Author's Accepted Manuscript

Methodological concerns for analysis of phytolith assemblages: Does count size matter?

Caroline A.E. Strömberg

PII: S1040-6182(07)00345-X
DOI: doi:10.1016/j.quaint.2007.11.008
Reference: JQI 1712

To appear in: *Quaternary International*

Cite this article as: Caroline A.E. Strömberg, Methodological concerns for analysis of phytolith assemblages: Does count size matter?, *Quaternary International* (2007), doi:10.1016/j.quaint.2007.11.008

Methodological concerns for analysis of phytolith assemblages: Does count size matter?

Caroline A.E. Strömberg[1]

[1] Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560-0121, USA. *Phone:* +1 (202) 633 1372. *Fax:* +1 (202) 786 2832. *E-mail:* strombergc@si.edu

Abstract

In quantitative phytolith analysis, chance error associated with insufficient counts can affect the robustness of the interpretation, whether it is vegetation reconstruction or taxonomic differentiation. It is therefore vital to choose a count size that will ensure statistically reliable results, while minimizing the time expended. Numerical statistical methods (bootstrapping) that have become available over the past few decades have made it possible to model even complex phytolith assemblages with relative ease. This study used bootstrapping as well as analytic statistical formulas to evaluate the influence of count size on vegetation reconstruction by means of two commonly used indices, D/P (tree cover index) and Iph (aridity index). The analysis indicates that the count size needed to ensure statistical precision depends on the question as well as the observed assemblage composition. Importantly, it is the count of specimens relevant to a specific ratio or other index ("index-specific" count) that matters, whereas the total count is less important. Based on these results, some general guidelines for choice of count size and for the use of statistics in phytolith analysis are suggested.

Keywords: Phytolith analysis; sampling; count size, statistical inference, bootstrapping.

1. Introduction

Applying phytolith analysis to understand vegetation structure or temporal/spatial changes in vegetation structure inherently relies on quantitative comparison between assemblages. More precisely, it involves comparison of counts of different morphotypes or classes of morphotypes to summarize the information stored in the soil record of biosilica. Thus, a change from grassland to forest is reflected in phytolith assemblages as an increase in forest indicator forms relative to grass phytoliths; a change from cool-season grassland to warm-season grassland is indicated by a change in the ratio of morphotypes typical of these grass groups; an increase in land-use may be marked by a relative increase in phytoliths from domesticated palm species (e.g., Barboni et al., 2007; Kondo et al., 1994; Runge, 1999; Shulmeister et al., 2001). Several indices have been devised to express both absolute and relative aspects of vegetation type. For example, the tree cover index, D/P, uses the ratio of certain forest indicator phytoliths to grass phytoliths to give a measure of vegetation structure in absolute terms (a D/P ratio >1.82=forest; Bremond et al., 2005a) and in relative terms (a decrease in the D/P ratio signals an opening up of the landscape; e.g. Alexandre et al., 1997; Bremond et al., 2005a). Other commonly used indices include the aridity index (Iph=tall vs. short grass savannah), and the climatic index (Ic=C_3 vs. C_4 grasslands) (e.g., Alexandre et al., 1999; Bremond et al., 2005b; Diester-Haass et al., 1973; Twiss et al., 1987). A variety of multivariate analytic methods currently in use (e.g., Blinnikov et al., 2001; Boyd et al., 1998; Kealhofer, 1996; Prebble and Shulmeister, 2002) likewise rely on differences between absolute or relative abundances among assemblages for reconstructing vegetation

change, as do the traditional relative frequency diagrams/tables (e.g., Rovner, 2001; Runge, 1999; Strömberg, 2004).

Similarly, plant taxonomic differentiation based on phytoliths often uses quantitative comparisons of relative frequencies of morphotypes. For example, archaeologists seeking to distinguish the phytoliths of a domesticated grass from those of its wild sister taxa employ comparative size statistics, relative frequencies, or morphotype ratios (Pearsall, 1978; Pearsall, 1982; Piperno, 1984). Several recent studies have used explicitly quantitative approaches to estimate morphotype frequencies or morphotype shape and size as a means to separate highly complex and redundant phytolith assemblages produced by different taxa (e.g., Ball et al., 1996; Carnelli et al., 2004; Fujiwara, 1993; Pearsall and Piperno, 1990; Piperno and Stothert, 2003; Zhao et al., 1998).

For both of these applications of quantitative phytolith analysis, the validity of the inference hinges on the statistical significance of a certain abundance, ratio, or difference between samples (Ball et al., 1996, 1999). This needed statistical significance depends largely on the number of phytoliths used for calculation, making count size an essential consideration for the phytolith researcher.

1.1. Biases produced by insufficient count size

Count size, the number of phytolith specimens tallied on a slide, is the last in the long series of steps of sub-sampling that occurs when a soil assemblage or a plant sample is processed and analyzed for phytoliths. Each of these steps may introduce biases and errors that render the final sub-sample non-representative of the original assemblage (e.g., Fredlund, 1986; Lentfer and Boyd, 1999; Strömberg, 2007); the phytolith count

is no different. Simply put—and assuming that errors during sub-sampling in the field and laboratory can be eliminated and "truly" random counts are obtained (such as in the 60-field scan procedure; see Dinan and Rowlett, 1993; Pearsall, 2000)—an insufficient number of specimens can, purely by chance, lead to misrepresentation of forms. This, in turn, may result in erroneous vegetation inference (in the case of paleoecological analysis) or size statistics (in the case of morphological quantification).

Hence, the goal must be to reach a count that will generate statistically robust measurements. However, a single phytolith extraction commonly yields thousands, if not millions, of specimens, making it an enormous chore to count them all. The desire to achieve reliable data must therefore be weighed against the time invested in counting and return in terms of information. Concerns regarding insufficient count size (usually referred to as "sample size") are shared with other disciplines of paleoecology and paleontology, where it has attracted considerable attention in the literature (e.g., Alroy, 2000; Foote, 1992; Raup, 1975; Stanton and Evans, 1972). Indeed, several publications have specifically addressed the task of finding the optimal count size given the trade-off between statistical precision and efficiency (Birks and Birks, 1980; Faegri et al., 1989; Jamniczky et al., 2003; Maher, 1972; Rull, 1987; Wolff, 1975).

1.2. Previous work on count size in phytolith analysis

Among phytolith researchers using phytoliths to differentiate plant taxa, Ball and colleagues (Ball et al., 1996, 1999) have devised formulas to quantitatively estimate the count sizes needed to ensure the statistical significance of their results. They

calculated the minimum number of phytoliths they had to measure to yield a ±5% wide 90% confidence interval around the sample mean for a give variable. Depending on the variable, this number varied between 5 and 160 measured phytoliths for each taxon. Other authors have applied statistical measures such as confidence intervals to ensure the robustness of their phytolith-based taxonomic interpretation (e.g., Pearsall and Piperno, 1990; Piperno, 2006).

In contrast, workers using phytoliths for paleoecological reconstruction seldom treat the issue of count size in much detail (but see Albert and Weiner, 2001; Pearsall, 2000; Zurro and Madella, this issue). Moreover, counts differ widely among studies (see also review in Piperno, 2006). Count sizes range from approximately 100 to 800 per sample, although counts of 5,000 phytoliths have been reported (Piperno and Becker, 1996). Most commonly, the number of specimens tallied per sample lie in the 200-400-range (e.g., Alexandre et al., 1997; Blinnikov et al., 2002; Carter, 2000). Authors use different, or no specific, criteria for when to stop counting. Some count a certain number of grass silica short cell phytoliths, a specified number of glass beads for absolute estimates of phytolith abundance, or an *a priori* assigned total number of phytolith morphotypes. However, it is rare that a statistical motivation for the chosen count size is provided. An exception is Pearsall (2000), who judged that 200 grass silica short cell phytoliths is an acceptable minimum number to get a representative count. She based this estimate on comparisons between tallies of 100, 150, 200, 250, 300, 350, and 400 silica short cells from Hawaiian vegetation types, and concluded that assemblage patterns of occurrences remained fairly constant for counts >200. In their study of cave deposits in Israel, Albert and Weiner (2001) arrived at a similar number by tallying incrementally higher counts of phytoliths and comparing with results from a very high count size (800 specimens). They noted that many less

abundant forms were lost in sums below 200 specimens. Piperno (1988) has also pointed out that tallies in excess of 200 may be necessary for certain samples to capture rare but ecologically significant forms.

In addition, few phytolith workers today use statistics to validate their estimates (see discussion in Pearsall, 2000). The various indices discussed above are in most cases reported without standard error or confidence intervals (e.g., Alexandre et al., 1999; Fearn, 1998; Strömberg, 2002, 2003; Twiss et al., 1987, but see Bremond et al., 2005a, 2005b). The same is true for the traditional relative frequency diagrams and tables (e.g., Kondo et al., 1994; Rovner, 2001), and assemblage data used in paleoecological multivariate analysis (e.g., Blinnikov et al., 2001; Kealhofer, 1996). Similarly, statistical hypothesis testing is rarely employed to underscore a certain interpretation (but see Chiswell, 1984, cited in Pearsall, 2000).

Piperno (2006) provided general guidelines for appropriate count sizes for phytolith analysis based on empirical experience and comparisons with the palynological literature. However, as I will argue below, palynological methods for determining adequate count sizes are not always appropriate for the types of questions that concern phytolith researchers. Therefore, a quantitative study to estimate the minimum phytolith count that will secure a reliable (statistically precise) vegetation inference in a variety of cases is sorely needed. The purpose of this paper is to provide such a study.

1.3. Concerns for determining adequate count size in phytolith analysis

Count size is a function of the structure of the data in hand (see discussion below), but also the question and the type of measurement that will provide the appropriate

answer. Thus, the quantities regularly employed in paleoecology and paleontology—absolute frequencies, diversity estimates, taxon/type richness, relative abundances—all require different count sizes for statistical robustness (Rull, 1987, see also Jiroutek et al., 2003, for discussion). The focus in paleozoology and traditional paleobotany (macrofossil and palynomorph studies) is often on documenting diversity (taxon/type richness + type equitability/evenness) (e.g., Falcon-Lang, 2003; Wing et al., 1995). Adequate count sizes for these types of studies are usually determined by means of sampling curves, where the number of taxa or types making up an assemblage is plotted against the number of specimens counted (see discussion in Jamniczky et al., 2003; Rull, 1987; Zurro and Madella, this issue). An adequate count size is determined as the point where a large fraction of the taxa/types have been encountered, causing the curve representing the sampling effort to level out. In other words, significantly higher count sizes will not return much in terms of more counted taxa/types. The sufficient count size is thus established both on the number of taxa/types in the assemblage (richness) and on their relative abundance distribution (evenness). In assemblages with high evenness, the appropriate count size can be expected to be lower, whereas in assemblages with a few very common taxa and many rare taxa, the sampling curve levels off very slowly, requiring higher count sizes (Jamniczky et al., 2003; Rull, 1987).

Zurro and Madella (this issue) explore the use of sampling curves to estimate appropriate count sizes. However, there are several reasons why a different approach may be more useful for applications within phytolith analysis such as vegetation reconstruction. First, due to the multiplicity and redundancy inherent in phytolith production, there is not a one-to-one correspondence between number of phytolith morphotypes and the number of plant taxa represented in an assemblage (Piperno,

1988; Rovner, 1971). Therefore, the count size determined by sampling curves is a function of the number of different phytolith *morphotypes*, not plant taxa, contained in an assemblage. An inherent problem with this method is that phytolith workers differ widely in how finely they subdivide the morphological spectrum encountered in phytolith assemblages in plants or in soil samples, and also in which morphotypes they count. Some authors consider only a very low number of morphotype classes for vegetation inference (e.g., 7 in Alexandre et al., 1997); others subdivide all, or certain, morphotypes into a large number of classes (e.g., Brown, 1984; Carnelli et al., 2004). For example, Bremond et al. (2005a, 2005b) combine all dumbbell-shaped grass silica short cells into one class, whereas other authors split this morphotype class into several subclasses (e.g., Fredlund and Tieszen, 1997; Piperno and Pearsall, 1998; Strömberg, 2005). Different numbers of morphotype classes will undoubtedly result in different sampling curves and consequently arrive at different optimal count sizes, in particular because a lower number of (broader) morphological classes will also likely correspond to higher evenness. This variation will lead to some arbitrariness in assigning count size. Another more critical problem is that not all morphotypes are necessarily relevant for vegetation reconstruction. Consequently, the count size achieved using sampling curves may be unnecessarily high, especially if the assemblage has low evenness. Capturing every morphotype may not be required for the research question. If the question at hand is "how many types of phytoliths are present in this assemblage?," sampling curves are clearly the right tool for finding the adequate count size. However, for the questions that are often central to phytolith analysis, such as "what is the proportion of phytoliths indicative of cool-temperate grasses to phytoliths of tropical grasses?" the use of sampling curves to determine the

count sizes that will provide robust vegetation or taxonomic inference seems inappropriate.

Another common focus within palynology has been to find the count size that allows statistical precision of relative abundances or absolute frequencies (e.g., Birks and Birks, 1980; MacDonald, 1996; Mosimann, 1965). Note that absolute frequencies will not be treated herein; see Pearsall (2000) for further discussion. Traditionally, palynologists have used confidence intervals (usually 95%) for relative frequencies to indicate at which grain count reliable estimates of assemblage composition have been reached (e.g., Faegri et al., 1989; MacDonald, 1996; Maher, 1972). Counts are viewed as sufficient when the width of the confidence interval does not change significantly with increasing counts or when it is "acceptable" following some *a priori* criterion determined by the investigator (Maher, 1972; Rull, 1987). This stopping point will depend on the overall count size and the relative frequency of the palynomorph type of interest. If more than one type is considered, both richness and evenness of taxa in the sample as a whole has to be taken into account (Maher, 1972; Mosimann, 1965; Rull, 1987). Thus, samples with low taxon richness and high evenness require lower counts (on the order of 200 grains, Rull, 1987), while large count sizes are essential for reliable estimates of taxa with very low frequencies (1620 grains for a taxon with a relative abundance of 1%, Rull, 1987). Mosimann (1965), Maher (1972), and Rull (1987) described practical methods for finding the appropriate tally and evaluating statistical precision of percentages for palynomorph assemblages modeled as multinomial distributions.

Phytolith analysis builds on the same general principles as the study of spores and pollen. Nevertheless, it is not necessarily clear how Maher's (1972) nomograms and Mosimann's (1965) equations can be used for the questions that interest the

phytolith researcher. Recently published, more complex criteria for optimizing count size within palynology are also not directly applicable to phytolith analysis (e.g., Lytle and Wahl, 2005).The rare use of statistics in published phytolith work is a likely indication of this circumstance.

How should we determine what is a sufficient count size for phytolith-based vegetation reconstruction or taxonomic differentiation? Both of these endeavors involve quantifying the relative abundances of sets of morphotypes, which, in a finite phytolith count, may vary due to sampling error. Accordingly, phytolith researchers must ask themselves: how many phytoliths must be counted until the relationship between these morphotypes in the sample is precisely recorded, or at least robust enough that the resulting vegetation/taxonomic inference is not substantially influenced by chance error? Statistical precision can be measured using confidence intervals. Thus, in statistical terms, the question becomes: what count size is necessary to obtain an "acceptable" confidence interval on the ratio/proportion of interest? What is "acceptable" depends on the statistical precision needed to answer a particular question.

The dilemma facing the phytolith worker when trying to assign confidence intervals is that statistical analytic modeling of indices such as D/P, Ic, and Iph is not trivial. Ratios and proportions compound the sampling error of each of the measurements that go into calculating them, making them relatively inaccurate; in addition, the resulting metrics are often not normally distributed (Sokal and Rohlf, 1981). Another problem arises because phytolith assemblages consist both of morphotypes relevant to the question at hand, "diagnostic" morphotypes, and of other, "non-diagnostic" morphotypes. For example, for D/P, "diagnostic" morphotypes consist of globular granulates of woody dicotyledons, grass silica short cells,

cuneiform bulliform cells, and acicular hair cells; "non-diagnostic" morphotypes may include various morphotypes from dicotyledons, palms, sedges, as well as ubiquitous or unknown forms (Table 1). As a result, the sum and frequency of the diagnostic morphotypes may vary in each counted sample (e.g., table 2 in Bremond et al., 2005b). This additional variation is referred to as an unconditional case in statistical terms. Taking the unconditional case into account is important because it may yield slightly wider confidence intervals, at least at low counts (D. Freedman, University of California at Berkeley, 2004, personal communication). However, because it substantially complicates calculations of confidence intervals using conventional analytic statistical formulas (Freedman et al., 1998; D. Freedman, University of California at Berkeley, 2004, personal communication), resampling statistics has emerged in the past few decades as an alternative approach (e.g., Efron, 1979; Mooney and Duval, 1993).

1.4. Resampling

Resampling is a set of non-parametric statistical methods by which data simulation instead of analytic formulas is used to estimate the precision of a parameter of the population (the "true" phytolith assemblage), for example the mean, median, and variance (Good, 2006; Simon, 1997). In bootstrapping, the resampling technique used herein, samples are drawn randomly and with replacement from the available data (the phytolith count). This is repeated thousands of times, modeling the physical process of conducting the same experiment or sample protocol over and over again (e.g., tossing a coin, or counting 100 phytoliths). Unlike conventional analytic statistics, bootstrapping involves no assumptions about the background universe (e.g.,

normality and homogeneity of variance), other than the assumption of random sampling (Mooney and Duval, 1993; Simon, 1997; Simon and Bruce, 1991). Because bootstrapping is numeric, rather than analytic, it is by definition inexact, but it has the advantage of being less abstract than conventional statistical methods. For these reasons, bootstrapping can deal with complex distributions that are hard to model mathematically using analytic statistical approaches and data for which parametric assumptions cannot be made. For most probabilistic and statistical questions, resampling methods such as bootstrapping have been shown to give equally good or superior results compared to conventional statistics; exceptions include situations dealing with low probabilities, low count sizes, and missing data (Mooney and Duval, 1993; Simon, 1997; Simon and Bruce, 1991).

The primary goal of this paper is to assign general guidelines for determining appropriate count sizes for paleoecological analysis using phytoliths; a secondary goal is to provide suggestions for how to add statistical power to phytolith assemblage interpretation. To accomplish these goals, I apply analytic statistical formulas as well as bootstrapping to theoretical phytolith assemblages to determine how the precision of vegetation inference based on phytolith ratios varies with count size. I used the tree cover index (D/P) and the aridity index (Iph) as examples of commonly used phytolith indices for paleoecological reconstruction. However, the results can be applied generally within phytolith analysis, whether concerning vegetation inference or taxonomic differentiation. To explore how the value of the "true" assemblage D/P and Iph affects confidence interval width, several different hypothetical phytolith assemblage compositions are examined. The results are discussed with reference to recent studies employing D/P and Iph and similar indices for reconstructing vegetation (Bremond et al., 2005a, 2005b; Strömberg, 2005). Specifically, I

investigate how the addition of confidence intervals to published D/P and Iph values affects vegetation interpretation (Bremond et al., 2005b).

2. Methods

It can be hypothesized that the width (and shape) of the confidence interval (C.I.) for a certain index (D/P, Iph) measured in a phytolith assemblage varies depending on:
(1) the level of confidence (90%, 95%, 99% etc.).
(2) the count of "diagnostic" morphotypes; that is, the count used to estimate the assemblage index, hereafter referred to as the *index-specific count* (n_i);
(3) the total count size, that is, how many "non-diagnostic" (non-index-specific) morphotypes (ND) are included (conditional case vs. unconditional case);
(4) the method of calculating the C.I. (analytic formulas vs. bootstrapping); and
(5) the "true" value of the index being measured (D/P, Iph).

The level of confidence is set to 95% in this study, in accordance with its common use in other disciplines of paleoecology (e.g., Lytle and Wahl, 2005).The confidence intervals were simulated for several different index-specific count sizes (n_i=10, 25, 50, …, 2000; see Figs. 1,2, Supplementary Tables 1,2).

For both indices, several different cases were investigated to test the influence of variables (3)-(5) above on the C.I. (Table 1). For the D/P index, the effect of non-index-specific phytoliths was tested by calculating the C.I.s and sampling errors for two different cases. In the first case, only the index-specific morphotypes (conditional case) were included; in the second case, the index-specific morphotypes and a chosen percentage of non-index-specific morphotypes were included (unconditional case). For the unconditional case, the proportion of non-index-specific phytoliths (ND) was

set to 50%, but note that this will vary with chance error. This value is a reasonable estimate in light of a modern dataset consisting of 61 samples from West Africa, in which non-index-specific phytoliths for D/P made up 53% of the total count (Bremond et al., 2005b). Similarly, in 94 Cenozoic phytolith assemblages from North America the count of morphotypes not used to reconstruct tree cover constitutes 19-67%, with an average of 47% (Strömberg, 2005). To explore the influence of calculation method, C.I.s were computed using both analytic formulas and bootstrapping for both the conditional and unconditional case.

For the Iph index, bootstrapping was used to simulate the C.I.s for the conditional case and for two different unconditional cases, ND=70% and ND=90%, to further explore variation due to inclusion of non-index-specific morphotypes in the overall count. The latter values are based on Bremond et al. (2005b), who reported that non-index-specific morphotypes for Iph make up 50-93% (mean=71%; 24 samples) in assemblages from shrub savanna with short grasses and 42-86% (mean=67%; 28 samples) in assemblages from tree/shrub tall-grass savanna. In short-grass steppe with shrubs, non-index-specific morphotypes for Iph constitute 83-97% (mean=90%; 7 samples); in semi-deciduous forest, the non-index-specific count for Iph amounts to 95-97% (mean=96%; 2 samples); these numbers are similar to values in Alexandre et al. (1997). In addition, C.I.s for the conditional case was calculated using analytic statistical formulas (see Table 1).

The variation in C.I.s due to the "true" value of D/P and Iph was explored using a range of representative values for each index. Typical values for D/P and Iph are given by recently published analyses of phytolith assemblages from modern vegetation types in West Africa (Bremond et al., 2005a, 2005b). Emphasis was placed on these studies because they report a large number of assemblages, provide all, or

part of, the raw data, and link their phytolith data with estimates of tree cover based on Leaf Area Index (LAI), pollen data, Moderate Resolution Imaging Spectroradiometer satellite data (MODIS), or all three (Barboni et al., 2007; Bremond et al., 2005a, 2005b; Cournac et al., 2002; Vincens et al., 2000). Also, the definition of D/P has changed since earlier modern analogue work in Africa (e.g., Alexandre et al., 1997). The definition of the tree cover index used by Bremond et al. (2005a, 2005b) follows the original use in Alexandre et al. (1997), but excludes smooth and wavy elongates. Thus, it is now defined as D/P=ligneous dicotyledon phytoliths / Poaceae phytoliths, where ligneous dicotyledon phytoliths=globular granulate, and Poaceae phytoliths=grass silica short cells (bilobate, cross, saddle, rondel, crenate)+cuneiform bulliform cell+acicular hair cell (Table 1) (Bremond et al., 2005a).

According to these studies, tall-grass savanna with a MODIS tree cover value of 1-34% (mean=13%; 28 samples) yields a D/P ranging from 0 to 0.49 (mean=0.07) (Bremond et al., 2005b). Dense tree/shrub tall-grass savanna, with a LAI of 0-~1.2 and a MODIS tree cover value of 46-60% (Barboni et al., 2007), yields D/P values of 0.65-1.16 (mean=0.92; 5 samples) (Bremond et al., 2005a). Forest vegetation, ranging from young *Albizia* forest to mature *Rinorea* forest with an LAI of ~1.5-~4 and MODIS tree cover values of 79-82% (Barboni et al., 2007), produces D/P values of 1.82-5.13 (mean=2.86; 15 samples) (Bremond et al., 2005a). Alexandre et al. (1997) reported D/P as high as 13.1 (recalculated using the current definition of D/P).

Based on this research, the following approximate values can be stated as rough guidelines for use of D/P in Africa: D/P<0.5 denotes a savanna or grassland (tree cover<40%); D/P~1 (0.65-1.16) corresponds to a dense savanna or woodland (tree cover~50%), and D/P=2 (>1.82) denotes a forest (tree cover>>~60%). To take

into account the typical distribution of phytolith morphotypes in these assemblages, five theoretical D/P ratios were chosen for the study: 0.1, 0.5, 1, 2, and 9.

The aridity index is defined as Iph=Chloridoideae phytoliths / (Chloridoideae phytoliths + Panicoideae phytoliths), where Chloridoideae phytoliths=saddle and Panicoideae phytoliths=bilobate+cross (Table 1) (Bremond et al., 2005b). Diester-Haass et al. (1973) determined that an Iph≥40% signified dry grasslands, whereas Alexandre et al. (1997) stated that an Iph of approximately 30% separates tall and short grass communities in West Africa. I use Iph=20(±1.4) % as the boundary between tall grass savannas (Iph=2.5-24%, mean=10%; 28 samples) and short grass savannas (Iph=12-73%, mean=31%; 24 samples), based on the analysis of 61 assemblages in Bremond et al. (2005b). The theoretical values of Iph chosen for this study were 10%, 20%, and 30%, respectively.

The bootstrapping simulations were preformed using Resampling Stats 5.0 software (www.resample.com) with 10,000 replicates for each D/P and Iph value and count size, and 95% C.I.s were determined in each case. The scripts for these simulations in Resampling Stats are available upon request from the author. The deviation due to sampling error from the sample D/P and Iph was calculated as (index $_{95\%max}$ – index $_{95\%min}$)/2. The analytic statistical formulas used to calculate the 95% C.I.s are given in Table 1.

For each theoretical D/P and Iph value and count size, the upper and lower limits of the C.I.s were examined in light of the interpretive scheme reviewed above (Bremond et al., 2005a, 2005b), to evaluate how the interpretation of habitat structure may change due to sampling error.

To further demonstrate how the addition of C.I.s to phytolith analysis may affect the robustness of paleoecological interpretation, C.I.s were calculated for

published D/P and Iph index values using both analytic formulas (conditional case) and bootstrapping (unconditional case). For reasons elaborated above, focus was placed on the well-studied West African modern analogue dataset published by Bremond et al. (2005b).

3. Results

3.1. Effect of index-specific count size on C.I. width

For both the D/P and Iph index, the width of the C.I. decreases asymptotically with the number of diagnostic phytoliths counted (Figs. 1,2, Supplementary Tables 1,2). For the values of the Iph index studied herein and for values of D/P≤0.5, the C.I. curves flatten out at index-specific count sizes above approximately 200 diagnostic phytoliths (Figs. 1d,e,2) (c.f. Rull, 1987). This marks the point at which the C.I. interval width does not change much with increasing count size; thus, counting more diagnostic phytoliths will not markedly improve the accuracy of the index measurement. For values of D/P≤2 and D/P>>2, this index-specific count is closer to 250 and 300 diagnostic phytoliths, respectively (Fig. 1a-c).

3.2. Effect of total count size on C.I. width

For D/P, calculation of 95% C.I. using analytic formulas for the conditional and unconditional case gave identical results at all count sizes (Fig. 1, Supplementary Table 1). When bootstrapping was used, simulation for the conditional and

unconditional case, respectively, also yielded matching C.I.s at count sizes above 25-50 diagnostic phytoliths. At lower index-specific count sizes, unconditional bootstrapping gave C.I.s that are often substantially wider than conditional bootstrapping, but note that there is also some fluctuation due to the inexact nature of the resampling procedure. For Iph, the relative proportion of non-diagnostic phytoliths [ND=0% (conditional), ND=70% or ND=90%] does not appear to affect the width of the bootstrapped C.I. at any of the index-specific count sizes investigated (Fig. 2, Supplementary Table 2). Importantly, at index-specific count sizes of ≥200, total count size (conditional vs. unconditional case) is irrelevant.

3.3. Effect of C.I. calculation method (analytic formulas vs. bootstrapping) on C.I. width and shape

For the Iph index, analytic formulas and bootstrapping yielded C.I.s of roughly equal width above the lowest index-specific count-sizes (Fig. 2, Supplementary Table 2). However, the discrepancy in width between C.I.s produced by the two calculation methods differs slightly depending on the value of the Iph index itself. Thus, at higher values of Iph (30% and 50%, the latter not shown) and very low counts (≤25), bootstrapping produced C.I.s that are 2-4 percentage units wider than C.I.s computed by analytic formulas. Note that, at low values of Iph (10%; Fig. 2c), the finite samples involved in bootstrapping led to C.I.s that are truncated at 0 and, therefore, more narrow than analytic C.I.s.

Similarly, for D/P, analytic formulas and bootstrapping generated C.I.s of equal width above a certain index-specific count. Below this count, bootstrapping produced C.I.s that are markedly wider (Fig. 1, Supplementary Table 1). The critical

index-specific count varies strongly depending on the value of D/P: the higher the value of D/P, the higher number of diagnostic phytoliths that must be counted to yield C.I.s of similar width as those calculated through analytic statistics. For example, at D/P=0.1, the critical index-specific count is ~25; at D/P=2, it is ~75; and at D/P=9, it is ~300 (Fig. 1, Supplementary Table 1).

Not only the width, but also the shape, varies somewhat between C.I.s computed using the two different methods (Figs. 1,2). Whereas analytic formulas generated symmetrical C.I.s, bootstrapping simulation resulted in C.I.s that were more or less asymmetrical, and more so at lower index-specific counts. The reasons for this difference are that bootstrapped C.I.s must have a lower boundary of 0 and they tend to be somewhat skewed. This tendency for bootstrapped C.I.s (along with the random fluctuation of the C.I. width, producing an asymptote that is not entirely smooth) is a result of the finite sample size at low counts and thus reflects a "real" aspect of the sampling procedure. For most values of Iph (~20-80%), this phenomenon does not lead to any major differences between C.I.s generated through analytic and resampling methods, respectively, above very low index-specific count sizes (>10). When Iph is close to 0%, the trend is more pronounced, resulting in markedly larger sampling error towards higher values. For Iph values close to 100% (data not shown), the sampling error is larger towards smaller values. Note that these differences are only important at low index-specific counts (<50).

For D/P, the C.I.s are always skewed towards the higher values and become more so at higher values of D/P, even at high index-specific count sizes (>100). Thus, the boundaries of the bootstrapped C.I.s can be substantially shifted relative to those generated through analytic statistics. For example, for D/P=2 and at a count of 200 diagnostic phytoliths, the 95% C.I. calculated with analytic formulas is 1.4-2.6,

compared to a bootstrapped 95% C.I. of 1.5-2.7. Note that, for D/P=9, the upper limit of the bootstrapped 95% C.I. *increases* from the smallest index-specific count sizes; in reality, it should start from positive infinity. This phenomenon is likely a function of the way the bootstrapping algorithm handles division by 0.

It can be concluded that, in terms of C.I. width, the choice of calculation method is less important above lower index-specific count sizes; at a count size of 200 diagnostic phytoliths analytic statistics and bootstrapping yield C.I.s of comparable width. On the other hand, the asymmetrical nature of bootstrap C.I.s may create marked differences in C.I. boundary positions even at higher index-specific count sizes (e.g., 200).

3.4. Effect of the value of the index being measured (D/P, Iph) on C.I. width

The width of the C.I. for Iph is strongly dependent on the value of Iph itself, at any given index-specific count size and regardless of calculation method (Fig. 2, Supplementary Table 2). It is largest for Iph=50% (data not shown), and smallest when Iph approaches 0 or 100% (data not shown for the latter). Thus, the width of the 95% C.I. at an index-specific count size of 50 is over 50% wider for Iph=30% (C.I.=26%) than for Iph=10% (C.I.=17%), using unconditional bootstrapping (Fig. 2a,c). This is true even at very high index-specific count sizes (Supplementary Table 2); at an index-specific count of 2000, the C.I. is 40% wider at Iph=30% (C.I.=3.8%) than at Iph=10% (C.I.=2.7%).

The C.I.s of D/P get wider for higher values of the D/P ratio, at any given index-specific count size. However, if calculated as a percentage of the D/P index value, the C.I. is at its most narrow when D/P=1. Thus, at an index-specific count size

of 100, the width of the C.I. for D/P=1 is 1.2 (corresponding to 121% of the D/P ratio) whereas for D/P=0.1, it is 0.2 (196% of the D/P ratio), and for D/P=9, it is 38.7 (430% of the D/P ratio); a difference in C.I. width persists at higher index-specific count sizes.

3.5. Application to previously published studies

Calculations of 95% C.I.s for D/P ratios reported in Bremond et al. (2005b) resulted in relatively small sampling errors for most of the samples from open vegetation (savanna and steppe), despite index-specific count sizes that are often well below 200 (Fig. 3a). This relates to the fact that most counts are from assemblages with D/P<<0.1. In contrast, the samples with higher D/P (>0.4) have very wide C.I.s, high index-specific count sizes notwithstanding. In the case of the assemblages with D/P close to 0.5 (samples S.91, 83-127), the sampling error is large enough that the upper limits of the 95% confidence intervals fall outside the range of values characteristic of "open savanna" (>0.5) (Fig. 3a). For the samples with D/P>4, the C.I.s are very large (±~40%), but are above the limit for closed forest vegetation (D/P>1.82) (Fig. 3b).

Calculation of the 95% C.I.s for Iph indices for the same dataset (Bremond et al., 2005b) showed that, out of 31 samples from vegetation dominated by short grasses [disregarding the samples with non-typical Iph values (<20%)], 13 have C.I.s with lower limits below the boundary Iph value of 20% (Fig. 4). The same is true for one of the two samples from forest vegetation, which has an Iph value of 35%. Similarly, of the 28 samples from tall-grass savanna (ignoring samples with Iph values >20%), four samples have wide enough C.I.s that their upper C.I. limits fall above the 20% boundary. In many cases, this phenomenon is a consequence of low

index-specific count sizes, that is, less than ~100 Chloridoideae + Panicoideae phytoliths (e.g., the forest sample, S.155, for which n_i=10, Fig. 4; Bremond et al., 2005b). In other cases, the C.I.s are reasonably narrow, but the Iph value is close to the 20% boundary (e.g., sample 82-79, with n_i=165, Fig. 4).

Note that the C.I.s calculated using analytic formulas (conditional case) and bootstrapping (unconditional case), respectively, do not differ substantially in width (Figs. 3,4). However, for D/P, the bootstrapped C.I.s are regularly shifted towards higher values, a tendency that is particularly noticeable for high D/P values (Fig. 3b).

4. Discussion

4.1. Confidence intervals

The accuracy of vegetation reconstruction or taxonomic differentiation using phytoliths depends on the sampling error of the indices or other quantitative measurements being used. The sampling errors and corresponding confidence intervals vary in width, shape and symmetry as a function of several factors. One of the factors affecting the size of the sampling error is the level of confidence (90%, 95%, etc.), another is the type of measurement used (e.g., proportions such as Iph vs. ratios such as D/P). The simulation study conducted in this paper highlighted two other factors that vitally influence confidence interval width, namely (1) the number of diagnostic specimens counted (index-specific count) and (2) the value of the index under investigation (see also Jiroutek et al., 2003; Maher, 1972; Rull, 1987). In contrast, at reasonably large index-specific count sizes (>>25), the total specimen

count (conditional or unconditional case) does not significantly affect the width of the confidence interval; neither does the C.I. calculation method (analytic statistics vs. bootstrapping) (Figs. 1,2).

The symmetry and shape of the confidence interval, on the other hand, depend on how the sampling errors are computed. Both analytic formulas and resampling methods are approximations that work best at high index-specific count sizes (e.g., Freedman et al., 1998; Simon, 1997), but bootstrapping has the advantage that it captures the skewed distribution of sampling errors that results from dealing with finite samples. Note that C.I.s calculated using an alternative analytic method, log-odds ratio, are also skewed to some extent (data not shown). In many cases, for example when proportions are examined, these differences are a potential problem only for boundary values of the index, close to 0 and 100%, and at low index-specific count sizes (Fig. 2c, see also Fig. 4). In other cases, such as for the D/P ratio, variation in confidence interval shape and symmetry is apparent for all values of the index and the choice of C.I. calculation method becomes important for robust assemblage interpretation (Fig. 1, see also Fig. 3).

There are several implications of these results. First, it is not possible to report a fixed sampling error for a particular index (e.g., Bremond et al., 2005a, 2005b) because the sampling error will vary with index-specific count size and value of the index in question. Second, and more importantly, determination of adequate count sizes should focus on the index-specific count (n_i), not the overall tally. Even if as much as 600 phytoliths are counted overall, the confidence interval of the particular index might be unacceptably wide if the index-specific count is substantially below 200. Furthermore, if several assemblage indices are to be investigated, the index-

specific count size should be adjusted to that which involves the smallest fraction of the overall count.

So how many index-specific morphotypes should be counted? Tallies of at least 200-250 diagnostic specimens ensure relatively constant confidence interval widths for a wide range of values for indices such as Iph and D/P (Figs. 1,2) (Rull, 1987) and should therefore generally produce statistically reliable estimates for assemblage analysis. Exceptions are D/P values >>2, for which counts in excess of 300 diagnostic phytoliths are necessary to meet the criterion of stable C.I. width (Fig. 1a). A count of 200 phytoliths correspond broadly with previous recommendations (Piperno, 1988; Pearsall, 2000; Albert and Weiner, 2001), but note that these authors did not distinguish index-specific count from total count in their analyses.

Although assigning index-specific counts of >200 as a general guideline for phytolith analysis may seem straightforward, it is not without problems. A count of 200 Chloridoideae + Panicoideae phytoliths for calculation of the Iph index could correspond to overall counts of 2000 phytoliths, depending on the relative abundance of index-specific phytoliths. To count that many phytoliths seems unreasonable. Instead, focus must be placed on how much statistical robustness is needed to yield an unambiguous interpretation of the data, given a particular set of analytic rules. Rules for analysis may include that the confidence interval for an index must be of a predetermined, "acceptable" width, or that it should not cross a certain boundary value (e.g., Iph=20%; Bremond et al., 2005b). To decide on which analytic rules to use, it is important to understand how the variation of confidence interval width and symmetry affects the interpretation of phytolith assemblages. Because the influence of index-specific count size and value of index differ when the index of interest is a ratio

(e.g., D/P) and a proportion (e.g., Iph), respectively, I will discuss these cases separately.

The ability to use Iph for vegetation reconstruction is influenced by index-specific count size and the value of the index. For assemblages with an Iph value close to 0% (or 100%), chance error associated with low index-specific count size (but ≥50) does not significantly alter vegetation inference. At an index-specific count of 100, the sampling error is as low as ±6%; at 200, it is at ±4%, providing a fairly precise interpretation of the abundance of short grasses. Moreover, the narrow C.I.s calculated for such low values of Iph ensures a consistent interpretation (correct or not), whether 50 or 2000 diagnostic phytoliths are counted, namely that the phytolith assemblage reflects a tall grass savanna (Iph<20%; Bremond et al., 2005b) (Fig. 2c). Note that, because 10% is fairly close to 0%, the exact confidence interval limits computed for the smaller count sizes are at best approximate; this applies irrespective of whether bootstrapping methods or analytic statistical formulas are used (Brown et al., 2001; Simon and Bruce, 1991; P. Bruce, 2004, personal communication).

When the Iph index is 30% (or 50%), the deviation due to sampling error is larger for all index-specific count sizes making the interpretation more dependent on the number of diagnostic specimens counted (Fig. 2a, Supplementary Table 2). Thus, at a count of 100 index-specific phytoliths, the sampling error is ±9%, and at 200, it is still high, ±6%. As a consequence, at least 100 index-specific phytoliths must be counted to ensure that the interpretation is unambiguously short-grass savanna (Iph>20%; Bremond et al., 2005b). For phytolith assemblages with values of Iph close to the boundary value of 20%, interpretation is particularly sensitive to index-specific count size. Even at index-specific counts of >>200, the sampling error is >±5% (corresponding to a standard error of roughly 3%) and a high index-specific

count is needed for unambiguous interpretations. For instance, for an assemblage with Iph=15%, an index-specific count of at least 200 (resulting in a total count of >800 phytoliths) is required to confidently classify it as this vegetation type (see below for further discussion).

The interpretation of D/P is also dependent on the value of the index in addition to index-specific count size, but the sampling errors behave differently than for Iph. At very high and very low values of the index (D/P>>1 and D/P<<1), the size of the sampling error relative to the D/P value itself increases dramatically (Fig. 1a,e, Supplementary Table 1). This tendency is particularly striking for high values of D/P, where it is coupled with a protracted asymptote of the C.I. curve (higher index-specific counts are needed to reach constant C.I. widths; Fig. 1a). Furthermore, the C.I.s for these "extreme" D/P ratios remain very wide relative to the D/P value even at very high index-specific count sizes. At D/P=9 the sampling error is ±2.1 at an index-specific count of 800 (well beyond the point where C.I. width stabilizes), corresponding to ±24%. In these cases, exact estimations of D/P values will simply not be possible. It follows that precise evaluation of the degree of tree cover reflected in an assemblage may also not be realistic. However, although the high (and low) D/P ratios are imprecise, it is possible to make unambiguous interpretations at relatively low index-specific count sizes because the range of values that define the vegetation types are also wide relative to the D/P values. For instance, for a typical forest assemblage with a D/P of about 3, a count as low as 75 diagnostic phytoliths will produce a 95% confidence interval of 1.85-5.41, falling above the defined D/P boundary value for closed forests (>1.82) (Bremond et al., 2005a). In contrast, for a count with a D/P close to one (e.g., D/P=0.9, average for phytolith assemblages from dense savanna), it is necessary to count >300 diagnostic morphotypes to ensure that

the 95% C.I.s falls within the interpretive boundaries of this vegetation category (0.65-1.16; Bremond et al., 2005a). Again, when interpretation hinges on a narrowly defined range of values, many more diagnostic morphotypes may have to be counted to achieve an adequate confidence interval width.

This is particularly true when the value of D/P approaches the interpretive limits of vegetation types. For example, for a forest assemblage with D/P=2, the assemblage composition estimate varies between 1.5 and 2.7 due to chance error at an index-specific count of 200. The lower end of this 95% confidence interval falls well below the range of the D/P values interpreted as reflecting a closed forest by Bremond et al. (2005b). It would likely not be interpreted as forest, but perhaps as some kind of forest-grassland intermediate, such as the *Margarita* ecotone (Bremond et al., 2005a). Indeed, a count size of 2000 diagnostic phytoliths is needed to ensure that the 95% confidence interval falls above this interpretive boundary.

The possible risks of not making certain that index-specific count sizes are adequate for the index being used for vegetation reconstruction are illustrated by previously published work using D/P and Iph (Figs. 3,4) (Bremond et al., 2005b). Calculations of 95% C.I.s for the Iph indices reported in Bremond et al. (2005b) showed that, in several cases, a combination of low index-specific counts (<<100 Chloridoideae + Panicoideae phytoliths) and narrow interpretive boundaries results in C.I.s that encompass Iph values typical of both short grass-dominated vs. tall grass-dominated vegetation. Had these been fossil samples with unknown source vegetation, there is a significant risk that they would have been misclassified simply due to chance error. On the other hand, the 95% C.I.s calculated for the D/P values of the two forest samples (S.155, 83-151) in Bremond et al. (2005b) demonstrates how even very wide C.I.s can sometimes be adequate for a robust assemblage

interpretation, as long as the interpretative boundaries are generous (Fig. 3b). These differences in the sensitivity of the analysis linked to count size and interpretative rules emphasize the necessity to determine satisfactory count sizes on a case-by-case basis. This is especially important in modern analogue studies attempting to devise guidelines for analysis of fossil phytolith samples (e.g., Barboni et al., 2007; Bremond et al., 2005b).

4.1. Statistical tests

Whereas the width of the confidence interval may be of importance for precision of an estimated ratio/proportion, another perhaps more important concern is that different interpretations of a sample are mutually exclusive. That is, the count size has to be large enough and the associated confidence interval narrow enough that an alternative interpretation can be ruled out, or that the difference between two samples can be detected ("event rejection"; Jiroutek et al., 2003).

Appropriate count sizes for these types of problems can be ascertained through statistical hypothesis testing, which also has the benefit of being less theoretically ambiguous than C.I.s (e.g., Freedman et al., 1998). Bootstrapping is a simple way to accomplish this for both the conditional and unconditional case, and is recommended for ratios such as D/P to account for their skewed C.I. shape. However, for conditional inference (at reasonable index-specific count sizes, see above), various analytic statistical methods are also available (see e.g., Freedman et al., 1998; Sokal and Rohlf, 1981).

The first case of event rejection is when a sample is compared to an alternative interpretation which has been determined *a priori* based on, for example, modern

analogue work. For instance, one may ask: what is the probability that a phytolith count with an D/P-ratio of 0.7 does not in reality derive from a soil phytolith assemblage from a densely wooded savanna, but from an open savanna (an assemblage with D/P<0.5)? In other words, how likely is it that chance error related to count size is throwing off the paleovegetational interpretation? This question can be investigated by performing a one-sided test, in which the null hypothesis is that D/P=0.5 (or less, but D/P=0.5 represents the null hypothesis that is closest to the sampled value of D/P). The analysis can be performed for either the unconditional or conditional case by means of bootstrapping. Thus, unconditional inference via bootstrapping, with ND=50% and α=0.025, shows that a count of 150 diagnostic phytoliths is sufficient (p-value is consistently <0.025) to reject the null hypothesis that the sample represents open savanna vegetation. In Fig. 1d, the distinction is illustrated by a lack of overlap between the 95% confidence interval for D/P=0.5 and a line representing D/P=0.7; at lower counts, there is a clear overlap. The hypothesis can also be tested using analytic statistics, by means of a Chi-square test (conditional case) or a one-sample t-test (conditional or unconditional case) (e.g., Freedman et al., 1998; Sokal and Rohlf, 1981; Zar, 1999). Note that all these tests, including bootstrapping, require reasonably large index-specific count sizes and index values that are not "extreme" to give precise estimates (D. Freedman, 2007, University of California, Berkeley, personal communication). "Extreme" values for D/P are close to 0 or infinity and for Iph close to 0 or 100%.

One might also have two samples with different observed D/P ratios (e.g., 0.5 and 1, respectively), and be interested in knowing what the probability is that they derive from soil phytolith assemblages with identical ratios of ligneous dicotyledon phytoliths to Poaceae phytoliths. This type of event rejection tests might be relevant

for comparisons between samples hypothesized to derive from vegetation growing under different climates or under varied human disturbance (e.g., Piperno and Becker, 1996; Scott, 2002). The null hypothesis could be that the "true" D/P ratio is intermediate between the two observed values, in this case that D/P is approximately 0.75. A one-tailed test for the difference in means, using unconditional inference (bootstrapping; ND=50%; α=0.025) shows that an index-specific count on the order of 140 for each sample is sufficient to reject the null hypothesis (p-value consistently <0.025). That is, it demonstrates that the samples likely contain different ratios of ligneous dicotyledon phytoliths to Poaceae phytoliths. Analytic statistical formulas can also be used for these tests, for example Fisher's exact test (conditional case) or two sample t-tests (conditional or unconditional case) (e.g., Freedman et al., 1998; Sokal and Rohlf, 1981; Zar, 1999). The same reservations regarding count size as stated above apply for this kind of test (D. Freedman, 2007, University of California, Berkeley, personal communication).

These comparisons confirm the statements of previous authors that different problems require different count sizes (e.g., Jiroutek et al., 2003; Piperno, 1988; Rull, 1987), with event rejection requiring somewhat smaller count sizes than if the aim is a particular confidence interval width (or both, Jiroutek et al., 2003).

4.2. General implications for phytolith analysis

Although vegetation analysis or taxonomic differentiation based on quantitative considerations are often supplemented with qualitative indices, such as presence of key indicator taxa (e.g., *Trichomanes*, Piperno, 1993), paleoecological analysis using phytoliths is fundamentally a quantitative exercise. Consequently, the findings

presented herein should be kept in mind. Close attention to count size is especially important for vegetation or climate indices with narrowly defined threshold values [D/P, Ic, Iph, fan-shaped index (Fs); e.g., Barboni et al., 1999; Bremond et al., 2005b]. In these cases, unambiguous interpretation requires a very high degree of statistical precision. It may be argued that quick-scanning, which is often used to confirm a vegetation pattern established from a smaller number of specimens (e.g., Pearsall, 2000), would alleviate the need for high counts of diagnostic specimens. However, eyeballing is effective only for samples with very clear distribution of morphotypes (with e.g., 90% grass phytoliths); it is less helpful in more unclear cases, such as for assemblages with a D/P close to one.

Considering the trade-off between high count sizes and scanning time, it should be of interest to phytolith workers to determine the count size that will ensure statistically significant results. At the very least, we should strive to put standard errors or confidence intervals on counts or otherwise provide statistical power to observed patterns and interpretations. As even complex assemblages can be reasonably modeled using resampling, the ardor and complex assumptions of analytic statistics that long hampered palynologists (see Maher, 1972 for discussion) should no longer stand in the way. The flexibility of numeric methods allows count size to be determined on a case-by-case basis, as recommended by Piperno (1988), whether dealing with questions requiring the use of confidence intervals or hypothesis testing.

However, the current study also suggests some general guidelines for phytolith analysis that require statistical precision in the form of confidence intervals, or a combination of confidence intervals and event rejection. Thus, it seems that an index-specific count of 200 phytoliths is a reasonable minimum number to aim for. When this count is reached, if the assemblage appears to be clearly dominated by one

morphotype class over another (e.g., has an D/P value close to 0), it is likely safe to stop counting and just quick-scan the slide for rare morphotypes. The robustness of the pattern could be checked using, for example, bootstrapping methods. Note that quick-scanning can be done on a more qualitative basis, or the counts of rare diagnostic morphotypes can be incorporated into the total count by calculating a ratio between the rare morphotypes and one of the more abundant morphotypes (see e.g., Lytle and Wahl, 2005). By contrast, if the D/P value is closer to 1, it might be necessary to count in excess of 200 diagnostic phytoliths. The general rule is that quadrupling of the sample size of diagnostic specimens cuts the standard error in half (Freedman et al., 1998). Alternatively, the vegetation interpretation must be made more conservative or general, to allow for some chance error. For example, a sample with an observed D/P ratio of 0.67 at an index-specific count of 200 can only safely be interpreted as either a densely wooded savanna or more open savanna (95% confidence interval=0.5-0.9). I took this more general approach in two recent studies on Cenozoic vegetation change in the Great Plains of North America (Strömberg, 2005) and Turkey and surrounding areas (Strömberg et al., 2007). The interpretation of vegetation was fundamentally based on the *relative change* in abundance of phytoliths from forest indicator taxa and from open-habitat grasses (see Strömberg, 2004), respectively, rather than on fixed proportions. Although the individual habitat descriptions were by necessity fairly vague, the pattern of change is statistically robust, showing an unambiguous shift from relatively closed habitats to more open vegetation (Fig. 5) (Strömberg, 2005; Strömberg et al., 2007: Fig. 5).

Note that these guidelines are valid not just for use of indices in phytolith analysis, but also for calculating C.I.s for the relative frequency of individual morphotypes (it is the same as calculating Iph for the conditional case), and

interpretation of relative frequency diagrams (e.g., Runge, 1999). For determination of sufficient count sizes for more complicated quantitative applications, such as multivariate statistical analysis (e.g., Blinnikov et al., 2001), see for example Lytle and Wahl (2005).

5. Conclusions

In this study I evaluated what factors influence measures of statistical precision, in the form of confidence intervals, for two indices that are commonly used in vegetation and climate reconstruction, the tree cover index (D/P) and the aridity index (Iph). The factors investigated were count size, index value, and confidence interval calculation method (conventional analytic statistical methods and bootstrapping). The simulations showed that the factors that are most important for determining C.I. width are (1) the number of diagnostic phytoliths (index-specific count), and (2) the value of the index. In contrast, the total tally of phytoliths is not important for statistical power. C.I. shape depends on calculation method used, with bootstrapping producing more realistic, skewed C.I.s.

Because of this variation in statistical precision, the interpretation of an assemblage can vary widely simply due to chance error if an insufficient count of diagnostic phytoliths is made. Similarly, if too few diagnostic phytoliths have been tallied, a postulated difference between samples may not be statistically significant. A count of 200 diagnostic phytoliths appears to be a good starting point, but it is not necessarily appropriate or sufficient. This is because what constitutes an adequate count size is influenced both by the question at hand and the observed assemblage

composition. Thus, analyses that demand precision in the form of narrow C.I.s for indices require higher tallies of diagnostic phytoliths than studies that seek simply to establish the difference between phytolith assemblages (event rejection). Also, vegetation inference is more statistically robust for assemblages with a clearly skewed morphotypes distribution (90% ligneous dicotyledons phytoliths vs. 10% Poaceae phytoliths) than for more evenly composed assemblages.

For these reasons, it is essential to determine the appropriate count size for each study individually (see also Piperno, 1988). As shown herein, bootstrapping is a simple way to do so. If high index-specific counts are not an option, bootstrapping methods are also suitable to evaluate the precision of a given interpretation. More generally, numeric approaches allow for the incorporation of statistics as a basic ingredient in phytolith analysis.

Acknowledgements

This work builds upon dissertation research conducted in the Department of Integrative Biology and Museum of Paleontology, University of California at Berkeley (UCB). I thank D. Freedman and P. Holroyd at UCB for discussions around statistics and P. Bruce at www.resampling.com for answering questions concerning the software used herein (Resampling Stats). I am grateful to D. Lindberg, C. D'Antonio, D. Kaplan (UCB), N.C. Arens (Hobart & William Smith Colleges), I. Rovner (Binary Analytical Consultants), and D. Piperno (Smithsonian Institution and Smithsonian Tropical Research Institute) for comments on an early version of this study. I thank G. Wilson (Denver Museum of Nature and Science) for improving the manuscript. My dissertation work was funded by research grants from the Department of Integrative Biology (Summer Fellowships), UCMP, von Beskow's Fund, Royal

Academy of Science (Sweden), Geological Society of America, Paleontological Society of America, Sigma Xi, and National Science Foundation (Dissertation Improvement Grant DEB-1-0104975).

References

Albert, R.M. and Weiner, S., 2001. Study of phytoliths in prehistoric ash layers from Kebara and Tabun caves using a quantitative approach. In: J.D. Meunier and F. Colin (Eds.), Phytoliths: applications in Earth sciences and human history. A.A. Balkema Publishers, Lisse, pp. 251-266.

Alexandre, A., Meunier, J.-D., Lezine, A.-M., Vincens, A. and Schwartz, D., 1997. Phytoliths: indicators of grassland dynamics during the late Holocene in intertropical Africa. Palaeogeography, Palaeoclimatology, Palaeoecology 136, 213-229.

Alexandre, A., Meunier, J.-D., Mariotti, A. and Soubies, F., 1999. Late Holocene phytolith and carbon-isotope record from a Latosol at Salitre, South-Central Brazil. Quaternary Research 51, 187-194.

Alroy, J., 2000. New methods for quantifying macroevolutionary patterns and processes. Paleobiology 26(4), 707-733.

Ball, T.B., Gardner, J.S. and Anderson, N., 1999. Identifying inflorescence phytoliths from selected species of wheat (*Triticum monococcum*, *T. dicoccon*, *T. dicoccoides*, and *T. aestivum*) and barley (*Hordeum vulgare* and *H. spontaneum*) (Graminae). American Journal of Botany 86(11), 1615-1623.

Ball, T.B., Gardner, J.S. and Brotherson, J.D., 1996. Identifying phytoliths produced by the inflorescence bracts of three species of wheat (*Triticum monococcum* L., *T. dicoccon* Schrank., and *T. aestivum* L.) using computer-assisted image and statistical analyses. Journal of Archaeological Science 23, 619-632.

Barboni, D., Bonnefille, R., Alexandre, A. and Meunier, J.-D., 1999. Phytoliths as paleoenvironmental indicators, West Side Middle Awash Valley, Ethiopia. Palaeogeography, Palaeoclimatology, Palaeoecology 152, 87-100.

Barboni, D., Bremond, L. and Bonnefille, R., 2007. Comparative study of modern phytolith assemblages from inter-tropical Africa. Palaeogeography Palaeoclimatology Palaeoecology 246, 454-470.

Birks, H.J.B. and Birks, H.H., 1980. Quaternary palaeoecology. Edward Arnold Ltd, London, 289 pp.

Blinnikov, M.S., Busacca, A. and Whitlock, C., 2001. A new 100,000-year phytolith record from the Columbia Basin, Washington, U.S.A. In: J.D. Meunier and F. Colin (Eds.), Phytoliths: applications in Earth sciences and human history. A.A. Balkema Publishers, Lisse, pp. 27-55.

Blinnikov, M.S., Busacca, A. and Whitlock, C., 2002. Reconstruction of the late Pleistocene grassland of the Columbia basin, Washington, USA, based on phytolith records in loess. Palaeogeography Palaeoclimatology Palaeoecology 177(1-2), 77-101.

Boyd, W.E., Lentfer, C.J. and Torrence, R., 1998. Phytolith analysis for a wet tropics environment: methodological issues and implications for the archaeology of Garua Island, West New Britain, Papa New Guinea. Palynology 22, 213-228.

Bremond, L., Alexandre, A., Hély, C. and Guiot, J., 2005a. A phytolith index as a proxy of tree cover density in tropical areas: Calibration with Leaf Area Index along a forest–savanna transect in southeastern Cameroon. Global and Planetary Change 45, 277-293.

Bremond, L., Alexandre, A., Peyron, O. and Guiot, J., 2005b. Grass water stress estimated from phytoliths in West Africa. Journal of Biogeography 32, 311-327.

Brown, D.A., 1984. Prospects and limits of a phytolith key for grasses in the Central United States. Journal of Archaeological Science 11, 345-368.

Brown, L.D., Cai, T.T. and DasGupta, A., 2001. *Interval estimation for a binomial proportion*. Statistical Science 16(2), 101-133.

Carnelli, A.L., Theurillat, J.-P. and Madella, M., 2004. Phytolith types and type-frequencies in subalpine–alpine plant species of the European Alps. Review of Palaeobotany and Palynology 129, 39-65.

Carter, J.A., 2000. Phytoliths from loess in Southland, New Zealand. New Zealand Journal of Botany 38, 325-332.

Chiswell, C.E., 1984. A study of prehistoric Andean storage by means of phytolith analysis. M.A. Thesis, Trenton University, Petersborough, Canada.

Cournac, L., Dubois, M.-A., Chave, J. and Riéra, B., 2002. Fast determination of light availability and leaf area index in tropical forests. Journal of Tropical Ecology 18, 295- 302.

Diester-Haass, L., Schrader, H.J. and Thiede, J., 1973. Sedimentological and paleoclimatological investigations of two pelagic ooze cores off Cape Barbas, North-West Africa. "Meteor" Forschungsergebnisse C 16, 19-66.

Dinan, E.H. and Rowlett, R.M., 1993. Vegetational changes at the Shriver paleo-Indian site, N. W. Missouri: Phytolith analysis as an aid in environmental reconstruction. In: D.M. Pearsall and D.R. Piperno (Eds.), Current Research in Phytolith Analysis: Applications in Archaeology and Paleoecology. MASCA Research Papers in Science and Archaeology. University Museum of

Archaeology and Anthropology, University of Pennsylvania, Philadelphia, pp. 73-82.

Efron, B., 1979. Bootstrap Methods: Another Look at the Jackknife. Annals of Statistics 7, 1-26.

Faegri, K., Kaland, P.E. and Krzywinski, K., 1989. Textbook of pollen analysis. Wiley, Chichester.

Falcon-Lang, H.J., 2003. Late Carboniferous tropical dryland vegetation in an alluvial-plain setting, Joggins, Nova Scotia, Canada. Palaios 18, 197-211.

Fearn, M.L., 1998. Phytoliths in sediment as indicators of grass pollen source. Review of Palaeobotany and Palynology 103(1998), 75-81.

Foote, M., 1992. Rarefaction analysis of morphological and taxonomic diversity. Paleobiology 18, 1-16.

Fredlund, G.G., 1986. Problems in the simultaneous extraction of pollen and phytoliths from clastic sediments. In: I. Rovner (Ed.), Plant opal phytolith analysis in archaeology and paleoecology. Proceedings of the 1984 phytolith research workshop. Occasional Papers of the Phytolitharien. North Carolina State University, Raleigh, North Carolina, pp. 102-111.

Fredlund, G.G. and Tieszen, L.L., 1997. Calibrating grass phytolith assemblages in climatic terms: Application to late Pleistocene assemblages from Kansas and Nebraska. Palaeogeography, Palaeoclimatology, Palaeoecology 136, 199-211.

Freedman, D., Pisani, R. and Purves, R., 1998. Statistics. W.W. Norton & Company, New York, 578 pp.

Fujiwara, H., 1993. Research into the history of rice cultivation using plant opal analysis. In: D.M. Pearsall and D.R. Piperno (Eds.), Current research in phytolith analysis: applications in archaeology and paleoecology. MASCA

Research Papers in Science and Archaeology. University Museum of Archaeology and Anthropology, University of Pennsylvania, Philadelphia, pp. 147158.

Good, P., 2006. Resampling methods. Birkhauser.

ICPN Committee 1.0: Madella, M., Alexandre, A. and Ball, T.B., 2002. International Code for Phytolith Nomenclature. The Phytolitharien. Bulletin of the Society for Phytolith Research 15(1), 7-16.

Jamniczky, H.A., Brinkman, D.B. and Russell, A.P., 2003. Vertebrate microsite sampling: how much is enough? Journal of Vertebrate Paleontology 23(4), 725-734.

Jiroutek, M.R., Muller, K.E., Kupper, L.L. and Stewart, P.W., 2003. A new method for choosing sample size for confidence interval-based inference. Biometrics 59, 580-590.

Kealhofer, L., 1996. The human environment during the terminal Pleistocene and Holocene in northeastern Thailand: phytolith evidence from Lake Kumphawapi. Asian Perspectives 35(2), 229-254.

Kondo, R., Childs, C. and Atkinson, I., 1994. Opal phytoliths of New Zealand. Manaaki Whenua Press, Canterbury, New Zealand, 85 pp.

Lentfer, C.J. and Boyd, W.E., 1999. An assessment of techniques for the deflocculation and removal of clays from sediments used in phytolith analysis. Journal of Archaeological Science 26(31-44).

Lytle, D.E. and Wahl, E.R., 2005. Palaeoenvironmental reconstructions using the modern analogue technique: the effects of sample size and decision rules. The Holocene 15(4), 554-566.

MacDonald, G.M., 1996. Non-aquatic Quaternary. In: J. Jansonius and D.C. McGregor (Eds.), Palynology: principles and applications. Volume 2: Applications. American Association of Stratigraphic Palynologists Foundation. Publishers Press, Salt Lake City, Utah, pp. 879-910.

Maher, L.J., Jr., 1972. Nomograms for computing 0.95 confidence intervals of pollen data. Review of Palaeobotany and Palynology 13, 85-93.

Mooney, C.Z. and Duval, R.D., 1993. Bootstrapping: a non-parametric approach to statistical inference. Sage Quantitative Applications in the Social Science Series, No. 95. Sage Publications, Newbury Park, CA.

Mosimann, J.E., 1965. Statistical methods for the pollen analyst: multinomial and negative multinomial techniques. In: B. Kummel and D. Raup (Eds.), Handbook of Paleontological Techniques. W. Freeman, San Francisco, pp. 636-673.

Pearsall, D.M., 1978. Phytolith analysis of archaeological soils: evidence for maize cultivation in formative Ecuador. Science 199, 177-178.

Pearsall, D.M., 1982. Maize phytoliths: a clarification. Phytolitharien Newsletter 1(2), 3.

Pearsall, D.M., 2000. Paleoethnobotany: A Handbook of Procedures. Academic Press, San Diego, 470 pp.

Pearsall, D.M. and Piperno, D.R., 1990. Antiquity of maize cultivation in Ecuador: summary and reevaluation of the evidence. American Antiquity 55(2), 324-337.

Piperno, D.R., 1984. A comparison and differentiation of phytoliths from maize and wild grasses: use of morphological criteria. American Antiquity 49(2), 361-383.

Piperno, D.R., 1988. Phytolith Analysis, an Archaeological and Geological Perspective. Academic Press, San Diego.

Piperno, D.R., 1993. Phytolith and charcoal records from deep lake cores in the American tropics. In: D.M. Pearsall and D.R. Piperno (Eds.), Current research in phytolith analysis: applications in archaeology and paleoecology. MASCA Research Papers in Science and Archaeology. University Museum of Archaeology and Anthropology, University of Pennsylvania, Philadelphia, pp. 58-71.

Piperno, D.R., 2006. Phytoliths: a comprehensive guide for archaeologists and paleoecologists AltaMira Press, New York, 304 pp.

Piperno, D.R. and Becker, P., 1996. Vegetational history of a site in the central Amazon basin derived from phytolith and charcoal records from natural soils. Quaternary Research 45, 202-209.

Piperno, D.R. and Pearsall, D.M., 1998. The silica bodies of tropical American grasses: Morphology, taxonomy, and implications for grass systematics and fossil phytolith identification. Smithsonian Contributions to Botany 85, 1-40.

Piperno, D.R. and Stothert, K.E., 2003. Phytolith evidence for early Holocene *Cucurbita* domestication in Southwest Ecuador. Science 299(5609), 1054-1057.

Prebble, M. and Shulmeister, J., 2002. An analysis of phytolith assemblages for the quantitative reconstruction of late Quaternary environments of the Lower Taieri Plain, Otago, South Island, New Zealand II. Paleoenvironmental reconstruction. Journal of Paleolimnology 27(4), 415-427.

Raup, D., 1975. Taxonomic diversity estimation using rarefaction. Paleobiology 1, 333-342.

Rovner, I., 1971. Potential of opal phytoliths for use in paleoecological reconstruction. Quaternary Research 1, 343-359.

Rovner, I., 2001. Cultural behavior and botanical history: phytolith analysis in small places and narrow intervals. In: J.D. Meunier and F. Colin (Eds.), Phytoliths: applications in Earth sciences and human history. A.A. Balkema Publishers, Lisse, pp. 119-127.

Rull, V., 1987. A note on pollen counting in palaeoecology. Pollen et spores 29(4), 471-480.

Runge, F., 1999. The opal phytolith inventory of soils in central Africa - quantities, shapes, classification, and spectra. Review of Palaeobotany and Palynology 107, 23-53.

Scott, L., 2002. Grassland development under glacial and interglacial conditions in southern Africa: review of pollen, phytolith and isotope evidence. Palaeogeography Palaeoclimatology Palaeoecology 177(1-2), 47-57.

Shulmeister, J., Shane, P., Lian, O.B., Okuda, M., Carter, J.A., Harper, M., Dickinson, W., Augustinus, P. and Heijnis, H., 2001. A long late-Quaternary record from Lake Poukawa, Hawke's Bay, New Zealand. Palaeogeography Palaeoclimatology Palaeoecology 176, 81-107.

Simon, J.L., 1997. Resampling: The New Statistics. Duxbury Press, Belmont, CA, 436 pp.

Simon, J.L. and Bruce, P., 1991. Resampling: a tool for everyday statistics. Chance - New Directions For Statistics And Computing 4(1), 22-32.

Sokal, R.R. and Rohlf, F.J., 1981. Biometry. The principles and practice of statistics in biological research. W. H. Freeman & Company, New York, 859 pp.

Stanton, R.J., Jr and Evans, I., 1972. Community structure and sampling requirements in paleontology. Journal of Paleontology 46, 845-858.

Strömberg, C.A.E., 2002. The origin and spread of grass-dominated ecosystems in the Late Tertiary of North America: Preliminary results concerning the evolution of hypsodonty. Palaeogeography Palaeoclimatology Palaeoecology 177(1-2), 59-75.

Strömberg, C.A.E., 2003. The origin and spread of grass-dominated ecosystems during the Tertiary of North America and how it relates to the evolution of hypsodonty in equids. Ph.D. Thesis, University of California, Berkeley, 779 pp.

Strömberg, C.A.E., 2004. Using phytolith assemblages to reconstruct the origin and spread of grass-dominated habitats in the Great Plains during the late Eocene to early Miocene. Palaeogeography Palaeoclimatology Palaeoecology 207(3-4), 239-275.

Strömberg, C.A.E., 2005. Decoupled taxonomic radiation and ecological expansion of open-habitat grasses in the Cenozoic of North America. Proceedings of the National Academy of Sciences of the Unites States of America 102(34), 11980-11984.

Strömberg, C.A.E., 2007. Can slide preparation methods cause size biases in phytolith assemblages?: Results from a preliminary study. In: M. Madella and D. Zurro (Eds.), Plants, People and Places: Recent Studies in Phytolithic Analysis. Oxbow Books.

Strömberg, C.A.E., Werdelin, L., Friis, E.M. and Saraç, G., 2007. The spread of grass-dominated habitats in Turkey and surrounding areas during the

Cenozoic: phytolith evidence. Palaeogeography, Palaeoclimatology, Palaeoecology 250, 18-49.

Twiss, P.C., Dort, W., Jr. and Sorenson, C.J., 1987. Grass-opal phytoliths as climatic indicators of the Great Plains Pleistocene. Quaternary environments of Kansas, Guidebook Series Kansas Geological Survey 5, 179-188.

Vincens, A., Dubois, M.-A., Guillet, B., Achoundong, G., Buchet, G., Kamgang Kabeyene Beyala, V., de Namur, C. and Riera, B., 2000. Pollen-rain–vegetation relationships along a forest–savanna transect in southeastern Cameroon. Review of Palaeobotany and Palynology 110(2000), 191-208.

Wing, S.L., Alroy, J. and Hickey, L.J., 1995. Plant and animal diversity in the Paleocene to early Eocene of the Bighorn Basin. Palaeogeography Palaeoclimatology Palaeoecology 115(117-156).

Wolff, R.G., 1975. Sampling and sample size in ecological analyses of fossil mammals. Paleobiology 1, 195-204.

Zar, J.H., 1999. Biostatistical analysis. Prentice Hall, Upper Saddle River, New Jersey, 663 pp.

Zhao, Z.J., Pearsall, D.M., Benfer, R.A. and Piperno, D.R., 1998. Distinguishing rice (*Oryza sativa* Poaceae) from wild *Oryza* species through phytolith analysis, II - Finalized method. Economic Botany 52(2), 134-145.

Zurro, D. and Madella, M., this issue. Statistical significance of phytolith assemblages. Quaternary International.

Tables:

Table 1. Cases and methods tested in this study.

Index	Definition [a]	Index-specific morphotypes [b]	Non-index specific morphotypes (ND)	Index value	Analytic statistical formula [c]		Bootstrapping	
					Conditional case	Unconditional case	Conditional case	Unconditional case
D/P	D/P; where D=Ligneous dicotyledon phytoliths and P=Poaceae phytoliths	D=globular granulates; P=grass silica short cells (bilobate, cross, saddle, rondel, crenate)+cuneiform bulliform cell+acicular hair cell	Palm phytoliths (globular echinate); various phytoliths from forest indicator taxa (see Strömberg, 2004); sedge phytoliths (epidermal plate with cone-shapes), phytoliths that are produced by various plants (e.g., elongate psilate), various unknown morphotypes, etc.	0.1; 0.5; 1; 2; 9	C.I. = ±1.96*SE; SE= SQRT((D/P)^2 * (1/(D*P)) * ($1/n_i$))	C.I. = ±1.96*SE; SE= SQRT((D/P)^2 * (1/(D*P) * (1/N) * (D + P))	X	X (ND=50%)
Iph	Chlor/(Chlor+Pan); where Chlor=Chloridoideae phytoliths and Pan=Panicoideae phytoliths	Chlor=saddle; Pan=bilobate+cross	All non-Chloridoideae and Panicoideae phytoliths	10%; 20%; 30%	C.I.= ±1.96*SE; SE= 100 * (SQRT(((1-(Iph/100)) * (Iph/100))/n_i))	n/a	X	X (ND=70%; ND=90%)

[a] Definitions for D/P and Iph taken from Bremond et al. (2005a, 2005b).

[b] Morphotypes as far as possible described using the International Code for Phytolith Nomenclature (ICPN Committee 1.0: Madella et al., 2002).

[c] Formulas from Zar, 1999; D. Freedman, 2007, University of California, Berkeley, personal communication. SE = standard error; n_i = index-specific count; N = total count (diagnostic + non-diagnostic); SQRT = square root; ^ = power of. For other abbreviations, see text.

Figure captions:

Fig. 1. The influence of count size on the sampling error for the tree cover index, D/P, estimated by 95% confidence intervals (C.I.s). Above a certain index-specific count size, bootstrapping simulation produces C.I.s of roughly equal width as analytic statistical formulas and for both conditional and unconditional cases. However, the count size necessary to achieve similar C.I. width increases with higher values of D/P. Also, whereas analytically estimated C.I.s are symmetrical around the D/P value, bootstrap C.I.s are skewed towards higher values. **a.** D/P=9. **b.** D/P=2. **c.** D/P=1. **d.** D/P=0.5. **e.** D/P=0.1. Note that the scale of the Y-axis (D/P) varies among images. See Supplementary Table 1 and text for explanation.

Fig. 2. The influence of count size on the sampling error for the aridity index, Iph, estimated by 95% C.I.s. C.I.s calculated using analytic statistical formulas (conditional case) and bootstrapping (unconditional case) have similar width above the lowest index-specific count sizes (>10). The bootstrapped C.I.s for the conditional case are not shown, but are very similar to other bootstrapped C.I.s (see Supplementary Table 2). As for the D/P index, analytically estimated C.I.s are symmetrical around the Iph value, and bootstrap C.I.s are skewed towards higher values of Iph close to 0% (the opposite is true of values of Iph close to 100%). **a.** Iph=30%. **b.** Iph=20%. **c.** Iph=10%. Note that the scale of the Y-axis (Iph) varies among images. See Supplementary Table 2 and text for explanation.

Fig. 3. Tree cover index (D/P) values in West African phytolith assemblages from Bremond et al. (2005b) with 95% confidence intervals calculated using analytic formulas (conditional case; see Table 1) and bootstrapping (unconditional case). The vegetation type observed at the sampled site is noted above the sample values. **a.** Samples from savanna and steppe vegetation. Samples S.91 and 83-127 have C.I.s that are wide enough that their interpretation is ambiguous. **b.** The two samples from semi-deciduous forest have very wide C.I.s but can be unambiguously interpreted, because their C.I. limits fall well above the defined D/P boundary value for forest. See text for further explanation.

Fig. 4. Aridity index (Iph) values in West African phytolith assemblages from Bremond et al. (2005b) with 95% confidence intervals calculated using analytic formulas (conditional case; see Table 1) and bootstrapping (unconditional case). The vegetation type observed at the sampled site is noted above the sample values. Several Iph values have C.I.s that are wide enough that their interpretation is ambiguous. Note that sample S.155 from semi-deciduous forest is not an example of this despite its very wide C.I.s due to low index-specific counts (n_i=<20); this sample has very high D/P value (Fig. 3b) and would unambiguously be interpreted as forest. See text for further explanation.

Fig. 5. Habitat openness estimates (FI-t ratio) in Cenozoic phytolith assemblages from Nebraska/eastern Wyoming with 95% confidence intervals calculated using bootstrapping (unconditional case). Samples are arranged roughly in stratigraphical order: NE1-NE5 = Late Eocene; NE7-NE21 = Oligocene; NE22-NE52 = Miocene. Younger assemblages have FI-t ratios around or below 50%, indicating a shift

towards more open vegetation, such as savanna or woodland. The FI-t ratio is calculated as the proportion of phytoliths indicative of forest indicator taxa (e.g., woody and herbaceous dicotyledons, palms, conifers, ferns) and the sum of these phytoliths and grass silica short cells (FI-t ratio) (Strömberg, 2005; Strömberg et al., 2007). See text for further explanation.

Supplementary table captions:

Supplementary Table 1. The influence of count size on sampling error for the tree cover index, the D/P ratio [a].

[a] Sampling error estimated as half of the 95% C.I. width. n_i = index-specific count; N = total count; SE= standard error. See Table 1 for analytic formulas used and text for further explanation.

Supplementary Table 2. The influence of count size on sampling error for the aridity index, Iph [a].

[a] Sampling error estimated as half of the 95% C.I. width. n_i = index-specific count; N = total count; SE= standard error. See Table 1 for analytic formulas used and text for further explanation.

Supplementary Table 1. The influence of count size on sampling error for the tree cover index, the D/P ratio [a]. Strömberg

		Analytic statistical formulas									Bootstrapping								
		Conditional case (ND=0%)				Unconditional case (ND=50%)					Conditional case (ND=0%)				Unconditional case (ND=50%)				
D/P	n_i	$D/P_{95\%min}$	$D/P_{95\%max}$	Sampling error (±1.96SE)	Sampling error (±%)	N	$D/P_{95\%min}$	$D/P_{95\%max}$	Sampling error (±1.96SE)	Sampling error (±%)	$D/P_{95\%min}$	$D/P_{95\%max}$	Sampling error (±)	Sampling error (±%)	N	$D/P_{95\%min}$	$D/P_{95\%max}$	Sampling error (±)	Sampling error (±%)
0,1	10	-0,12	0,32	0,22	215,5	20	-0,12	0,32	0,22	215,5	0,00	0,43	0,21	214,3	20	0,00	0,50	0,25	250,0
0,1	25	-0,04	0,24	0,14	136,3	50	-0,04	0,24	0,14	136,3	0,00	0,25	0,13	125,0	50	0,00	0,28	0,14	138,9
0,1	50	0,00	0,20	0,10	96,4	100	0,00	0,20	0,10	96,4	0,02	0,22	0,10	99,6	100	0,02	0,22	0,10	97,9
0,1	75	0,02	0,18	0,08	78,7	150	0,02	0,18	0,08	78,7	0,03	0,19	0,08	81,5	150	0,03	0,19	0,08	81,0
0,1	100	0,03	0,17	0,07	68,1	200	0,03	0,17	0,07	68,1	0,04	0,18	0,07	67,4	200	0,04	0,18	0,07	68,5
0,1	150	0,04	0,16	0,06	55,6	300	0,04	0,16	0,06	55,6	0,05	0,16	0,06	56,9	300	0,05	0,16	0,06	55,1
0,1	200	0,05	0,15	0,05	48,2	400	0,05	0,15	0,05	48,2	0,06	0,16	0,05	48,9	400	0,06	0,15	0,05	48,9
0,1	250	0,06	0,14	0,04	43,1	500	0,06	0,14	0,04	43,1	0,06	0,15	0,04	43,7	500	0,06	0,15	0,04	43,2
0,1	300	0,06	0,14	0,04	39,3	600	0,06	0,14	0,04	39,3	0,06	0,14	0,04	38,4	600	0,06	0,14	0,04	39,2
0,1	350	0,06	0,14	0,04	36,4	700	0,06	0,14	0,04	36,4	0,07	0,14	0,04	36,5	700	0,07	0,14	0,04	36,9
0,1	400	0,07	0,13	0,03	34,1	800	0,07	0,13	0,03	34,1	0,07	0,14	0,03	33,4	800	0,07	0,14	0,03	34,2
0,1	450	0,07	0,13	0,03	32,1	900	0,07	0,13	0,03	32,1	0,07	0,13	0,03	32,3	900	0,07	0,13	0,03	32,2
0,1	500	0,07	0,13	0,03	30,5	1000	0,07	0,13	0,03	30,5	0,07	0,13	0,03	31,6	1000	0,07	0,13	0,03	31,3
0,1	550	0,07	0,13	0,03	29,1	1100	0,07	0,13	0,03	29,1	0,07	0,13	0,03	29,2	1100	0,07	0,13	0,03	29,6
0,1	600	0,07	0,13	0,03	27,8	1200	0,07	0,13	0,03	27,8	0,07	0,13	0,03	28,3	1200	0,07	0,13	0,03	27,6
0,1	650	0,07	0,13	0,03	26,7	1300	0,07	0,13	0,03	26,7	0,07	0,13	0,03	27,0	1300	0,07	0,13	0,03	26,7
0,1	700	0,07	0,13	0,03	25,8	1400	0,07	0,13	0,03	25,8	0,08	0,13	0,03	25,1	1400	0,08	0,13	0,03	25,5
0,1	750	0,08	0,12	0,02	24,9	1500	0,08	0,12	0,02	24,9	0,08	0,13	0,03	25,0	1500	0,08	0,13	0,03	25,3
0,1	800	0,08	0,12	0,02	24,1	1600	0,08	0,12	0,02	24,1	0,08	0,13	0,02	23,5	1600	0,08	0,13	0,02	23,9
0,1	850	0,08	0,12	0,02	23,4	1700	0,08	0,12	0,02	23,4	0,08	0,12	0,02	23,5	1700	0,08	0,12	0,02	23,2
0,1	900	0,08	0,12	0,02	22,7	1800	0,08	0,12	0,02	22,7	0,08	0,12	0,02	22,9	1800	0,08	0,12	0,02	22,9
0,1	950	0,08	0,12	0,02	22,1	1900	0,08	0,12	0,02	22,1	0,08	0,12	0,02	21,7	1900	0,08	0,12	0,02	21,9
0,1	1000	0,08	0,12	0,02	21,6	2000	0,08	0,12	0,02	21,6	0,08	0,12	0,02	21,2	2000	0,08	0,12	0,02	21,6
0,1	1500	0,08	0,12	0,02	17,6	3000	0,08	0,12	0,02	17,6	0,08	0,12	0,02	17,8	3000	0,08	0,12	0,02	17,7
0,1	2000	0,08	0,12	0,02	15,2	4000	0,08	0,12	0,02	15,2	0,09	0,12	0,02	15,3	4000	0,09	0,12	0,02	15,5
0,5	10	-0,16	1,16	0,66	131,5	20	-0,16	1,16	0,66	131,5	0,11	1,50	0,69	138,9	20	0,00	1,75	0,88	175,0
0,5	25	0,08	0,92	0,42	83,2	50	0,08	0,92	0,42	83,2	0,19	1,08	0,45	89,3	50	0,17	1,17	0,50	100,1
0,5	50	0,21	0,79	0,29	58,8	100	0,21	0,79	0,29	58,8	0,25	0,85	0,30	60,2	100	0,26	0,88	0,31	62,3
0,5	75	0,26	0,74	0,24	48,0	150	0,26	0,74	0,24	48,0	0,29	0,79	0,25	49,3	150	0,29	0,77	0,24	47,8
0,5	100	0,29	0,71	0,21	41,6	200	0,29	0,71	0,21	41,6	0,33	0,75	0,21	42,1	200	0,32	0,74	0,21	42,3
0,5	150	0,33	0,67	0,17	34,0	300	0,33	0,67	0,17	34,0	0,35	0,69	0,17	33,4	300	0,35	0,68	0,17	33,5
0,5	200	0,35	0,65	0,15	29,4	400	0,35	0,65	0,15	29,4	0,36	0,67	0,15	30,6	400	0,36	0,66	0,15	29,7
0,5	250	0,37	0,63	0,13	26,3	500	0,37	0,63	0,13	26,3	0,38	0,64	0,13	26,4	500	0,37	0,64	0,14	27,1
0,5	300	0,38	0,62	0,12	24,0	600	0,38	0,62	0,12	24,0	0,39	0,62	0,12	23,3	600	0,39	0,62	0,12	23,3
0,5	350	0,39	0,61	0,11	22,2	700	0,39	0,61	0,11	22,2	0,40	0,62	0,11	22,0	700	0,40	0,62	0,11	22,0
0,5	400	0,40	0,60	0,10	20,8	800	0,40	0,60	0,10	20,8	0,41	0,61	0,10	20,0	800	0,40	0,61	0,10	20,8
0,5	450	0,40	0,60	0,10	19,6	900	0,40	0,60	0,10	19,6	0,40	0,60	0,10	20,0	900	0,41	0,60	0,10	19,0
0,5	500	0,41	0,59	0,09	18,6	1000	0,41	0,59	0,09	18,6	0,41	0,60	0,10	19,0	1000	0,41	0,60	0,10	19,0
0,5	550	0,41	0,59	0,09	17,7	1100	0,41	0,59	0,09	17,7	0,41	0,59	0,09	18,0	1100	0,42	0,59	0,09	17,7
0,5	600	0,42	0,58	0,08	17,0	1200	0,42	0,58	0,08	17,0	0,43	0,59	0,08	16,1	1200	0,42	0,58	0,08	16,0
0,5	650	0,42	0,58	0,08	16,3	1300	0,42	0,58	0,08	16,3	0,42	0,58	0,08	15,9	1300	0,42	0,59	0,08	16,6
0,5	700	0,42	0,58	0,08	15,7	1400	0,42	0,58	0,08	15,7	0,43	0,58	0,08	15,2	1400	0,43	0,58	0,08	15,1
0,5	750	0,42	0,58	0,08	15,2	1500	0,42	0,58	0,08	15,2	0,43	0,58	0,07	14,9	1500	0,42	0,58	0,08	15,2
0,5	800	0,43	0,57	0,07	14,7	1600	0,43	0,57	0,07	14,7	0,43	0,58	0,07	14,7	1600	0,43	0,58	0,08	15,7
0,5	850	0,43	0,57	0,07	14,3	1700	0,43	0,57	0,07	14,3	0,43	0,58	0,07	14,7	1700	0,43	0,58	0,08	15,4
0,5	900	0,43	0,57	0,07	13,9	1800	0,43	0,57	0,07	13,9	0,43	0,57	0,07	14,0	1800	0,43	0,57	0,07	14,0
0,5	950	0,43	0,57	0,07	13,5	1900	0,43	0,57	0,07	13,5	0,43	0,57	0,07	14,0	1900	0,44	0,57	0,07	13,4
0,5	1000	0,43	0,57	0,07	13,2	2000	0,43	0,57	0,07	13,2	0,43	0,57	0,07	13,4	2000	0,44	0,57	0,06	12,8
0,5	1500	0,45	0,55	0,05	10,7	3000	0,45	0,55	0,05	10,7	0,45	0,56	0,06	11,0	3000	0,45	0,55	0,05	10,2
0,5	2000	0,45	0,55	0,05	9,3	4000	0,45	0,55	0,05	9,3	0,45	0,54	0,05	9,2	4000	0,45	0,55	0,05	9,3

Supplementary Table 1. (cont'd)

		Analytic statistical formulas									Bootstrapping								
		Conditional case (ND=0%)				Unconditional case (ND=50%)					Conditional case (ND=0%)				Unconditional case (ND=50%)				
D/P	n_i	$D/P_{95\%min}$	$D/P_{95\%max}$	Sampling error (±1.96SE)	Sampling error (±%)	N	$D/P_{95\%min}$	$D/P_{95\%max}$	Sampling error (±1.96SE)	Sampling error (±%)	$D/P_{95\%min}$	$D/P_{95\%max}$	Sampling error (±)	Sampling error (±%)	N	$D/P_{95\%min}$	$D/P_{95\%max}$	Sampling error (±)	Sampling error (±%)
1	10	-0,24	2,24	1,24	124,0	20	-0,24	2,24	1,24	124,0	0,25	4,00	1,88	187,5	20	0,22	4,50	2,14	213,9
1	25	0,22	1,78	0,78	78,4	50	0,22	1,78	0,78	78,4	0,47	2,13	0,83	82,7	50	0,44	2,29	0,92	92,4
1	50	0,45	1,55	0,55	55,4	100	0,45	1,55	0,55	55,4	0,56	1,78	0,61	60,8	100	0,57	1,78	0,61	60,6
1	75	0,55	1,45	0,45	45,3	150	0,55	1,45	0,45	45,3	0,63	1,59	0,48	47,8	150	0,63	1,59	0,48	48,1
1	100	0,61	1,39	0,39	39,2	200	0,61	1,39	0,39	39,2	0,67	1,50	0,42	41,7	200	0,67	1,49	0,41	41,1
1	150	0,68	1,32	0,32	32,0	300	0,68	1,32	0,32	32,0	0,72	1,38	0,33	32,8	300	0,72	1,37	0,33	32,6
1	200	0,72	1,28	0,28	27,7	400	0,72	1,28	0,28	27,7	0,75	1,33	0,29	28,6	400	0,76	1,32	0,28	28,1
1	250	0,75	1,25	0,25	24,8	500	0,75	1,25	0,25	24,8	0,77	1,29	0,26	26,0	500	0,78	1,27	0,25	24,6
1	300	0,77	1,23	0,23	22,6	600	0,77	1,23	0,23	22,6	0,78	1,27	0,25	24,9	600	0,80	1,27	0,23	23,3
1	350	0,79	1,21	0,21	21,0	700	0,79	1,21	0,21	21,0	0,82	1,24	0,21	21,3	700	0,80	1,24	0,22	21,8
1	400	0,80	1,20	0,20	19,6	800	0,80	1,20	0,20	19,6	0,82	1,22	0,20	20,2	800	0,82	1,21	0,20	19,5
1	450	0,82	1,18	0,18	18,5	900	0,82	1,18	0,18	18,5	0,84	1,20	0,18	17,9	900	0,83	1,20	0,19	18,6
1	500	0,82	1,18	0,18	17,5	1000	0,82	1,18	0,18	17,5	0,84	1,20	0,18	18,2	1000	0,83	1,19	0,18	18,1
1	550	0,83	1,17	0,17	16,7	1100	0,83	1,17	0,17	16,7	0,84	1,19	0,17	17,4	1100	0,84	1,18	0,17	17,0
1	600	0,84	1,16	0,16	16,0	1200	0,84	1,16	0,16	16,0	0,85	1,17	0,16	16,0	1200	0,85	1,17	0,16	16,0
1	650	0,85	1,15	0,15	15,4	1300	0,85	1,15	0,15	15,4	0,86	1,17	0,15	15,2	1300	0,86	1,16	0,15	15,2
1	700	0,85	1,15	0,15	14,8	1400	0,85	1,15	0,15	14,8	0,86	1,15	0,14	14,4	1400	0,86	1,15	0,14	14,5
1	750	0,86	1,14	0,14	14,3	1500	0,86	1,14	0,14	14,3	0,87	1,15	0,14	13,8	1500	0,87	1,16	0,15	14,7
1	800	0,86	1,14	0,14	13,9	1600	0,86	1,14	0,14	13,9	0,87	1,14	0,14	13,6	1600	0,87	1,16	0,14	14,4
1	850	0,87	1,13	0,13	13,4	1700	0,87	1,13	0,13	13,4	0,87	1,14	0,13	13,2	1700	0,87	1,14	0,14	13,6
1	900	0,87	1,13	0,13	13,1	1800	0,87	1,13	0,13	13,1	0,87	1,14	0,13	13,2	1800	0,88	1,14	0,13	12,8
1	950	0,87	1,13	0,13	12,7	1900	0,87	1,13	0,13	12,7	0,88	1,14	0,13	12,9	1900	0,89	1,14	0,13	12,6
1	1000	0,88	1,12	0,12	12,4	2000	0,88	1,12	0,12	12,4	0,88	1,14	0,13	12,9	2000	0,88	1,13	0,12	12,3
1	1500	0,90	1,10	0,10	10,1	3000	0,90	1,10	0,10	10,1	0,90	1,11	0,10	10,5	3000	0,90	1,11	0,10	10,4
1	2000	0,91	1,09	0,09	8,8	4000	0,91	1,09	0,09	8,8	0,92	1,09	0,08	8,4	4000	0,91	1,09	0,09	8,8
2	10	-0,63	4,63	2,63	131,5	20	-0,63	4,63	2,63	131,5	0,67	9,00	4,17	208,3	20	0,57	9,00	4,21	210,7
2	25	0,34	3,66	1,66	83,2	50	0,34	3,66	1,66	83,2	0,92	5,25	2,16	108,2	50	0,92	5,67	2,38	118,8
2	50	0,82	3,18	1,18	58,8	100	0,82	3,18	1,18	58,8	1,17	4,00	1,41	70,7	100	1,14	3,85	1,36	67,9
2	75	1,04	2,96	0,96	48,0	150	1,04	2,96	0,96	48,0	1,27	3,41	1,07	53,5	150	1,26	3,40	1,07	53,6
2	100	1,17	2,83	0,83	41,6	200	1,17	2,83	0,83	41,6	1,33	3,00	0,77	38,5	200	1,35	3,12	0,89	44,4
2	150	1,32	2,68	0,68	34,0	300	1,32	2,68	0,68	34,0	1,46	2,85	0,67	33,7	300	1,44	2,86	0,71	35,6
2	200	1,41	2,59	0,59	29,4	400	1,41	2,59	0,59	29,4	1,50	2,77	0,64	31,8	400	1,50	2,71	0,61	30,3
2	250	1,47	2,53	0,53	26,3	500	1,47	2,53	0,53	26,3	1,55	2,62	0,54	26,8	500	1,55	2,63	0,54	26,9
2	300	1,52	2,48	0,48	24,0	600	1,52	2,48	0,48	24,0	1,59	2,57	0,49	24,6	600	1,58	2,56	0,49	24,7
2	350	1,56	2,44	0,44	22,2	700	1,56	2,44	0,44	22,2	1,61	2,54	0,46	23,1	700	1,60	2,51	0,45	22,7
2	400	1,58	2,42	0,42	20,8	800	1,58	2,42	0,42	20,8	1,63	2,48	0,42	21,2	800	1,64	2,48	0,42	21,1
2	450	1,61	2,39	0,39	19,6	900	1,61	2,39	0,39	19,6	1,65	2,46	0,41	20,4	900	1,66	2,45	0,40	19,8
2	500	1,63	2,37	0,37	18,6	1000	1,63	2,37	0,37	18,6	1,66	2,42	0,38	19,1	1000	1,67	2,43	0,38	18,9
2	550	1,65	2,35	0,35	17,7	1100	1,65	2,35	0,35	17,7	1,68	2,40	0,36	17,8	1100	1,68	2,41	0,36	18,1
2	600	1,66	2,34	0,34	17,0	1200	1,66	2,34	0,34	17,0	1,70	2,39	0,34	17,2	1200	1,70	2,38	0,34	17,1
2	650	1,67	2,33	0,33	16,3	1300	1,67	2,33	0,33	16,3	1,71	2,37	0,33	16,5	1300	1,71	2,37	0,33	16,6
2	700	1,69	2,31	0,31	15,7	1400	1,69	2,31	0,31	15,7	1,72	2,35	0,31	15,6	1400	1,71	2,35	0,32	15,8
2	750	1,70	2,30	0,30	15,2	1500	1,70	2,30	0,30	15,2	1,73	2,33	0,30	15,2	1500	1,73	2,34	0,30	15,2
2	800	1,71	2,29	0,29	14,7	1600	1,71	2,29	0,29	14,7	1,73	2,33	0,30	15,1	1600	1,73	2,33	0,30	15,0
2	850	1,71	2,29	0,29	14,3	1700	1,71	2,29	0,29	14,3	1,73	2,32	0,29	14,7	1700	1,74	2,31	0,29	14,4
2	900	1,72	2,28	0,28	13,9	1800	1,72	2,28	0,28	13,9	1,74	2,31	0,28	14,1	1800	1,75	2,30	0,28	13,9
2	950	1,73	2,27	0,27	13,5	1900	1,73	2,27	0,27	13,5	1,75	2,30	0,27	13,6	1900	1,75	2,29	0,27	13,6
2	1000	1,74	2,26	0,26	13,2	2000	1,74	2,26	0,26	13,2	1,75	2,29	0,27	13,4	2000	1,76	2,29	0,27	13,3
2	1500	1,79	2,21	0,21	10,7	3000	1,79	2,21	0,21	10,7	1,80	2,24	0,22	10,9	3000	1,80	2,23	0,21	10,7
2	2000	1,81	2,19	0,19	9,3	4000	1,81	2,19	0,19	9,3	1,82	2,20	0,19	9,3	4000	1,83	2,20	0,19	9,4

Supplementary Table 1. (cont'd)

		Analytic statistical formulas									Bootstrapping								
		Conditional case (ND=0%)				Unconditional case (ND=50%)					Conditional case (ND=0%)				Unconditional case (ND=50%)				
D/P	n_i	$D/P_{95\%min}$	$D/P_{95\%max}$	Sampling error (±1.96SE)	Sampling error (±%)	N	$D/P_{95\%min}$	$D/P_{95\%max}$	Sampling error (±1.96SE)	Sampling error (±%)	$D/P_{95\%min}$	$D/P_{95\%max}$	Sampling error (±)	Sampling error (±%)	N	$D/P_{95\%min}$	$D/P_{95\%max}$	Sampling error (±)	Sampling error (±%)
9	10	-9,59	27,59	18,59	206,6	20	-9,59	27,59	18,59	206,6	2,33	9,00	3,33	37,0	20	1,75	13,00	5,63	62,5
9	25	-2,76	20,76	11,76	130,7	50	-2,76	20,76	11,76	130,7	3,17	24,00	10,42	115,7	50	3,29	27,00	11,86	131,7
9	50	0,68	17,32	8,32	92,4	100	0,68	17,32	8,32	92,4	4,00	49,00	22,50	250,0	100	4,30	43,00	19,35	215,0
9	75	2,21	15,79	6,79	75,4	150	2,21	15,79	6,79	75,4	4,77	24,00	9,62	106,8	150	4,83	24,67	9,92	110,2
9	100	3,12	14,88	5,88	65,3	200	3,12	14,88	5,88	65,3	5,25	19,00	6,88	76,4	200	5,09	20,80	7,85	87,3
9	150	4,20	13,80	4,80	53,3	300	4,20	13,80	4,80	53,3	5,82	17,75	5,97	66,3	300	5,61	17,25	5,82	64,7
9	200	4,84	13,16	4,16	46,2	400	4,84	13,16	4,16	46,2	5,90	15,67	4,89	54,3	400	5,97	15,62	4,82	53,6
9	250	5,28	12,72	3,72	41,3	500	5,28	12,72	3,72	41,3	6,14	14,63	4,24	47,1	500	6,16	14,45	4,15	46,1
9	300	5,61	12,39	3,39	37,7	600	5,61	12,39	3,39	37,7	6,50	14,00	3,75	41,7	600	6,40	13,95	3,78	41,9
9	350	5,86	12,14	3,14	34,9	700	5,86	12,14	3,14	34,9	6,61	13,00	3,20	35,5	700	6,52	13,39	3,43	38,2
9	400	6,06	11,94	2,94	32,7	800	6,06	11,94	2,94	32,7	6,69	12,79	3,05	33,9	800	6,71	13,00	3,14	34,9
9	450	6,23	11,77	2,77	30,8	900	6,23	11,77	2,77	30,8	6,76	12,64	2,94	32,7	900	6,79	12,66	2,93	32,6
9	500	6,37	11,63	2,63	29,2	1000	6,37	11,63	2,63	29,2	6,81	12,51	2,85	31,7	1000	6,85	12,45	2,80	31,1
9	550	6,49	11,51	2,51	27,9	1100	6,49	11,51	2,51	27,9	6,97	12,10	2,56	28,5	1100	6,96	12,25	2,65	29,4
9	600	6,60	11,40	2,40	26,7	1200	6,60	11,40	2,40	26,7	7,00	12,04	2,52	28,0	1200	7,03	11,98	2,48	27,5
9	650	6,69	11,31	2,31	25,6	1300	6,69	11,31	2,31	25,6	7,02	12,00	2,49	27,6	1300	7,07	11,82	2,38	26,4
9	700	6,78	11,22	2,22	24,7	1400	6,78	11,22	2,22	24,7	7,14	11,73	2,29	25,5	1400	7,15	11,73	2,29	25,5
9	750	6,85	11,15	2,15	23,9	1500	6,85	11,15	2,15	23,9	7,15	11,50	2,17	24,2	1500	7,22	11,68	2,23	24,8
9	800	6,92	11,08	2,08	23,1	1600	6,92	11,08	2,08	23,1	7,25	11,50	2,13	23,6	1600	7,25	11,51	2,13	23,7
9	850	6,98	11,02	2,02	22,4	1700	6,98	11,02	2,02	22,4	7,33	11,50	2,08	23,1	1700	7,28	11,40	2,06	22,9
9	900	7,04	10,96	1,96	21,8	1800	7,04	10,96	1,96	21,8	7,33	11,33	2,00	22,2	1800	7,30	11,32	2,01	22,3
9	950	7,09	10,91	1,91	21,2	1900	7,09	10,91	1,91	21,2	7,41	11,34	1,97	21,8	1900	7,38	11,32	1,97	21,9
9	1000	7,14	10,86	1,86	20,7	2000	7,14	10,86	1,86	20,7	7,47	11,20	1,86	20,7	2000	7,43	11,25	1,91	21,2
9	1500	7,48	10,52	1,52	16,9	3000	7,48	10,52	1,52	16,9	7,67	10,72	1,52	16,9	3000	7,68	10,73	1,52	16,9
9	2000	7,69	10,31	1,31	14,6	4000	7,69	10,31	1,31	14,6	7,81	10,49	1,34	14,9	4000	7,84	10,49	1,32	14,7

Supplementary Table 2. The influence of count size on sampling error for the aridity index, Iph [a].

		Analytic statistical formulas			Bootstrapping										
		Conditional case (ND=0%)			Conditional case (ND=0%)			Unconditional case (ND=70%)				Unconditional case (ND=90%)			
Iph (%)	n_i	$Iph_{95\%min}$ (%)	$Iph_{95\%max}$ (%)	Sampling error (±1.96SE%)	$Iph_{95\%min}$ (%)	$Iph_{95\%max}$ (%)	Sampling error (±%)	N	$Iph_{95\%min}$ (%)	$Iph_{95\%max}$ (%)	Sampling error (±%)	N	$Iph_{95\%min}$ (%)	$Iph_{95\%max}$ (%)	Sampling error (±%)
30	10	1,6	58,4	28,4	0,0	60,0	30,0	33	0,0	60,8	30,4	100	0,0	60,0	30,0
	25	12,0	48,0	18,0	12,0	48,0	18,0	83	12,1	50,0	19,0	250	11,3	50,0	19,3
	50	17,3	42,7	12,7	18,0	42,0	12,0	167	17,3	43,2	12,9	500	17,9	43,8	13,0
	75	19,6	40,4	10,4	20,0	40,0	10,0	250	19,7	40,9	10,6	750	19,0	40,2	10,6
	100	21,0	39,0	9,0	22,0	39,0	8,5	333	20,4	39,7	9,7	1000	21,3	38,9	8,8
	150	22,7	37,3	7,3	22,7	37,3	7,3	500	22,4	37,2	7,4	1500	22,1	38,0	8,0
	200	23,6	36,4	6,4	23,5	36,5	6,5	667	23,6	35,8	6,1	2000	24,2	36,3	6,0
	250	24,3	35,7	5,7	24,0	36,0	6,0	833	24,2	36,0	5,9	2500	24,1	35,6	5,7
	300	24,8	35,2	5,2	25,0	34,8	4,9	1000	24,8	35,1	5,1	3000	24,8	35,4	5,3
	350	25,2	34,8	4,8	24,7	34,9	5,1	1167	25,6	34,9	4,6	3500	25,4	34,5	4,6
	400	25,5	34,5	4,5	25,5	34,5	4,5	1333	25,4	34,8	4,7	4000	25,5	34,6	4,6
	450	25,8	34,2	4,2	25,6	34,7	4,6	1500	25,8	33,9	4,0	4500	26,2	34,3	4,0
	500	26,0	34,0	4,0	25,4	34,3	4,5	1667	26,1	33,9	3,9	5000	26,1	33,8	3,9
	550	26,2	33,8	3,8	26,2	33,9	3,9	1833	26,1	33,8	3,9	5500	26,4	34,1	3,8
	600	26,3	33,7	3,7	26,2	33,8	3,8	2000	26,5	33,7	3,6	6000	26,3	33,5	3,6
	650	26,5	33,5	3,5	26,6	33,5	3,5	2167	26,4	33,6	3,6	6500	26,5	33,5	3,5
	700	26,6	33,4	3,4	26,7	33,5	3,4	2333	26,4	33,2	3,4	7000	26,7	33,5	3,4
	750	26,7	33,3	3,3	26,6	33,3	3,3	2500	26,6	33,2	3,3	7500	27,0	33,3	3,1
	800	26,8	33,2	3,2	26,9	33,4	3,3	2667	26,8	33,5	3,4	8000	27,0	33,0	3,0
	850	26,9	33,1	3,1	26,9	32,9	3,0	2833	26,8	33,2	3,2	8500	27,1	33,2	3,0
	900	27,0	33,0	3,0	27,0	33,1	3,0	3000	26,9	33,3	3,2	9000	27,0	33,3	3,2
	950	27,1	32,9	2,9	27,1	32,9	2,9	3167	27,0	32,8	2,9	9500	27,3	33,0	2,8
	1000	27,2	32,8	2,8	27,3	33,1	2,9	3333	27,3	32,9	2,8	10000	27,2	32,8	2,8
	1500	27,7	32,3	2,3	27,9	32,3	2,2	5000	27,5	32,2	2,4	15000	27,9	32,4	2,3
	2000	28,0	32,0	2,0	28,0	32,0	2,0	6667	27,9	32,0	2,0	20000	28,2	32,0	1,9
20	10	-4,8	44,8	24,8	0,0	50	25,0	33	0,0	50	25,0	100	0,0	50	25,0
	25	4,3	35,7	15,7	8,0	36,0	14,0	83	4,8	35,9	15,5	250	5,1	35,1	15,0
	50	8,9	31,1	11,1	10,0	31,0	10,5	167	9,4	32,6	11,6	500	9,4	32,0	11,3
	75	10,9	29,1	9,1	12,0	29,3	8,7	250	11,1	29,9	9,4	750	11,1	29,4	9,1
	100	12,2	27,8	7,8	12,5	28,0	7,8	333	12,6	28,7	8,1	1000	12,5	27,8	7,7
	150	13,6	26,4	6,4	13,3	26,7	6,7	500	13,9	26,2	6,2	1500	14,1	26,9	6,4
	200	14,5	25,5	5,5	14,5	26,0	5,8	667	14,3	25,7	5,7	2000	14,4	25,7	5,6
	250	15,0	25,0	5,0	15,2	25,2	5,0	833	15,0	25,0	5,0	2500	15,3	25,1	4,9
	300	15,5	24,5	4,5	15,7	24,7	4,5	1000	15,6	25,1	4,8	3000	15,5	24,7	4,6
	350	15,8	24,2	4,2	15,9	24,3	4,2	1167	16,1	24,1	4,0	3500	16,0	24,5	4,3
	400	16,1	23,9	3,9	16,3	23,8	3,8	1333	16,1	24,0	4,0	4000	16,0	24,0	4,0
	450	16,3	23,7	3,7	16,2	23,8	3,8	1500	16,3	23,6	3,7	4500	16,4	23,9	3,8
	500	16,5	23,5	3,5	16,2	23,4	3,6	1667	16,5	23,4	3,5	5000	16,6	23,8	3,6
	550	16,7	23,3	3,3	16,5	23,6	3,5	1833	16,8	23,1	3,2	5500	16,7	23,5	3,4
	600	16,8	23,2	3,2	16,8	23,3	3,3	2000	16,9	23,4	3,2	6000	16,9	23,2	3,2
	650	16,9	23,1	3,1	17,0	22,9	3,0	2167	16,9	23,3	3,2	6500	16,8	23,2	3,2
	700	17,0	23,0	3,0	16,9	23,1	3,1	2333	17,1	23,1	3,0	7000	17,2	23,2	3,0
	750	17,1	22,9	2,9	17,2	22,7	2,7	2500	17,2	22,9	2,8	7500	17,2	22,9	2,9
	800	17,2	22,8	2,8	17,3	22,9	2,8	2667	17,1	22,7	2,8	8000	17,5	22,7	2,6
	850	17,3	22,7	2,7	17,4	22,8	2,7	2833	17,4	22,9	2,7	8500	17,6	22,9	2,7
	900	17,4	22,6	2,6	17,3	22,8	2,7	3000	17,3	22,5	2,6	9000	17,3	22,7	2,7
	950	17,5	22,5	2,5	17,5	22,8	2,7	3167	17,5	22,6	2,6	9500	17,4	22,4	2,5
	1000	17,5	22,5	2,5	17,6	22,5	2,5	3333	17,5	22,4	2,5	10000	17,4	22,6	2,6
	1500	18,0	22,0	2,0	18,1	22,1	2,0	5000	18,1	22,1	2,0	15000	18,0	22,1	2,1
	2000	18,2	21,8	1,8	18,2	21,9	1,8	6667	18,3	21,7	1,7	20000	18,2	21,7	1,7
10	10	-8,6	28,6	18,6	0,0	30	15,0	33	0,0	33,3	16,7	100	0,0	30	15,0
	25	-1,8	21,8	11,8	0,0	24,0	12,0	83	0,0	23,8	11,9	250	0,0	21,7	10,9
	50	1,7	18,3	8,3	2,0	18,0	8,0	167	3,1	19,2	8,1	500	2,0	18,8	8,4
	75	3,2	16,8	6,8	4,0	17,3	6,7	250	3,9	17,1	6,6	750	3,8	17,0	6,6
	100	4,1	15,9	5,9	5,0	16,0	5,5	333	4,2	16,2	6,0	1000	4,4	16,0	5,8
	150	5,2	14,8	4,8	5,3	15,3	5,0	500	5,2	15,1	4,9	1500	5,6	15,0	4,7
	200	5,8	14,2	4,2	6,0	14,5	4,3	667	6,1	14,6	4,2	2000	6,0	14,6	4,3
	250	6,3	13,7	3,7	6,4	14,0	3,8	833	6,3	13,9	3,8	2500	6,4	13,8	3,7
	300	6,6	13,4	3,4	7,0	13,3	3,2	1000	6,8	13,6	3,4	3000	7,0	13,5	3,3
	350	6,9	13,1	3,1	6,9	13,4	3,3	1167	7,1	13,1	3,0	3500	7,0	13,1	3,1
	400	7,1	12,9	2,9	7,3	13,0	2,9	1333	7,0	12,9	3,0	4000	7,3	12,9	2,8
	450	7,2	12,8	2,8	7,1	12,9	2,9	1500	7,2	13,0	2,9	4500	7,3	12,7	2,7
	500	7,4	12,6	2,6	7,6	12,6	2,5	1667	7,4	12,8	2,7	5000	7,4	12,5	2,5
	550	7,5	12,5	2,5	7,5	12,5	2,5	1833	7,6	12,5	2,5	5500	7,4	12,5	2,5
	600	7,6	12,4	2,4	7,8	12,3	2,2	2000	7,7	12,5	2,4	6000	7,6	12,6	2,5
	650	7,7	12,3	2,3	7,8	12,2	2,2	2167	7,7	12,4	2,3	6500	7,8	12,4	2,3
	700	7,8	12,2	2,2	7,9	12,3	2,2	2333	7,9	12,3	2,2	7000	7,8	12,2	2,2
	750	7,9	12,1	2,1	8,0	12,3	2,1	2500	7,9	12,3	2,2	7500	8,0	12,3	2,2

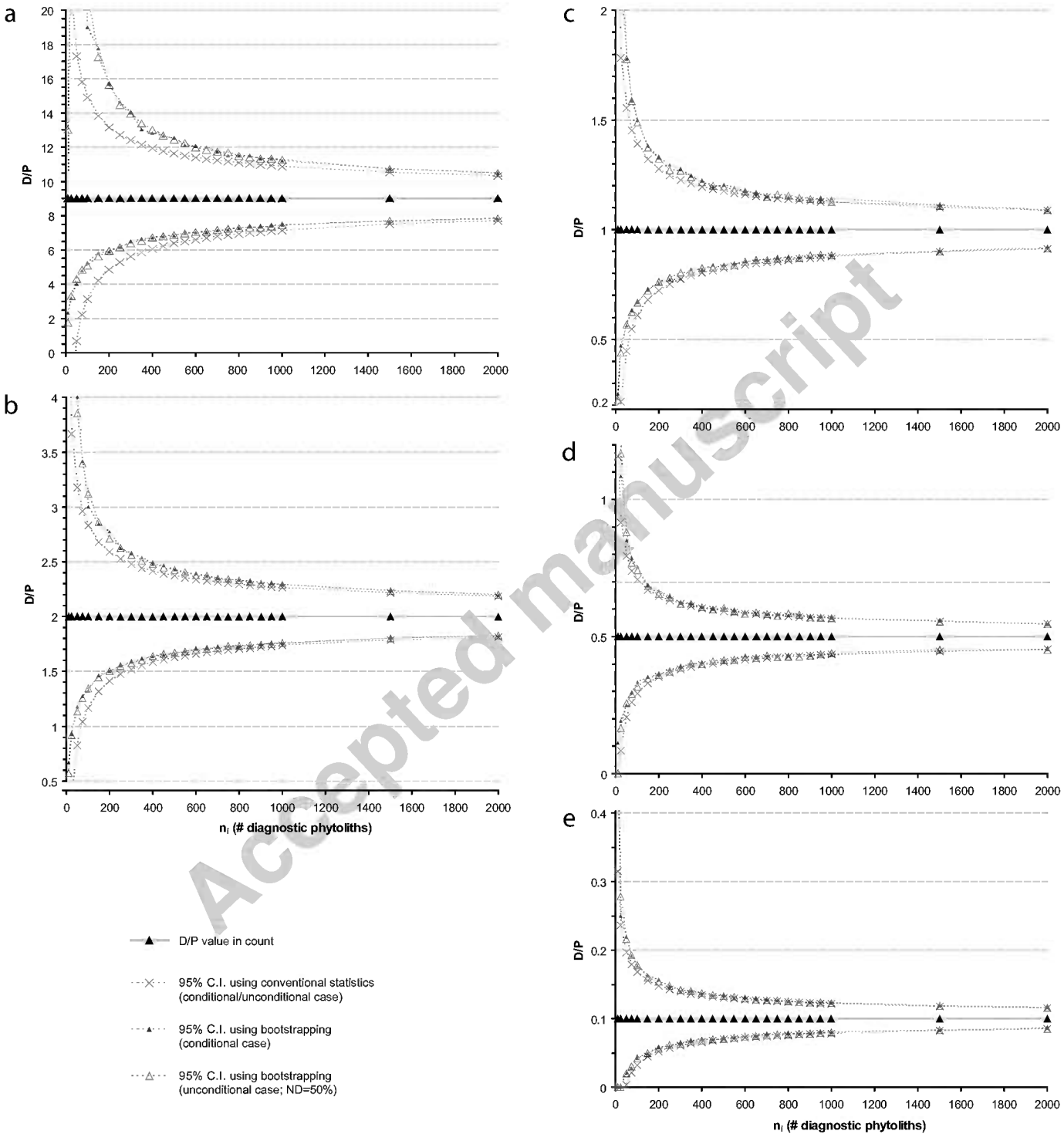

Fig. 1.

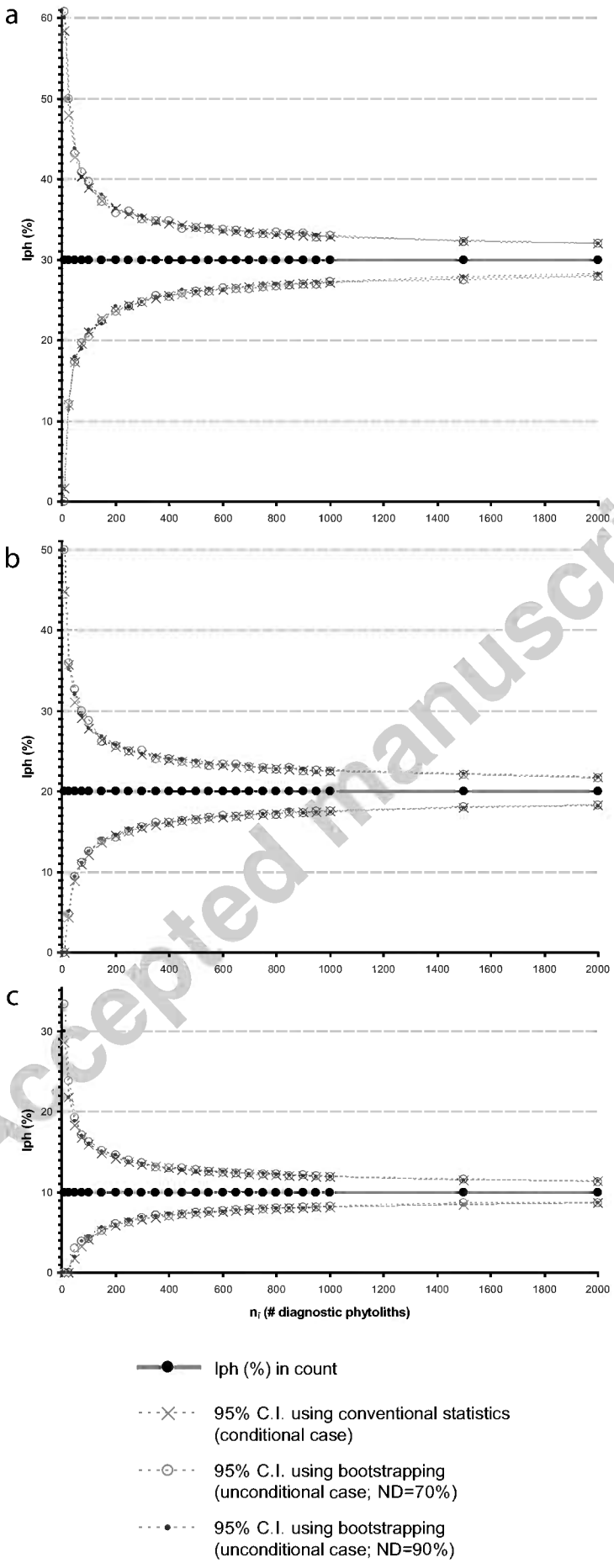

Fig. 2.

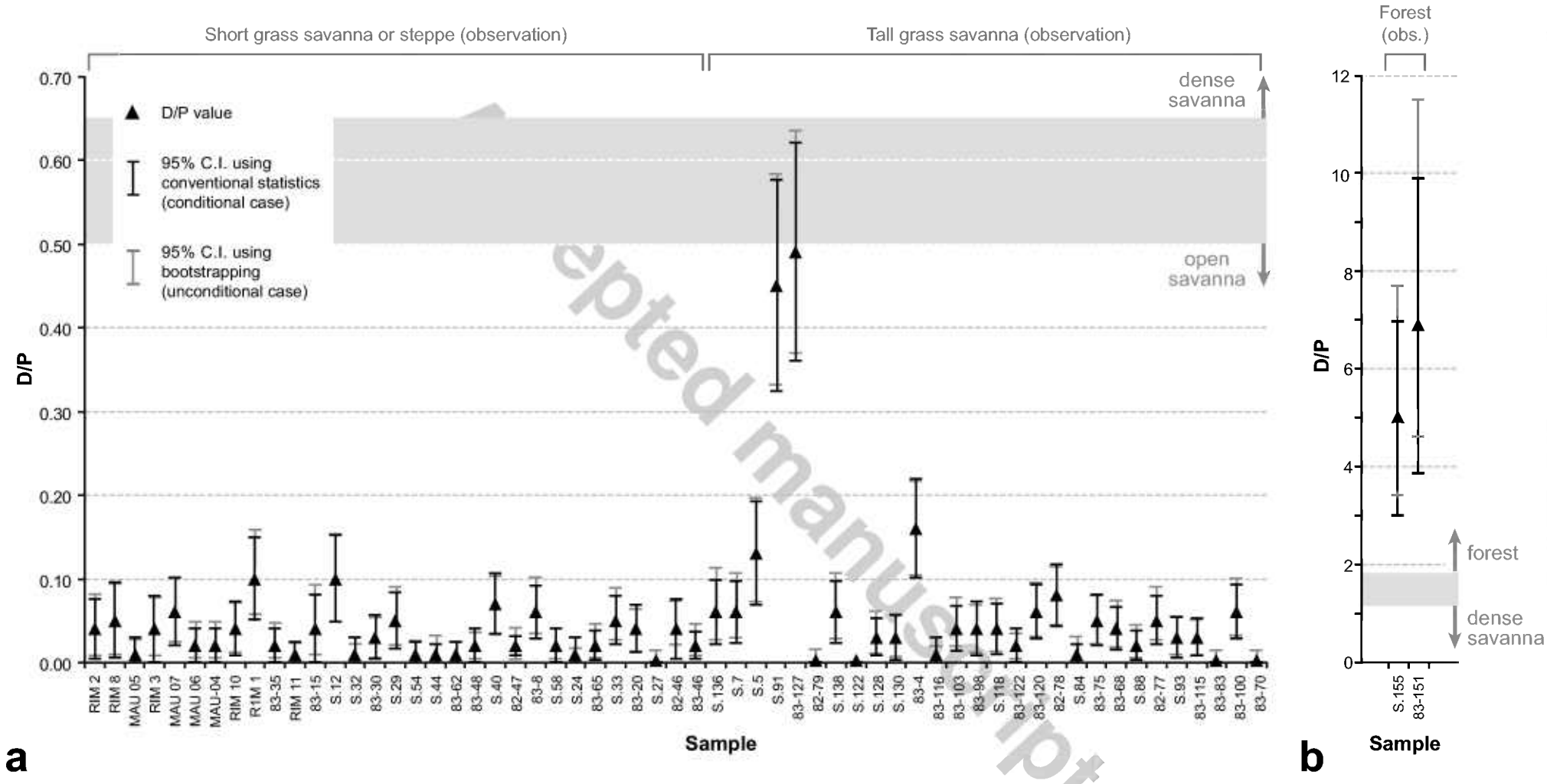

Fig. 3.

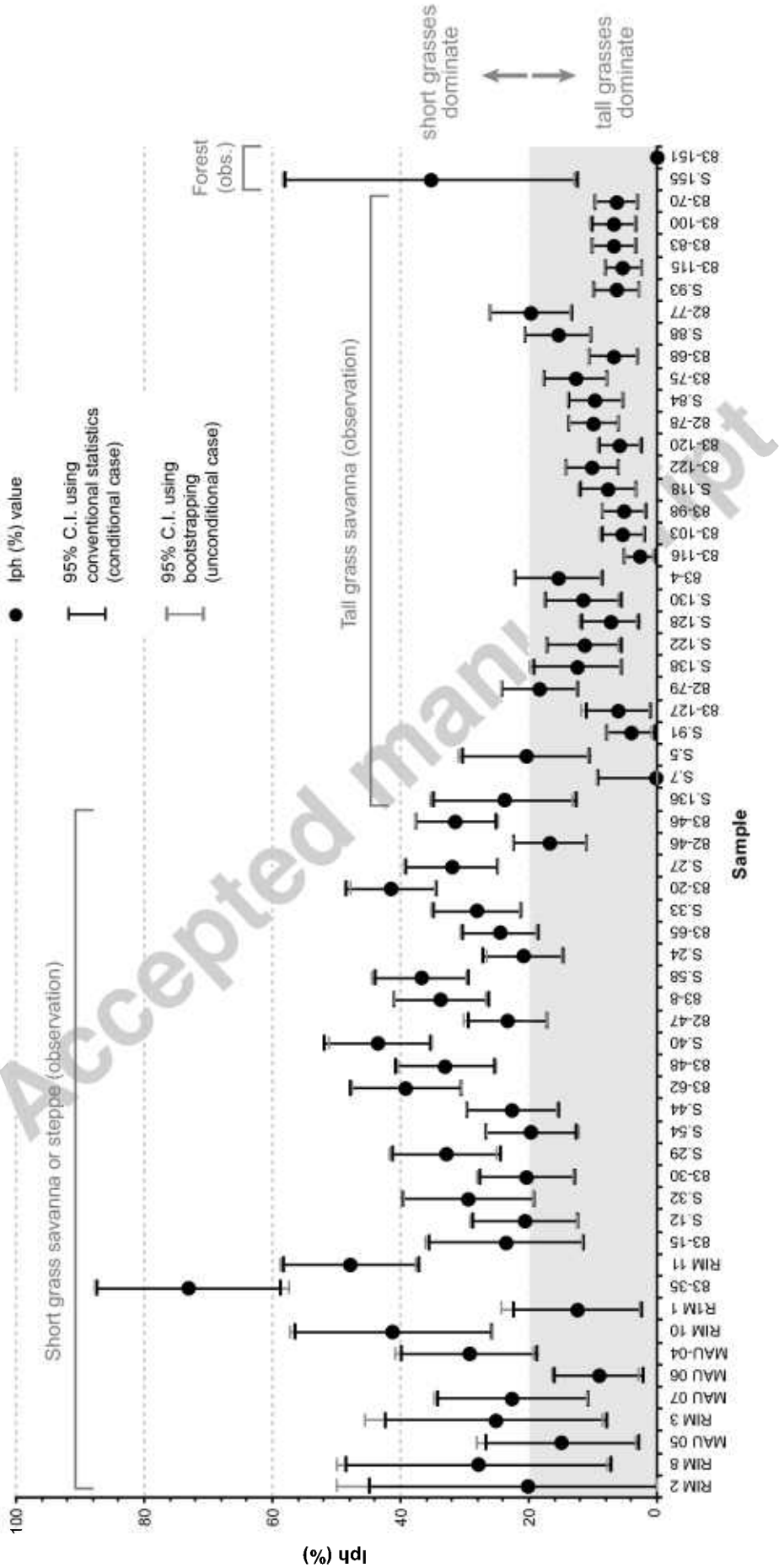

Fig. 4.

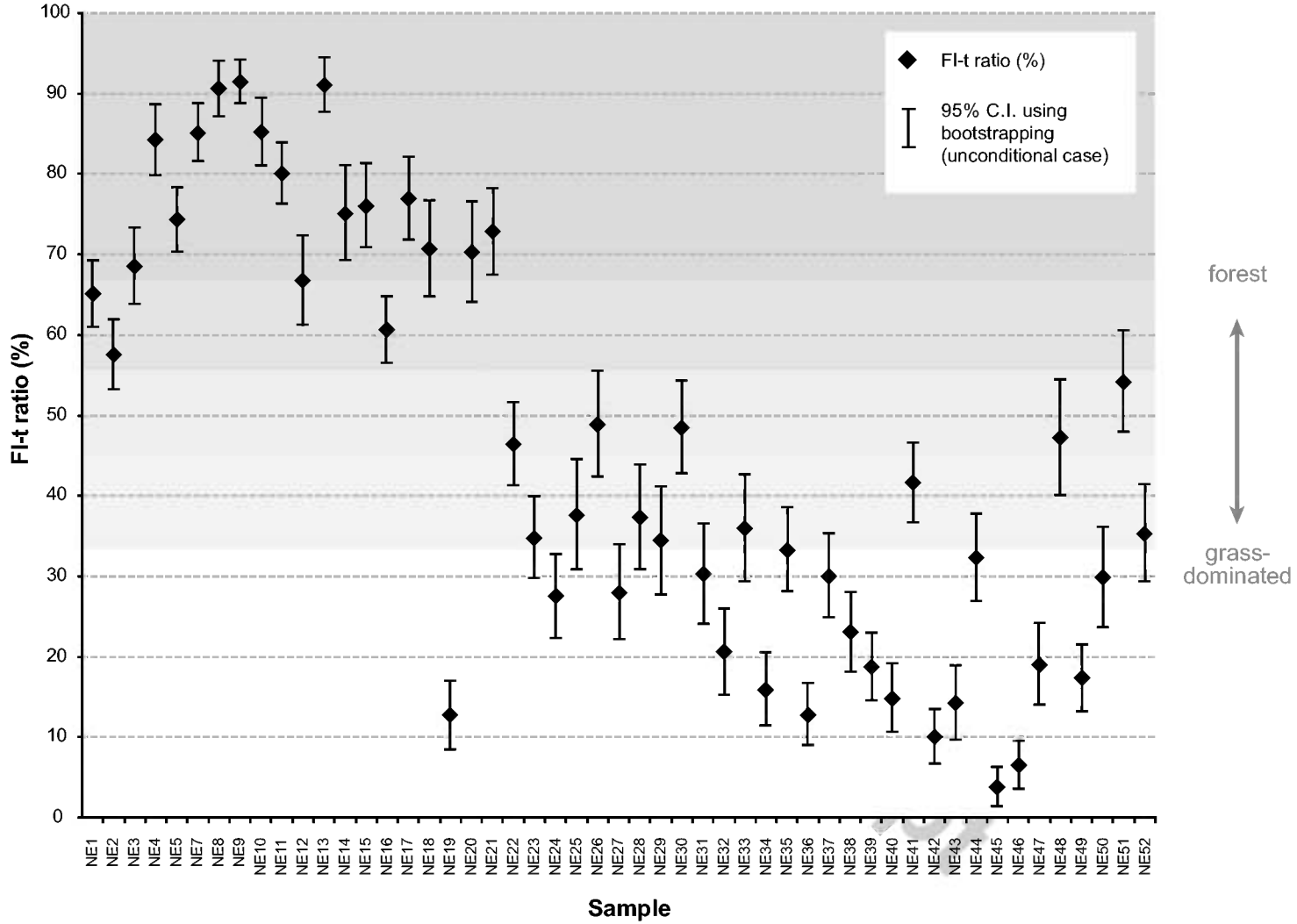

Fig. 5.