AutoBeale: the Basics

R. Peter Richards
peterichards@oberlin.net

Outline

- Beale Ratio Estimator
- History and Use
- AutoBeale

Beale Ratio Estimator

- I have n days with observations of load (l) and discharge (q)
- I have N total days with observations of discharge on all of them
- How can I adjust my observed daily load to estimate the true daily load for the period N?

Aside: Estimators and NCWQR data

- Estimators attempt to “fill in or adjust for missing data”
- In WQ work, data are usually daily values
- NCWQR is saturation sampling – other than rare failures, there are no missing data
- Load computation for such data does not require an estimator!
- But – we get a useful measure of uncertainty from such estimators

Beale Ratio Estimator

- Assumes only that daily load increases with daily discharge

$$\bar{l}_a = \bar{l}_o \frac{q_a}{q_o} \left[\frac{1 + \left(\frac{1}{n} - \frac{1}{N} \right) \frac{s_{lq}}{\bar{l}_o \bar{q}_o}}{1 + \left(\frac{1}{n} - \frac{1}{N} \right) \frac{s_{qq}}{\bar{q}_o^2}} \right]$$

Beale Ratio Estimator

The ratio part:

$$\frac{\bar{l}_a}{\bar{q}_a} = \frac{\bar{l}_o}{\bar{q}_o}$$

or

$$\bar{l}_a = \bar{l}_o \frac{q_a}{q_o} \left[\frac{1 + \left(\frac{1}{n} - \frac{1}{N} \right) \frac{s_{lq}}{\bar{l}_o \bar{q}_o}}{1 + \left(\frac{1}{n} - \frac{1}{N} \right) \frac{s_{qq}}{\bar{q}_o^2}} \right]$$

Beale Ratio Estimator

The Beale part:

$$\bar{l}_a = \bar{l}_o \frac{q_a}{q_o} \left[\frac{1 + \left(\frac{1}{n} - \frac{1}{N} \right) \frac{s_{lq}}{\bar{l}_o \bar{q}_o}}{1 + \left(\frac{1}{n} - \frac{1}{N} \right) \frac{s_{qq}}{\bar{q}_o^2}} \right]$$

History and Use

- Tin (1965), Cochran (1963)
- Wisconsin DNR – Bannerman, Baun (1982)
- IJC – John Clark
- In GL, always been used in stratified mode
  - Stratification by flow or by time
  - Primarily to reduce variance of the estimate
  - Also reduces the effects of curvilinearity in l/q relationship

More formulas

$$\mathrm{MSE} = \bar{l}^2\left[\left(\frac{1}{n}-\frac{1}{N}\right)\left(\frac{s_{qq}}{\bar{q}^2}+\frac{s_{ll}}{\bar{l}^2}-2\frac{s_{lq}}{\bar{l}\bar{q}}\right)+\left(\frac{1}{n}-\frac{1}{N}\right)^2\left(2\frac{s_{qq}^2}{\bar{q}^4}-4\frac{s_{qq}}{\bar{q}^2}\frac{s_{lq}}{\bar{l}\bar{q}}+\frac{s_{lq}^2}{\left(\bar{l}\bar{q}\right)^2}+\frac{s_{qq}}{\bar{q}^2}\frac{s_{ll}}{\bar{l}^2}\right)\right]$$

(in each stratum)

$$f_{\mathrm{eff}} = \frac{\left(\sum_i \frac{N_i^2}{f_i+1} s_i^2\right)^2}{\sum_i \frac{\left(\frac{N_i^2}{f_i+1}\right)^2 s_i^4}{f_i}}$$

Effective degrees of freedom

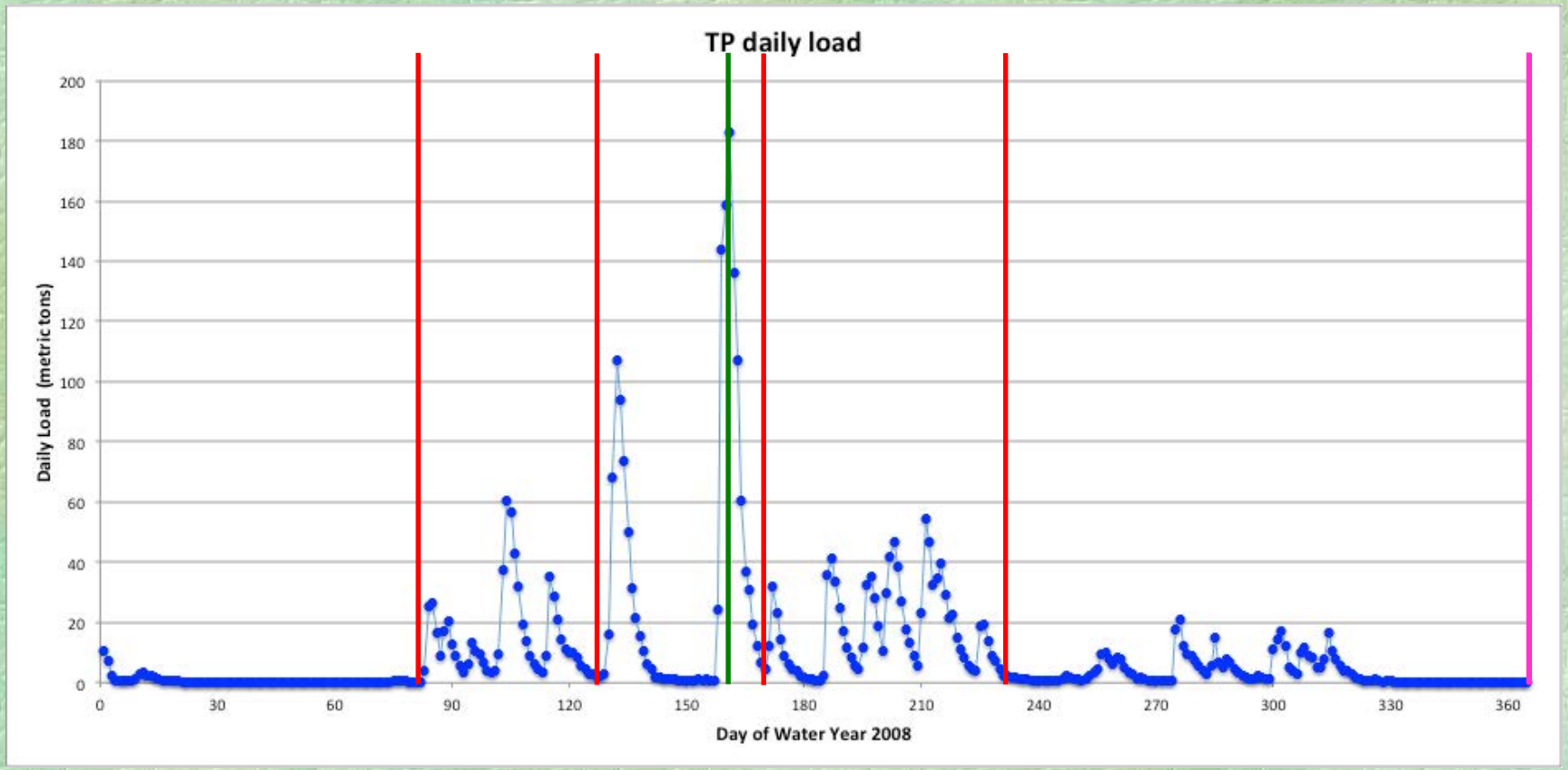
$$n_i = n\frac{N_i s_i}{\sum N_i s_i}$$

Optimal allocation of samples among strata (usually not relevant because sampling has already been done!)

AutoBeale

- At IJC, stratification was done by John Clark using BPJ – i.e. subjectively
- AutoBeale (ca. 1995) is an attempt to automate stratification – i.e. objectively
- Objective – identify time stratification that minimizes overall MSE for a given set of data
- Approach – sequential systematic search
- Quit criterion – less than 0.5% reduction in MSE

AutoBeale


TP daily load
Daily Load (metric tons)
0
20
40
60
80
100
120
140
160
180
200
0
30
60
90
120
150
180
210
240
270
300
330
360
Day of Water Year 2008

AutoBeale Weaknesses

- FORTRAN
- Occasional obscure bugs
- Sequential approach to a simultaneous problem
- What other approaches could be used to do autostratification?

Citations

- Baun, K. 1982. *Alternative Methods of Estimating Pollutant Loads in Flowing Water*. Tech. Bulletin 133, Dept. Natural Resources, Madison, Wisconsin. 11 pages.
- Cochran, W.G. 1963. *Sampling Techniques* (2nd edition). Wiley Publications in Statistics. John Wiley and Sons, New York
- Tin, M. 1965. Comparison of some ratio estimators. *J. Am. Stat. Assoc. 60:* 294-307.


Rem…