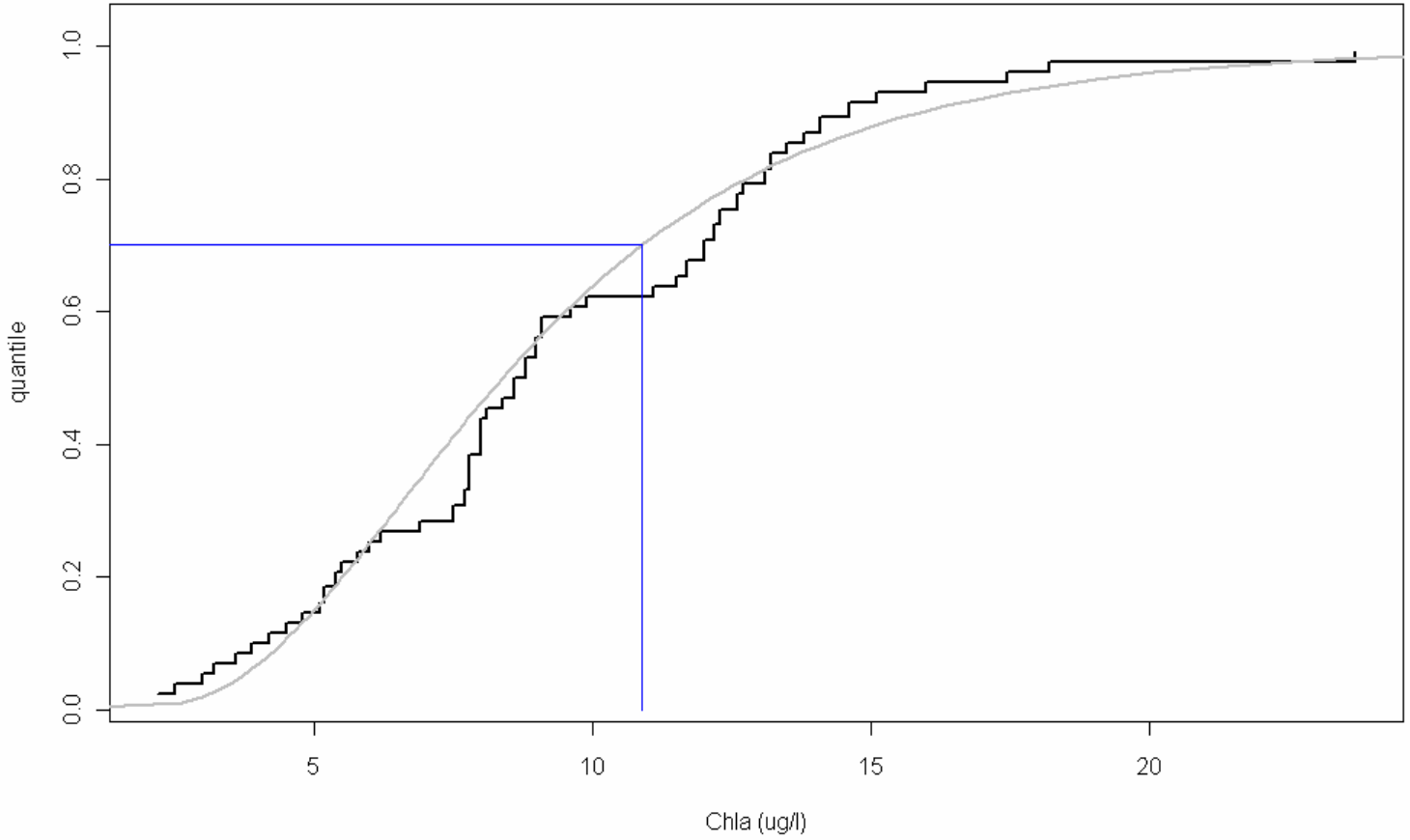
Notes on absolute status based on CDF scoring functions

Elgin Perry

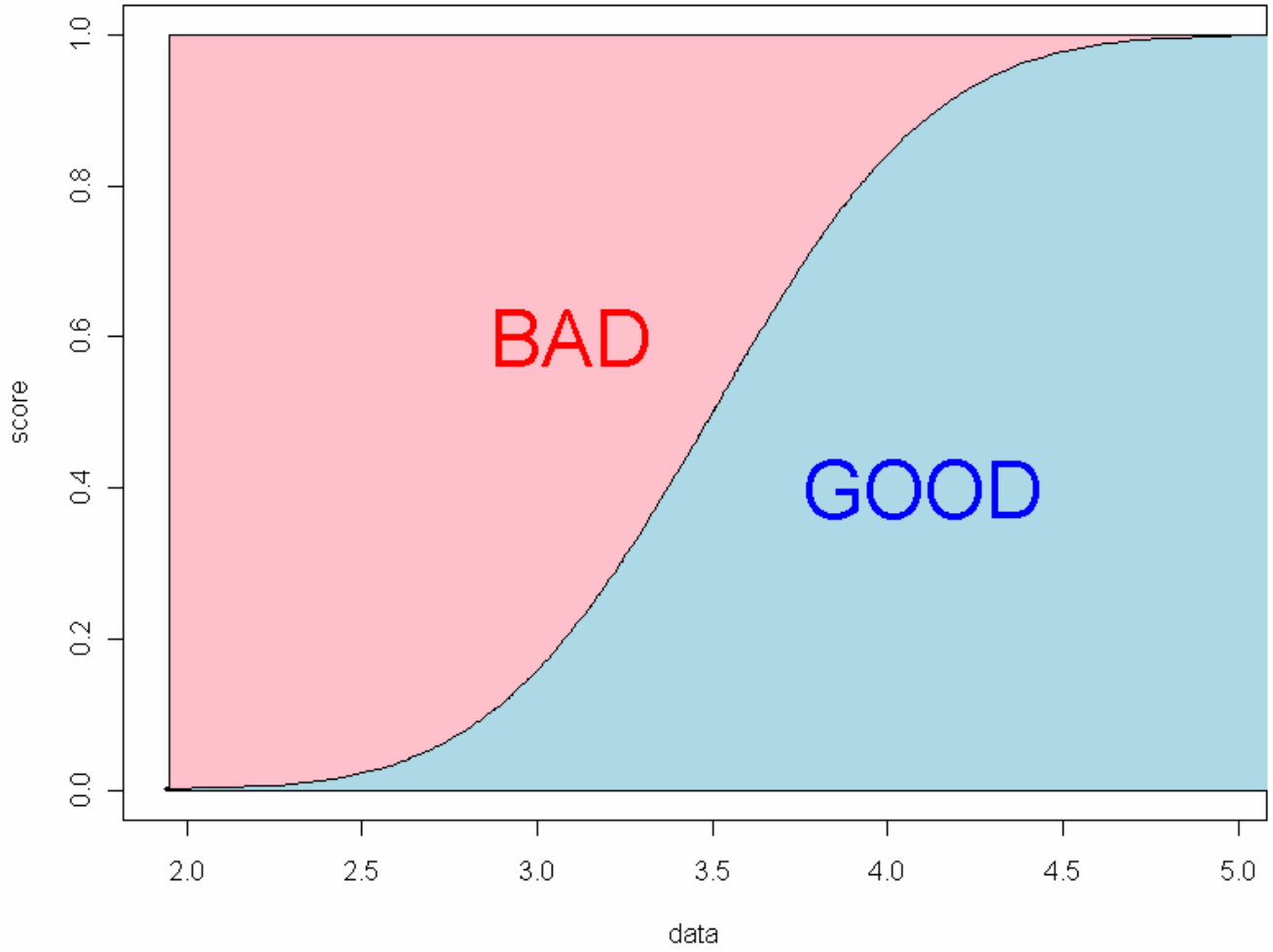
6/12/2012

Presented to TMAW 10/07/2010
reviewed 5/4/2011
refreshed 6/12/2012

Basic Concept -
Develop a Scoring Function based on a <u>parametric</u> Cumulative Distribution Function (CDF).



This function should describe the distribution that is right on the cusp of being impaired



Scoring Functions

Normal CDF

$$F(x;\mu,\sigma) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\mu)^2}{2\sigma^2}} dz$$

Logistic CDF

$$F(x;\alpha,\beta) = \frac{1}{(1+e^{-(x-\alpha)/\beta})^2}$$

LogNormal Distribution.

$$F(Y;\mu,\sigma) = \int_{y=0}^{Y} \frac{1}{\sigma y\sqrt{2\pi}} e^{-\frac{(\log(y)-\mu)^2}{2\sigma^2}} dy$$

Poisson Distribution:

$$F(X;\lambda) = \sum_{x=0}^{X} \frac{e^{-\lambda}\lambda^x}{x!} \quad x = 0,1,2... \quad \lambda > 0$$

Binomial Distribution:

$$F(X;p)=\sum_{x=0}^{X}\binom{n}{x}p^{x}(1-p)^{n-x} \qquad x=0,1,2\ldots n \quad 0<p<1$$

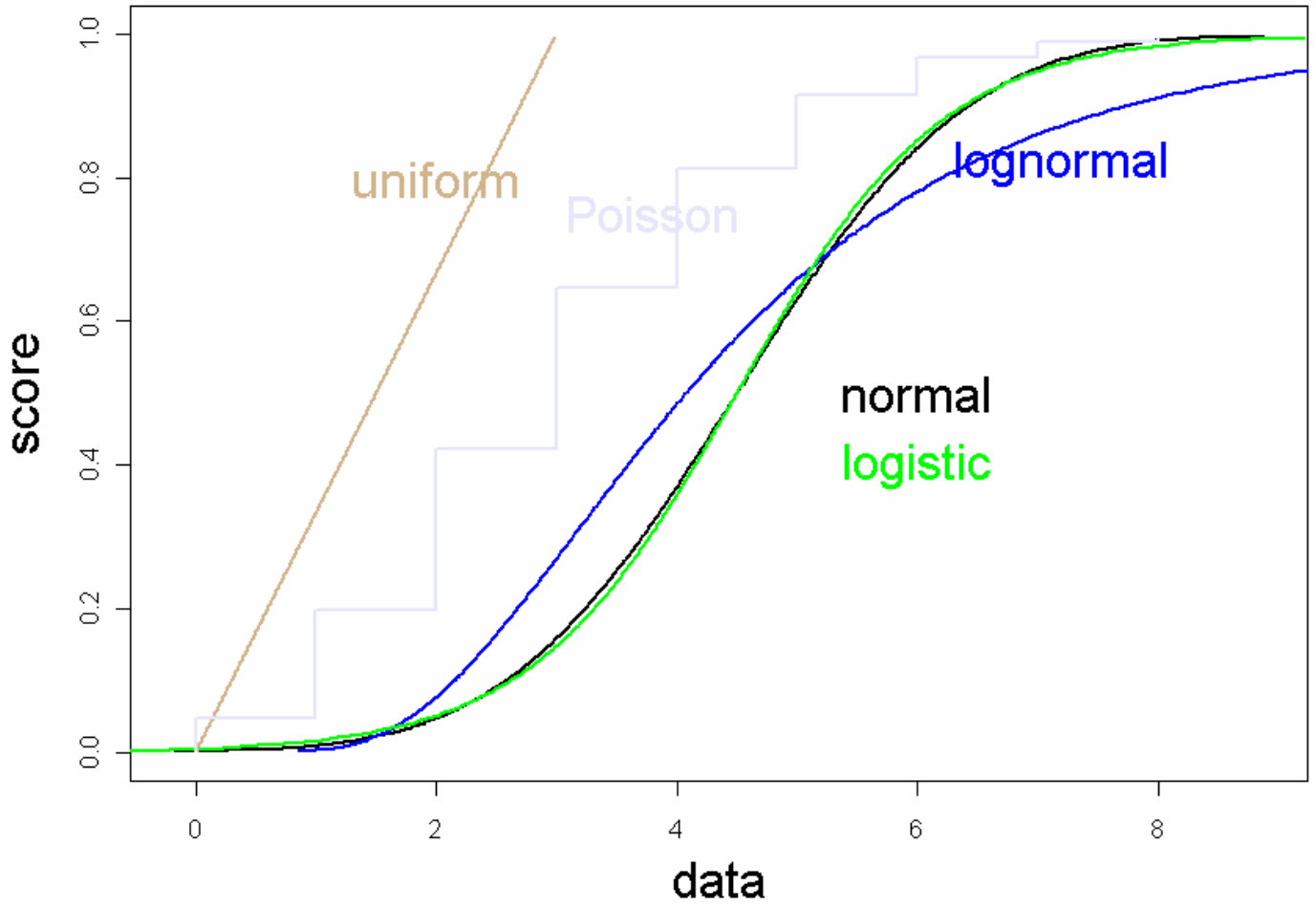
or $\log\left(\frac{p}{1-p}\right)$ **+ logistic CDF**

or $y=\sin^{-1}\sqrt{(r+3/8)/(n+3/4)}$ **+ normal CDF**

or **Beta CDF**

$$F(X;a,b)=\int_{0}^{X}\frac{x^{(a-1)}(1-x)^{(b-1)}}{B(a,b)}dx \qquad 0\le x\le 1, \qquad a>0,b>$$

Pictures of scoring functions



How to quantify scoring function

- reference data,
- the water quality model,
- values taken from literature review,
- best professional judgment,
- shifted distribution

minimum requirements are central tendency and measure of variance.

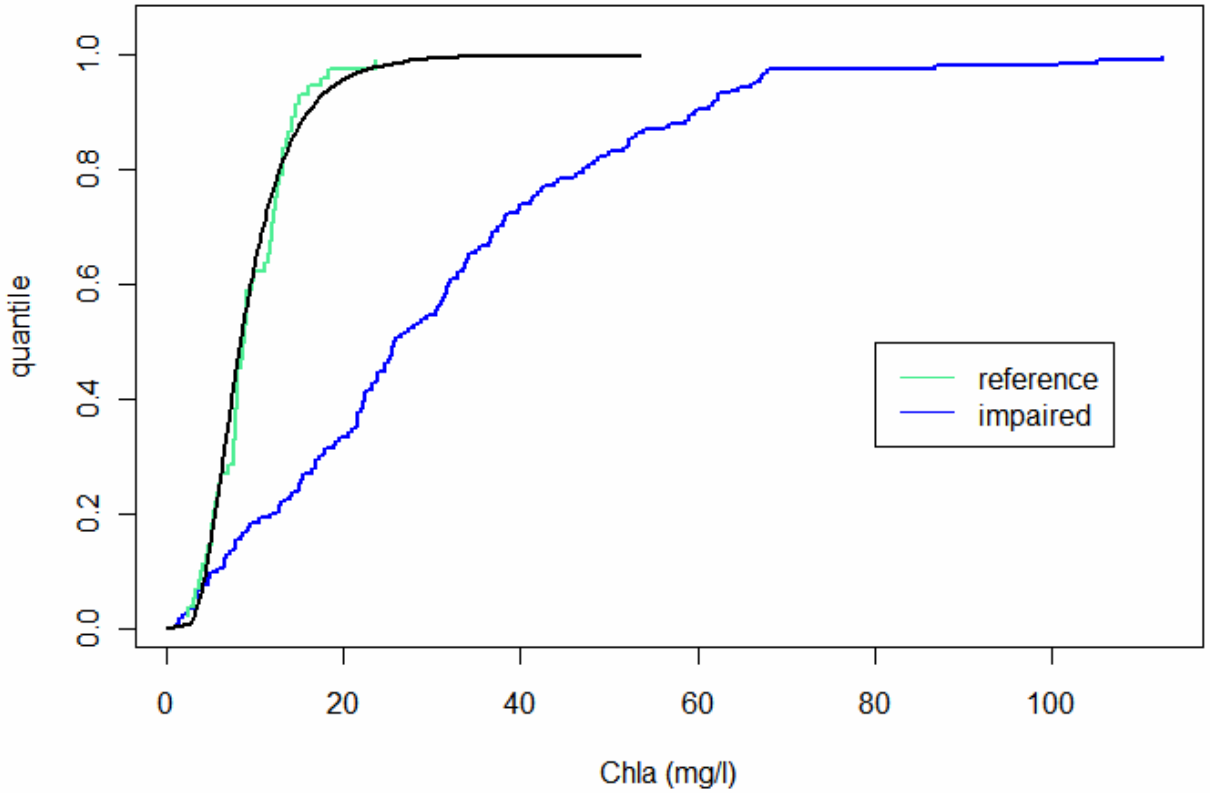


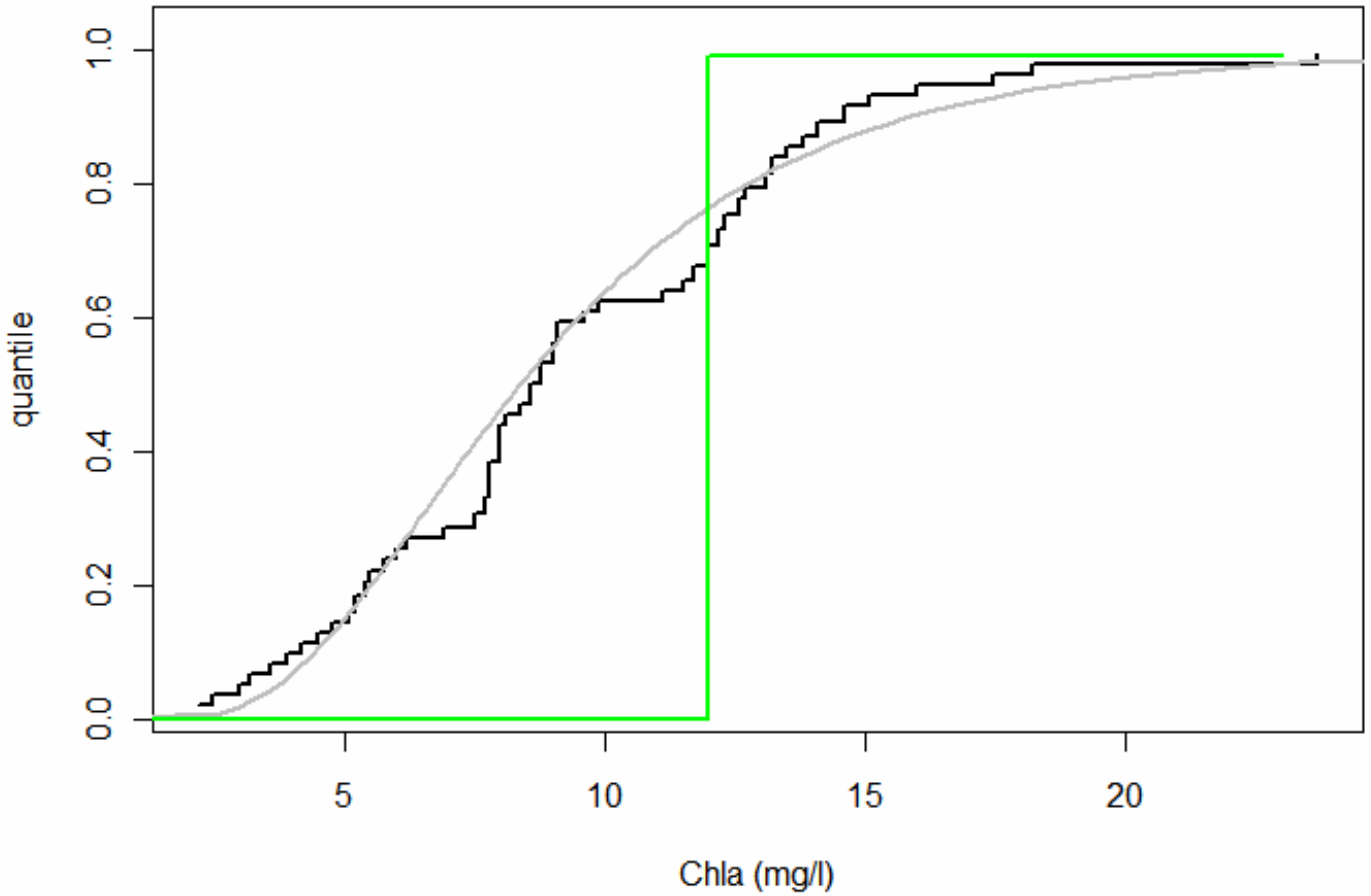
Figure 1. The empirical distribution function of chlorophyll a from 'good' regions (green step function) with an overlay of the log-normal scoring function (smooth black curve) and the ECDF of chlorophyll a from 'bad' regions (blue step function).

Assessment Endpoints and inference:

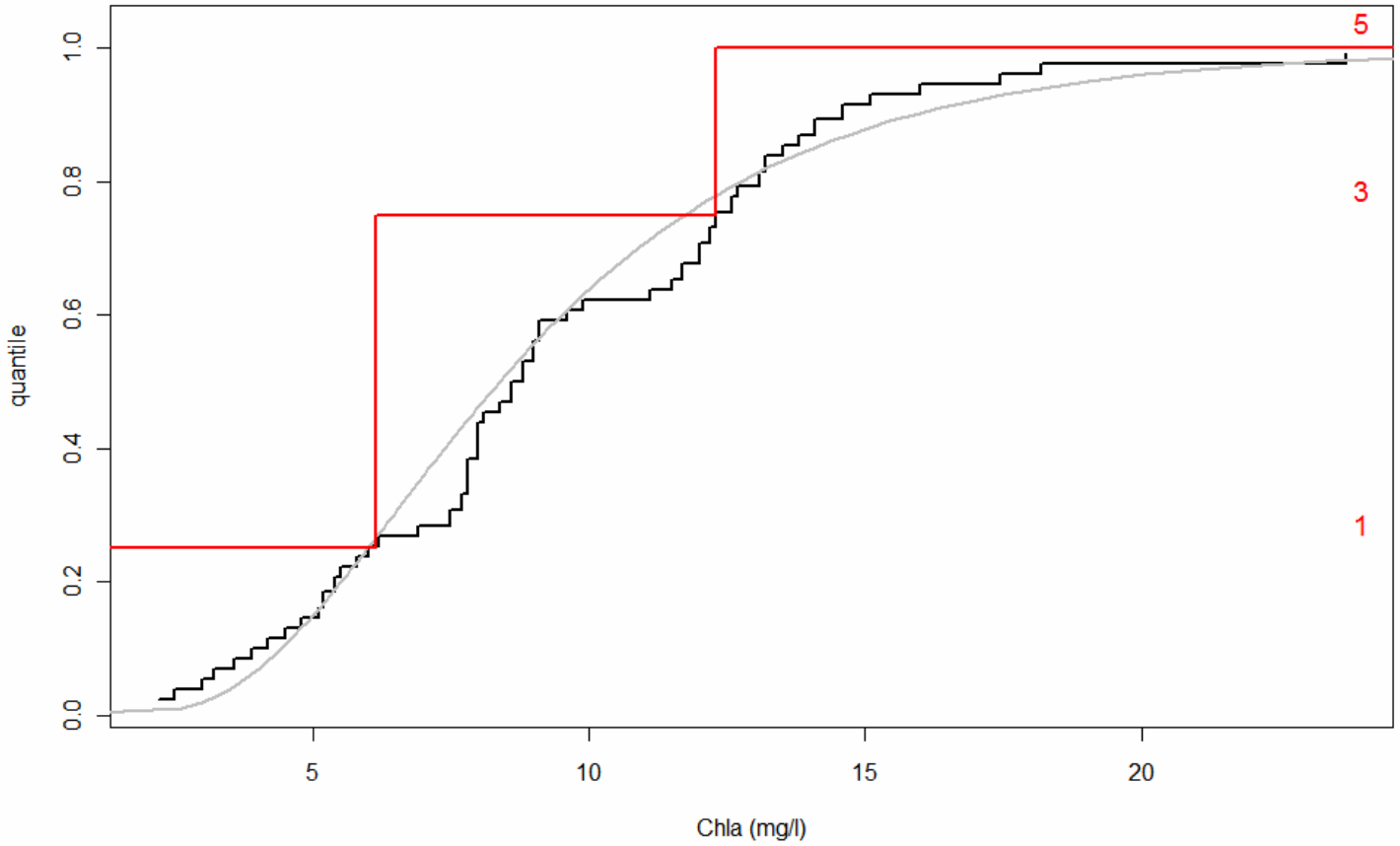
- median quantile score / beta distribution
- mean quantile score / ? distribution
- distribution tests, e.g. Kolmogorov-Smirnov, Shapiro-Wilks
- bright line criteria with scientific basis

Comparison to other approaches:

Single number criteria:

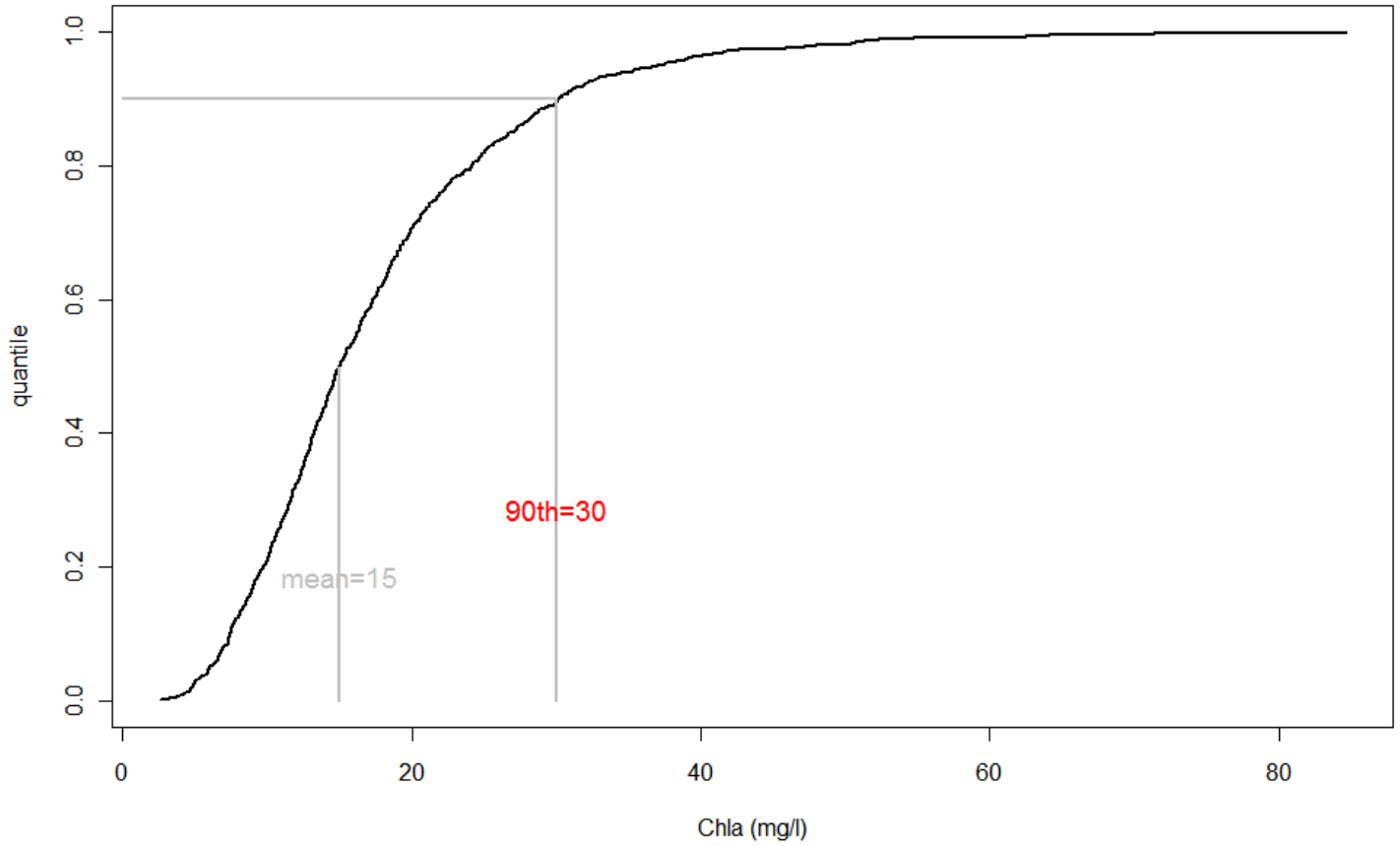


IBI scoring approach:

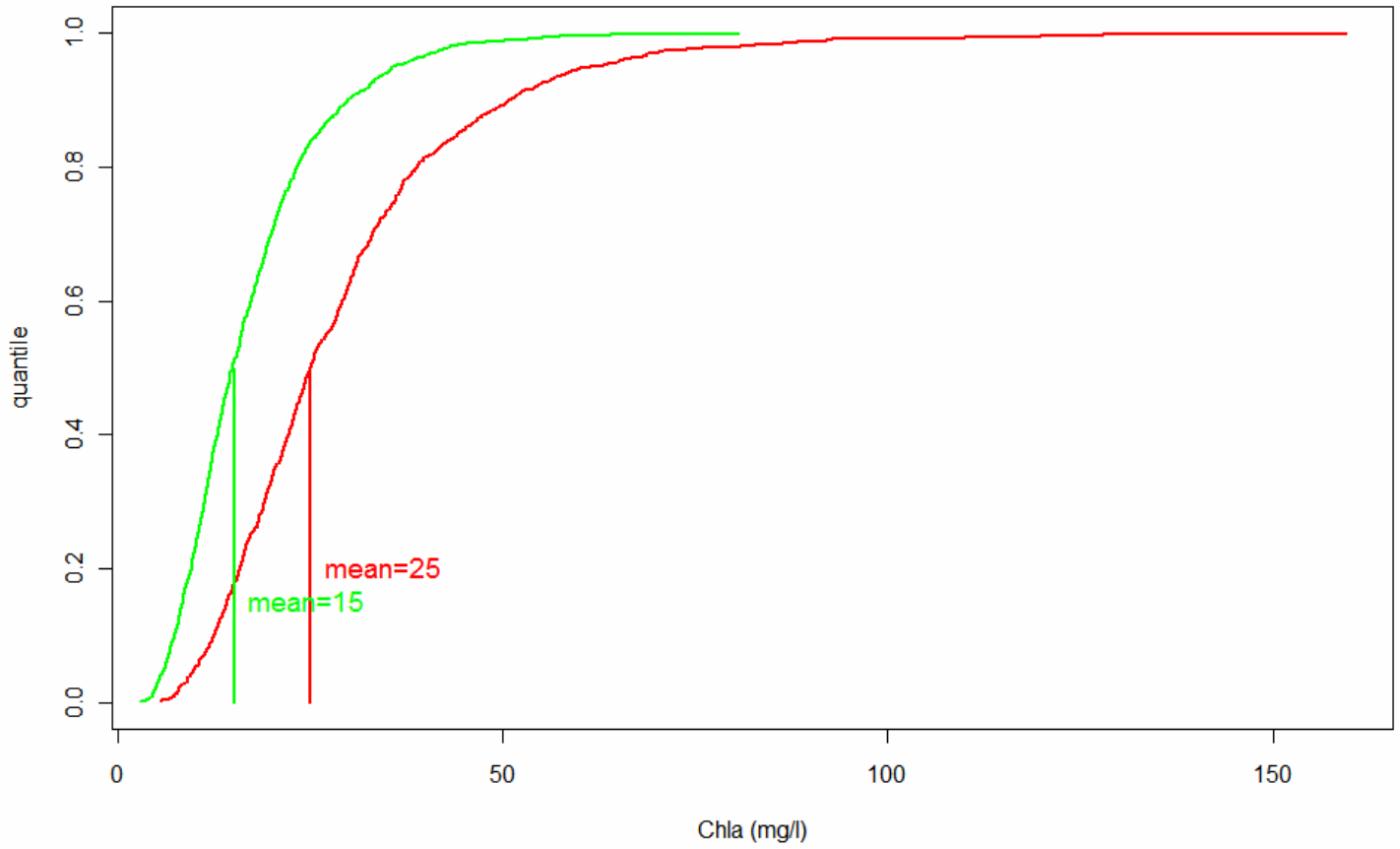


Examples of specifying Reference Distribution:

Mean and upper bound: mean = 15, exceed 30 not more than 10%



Distribution Shift:



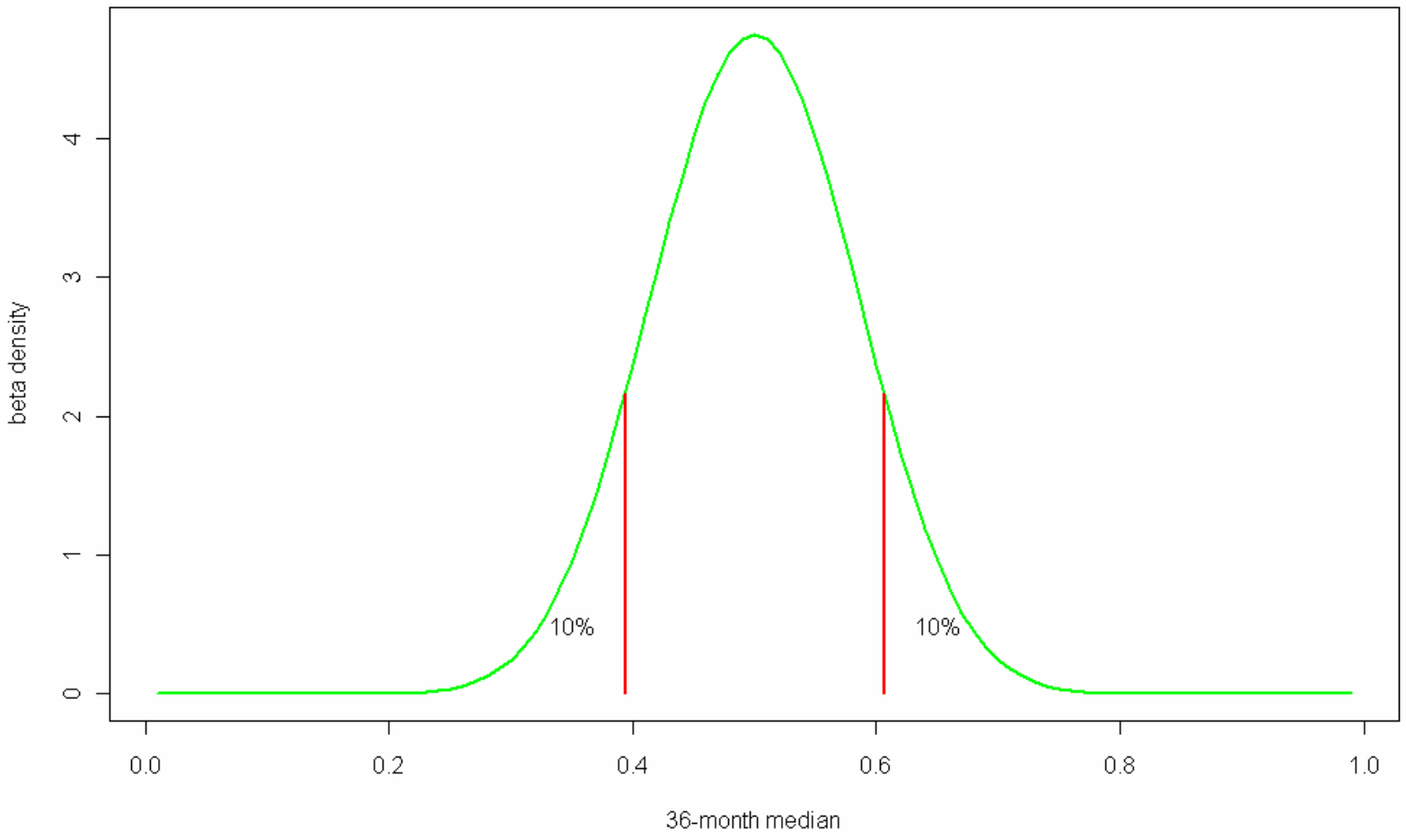


Figure 2. Beta density (B,18,18) showing 90% critical point

Advantages:

- **preserves continuity and ranking of original data**
- **acknowledges overlap of good and bad**
- **scores data to universal 0-1 scale to facilitate averaging over dissimilar scales**
- **scored data have uniform 0-1 statistical distribution under probability integral transform**
- **medians of scored data follow beta distribution**
- **easy to explain through grading on a curve analogy**
- **minimal requirements for specifying reference distribution**
- **has the potential to accommodate context sensitive criteria**

Outstanding issues:

- Choice of scoring function
- Nonparametric density scoring
- Risk if density if mis-specified
- space-time issues
- potential for context sensitive status assessment
- outside review

Example of Shifted Distribution Scoring Function for 7-day mean of Dissolved Oxygen Using York River Basin ConMon Data.

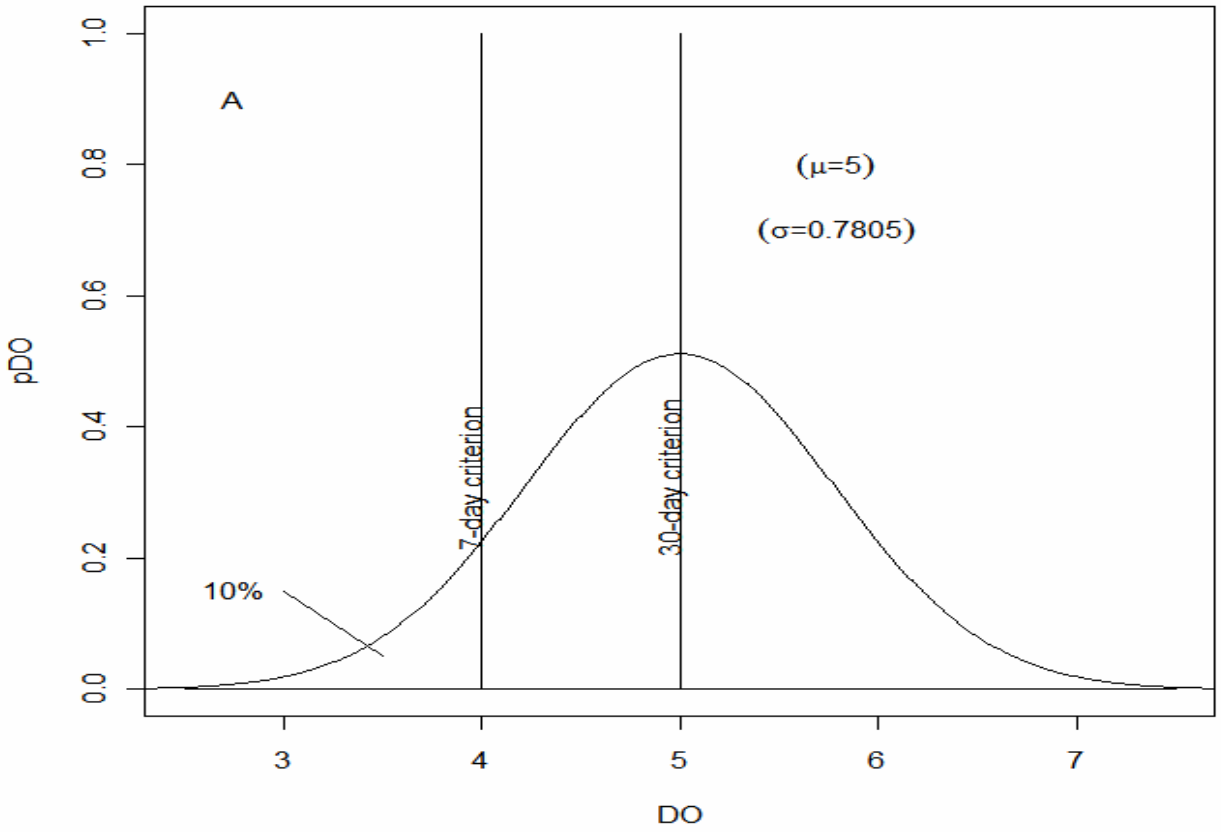
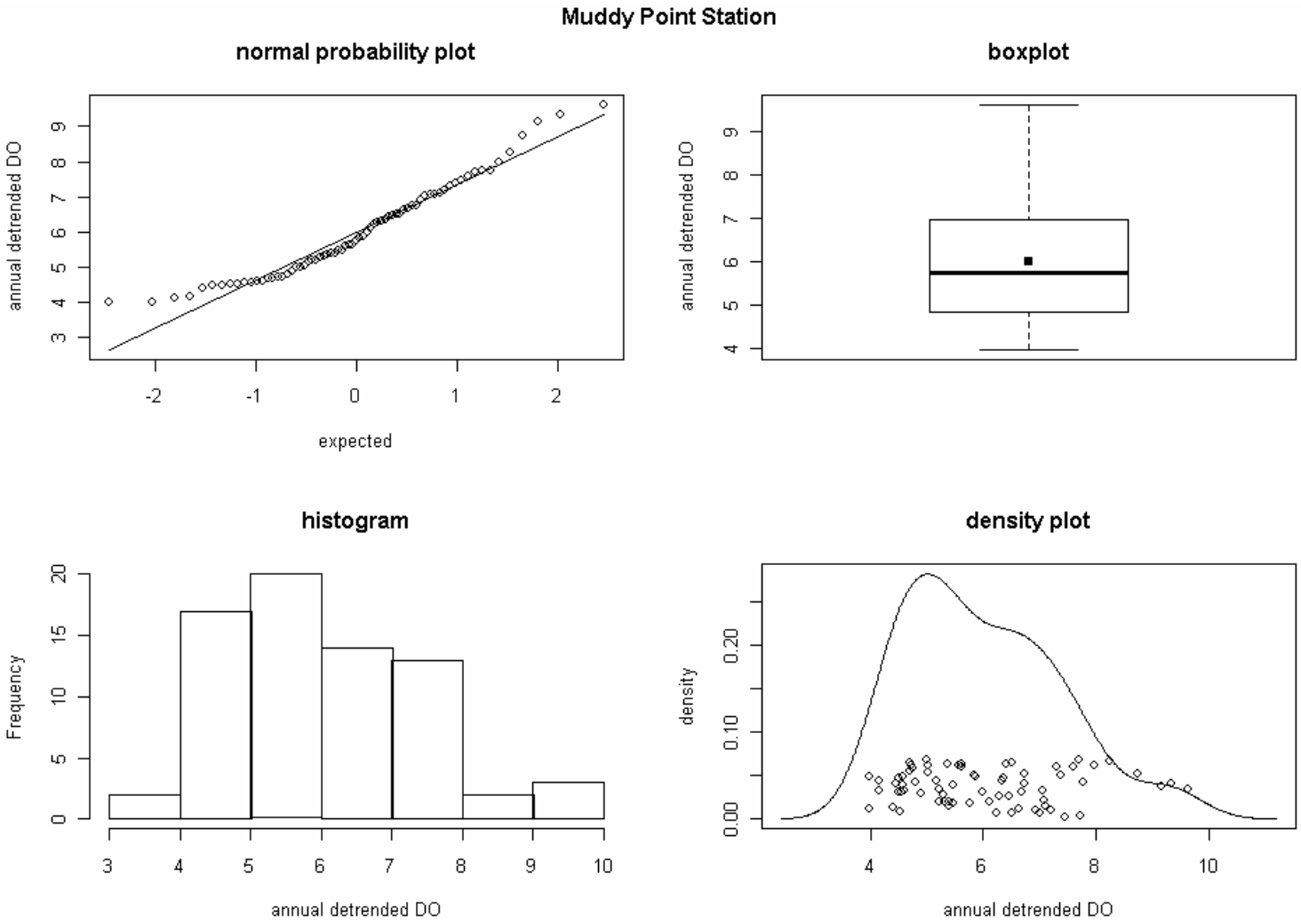


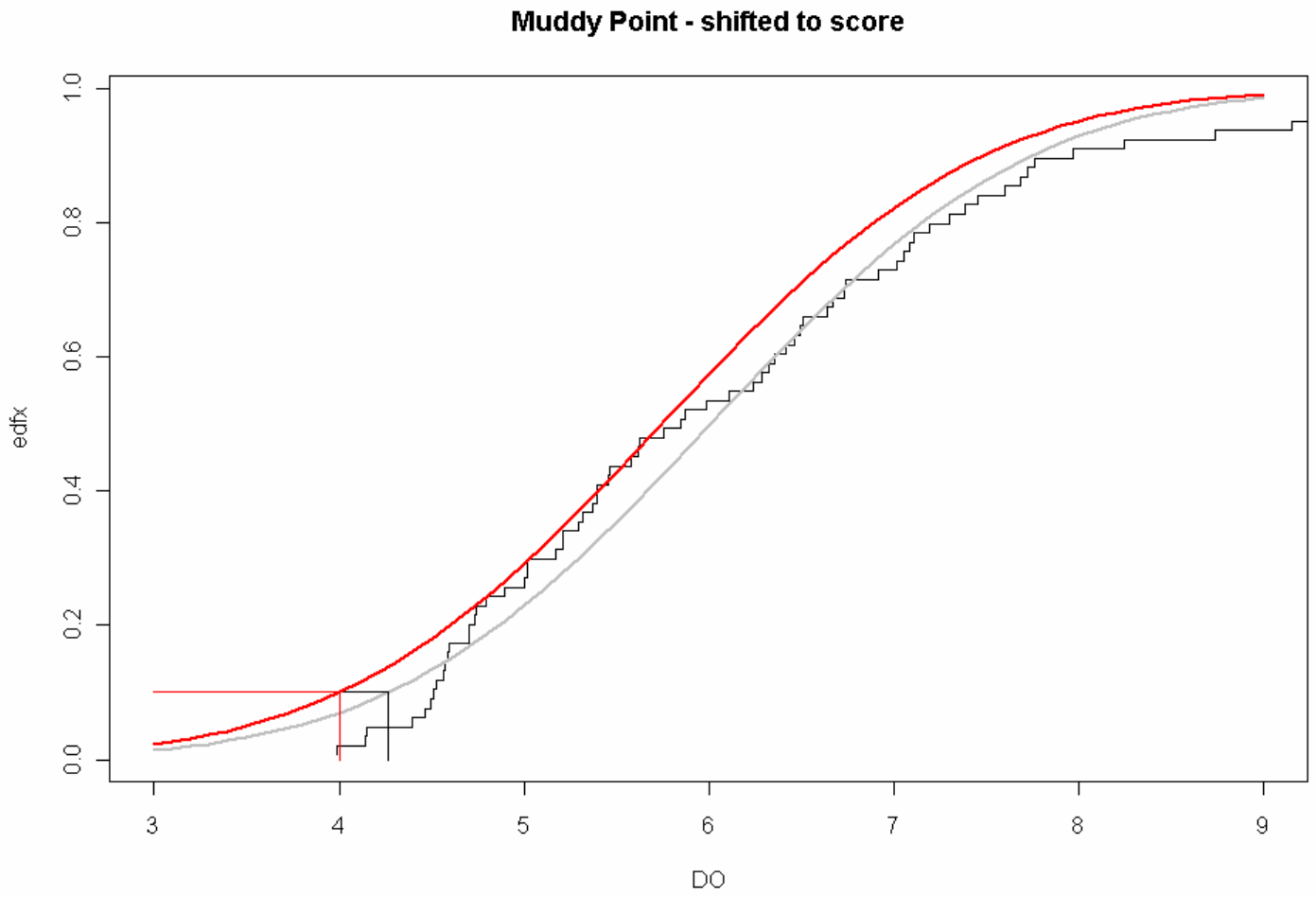
Table 1. Locations of continuous monitors in the York river and tributaries.

River	Location Name	Name Abbrev.	Data Code	Years
Pamumkey	White House Landing	wh	PMK034.00	2003-2010
Pamumkey	Sweet Hall Marsh	sh	PMK012.18	2003-2010
Mattaponi	Walkerton Landing	wl	MPN028.86	2003-2005
Mattaponi	Muddy Point	mp	MPN005.04	2003-2005
York	Taskinas Creek	tc	TSK000.23	2003-2010
York	Claybank	cb	YRK015.09	2003-2010
York	Yorktown	yk	YRK006.77	2003-2005
York	Gloucester Point	gp	YRK005.40	2003-2010
York	Goodwin Islands	gi	CHE019.38	2006-2010

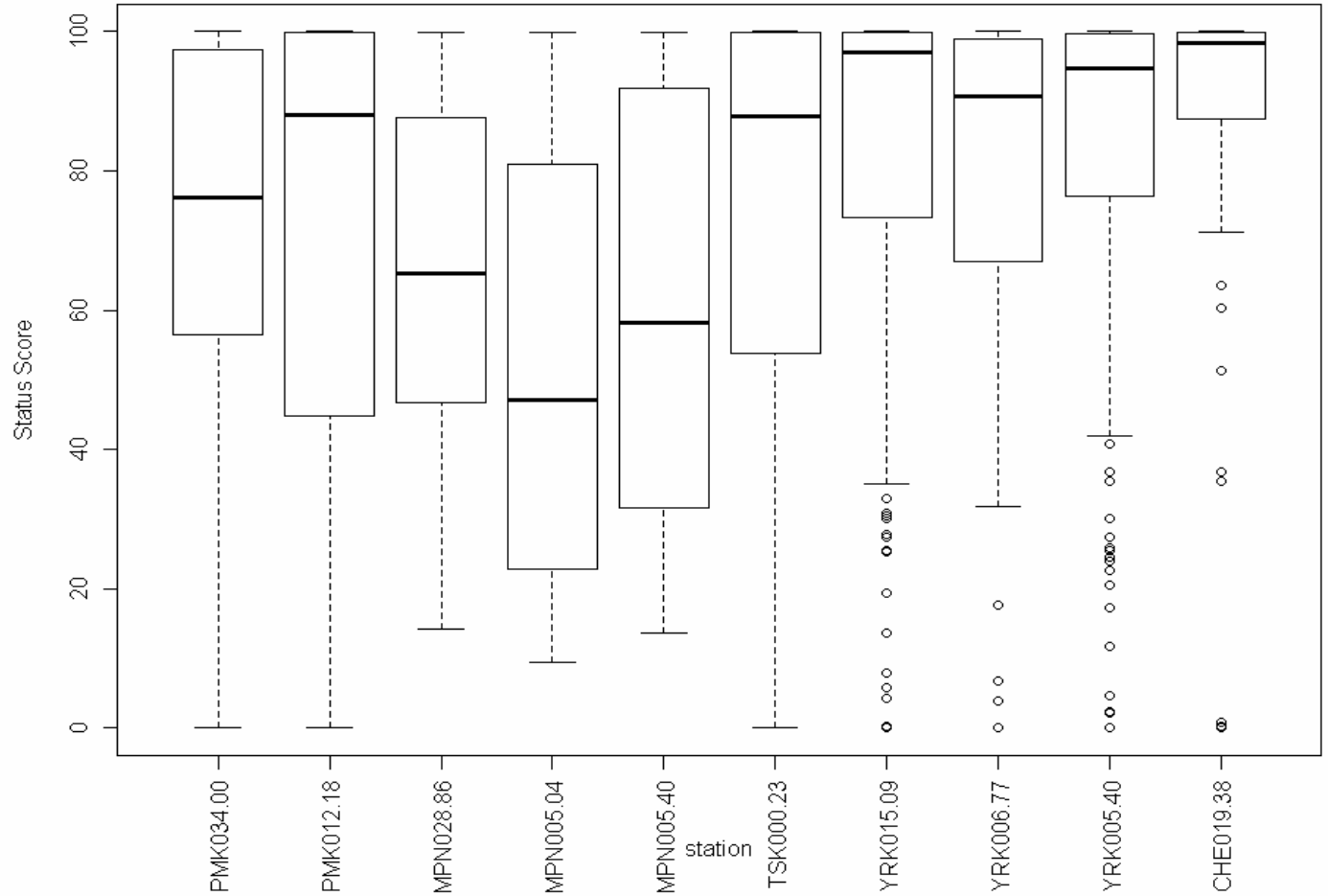
Walkerton
Stevensville
Manquin
King William
King and Queen Court House
Mattaponi Indian Reservation
Jamaica
Water View
Mollusk
Merry Point
Eastern Branch
Kilmarnock
Church View
Corrotoman River
Dymer Creek
Irvington
Weems
White Stone
Warner
Urbanna
Little Plymouth
Mascot
Christchurch
Locust Hill
Topping
17S
Pamunkey Indian Reservation
wh
sh
mp
Wake
Grafton
Hardyville
Hartfield
Deltaville
Shacklefords
Quinton
West Point
Mattaponi
Cologne
Piankatank River
New Kent
Cobbs Creek
Dutton
Hallieford
64
Church Hill
James Store
Hudgins
Barhamsville
Foster
Providence Forge
Mt Folly
Ark
North
Lanexa
te
Cardinal
17
East River
Ware Neck
Schley
Bohannon
Zanoni
Toano
Norge
Mobjack
Chickahominy River
Ruthville
Charles City
eb
White Marsh
Naxera
Camp Peary
60
Mobjack Bay
Chickahominy State Wildlife Management Area
Ordinary
Severn
Wicomico
Achilles
Maryus
Hayes
Bena
Gordon Creek
Williamsburg
gp
Gloucester Point
yk
Yorktown
Chickahominy River
Lackey
Claremont
Colonial National Historical Park
gi
64
Jamestown
Seaford


Muddy Point Station
normal probability plot
annual detrended DO
expected
boxplot
annual detrended DO
histogram
Frequency
annual detrended DO
density plot
density
annual detrended DO

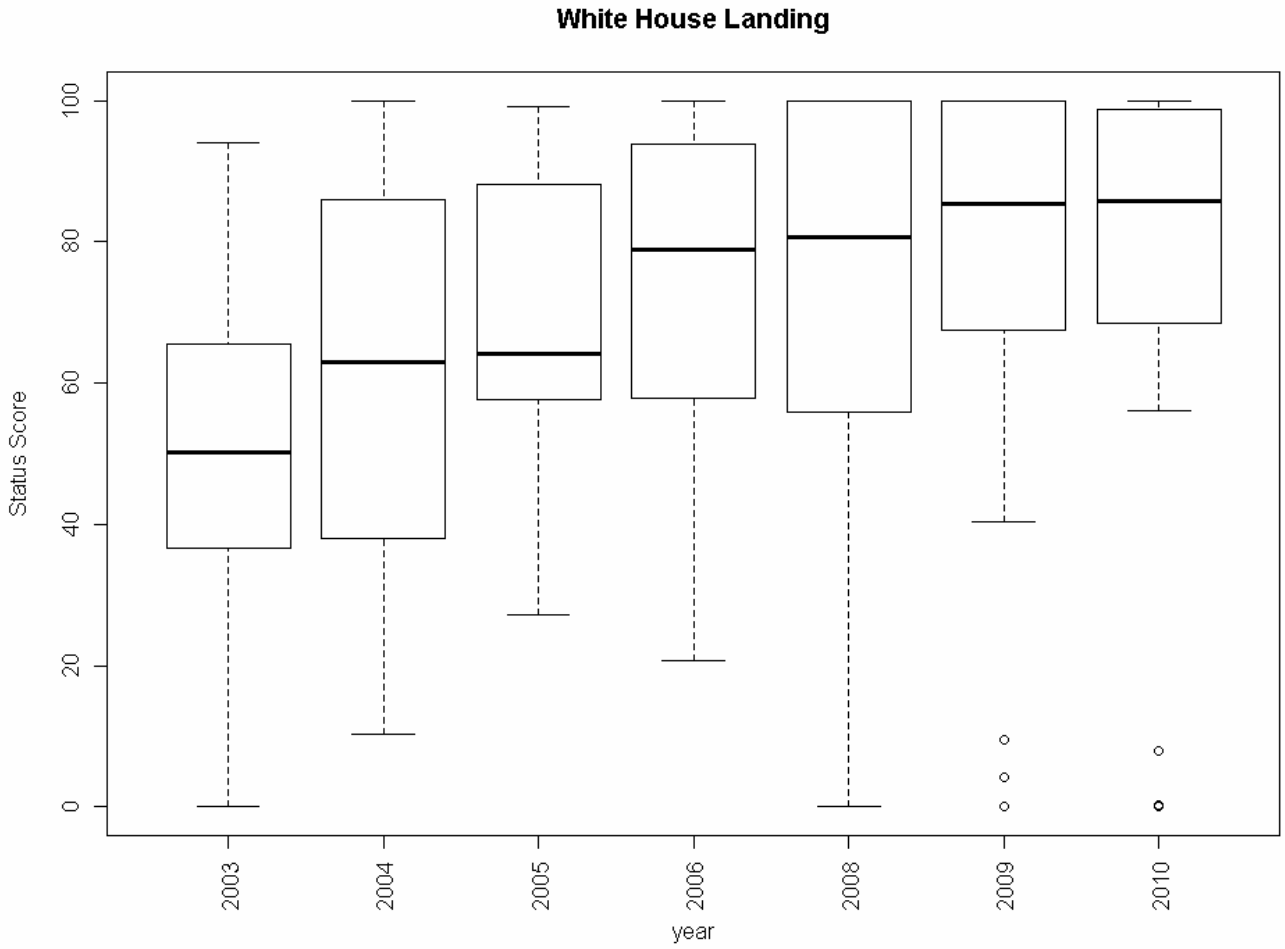
Shifting the Muddy Point Weekly Mean Distribution to Create a Scoring Function.



Comparison among stations of absolute status scores for ConMon weekly means in the York River Basin.



River	Location Name	Data Code
Pamumkey	White House Landing	PMK034.00
Pamumkey	Sweet Hall Marsh	PMK012.18
Mattaponi	Walkerton Landing	MPN028.86
Mattaponi	Muddy Point	MPN005.04
York	Taskinas Creek	TSK000.23
York	Claybank	YRK015.09
York	Yorktown	YRK006.77
York	Gloucester Point	YRK005.40
York	Goodwin Islands	CHE019.38


White House Landing
Status Score
0
20
40
60
80
100
2003
2004
2005
2006
2008
2009
2010
year