The Holocene 15,2 (2005) pp. 293–297

Minimum count sums for charcoal-concentration estimates in pollen slides: accuracy and potential errors

Walter Finsinger* and Willy Tinner

(Institute of Plant Sciences, Universität Bern, Altenbergrain 21, CH-3013 Bern, Switzerland)

Received 19 November 2003; revised manuscript accepted 27 February 2004

A
HOLOCENE
RESEARCH
REPORT

Abstract: Charcoal particles in pollen slides are often abundant, and thus analysts are faced with the problem of setting the minimum counting sum as small as possible in order to save time. We analysed the reliability of charcoal-concentration estimates based on different counting sums, using simulated low- to high-count samples. Bootstrap simulations indicate that the variability of inferred charcoal concentrations increases progressively with decreasing sums. Below 200 items (i.e., the sum of charcoal particles and exotic marker grains), reconstructed fire incidence is either too high or too low. Statistical comparisons show that the means of bootstrap simulations stabilize after 200 counts. Moreover, a count of 200–300 items is sufficient to produce a charcoal-concentration estimate with less than ±5% error if compared with high-count samples of 1000 items for charcoal/marker grain ratios 0.1–0.91. If, however, this ratio is extremely high or low (> 0.91 or < 0.1) and if such samples are frequent, we suggest that marker grains are reduced or added prior to new sample processing.

Key words: Charcoal analysis, palaeoecology, random sampling, sample size, error estimation.

Introduction

Charcoal analysis is widely used to reconstruct past forest-fire occurrence and to study the effects of fire on ecosystems (e.g., Iversen, 1941; Odgaard, 1992; Whitlock and Larsen, 2001). Sedimentary charcoal records have also been used to assess whether carbon dioxide released by forest fires affects the global atmospheric CO_2 concentration (Carcaillet *et al.*, 2002). Such large-scale comparisons rely on data sets gathered by different analysts using different preparation and counting methods, hampering the comparability of results. At present, various preparation methods are used to estimate charcoal abundance in sediments (e.g., pollen slides, thin sections, sieving, combustion/digestion), resulting in different spatial and temporal resolutions of this fire proxy (e.g., MacDonald *et al.*, 1991; Carcaillet *et al.*, 2001). Nevertheless, analysis of charcoal content in pollen slides is one of the most commonly used methods, probably also because pollen slides are widely available.

Comparison of recent fire activity deduced from fire scars or written historical sources with charcoal records from lake sediments show that pollen-slide charcoal represents regional source areas (MacDonald *et al.*, 1991; Tinner *et al.*, 1998). These empirical results confirm previous theoretical modelling of charcoal-particle transport (Clark, 1988a). It has often been assumed that large charcoal particles derive from a local source (i.e., within the catchment) while smaller ones are transported longer distances (e.g., Patterson *et al.*, 1987). These assumptions led many charcoal analysts to tediously estimate in a time-consuming manner the charcoal area of particles in pollen slides (e.g., Waddington, 1969; Swain, 1973). However, a recent comparison between particle-area concentration (mm^2/cm^3) and particle-number concentration ($no./cm^3$) for different vegetation types showed that it is unnecessary to measure charcoal areas in standard pollen slides (Tinner and Hu, 2003).

An important problem of charcoal analysis, however, is not yet solved: charcoal analysts using the pollen-slide technique have to decide about the minimum counting sum needed to produce accurate results in pollen slides. In this respect, different authors have made different choices: for instance, Carcaillet *et al.* (2001) and Carrión *et al.* (2003) scanned the total surface of each pollen slide, while Odgaard (1992) analysed for each sample at least one entire slide, or, in cases of abundant charcoal particles, half a slide. In contrast, Tinner and Conedera (1995) chose a minimum counting sum of at least 200 charcoal particles for each sample, while others (e.g., MacDonald *et al.*, 1991; Sarmaja-Korjonen, 1998; Tinner *et al.*, 1999; Tinner *et al.*, 2000; Darbyshire *et al.*, 2003) measured or counted at least 100 individual fragments. On the

---

*Author for correspondence (e-mail: walter.finsinger@ips.unibe.ch)

10.1191/0959683605hl808rr

other hand, some authors do not reveal the counting sum they used for charcoal analysis.

Previous authors (e.g., Mosimann, 1965; Maher, 1972) have already addressed the effect of count sums on the accuracy of percentage and concentration data in palaeoecology. They used the negative binomial distribution to calculate confidence intervals for particle concentrations, which is subsequently used to estimate the data reliability for counts outside the sum. Nomograms presented by Maher (1972) are useful to calculate confidence limits for different ratios of charcoal particles and *Lycopodium* spores; thus it is possible to evaluate the errors of counted samples (e.g., by plotting confidence intervals of charcoal counts). However, another question is when to stop charcoal counts after sufficient precision is reached. This question is important, because low-count sums as used by, for example MacDonald *et al.* (1991) and Darbyshire *et al.* (2003) may correspond on an average to about 10–20% of a pollen-slide surface, allowing considerable time to be saved. Potential inaccuracies due to low-count sums can be addressed by comparison with high-count sums. Here we measure and evaluate the means and relative errors for different sums and define a minimum counting limit (beyond which additional counting may not be warranted).

Material

We extracted a sediment core (AVG 07/02-2A) from the Lago Grande di Avigliana, northwestern Italy (350 m a.s.l.; 45°04.252′, EO 07°23.131′), with a freeze-corer containing dry ice and alcohol (Wright, 1991) and selected five subsamples (1 cm^3 each) from it. After adding *Lycopodium* tablets for the estimation of CharConc (Stockmarr, 1971), the sediment samples were treated for pollen analysis chemically (HCL, KOH, HF, HCL, acetolysis and KOH) and physically (500 μm sieving and decanting). Charcoal particles $> 75\ \mu m^2$ (or c. 10 μm long) were counted in these slides with a light microscope at 250 × magnification. Particles were selected following the criteria of Clark (1988b): black, completely opaque and angular.

Methods

In each sample, 1000 items (i.e., the sum of charcoal particles and *Lycopodium* spores) were counted, the counts being entered directly during the microscope analysis into an electronic spreadsheet as charcoal particle or *Lycopodium* grain. Simulation of low-count sums was made by random sampling with replacement. When sampling with replacement, the probability of each object to be drawn is proportional to its abundance and does not change during the simulation process; this procedure is analogous to the bootstrap algorithm (Efron and Tibshirani, 1993). As the estimated standard error is larger for sampling with replacement than for sampling without replacement, this method can be considered to give relatively robust results (Efron and Tibshirani, 1993). The same method was used recently by Heiri and Lotter (2001) for studying the effect of low counts on multivariate data, e.g., chironomid inferred temperatures.

For each sample, 18 count sums were simulated with sums of 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 300, 400 and 500 items. This procedure was repeated 100 times for each simulated count sum, thus obtaining 100 bootstrap replicates for each sum (Figure 1). Intuitively, the difference between such analyses will decrease the higher their count sum is.

Given the assumption of homogeneous distribution of particles on a pollen slide, the inferred particle abundance of low counts was compared with the expected abundance of the complete sample ($\mathrm{CharConc_{TOT}}$). We applied two tests to evaluate the precision of particle-concentration estimates with different count sums. First, we compared the mean values of the 100 bootstrap replicates with the expected particle concentration of the complete sample using a Z-test (two-tailed; $\alpha = 5\%$) in order to verify at which count sum they do not differ significantly (Table 1). We selected this parametric test because the test involves pair wise comparisons for random samples of whichever distribution for $n > 100$ (Riedwyl, 1992). Secondly, to check the hypothesis that the variance becomes smaller than a defined threshold with increasing count sums, a variance analysis involving an F-test was performed (Table 2). The null hypothesis is that the variance of the inferred low count is not different from the threshold variance.

Results and discussion

In our five samples (Figure 1), the ratio p of charcoal particles over marker grains ($p = x/y$) is on average, if compared to values obtained in other studies (e.g., Tinner *et al.*, 1999). In our test samples p is low for sample 30 cm (0.275), medium for samples 2 cm, 29 cm and 26 cm (0.367, 0.338 and 0.378, respectively); and high for sample 18 cm (0.625). In addition, two samples, ex1 ($p = 0.91$) and ex2 ($p = 0.1$), were created with extremely high and low ratios. In these seven samples, the means of inferred particle-number concentration for different count sums is similar (Figure 1). With increasing count sums the total and interquartile range decrease towards $\mathrm{CharConc_{TOT}}$, and in comparison to lower counts the sums of 300, 400 and 500 items seem to be the most similar in respect of the interquartile range. This result suggests that in general the means and distributions of sums > 100–200 items are similar to those of high counts with 1000 items (HCS).

This optical impression is sustained by results of statistical analysis. The comparison of means shows that, with the exception of sample ex1 (Table 1), all means of simulated charcoal concentrations of low counts do not differ from the mean of the HCS for counting sums ≥ 200 items (Table 1). The exceptional behaviour of sample ex1 is possibly related to the problem of binomial distribution sampling: when p takes extreme values, the distribution is asymmetric, and higher counts are needed to obtain an acceptable normality approximation. However, in this extreme case, means do not differ for counting sums ≥ 400 items. It is obvious that low-count means are tending to HCS with increasing n. Using the mean absolute percentage error ($\mathrm{MAPE_i}$; Jensen, 2003), it is possible to calculate and visualize a relative value for dissimilarity. The $\mathrm{MAPE_i}$ of the inferred concentration for low counts in relation to the expected concentration of the HCS is defined as follows (Figure 2):

$$MAPE_i = \left[\sum_{j=1}^{n} \left(\left| \frac{CharConc_{ji} - CharConc_{TOT}}{CharConc_{TOT}} \right| \right) \cdot 100 \right] 1/n \quad (1)$$

where $\mathrm{MAPE_i}$ is the mean absolute percentage error of a given sample for a count sum i (ranging from 10 to 500, see above), n is the number of bootstrap replicates, $\mathrm{CharConc_{ji}}$ ($j = 1,2 \ldots n$; $n = 100$) is one inferred particle concentration for the count sum i, and $\mathrm{CharConc_{TOT}}$ is the particle concentration of the complete sample. MAPE is a comparative measure that does

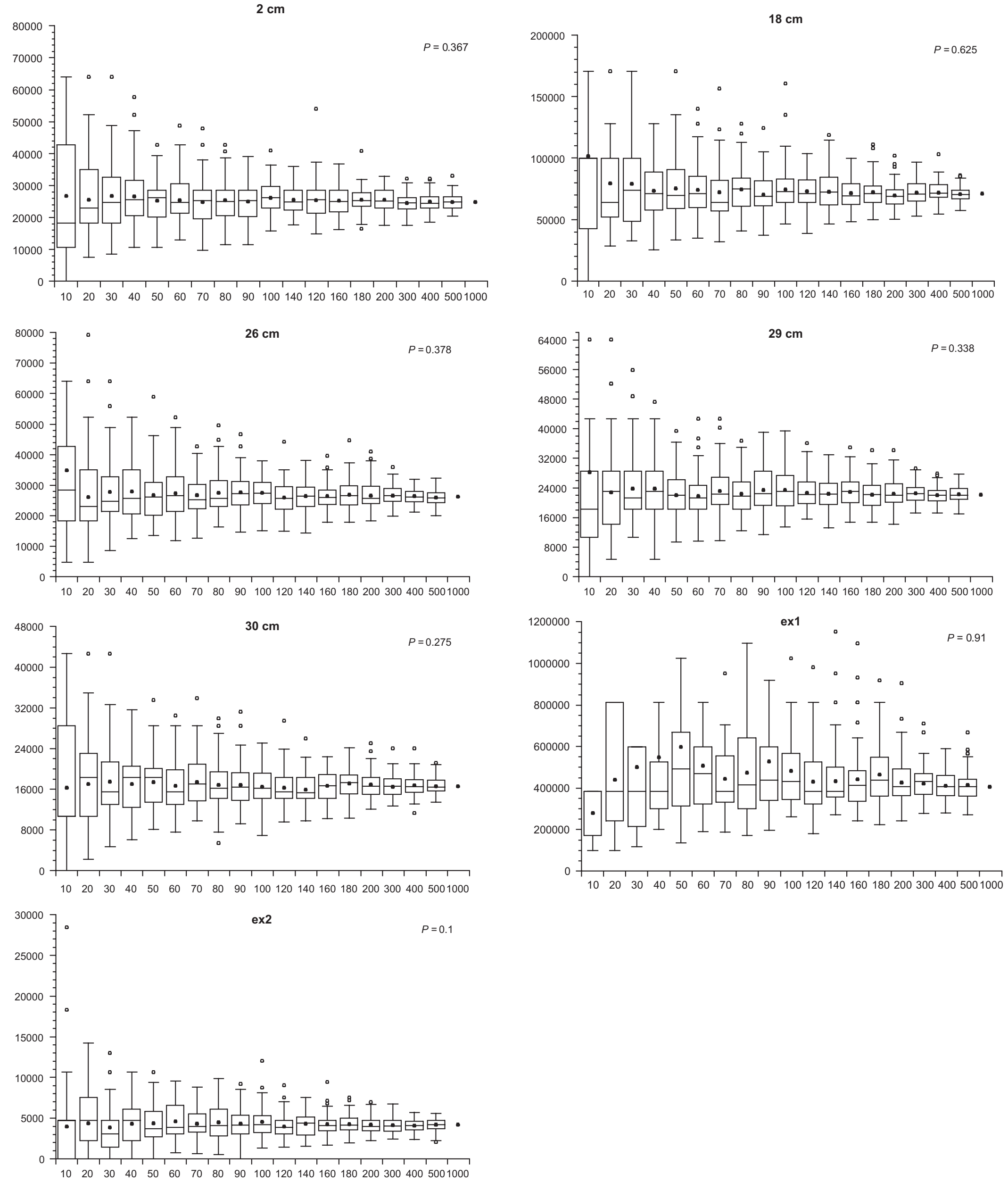

Figure 1 Distribution of inferred particle number concentration (number of particles cm^{-3}) for simulated low-count sums. Each box-plot displays the data of 100 bootstrap replicates (see text for details). The boxes enclose 50% of the data with the median value of the variable displayed as a line. Vertical lines display 1.5 the interquartile distance from the box. Open circles = points greater or smaller than 1.5 times the interquartile distance. Closed circles = mean. For low counts of samples ex1 and ex2, a number of data points are missing, given the low probability of sampling at least one *Lycopodium* grain (ex1) or charcoal particle (ex2).

Table 1 Comparison between the mean of inferred particle concentration for low counts and particle concentration of the complete sample (n = 1000 items)

Count sum (n)	Sample label						
	2 cm	18 cm	26 cm	29 cm	30 cm	ex1	ex2
100	2.613*	1.948	2.636*	2.349*	– 0.245	4.036*	1.87344
200	1.909	– 1.859	0.706	0.794	1.305	1.966*	– 0.39828
300	– 0.897	0.705	1.256	1.565	– 0.515	2.067*	– 0.81613
400	0.448	0.886	0.431	– 0.535	0.902	0.658	– 1.72881
500	0.015	– 1.291	– 1.073	0.842	0.276	1.255	– 0.08822

Z-test of equality of means: comparison between means of different counting sums (bootstrap sample mean) and an expected mean (CharConc$_{TOT}$). Null hypothesis for comparison: the means are not different, H$_0$: $y_1 = y_0$, two-tailed; confidence limit $\alpha = 5\%$; critical value for z is $z_{0.975} = 1.960$ (* = statistically different; $p < 0.05$).

Table 2 Comparison between the variance of inferred particle concentration for low counts and a threshold variance

Count sum (n)	Sample label						
	2 cm	18 cm	26 cm	29 cm	30 cm	ex1	ex2
100	0.6173	0.9823	0.5666	0.9575	0.7102	3.3753*	0.00026
120	0.8200	0.4823	0.5445	0.5798	0.6295	2.4434*	0.00018
140	0.4729	0.6275	0.4962	0.5567	0.5173	1.8164*	0.00017
160	0.4896	0.3961	0.3616	0.5736	0.5025	1.7308*	0.00014
180	0.3342	0.4181	0.4348	0.4705	0.4372	1.6152	0.00011
200	0.3221	0.2612	0.4510	0.4491	0.3859	1.0174	0.00010
300	0.2134	0.2917	0.2153	0.2242	0.2595	0.5383	0.00009
400	0.1943	0.2177	0.1158	0.1566	0.2220	0.4314	0.00004
500	0.1624	0.0915	0.1298	0.1602	0.1636	0.4906	0.00004

F-test of equality of variances: comparison of variances between different counting sums (s_n^2) and a variance (s_0^2) calculated by setting 2 standard deviations = 5% CharConc$_{TOT}$.
Null hypothesis for comparison: the standard deviations are not different, H_0: $s_n = s_0$, one-tailed; $\alpha = 0.05$, $\alpha_{\text{Bonferroni adjusted}} = 0.05/9$; critical value for F is thus $F_{0.95(99,99)} = 1.685$ (* = statistically different).

not have the problem of averaging positive and negative errors such as is the case with the mean error (Jensen, 2003). In theory, with increasing n and thus increasing similarity between CharConc$_{ji}$ and CharConc$_{TOT}$, MAPE$_i$ tends to zero (Figure 2). In our simulations, the mean error is highly unstable at low-count sums especially for samples ex1 and ex2. In these samples, only after a count sum of 200 items is the error lower than a value of 20%. In contrast, MAPE$_i$ of the other samples shows less variability and is lower than 20% already below a count sum of 100 items. In general the curves stabilize with counting sums >200–300 items.

Akin to the comparison of means, the F-test of equality of variances (Davis, 2002) shows that the variance of counts does not differ from an interval spanning ±5% of CharConc$_{TOT}$ for counts ≥ 180 items in all cases (Table 2). We have chosen 2 standard errors as the degree of variance to be compared with ±5% of CharConc$_{TOT}$ (Table 2). The rationale behind this choice is that in the case of a normal distribution (here n = 100) 95% of the expected mean values lie within the defined threshold interval. If the standard error of the mean for a simulated low-count sum is larger than the threshold interval, on average less than 95% of the low counts are within this interval. Thus, the associated error of the low-count sum is too high and further counting would be needed. It is possible to visualize this concept by plotting the ratio 4se:10% CharConc$_{TOT}$. In our simulations, values < 1 (indicating 4se $<$ 10% CharConc$_{TOT}$) are reached for counting sums 200–300 (Figure 3).

Conclusion

A question of charcoal analysis, which has not yet been addressed in detail, is how many charcoal particles must be counted in pollen slides to reach accurate estimates. The importance of this question is underscored by the wide and increasing use of charcoal analysis in palaeoecology. With bootstrap simulations and the F-test of equality of variances, we recommend that if the goal is to estimate the charcoal-particle concentration in pollen slides with an error

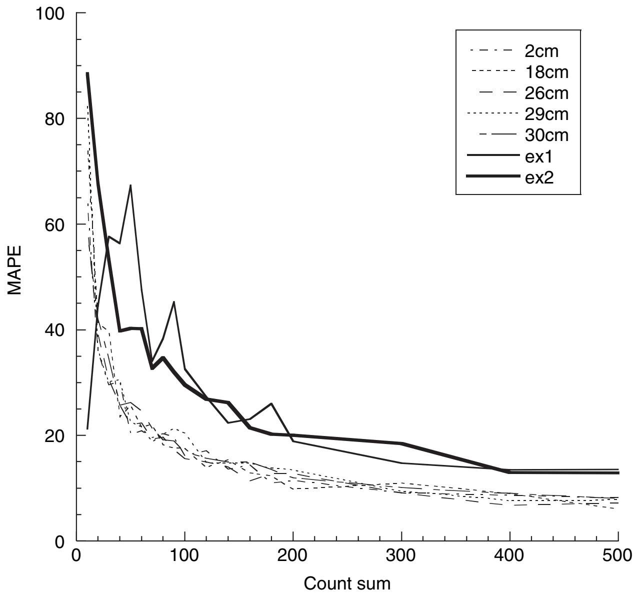

Figure 2 Mean absolute percentage error (MAPE$_i$) of the inferred low counts.

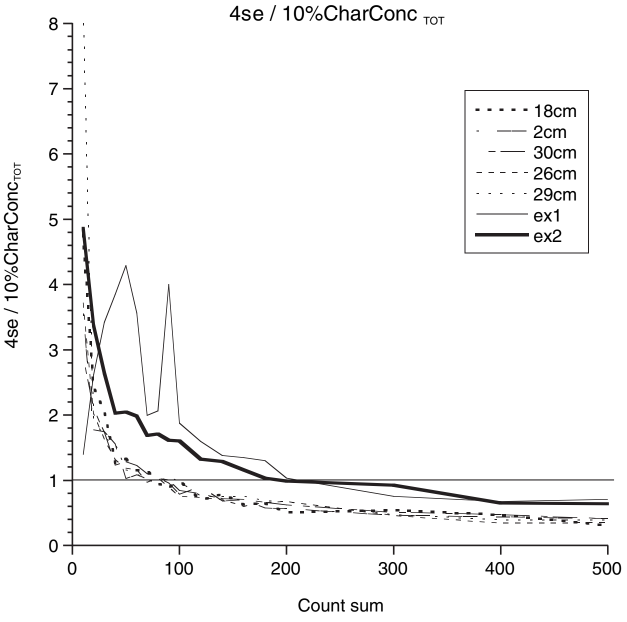

Figure 3 Standard error variation for different count sums in relationship to a threshold of ±5% CharConc$_{TOT}$ error (see text for details).

$\leq 5\%$ in regard to a high-count sum value (Figure 3), then a count of 200 items is sufficient. This finding is supported by quantitative comparisons of means that show that particle-concentration estimates inferred from count sums as low as 200–300 items are statistically not different from concentrations obtained with HCS of 1000. In addition, graphical comparisons indicate that, in most cases, the means as well as the standard error of bootstrap simulations stabilizes after 200 counts.

This implies that if a minimum count sum of 100 individual charcoal fragments is used (e.g. MacDonald *et al.*, 1991; Sarmaja-Korjonen, 1998; Tinner *et al.*, 1999; 2000; Darbyshire *et al.*, 2003), then a sum of at least 100 marker grains is needed. Conversely, minimum count sums of 100 charcoal fragments may be too high. For example, with a ratio p of 55 charcoal particles to 145 *Lycopodium* grains in the pollen slide ($p = 0.275$; sample 30 cm), the count is sufficiently precise. With extremely high or low p values exceeding our ex1 and ex2 simulations ($p = 0.91$ and 0.1, respectively), particle-concentration estimates are less precise and thus require higher count sums. However, such extraordinary ratios are extremely rare in charcoal records, such that they can still be interpreted as outstanding fire events (p high) or lack of fires (p low). In case samples such excessive p values are frequent, we suggest that marker grains are reduced or added prior to new sample processing. Therefore, after an initial charcoal analysis in a stratigraphic record, it may be advisable to re-valuate the number of marker grains to be added to the sediment sample in order to get moderate charcoal/marker ratios (i.e., between 0.1 and 0.9).

Acknowledgements

We are very thankful to Oliver Heiri and Christian Bigler for discussions and comments. We are also indebted to Daniel Dietrich for statistical advice, to Herb Wright for revising the manuscript, and to Frank Oldfield for the editing work. Cathy Whitlock and José S. Carrión made valuable comments.

References

Carcaillet, C., Almquist, H., Asnong, H., Bradshaw, R.H.W., Carrión, J.S., Gaillard, M.J., Gajewski, K., Haas, J.N., Haberle, S.G., Hadorn, P., Muller, S.D., Richard, P.J.H., Richoz, I., Rosch, M., Goñi, M.F.S., von Stedingk, H., Stevenson, A.C., Talon, B., Tardy, C., Tinner, W., Tryterud, E., Wick, L. and **Willis, K.J.** 2002: Holocene biomass burning and global dynamics of the carbon cycle. *Chemosphere* 49, 845–63.

Carcaillet, C., Bouvier, M., Fréchette, B., Larouche, A.C. and **Richard, P.J.H.** 2001: Comparison of pollen-slide and sieving methods in lacustrine charcoal analyses for local and regional fire history. *The Holocene* 11, 467–76.

Carrión, J.S., Sánchez-Gómez, P., Mota, J.F., Yll, R. and **Chaín, C**. 2003: Holocene vegetation dynamics, fire and grazing in the Sierra de Gador, southern Spain. *The Holocene* 13, 839–49.

Clark, J.S. 1988a: Particle motion and theory of charcoal analysis: source area, transport, deposition, and sampling. *Quaternary Research* 30, 67–80.

—— 1988b: Stratigraphic charcoal analysis on petrographic thin sections: application to fire history in northwestern Minnesota. *Quaternary Research* 30, 81–91.

Darbyshire, I., Lamb, H. and **Umer, M.** 2003: Forest clearance and regrowth in northern Ethiopia during the last 3000 years. *The Holocene* 13, 537–46.

Davis, J.C. 2002: *Statistics and data analysis in geology.* New York: Wiley.

Efron, B. and **Tibshirani, R.J.** 1993: *An introduction to the bootstrap.* New York: Chapman and Hall.

Heiri, O. and **Lotter, A.F.** 2001: Effect of low count sums on quantitative environmental reconstructions: an example using subfossil chironomids. *Journal of Paleolimnology* 26, 343–50.

Iversen, J. 1941: Land occupation in Denmark's Stone Age. *Danmarks Geologiske Undersøgelse II* 66, 1–68.

Jensen, A.N. 2003: Introduction to naive forecasting techniques. http://www.csus.edu/indiv/j/jensena/mgmt105/naie_01.ppt (last accessed 12 November 2003).

MacDonald, G.M., Larsen, C.P.S., Szeicz, J.M. and **Moser, K.A.** 1991: The reconstruction of boreal forest fire history from lake sediments-a comparison of charcoal, pollen, sedimentological, and geochemical indexes. *Quaternary Science Reviews* 10, 53–71.

Maher, L.J.J. 1972: Nomograms for counting 0.95 confidence limits of pollen data. *Review of Palaeobotany and Palynology* 13, 85–93.

Mosimann, J.E. 1965: Statistical methods for the pollen analyst: multinomial and negative multinomial techniques. In Kummel, B. and Raup, D., editors, *Handbook of paleontological techniques*, San Francisco: Freeman and Co., 636–73.

Odgaard, B.V. 1992: The fire history of Danish heathland areas as reflected by pollen and charred particles in lake sediments. *The Holocene* 2, 218–26.

Patterson, W.A., Edwards, K.J. and **Maguire, D.J.** 1987: Microscopic charcoal as a fossil indicator of fire. *Quaternary Science Reviews* 6, 3–23.

Riedwyl, H. 1992: *Angewandte Statistik*. Bern and Stuttgart: Haupt.

Sarmaja-Korjonen, K. 1998: Latitudinal differences in the influx of microscopic charred particles to lake sediments in Finland. *The Holocene* 8, 589–97.

Stockmarr, J. 1971: Tablets with spores used in absolute pollen analysis. *Pollen et Spores* 13, 615–21.

Swain, A.M. 1973: A history of fire and vegetation in northeastern Minnesota as recorded in lake sediment. *Quaternary Research* 3, 383–96.

Tinner, W. and **Conedera, M.** 1995: Indagini paleobotaniche sulla storia della vegetazione e degli incendi forestali durante l'Olocene al Lago di Origlio (Ticino meridionale). *Bollettino della Società Ticinese di Scienze Naturali* 83, 91–106.

Tinner, W. and **Hu, F.S.** 2003: Size parameters, size-class distribution and area-number relationship of microscopic charcoal: relevance for fire reconstruction. *The Holocene* 13, 499–505.

Tinner, W., Conedera, M., Ammann, B., Gäggeler, H.W., Gedye, S., Jones, R. and **Sägesser, B.** 1998: Pollen and charcoal in lake sediments compared with historically documented forest fires in southern Switzerland since AD 1920. *The Holocene* 8, 31–42.

Tinner, W., Conedera, M., Gobet, E., Hubschmid, P., Wehrli, M. and **Ammann, B.** 2000: A palaeoecological attempt to classify fire sensitivity of trees in the southern Alps. *The Holocene* 10, 565–74.

Tinner, W., Hubschmid, P., Wehrli, M., Ammann, B. and **Conedera, M.** 1999: Long-term forest fire ecology and dynamics in southern Switzerland. *Journal of Ecology* 87, 273–89.

Waddington, J.C.B. 1969: A stratigraphic record of the pollen influx to a lake in the Big Woods of Minnesota. *Geological Society of America, Special Paper* 123, 263–83.

Whitlock, C. and **Larsen, C.** 2001: Charcoal as a fire proxy. In Smol J.P., Birks, H.J.B. and Last, W.M., editors, *Terrestrial, algal, and siliceous indicators*, Dordrecht: Kluwer, 75–97.

Wright, H.E. 1991: Coring tips. *Journal of Paleolimnology* 6, 37–49.