Definition and Number of Subsamples for Using Mosses as Biomonitors of Airborne Trace Elements

J. R. Aboal, J. A. Couto, J. A. Fernández, A. Carballeira

Área de Ecología, Facultad de Biología, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain

Abstract. We collected 50 subsamples of the moss *Scleropodium purum* from each of three sampling sites and determined the concentrations of As, Cd, Cu, K, Hg, Ni, Pb, Se, and Zn in each subsample. We then calculated the number of subsamples required to determine significant differences in the mean concentrations of two sampling sites. We found that to differentiate between an uncontaminated sampling site and another, slightly contaminated site, 30 subsamples are required from each. On the basis of these results and because, to date, there are no studies that justify the application of the previously proposed recommendations, studies of local variability in other areas, and under different conditions, with other contaminants and moss species must be undertaken. For posterior comparison of data on variability as well as application of the recommendations associated with such results, the general use of a single definition of a subsample is required. We propose the following definition of a subsample: "1 g dry weight (approx.), collected within a 25-cm radius of a node, selected at random from a 1×1-m sampling grid placed in a sampling site."

Some 30 years have passed since the moss technique for biomonitoring atmospheric contaminants was first applied (Rühling and Tyler 1968), and despite its extensive use—with more than 7000 sampling sites included in the last European Survey (Buse *et al.* 2003)—scientific studies that support or revise the existing protocols are still required.

The validity of the technique depends on obtaining a representative sample of the tissue concentrations of different elements in moss growing at a sampling site. To obtain the said sample from all of the moss present at the sampling site, a series of subsamples is collected and combined to make a single composite sample, with the aim of representing the existing local variability in concentrations of elements. However, despite the apparent simplicity of the procedure, the definition of a subsample, as well as its location, the area of collection (extension of the sampling site), and the number of samples will determine the concentration in the composite sample (Fernández *et al.* 2002). From among all of these aspects, the present study centers specifically on questions related to the definition and number of subsamples. Further information about sampling statistics and sampling procedures can be found in Sansoni and Iyengar (1978), Gy (1982), Keith (1988, 1991), Markert (1994, 1996), Quevauvillier (1995), and Markert *et al.* (2003).

Until now, no precise definition of a subsample has been given, in terms of extension, weight, and distance from other subsamples, and the concept has been left open to interpretation by individual researchers. The establishment of a definition of the subsample is essential for correct standardization of the technique with a view to homologation or implementation of a quality control system. For example, at present two subsamples can be considered as a few moss shoots collected at tens of meters from each other, or as large quantities of moss from the same clump and separated by a few centimeters.

As regards the number of subsamples required, it has been recommended that between 5 and 10 subsamples be collected from an area of 50×50 m (Rühling 1989, 1994a); this recommendation has been applied in almost all studies using the moss technique, independently of the moss species, the level of contamination, or the latitude of the sampling site. However, indiscriminate application of this recommendation is not valid because the number of subsamples that should be collected from a sampling site depends on the variability in the concentrations of contaminants in the moss (the greater the variability, the greater the number of samples required). Paradoxically, there are very few studies in which data are provided on the variability of concentrations of different elements within a sampling site (Cenci 1999; Fernández *et al.* 2002), and there are none that justify the recommendations made by Rühling (1989, 1994a). In a detailed study of local variability (in a sampling site of 35×35 m), Fernández *et al.* (2002) demonstrated that for a given element, it was necessary to collect a minimum of 30 subsamples, for an error of 20% in determining the mean, clearly more than the number generally recommended.

Correspondence to: J. R. Aboal; *email:* bfjaboal@usc.es

Finally, we believe that definition of the number of sub-samples required to demonstrate significant differences between two sampling sites, in terms of the concentrations of contaminants in the moss growing at each, is a fundamental aspect of the moss technique. In biomonitoring studies, sampling sites are commonly characterized by the differences in the tissue concentrations in the mosses; however, the researcher is not usually aware that the capacity to differentiate between them, with statistical certainty, is directly related to the number of subsamples collected. The aims of the present study are, therefore, (i) to develop a valid method of calculating the number of subsamples required to differentiate sampling sites on the basis of the concentrations of elements in mosses growing there; (ii) to apply this method to metals and metalloids data from mosses collected at three sampling sites to calculate the number of subsamples required as stated above; (iii) to study the effect of the differences between the concentrations of the pairs of sampling sites being compared on the number of subsamples required; and (iv) to propose a standardized definition of a subsample.

Materials and Methods

Sampling

Three sampling sites (SS) were selected on the basis of the degree of contamination, known from previous studies. One site (30 × 30 m; X = 49,300, Y = 475,834, UTM 29N ED50) was located in the surroundings of a FeSi smelter and showed exceptionally high levels of As and Cu (Fernández *et al*. 2004). The second site (25 × 20 m; X = 60,706, Y = 479,497, UTM 29N ED50) was located in the surroundings of a coal-fired power plant and showed intermediate levels of contamination by several contaminants (Couto 2002; Fernández *et al*. 2002; Couto *et al*. 2003). The third site (25 × 30 m; X = 58,640, Y = 471,040, UTM 29N ED50) was located in a rural area, far from industries and population nuclei and showed low levels of contamination (Aboal *et al*. 2004). The SS were located far from trees and there was only grass cover and/or sparsely distributed shrubs (*Ulex* spp., *Erica* spp., and *Pteridium aquilinum*).

At each SS, we collected 50 subsamples, defined as all of the moss present within a 25-cm radius of a determined point. We placed a regular, 1 × 1-m sampling grid (using metric tapes), and the position of each point within the grid was chosen using random numbers. If at any of the points there was no moss growing, the point was discounted and another selected, until the 50 subsamples were obtained. All sampling was carried out between 15 May and 15 June 2003 to minimize the possible effect of temporal variability on comparing the sampling sites.

Chemical Analysis

Prior to extraction, apical sections (3–4 cm) were separated from the moss shoots. After manual removal of all adhering material (plant remains, soil particles, etc.), the apices were washed for 30 sec with bidistilled water, with shaking, to remove mineral particles deposited on the surface, as well as plant remains, epiphytes, etc. Once cleaned and dried (20°C), the samples were homogenized, dried to constant weight (45°C) in a forced-air oven, and 1 g d.w. was digested with HNO_3 (65%) in a microwave oven (CEM MDS2100). The concentrations of Cu and Zn were determined by flame absorption spectrophotometry (Perkin Elmer 2100), and K by emission. Cd, Ni, and Pb were determined by graphite furnace spectrophotometry (Perkin Elmer AAnalyst 600) and As, Hg, and Se were determined by atomic fluorescence spectrophotometry (PSA Millenium Excalibur).

To monitor the processes of extraction and determination of the metal contents, certified reference material (GBW07604, poplar leaves) and an internal reference material (O.1. *Scleropodium purum*) were analyzed, one or both every 9/10 samples. The recoveries were satisfactory in all cases and ranged between 80% and 101%. The results of the internal reference material analysis were also used to calculate the variability associated with the extraction and determination processes. This variability, expressed as the coefficient of variation, was 4% for K; 6% for Cu, Pb, and Zn; 7% for Hg and Ni; 8% for As and Cd; and 10% for Se. Furthermore, the existence of contaminating material during processing and extraction was controlled for using analytical blanks, 1 every 9/10 samples.

Statistical Analysis

One important consideration in determining the statistical analysis to be carried out is that, in biomonitoring surveys with terrestrial mosses, the concentrations of contaminants in composite samples from each SS are used (Rühling 1994a, 1994b). For any given contaminant, the data are therefore mean concentrations of all subsamples collected from each SS. The hypothesis to be tested is therefore $\mu_1 = \mu_2$, where 1 and 2 represent the concentrations of the different SS. For this, we applied the calculations for the power of the test described as follows.

Power and Sample Size in Testing for Difference Between Two Means in Normal Populations. We may ask how many subsamples are required to find significant differences between μ_1 and μ_2. We used the sample size calculations recommended by Zar (1984) to test for difference between two means. Assuming each subsample comes from a normal population, we can estimate the minimum sample size to achieve the desired test characteristics (Cochran and Cox 1957, in Zar 1984):

$$n \geq \frac{2\sigma_p^2}{\delta^2} \left(t_{\alpha,v} + t_{\beta(1),v} \right)^2 \tag{1}$$

where n is the sample size; δ is the minimum detectable difference between population means; α is the significance level ($\alpha = 0.05$); $1 - \beta$ is the power of the test ($\beta = 0.1$); v is the number of degrees of freedom ($2*(n - 1)$); and σ_p^2 is the population variance, estimated by the pooled variance:

$$\sigma_p^2 = \frac{\sum_{i=1}^{n} \left(x_i 1 - \overline{X_1} \right)^2 + \sum_{i=1}^{n} \left(x_i 2 - \overline{X_2} \right)^2}{v_1 + v_2} \tag{2}$$

Four examples in which two populations are compared are shown in Figure 1. In each of these there is variation in the mean values (μ_1 and μ_2) and/or the deviations (σ_1 and σ_2) of the populations compared. In this figure, d represents the difference in the mean values for the populations being compared ($d = \mu_2 - \mu_1$). The values of δ when $n = 5$ and $n = 25$ (δ_5 and δ_{25}), corresponding to the four comparisons, are also shown. If $\delta \geq d$, no significant differences between the means will be found; if $\delta < d$, significant differences will be found. In all comparisons, an increase in n causes a reduction in δ, and thereby an increase in the capacity to identify significant differences between the populations being compared.
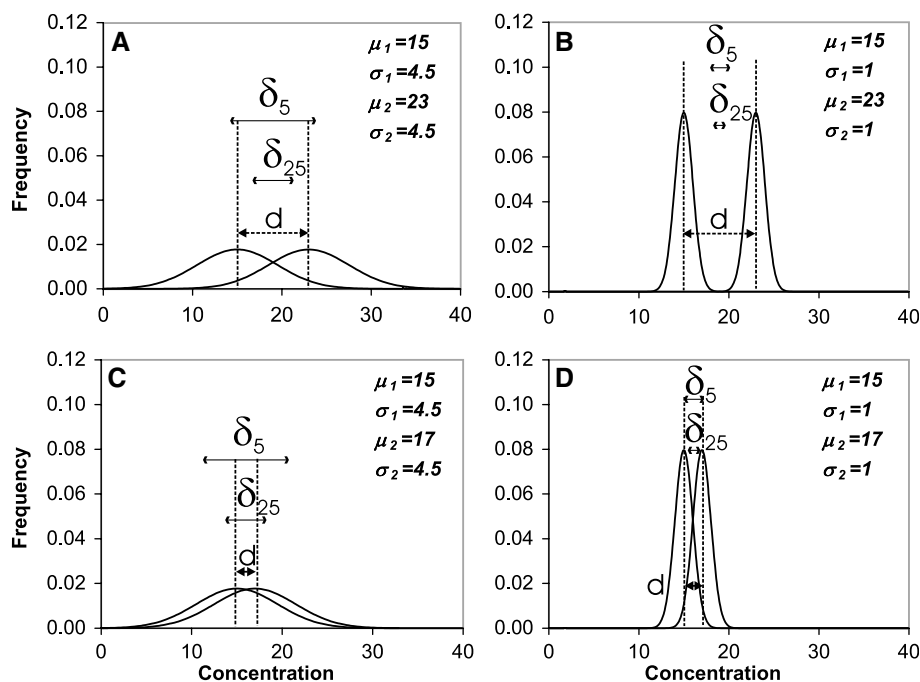
Fig. 1. Diagram showing the four comparisons made between two populations that differ in terms of mean concentrations (μ_1 and μ_2) and/or deviations (σ_1 and σ_2). The size of the arrows indicates the minimum difference in concentration of δ, detectable (for $n = 5$ and $n = 25$) and d, the differences in the mean values corresponding to the two populations

If in Figure 1 we compare pairs of populations that have the same σ, but different values of d (Figure 1A and 1C; Figure 1B and 1D), we observe that the higher the value of d, the easier it is to detect the differences between them, because a lower value of n is required. For example, in Figure 1A and 1C $\delta_{25} = 4.12$; thus, in Figure 1A δ is less than d (4.12 < 8), which allows significant differences to be found for this value of n. In contrast, in Figure 1C, δ is greater than d (4.12 > 2) and no significant differences are found. Thus, if we wish to detect significant differences between means that are close in value, d (Figure 1C or 1D), then we require larger n than if we wanted to detect only large differences (Figure 1A or 1B).

If we compare populations with identical d and different σ (Figure 1A and 1B; Figure 1C and 1D), we observe that the smaller the value of σ, the lower the value of n required to detect significant differences (e.g., Figure 1A and 1B $d = 8$, as well as in Figure 1A $\delta_5 = 9.12$ and in Figure 1B $\delta_5 = 2.06$, so that $\delta > d$ only in Figure 1A in which no significant differences will be found). If the variability of the samples is large, then a larger sample size is required to achieve a given ability to detect differences between means.

It must be taken into account that the value of δ depends on the size of μ_1 and μ_2, and not only the value of the differences between the population means. For example, in Figure 1D, where $\mu_1 = 15$ and $\mu_2 = 17$, the value of δ when $n = 25$ is 0.92, whereas if 1000 units are added to the means, thereby proportionally increasing the variances while maintaining the coefficient of variation, the value of δ for the same n would be 58.67.

Power and Sample Size in Testing for Difference Between Two Means in Non-Normal Populations. As indicated in the heading of the previous section, the previous calculation is valid only for data from populations with an underlying normal distribution. However, in nature the distribution of contaminant concentrations is not usually normal (Olsson and Biegnert 1997), but rather strongly skewed to the right, because negative concentrations do not exist. A nonparametric test must therefore be applied to compare the data. The calculation of the power of the test is based on characterization of the distributions of the two populations being compared, which if they are normal will be defined by their means and standard deviations. However, if the distribution is not normal, it is not possible to characterize the population only with these statistics, which creates an insolvable *n*-infinite dimensional problem. It is thus impossible to design a nonparametric test, and the data must be transformed to allow application of a parametric test. For this, we propose a new test based on the application of a transformation that allows the distribution to be normalized. Thus, from the mean and the standard deviation of the normalized distribution, we obtain a set of data of size *n*, belonging to a random normal distribution, which can later be back-transformed. Finally, the statistic *U* will be used to find whether the two sets of data obtained comply with the null hypothesis $\mu_{X1} = \mu_{X2}$.

Data Transformation

From two populations of non-normally distributed data, with means μ_{X1} and μ_{X2} and variances σ^2_{X1} and σ^2_{X2}, and a difference between the means of $d = \mu_{X2} - \mu_{X1}$, we can obtain the corresponding random normally distributed populations with means μ_{Y1} and μ_{Y2} and variances σ^2_{Y1} and σ^2_{Y2}, using any of the procedures described as follows.

We used a Box-Cox transformation (Legendre and Legendre 1998) to normalize the data by calculation of λ, which minimizes the associated variance. For this we used the Minitab 14 statistical package, which, as well as the best fit of λ, provides a range of adequate λ values. The best-fit values of λ for the concentrations of a given contaminant did not coincide for the three SS sampled (*e.g.,* for Ni the best-fit values of λ were −0.5, −0.5, −1). However, for each element the range of suitable values of λ overlapped; therefore, we chose a value within the range that minimized each variance. Posteriorly, after applying the transformation chosen, we tested the normality of the data by means of Lilliefor's modification of the Kolmogorov-Smirnov test, using SPSS 11.5.

When the chosen transformation was $\lambda = 0$, *i.e.,* a logarithmic transformation, then $Y = \log X$ and we obtained a normal distribution with mean μ_Y and variance σ^2_Y, and thus the mean and variance of X are calculated as:

$$\mu_X = 10^{\left(\mu_Y + \frac{\sigma_Y^2}{2}\right)} \qquad (3)$$

and

$$\sigma_X^2 = 10^{\left(2\mu_Y + \sigma_Y^2\right)} * \left(10^{\sigma_Y^2} - 1\right) \qquad (4)$$

If we wish to obtain the mean (μ_Y) and variance (σ_Y^2) of the random normal variable Y on the basis of the mean and variance of the original random variable X (without transformation), we resolve the previous linear system (Eqs. 3 and 4) to obtain:

$$\sigma_Y^2 = 10^{\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)} \qquad (5)$$

and

$$\mu_Y = 10^{(\mu_X)} - \frac{\sigma_Y^2}{2} \qquad (6)$$

When the transformation chosen was $\lambda = 0.5$, i.e., a square root transformation, then $Y = X^{0.5}$, which corresponds to a normal distribution of mean μ_Y and variance σ_Y^2; thus the mean and variance of X are calculated as:

$$\mu_X = \mu_Y^2 + \sigma_Y^2 \qquad (7)$$

and

$$\sigma_X^2 = \left(2\mu_Y^2 + \sigma_Y^2\right) * 2\sigma_Y^2 \qquad (8)$$

As before, by resolving this system of equations (Eqs. 7 and 8) we can obtain the mean and variance of Y:

$$\mu_Y = \sqrt[4]{\mu_X^2 - \frac{1}{2}\sigma_X^2} \qquad (9)$$

and

$$\sigma_Y = \sqrt{\mu_X - \sqrt{\mu_X^2 - \frac{1}{2}\sigma_X^2}} \qquad (10)$$

For the transformations $\lambda = -0.5$ and $\lambda = -1$, explicit expressions for μ_Y and σ_Y^2 in terms of μ_X and σ_X^2 are not available, because $X^{-0.5}$ and X^{-1} are never normally distributed. Thus, for these two values of λ, we calculated μ_Y and σ_Y^2 as the mean and variance of the data transformed from the original variable ($y_i = x_i^{-0.5}$ or $y_i = x_i^{-1}$), rather than directly estimating μ_Y and σ_Y^2 from μ_X and σ_X^2, a solution which, although not strictly correct, was the only possible alternative.

The U Test

The transformations used allow us to obtain data corresponding to a normal population of mean μ_Y and variance σ_Y^2. Using probability values generated as random numbers, we can obtain data corresponding to the normally distributed populations with mean values μ_{Y1} and μ_{Y2} and variances σ_{Y1}^2 and σ_{Y2}^2. For each of the two populations, the same number of data ($y_i...y_n$) is generated as the value of the sample size (n) that we want to test. All of the data generated ($y_i...y_n$) are then conveniently back-transformed (e.g., if we apply a logarithmic transformation, then $x_i = 10^{y_i}$). From the back-transformed data ($x_i...x_n$) the means $\overline{X_1}$ and $\overline{X_2}$ and the standard deviations (σ_{X1}^2 and σ_{X2}^2) of both populations can then be calculated. Finally, we can calculate the statistic U defined as:

$$U = \frac{\overline{X_1} - \overline{X_2}}{\frac{\sigma_{X_1} + \sigma_{X_2}}{\sqrt{n}}} \qquad (11)$$

This series of calculations is repeated 500 times, thereby generating 500 different values of U for the sample size tested.

As with any statistical test, the values of the statistic obtained must be compared with the tabulated values. Therefore, the tabulated values of the statistic U were also calculated in a similar way, but complying with the null hypothesis of $\mu_{X1} = \mu_{X2}$. The values of μ_{X1} and μ_{X2} correspond to the mean of population 1 and the variances σ_{X1}^2 and σ_{X2}^2, to the variance that we consider to be representative of the process, and the calculus can be repeated in the range of variances obtained; calculation of the tabulated value of U was repeated 1000 times. Thus, a normal distribution of tabulated values of U was obtained and the percentile corresponding to a level of significance of $\alpha \leq 0.05$ calculated, which is the tabulated U for comparison.

The 500 calculated values of U were compared with the corresponding tabulated U values ($\alpha = 0.05$), and the null hypothesis (H_0), $\mu_{X1} = \mu_{X2}$, was rejected in those cases in which the calculated U was greater than the tabulated U. Finally, the percentage number of rejections was calculated and if >95% ($\alpha = 0.05$); this indicated the final rejection of the H_0 for this n. The values of n tested varied between a minimum of 2 and a maximum of 200; for operational reasons, it is impracticable to go beyond this number. From all of the values of n tested and for those for which the H_0 was rejected, the lowest value of n was chosen as the sample size.

Finally, we calculated the sample size required (by means of calculation of U) to differentiate between a pair of sites for which the value of d was 1, 2.5, or 7 times the regional background level (BL) of each element (unpublished data). This was carried out with respect to the type of distribution of each of the elements, and the mean of one of the populations being compared was always equal to BL (μ_1 =BL and μ_2 = 2*BL, 3.5*BL or 8*BL). The calculation was repeated using three coefficients of variation (25%, 35%, and 50%) of the element concentrations. In the case of Cu, Hg, and Ni it was not possible to calculate U because there are neither explicit transformations nor sets of original data corresponding to these elements on which to apply direct transformations. Despite this, we calculated the sample size for the three elements, using the method of Zar, and thus obtained a value, which although incorrect because normality is assumed, allowed us to discuss the results.

Prior to the analysis of the results, the correct functioning of the U test was evaluated by applying it to a normally distributed data set (without applying any transformation), and it was found that the results obtained were identical to those obtained on applying Zar's test.

Results

The descriptive statistics for the elements determined at the 3 SS studied are shown in Table 1. The concentrations of these elements are not normally distributed, and most fitted significantly to log x and $\sqrt{x}$ distributions, some to $1/\sqrt{x}$ distribution, and in one case to a $1/x$ distribution. In all cases, the data showed positive asymmetry (distributions skewed to the right).

As regards the variability of the results, the coefficients of variation ranged between 16% and 41% (most being around 25%) without any pattern being observed for elements or SS. The coefficients of variation did not follow any order in terms of the corresponding concentrations of elements.

The median values of the bioconcentrations of As, Cu, Hg, Ni, Pb, and Se obtained at SS1—in the surroundings of the FeSi smelter—were much higher than those corresponding to

Table 1. Descriptive statistics for the tissue concentrations (µg g^{-1}) of the elements measured in subsamples of the moss *Scleropodium purum* collected at the sampling sites (SS) under study

	Analytical CV	λ	SS1					SS2					SS3				
			Mean	Median	CV	Min	Max	Mean	Median	CV	Min	Max	Mean	Median	CV	Min	Max
As[a]	8	0	177.3	20.6	41	34.1	394.5	46.9	4.2	31	22.9	94.7	42.5	3.9	33	21.4	80.3
Cd[a]	8	0	115.9	4.2	31	64.1	169.9	90.1	3.9	33	49.2	253.9	42.5	6.5	20	28.3	74.1
Cu	6	−0.5	23.4	3.9	33	16.3	43.7	3.1	6.5	20	2.3	4.9	3.9	9.2	37	2.6	5.3
Hg[a]	7	−1	33.8	6.5	20	12.2	86.7	17.5	9.2	37	10.6	29.6	17.9	2.8	24	12.0	45.8
K	4	0	4368	9.2	37	2618	6357	4924	2.8	24	3731	6739	6421	1.5	24	4513	9247
Ni	7	−0.5	7.34	2.8	24	3.64	13.85	0.73	1.5	24	0.44	2.16	0.71	0.1	16	0.40	1.51
Pb	6	0.5	3.94	1.5	24	0.94	6.23	1.79	0.1	16	0.80	3.24	1.08	0.2	20	0.30	2.32
Se[a]	10	0.5	287.0	0.1	16	170.3	409.0	87.2	0.2	20	43.8	159.6	94.5	3.7	39	51.7	146.4
Zn	6	0.5	45.6	0.2	20	32.7	60.4	46.8	3.7	39	28.1	64.2	35.1	1.0	20	24.2	50.6

Notes: The analytical error (expressed as the coefficient of variation, CV, as a percentage) and the optimum value of λ that allows normalization of the data set are included.

CV = coefficient of variation as a percentage; Min = minimum; Max = maximum.

[a] ng g^{-1}

the other two SS, for which similar values were obtained. The concentrations corresponding to SS1 were between 2 and 10 times higher than those corresponding to the other sites, depending on the element considered (Table 1). However, there was a clear gradient of concentration of Cd among the three SS (SS1 > SS2 > SS3). The concentrations of K and Zn at the three sites were similar.

Figure 2 shows diagrams of the distribution of the concentration of As in the pairs of sites SS1-SS2 (Figure 2A) and SS2-SS3 (Figure 2B), as an example of what occurs with the distributions of most of the elements. As explained in Materials and Methods, the different sample sizes will be determined by the value *d*, because the dispersion (expressed as coefficients of variations) of the distributions of the concentrations of a given element are similar in all the SS. Figure 2A shows a case in which the two distributions are clearly differentiated (high value of *d*), whereas in Figure 2B the distributions shown overlap (low value of *d*). A larger number of samples would therefore be required to differentiate the SS in Figure 2B than to differentiate those in Figure 2A.

Table 2 shows the number of samples required to be collected to differentiate ($p \leq 0.05$) the SS compared, both when applying the *U* statistic (Eq. 11) and when applying the test proposed by Zar (1984) for normal populations (Eq. 1). In general terms, the values obtained by both methods were the same or very similar, and notable differences were appreciated only in the cases of Cd and K (both with λ = 0) when comparing SS1 and SS2. In all cases in which explicit transformations were carried out, the results were consistent with those obtained using only the method proposed by Zar, as occurs when there are not explicit transformations for λ = −0.5. However, when the transformations for λ = −1 do not exist, as with Hg, the results obtained with the *U* test were totally incoherent, indicating that the test is sensitive to the type of transformation used.

If we take the value of *n* (from Table 2) corresponding to the example for As shown in Figure 2, we see that by collecting six subsamples from SS1 and SS2 we can differentiate ($p \leq 0.05$) between the corresponding mean concentrations (Figure 2A). However, if we want to differentiate between SS2 and SS3 (Figure 2B), the number of subsamples required
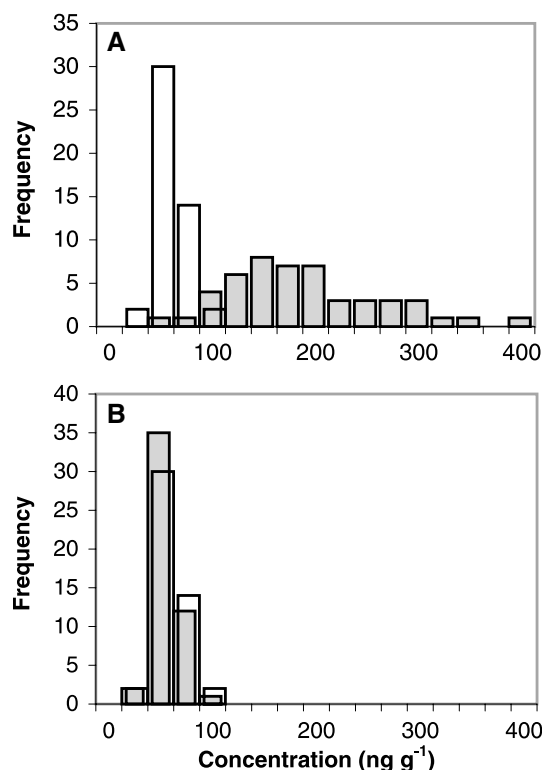

Fig. 2. Comparison of the distributions of the tissue concentrations of As (ng g^{-1}) in the moss *Scleropodium purum* at sampling sites 1 and 2 (a) and at sampling sites 2 and 3 (b)

would be more than 200. Likewise, for the remaining elements studied, when we compared pairs of SS with very different means, the number of subsamples ranged between 2 and 11, whereas if the mean values were similar, as in the previous example, the *n* required would be more than 200, without the exact number required being known.

The numbers of samples required to differentiate between two populations, one of which has a mean concentration equal to the BL and the other of which has a mean equal to 2*BL,

Table 2. The sample sizes required to differentiate ($p \leq 0.05$) between the mean tissue concentrations of elements in moss (*Scleropodium purum*) growing at the sampling sites (SS) compared, calculated by applying either the U test or the test proposed by Zar for normally distributed populations

Method of calculation	Pairs of SS compared	As	Cd	Cu	Hg	K	Ni	Pb	Se	Zn
Zar	SS1-SS2	4	20	2	9	20	2	5	2	>200
	SS1-SS3	4	2	2	8	2	2	3	2	9
	SS2-SS3	>200	6	20	>200	6	20	21	>200	17
U test	SS1-SS2	6	102	2	n. c.	102	2	5	2	>200
	SS1-SS3	5	2	2	n. c.	2	2	2	2	11
	SS2-SS3	>200	10	13	n. c.	10	13	11	>200	12

n.c. = not calculated.

3.5*BL, or 8*BL, are shown in Table 3. The variance of each population was determined by three selected coefficients of variation (25%, 35%, and 50%). In the cases of Cu, Hg, and Ni, the value obtained is merely a guideline (see The U Test section above) and underestimates the real value because a normal distribution was assumed.

Discussion

The variability in the concentrations of elements under study (Table 1) did not differ greatly from the results obtained by Cenci (1999), between 7% and 32%; Fernández *et al.* (2002), between 21% and 54%, or those of Rühling (1994b), who reported between 10% and 20% of local uncertainties. Comparison of results from different studies should be carried out with caution because of the use of different species, and also because of the way in which the variability was studied in each case. Cenci (1999) investigated the variability by collecting nine subsamples corresponding to the nine equal parts into which a sampling area of 1 m² was divided. Fernández *et al.* (2002) collected 50 isolated moss clumps, separated by more than 0.45 m from other clumps, in an area of 35 × 35 m. Finally, Rühling (1994b) did not specify how the variability was studied, and only indicated that the data referred to plots of 50 × 50 m.

The variability considered in the previous paragraph includes various sources: local, sampling, processing, and analytical. Assuming that these sources of variability are due to independent processes, we can assume that the variances are additive, so that in the present study, the contribution of the analytical variance (see Chemical Analysis in Materials and Methods) to the total variance is, in general terms, less than 30%. Most of the existing variability corresponds to sampling and local variability, the latter being generated by processes of deposition and intrapopulation differences, whereas the variability associated with processing is almost negligible, because any possible error would be systematic and would be included in all samples to the same extent. The magnitude of the variability in the concentrations on a local scale is very high compared with the variability in the concentrations of most of the elements on a regional scale in Galicia (Buse *et al.* 2003). Thus, for example, in Zn the regional variability (42% as coefficient of variation) was very similar to the local variability (20%–39%, Table 1). However, for other elements, such as As and Ni, the variability differed by an order of magnitude. This is consistent with the results of Sloof and Wolterbeek (1991), who found that in lichens, for all of the

metals determined, there was an order of magnitude of difference between local and site variations.

The variability in the concentrations of an element and the difference in the mean concentrations corresponding to two SS will determine the number of samples required to differentiate between them. Thus, when the objective is differentiation between concentrations of elements at SS, studies of the underlying variability in different conditions and for the species selected must be carried out in the area in which the air quality is to be monitored. This allows establishment of the maximum variability (expressed as the coefficient of variation) associated with each of the elements selected—which is independent of the mean concentration of the element, as occurs in lichens (Sloof and Wolterbeek 1991)—and from these data (adding 10% to give a certain level of security) to estimate the sample size that ensures differentiation between SS. However, it is possible that on occasion collection and/or processing of the number of samples required would be impracticable, and the mean values corresponding to the SS being compared would not be distinguishable. Comparison of data does not necessarily imply comparison of two SS separated in space, but also the same SS at two different times, *e.g.,* in studies of the temporal changes in concentrations of elements. In this case, the power to identify significant differences, however small, becomes more obvious.

One fundamental objective of biomonitoring studies with mosses is to be able to differentiate contaminated SS from other sites with element concentrations that are similar to the corresponding BL. Using the terminology of Fernández and Carballeira (2001), a SS is slightly contaminated when the concentrations of elements in the mosses collected are twice those of the BL. From the results of the present study, if we want to differentiate SS that are not contaminated from others that are slightly contaminated in this region, and applying the level of security proposed (coefficients of variation of 50% for As, Cd, Cu, Hg, Se, and Zn, and of 35% for Ni and Pb), the numbers of samples required are those shown in Table 3. Because these values differ depending on the element under consideration, and given that in biomonitoring studies various elements are usually determined in a single sample, the largest number of samples established should be collected, *i.e., n ≈* 30. If differentiation between uncontaminated and contaminated sites is required, to a greater or lesser degree (>3.5*BL = moderate contamination; >8*BL = serious contamination), the number of samples required decreases as the levels of contamination increases (Table 3). However, there may be difficulty in differentiating two uncontaminated sites (with concentrations of elements less than twice the BL), *e.g.,*

Table 3. Number of subsamples required to differentiate ($p \leq 0.05$) two sampling sites, in one of which the mean tissue concentration of an element ($\mu g\ g^{-1}$) in moss (*Scleropodium purum*) is equal to the background level (BL), and in the other in which the mean concentration is equal to 2*BL, 3.5*BL, or 8*BL

		2*BL			3.5*BL			8*BL		
	BL	25%	35%	50%	25%	35%	50%	25%	35%	50%
As	130[a]	8	16	27	2	6	10	2	2	5
Cd	40[a]	8	13	29	2	6	9	2	2	5
Cu	4.2	2	7	12	2	3	6	2	2	4
Hg	50[a]	2	7	12	2	3	6	2	2	4
Ni	2.4	2	7	12	2	3	6	2	2	4
Pb	1.3	4	7	15	2	2	6	2	2	2
Se	70[a]	5	8	13	2	2	5	2	2	2
Zn	25	4	7	13	2	2	5	2	2	2

Note: The number of subsamples required, calculated using three different coefficients of variation (25%, 35% and 50%) for the concentrations of elements, is shown.
[a] $ng\ g^{-1}$.

to differentiate levels of As, Hg, and Se in SS2 and SS3 (Table 2), more than 200 samples would be required.

In previous studies (Rühling 1989, 1994a, 1994b; Rühling and Steinnes 1998; Bargagli 1998), the authors proposed the collection of between 5 and 10 subsamples from a SS of 50 × 50 m. These recommendations are followed in Pan-European surveys. According to the results obtained in the present study (Table 2), if these recommendations had been followed (10 subsamples), we would not have been able to differentiate 33% of the cases compared, in which the required *n* was greater than 10. However, if the recommendation given in the previous paragraph had been followed (collection of 30 subsamples), the number of cases that would not be differentiated would be reduced to 18%. Likewise, using only 10 subsamples we would be able to differentiate a SS with BL concentrations from another that had more than 3.5 times this level (CV, 50%; Table 3). Thus, the standard recommendation (5–10 subsamples) does not allow differentiation between uncontaminated SS and slightly contaminated SS.

At this point, it must be taken into consideration that the results obtained (sample size required to differentiate SS) are strictly only applicable to studies carried out in Galicia, using *S. purum* sampled at random, and applying the proposed definition of a subsample (see Materials and Methods). However, the methodology proposed in the present study is totally valid and recommendable for studies carried out in other regions, under different settings, with different contaminants and other moss species. The generalized use of a single definition of a subsample in different studies provides a basis for posterior comparison of data on local variability, as well as interpretation of the data.

The local variability associated with a SS will be the sum of the intersubsample and intra-subsample variability (comprising the interindividual variability—between moss shoots—and the variability associated with the deposition processes at distances less than the extension of the subsample). The definition of a subsample delimits the distribution of variability (the greater the extension of a subsample, the greater the variability absorbed at the intra-subsample level, and thus the lower the intra-subsample variability and vice versa), thereby affecting the number of samples.

On the basis of a series of operational and scientific aspects explained below, we propose the following definition for a subsample: "1 g (approx.) dry weight, collected within a 25-cm radius around a node, selected at random from within a 1 × 1-m sampling grid in a sampling site." On some occasions it may not be possible to use a regular sampling grid and to select points at random, in which case scattered subsamples should be collected from the entire SS, each separated by 1 m (avoiding an aggregated distribution).

The basic reason for selecting 25 cm around a node is the homogeneity of the subsample, because the more homogeneous the sample, the easier to capture the variability. From a previous study (Fernández *et al.* 2002), we know that for some elements the variability at a level of less than 45 cm is very low compared with the total variability associated with the SS. The homogeneity will depend on the type of reproduction of the moss species used and on the deposition process. If the species selected reproduces mainly by asexual means (the adjoining gametophores originate from a single protonema), the variability will be less than in species that reproduce mainly by sexual means. Most of the moss species used in biomonitoring studies are epigeous and form mats comprising different clones, so that the variability between adjoining moss shoots is probably low. As regards the deposition processes, and taking into account how they are produced, we do not expect that the variability would be affected within such short distances.

In our opinion, the collection of 1 g d.w. of moss, which is the equivalent of approximately 100 apices of *S. purum* (of 3–4 cm length), is more than sufficient to represent the area sampled (0.2 m^2 = π*0.25^2). Even if moss species that are morphologically different from *S. purum*, or the entire green part of the moss shoots are used, we consider that 1 g d.w. meets the criterion of representativeness. In the present study, in which all the moss present within an area of 0.2 m^2, the corresponding dry weight varied between 0.5 and 5.8 g (median, 1.3 g), which shows that the criterion proposed can be met in practice, without any difficulty. Furthermore, in studies of local variability, 1 g d.w. of material is sufficient to allow individual analysis of each of the subsamples.

On combining all of the subsamples collected to form a composite sample, it is of vital importance that the weight of each be as similar as possible, because differences in the weight of the subsamples would produce an error in the estimation of the mean (the concentration of an element in the

composite sample is the mean concentration of the subsamples, weighted by their weights) (Fernández *et al.* 2002).

The number of subsamples combined determines the weight of the composite sample, which must be sufficient to allow all of the analyses required, with a portion left over for posterior storage. As regards the analyses, if we require 10–15 g d.w. of moss for analysis of dioxins and furans (Abad *et al.* 2003), 5 g d.w. for PAHs (Gerdol *et al.* 2002), and approximately 0.5 g for the analysis of metals and metalloids, at least 20 g d.w. of material is required for a single analytical determination, if no repetition of any is required. Even when only a low weight of composite sample is required for a specific analytical method (*e.g.,* 0.5 g d.w. for the analysis of heavy metals), the number of subsamples to be combined should not be less than the sample size estimated to be necessary, either for differentiation between SS, as in the present study, or to satisfy the assumed error of the mean (Fernández *et al.* 2002). Given that terrestrial mosses are routinely stored in Environmental Specimen Banks, in Finland, Denmark, Sweden, Iceland, Norway, Japan, the United States, and Spain (Galicia) (Aboal *et al.* 2001), the final weight of the sample must also include the amount of moss destined for storage. As an example, the number of samples determined in the present study to be necessary to calculate the mean concentration of elements at a SS, with an associated error of 20%, was 30 (Fernández *et al.* 2002); this involves obtaining a sample of 30 g d.w. of moss, equivalent to a volume of 4 L, which is only twice the amount previously recommended (Rühling 1994a) and which is sufficient for analytical needs and storage purposes.

Conclusions

The high local variability existing at a SS is not explained by analytical variability but rather by interindividual differences and deposition processes. The magnitude of this variability is very high, on the same order as the variability of some elements among SS, on a regional scale.

The high value of the variability and the lack of studies investigating this have led us to reflect on the validity of the indiscriminate application of the sample size proposed by Rühling (1994b): between 5 and 10 subsamples used to form a composite sample for each SS. In the present study, we have demonstrated that for some elements, more than 30 subsamples per SS are required to differentiate uncontaminated SS from other slightly contaminated sites.

It is therefore clear that pilot studies of the variability in concentrations of elements are necessary before selecting the required number of samples. To determine the number, the methodology proposed here or that proposed in a previous study (Fernández *et al.* 2002) will be useful, depending on the objectives. Furthermore, in light of these studies, we now know the power of the moss technique to differentiate between SS with similar levels of contamination.

With the aim of being able to compare the results obtained in diverse studies, a standardized definition of a subsample is required. For the reasons explained in the Discussion section, we propose the following definition: ''1 g dry weight of moss (approx.), collected within a 25-cm radius of a node, selected at random from a 1 × 1-m sampling grid placed in a sampling site.''

References

Abad E, Caixach J, Rivera J, Real C, Aboal J, Fernández A, Carballeira A (2003) Study on the use of mosses as biomonitors to evaluate the environmental impact of PCDDs/PCDFs from combustion processes—preliminary results. Organohalogen Compounds 60:283–286

Aboal JR, Fernández JA, Carballeira A (2001) Sampling optimization, at station scale, in moss, pine and oak contamination monitoring. Environ Pollut 115:313–316

Aboal JR, Fernández JA, Carballeira A (2004) Oak leaves and pine needles as biomonitors of airborne trace elements pollution. Environ Exp Bot 51:215–225

Bargagli R (1998) Trace elements in terrestrial plants. Springer Verlag, Berlin

Buse A, Norris D, Harmens H, Büker P, Ashenden T, Mills G (2003) Heavy metals in European mosses: 2000/2001 survey. UNECE ICP Vegetation

Cenci RM (1999) L'utilizzo di muschi indigeni e trapiantati per valutare in micro e macro aree le reicadute al suolo di elementi in tracce: proposte metodologiche. Proceedings of Workshop ''Biomonitoraggio della qualitá dell'aria sul territorio nazionale.'' Agenzia Nazionale per la Protezione Ambientale, Rome, pp 241–263

Cochran WG, Cox GM (1957) Experimental designs. 2nd ed. John Wiley, New York

Couto JA (2002) Implicaciones metodológicas en la biomonitorización activa y pasiva de la calidad del aire mediante musgos terrestres. Ph.D. Thesis, Universidad de Santiago de Compostela

Couto JA, Fernández JA, Aboal JR, Carballeira A (2003) Annual variability in heavy metal bioconcentration in moss. Sampling protocol optimization. Atmos Environ 37:3517–3527

Fernández JA, Carballeira A (2001) Evaluation of contamination, by different elements, in terrestrial mosses. Arch Environ Contam Toxicol 40:461–468

Fernández JA, Aboal JR, Couto JA, Carballeira A (2002) Sampling optimization at the sampling-site scale for monitoring atmospheric deposition using moss chemistry. Atmos Environ 36:1163–1172

Fernández JA, Aboal JR, Couto JA, Carballeira A (2004) Moss bioconcentration of trace elements around a FeSi smelter: modelling and cellular distribution. Atmos Environ 38:4319–4329

Gerdol R, Bragazza L, Marchesini R, Medici A, Pedrini P, Benedetti S, Bovolenta A, Coppi S (2002) Use of moss (*Tortula ruralis* Hedw.) as biomonitor of organic and inorganic air pollution in urban and rural areas in N Italy. Atmos Environ 36:4069–4075

Gy PM (1982) Sampling of particulate materials. Elsevier, Amsterdam

Keith LH (ed.) (1988) Principles of environmental sampling. ACS professional reference book. American Chemical Society, Washington, D.C.

Keith LH (1991) Environmental sampling and analysis, a practical guide. Lewis, Chelsea, MI

Legendre P, Legendre L (1998) Numerical ecology. Elsevier, Amsterdam

Markert B (ed.) (1994) Environmental sampling of trace analysis. VCH Publisher, Weinheim

Markert B (1996) Instrumental element and multielement analysis of plant samples—methods and applications. Wiley, Chichester

Markert B, Breure T, Zechmeister H (eds.) (2003) Bioindicators and biomonitors, principles, concepts and applications. Elsevier, Amsterdam

Olsson M, Biegnert A (1997) Specimen banking—a planning in advance. Chemosphere 34:1961–1974

Quevauviller P (ed) (1995) Quality assurance in environmental monitoring/sampling and sample preparation. VCH Publisher, Weinheim

Rühling Å (1989) Survey of the heavy-metal deposition in Europe using bryophytes as bioindicators. Proposal for an international programme. IVL/Dept. of Plant Ecology, University of Lund (mimeo)

Rühling Å (1994a) Monitoring of atmospheric heavy-metal deposition in Europe using bryophytes and humus samples as indicators. Proposal for international programme 1995. Dept. of Plant Ecology, University of Lund (mimeo 1994-11-21)

Rühling Å (ed.) (1994b) Atmospheric heavy metal deposition in Europe—estimation based on moss analysis. NORD 1994; 1994:9, Copenhagen

Rühling Å, Tyler G (1968) An ecological approach to the lead problem. Botaniska Notiser 122:321–342

Rühling Å, Steinnes E (eds.) (1998) Atmospheric heavy metal deposition in Europe 1995–1996. NORD, 1998:15, Copenhagen

Sansoni B, Iyengar V (1978) Sampling and sample preparation methods of trace elements in biological materials. Forschungszentrum Jlich, Jl Spez., 13, Jlich, Germany

Sloof JE, Wolterbeek HT (1991) National trace-element air pollution monitoring survey using epiphytic lichens. Lichenologist 23:139–165

Zar JH (1984) Biostatistical analysis. Prentice-Hall International Editions, London