

王斌, 朱雷, 宫照恒, 等. 2009 含未检出值的环境监测数据集的处理方法 [J]. 环境科学学报, 29(7): 1345-1350

Wang B, Zhu L, Gong ZH, et al. 2009 Introduction to the methods of parameter estimation for environmental monitoring data set with truncated data below a detection limit [J]. Acta Scientiae Circumstantiae, 29(7): 1345-1350

# 含未检出值的环境监测数据集的处理方法

王斌<sup>1</sup>, 朱雷<sup>2</sup>, 宫照恒<sup>1</sup>, 王戎<sup>1</sup>, 陶澍<sup>1,\*</sup>

1 地表过程分析与模拟教育部重点实验室, 北京大学城市与环境学院, 北京 100871

2 北京大学环境科学与工程学院, 北京 100871

收稿日期: 2009-02-11 录用日期: 2009-04-20

**摘要:** 环境样品分析过程中经常遇到包含未检出值的数据集, 这在对含有微量金属和微量有机污染物样本的分析研究中更为常见. 目前处理未检出数据的方法通常是忽略或者用  $Q$  检出下限的  $1/2$  或  $1/3$  进行简单替代. 然而, 当未检出数据比例较高时, 忽略或简单替代可能给数据表征带来较大偏差, 也不利于更深入的统计分析. 为改善对包含未检出数据的分析效果, 目前已经建立了多种统计方法. 本文系统介绍了这些方法, 包括总体均值和方差估计的参数和非参数方法以及关于均值的 95% 置信区间上界的估计方法. 此外, 本文利用虚拟数据和实测数据给出了主要方法的应用实例.

**关键词:** 未检出数据; 低于检出下限; 数据处理

文章编号: 0253-2468(2009)07-1345-06 中图分类号: X830.3 文献标识码: A

## Introduction to the methods of parameter estimation for environmental monitoring data set with truncated data below a detection limit

WANG Bin<sup>1</sup>, ZHU Lei<sup>2</sup>, GONG Zhaoheng<sup>1</sup>, WANG Rong<sup>1</sup>, TAO Shu<sup>1,\*</sup>

1 Laboratory for Earth Surface Processes College of Urban and Environmental Sciences Peking University Beijing 100871

2 College of Environmental Sciences and Engineering Peking University, Beijing 100871

Received 11 February 2009; accepted 20 April 2009

**Abstract** Data sets containing nondetects that are below detection limits are inevitable in analytical results for environmental samples especially for trace elements and trace organic pollutants. Commonly, the nondetects were either ignored or simply substituted by zero, half or one third of the detection limits. However, these approaches would result in biased representative statistics and difficulty in further statistical analysis especially when there was a large percentage of nondetects. In fact, a number of methods have been developed to handle the data and to provide the best estimation of the population parameters based on the data above the detection limits. These methods including both parametric and distribution-free are introduced in this article. The methods for calculating the upper limits of the 95% confidence intervals of population means are also presented. The method application is illustrated based on both virtual and field data sets.

**Keywords** nondetects; below detection limit; data processing

### 1 前言 (Introduction)

任何对环境样品的分析检测方法都有其极限, 即特定的检出限 (DL). 因此, 环境样品分析过程中不可避免地会遇到未检出 (或不能可靠检出) 问题 (Cohen, 1950; Gleit, 1985; Helsel, 1990). 在测定样品浓度高于方法检出上限的情况下, 减少样品量或对样品进行稀释仍可获得可靠的结果. 当测定结果

低于方法检出下限时, 虽然有时可以通过增加样品量或对待测物进行富集获取结果, 但在很多情况下, 限于样品量、时间和经济成本或者富集过程本身引入的误差, 研究者不得不接受不能获得实测数据的结果, 并将相关数据记为未检出. 这在环境样品的痕量和超痕量污染物分析中十分常见.

未检出数据对数据分析结果的影响在很大程度上取决于未检出数据的多寡. 在检出率很高的情

基金项目: 国家自然科学基金 (No. 40730737, 40771179); 国家重点基础研究发展计划 (No. 2007CB407301)

Supported by the National Natural Science Foundation of China (No. 40730737, 40771179) and the National Basic Research Program of China (No. 2007CB407301)

作者简介: 王斌 (1985-), 男, 博士研究生; \* 通讯作者 (责任作者), E-mail: tao@urban.pku.edu.cn

Biography: WANG Bin (1985-), male, Ph.D. candidate; \* Corresponding author, E-mail: tao@urban.pku.edu.cn

况下,无论是忽略未检出数据的存在,还是对未检出数据进行简单替代,都不会对数据分析结果产生显著影响.然而,在未检出率比较高时,其影响就不能忽略.特别是对毒性很强的微量污染物而言,对结果估计的偏差会严重影响风险分析的可靠性.

处理未检出数据的常用方法包括忽略和替代.一般的简单替代方式包括用  $Q$  检出下限的  $1/2$   $1/3$  或者检出下限替代未检出数据 (Helsej 1990; USEPA, 2006).在样本量较小,且未检出比例较高时,忽略未检出结果实质上相当于忽略最低浓度段数据的贡献,从而导致统计结果高估.如果用 0 替代,则必然导致对结果的低估,其程度取决于未检出数据比例、数据分布形式以及总体期望与未检出下限的差别.显然,用检出下限的  $1/2$  或  $1/3$  作为替代做法的风险似乎低于忽略或用 0 替代,但由于这样的取值十分主观,且无从判断这一取值和未检出数据的真实值的相对差别,有些情况下不一定优于 0 替代或者忽略.譬如,在未检出真实值非常接近 0 时,按检出下限的特定分数取值产生的误差可能会更大.

由于含量很低,且不可能为负,环境中的微量污染物等参数通常呈对数正态分布 (陶澍, 1994).这样的例子既有主要环境介质中的微量金属 (李淑媛等, 1995) 和微量有机污染物 (刘凌和陆桂华, 2003; Wang 2009), 也包括诸如风速 (周军莉, 2001)、气溶胶粒径 (王海霞, 2007) 等环境变量.在对右偏 (如对数正态分布) 数据进行分析时,由于检出下限比正态分布接近期望 (如几何均值),较大比例的未检出数据更容易导致数据分析偏差.如果简单忽略未检出数据的存在,很可能显著高估总体大小,若用 0 替代,则可能造成对总体大小的显著低估.即便采用检出下限的  $1/2$  或  $1/3$  进行替代,仍然很难避免分析结果的显著偏差.除了可能造成的期望估计偏差外,任何采用相同的取值替代方法的另一个问题是会造成整体数据分布畸形,不利于进一步的统计检验.对参数方法而言,对正态分布的偏离可能引起检验结果失真;对非参数方法而言,这样的畸形分布会引起大量同分,从而模糊显著性检验的判断 (陶澍, 1994).因此,忽略和简单替代的做法都不是好的选择 (Singh *et al.*, 2006; USEPA, 2006).

从理论上说,作为随机变量,所有未检出数据相当于从总体特定分布的检出下限区间中随机抽

样.而关于总体分布的信息应当部分反映在检出下限以上的实测数据中.因此,在考虑部分未检出的前提下,利用实测结果信息对总体分布进行估计,不仅可以在一定程度上对总体分布进行判断,而且可以在获得总体分布有效描述的前提下,逐一对未检出数据进行随机赋值.近年来,许多研究者在这方面进行了大量探讨,并建立了多种行之有效的办法,包括参数方法和非参数方法 (Helsej 2005a; 2005b; Singh *et al.*, 2006).参数方法可以根据高于检出上限的数据估计总体分布参数 (比如正态分布的均值和方差),继而根据分布对未检出数据赋值;而非参数的方法可不依赖于数据分布进行总体参数估算.

尽管近年来关于未检出数据处理方法的研究取得了较大进展,目前已建立了很多成熟方法,且多数方法有相应软件支持,但这些方法在国内甚至国外环境研究领域并未得到普遍应用.本文的目的在于通过对具体方法的介绍,推动这些方法在我国环境研究中的应用,提高环境监测和研究的数据处理水平.

## 2 方法介绍 (Available methods)

以下分别介绍对包含未检出数据的数据集进行分析的主要方法.

### 2.1 含未检出数据的总体均值和方差的估计方法

估算总体均值和方差的方法按照是否依赖于数据的分布形式可分为参数和非参数 2 类.参数方法中常用的有最大似然函数法、期望最大化法、顺序回归统计法和 EPA Delta 对数正态方法等,非参数方法常见的有 Winsorization 法和 Kaplan-Meier 法 (Singh *et al.*, 2006).

根据最大似然原理建立的最大似然函数法是估算包括未检出数据的总体均值和方差的经典方法.此类方法包括 Cohen 最大似然函数法、无偏最大似然函数法和约束最大似然函数法. Cohen (1950; 1959) 在因计算机未得到普遍应用,最大似然函数求解过于复杂的情况下,发展了经典的 Cohen 最大似然函数法并建立了与之配套的最大似然函数求解常数表.随着计算能力的迅速提高,该方法逐渐被淘汰.无偏最大似然函数法修正了 Cohen 最大似然函数法的有偏估计 (Saw, 1961; Schneider 1986).约束似然函数法则通过对远离中心的数据点赋予较低权重的方法简化了最大似然函数计算

(Perrson and Rootzen 1977), Singh 和 Nocerino (2002)则进一步给出该方法的无偏估计。

另一类方法建立在期望最大化算法的基础上 (Dempster *et al.*, 1977). Gleit (1985)发现, 用迭代的期望最大化算法处理包含低于检出下限数据的最大化似然函数可以得到比其它方法更优的结果。迭代的初始赋值利用检出数据的样本均值和方差, 迭代次数则通过最大化似然函数确定。在每步的迭代中, 未检出数据用条件期望值替代。遗憾的是, 这样的迭代算法容易受异常值干扰, 有时甚至可能得到负值或大于检出下限的替代值, 由此导致较高的估计偏差。相比之下, 稳健的期望最大化算法则可以克服一般期望最大化算法的上述缺陷 (Singh *et al.*, 2006)。

最大似然函数法和期望最大化算法都是在假定总体服从正态分布的前提下求算均值和方差, 据此构建总体分布。对于非正态分布的情况需要对数据进行正态变换。Singh等 (2006)认为, 将非正态分布数据变换为对称分布 (近似正态分布)的结果可能导致难以对逆变换后得到的结果进行解释, 带来无法估计的误差。

作为一类方法, 顺序回归统计法假设数据服从某种特定分布 (如正态分布), 然后根据检出数据的大小以及按其所占总体概率的比例建立的函数关系外推出未检出数据。此类方法本质上并没有差别 (Gleit 1985), 可根据分布形式细分为正态分布、对数正态分布和  $\gamma$ 分布的顺序回归统计方法, 也可以根据原始数据是否经过正态变换分为基于原始数据和基于变换数据的统计方法。目前常用的基于相同原理发展出来的具体方法包括原始数据服从正态分布的回归法 (Newman *et al.*, 1990)、结合最大似然估计的稳健顺序回归统计-最大似然函数法 (Kroll and Stedinger 1996)、基于对数变换的顺序回归统计 (Uncerson 5 1, 2003)和基于对数变换的稳健的顺序回归法 (Helsel 2005a)。需要指出的是, 对原始数据进行对数变换, 再对计算结果进行逆变换可能带来未知误差。特别是在未检出数据较多或分布呈中等或较大程度右偏时, 对数变换的方法可能导致结果误差增大。

对于对数正态分布数据, USEPA (1991)提出了专门的 EPA Delta对数正态分布法。这种方法假定检出数据服从对数正态分布, 而占总体个数比例为 Delta的未检出数据则服从离散均匀分布, 继而根据

这种混合分布建立迭代方程进行求解。尽管该方法也可用于估算置信上限, 但由于假定检出数据服从对数正态分布, 计算置信上限没有太大的实际价值 (Singh *et al.*, 2006)。

用于包含未检出数据的数据集均值和方差估计的非参数方法主要有 Winsorization法和生存曲线法。Winsorization法 (Gilbert 1987)对对称分布数据的估算更为可靠, 但不能用于未检出率高于 50%的情况 (Dixon and Tukey 1968)。生存曲线法中常用的是 Kaplan-Meier法 (Kaplan and Meier 1958)。Helsel认为, 这是估算含未检出数据样本, 特别是未知分布样本的总体均值和方差的最合理的方法 (Helsel 2005a)。

## 2.2 总体均值置信区间上界的估计方法

估计均值的可靠性可以通过计算置信区间上界 (UCL)进行评估, 多数情况下取 95% 置信区间上界, 记为 UCL95。Helsel提出, 以学生  $t$ 分布为包含未检出数据的可能分布 (Helsel 2005a)。基于这种分布可以使用最大似然函数法或期望最大化算法估计样本的均值和方差以及相应的 UCL95 (Singh *et al.*, 2006)。对于正态分布数据, 基于最大似然函数法计算的 UCL95可以允许低于检出限数据占样本数据的 10% ~ 70%, 但对于不对称分布的数据, 则不能得到较为可靠的估计, 此时最好选择非参数方法。

使用非参数方法往往可以得到更可靠的估计结果。譬如, 对任何均值和方差估计方法均可以采用 Chebyshev不等式估计 UCL95 对中等偏斜的数据, 这种方法较为保守, 所以应当尽量结合 Kaplan-Meier法使用。同时也可以采用更保守的置信区间 (97.5% 或 99%), 在偏斜程度较大情况下更是如此。利用 Jackknife法可以结合顺序回归统计或期望最大化算法获得学生  $t$ 分布的 UCL95, 但该方法不适用于偏度较大的情况 (Singh *et al.*, 2006)。Bootstrap方法包括标准 Bootstrap法、Bootstrap  $t$ 法、百分位数 Bootstrap法和偏差校正的 Bootstrap方法 (Efron and Tibshirani 1993; Many 1997)。这些方法在分布、检测统计量和未检出数据个数未知的情况下特别有用。它们更适用于数据分布比较对称的情况。在用 Kaplan-Meier方法估算均值和方差时, Helsel建议采用正态分布或学生  $t$ 分布来截取数据的 UCL95 (Helsel 2005a)。如上所述, Kaplan-Meier方法也可以结合 Chebyshev不等式、Jackknife和

Bootstrap抽样方法等一起确定 UCL<sub>95</sub>

### 3 方法应用 (Application of the methods)

#### 3.1 未检出数据估计方法的应用

环境研究中涉及的数据可大致分为正态分布、近似正态分布、对数正态分布、gamma分布和未知分布等不同类型. 由于环境研究中涉及的样本量通常较小, 有时很难根据统计检验对总体分布特征进行可靠判断. 然而, 如果对未检出数据处理不限于获得总体算术均值和方差, 确定总体分布至关重要. 此外, 还需要确认所有数据确实来自同一总体. 此外, 若有可能, 应当避免不必要的的数据变换. 这是因为环境标准的制定通常基于原始尺度, 经变换和逆变换很容易导致难以确定的偏差. 因此, 如果数据服从正态分布、近似正态分布或其它对称分布, 应当直接选用参数方法, 以提高信息利用率. 值得一提的是, 许多常见环境变量 (如微量金属含量、微量有机污染物含量等) 多属于对数正态分布; 遗憾的是, 模拟结果证实, 这些估计方法对经过对数变换数据的估算结果一般有较大误差 (Singh *et al.*, 2006). 这种情况下如果仅仅需要获得对总体进行统计描述, 即仅限于获得总体算术均值和方差, 与其在数据变换的基础上用参数方法, 不如直接使用非参数方法, 如 Kaplan-Meier法. 但是, 此时数据的未检出率不宜太高. 例如, Helsel认为未检出率超过 40% ~ 50% 时结果会极不可靠 (Helsel, 2005a). 如果不仅需要得到算数均值和方差的估计值, 还需要计算几何均值和几何标准差 (陶澍, 1994), 并逐一估计未检出数据, 则不可避免地需要确定数据分布, 进行数据变换, 继而用相应的参数方法.

#### 3.2 相关软件

目前研究者已经开发出许多相关软件, 可以直

接利用这些软件估算含未检出数据的数据集的均值、方差和均值的置信区间上界. 大部分软件可以在网上免费下载. 例如, USEPA 开发的 ProUCL Version 4.00.02 可在网址: <http://www.epa.gov/esd/tsc/images/setup.zip> 下载; NADA for R 工具包的下载网址为: <http://www.practicalstats.com/nada/nada/downloads.html>; UnCensor Version 5.1 for Windows 软件的下载网址是: [http://web.vims.edu/env/research/software/vims\\_software.html?svr=www](http://web.vims.edu/env/research/software/vims_software.html?svr=www). 需要说明的是, ProUCL 几乎包含了 NADA for R 的全部功能, 且 NADA for R 是命令行式程序, 故不推荐使用 NADA for R.

### 4 应用实例 (Case study)

#### 4.1 模拟数据实例

在均值为 5.00 方差为 2.00 的虚拟正态分布中随机抽取 1000 个体 (忽略单位), 分别从中删去 5%、20% 和 40% 低值 (虚拟未检出) 构成 3 个包含不同比例未检出数据的模拟数据集, 对应的检出下限分别为 2.674、3.825 和 4.602. 对于正态分布数据, 为充分利用信息, 采用期望最大化迭代算法估算原始总体的均值和方差, 由此得到的总体均值和 95% 置信区间上界分别为:  $5.010 + 0.090$ ,  $5.020 + 0.087$  和  $5.040 + 0.082$ , 方差分别是 2.09, 1.97 和 1.77. 显然, 较高的未检出比例可能导致较大的误差. 但是, 即便在未检出率高达 40% 的情况下, 模型仍能较好地还原总体均值, 但方差显著偏低. 根据计算得到的分布参数和对应的检出下限, 可在正态分布这个假设下估算出未检出个体取值. 恢复数据和删除数据的关系 (按大小顺序排列) 如图 1 所示. 从图中可以看到, 迭代算法总体可以较准确地恢复原始数据, 仅极低值数据段的个别点有较明显偏差.

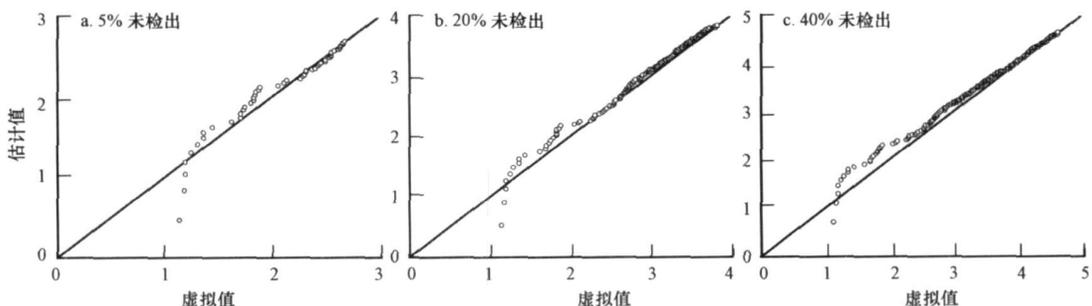


图 1 5%、20% 和 40% 未检出条件下采用期望最大化迭代算法获得的未检出数据估计值和虚拟未检出数据的关系 (虚拟正态分布)

Fig 1 Relationship between the virtual and the estimated values of the data below the detection limit. A normal distribution with 5%, 20% and 40% of the truncated data below the detection limit was assumed for the simulation, respectively. The expectation maximization method was applied.

用类似方式模拟对数正态分布数据集中未检出数据的估算. 模拟数据集的几何均值和几何标准差分别为 148.00 和 4.11. 仍然从随机生成数据中删去 5%、20% 和 40% 低值. 此时的检出下限分别为 14.5、46.0 和 99.6. 分别用稳健顺序回归法和无偏最大似然函数法估算经过对数变换数据的均值和标准差 (稳健顺序回归计算的几何均值和几何标准差分别是  $149.00 \pm 4.39$ 、 $148.00 \pm 4.30$  和  $148.00 \pm 4.32$ ; 无偏最大似然函数法计算的几何均值和几何标准差分别是  $149.00 \pm 4.27$ 、 $150.00 \pm 4.24$  和

$146.00 \pm 4.39$ . 两个方法的结果非常类似. 根据计算分布参数和对应的检出下限可在正态分布假设下估算出未检出个体取值. 由于两个方法结果非常相似, 图 2 仅仅给出了用无偏最大似然函数法得到的恢复数据和删除数据的关系 (按大小顺序排列). 对于确实符合对数正态分布的数据, 上述方法均能获得很好的结果. 相比之下, 尝试用两种非参数方法 (Kaplan-Meier法和 Winsorization法) 获得的结果远不如参数方法.

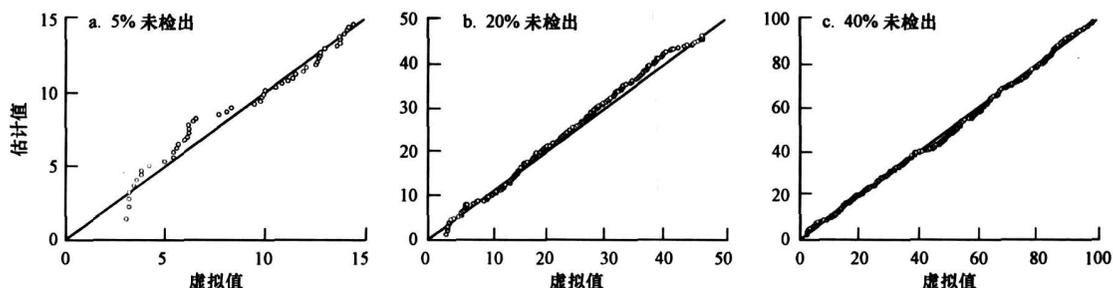


图 2 5%、20% 和 40% 未检出条件下采用无偏最大似然函数法获得的未检出数据估计值和虚拟未检出数据的关系 (虚拟对数正态分布)

Fig 2 Relationship between the virtual and the estimated values of the data below the detection limit. A log-normal distribution with 5%, 20% and 40% of the truncated data below the detection limits was assumed for the simulation, respectively. The bias-corrected maximum likelihood method was applied

### 4.2 实测数据举例

为演示方法的使用, 选用两组包含不同比例未检出数据的实测数据集进行未检出估值. 这些数据分别为华北平原表土狄氏剂 (样本量 379, 检出下限 0.004, 未检出率 8.4%) 和硫丹 (样本量 379, 检出下限 0.0005, 未检出率 3.7%). 样品的采集和测定方法在相关文献中有详细介绍 (Tao et al., 2008). 虽然在数据缺失且样本量不大的条件下很难进行分

布检验, 但根据经验, 微量有机污染物的含量大致呈对数正态分布, 因此, 选用参数的期望最大化迭代算法估计经对数变换后数据的总体均值和方差. 根据获得的分布参数 (变换后数据的几何均值和几何标准差: 狄氏剂为  $0.158 \pm 4.170$ , 硫丹为  $0.0375 \pm 3.9100$ ) 以及方法的检出下限对未检出数据集进行逐一估计. 对检出限下数据进行赋值后的两组数据集的频数分布如图 3 所示.

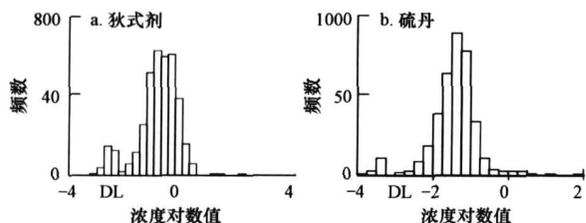


图 3 包括估算未检出数据的华北平原表土狄氏剂和硫丹含量的频数分布图 (数据经自然对数变换) (低于检出下限 (DL) 的数据来自估值)

Fig 3 Histogram of the concentrations of dieldrin and endosulfan in surface soil from Northern China Plain. The data are presented in log-scale. The data below the detection limits (DL) were those estimated

两个例子中估计的未检出数据分布与检出数据分布似乎不能很好地吻合, 未检出结果更像属于另一个独立的分布. 其可能的原因包括: ①原始数据并非严格的对数正态分布; ②对个体未检出数据的估值对于对数变换非常敏感; ③实测数据的误差不一定随机出现, 即总体中低于且接近检出下限的数据出现在检出数据中的可能与总体中高于且接近检出下限的数据出现在未检出数据中的可能性并非相同. 关于这一现象的可靠解释有待更多深入研究. 尽管如此, 这样的估计已是实际条件下能充分利用数据信息的最佳方法, 并为进一步深入进行数据分析提供了可能.

**责任作者简介:** 陶澍 (1950—), 男, 长江特聘教授, 博士生导师, 研究方向为微量污染物有毒污染物的排放、环境行为、归趋和效应。

#### 参考文献 (References):

- Cohen A C. 1950 Estimating the mean and variance of normal populations from singly truncated and double truncated samples [J]. *Ann Math Statist* 21: 557—569
- Cohen A C. 1959 Simplified estimators for the normal distribution when samples are singly censored or truncated [J]. *Technometrics* 1 (3): 217—237
- Dixon W J, Tukey J W. 1968 Approximate behavior of the distribution of in size-d-T (trimming/winsorization) [J]. *Technometrics* 10: 83—98
- Dempster A P, Laird N M, Rubin D B. 1977 Maximum likelihood from incomplete data via the EM algorithm [J]. *Journal of the Royal Statistical Society (Series B Methodological)*, 39(1): 1—38
- Efron B, Tibshirani R J. 1993. *An introduction to the bootstrap* [M]. New York: Chapman & Hall
- Gleis A. 1985. Estimation for small data sets with detection limits [J]. *Environ Sci Technol* 19 (12): 1201—1206
- Gilbert R O. 1987. *Statistical methods for environmental pollution monitoring* [M]. New York: Van Nostrand Reinhold
- Hebel D R. 1990 Less than obvious—statistical treatment of data below the detection limit [J]. *Environ Sci Technol* 24(12): 1766—1774
- Hebel D R. 2005a Nondetects and data analysis: Statistics for censored environmental data [M]. New York: John Wiley and Sons
- Hebel D R. 2005b More than obvious: Better methods for interpreting nondetect data [J]. *Environ Sci Technol* 39(20): 419—423
- Kaplan E L, Meier O. 1958. Nonparametric estimation from incomplete observations [J]. *J Amer Statist Assn* 53: 457—481
- Kroll C N, Stedinger J R. 1996. Estimation of moments and quantiles using censored data [J]. *Water Resources* 32: 1005—1012
- Manly B F J. 1997. *Randomization bootstrap and Monte Carlo methods in biology* (Second Edition) [M]. London: Chapman & Hall
- Newman M C, Dixon P M, Pinder J E. 1990 Estimating mean and variance for environmental samples with below detection limit observations [J]. *Water Resources Bulletin* 25(4): 905—916
- 李淑媛, 苗丰民, 刘国贤, 等. 1995. 渤海底质重金属环境背景值初步研究 [J]. *海洋学报 (中文版)*, 17(2): 78—85
- Li S Y, Miao F M, Liu G X, *et al*. 1995. The pilot study of the environmental background heavy metal concentration in the Bohai sediment [J]. *Acta Oceanologica Sinica (Chinese version)*, 17(2): 78—85 (in Chinese)
- 刘凌, 陆桂华. 2003. 多氯联苯湿地生物降解规律预测及污染风险分析 [J]. *水科学进展*, 14(1): 79—84
- Liu L, Lu G H. 2003. PCB biodegradation in constructed wetland and contamination risk analysis [J]. *Advances In Water Science* 14(1): 79—84 (in Chinese)
- Perrson T, Rootzen H. 1977. Simple and highly efficient estimators for a type I censored normal sample [J]. *Biometrika* 64: 123—128
- Saw J G. 1961. The bias for the maximum likelihood estimates of location and scale parameters given a type II censored normal sample [J]. *Biometrika* 48: 448—451
- Schneider H. 1986. *Truncated and censored samples from normal populations* [M]. New York: Marcel Dekker Inc
- Singh A K, Singh A, Engelhardt M. 1997. The lognormal distribution in environmental applications [R]. EPA/600/R-97/006 Technology Support Center Issue Paper
- Singh A, Nocerino J M. 2002. Robust estimation of the mean and variance using environmental data sets with below detection limit observations [OL]. <http://www.epa.gov/nerkesd1/omb/research/papers/jn103.pdf> 60: 69—86
- Singh A, Robert M, Sanghee E L. 2006. On the Computation of a 95% Upper Confidence Limit of the Unknown Population Mean Based Upon Data Sets with Below Detection Limit Observations [OL]. EPA/600/R-06/022. Washington D C: US EPA, <http://epa.gov/esd/tsc/indices/EPA%20600%20R-06%20022.pdf>
- 陶澍. 1994. *应用数理统计方法* [M]. 北京: 中国环境科学出版社
- Tao S. 1994. *Mathematical methods of statistics* [M]. Beijing: China Environmental Science Press (in Chinese)
- Tao S, Liu W X, Li Y, *et al*. 2008. Organochlorine pesticides contaminated surface soil as remission source in the Hahe Plain, China [J]. *Environ Sci Technol* 42: 8395—8400
- USEPA. 1991. Technical support document for water quality based toxics control [OL]. Washington D C: Office of Water Enforcement and Permitting <http://www.epa.gov/npd/es/pubs/owm0264.pdf>
- UNCENSOR 5.1. 2003. A statistical program for left-censored data sets [Z]. University of Georgia Savannah River Ecology Laboratory
- USEPA. 2006. Data quality assessment: Statistical methods for practitioners [OL]. EPA QA/G-9S, EPA/240/B-06/003. Washington D C: US EPA. <http://www.epa.gov/quality/qs-docs/g9s-final.pdf>
- 王海霞. 2007. 亚微米气溶胶的粒径谱及其中多环芳烃的研究 [D]. 大连: 中国科学院大连化学物理研究所, 7—8
- Wang H X. 2007. Studies on the Size Distribution and Polycyclic Aromatic Hydrocarbon Contents of Sulphuric Acid Aerosols [D]. Dalian: Dalian Institute of Chemical Physics, Chinese Academy of Sciences, 7—8 (in Chinese)
- Wang Z, Chen J W, Yang P, *et al*. 2009. Distribution of PAHs in pine (*Pinus thunbergii*) needles and soils correlates with their gas-particle partitioning [OL]. *Environ Sci Technol* <http://pubs.acs.org/doi/abs/10.1021/es802067e>
- 周军莉. 2001. 林带防护效应的数值模拟研究 [D]. 西安: 西安建筑科技大学, 19—36
- Zhou J L. 2001. The numerical simulation to the defending effect of windbreaks [D]. Xi'an: Xi'an University of Architecture and Technology, 19—36 (in Chinese)