Stat 505 - Environmental Statistics - References for part 1
Analysis of data with below detection limit values

General Books / Book Chapters

Gibbons, R.D. 1994. Statistical Methods for Groundwater Monitoring. Wiley.
Focuses on analysis of groundwater well data (repeated samples from fixed points) and regulatory statistics. Chapter 5 is a very nice summary of different ways to compute the d.l. Chapter 10 describes analysis of below dl values, emphasizing P[est. mean > action level].

Gilbert, R.O. 1987. Statistical Methods for Environmental Pollution Monitoring. Wiley
My favorite introductory book, but now a bit out of date. Clearly written. Focuses on useful techniques and avoids being an encyclopedia. Chapter 14 describes < dl values.

Helsel, D.R. 2005. Nondetects and Data Analysis. Wiley.
Applied, emphasizing how to carry out various methods, and comparison between methods. More details on computations than Helsel and Hirsch

Helsel, D.R. and Hirsch, R.M. 1992. Statistical Methods in Water Resources. Elsevier.
Now out of print. Applied, written for hydrologists. Chapter 13 covers below detection limit obs. Good basic intro. On line at: http://pubs.usgs.gov/twri/twri4a3/

Millard, S.M. and Neerchal, N. 2001. Environmental Statistics with S-plus. CRC/Chapman and Hall
Diverse methods in environmental statistics, same technical level as Helsel. Chapter 10 deals with < dl values.

Definition of detection limits

Currie, L.A. 1968. Limits for qualitative detection and quantitative determination. *Analytical Chemistry* 48:586-593.
Source of the 3σ and 10σ limits.

Currie, L.A. 2004. Detection and quantification limits: basic concepts, international harmonization, and outstanding ("low-level") issues. *Applied Radiation and Isotopes* 61:145-149.

Hubaux, A. and Vos, G. 1970. Decision and detection limits for linear calibration curves. *Analytical Chemistry* 42:849-855.
Alt. method of computing detection limit, allowing error variance to vary with the mean concentration.

Alvarez, J.L. 2007. Poisson-based detection limit and signal confidence intervals for few total counts. *Health Physics* 93:120-126
Radionuclide analyses are based on counts of events and Poisson distributions. Here is an entry to the literature on detection limits for Poisson events.

Overview papers

Akritas, M. G., Russell, T.F., and Patil, G.P. 1994. Statistical analysis of censored environmental data. pp 221-242 In Patil, G.P. and Rao, C.R. (eds) Handbook of Statistics 12, *Environmental Statistics*. North-Holland, New York.
Emphasizes statistical theory underlying various estimation and test methods.

Helsel, D.R. 1990. Less than obvious - statistical treatment of data below the detection limit. *Environmental Science & Technology* 24: 1766-1774.
An attempt to convince environmental scientists to use methods other than substitution.

Helsel, D.R. 2005. More Than Obvious: Better methods for interpreting nondetect data. Environmental Science and Technol. 39:419A-423A.
Argues for Kaplan-Meier based approaches. I don't agree with all Dennis's conclusions.

Helsel, D.R. 2006. Fabricating data: how substituting values for nondetects can ruin results, and what can be done about it. *Chemosphere* 65:2434-2439.
A recent attempt to convince environmental scientists to use methods other than substitution.

Porter, P.S., Ward, R.C. and Bell, H.F. 1988. The detection limit. *Environmental Science and Technology* 22:856-861.

Maximum likelihood estimation

Cohen, A.C. 1959. Simplified estimators for the normal distribution when samples are singly censored or truncated. *Technometrics* 1:217-237
Original derivation of the m.l.e. Includes a "simple" adjustment method for hand calculation using tables (Cohen 1961).

Cohen, A.C. 1961. Tables for maximum likelihood estimates: singly truncated and singly censored samples. *Technometrics* 3:535-541.

Cohen, A.C. 1991. *Truncated and Censored Samples, Theory and Applications*. Marcel-Dekker, New York
Book length treatment, emphasizing normal populations

Lyles, R., Williams, J., and Chuachoowong, R. 2001. Correlating two viral load assays with known detection limits. *Biometrics* 57:1238–1244.
mle of correlation coefficient when both X and Y include <dl values.

Newman, E. and Rudel, R. 2007. Estimating correlation with multiply censored data arising from the adjustment of singly censored data. *Environmental Science and Technology* 41:221-228.
Multiply censored obs often arise from a lab measurement with a single dl divided by sample weight that varies between observations. Here's an ingeneous way to incorporate that info. I haven't read the paper in detail, but I suspect this would be a very promising MS or perhaps PhD topic.

Tobin, J. 1958. Estimation of relationships for limited dependent variables. *Econometrica* 26:24-26.
Describes tobit regression, i.e. mle

Robust order statistics estimator

Gilliom, R.J. and Helsel, D.R. 1986. Estimation of distributional parameters for censored trace level water quality data. I. Estimation techniques. *Water Resources Research* 22:135-146.
Original paper describing the ROS (robust order statistics) method

Helsel, D.R. and Gilliom, R.J. 1986. Estimation of distributional parameters for censored trace level water quality data. II. Verification and applications. *Water Resources Research* 22:147-155.
Simulation evaluation of ROS and various other estimators

Hirsch, R.M. and Stedinger, J.R. 1987. Plotting positions for historical floods and their precision. *Water Resources Research* 23:715-727.
Method for determining plotting positions (quantiles) when multiple detection limits. Motivating example is flood height, but method applies to detection limits.

Log normal distributions

Finney, D.J. 1941. On the distribution of a variate whose logarithm is normally distributed. *Journal of the Royal Statistical Society* (suppl.) 7:155-161.
Derived the UMVUE of the mean of a lognormal distribution.

Land, C.E. 1971. Confidence interals for linear functions of the normal mean and variance. *Annals of Mathematical Statistics* 42:1187-1205.
Derivation of the distribution of sum of normal + k*Chi-square random variables. Useful for confidence intervals for the log normal mean.

Land, C.E. 1972. Tables of confidence limits for linear functions of the normal mean and variance. *Selected Tables in Mathematical Statistics* 3:385-419.
Tables of coefficients used to calculate quantiles and confidence intervals.

Reimann, C. and Filzmoser, P. 2000. Normal and lognormal data distribution in geochemistry: death of a myth. Consequences for the statistical treatment of geochemical and environmental data. *Environmental Geology* 39:1001-1014.
Uses data sets from large regional surveys to argue that environmental chemical data are neither normally or lognormally distributed. "Distributions are usually skewed, have outliers, and originate from more than one process".

Note: I suspect the last phrase above is the key: how much of their conclusion arises because a regional survey includes many different environments, each with its own distribution of the chemical.

Shumway, R.H., Azari, A.S., and Johnson, P. 1989. Estimating mean concentrations under transformation for environmental data with detection limits. *Technometrics* 31:347-356.

Stoline, M.R. 1991. An examination of the Lognormal and Box and Cox family of transformations in fitting environmental data. *Environmetrics* 2:85-106
Both Shumway and Stoline argue that log() is too restrictive and Box-Cox family should be used instead.

Tests

Latta, R.B. 1981. A Monte-Carlo study of some two-sample rank tests with censored data. *Journal of the American Statistical Association* 76:713-719.
Compares empirical type I error rate and power for various score tests, especially when censoring patterns are not the same in the two groups.

Millard, S.P. and Deverel, S.J. 1988. Nonparametric statistical methods for comparing two sites based on data with multiple nondetect limits. *Water Resources Research* 24:2087-2098.
Original paper applying score tests to two sample environmental problems. Actual method proposed by M and D uses normal scores, a slightly different score function than the one I discussed in class.

Chu, H.T., Nie, L., and Cole, S.R. 2006. Sample size and statistical power assessing the effect of interventions in the context of mixture distributions with detection limits. *Statistics in Medicine* 25:2647-2657.
A biomedical application where distribution is assumed a mixture of a log normal and extra zeros.

Evaluation of methods

There are many, many of these. All use numerical simulation to evaluate performance (bias, variance, mse, type I error rate, power, or perhaps something else) for a collection of estimators and/or tests. Helsel, p 74-77, summarizes many of these.

Spatial analysis of below-detection limit data

DeOliveira, V. 2005. Bayesian inference and prediction of Gaussian random fields based on censored data. *Journal of Computational and Graphical Statistics* 14:95-115.

Fridley, B.L. and Dixon, P. 2007. Data augmentation for a Bayesian spatial model involving censored observations *Environmetrics* 18:107-123.