THE STANDARD ERROR OF A WEIGHTED MEAN CONCENTRATION—I. BOOTSTRAPPING VS OTHER METHODS

DONALD F. GATZ

Illinois State Water Survey, 2204 Griffith Drive, Champaign, IL 61820, U.S.A.

and

LUTHER SMITH

ManTech Environmental Technology, P.O. Box 12313, 2 Triangle Drive, Research Triangle Park, NC 27709, U.S.A.

Abstract—Concentrations of chemical constituents of precipitation are frequently expressed in terms of the precipitation-weighted mean, which has several desirable properties. Unfortunately, the weighted mean has no analytical analog of the standard error of the *arithmetic* mean for use in characterizing its statistical uncertainty. Several approximate expressions have been used previously in the literature, but there is no consensus as to which is best. This paper compares three methods from the literature with a standard based on bootstrapping. Comparative calculations were carried out for nine major ions measured at 222 sampling sites in the National Atmospheric Deposition/National Trends Network (NADP/NTN). The ratio variance approximation of Cochran (1977) gave results that were not statistically different from those of bootstrapping, and is suggested as the method of choice for routine computing of the standard error of the weighted mean. The bootstrap method has advantages of its own, including the fact that it is nonparametric, but requires additional effort and computation time.

Key word index: Precipitation chemistry, bootstrap methods, statistics, wet deposition, precipitation-weighted mean.

INTRODUCTION

Central tendencies of ion concentrations in precipitation are often expressed in terms of the weighted mean, M_w, with precipitation amount, or the roughly equivalent quantity, sample volume, as the weighting factor. M_w is useful because it limits the influence of the very high concentrations that can occur in very light rains, and yields an overall concentration that is proportional to ion mass deposition (also known as loading). This proportionality makes it possible to compute spatial deposition fields using the "indirect" method (Vong *et al.*, 1989), in which spatial fields of ion concentration and precipitation amount are estimated separately and then multiplied together at grid intersections to produce the deposition field. This "indirect" method has the advantage of allowing us to estimate the precipitation field from all available precipitation measuring stations, which typically form a much denser network than do the precipitation chemistry sampling stations.

As with any measured or derived quantity, the statistical uncertainty of M_w is an important issue.

When Hawley *et al.* (1988) discussed the standard error of M_w, they pointed out that the method for estimating its value varies according to the purpose for which it is to be used. For example, in the *complete sampling* case, all of the possible samples for a specific location and time interval were collected and analyzed. Method for estimating the standard error of the weighted mean, SEM_w, for this case were dealt with in detail by Hawley *et al.* (1988).

The *representative sampling* case, however, is more often encountered in research, where observations from a finite set of samples are used to estimate some feature (such as M_w) of a broader population. This paper describes methods of estimating SEM_w that pertain to the representative sampling case.

Although M_w is a very useful concept in precipitation chemistry, it has one notable drawback: there is no readily derivable, generally applicable, analog of the standard error of the mean to express the uncertainty of the precipitation-weighted mean. A theoretical mathematical-statistical development of a formula for SEM_w would require knowledge of the statistical distributions of ionic concentration and

precipitation amount, and possibly the relationship of these two quantities to each other. Thus, a theoretical formula that would be valid for the conditions at one location would not necessarily hold at another if the distribution of precipitation were different. This would also be true for another place or another ion with a different distribution of concentration, or for a different relationship between precipitation and concentration.

Several expressions for SEM_w have been proposed and used in varying degrees in the atmospheric chemistry literature. Miller (1977) suggested an expression that was also used by Liljestrand and Morgan (1979) and by Topol et al. (1985) to express the uncertainty of the M_w. Galloway et al. (1984) used a somewhat different equation for SEM_w, and Endlich et al. (1988) computed confidence intervals for M_w based on the formula for the ratio variance given by Cochran (1977). Most of these proposed expressions for SEM_w were unaccompanied by any discussion (or justification) of the assumptions required in their derivation.

There is no general consensus among precipitation chemistry researchers on which expression is preferable, or the conditions under which one expression might be favored over another. Hawley (1985) compared three methods, including that of Miller (1977) mentioned earlier, and the jackknife technique described by Efron (1982). He concluded that each of the three methods for computing the variance of M_w had shortcomings, especially for sample sets < 50.

Fortunately, a very satisfactory new method appears to be available. Bootstrap methods (Efron and Tibshirani, 1986, 1991) were developed as a means of using today's readily available computational power to obtain a numerical estimate of the standard error of estimators other than the mean, such as, e.g. M_w. Efron and Tibshirani (1991) pointed out that the bootstrap algorithm can be applied to almost any statistical estimation problem. Further, they argued that the same logic that makes the usual expression of the standard error reasonable as an assessment of the standard error of the arithmetic mean applies equally well to the bootstrap method as an assessment of the standard error of any general statistic. Applied to ion concentrations in precipitation, the bootstrap provides a statistically valid *nonparametric* estimate of the SEM_w.

The purpose of this paper is to illustrate how the bootstrap technique can be used to estimate SEM_w for ion concentrations in precipitation, and to compare its results with SEM_w values estimated from the three literature methods listed earlier.

METHODS

Values of SEM_w were computed for the nine major ions in precipitation at each of 222 (i.e. essentially all) sites in the National Atmospheric Deposition Program/National Trends Network (NADP/NTN) database for the three years

1988–1990 (NADP/NTN, 1992). A map of the sampling site locations and descriptions of sampling and analytical methods have been published by the National Atmospheric Deposition Program (1992). Separate values were computed covering the entire 1988–1990 data period for each of the three formulations from the literature, and compared to those computed by the bootstrap algorithm.

This period of record was chosen to insure reasonable numbers of samples even at sites with limited precipitation; it was the latest three-year period available when the work was begun. The maximum number of samples at any sampling site was 147, and 90% of the sites had 45 samples or more. All available sites were included, without regard to data completeness criteria. Samples not collected under the standard wet-only protocol were excluded. Normal network sample duration is one week, but samples were not excluded based on duration. All calculations were performed using SYSTAT statistical software, version 5.03 (Wilkinson, 1990).

Methods Proposed in the Literature

Miller (1977) chose the following formula to compute SEM_w:

$$(SEM_w)^2 = \frac{1}{n}\left(\frac{1}{\sum P_i}\right) \sum P_i (X_i - \bar{X}_w)^2$$

where P_i is the precipitation amount for sample i, X_i the concentration in sample i, $\bar{X}_w$ the precipitation-weighted mean concentration over samples $i = 1 - n$, $\bar{X}_w$ the $\sum(P_i X_i)/\sum P_i$, and n the number of samples.

The choice of Galloway et al. (1984), using the same notation, was:

$$(SEM_w)^2 = \frac{n}{(\sum P_i)^2}\left[\frac{n\sum P_i^2 X_i^2 - (\sum P_i X_i)^2}{n(n-1)}\right].$$

No rationale, justification, or discussion of any assumptions or limitations inherent in these formulations was provided in either paper.

Endlich et al. (1988) expressed SEM_w as an approximate ratio variance given by Cochran (1977). The formula that follows is given in an SRI International (1986) report:

$$(SEM_w)^2 = \frac{n}{(n-1)(\sum P_i)^2}[\sum (P_i X_i - \bar{P}\bar{X}_w)^2$$
$$- 2\bar{X}_w \sum (P_i - \bar{P})(P_i X_i - \bar{P}\bar{X}_w)$$
$$+ \bar{X}_w^2 \sum (P_i - \bar{P})^2].$$

The Bootstrap Method

Efron and Tibshirani (1991) recently described the bootstrap method, and the following description borrows heavily from theirs. The name "bootstrap" refers to the repeated random resampling, with replacement, of some original data set $x = (x_1, x_2, \dots, x_n)$, to generate new data sets of size n. A bootstrap sample is denoted as $x^* = (x_1^*, x_2^*, \dots, x_n^*)$. Each x_i^* is one of the original x values, randomly selected.

The bootstrap estimate of SEM_w is computed as follows: (1) a large number B of independent bootstrap samples, each of size n, is generated using a random number generator, (2) M_w is computed for each bootstrap sample, and (3) the empirical standard deviation of the B bootstrapped weighted means is computed as the bootstrap estimate of SEM_w. In this work, B was 200. The value of n varied from site to site, with a maximum of 147 for weekly sampling over three years. At a few sites with limited precipitation, n was < 50.

Efron and Tibshirani (1991) also indicated that, while ideally B would be infinite, randomness in the bootstrap standard error that comes from using a finite B is usually negligible for $B > 200$, and that values of B as small as 25 can often give satisfactory results. For the special case of the arithmetic mean, as B increases without bound, the

bootstrap standard error estimate approaches the square root of $(n-1)/n$ times the usual formula for the standard error of the mean. If the same accuracy applied to all other applications of the bootstrap method, the method would be accurate to within 2% for $n > 25$.

The bootstrap method typically assumes that data are independent and identically distributed (iid), and that has been assumed in this work. Presumably, the previously published methods would make the same assumption, although only Cochran (1977) makes it clear. In the case of M_w, the relevant distribution should be the joint distribution of precipitation amount and ion concentration (Hawley, 1985). Because of the well-known inverse relationship (Junge, 1963) between ion concentration in precipitation and precipitation amount, and also because ion concentrations frequently have seasonal variations, the assumption that the single variable *concentration* is identically distributed is generally not valid. However, the literature does not discuss the matter of the joint distribution of ion concentration and precipitation amount.

RESULTS

Comparisons of SEM$_w$ values computed by the bootstrap method with those computed by the formulations of Miller (1977), Galloway *et al.* (1984), and Cochran (1977) are shown for all major ions at NADP/NTN sites in Figs 1–9 and Table 1. The bootstrap method is taken as the standard because of its established statistical validity and because it requires no assumptions about distributions. Each point in the three panels of these figures represents one NADP/NTN network sampling site. If the values computed by the respective literature formulations agreed perfectly with those computed by bootstrapping, all the points would fall on the 1:1 lines shown in the three panels. To the extent that the points fall predominantly above or below the 1:1 line, the method shows a bias with respect to the bootstrap method. The degree of scatter in the points represents the imprecision of the method, relative to the bootstrap results.

The hypothesis that the difference between the bootstrap and each alternative method was zero was tested for each ion. The paired *t*-test was used when the differences (computed for each sampling site) between the SEM$_w$ calculated by the bootstrap and the alternative method from the literature appeared to come from reasonably normally distributed population, as judged by inspection from a stem and leaf plot; it tests the hypothesis that the mean is zero. When the stem and leaf plots suggested that the distributions were reasonably symmetric, but not normally distributed, the Wilcoxon signed ranks test was used; when asymmetry appeared in the plots, Fisher's sign test was used. The signed rank and sign tests address whether the median difference is zero.

For each ion, the overall error rate of the testing was protected by employing Hommel's (1988) modification of the standard Bonferroni procedure (Miller, 1981) for simultaneous hypothesis testing. Hommel's modification is more powerful than the usual Bonferroni method or other modifications (Hommel, 1989). Hommel's procedure is conservative (i.e. the true overall error rate does not exceed the nominal rate) if Simes' (1986) test, on which Hommel based his test, is conservative. For the situation here, all test statistics are asymptotically distributed as standard normal variates, and Simes' (1986) simulation results indicate that the testing done here should be conservative.

The results of the hypothesis tests are presented in Table 1 in the form of adjusted *p*-values, as described by Wright (1992). The adjusted *p*-value for a particular hypothesis represents the smallest overall error rate for which that particular hypothesis can be rejected. Thus, the reader is free to choose any *overall* error rate, *p*, and read *directly* from Table 1 whether any particular test is statistically significant (i.e. if the value in Table 1 is $< p$, then that test is statistically significant at the overall error rate of *p*).

Table 1. Probabilities that the mean SEM$_w$ values (median values for the Wilcoxon and the Fisher sign tests) computed by three literature formulations are the same as those computed by the bootstrap method, given the observed differences in results, for 222 sampling sites in the NADP/NTN database

Ions	Tests	Miller (1977)	Galloway *et al.* (1984)	Cochran (1977)
Cl^-	Wilcoxon test	0.0000	0.0000	0.3404
SO_4^{2-}	*t*-test	0.0000	0.0000	0.1280
NO_3^-	*t*-test	0.0000	0.0000	0.3636
H^+	Fisher sign test	0.0000	0.0000	0.5028
NH_4^+	*t*-test	0.0000	0.0000	0.9268
Na^+	Wilcoxon test	0.0118	0.0000	0.9571
K^+	Wilcoxon test	0.0004	0.0000	0.2863
Ca^{2+}	Wilcoxon test	0.5045	0.0003	0.1824
Mg^{2+}	*t*-test	0.8502	0.0000	0.8502

Note: The paired *t*-test was used for ions where the distributions of the differences between methods were approximately normal; the distribution-free Wilcoxon signed rank test was used for non-normal symmetric distributions and the Fisher sign test otherwise.

Figure 1 shows results of the comparisons for chloride (Cl⁻). All three methods show a general overall agreement with the bootstrap results. The top panel, comparing bootstrap results at the NADP/NTN sites to those computed from the Miller (1977) formula, shows a predominance of points with bootstrap SEM_w values greater than $0.1\ mg\,\ell^{-1}$ falling below the 1:1 line, while for the Galloway et al. method the points in the same range are predominantly above the line. Table 1 indicates significant differences (all references to significance refer to the 5% level) from the bootstrap method for both the Miller and the Galloway et al. methods, based on the nonparametric Wilcoxon method. The Cochran formulation shows excellent agreement with the bootstrap method in Fig. 1, and correspondingly the probabilities show no significant bias in Table 1.

A comparison of the dispersion or degree of scatter of the points from the 1:1 line, in the comparisons of Figs 1–9 is shown in Figure 10. The degree of scatter is expressed by the pairwise root-mean square (RMS) differences between the SEM_w values of the respective literature methods and the bootstrap method. For Cl⁻, for example, Fig. 10 shows the greatest scatter for the Miller method, an intermediate value for the Galloway et al. method, and the smallest value for the

Cochran method. (With a single off-scale outlier eliminated, the Miller RMS value dropped from 0.057 to 0.016, while values of the other two methods changed very slightly.)

Figures 2 and 3 show similar results for sulfate (SO_4^{2-}) and nitrate (NO_3^-). Both ions show significance at the 5% level, based on the t-test, that the mean differences between SEM_w values computed by the Miller and Galloway et al. methods are not the same as those computed by bootstrapping. The biases are in the same *directions* as those seen for Cl⁻. No bias appears in the Cochran method, which also exhibits a much smaller relative scatter (Fig. 10) than the results of the other two methods. (Elimination of one outlier from the SO_4^{2-} data did not materially affect the probabilities; however, it reduced the Miller RMS value (Fig. 10) from 0.062 to 0.025, without substantially affecting those of the other methods.)

The results for hydrogen ion (H⁺) (Fig. 4) are also similar to those of the three anions. The Miller and Galloway et al. methods show apparent biases, in the same directions as seen above, and indeed, both methods yield SEM_w values different from the bootstrap values at the 5% level based on the t-test (Table 1). The RMS scatter (Fig. 10) is greatest for the Galloway et al. method, intermediate for the Miller
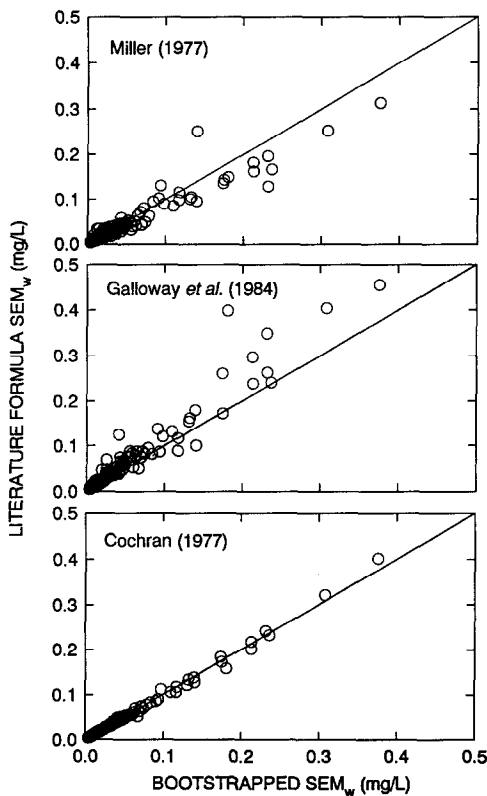

Fig. 1. Comparison of bootstrapped SEM_w values for Cl⁻ at 222 NADP/NTN sites with values computed by three formulations proposed previously in the literature. One off-scale outlier is not shown.
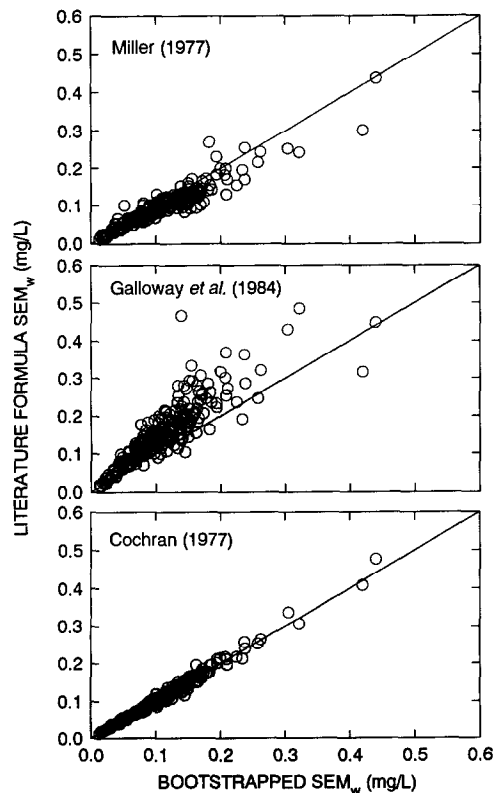

Fig. 2. Comparison of bootstrapped SEM_w values for SO_4^{2-} at 222 NADP/NTN sites with values computed by three formulations proposed previously in the literature. One off-scale outlier is not shown.
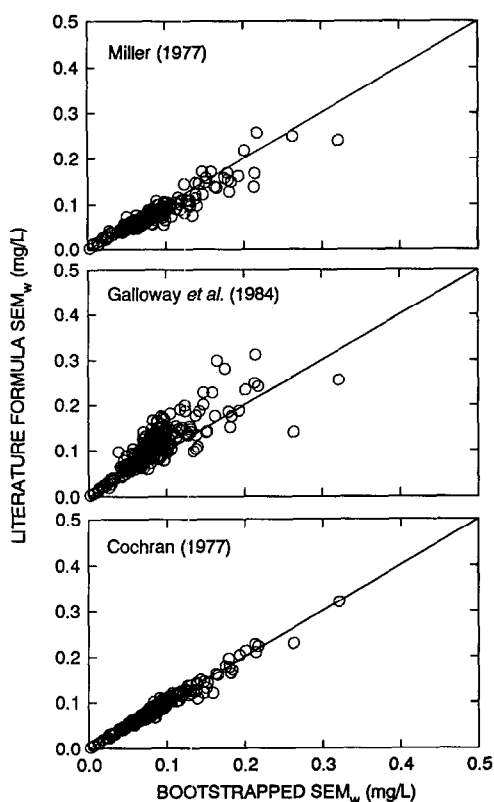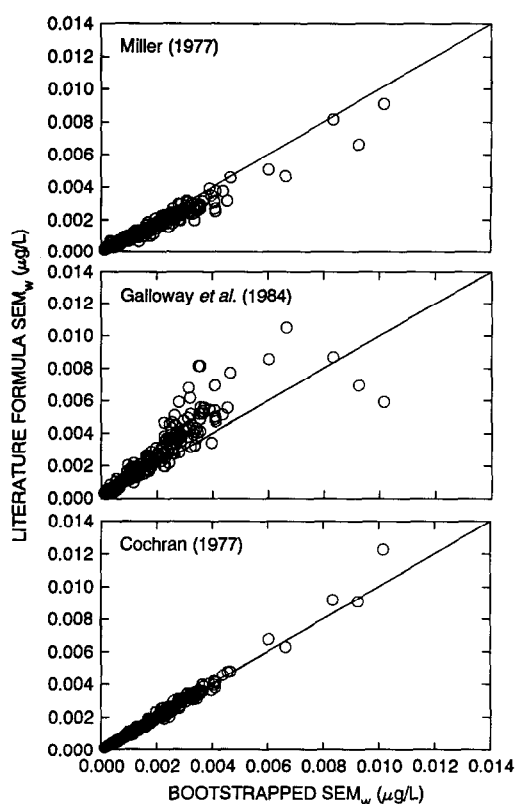
Fig. 3. Comparison of bootstrapped SEM_w values for NO_3^- at 222 NADP/NTN sites with values computed by three formulations proposed previously in the literature.

Fig. 4. Comparison of bootstrapped SEM_w values for H^+ at 222 NADP/NTN sites with values computed by three formulations proposed previously in the literature.

method, and smallest for the Cochran method. This same pattern is evident for ammonium (NH_4^+) as well (Fig. 5). The biases in the Miller and Galloway *et al.* methods are highly significant by the *t*-test (Table 1), while the Cochran method again shows no bias and the least scatter (Fig. 10).

For sodium (Na^+) and potassium (K^+), the pattern is somewhat different, however (Figs 6 and 7, respectively). The previous bias toward low SEM_w values from the Miller method is not evident in these figures, although Wilcoxon tests (Table 1) still found significant differences between the medians of the respective methods for both ions at the 5% level. The previously seen bias towards high SEM_w values in the Galloway *et al.* method is still evident for Na^+ in Fig. 6, however, and is confirmed by the Wilcoxon test (Table 1). The bias of the Galloway *et al.* method for K^+ is visually much smaller (Fig. 7), but the Wilcoxon test (Table 1) still found a significant difference from the bootstrap results.

The RMS scatter of both the Na^+ and K^+ plots is greatest for the Miller formulation (Fig. 10), although the Na^+ result is strongly affected by two off-scale data points. The Cochran method again achieved the best agreement with the bootstrap method for both ions, showing no bias and the least scatter.

For calcium (Ca^{2+}) and magnesium (Mg^{2+}) (Figs 8 and 9, respectively), the *t*-test (Table 1) indicated that only the Galloway *et al.* method is significantly biased (Table 1), although this bias is not obvious from the figures. The RMS scatter (Fig. 10) is greatest for the Miller method and least for the Cochran method for both ions.

Uncertainties in M_w ion concentrations may be expressed simply by stating the standard error. However, it is often more useful to use both M_w and SEM_w to express the uncertainty by the approximate confidence interval $M_w \pm SEM_w \, z^{(\alpha)}$, where $z^{(\alpha)}$ is the 100 α percentile point of a standard normal distribution. If the M_w concentrations were normally distributed, this interval would have the approximate coverage probability of $1 - 2\alpha$. For a 95% confidence interval ($\alpha = 0.025$, $1 - 2\alpha = 0.95$), the $z^{(\alpha)}$ is -1.96. While the assumption of normally distributed M_w's permits the convenient computation of confidence intervals, the normal assumption is not always justified. Hawley's (1985) analyses of data from several Multistate Atmospheric Power Production Pollution Study (MAP3S) sites indicated that the M_w is approximately lognormal for $n < 100$ and approximately normal for larger n. Figure 11 shows a histogram of the distribution of 200 bootstrapped VWM values for K^+ at the
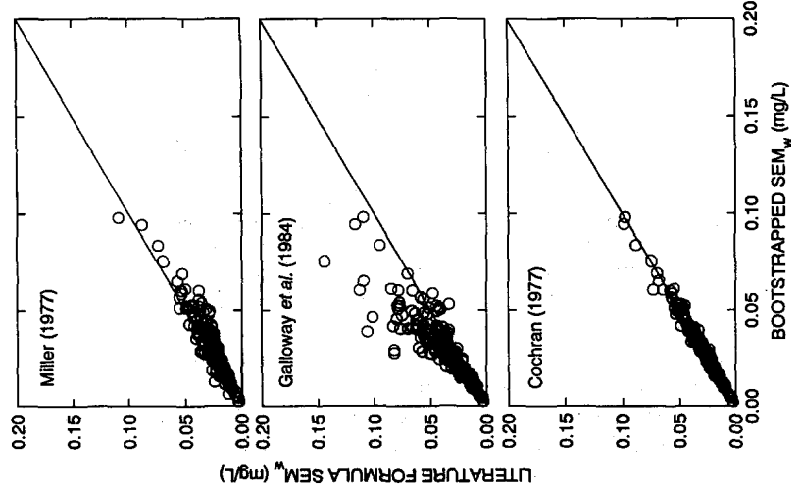
Fig. 7. Comparison of bootstrapped SEM$_w$ values for K$^+$ at 222 NADP/NTN sites with values computed by three formulations proposed previously in the literature. One off-scale outlier is not shown.
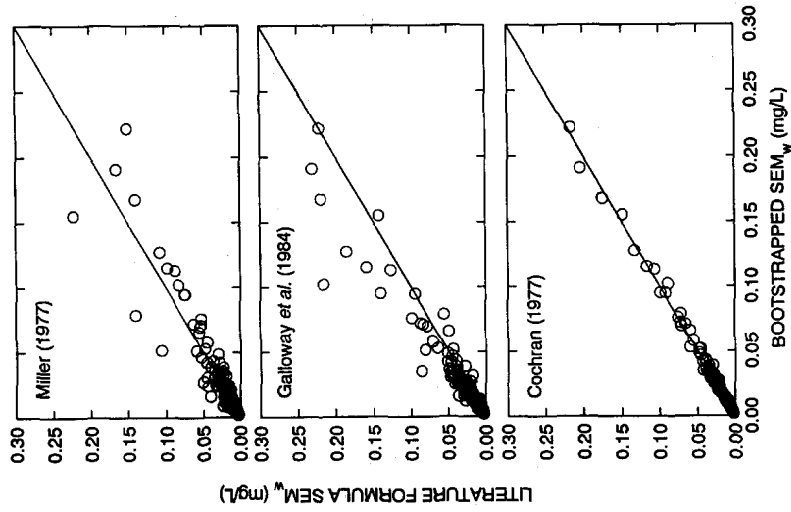
Fig. 6. Comparison of bootstrapped SEM$_w$ values for Na$^+$ at 222 NADP/NTN sites with values computed by three formulations proposed previously in the literature. Two off-scale outliers are not shown.
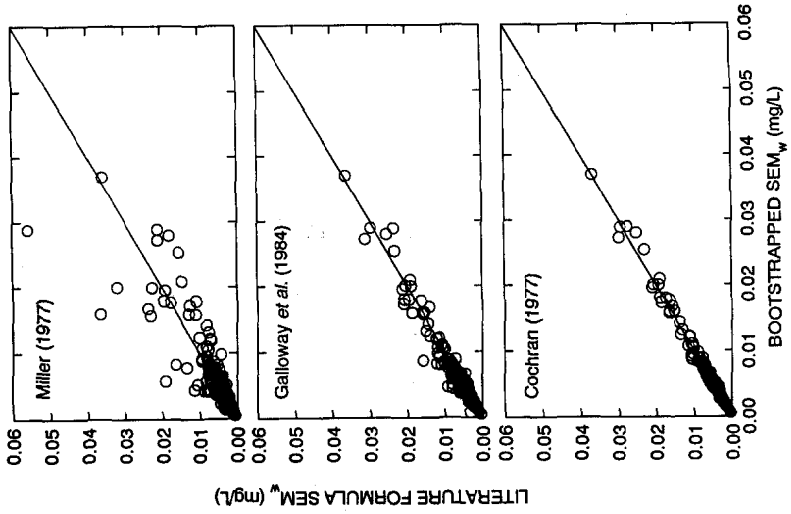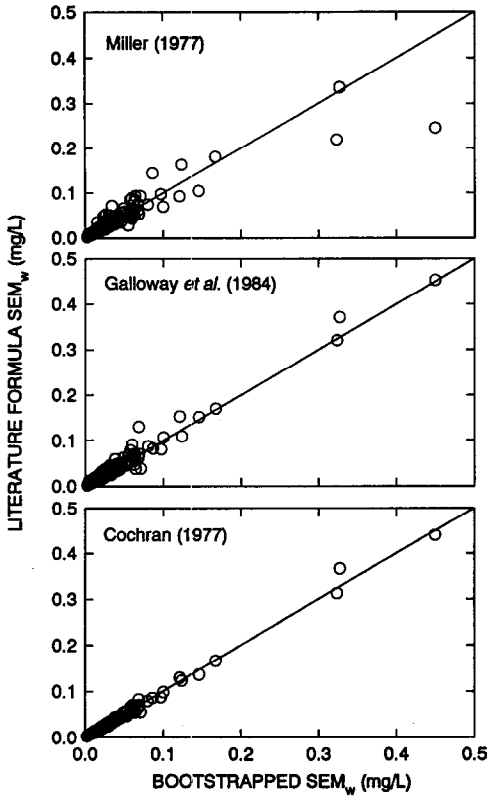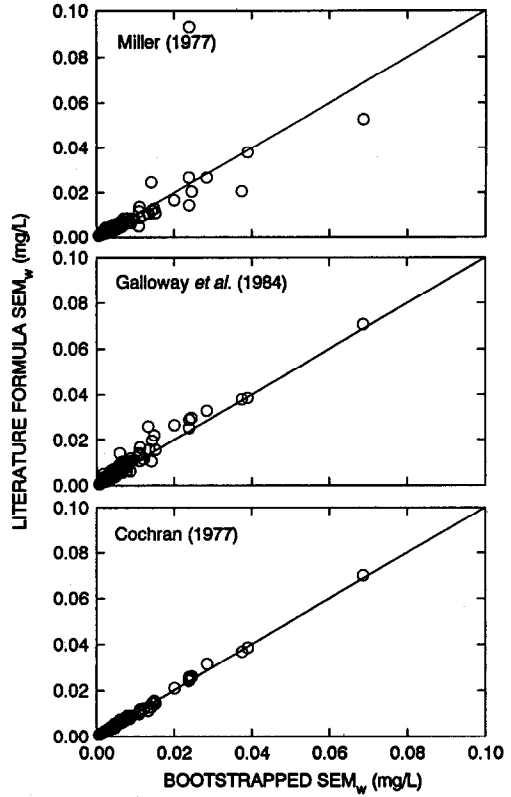
Fig. 5. Comparison of bootstrapped SEM$_w$ values for NH$_4^+$ at 222 NADP/NTN sites with values computed by three formulations proposed previously in the literature. One off-scale outlier is not shown.

Fig. 8. Comparison of bootstrapped SEM_w values for Ca^{2+} at 222 NADP/NTN sites with values computed by three formulations proposed previously in the literature.

Fig. 9. Comparison of bootstrapped SEM_w values for Mg^{2+} at 222 NADP/NTN sites with values computed by three formulations proposed previously in the literature.
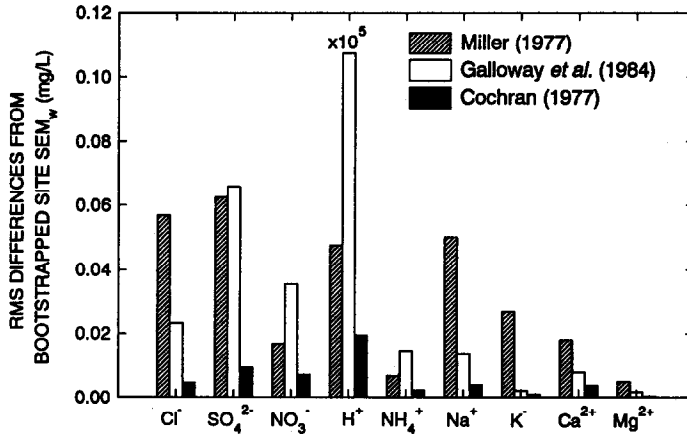

Fig. 10. Degree of scatter in the method comparisons in Figs 1–9, as expressed by the RMS differences of the SEM_w values of the respective literature methods from the corresponding bootstrap values. The H^+ SEM_w values have been multiplied by 10^5 to make them large enough for visual comparison. Effects of removing outliers from the data sets are discussed in the text.

Florida Everglades site, along with a smooth curve showing the normal distribution corresponding to the measured VWM and the bootstrapped SEM_w. Close inspection reveals that the assumption of a normal distribution in this case would lead to an inaccurate assessment of the 95% confidence interval. The following paper (Gatz and Smith, 1994) explores this issue in more detail.
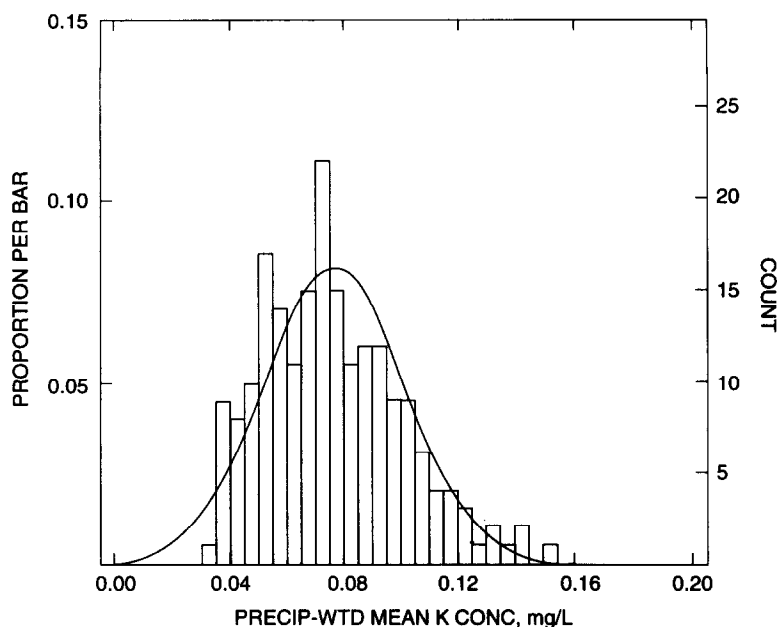
Fig. 11. Histogram of 200 bootstrapped precipitation-weighted mean K^+ concentrations at the Florida Everglades site in the NADP/NTN network. The data were collected over a 3 yr period, 1988–1990. The smooth curve is the normal distribution based on the mean and standard deviation of the 200 M_w values. Standard 95% confidence limits based on the assumption of a normal distribution are 0.026 and 0.126 mg ℓ^{-1}, and those estimated from the bootstrapped data by the percentile method (Efron and Tibshirani, 1986) are 0.037 and 0.136 mg ℓ^{-1}.

DISCUSSION AND CONCLUSIONS

The results presented graphically in Figs 1–9 and the statistical comparisons in Table 1 leave no doubt that the Cochran method gives results that are not statistically different from those produced by the bootstrap method, at least over the three year data period of this study. The Miller (1977) method produced mean SEM_w values that were significantly different from those of the bootstrap method for seven of the nine ions tested. The SEM_w values computed from the Miller method were typically underestimates, compared to those from the bootstrap method. This agrees with results reported earlier by Hawley (1985). The same comparisons between the bootstrap method and that of Galloway et al. (1984) produced significantly different results for all nine ions. The Galloway et al.. results were frequently biased high (overestimates) compared to those from the bootstrap method.

In contrast, the agreement between the Cochran method and the nonparametric bootstrap methods was excellent. These two completely different methods gave results that were not statistically different for nine different ions at more than 200 sampling locations over a three-year time period.

In terms of ease of computation, the Cochran method has an edge over the bootstrap method and would appear to be the method of choice for routine

computing of SEM_w. However, smaller sample sizes than those utilized here may warrant investment in the additional computation of the bootstrap, particularly when confidence limits are desired.

REFERENCES

Cochran W. G. (1977) *Sampling Techniques* (3rd Edn). Wiley, New York.
Efron B. (1982) *The Jackknife, the Bootstrap, and Other Resampling Plans.* Society for Industrial and Applied Mathematics, Philiadelphia.
Efron B. and Tibshirani R. (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.* 1(1), 54–77.
Efron B. and Tibshirani R. (1991) Statistical data analysis in the computer age. *Science* 253, 390–395.
Endlich R. M., Eynon B. P., Ferek R. J., Valdes A. D. and Maxwell C. (1988) Statistical analysis of precipitation

chemistry measurements over the eastern United States — I. Seasonal and regional patterns and correlations. *J. appl. Met.* **27**, 1326–1333.

Galloway J. N., Likens G. E. and Hawley M. E. (1984) Acid precipitation: natural versus anthropogenic components. *Science* **226**, 829–831.

Gatz D. F. and Smith L. (1995) The standard error of a weighted mean concentration — II. Estimating confidence intervals. *Atmospheric Environment* **29**, 1195–1200.

Hawley M. E. (1985) Comparison of volume-weighted mean concentrations in precipitation chemistry. *Ninth Conference on Probability and Statistics in the Atmospheric Sciences*, pp. 55–59. American Meteorological Society, Boston.

Hawley M. E., Galloway J. N. and Keene W. C. (1988) Standard error calculations for non-sea salt constituents in marine precipitation. *Water Air Soil Pollut.* **42**, 87–102.

Hommel G. (1988) A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* **75**, 383–386.

Hommel G. (1989) A comparison of two modified Bonferroni procedures. *Biometrika* **76**, 624–625.

Junge C. E. (1963) *Air Chemistry and Radioactivity*, pp. 311–315. Academic Press, New York.

Liljestrand H. M. and Morgan J. J. (1979) Error analysis applied to indirect methods for precipitation acidity. *Tellus* **31**, 421–431.

Miller J. M. (1977) A statistical evaluation of the U.S. precipitation chemistry network. In *Precipitation Scavenging (1974)* (edited by Semonin R. G. and Beadle R. W.), pp. 639–659. Available as CONF 74100 from National Technical Information Service, U.S. Dept. of Commerce, Springfield, VA.

Miller R. G. (1981) *Simultaneous Statistical Inference*, pp. 67–68. Springer, New York.

National Atmospheric Deposition Program (1992) NADP/NTN Annual Data Summary: Precipitation Chemistry in the United States. 1991. Natural Resources Ecology Laboratory, 475 pp. Colorado State University, Fort Collins.

Simes R. J. (1986) An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73**, 751–754.

SRI International (1986) Statistical analysis of precipitation chemistry measurements over the eastern United States. Final Report, Contract U101-3, Utility Acid Precipitation Study Program, pp. 3–9, UAPSP-112, Palo Alto, CA.

Topol L. E., Lev-On M., Flanagan J., Schwall R. J., and Jackson A. E. (1985) Quality assurance manual for precipitation measurement systems. Contract No. 68-02-3767. Environmental Monitoring Systems Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Vong R., Cline S., Reems G., Bernert J., Charles D., Gibson J., Husar R., Olson A., Simpson J. and Sielkop S. (1989) Regional analysis of wet deposition for effects research. Report EPA/600/3-89/030, Environmental Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Corvallis, OR.

Wilkinson L. (1990) *SYSTAT: The System for Statistics.* SYSTAT, Inc., Evanston, IL.

Wright S. P. (1992) Adjusted *p*-values for simultaneous inference. *Biometrics* **48**, 1005–1013.