59-

STATISTICAL METHODS FOR DETECTION AND QUANTIFICATION OF ENVIRONMENTAL CONTAMINATION

Robert D. Gibbons
University of Illinois at Chicago

David E. Coleman
Alcoa Technical Center

JOHN WILEY & SONS, INC.
New York / Chichester / Weinheim / Brisbane / Singapore / Toronto

7174

To Carol, m

To Sally, my

This book is printed on acid-free paper. ♾

Copyright © 2001 by John Wiley & Sons. All rights reserved.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4744. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, (212) 850-6011, fax (212) 850-6008, E-Mail: PERMREQ @ WILEY.COM.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional person should be sought.

Library of Congress Cataloging-in-Publication Data:

Gibbons, Robert D., 1955–

Statistical methods for detection and quantification of environmental contamination / Robert D. Gibbons, David E. Coleman.

p. cm.

Includes index.

ISBN 0-471-25532-7 (cloth: alk. paper)

1. Pollution—Measurement—Statistical methods. 2. Environmental monitoring—Statistical methods. I. Coleman, David E. II. Title.

TD193.G547 2001

628.5'0287—dc21 2001017855

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

7176

the measurements are put is mon-
ecision is good relative to the size
requency. A constant bias of 5 g
a shift by 100 g, or a trend where
ngle outlier measurement that is
ipling in the process standard de-

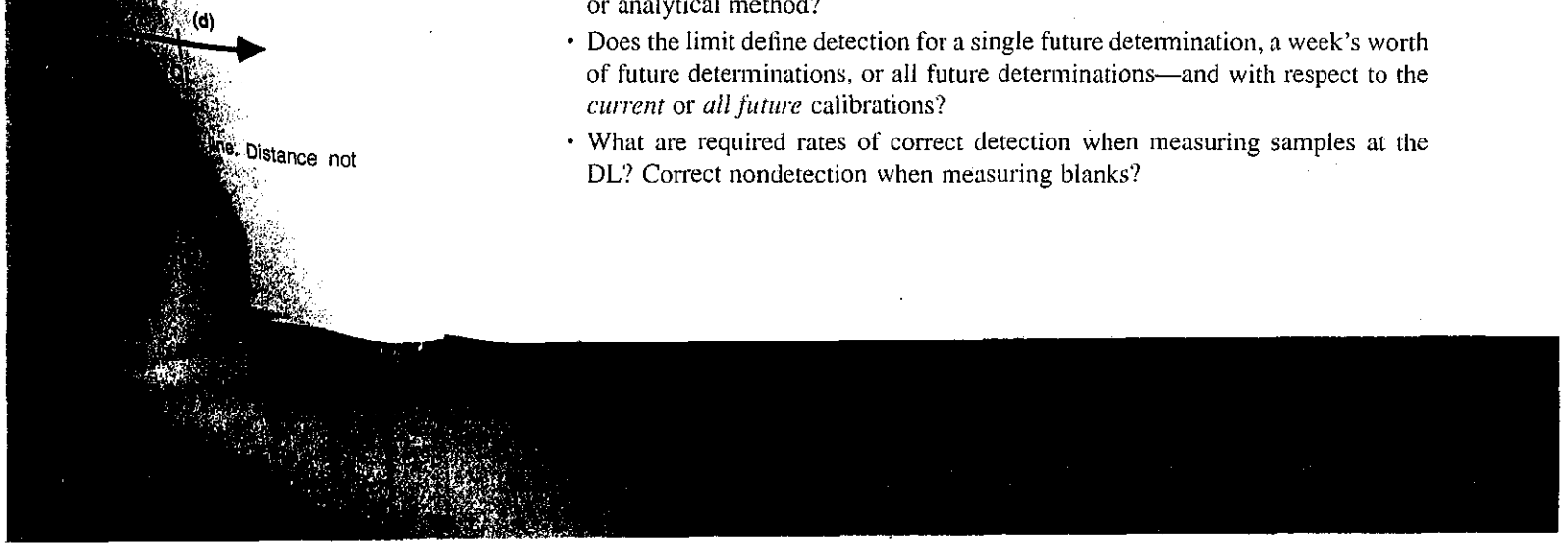
lue to an error in estimating the
ected simply by recalibrating, of-
, precision that changes with true
eighted least squares for calibra-
icated than ordinary least squares.

QUANTITATION)

ions and interpretations of *detec-*
neasurement (see Coleman et al.,
. *Critical limits* (also called *criti-*
ure but have often been confused
ree types of limits are simple. In
d by Currie (1968, 1995). Refer-
nits, from lowest to highest, then
f the *intervals* between the limits.
lures used to compute these limits.

e Limits

l level (L_C). L_C is the lowest meas-
tly assert that the analyte has been
likely to have been obtained from
confidence level (e.g., 99%) to de-
ence any measurement above L_C
is present (at least one mol-
the confidence level chosen.
ent (when a measurement
threshold.



Unfortunately, L_C has sometimes been confused with a detection limit, as in the EPA's '*method detection limit*' (MDL, as described in 40 CFR, Part 136, App. B). Without developing a comprehensive critique of the MDL in this chapter, let it be observed that the EPA MDL is defined basically as

$$\text{MDL} = t_{[\text{df}=6,1-\alpha=0.99]}s = 3.14s$$

where t is a Student's t critical value and s is a sample standard deviation from seven replicate samples of spiked reagent-grade water. Thus the MDL is a critical level that has been set up to be exceeded approximately 1% of the time when a blank sample is measured (actually, based on its construction, it is technically more correct to say that a measurement will be negative approximately 1% of the time when measuring a sample with concentration = MDL).

Note that there is little or no assurance that samples with real concentrations at or below L_C will be detected. For that, we need to consider a higher concentration: the detection limit.

DL (Also Called L_D) The middle limit among these three is the *detection limit* (DL). The DL is the lowest concentration at or above which an analyte can confidently be detected (i.e., distinguished from zero). Thus the detection limit defines the lowest concentration at which the measurement signal consistently emerges from the noise. That is, when the true concentration is at or above the DL, the reported measurement will, with high confidence, exceed L_C. At the DL, a measurement provides the most elementary form of information, a binary value indicating (at some level of confidence) whether the analyte is detected or *not* detected. See Chapter 9 for a discussion of how this basic information content is consistent with an alternative definition of the DL as "the minimum concentration at which one can be sure that a measurement will have at least zero significant digits."

There is tremendous diversity in proposed DLs. Some of this diversity is due to semantic confusion between L_C and DL. However, the diversity in proposed DLs is due primarily to differences in DL applications, differences in required confidence, and different statistical approaches to calculating DLs. Some of the key issues are:

- For which sources of bias and variation should the DL account (i.e., include in its estimate of measurement variation, e.g., σ)? Does the limit account for laboratory-to-laboratory biases, preparation variation, calibration error (lack of fit, coefficient error, error in the calibration standard concentrations), matrix effects, analyte impurity, differences in the identification or computation algorithm used by the measurement system, bias, and variation in the instrument or analytical method?
- Does the limit define detection for a single future determination, a week's worth of future determinations, or all future determinations—and with respect to the *current* or *all future* calibrations?
- What are required rates of correct detection when measuring samples at the DL? Correct nondetection when measuring blanks?

- Bias can be tolerated if the only use to which the measurements are put is monitoring a process for change, provided that precision is good relative to the size of change to be detected and the sampling frequency. A constant bias of 5 g may be immaterial if all one is interested in is a shift by 100 g, or a trend where the mass slowly increases by 100 g, or a single outlier measurement that is lower than the average by 100 g, or a quadrupling in the process standard deviation by a factor of 4, from 20 g to 80 g.
- Bias that changes with true concentration, due to an error in estimating the calibration function slope, can usually be corrected simply by recalibrating, often by using ordinary least squares. However, precision that changes with true concentration may force the routine use of weighted least squares for calibration, a technique that is somewhat more complicated than ordinary least squares.

2.5 DETECTION AND QUANTIFICATION (QUANTITATION)

Over the years, there have been numerous definitions and interpretations of *detection limits* and *quantitation limits* for trace-level measurement (see Coleman et al., 1997; Currie, 1968; or Oresic and Grdinic, 1990). *Critical limits* (also called *critical levels*) have been little discussed in the literature but have often been confused with detection limits. The basic concepts of all three types of limits are simple. In this section we once again use the foundation laid by Currie (1968, 1995). Referring to Figure 2.2, we first define each of the limits, from lowest to highest, then we discuss the technically sound interpretation of the *intervals* between the limits. In later chapters we deal with the statistical procedures used to compute these limits.

2.5.1 Conceptual Definitions of the Three Limits

L_C The lowest of these three limits is the *critical level* (L_C). L_C is the lowest measured concentration above which one can confidently assert that the analyte has been detected. It is the lowest measurement that is unlikely to have been obtained from a blank sample. We reserve the right to choose a confidence level (e.g., 99%) to define what *is* and what *is not* considered unlikely. Hence any measurement above L_C should be considered strong evidence that the analyte is present (at least one molecule), where "convincing" is only to the degree of the confidence level chosen. Because of this implicit decision that the analyte is present (when a measurement exceeds L_C), L_C is sometimes referred to as the *detection threshold*.

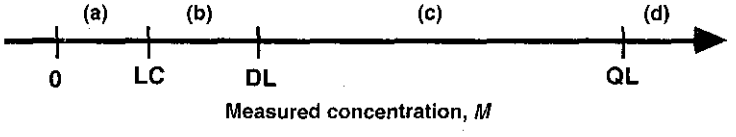


Figure 2.2 Metrological partitioning of the low end of the real number line. Distance not to scale.

- Is standard deviation of measurement error assumed known? Is it assumed constant within the range of concentrations of interest?

QL (Also Called L_Q) *Quantitation limits* (QLs) are also defined in diverse ways. In concept, a QL is the lowest concentration at which there is some confidence that the reported measurement is relatively close to the true value. This is what is truly meant by "to quantitate" or "to quantify." Some of the diverse definitions are given below.

The quantitation limit (QL) is the lowest concentration at or above which:

- *One can quantitate.* Amazingly, this definition has been proposed by some chemists in all seriousness, but the definition merely shifts the burden to defining *quantitate*.
- *One can have "assurance" of detection.* However, this is provided by the DL. It is appropriate for detection but is inconsistent with the common usage of the related words: *quantify*, *quantity*, *quantitative*, and is inconsistent with the historical use of the term, introduced by Currie (1968).
- *Measurements have a low, prescribed standard deviation* (e.g., 5 ppb). This is a reasonable definition for ensuring a certain number of significant digits of measurement for a known range of measurement values, say 100 to 999 ppb.
- *Measurements have limited relative standard deviation* (e.g., RSD < 10%). This also is a reasonable definition and is used by the American Chemical Society to define the *limit of quantitation* (LOQ): LOQ = 10σ, the solution to requiring that the 99%+ confidence interval about a measurement be within span ±30%: $\pm 3\sigma = \pm 30\%$. Additionally, it has been required to have RSD < 10%, in Gibbons (1994). Two weaknesses of the "10σ" approach are: (1) there is no indication of degrees of freedom in the estimate of the RSD, so it is not possible to determine the multiplier required to base the QL on a valid statistical interval; and (2) the percentage is arbitrary and typically has no relationship to significant digits.
- *Measurements have limited relative measurement uncertainty* (RMU) (e.g., RMU 5%) or some other prescribed proportion at some level of confidence (e.g., 95%). The RMU is the RSD times a multiple (>1) which is a function of concentration, and depends on the confidence levels and the calibration design. See Chapter 9 for a complete development of the QL as a guarantor of limited RMU, which can provide assurance of at least one significant digit in a measurement.

2.5.2 Interpretation of the Intervals Defined by the Critical Level, Detection Limit, and Quantitation Limit

As discussed in Section 2.5.1, conceptually, the critical level, detection level, and quantitation level partition the real number line into four intervals into which a trace-level measurement, M, can fall. Figure 2.2 provides a graphical representation of

7182

this partitioning of the number line. The intervals of the number line (a) to (d) as divided by the L_C, DL, and QL are listed below, with interpretation:

(a) The measurement M is less than the critical level ($M < L_C$). By definition of L_C, M is indistinguishable from zero concentration and hence should be considered a nondetect. M might even be less than zero. If $M < 0$, M can still be useful—if it is one of several measurements. It should not be discarded, labeled "nondetect," nor set to a prescribed nonnegative value (such as 0, DL/2, or DL).

(b) M is at or above the critical level but below the detection limit ($L_C \leq M < \text{DL}$). By definition of L_C, M is treated as a detection. However, by definition of DL, it should be realized that there is low confidence of detection in this interval. Note that the label "low confidence" is not a value judgment; it is a factual statement given that a level of confidence has been selected.

(c) M is at or above the detection limit but below the quantitation limit ($\text{DL} \leq M < \text{QL}$). By definition of L_C, M is treated as a detection, and by definition of DL, any true concentration in this interval is *likely* to be detected. By definition of QL, M is a very noisy measurement value which should only be reported with an error interval, and used only with extreme caution in comparison or computation.

(d) M is at or above the quantitation limit ($\text{QL} \leq M$). By definition of QL, M can be reported as a measurement (preferably with an error interval) and can generally be used for comparison and computation.

2.6 BETWEEN-LABORATORY ENVIRONMENT

Once again, we draw on the Currie (1995) insightful treatment of this topic, but we start at a more basic level and examine sources of bias and variation from several perspectives.

2.6.1 Bias and Variation Within a Single Laboratory

Chemical measurement is complex, even when one considers measuring a single analyte in a single type of matrix by a single method. Under these constraints, any given measurement still has several sources of bias and variation that may be nontrivial: intrinsic instrument noise, some "typical" amount of carryover error, plus differences in analysts, sample preparation, instrumentation, and even data-processing algorithms (thresholds, signal filters, etc.). Here we ignore sampling variation, which results in different true values for different samples.

A thorough approach to characterizing measurements within a laboratory would involve developing a plausible statistical model of measurement components (probably a complicated *mixed-effects model*, a model containing factors with random effects and factors with fixed effects), design a study to collect the necessary data,

7184

13

CENSORED DATA

13.1 INTRODUCTION

One of the most difficult problems in analysis of environmental monitoring data involves the incorporation of nondetects into estimates of summary statistics (e.g., mean and standard deviation) and corresponding tests of hypotheses and interval estimates. More often than not, environmental monitoring data consist of a mixture of results that can and cannot be quantified accurately. In practice, the censoring mechanism is the detection limit; values below are reported as ND or $<L_D$ to signify that they were not found in the sample. All other values are reported as a concentration. Based on earlier chapters, one should immediately note that this is the wrong procedure. The L_Q and not the L_D should be the censoring mechanism since values above the L_D and below the L_Q are detected but not quantifiable. Using the L_D as the censoring point produces data with widely varying levels of uncertainty violating the assumption of homoscedasticity (i.e., constant measurement variation) which is assumed by all of the previous statistical theory and methods. Even with an agreed-upon censoring point, there is considerable controversy regarding the appropriate method or methods for incorporating the censored data in computing summary statistics, testing hypotheses, and computing interval estimates. This is not at all surprising since the correct choice of method depends on both the degree of censoring (e.g., 20% versus 80% nondetects) and the type of application (e.g., computing the mean versus computing a prediction limit from data that are a mixture of quantifiable and nonquantifiable measurements), as well as ease of use. Additionally, the controversy can be fueled by an inclination toward a particular favorable outcome.

7186

13.2 CONCEPTUAL FOUNDATION

Assume that there is a population of true concentrations from which we have drawn a sample of size n. For convenience, also assume that variation in the sampled population can be represented by a continuous probability distribution for which a fraction of the true concentrations are essentially zero. This partial loss of information occurs because of censoring imposed by limits of detection and/or quantification.

For example, Davis (1994) points out that we may assume an underlying distribution as in Figure 13.1, but what we observe is the distribution in Figure 13.2, where the vertical line represents a point mass at $L_D/2$ containing the probability content of the region $<L_D$. In practice, the measurements are often coarsely rounded so that the observed frequency distribution looks like Figure 13.3. Davis (1993) points out that in real-world application the true underlying model in Figure 13.3 is unknown; therefore, different approaches will yield widely different results, depending on the degree to which they rely on the assumed distribution. This is even more critical in environmental monitoring applications (e.g., groundwater monitoring) in which repeated application of tail probabilities are used to control the overall site-wide false positive rate (i.e., prediction limits). How well a censored data estimator works in the center of the distribution (e.g., to estimate mean concentration) is often a poor index of how well that method will work in the tails of the distribution (e.g., to estimate a 99% confidence prediction limit for a new single measurement). In the following sections, several methods are described, and some general recommendations are provided.

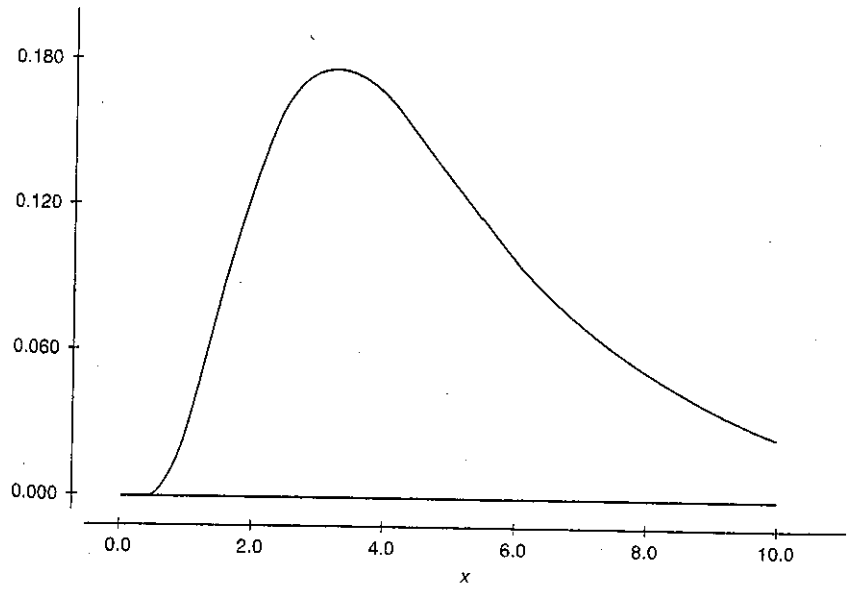


Figure 13.1 Assumed underlying distribution before censoring. Lognormal (1.5, 0.6), DL = 5.