patuxent flow
90
HPL par
90
Jan
Mar
May
Jul
Sep
Nov
Jan
Day of Year

Patterns

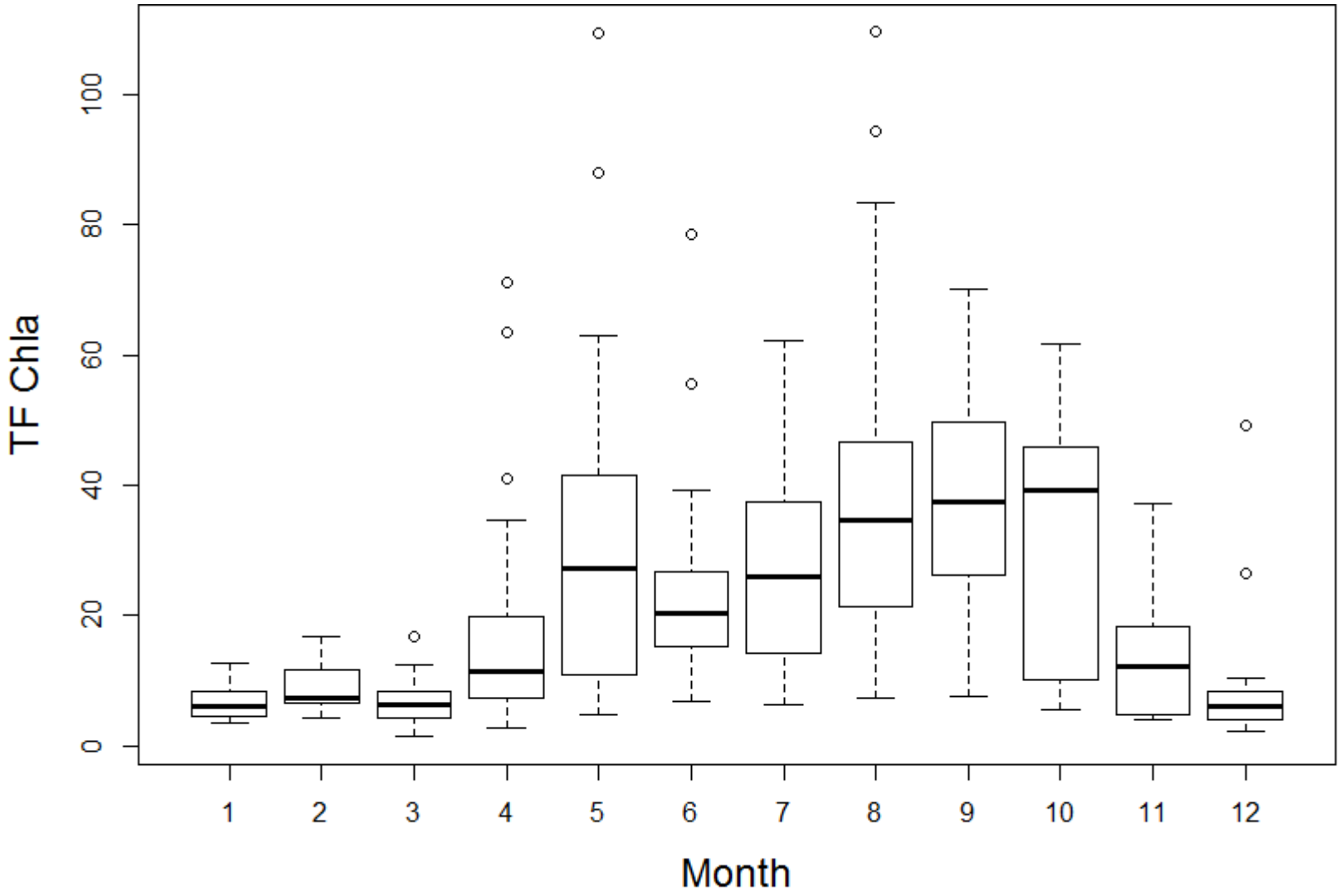
High flow reduces Chl in TF.
High Chl in TF sets up during low flow periods
Flow appears to have little effect downriver
Flow and light are connected
Maybe cloudy days lead to lower Chl.

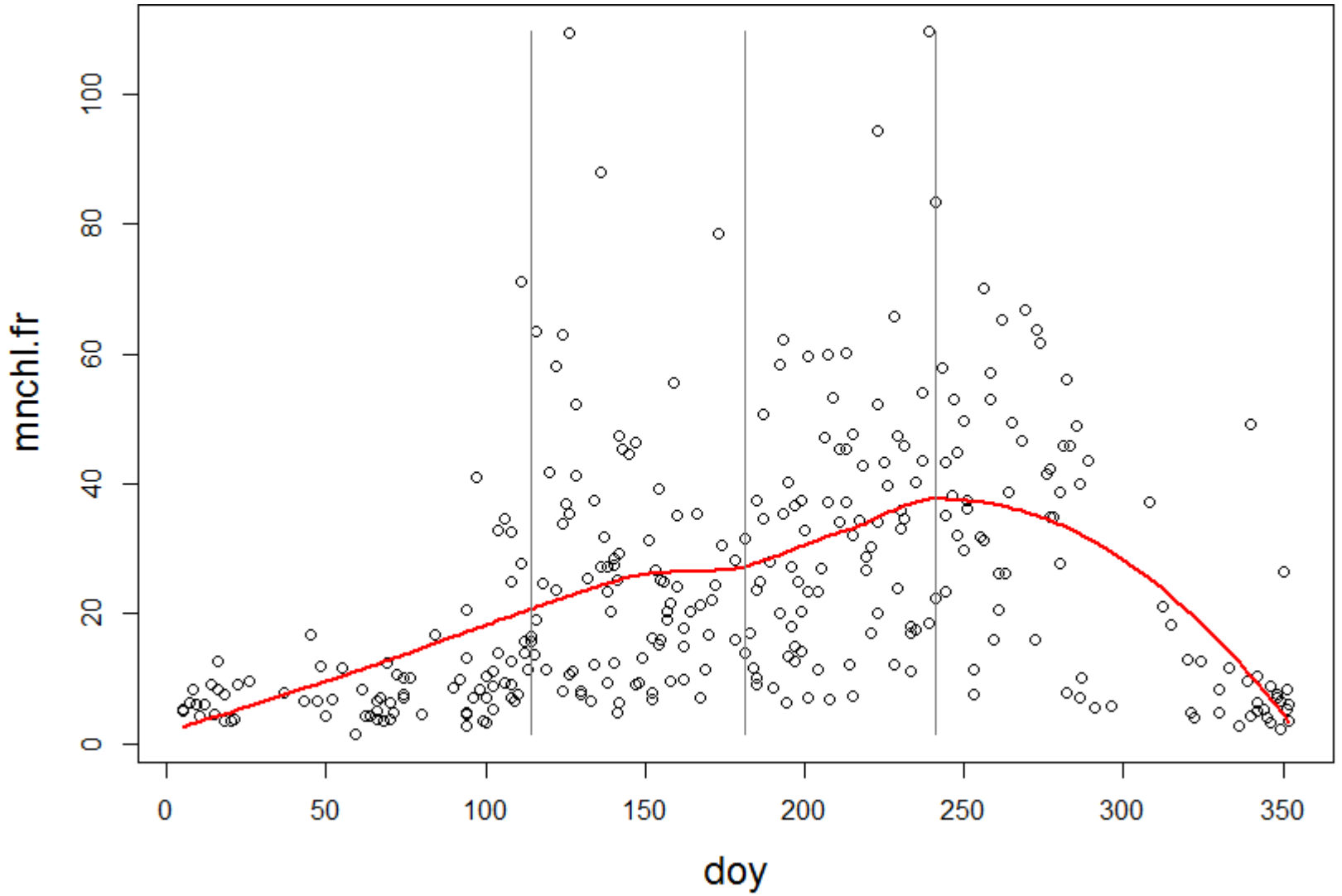
Quantitative Measures

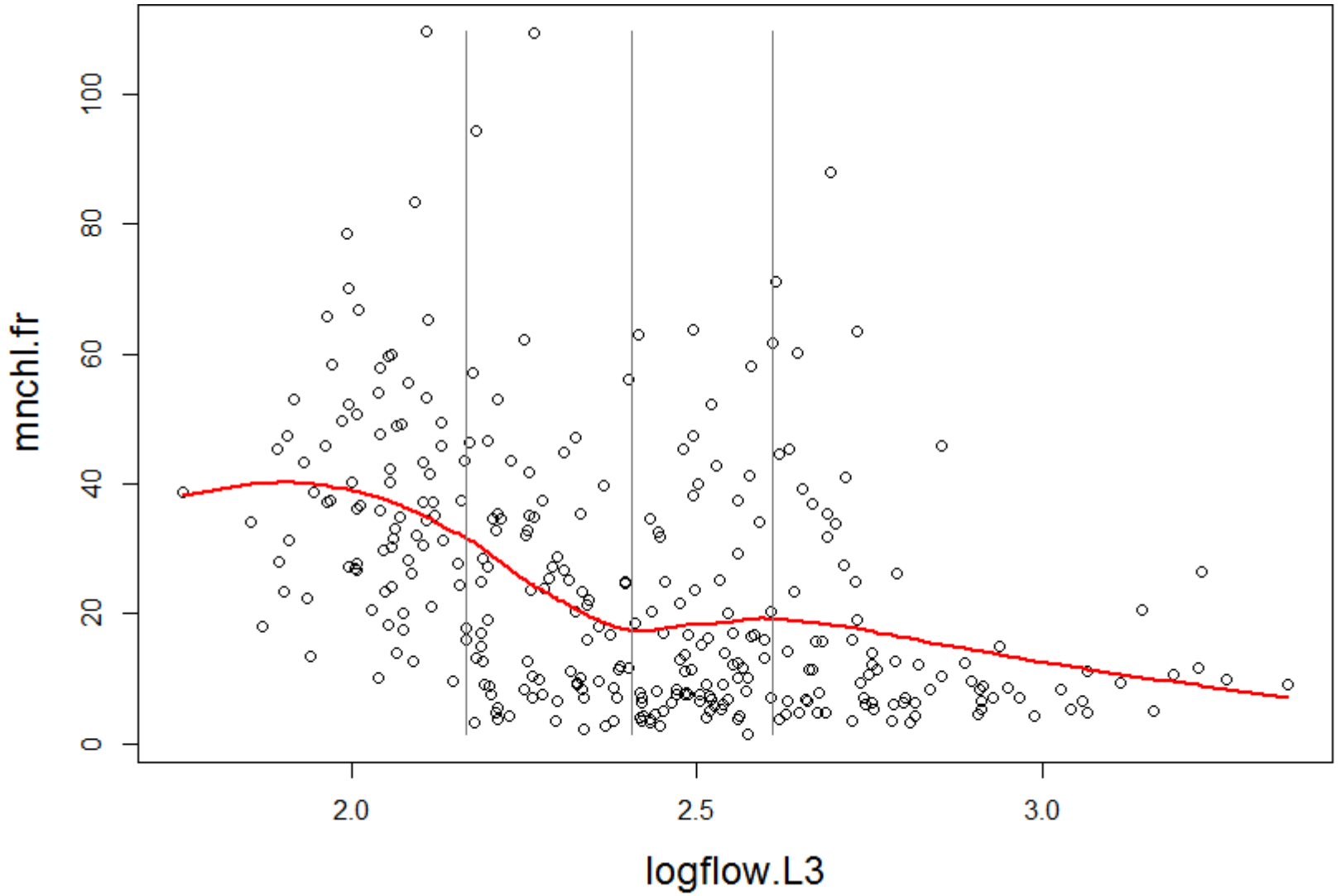
- Dependent – Integral of chl by salinity zone
- Season
  - Monthly means
  - Cyclical day of year (doy)
- Flow
  - Moving average of flow with lags (up to 7 weeks)
  - Days since flow event
  - Size of flow event

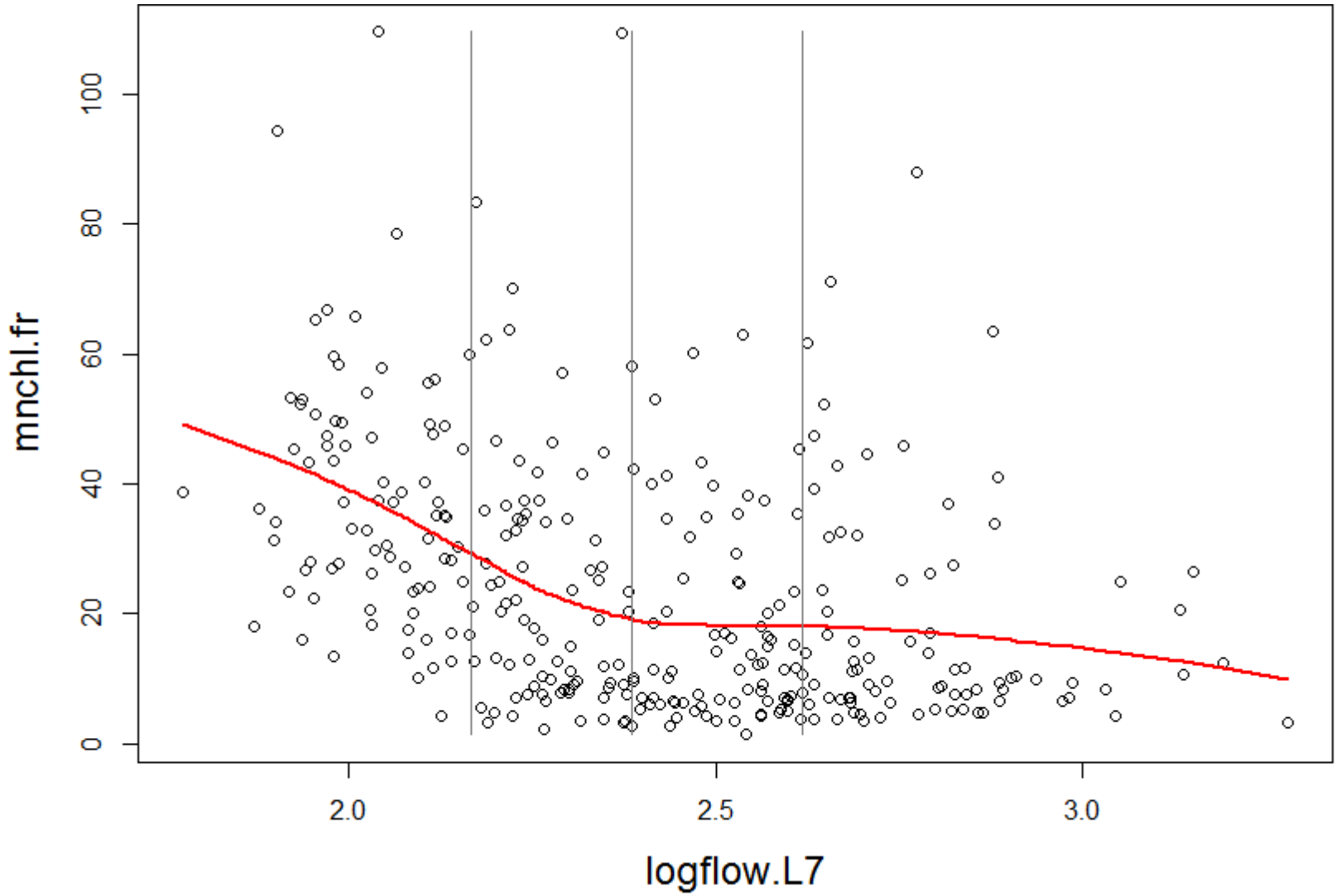
Quantitative Measures

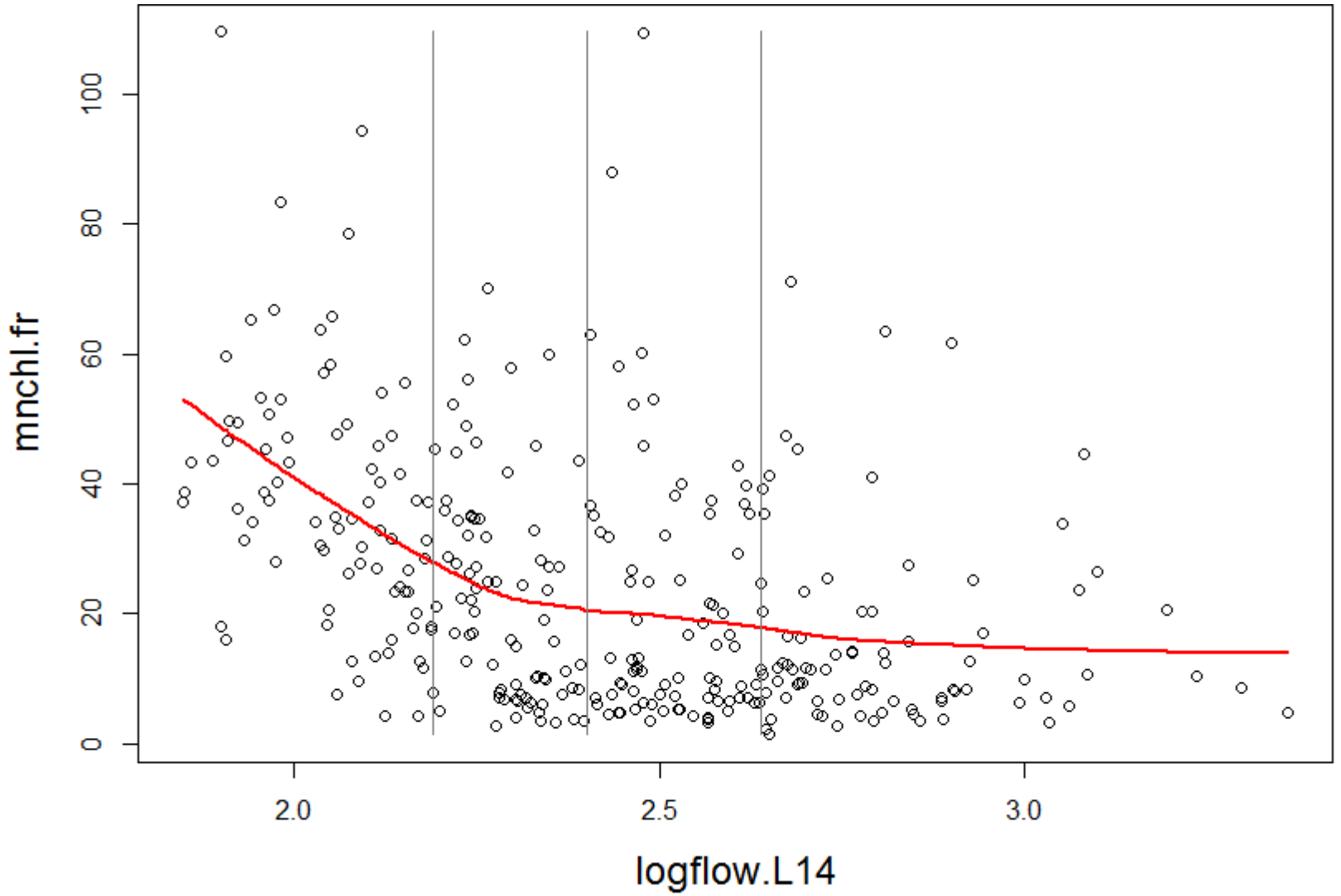
- Light
  - Series of averages (1 day, 2 day, . . . 7 day)
  - Sunny days (par > 0.75 quantile)
  - Cloudy days (par < 0.25 quantile)
  - Sunny days-cloudy days

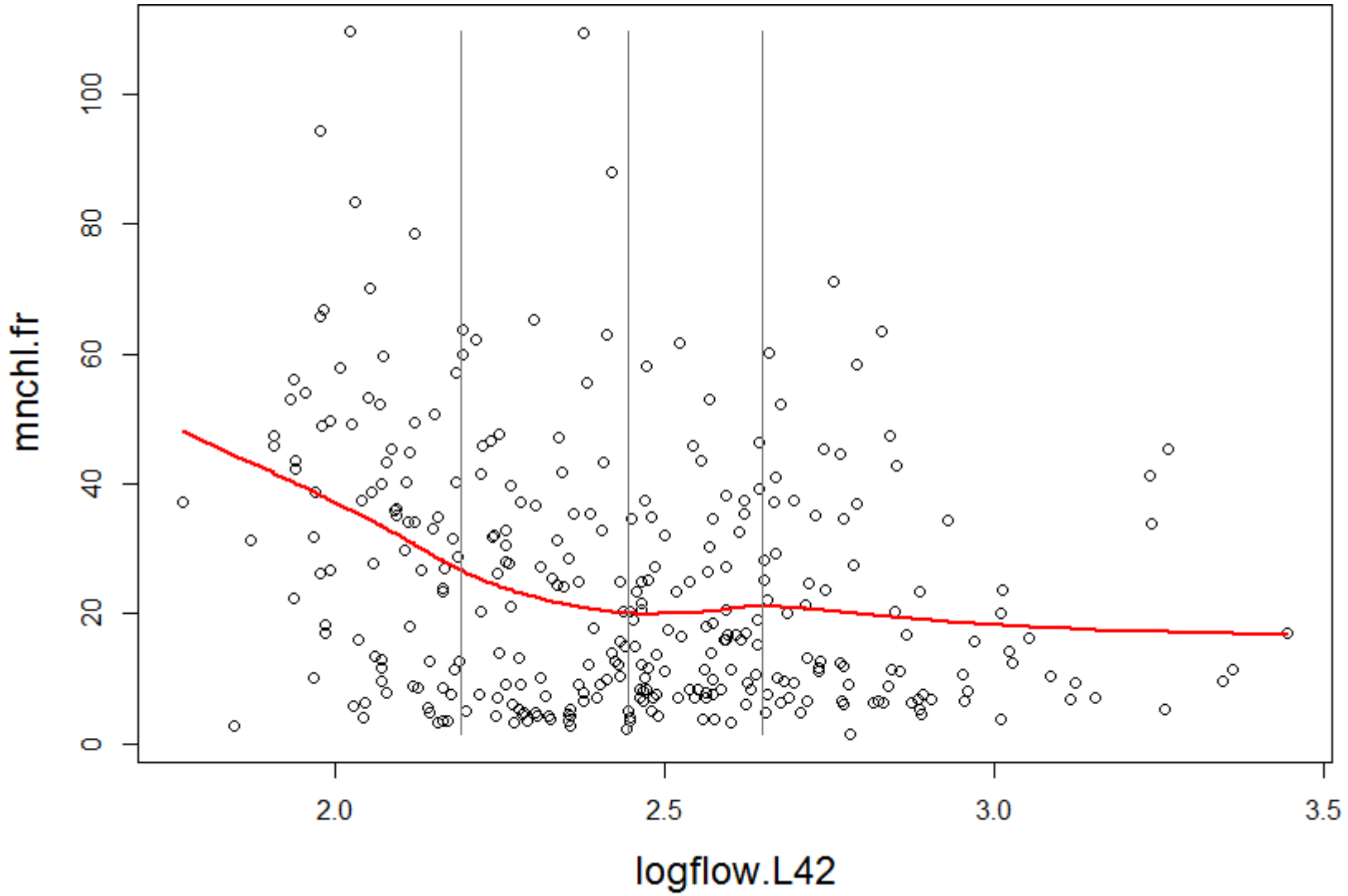

TF Chla
0
20
40
60
80
100
1
2
3
4
5
6
7
8
9
10
11
12
Month

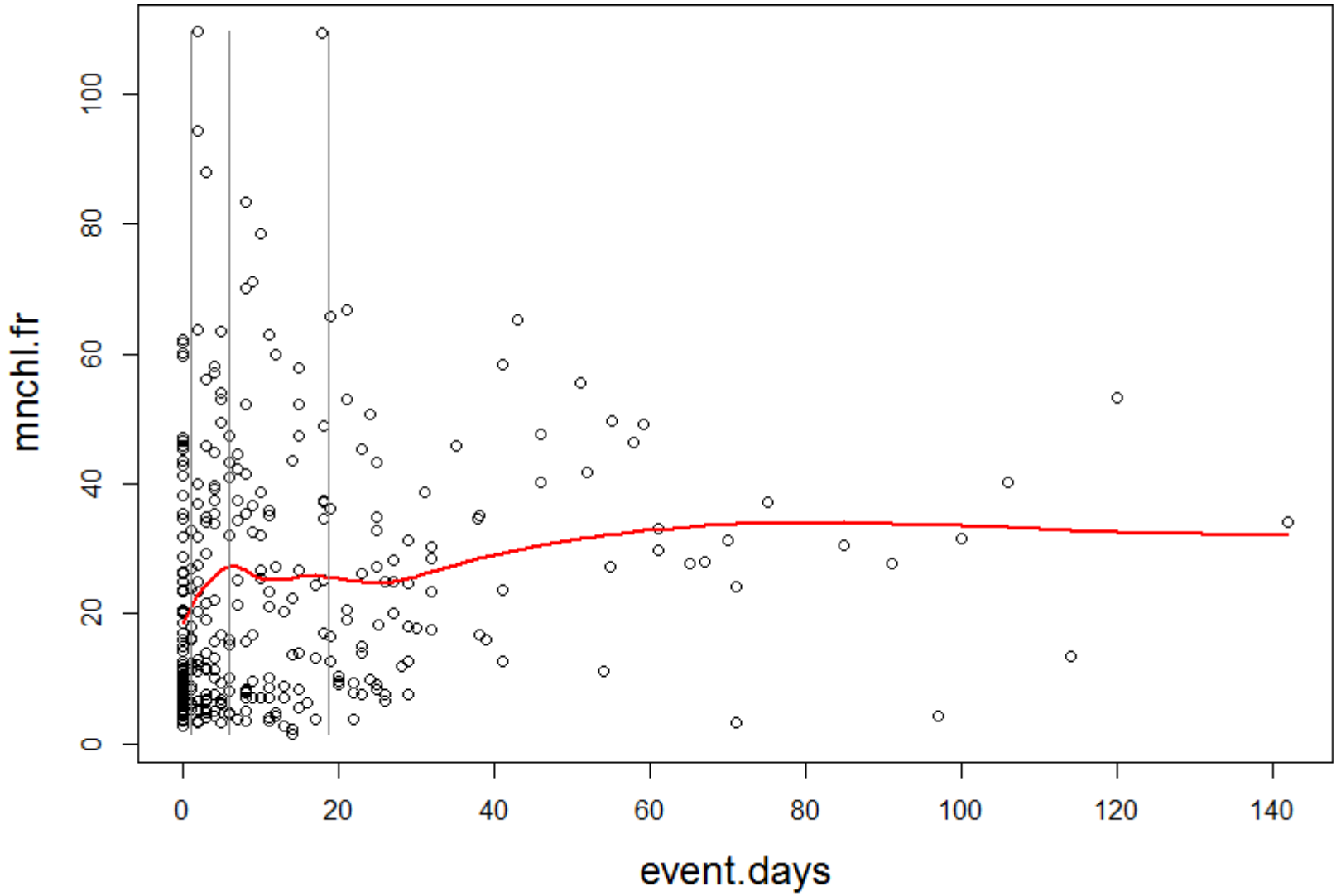

mnchl.fr
0
20
40
60
80
100
0
50
100
150
200
250
300
350
doy

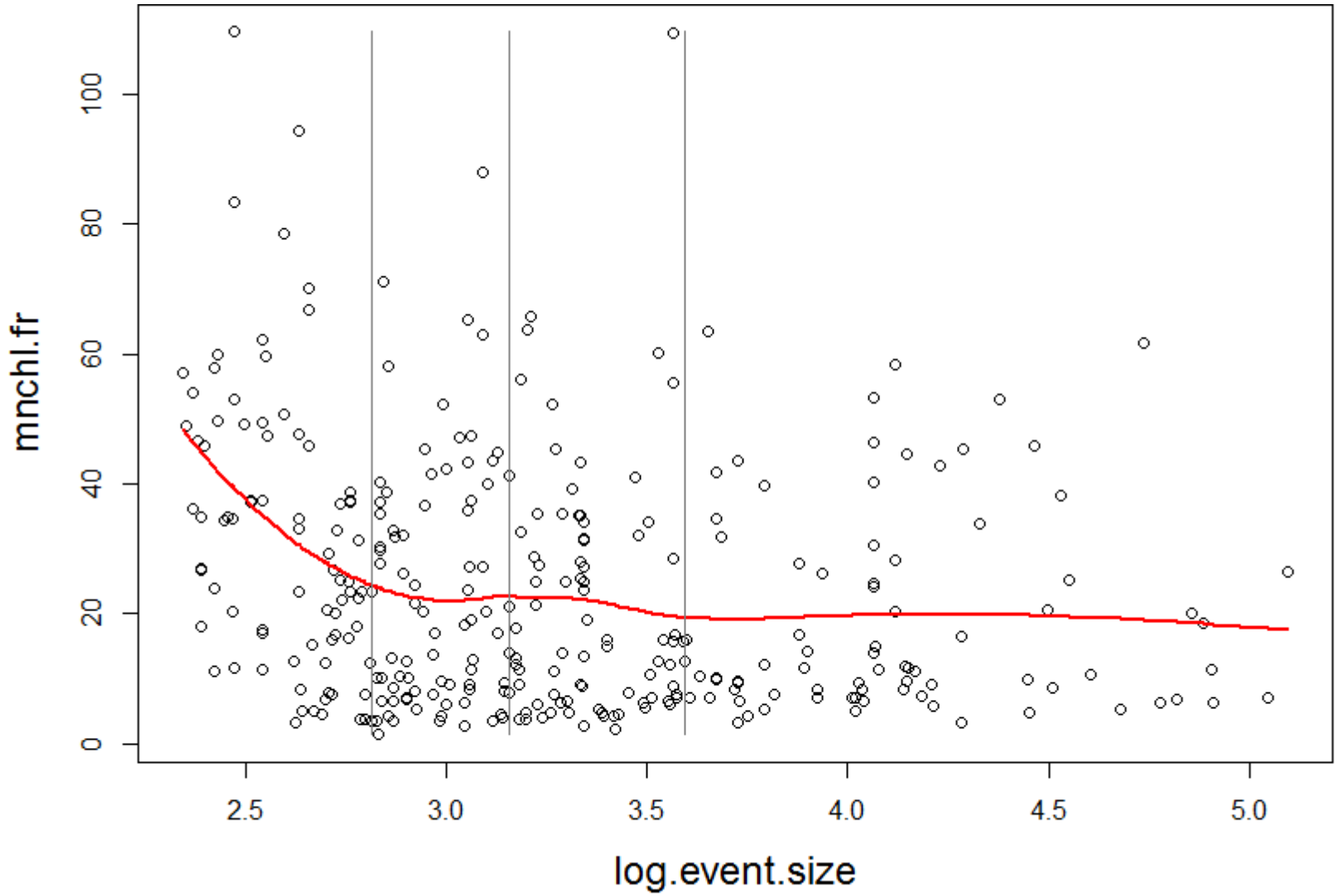

mnchl.fr
0
20
40
60
80
100
2.0
2.5
3.0
logflow.L3

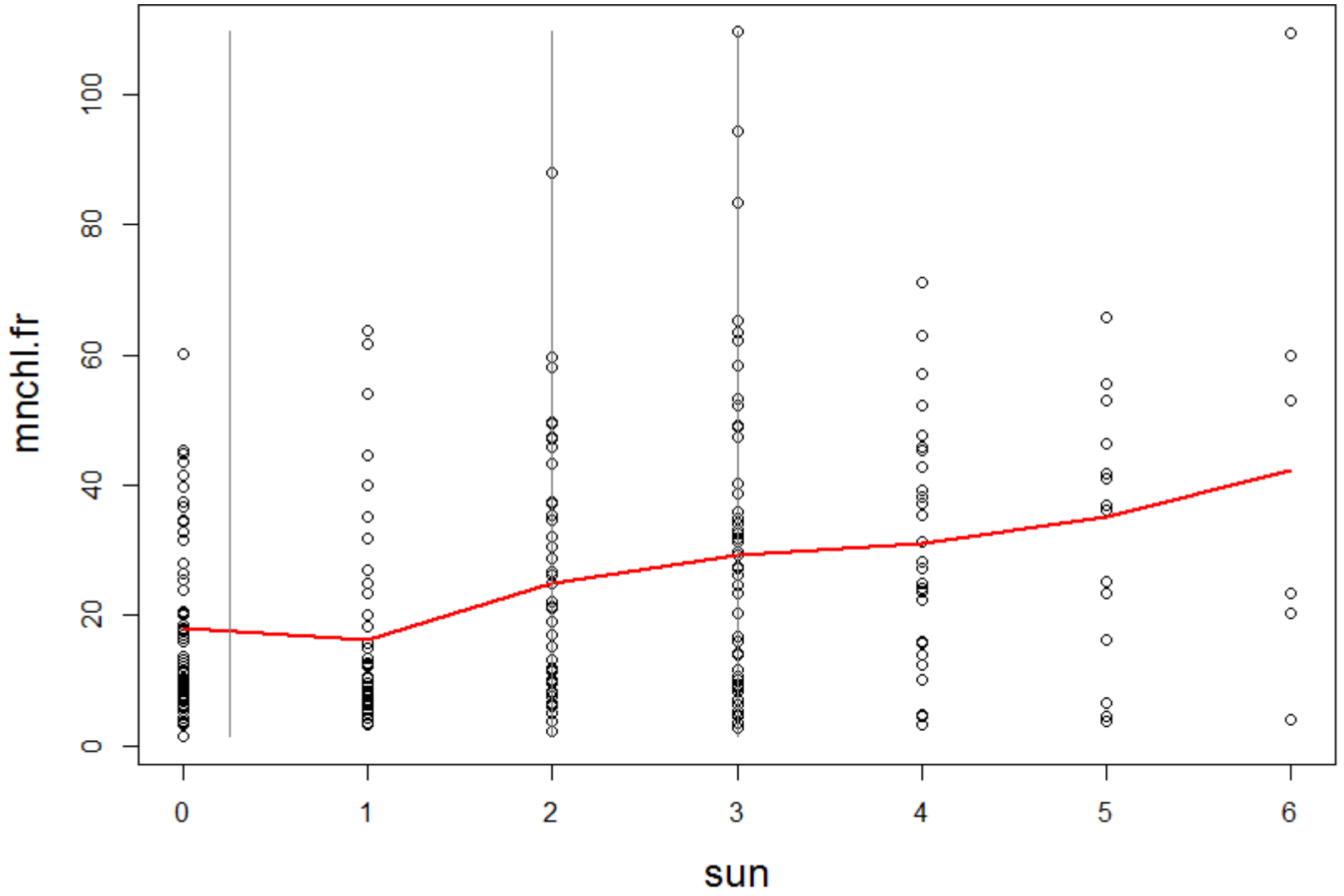

mnchl.fr
0
20
40
60
80
100
2.0
2.5
3.0
logflow.L7

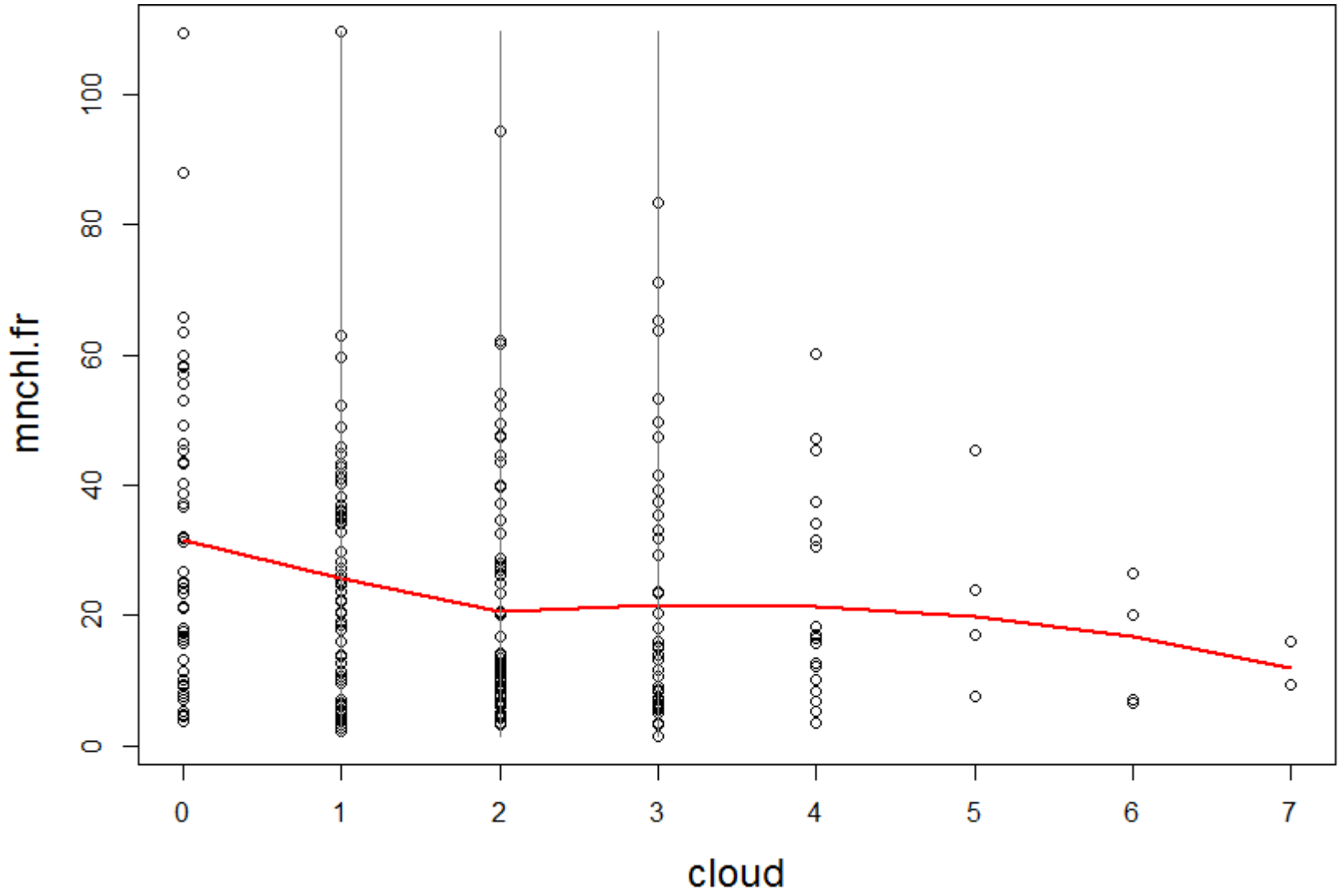

mnchl.fr
0
20
40
60
80
100
2.0
2.5
3.0
logflow.L14

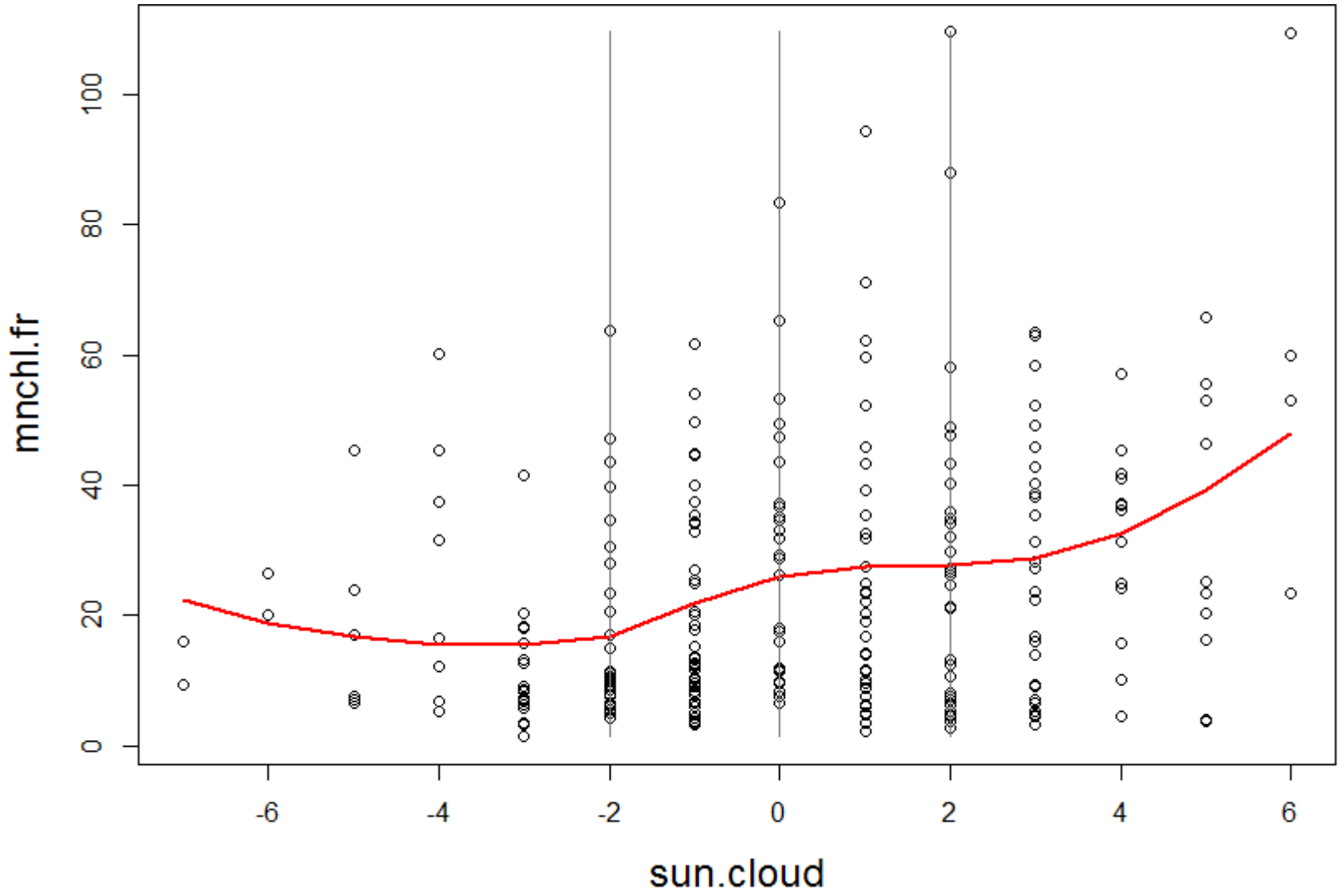

mnchl.fr
0
20
40
60
80
100
2.0
2.5
3.0
3.5
logflow.L42

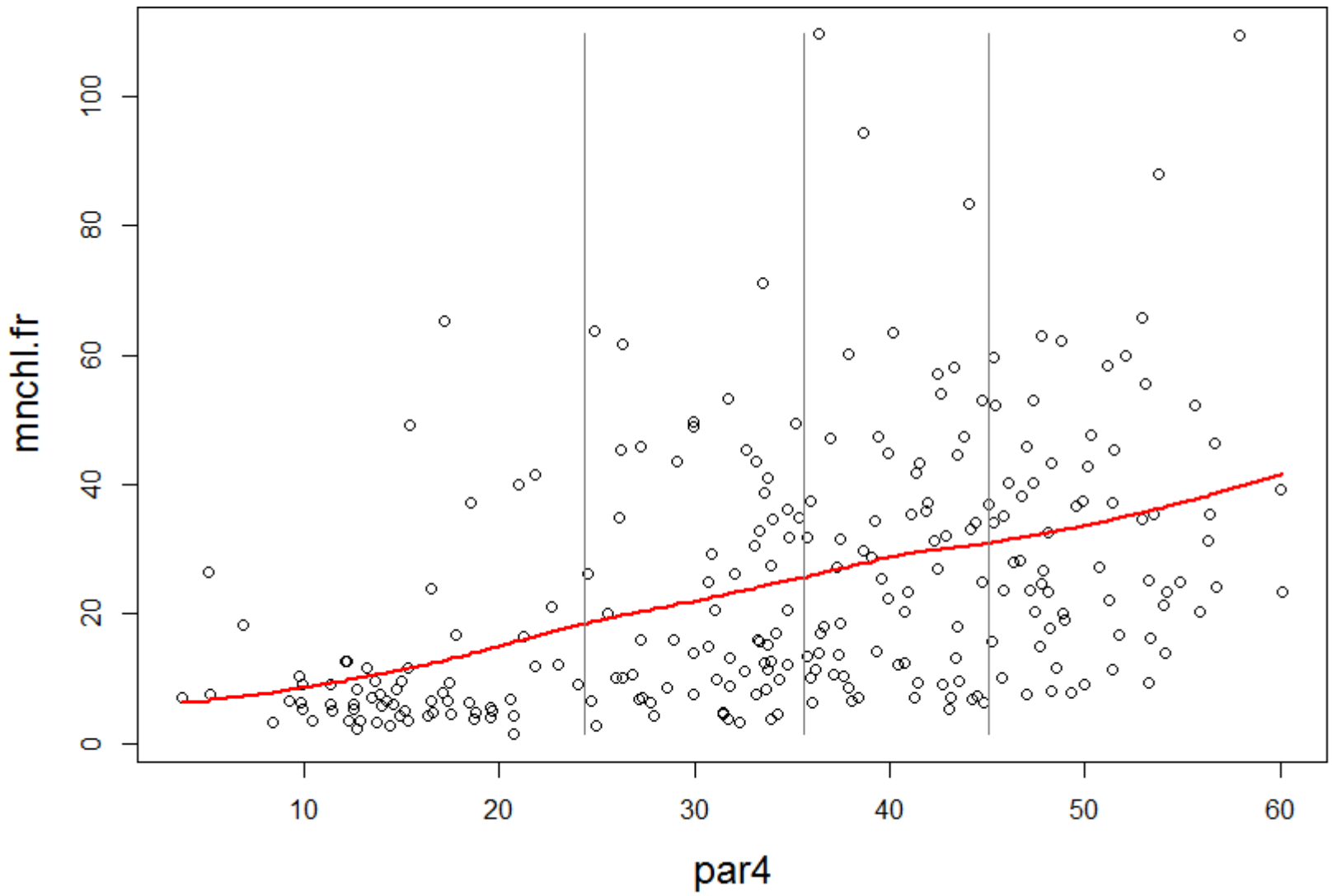

mnchl.fr
0
20
40
60
80
100
0
20
40
60
80
100
120
140
event.days


mnchl.fr
0
20
40
60
80
100
2.5
3.0
3.5
4.0
4.5
5.0
log.event.size

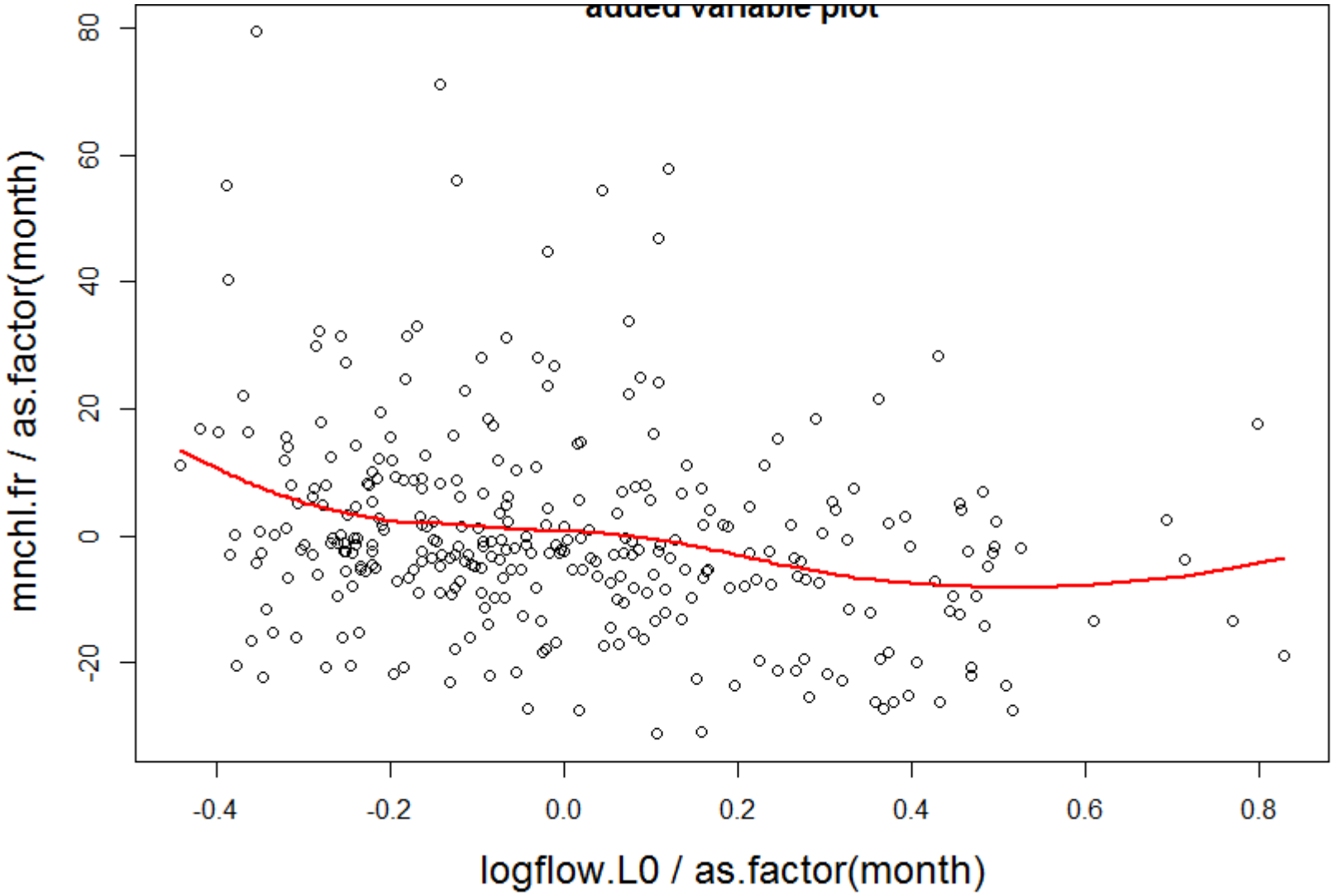

mnchl.fr
0
20
40
60
80
100
0
1
2
3
4
5
6
sun

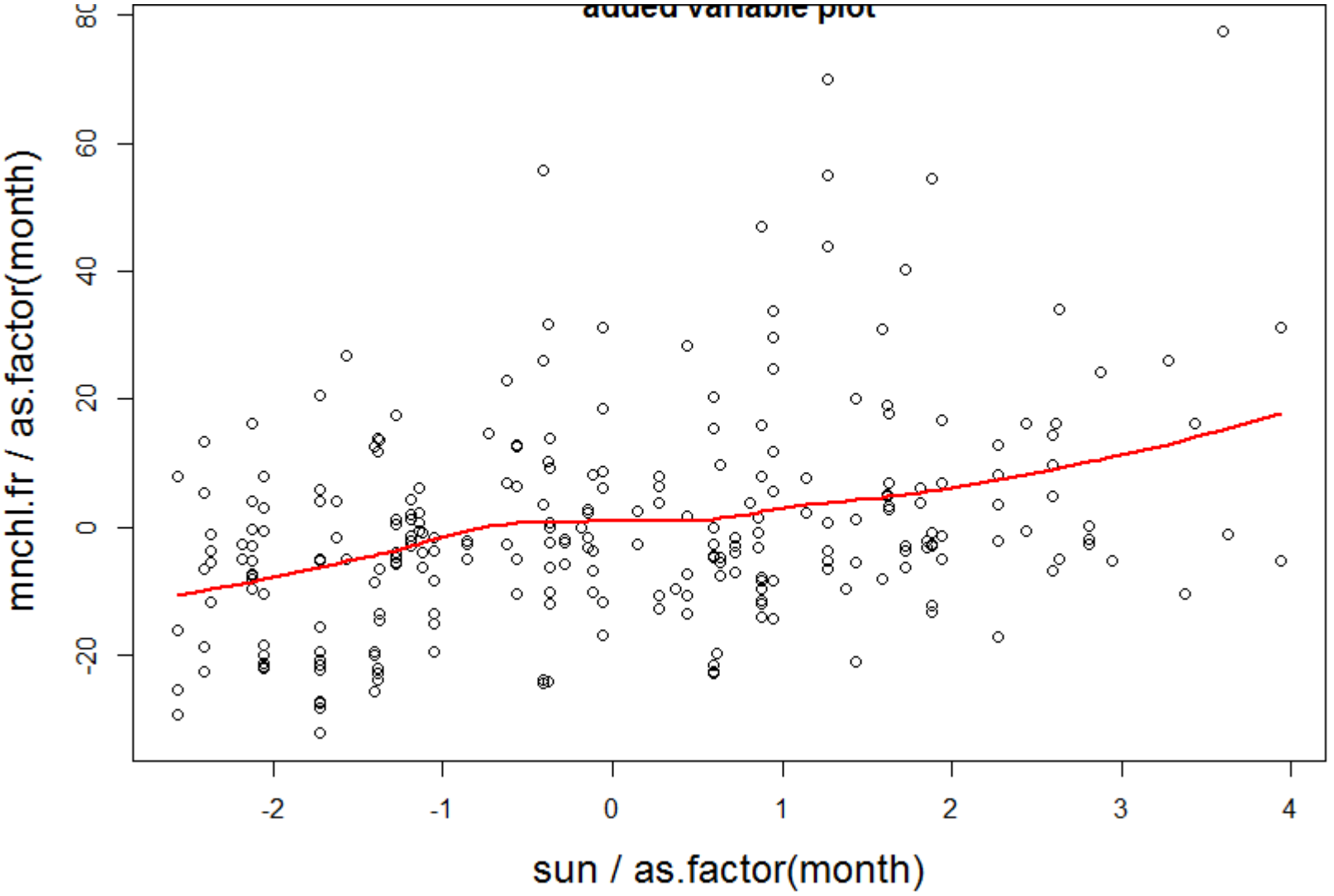

mnchl.fr
0
20
40
60
80
100
0
1
2
3
4
5
6
7
cloud

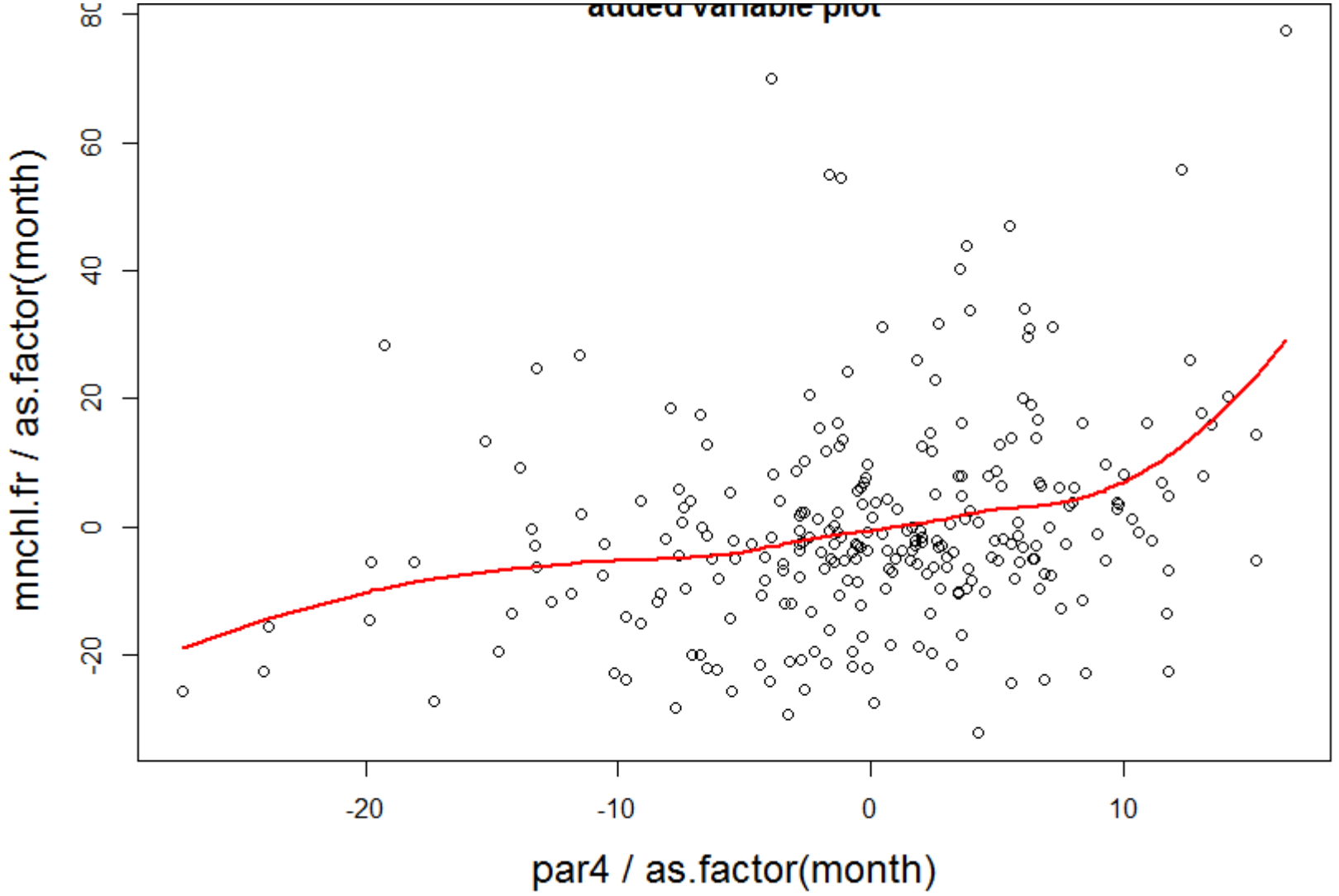

mnchl.fr
0
20
40
60
80
100
-6
-4
-2
0
2
4
6
sun.cloud


mnchl.fr
0
20
40
60
80
100
10
20
30
40
50
60
par4

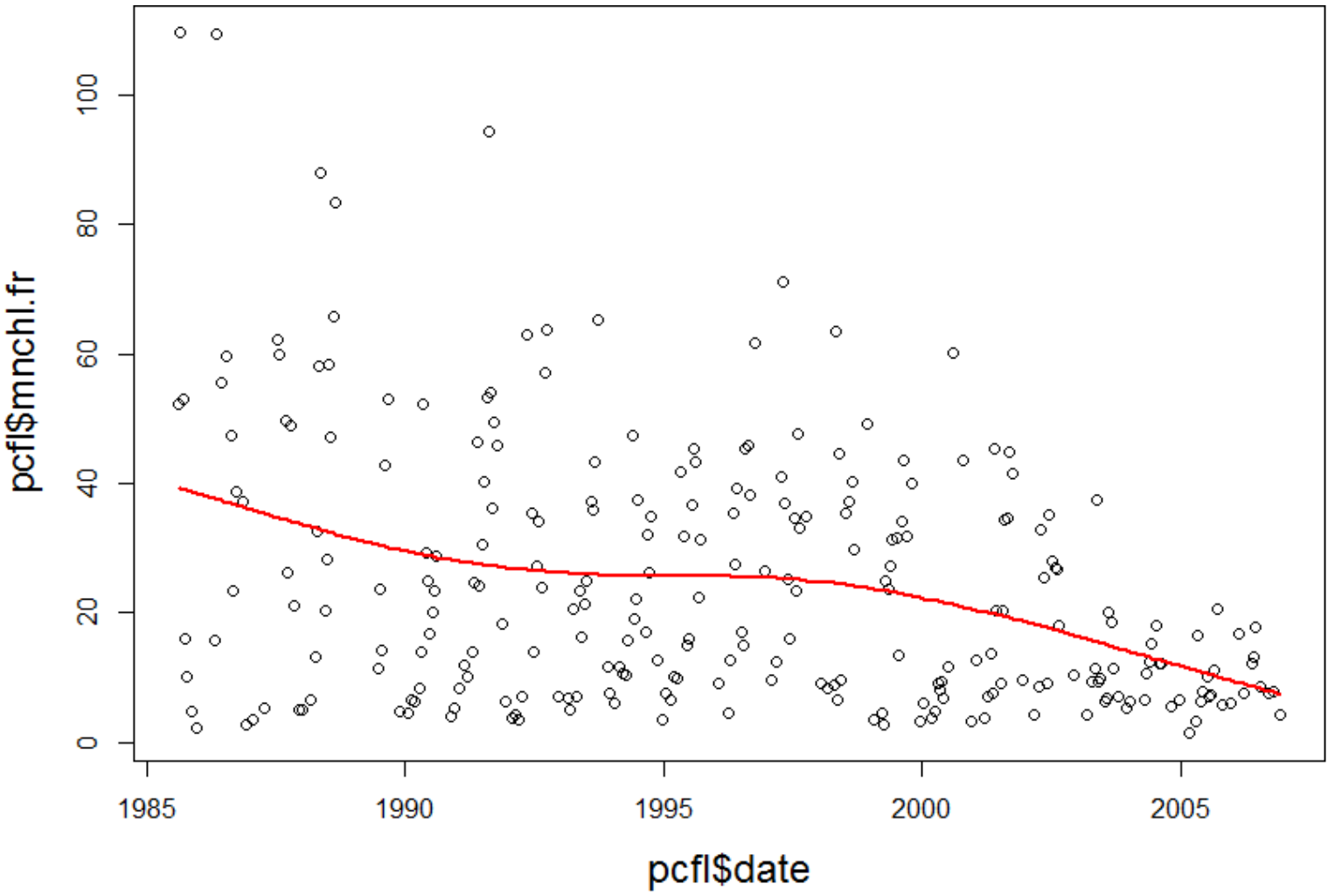
Variable Added Plots



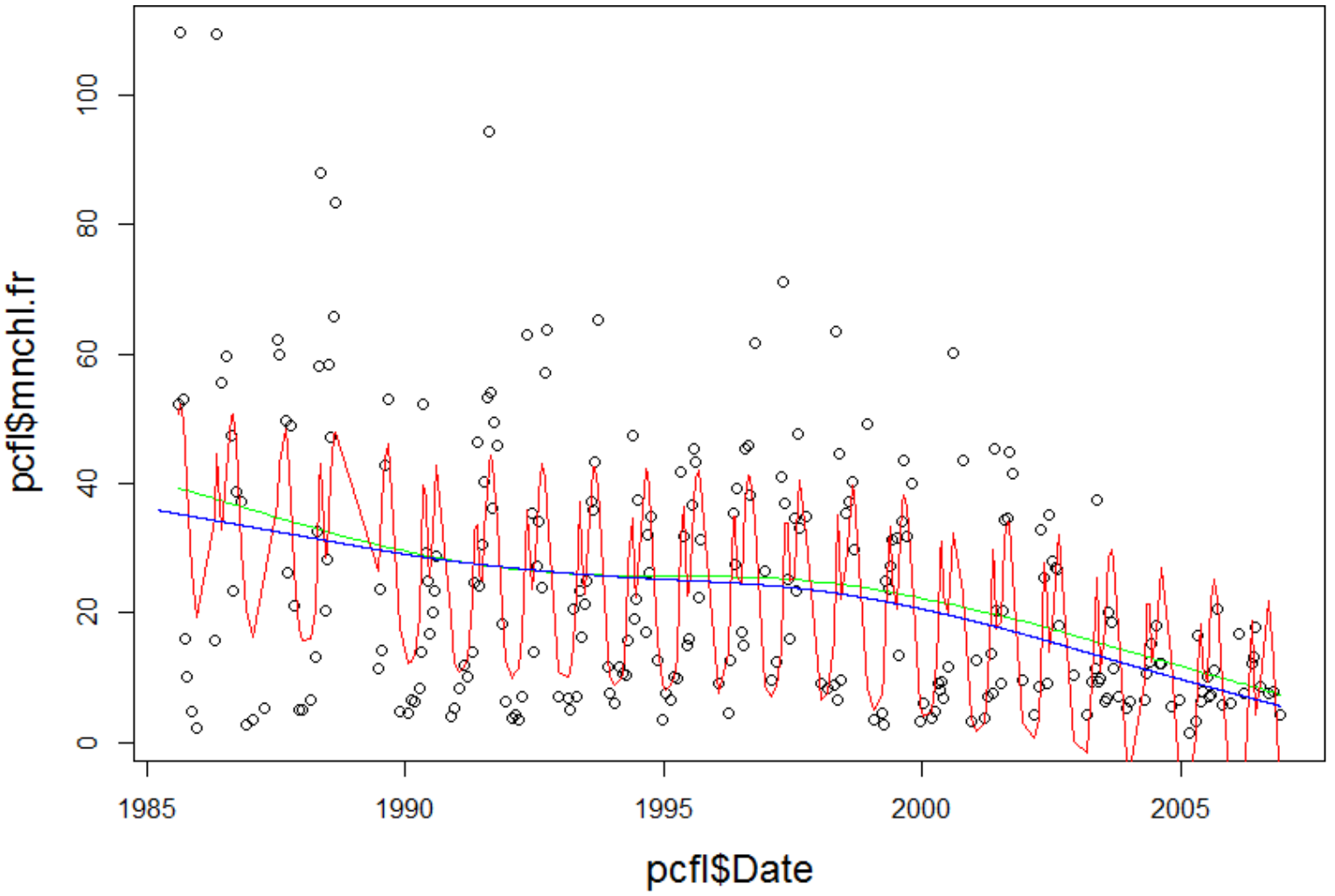

added variable plot
mnchl.fr / as.factor(month)
sun / as.factor(month)
-20
0
20
40
60
-2
-1
0
1
2
3
4


added variable plot
mnchl.fr / as.factor(month)
par4 / as.factor(month)
-20
-10
0
10
-20
0
20
40
60

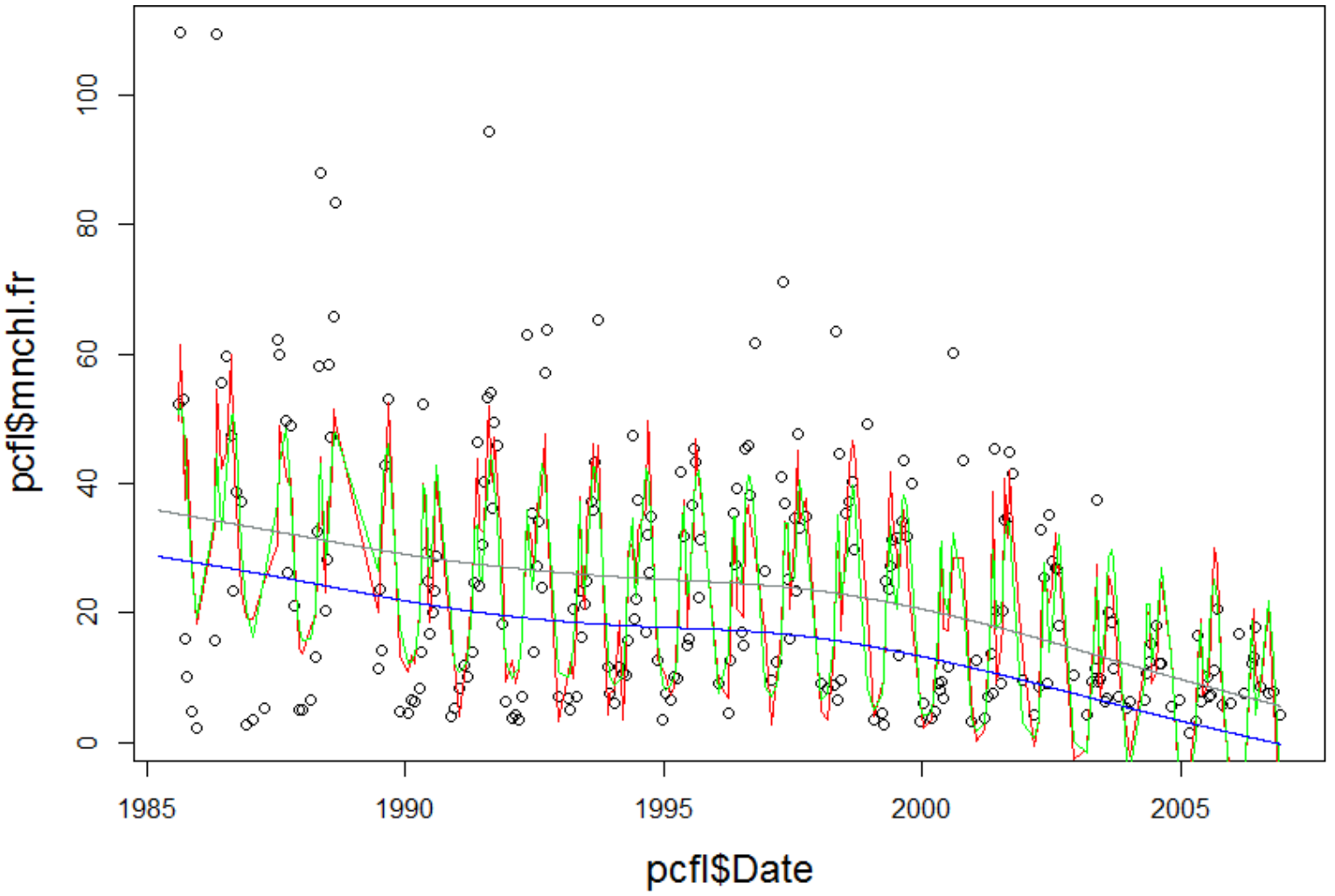
Finally a gam trend line



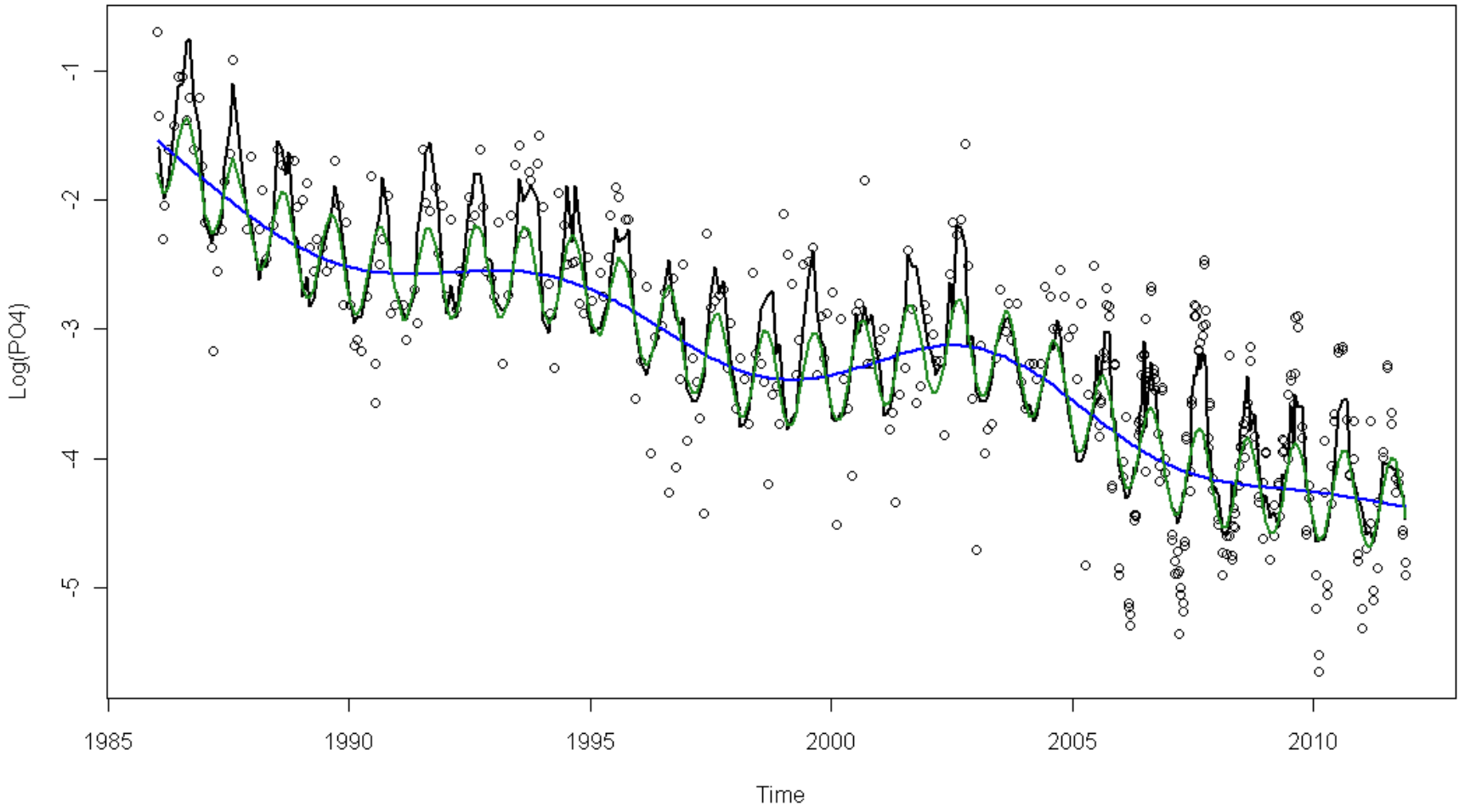
Seasonally adjusted trend



Trend(season,flow)



Pax fall line PO4 loads



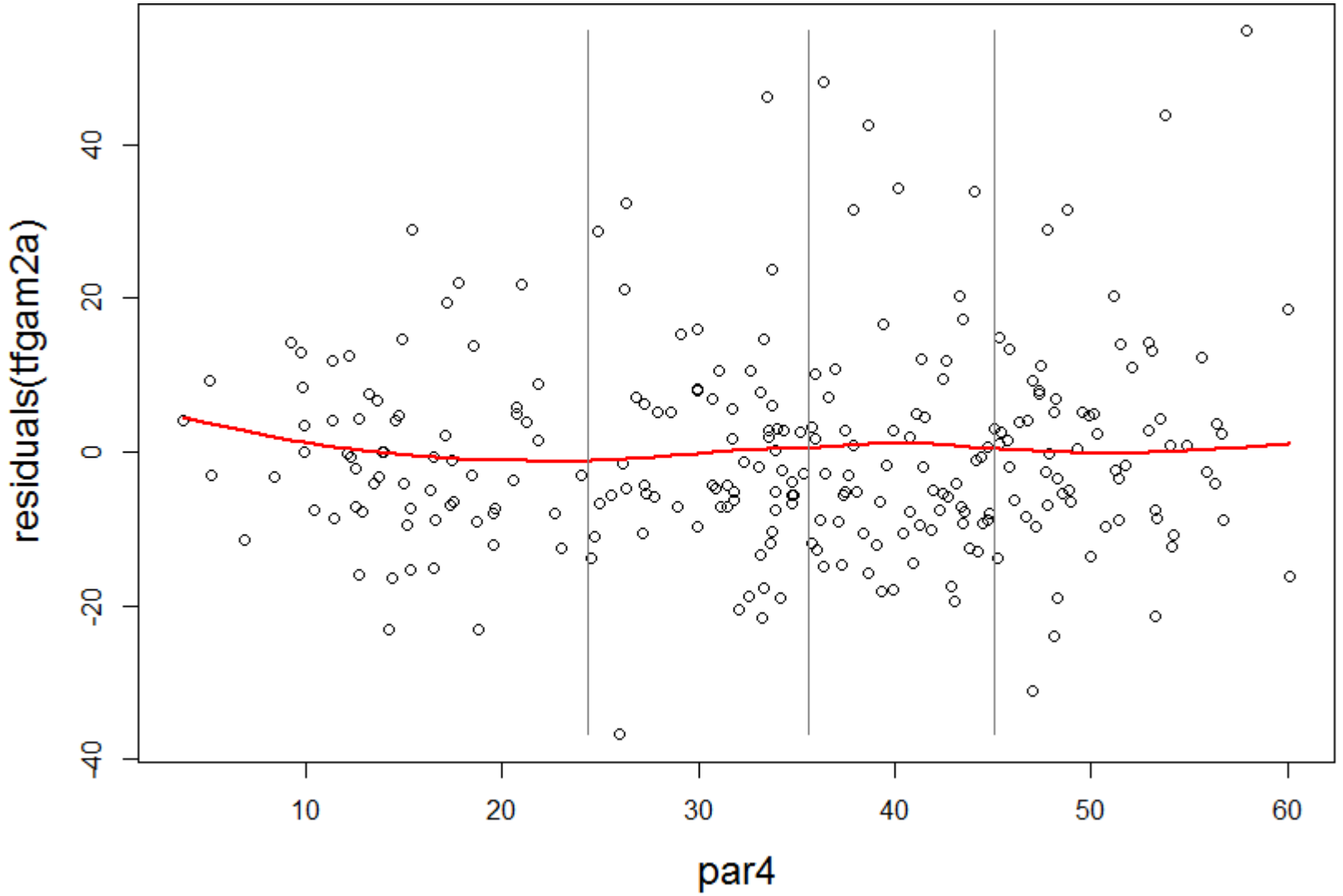
Conclusions

TF chorophyll is going down

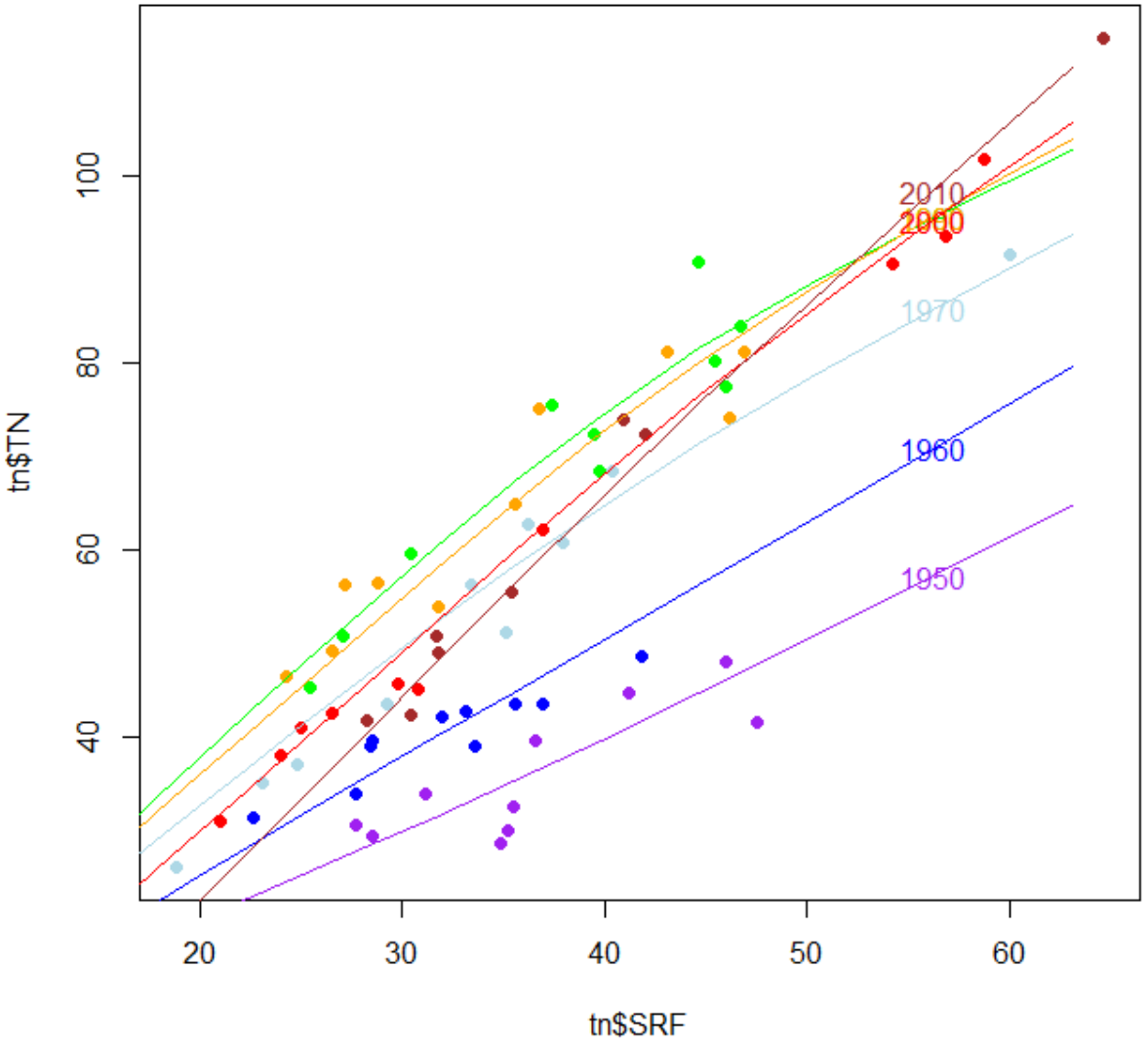
Effective management actions were in 1980's and post 2000.

Analysis Conclusions

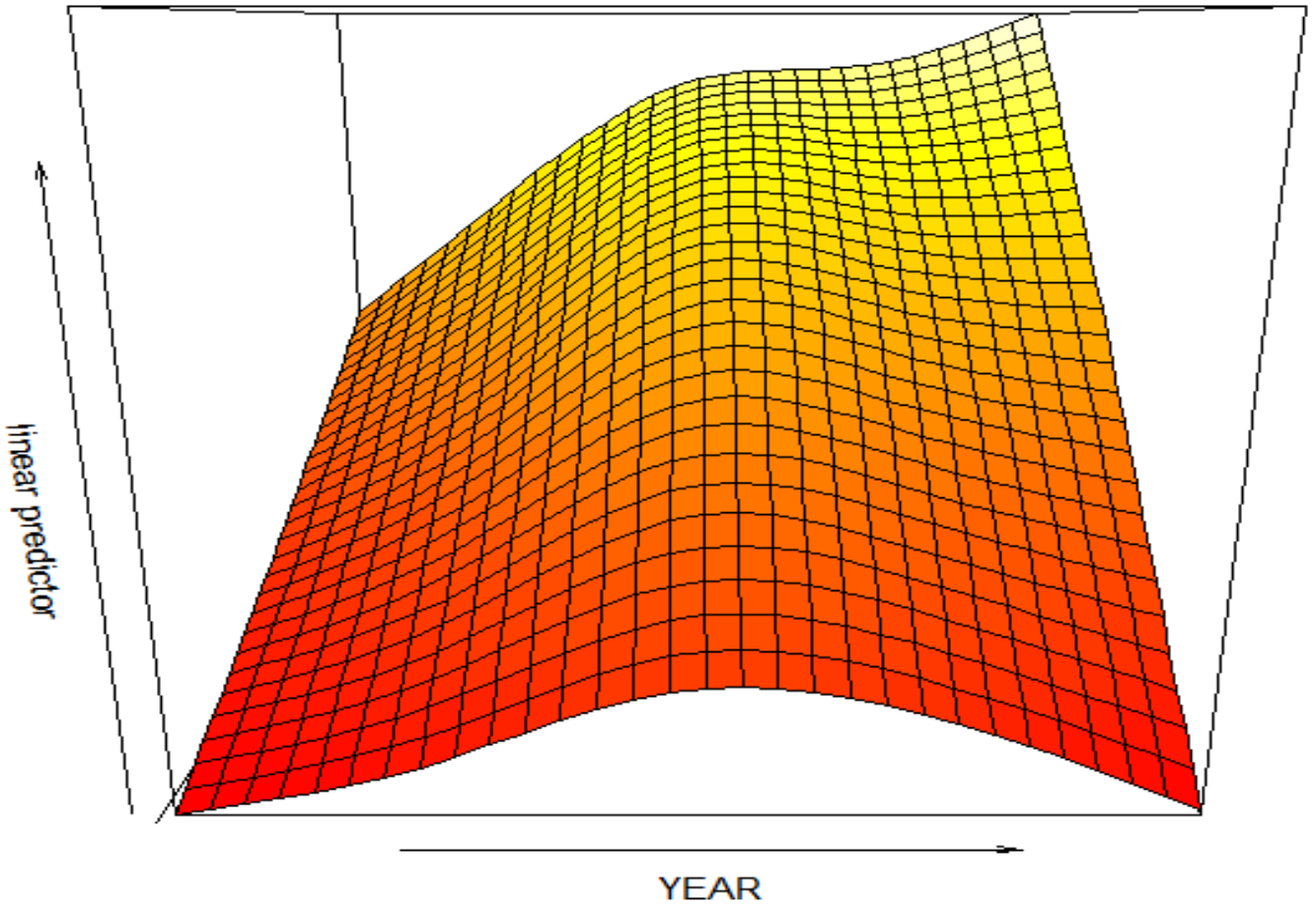
- Cyclical DOY is good seasonal method
- No season x time interaction
- No flow x time interaction
- Event.days, Event.size = deadend
- Light = deadend


residuals(tfgam2a)
40
20
0
-20
-40
10
20
30
40
50
60
par4

Susquehanna TN load versus flow over time



TN load = f(time,flow)



Data Needs

1. Flow
2. Nutrient Loads
3. Daily Loads
4. Point Source vs. Non-Point Source
5. Below Fall Line Component of Load