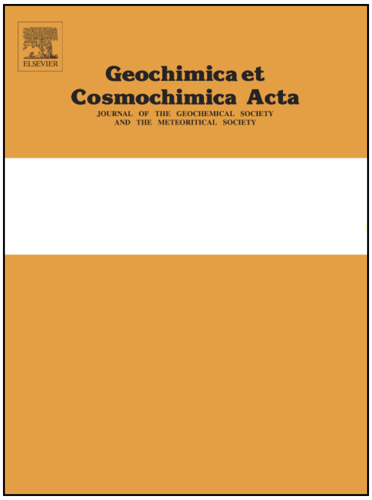


ELSEVIER

Geochimica et Cosmochimica Acta

JOURNAL OF THE GEOCHEMICAL SOCIETY AND THE METEORITICAL SOCIETY

A hierarchical Bayesian approach to the classification of C_3 and C_4 grass pollen based on SPIRAL $\delta^{13}C$ data

Michael A. Urban, David M. Nelson, Ryan Kelly, Tahir Ibrahim, Michael Dietze, Ann Pearson, Feng Sheng Hu

PII: S0016-7037(13)00399-2
DOI: http://dx.doi.org/10.1016/j.gca.2013.07.019
Reference: GCA 8369

To appear in: *Geochimica et Cosmochimica Acta*

Received Date: 13 December 2012
Accepted Date: 16 July 2013

Please cite this article as: Urban, M.A., Nelson, D.M., Kelly, R., Ibrahim, T., Dietze, M., Pearson, A., Hu, F.S., A hierarchical Bayesian approach to the classification of C_3 and C_4 grass pollen based on SPIRAL $\delta^{13}C$ data, *Geochimica et Cosmochimica Acta* (2013), doi: http://dx.doi.org/10.1016/j.gca.2013.07.019

A hierarchical Bayesian approach to the classification of C_3 and C_4 grass pollen based on SPIRAL δ^{13}C data

Michael A. Urban[1], David M. Nelson[2]*, Ryan Kelly[3], Tahir Ibrahim[3], Michael Dietze[4], Ann Pearson[5], Feng Sheng Hu[1,3,6]*

[1]Program in Ecology, Evolution and Conservation, University of Illinois, Urbana, IL, USA
[2]University of Maryland Center for Environmental Science, Appalachian Laboratory, Frostburg, MD, USA
[3]Department of Plant Biology, University of Illinois, Urbana, IL, USA
[4]Department of Earth and Environment, Boston University, Boston, MA, USA
[5]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA
[6]Department of Geology, University of Illinois, Urbana, IL, USA

*Corresponding authors: Feng Sheng Hu (fshu@life.illinois.edu) and David Nelson (dnelson@umces.edu)

Keywords: carbon isotopes, pollen, grasses, SPIRAL, Bayesian, grasslands

Abstract

Differentiating C_3 and C_4 grass pollen in the paleorecord is difficult because of their morphological similarity. Using a spooling wire microcombustion device interfaced with an isotope ratio mass spectrometer, Single Pollen Isotope Ratio AnaLysis (SPIRAL) enables classification of grass pollen as C_3 or C_4 based upon $\delta^{13}C$ values. To address several limitations of this novel technique, we expanded an existing SPIRAL training dataset of pollen $\delta^{13}C$ data from 8 to 31 grass species. For field validation, we analyzed $\delta^{13}C$ of individual grains of grass pollen from the surface sediments of 15 lakes in Africa and Australia, added these results to a prior dataset of 10 lakes from North America, and compared C_4-pollen abundance in surface sediments with C_4-grass abundance on the surrounding landscape. We also developed and tested a hierarchical Bayesian model to estimate the relative abundance of C_3- and C_4-grass pollen in unknown samples, including an estimation of the likelihood that either pollen type is present in a sample. The mean (±SD) $\delta^{13}C$ values for the C_3 and C_4 grasses in the training dataset were -29.6 ± 9.5‰ and -13.8 ± 9.5‰, respectively. Across a range of % C_4 in samples of known composition, the average bias of the Bayesian model was <3% for C_4 in samples of at least 50 grains, indicating that the model accurately predicted the relative abundance of C_4 grass pollen. The hierarchical framework of the model resulted in less bias than a previous threshold-based C_3/C_4 classification method, especially near the high or low extremes of C_4 abundance. In addition, the percent of C_4 grass pollen in surface-sediment samples estimated using the model was strongly related to the abundance of C_4 grasses on the landscape ($n= 24$, $p< 0.001$, $r^2= 0.65$). These results improve $\delta^{13}C$-based quantitative reconstructions of grass community composition in the paleorecord and demonstrate the utility of the Bayesian framework to aid the interpretation of stable isotope data.

1. INTRODUCTION

Knowledge of biotic responses to past climatic variability is important for anticipating future change (Flessa et al., 2005). Fossil pollen assemblages are a valuable indicator of spatiotemporal variation in plant community composition on the landscape (Williams et al., 2004). However, grass (Poaceae) pollen is typically morphologically indistinct below the family level (Fægri et al., 1989), rendering pollen analysis a blunt instrument for investigating past changes in grassland ecosystems. This problem hampers our understanding of the ecology and evolution of grasslands, which today cover a major portion of Earth's land surface and regulate key biogeochemical cycles (Saugier and Roy, 2000).

Carbon isotopic analysis of grass pollen offers an important tool for distinguishing C_3 and C_4 grasses in the paleorecord (Amundson et al., 1997; Descolas-Gros and Scholzel, 2007; Nelson et al., 2006). Recent technical advances include Single Pollen Isotope Ratio AnaLysis (SPIRAL), which involves the use of a spooling-wire microcombustion device interfaced with an isotope-ratio mass spectrometer (SWiM-IRMS) for the δ^{13}C analysis of individual grass pollen grains (Nelson et al., 2007). Nelson et al. (2007) showed that δ^{13}C values of pollen from known C_3 and C_4 grasses could be distinguished based on their distribution around a threshold δ^{13}C value of -19.2‰. Although high variability and overlapping ranges of δ^{13}C values for C_3 and C_4 grasses prevent perfect classification, a significant correlation was found between δ^{13}C-based estimates of % C_4-grass pollen in surface-sediment samples and the abundance of C_4 grasses on the landscape at ten sites in North America (Nelson et al., 2008).

Despite the useful paleoenvironmental information obtained from SPIRAL, the existing technique has several inherent limitations. First, SPIRAL was developed (Nelson et al., 2007) and validated (Nelson et al., 2008) with a small amount of data from North American grasses and grasslands. For example, only four C_4 grasses and four C_3 grasses were used to identify the threshold $\delta^{13}C$ value separating C_3 and C_4 (Nelson et al., 2007). Thus the applicability of this technique to a broader range of grassland ecosystems remains uncertain. Additionally, a fixed $\delta^{13}C$ threshold was selected to differentiate C_3 and C_4 grasses, which may be problematic because $\delta^{13}C$ values vary both within and among species (Cerling, 1999). Finally, there is no formal propagation of uncertainty for SPIRAL, which means that the precision of the technique is not well constrained. In this study, we address these problems by (1) expanding the reference $\delta^{13}C$ dataset for distinguishing C_3- from C_4-grass pollen, (2) improving the validation dataset from North America (Nelson et al., 2008) by adding new surface-sediment samples from lakes in Africa and Australia, and (3) developing and evaluating a hierarchical Bayesian model to estimate the percent of C_3- and C_4-grass pollen based on SPIRAL $\delta^{13}C$ data.

2. METHODOLOGY

2.1 Herbarium and surface-sediment samples

We performed $\delta^{13}C$ analyses on pollen from herbarium specimens of 28 grass species, including additional pollen from five of the eight species previously analyzed in Nelson et al. (2007) (Electronic Annex EA-1). Our expanded training dataset includes these new results and all of the $\delta^{13}C$ data reported in Nelson et al. (2007). These specimens were collected between 1927 and 1995 from Africa, Australia, and North America.

As a step to develop a global relationship between C_4 grass abundance and SPIRAL data, we performed $\delta^{13}C$ analysis of pollen in surface sediments from Africa and Australia to supplement the published $\delta^{13}C$ dataset from North America (Nelson et al., 2008). All of the surface-sediment samples from Africa and Australia come from lakes, with the exception of Rumuiku Swamp in Africa (Electronic Annex EA-2). The samples typically represent the upper ~5 cm of sediment, which likely accumulated during the past several decades. We do not have data on the composition and abundance of grasses around our African and Australian sites. Therefore, we estimated the relative abundance or productivity of C_4 grasses around each site based on the relationships of C_4 grasses with various environmental factors reported in the literature (Electronic Annex EA-2). In equatorial East Africa, C_4-grass abundance (Livingstone and Clayton, 1980) and productivity (Tieszen et al., 1979) are negatively correlated with elevation, with C_4 grasses predominating below ~1500 m. We used the relationship in Tieszen et al. (1979) to estimate C_4 grass abundance around each of our African sites. In Australian grasslands, minimum January temperatures (JANT; °C) and median August rainfall (AURF; cm) are strong predictors of C_4 grass abundance in the regional grass flora (Hattersley, 1983). We obtained JANT and AURF data from the Australian Bureau of Meteorology (www.bom.gov.au) and used the relationship in Hattersley (1983) to calculate C_4 grass abundance around each of our Australian sites. For each North American site the percent contribution of C_4 grasses to the total potential production of grasses was determined using the relationship between latitude and C_4-grass productivity (Tieszen et al., 1997).

2.2 Sample treatment and isotopic analysis

All samples were treated using standard pollen preparation techniques modified to exclude carbon-containing compounds (Nelson et al., 2006), except that hydrofluoric acid was not used for the herbarium specimens, which has little influence on pollen $\delta^{13}C$ (Jahren, 2004). Grass pollen gains were isolated in Nanopure water on a microscope slide at 200x magnification using an Eppendorf Transferman micromanipulation device. Individual grains were transferred to ~0.4 μL drops of Nanopure water and applied to a SWiM device interfaced with a ThermoFinnigan Delta V IRMS using a steel and glass syringe (Nelson et al., 2007; Nelson et al., 2008). Sample data were normalized to VPDB using a two-point normalization curve with in-house 2.5 nmol C standards of leucine (true $\delta^{13}C$ = -32.1‰), sorbitol (true $\delta^{13}C$= -16.2‰), serine (true $\delta^{13}C$ = -25.7‰), and/or glycine (true $\delta^{13}C$ = -37.9‰) that were calibrated against the USGS40 and USGS41 glutamic acid standards.

The number of individual grains of grass pollen applied to the SWiM device ranged from 88 to 239 per sample for the herbarium and surface-sediment samples. We followed Nelson et al. (2007; 2008) for the $\delta^{13}C$ analysis of individual pollen grains. Briefly, along with each sample, we analyzed blanks of Nanopure water to which a single pollen grain was added and then removed. The mean plus 2 standard deviations of blank CO_2 yields was set as a minimum size threshold; grains below this threshold were excluded. The final $\delta^{13}C$ data were corrected for blank ^{13}C content using isotopic mass balance. The $\delta^{13}C$ values of herbarium specimens were corrected to a pre-industrial $\delta^{13}C$ value of atmospheric CO_2 (-6.3‰; Friedli et al., 1986).

2.3 Statistical model

We chose a Bayesian approach for our statistical analysis. Bayesian methods differ theoretically from more widely-used frequentist approaches primarily in that Bayesian methods include *prior* distributions for all unknown parameters to be estimated. Following a fundamental theorem of probability known as Bayes' theorem, prior distributions can be combined with the likelihood of a given dataset (i.e., the probability of observing the dataset, given as a function of unknown parameters) to yield *posterior* parameter distributions. Formally and conceptually, a posterior distribution represents a prior notion of an unknown parameter value, updated with available data according to the proposed model. In many cases (e.g. linear regression), Bayesian and frequentist approaches yield essentially equivalent results when the prior distributions selected are uninformative (i.e. provide little constraint on the unknown parameters), and/or when the dataset is sufficiently large to overwhelm the priors. In other cases, however, the choice of priors can be influential, and the inherent subjectivity in assigning priors has been central to arguments for and against the use of Bayesian methods. For a summary of these theoretical considerations, see Savage (1962).

From a pragmatic standpoint, advances in computational methods have provided a consistent and convenient framework for fitting complex models from a Bayesian perspective, where a frequentist approach would be infeasible or impossible. This practical advantage is the motivation for our Bayesian model. The model we propose below is relatively straightforward, and is closely related to model-based clustering methods (Fraley and Raftery, 2002). Nevertheless, the exact model structure is specific to our context and goals, i.e. estimating C_4 grass abundance in unknown samples and the likelihood that they contain C_4 grass pollen. We know of no frequentist approach that would suffice to fit such a model, whereas in a Bayesian context it can be solved using generic numerical methods. For a practical introduction to such methods, we recommend Clark (2007) and Hoff (2009).

We designed a hierarchical Bayesian model to predict the percent of C_4 grains in samples of unknown composition based on the $\delta^{13}C$ values of individual grass pollen grains (Fig. 1). At the basis of the model is the likelihood function

$$y_i \sim \begin{cases} N\left(\mu_{C_3}, \sigma^2_{C_3}\right), & x_i = 0 \\ N\left(\mu_{C_4}, \sigma^2_{C_4}\right), & x_i = 1 \end{cases}$$

in which, for the i^{th} grain in the sample, y_i is the measured $\delta^{13}C$ of the grain, x_i is a binary variable identifying the grain as C_3 ($x_i = 0$) or C_4 ($x_i = 1$), μ and σ^2 represent the population means and variances (respectively) for C_3 and C_4 grains as indicated by subscripts, and $N(\mu, \sigma^2)$ denotes the normal (Gaussian) distribution with mean μ and variance σ^2. In other words, the likelihood is the conditional probability of observing the $\delta^{13}C$ value of an individual grain, given the classification of the grain and assuming normally-distributed $\delta^{13}C$ values for both C_3 and C_4. We calculated μ_{C_3}, μ_{C_4}, $\sigma^2_{C_3}$, and $\sigma^2_{C_4}$ from the herbarium dataset described above, and subsequently treated these variables as fixed in our predictive model.

Because the C_3/C_4 identity of the pollen grains in sediment samples is unknown, we added a second hierarchical level to model x, the indicator variable for C_4 presence, based on the unknown proportion of C_4 grains in the population, θ:

$$x_i \sim Bernoulli(\theta)$$

i.e.,

$$x_i = \begin{cases} 1 & \text{with probability } \theta \\ 0 & \text{with probability } (1-\theta) \end{cases}$$

The unknown parameter *θ* requires a prior distribution as well. In defining this prior, we introduced a final hierarchical level in the model to accommodate samples composed of (1) purely C_3, (2) purely C_4, or (3) both C_3 and C_4 pollen grains. We refer to these sample types as “C_3-only”, “C_4-only”, and “mixed”, respectively, and define the prior distribution of *θ* separately for each:

$$\theta \sim \begin{cases} 0 & \text{for } C_3\text{-only samples} \\ Uniform(0,1) & \text{for mixed samples} \\ 1 & \text{for } C_4\text{-only samples} \end{cases}$$

In other words, if a sample is identified as C_3-only or C_4-only, then *θ* is assigned a constant value of 0 or 1 (respectively). For mixed samples, *θ* must be estimated based on the data. In this case, the uniform prior represents our lack of knowledge of the true proportion of C_4 in the sample by assuming *a priori* that all values of *θ* are equally likely.

The compound prior on *θ* effectively defines three distinct sub-models. In a Bayesian framework, these models can be fit simultaneously to formally compare their ability to describe a given dataset. This simple form of Bayesian model selection (Dellaportas et al., 2002) treats the choice of model itself as an unknown parameter, which therefore requires its own prior distribution. We assumed that the sub-models were equally likely *a priori*, and thus assigned each a prior probability of 1/3. The posterior estimate of the model-selection parameter then yields “posterior model probabilities” representing the relative probability that each candidate model (i.e. sample type) is the true model. This allows for hypothesis testing analogous to the use of *p*-values (e.g. rejecting a candidate model if it has a posterior probability <0.05; Marden, 2000).

The division of the main hierarchy into three possible submodels serves two purposes. First, for samples that truly contain only one pollen type, the corresponding monotypic model is conceptually correct, and generally provides a better fit than if only the "mixed" model is allowed (data not shown). Second, fitting this model produces a posterior estimate of θ while simultaneously calculating the posterior probability of each sample type. In applications aimed primarily at assessing the relative abundance of C_4 grains in a sample (e.g. to compare C_4 abundance across space or time), θ will be of primary interest. However, in some cases the goal of SPIRAL may be to identify whether one pollen type is present or absent in a sample (e.g. Urban et al., 2010). For that purpose, the posterior model probabilities allow explicit quantification of the probability that either or both types are present.

We fit the model by Markov Chain Monte Carlo (MCMC) sampling using the software package JAGS (version 3.1.0; Plummer, 2011) interfaced through R (R Development Core Team, 2010) with the library *rjags* (Plummer, 2012). Briefly, the software uses a variety of MCMC algorithms to sample over possible values of the unknown parameters. For each parameter, the resulting posterior distribution (i.e. histogram of all values sampled during the MCMC sequence) is an approximation of the true probability density function of the parameter given the dataset of observations. Any population statistic of interest can then be estimated from the corresponding sample statistic for the MCMC sample. For example, we summarize θ by its posterior median, calculated as the sample median across the entire MCMC sequence.

We used pseudodata from the herbarium samples to verify the model. We produced samples with known composition of 0 to 100% C_4 in 10% increments, and sample sizes of 50, 100, and 150 grains. We randomly generated 1000 replicates of each % C_4 X sample-size combination, and fit the model to each replicate sample to generate posterior estimates of θ and posterior probabilities for each sample type (C_3-only, C_4-only, or mixed). For comparison, we also estimated % C_4 for each sample using the threshold-based classification method (i.e. Nelson et al., 2007), but with the threshold value (the midpoint between m_{C_3} and m_{C_4}) updated to reflect the expanded herbarium dataset. Finally, we used the model to estimate the percent of C_4 grains in the surface sediments of sites in Africa, Australia, and North America. For comparison of these estimates with the relative abundance of C_4 grasses on the landscape, we used reduced major axis regression because of symmetry in the variables on the *x* and *y* axes (Smith, 2009), and because both the *x* and *y* variables contain uncertainty (McArdle, 1988). The fit of this regression was compared with a 1:1 relationship following equations outlined in McArdle (1988). These regression analyses were performed in R (R Development Core Team, 2010).

3. RESULTS AND DISCUSSION

3.1 δ^{13}C of C_3 and C_4 grass pollen: an expanded training set

The expanded training set is based on pollen from 31 herbarium specimens. The number of grass pollen grains applied to the moving wire with peak areas exceeding the 2σ threshold of blanks ranges between 21 and 130 grains per sample, with an average of 62 grains per sample (Electronic Annex EA-1). The expanded training set therefore includes 1,921 δ^{13}C values, 1,402 of which were obtained as part of the present study. An average of 32% of applications of pollen from herbarium samples yield a peak area above the blank threshold, which is lower than results from surface-sediment samples from North American lakes (47%, Nelson et al., 2008) and

Miocene/Oligocene sediment samples (45%, Urban et al., 2010). The mean δ^{13}C values of grass pollen range between -42.7 and -24.0‰ for C_3 species and between -17.2 and -10.5‰ for C_4 species (Electronic Annex EA-1). A majority of the pollen δ^{13}C values fall within the typical δ^{13}C ranges for C_3 (-34 to -22‰) and C_4 (-15 to -10‰) plants (Fig. 2; Electronic Annex EA-1). However, similar to previous results, the δ^{13}C variation is large, with many individual data points exceeding these ranges, likely because of variability in the magnitude and composition of the analytical blank (Nelson et al., 2007).

The updated herbarium dataset yields somewhat different parameter estimates than those reported by Nelson et al. (2007). Estimates of $m_{C_3} = -29.6$‰ and $m_{C_4} = -13.8$‰ are more negative than previously determined values (–26.9‰ and –11.5‰, respectively), leading to an estimated threshold value of –21.7‰ that is also more negative than the original value (–19.2‰). Variability of δ^{13}C in the new dataset is similar between C_3 and C_4 grains (standard deviation = 9.5‰ for each), which is greater than previously determined for C_3 (6.3‰), but similar for C_4 (9.6‰). Based on the updated values, the probability of an individual grain being identified as C_4 by the Bayesian model varies smoothly over the range of possible δ^{13}C values (Fig. 2).

In terms of estimating the overall composition of unknown samples, the pseudodata experiments show a striking improvement of the Bayesian approach. Overall, results from samples of pseudodata randomly generated from the herbarium dataset illustrate that Bayesian estimates of % C_4 grass pollen are highly accurate (Fig. 3). For all sample sizes tested, bias (i.e., the mean deviation between the estimated and true %C_4) is ≤5.5%, with largest biases when true C_4 composition is 80% (n=50) or 10% (n≥100). Average biases across all true % C_4 values are only 2.9% for sample size n=50, and 2.4% for n=100 and n=150. By contrast, the original threshold-based methodology of Nelson et al. (2007) produces accurate estimates of sample composition when true composition is near 50%, but becomes increasingly biased towards underestimation (overestimation) as true % C_4 increases (decreases). Maximum bias of ~16% for the threshold-based approach occurs for purely C_3 or C_4 samples, and average bias across all true % C_4 values is 8.2%.

The improved accuracy of the Bayesian model for samples with low and high abundances of C_4 grass pollen is a function of its hierarchical structure. The model explicitly incorporates θ, the estimated relative abundance of C_4 grains in the population, as well as a model-selection parameter representing the possibility that either C_3 or C_4 can be entirely absent from a sample. The MCMC approach then solves for these parameters simultaneously while accounting for the fact that they both influence the likelihood of an individual grain being identified as C_3 or C_4. By contrast, the threshold method relies on a fixed threshold value with classification accuracies for C_3 and C_4 grains that are independent of sample composition. In practice, the threshold method misclassifies approximately the same percent C_3 and C_4 grains. Thus, near 50% true C_4 abundance, the number of misclassification errors for C_3 and C_4 are similar, which results in offsetting effects on estimated % C_4 and small net bias. However, when % C_4 is far from 50% the misclassification errors are imbalanced, which results in a biased estimate of % C_4.

To illustrate how the hierarchical Bayesian model overcomes this limitation, here we consider a hypothetical sample with low (<50%) C_4 abundance, and we note that the opposite rationale applies for samples with high C_4 abundance. For a low-C_4 sample, the data favor a correspondingly low estimate of θ. Consequently, the likelihood of any grain being identified as C_4 in the model is diminished, reflecting the reduced probability of a C_4 grain being found in a sample when the true abundance of C_4 grains is low. This in turn causes fewer C_3 grains with ambiguous $\delta^{13}C$ values to be misclassified as C_4. As the true percent of C_4 in the hypothetical sample approaches 0, the data will begin to favor selection of the C_3-only model, which prevents misidentification of C_3 grains. These same mechanisms lead to an increased proportion of C_4 grains misclassified as C_3 in low-C_4 samples. However, since a sample with low C_4 abundance has fewer C_4 than C_3 grains by definition, the net effect is an improvement in accuracy relative to the threshold-based method.

Our Bayesian model can also be used to assess the presence or absence of C_4 grasses on the landscape (Fig. 4). For example, for pseudodata samples containing 0% C_4, the posterior probability of the C_3-only model [P(C_3-only)] has a median value of >0.95, indicating strong preference for the correct model most of the time. Similarly, for pseudodata samples containing 100% C_4, median P(C_4-only) is ~0.94 indicating strong preference for the C_4-only model. Furthermore, our results suggest that the method has substantial power to reject the C_3-only model when C_4 grains are in fact present. For example, with a sample size of 100 grains, median P(C_3-only) is <0.01 for samples with only 20% C_4. Samples with C_4 present in lower abundance are more ambiguous. Among samples with 10% C_4, for instance, median P(C_3-only) of a 100-grain sample is 0.54. The ability to identify C_4 presence improves with sample size. For example, for a sample with 10% C_4, median P(C_3-only) is 0.23 with n=150 grains, compared to 0.73 with n=50 grains. Thus, for samples of relatively large size ($\geq$100 grains) the practical detection limit for reliably identifying the presence of C_4 grains in a sample is between 10-20% C_4.

3.2. Field validation of grass-pollen $\delta^{13}C$ as a proxy indicator of C_3/C_4 abundance

For the surface-sediment samples from Africa and Australia, the number of grass pollen grains with peak areas exceeding the 2σ threshold of blanks ranges between 30 and 142 grains per sample, with an average of 52 grains per sample (Electronic Annex EA-2). The total surface-sediment dataset therefore includes 1,522 $\delta^{13}C$ values, 773 of which were obtained as part of the present study. On average, 48% of applications of pollen from sediment samples yield a peak area above the blank threshold. A majority of the pollen $\delta^{13}C$ values fall within or between typical $\delta^{13}C$ ranges for C_3 and C_4 plants (Electronic Annex EA-2, EA-3, and EA-4). However, as with the expanded herbarium dataset, the $\delta^{13}C$ variation is large.

Bayesian estimates of the median % C_4 grass pollen from the surface-sediment samples range between 0 and 99% (Fig 5; Electronic Annex EA-2). Across the large spatial and environmental gradients represented by our surface-sediment sites, we expected that the abundance of C_3 and C_4 grass pollen in surface sediments would be overall similar to the abundance of C_3 and C_4 grasses on the landscape. Consistent with this expectation, there was a significant relationship between the Bayesian estimates of % C_4 grass pollen in the surface-sediment samples from Africa, Australia, and North America and C_4-grass abundance around these sites (Fig. 5; n= 24, p< 0.001, r^2= 0.65). Furthermore, this relationship does not differ from a 1:1 relationship (p= 0.45), indicating no consistent bias in the representation of C_3 and C_4 grasses that may be associated with factors such as pollen productivities or preservation in sediments. We excluded one site, Rumuiku Swamp, from the regression because it had unusually low % C_4 grass pollen for its elevation, probably because the local swamp environment supported a greater abundance of C_3 grasses. However, the regression remains significant even if Rumuiku swamp is included (n=25, p< 0.001, r^2= 0.54). Nelson et al. (2008) found a similar relationship in North America using the original (-19.2‰) threshold method, but lacked data from sites with <20% C_4 grass abundance on the landscape. The additional data in the present study helps to extend this range and further validates SPIRAL as a tool for paleoenvironmental reconstruction.

3.3 Application to the paleorecord: interpreting SPIRAL $\delta^{13}C$ data in the Bayesian framework

The improved estimates of C_4-grass abundance from incorporation of SPIRAL data into the Bayesian model can help to assess factors (e.g. atmospheric CO_2 concentrations) controlling the origin, expansion, and variations in abundance of C_4 grasses in Earth's history. To demonstrate the application of the model to the paleorecord, we reevaluated a published SPIRAL dataset (Urban et al., 2010). Briefly, Urban et al. (2010) measured $\delta^{13}C$ of grass pollen grains in sediments spanning the early-Oligocene to middle-Miocene from sites in southwestern Europe and used a threshold value of -19.2‰ (before modification for variations in $\delta^{13}C$ of atmospheric CO_2 and aridity) to detect the presence of pollen from C_4 grasses. The samples in that study contained between 63 and 100 grains. Results indicated that C_4 grasses appeared on the landscape of southwestern Europe no later than the early Oligocene, which suggests that low pCO_2 may not have been the main driver and/or precondition for the development of C_4 photosynthesis in the grass family.

We evaluated the probability that the $\delta^{13}C$ data in samples from Urban et al. (2010) support the C_3-only model in our Bayesian analysis. We adjusted the $\delta^{13}C$ values of the Urban et al. (2010) samples to that of pre-industrial $\delta^{13}C$ of atmospheric CO_2 (-6.3‰) using estimated values of $\delta^{13}C$ of atmospheric CO_2 during the Cenozoic based on benthic foraminifera $\delta^{13}C$ data (Tipple et al., 2010). The probability of a C_3-only model was <0.01 (indicating >99% probability that at least some C_4 grains were present) for all samples (Electronic Annex EA-5). However, low water availability may have caused the $\delta^{13}C$ values of C_3 plants to shift in the positive direction (Ehleringer and Cooper, 1988). To account for the potential influence of aridity we shifted the mean δ^{13}C value of our C_3 training set by 1-3‰ in the positive direction, as in Urban et al. (2010). All but one sample had a P(C_3-only) of <0.01 after addition of 1‰ to the mean δ^{13}C value of the C_3 training dataset. Six of the eight samples, including the oldest two, had a P(C_3-only) of <0.05 after addition of 3‰ to the mean δ^{13}C value of the C_3 training dataset (Electronic Annex EA-5). The mean Bayesian estimates of % C_4 grass pollen are particularly high in the oldest two samples, consistent with the identification of plant communities in regions where today C_4 grasses are dominant as the closest analogs for the corresponding pollen assemblages (Suc, 1984). Therefore, our Bayesian estimates of % C_4 grass pollen confirm the prior conclusion of Urban et al. (2010) that C_4 grasses occurred on the landscape of southwestern Europe by at least the early Oligocene. The main advantage of the Bayesian model over the threshold approach used the context of the Urban et al. (2010) study is that the former allows for an explicit estimate of the probability of C_4 grasses being present on the landscape, which is essential for quantitatively assessing the timing of C_4-grass origin in geological history.

Overall, our new δ^{13}C data along with the Bayesian framework improve quantitative reconstructions of variation in the relative abundance of C_3 and C_4 grasses in response to environmental changes in the paleorecord. The flexible and hierarchical nature of the Bayesian model yields more accurate estimation of the abundance of C_4 grass pollen than the simpler, but biased, threshold approach, and also provides posterior model probabilities that enable hypothesis testing. Thus we recommend that future estimates of C_3 and C_4 grass abundances should, when possible, be made using Bayesian methods rather than threshold-based counting approaches. Bayesian analyses have begun to have important applications in the interpretations of geochemical isotope data. For example, recent studies have used Bayesian analysis to develop probabilistic region-of-origin assignments in wildlife and human forensics (Kennedy et al., 2011;

Wunder, 2010), enhance radiocarbon-age modeling for sediment records (Blaauw et al., 2007; Blaauw and Christen, 2011), and enable detection of climate-related shifts in elemental and isotopic abundances in peat cores (Gallagher et al., 2011). The increased use of Bayesian approaches promises to transform the environmental interpretations of geochemical data, especially in cases where small samples are involved. We expect that Bayesian analyses will become a mainstay of geochemistry.

ACKNOWLEDGEMENTS

We thank Triet Vuong for assistance with sample preparation and isotopic analysis, and Peter Kershaw, Patrick Moss, Alayne Street-Perrott, John Tibby and Dirk Verschuren for providing surface-sediment samples. We are also grateful to three anonymous reviewers and the associate editor for feedback that improved the manuscript. Funding was provided by NSF DEB-0816610 (FSH, DMN, AP) and NSF EF-1065848 (MCD). Sample collection in lakes Naivasha and Challa was done under Kenya MOEST research permit 13/001/11C to Dirk Verschuren.

REFERENCES

Amundson R., Evett R. R., Jahren A. H., and Bartolome J. (1997) Stable carbon isotope composition of Poaceae pollen and its potential in paleovegetational reconstructions. *Rev. Palaeobot. Palyno.* **99**, 17-24.

Blaauw M., Bakker R., Christen J. A., Hall V. A., and van der Plicht J. (2007) A Bayesian framework for age modeling of radiocarbon-dated peat deposits: case studies from the Netherlands. *Radiocarbon* **49**, 357-367.

Blaauw M. and Christen J. A. (2011) Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian Anal.* **6**, 457-474.

Cerling T. E., 1999. Paleorecords of C_4 plants and ecosystems. In: R. F. Sage and R. K. Monson Eds.), *C_4 Plant Biology*. Academic Press, San Diego.

Clark J. S., 2007. *Models for ecological data : an introduction*. Princeton University Press, Princeton.

Dellaportas P., Forster J. J., and Ntzoufras I. (2002) On Bayesian model and variable selection using MCMC. *Stat. Comput.* **12**, 27-36.

Descolas-Gros C. and Scholzel C. (2007) Stable isotope ratios of carbon and nitrogen in pollen grains in order to characterize plant functional groups and photosynthetic pathway types. *New Phytologist* **176**, 390-401.

Ehleringer J. R. and Cooper T. A. (1988) Correlations between carbon isotope ratio and microhabitat in desert plants. *Oecologia* **76**, 562-566.

Fægri K., Iverson J., Kaland P. E., and Krzywinski K., 1989. *Textbook of Pollen Analysis*. Wiley, New York.

Flessa K. W., Jackson S. T., Aber J. D., Arthur M. A., Crane P. R., Erwin D. H., Graham R. W., Jackson J. B. C., Kidwell S. M., Maples C. G., Peterson C. H., and Reichman O. J., 2005. *The geological record of ecological dynamics: Understanding the biotic effects of future environmental change*. National Academies Press, Washington, D.C.

Fraley C. and Raftery A. E. (2002) Model-based clustering, discriminant analysis, and density estimation. *J. Am. Stat. Assoc.* **97**, 611-631.

Friedli H., Lotscher H., Oeschger H., Siegenthaler U., and Stauffer B. (1986) Ice core record of the $^{13}C/^{12}C$ ratio of atmospheric CO_2 in the past two centuries. *Nature* **324**, 237-238.

Gallagher K., Bodin T., Sambridge M., Weiss D., Kylander M., and Large D. (2011) Inference of abrupt changes in noisy geochemical records using transdimensional changepoint models. *Earth Planet. Sc. Lett.* **311**, 182-194.

Hattersley P. W. (1983) The distribution of C_3 and C_4 grasses in Australia in relation to climate. *Oecologia* **57**, 113-128.

Hoff P. D., 2009. *A First Course In Bayesian Statistical Methods*. Springer, New York.

Jahren A. H. (2004) The carbon stable isotope composition of pollen. *Rev. Palaeobot. Palyno.* **132**, 291-313.

Kennedy C. D., Bowen G. J., and Ehleringer J. R. (2011) Temporal variation of oxygen isotope ratios ($\delta^{18}O$) in drinking water: Implications for specifying location of origin with human scalp hair. *Forensic Sci. Int.* **208**, 156-166.

Livingstone D. A. and Clayton W. D. (1980) An altitudinal cline in tropical African grass floras and its paleoecological significance. *Quaternary Res.* **13**, 392-402.

Marden J. I. (2000) Hypothesis testing: From p values to Bayes factors. *J. Am. Stat. Assoc.* **95**, 1316-1320.

McArdle B. H. (1988) The structural relationship - regression in biology. *Can. J. Zool.* **66**, 2329-2339.

Nelson D. M., Hu F. S., and Michener R. H. (2006) Stable-carbon isotope composition of Poaceae pollen: an assessment for reconstructing C_3 and C_4 grass abundance. *Holocene* **16**, 819-825.

Nelson D. M., Hu F. S., Mikucki J., Tian J., and Pearson A. (2007) Carbon isotopic analysis of individual pollen grains from C_3 and C_4 grasses using a spooling wire microcombustion interface. *Geochim. Cosmochim. Acta* **71**, 4005-4014.

Nelson D. M., Hu F. S., Scholes D. R., Joshi N., and Pearson A. (2008) Using SPIRAL (Single Pollen Isotope Ratio AnaLysis) to estimate C_3- and C_4-grass abundance in the paleorecord. *Earth Planet Sc. Lett.* **269**, 11-16.

Plummer M., 2011. JAGS Version 3.1.0 user manual.

Plummer M., 2012. rjags: Bayesian graphical models using MCMC. R package version 3-7. http://CRAN.R-project.org/package=rjags.

R Development Core Team (2010). R: A language and environment for statistical computing, in *R Foundation for Statistical Computing*.

Saugier B. and Roy J., 2000. Estimations of global terrestrial productivity: converging towards a single number. In: J. Roy, B. Saugier, and H. A. Mooney Eds.), *Global Terrestrial Productivity: Past, Present, and Future*. Academic Press, New York.

Savage L. J., 1962. *The Foundations of Statistical Inference: A Discussion*. Methuen & Co., London.

Smith R. J. (2009) Use and misuse of the reduced major axis for line-fitting. *Am. J. Phys. Anthropol.* **140**, 476-486.

Suc J. P. (1984) Origin and evolution of Mediterranean vegetation and climate in Europe. *Nature* **307**, 429-432.

Tieszen L. L., Reed B. C., Bliss N. B., Wylie B. K., and DeJong D. D. (1997) NDVI, C_3 and C_4 production, and distributions in great plains grassland land cover classes. *Ecol. Appl.* **7**, 59-78.

Tieszen L. L., Senyimba M. M., Imbamba S. K., and Troughton J. H. (1979) The distribution of C_3 and C_4 grasses and carbon isotope discrimination along an altitudinal and moisture gradient in Kenya. *Oecologia* **37**, 337-350.

Tipple B. J., Meyers S. R., and Pagani M. (2010) Carbon isotope ratio of Cenozoic CO_2: A comparative evaluation of available geochemical proxies. *Paleoceanography* **25**, PA3202.

Urban M. A., Nelson D. M., Jiménez-Moreno G., Chateauneuf J.-J., Pearson A., and Hu F. S. (2010) Isotopic evidence of C_4 grasses in southwestern Europe during the early Oligocene–middle Miocene. *Geology* **38**, 1091-1094.

Williams J. W., Shuman B. N., Webb T., Bartlein P. J., and Leduc P. L. (2004) Late-Quaternary vegetation dynamics in North America: scaling from taxa to biomes. *Ecol. Monogr.* **74**, 309-334.

Wunder M. B., 2010. Using Isoscapes to Model Probability Surfaces for Determining Geographic Origins. In: J. B. West, G. J. Bowen, T. E. Dawson, and K. P. Tu Eds.), *Isoscapes: Understanding Movement, Pattern, and Process on Earth Through Isotope Mapping*. Springer.

Figure legends

Figure 1. Conceptual diagram of the hierarchical Bayesian model used in this study. The likelihood function describes the probability distribution of $\delta^{13}C$ values for each pollen grain in a sample (y_i), given its classification as C_3 or C_4 ($x_i = 0$ or $x_i = 1$, respectively). The distribution of x_i in turn depends on θ, the proportion of C_4 grains in the population. Finally, the prior distribution of θ varies among sub-models representing three possible sample types (C_3-only, mixed, C_4-only). See Section 2.3 for details.

Figure 2. Histograms of $\delta^{13}C$ values from individual grains of grass pollen (1‰ bins). The dashed grey line represents data from C_3 grasses and the black line data from C_4 grasses (y-axis on left). The solid grey line represents the calculated probability of individual grains being classified as C_4 as a function of $\delta^{13}C$ (y-axis on right).

Figure 3. Estimated vs. true % of C_4 grains in samples of pseudodata derived from the herbarium training dataset. Columns correspond to three sample sizes (n=50, 100, and 150 grains). Rows correspond to results from Bayesian (top) and threshold (bottom) methods. For each panel, the mean (thick black line) and 95% confidence intervals (thin black lines) of estimates from 1000 random samples are plotted. The solid grey lines represent 1:1 relationships.

Figure 4. Probability that each candidate model (rows: C_3-only, mixed, and C_4-only) is the true model for pseudodata samples of known size (columns: 50, 100, or 150 grains) and composition (x-axis: 0-100% C_4). The dashed grey horizontal lines represent $p = 0.05$. For each set of

pseudodata samples, the box represents the 25-75th percentiles of posterior probabilities, with median indicated by a heavy black line. The whiskers encompass all remaining points within 1.5 times the interquartile range of the box, and points outside this range are plotted individually.

Figure 5. Estimated C_4 coverage (%) on the landscape around lakes in Africa (diamonds), Australia (X symbol), and North America (triangles), compared to the abundance of C_4 grass pollen (%) in the surface-sediments of these sites, as estimated from $\delta^{13}C$ of individual grains of grass pollen using the Bayesian model. The major axis slope is 0.97 and 95% confidence interval of the slope is 0.75 - 1.24. The data point with an asterisk is excluded from the regression, as explained in section 3.2. The 1:1 line is the solid grey line; the regression line is represented by the black dashed line. Error bars on each data point represent 95% confidence intervals.

Figure 1

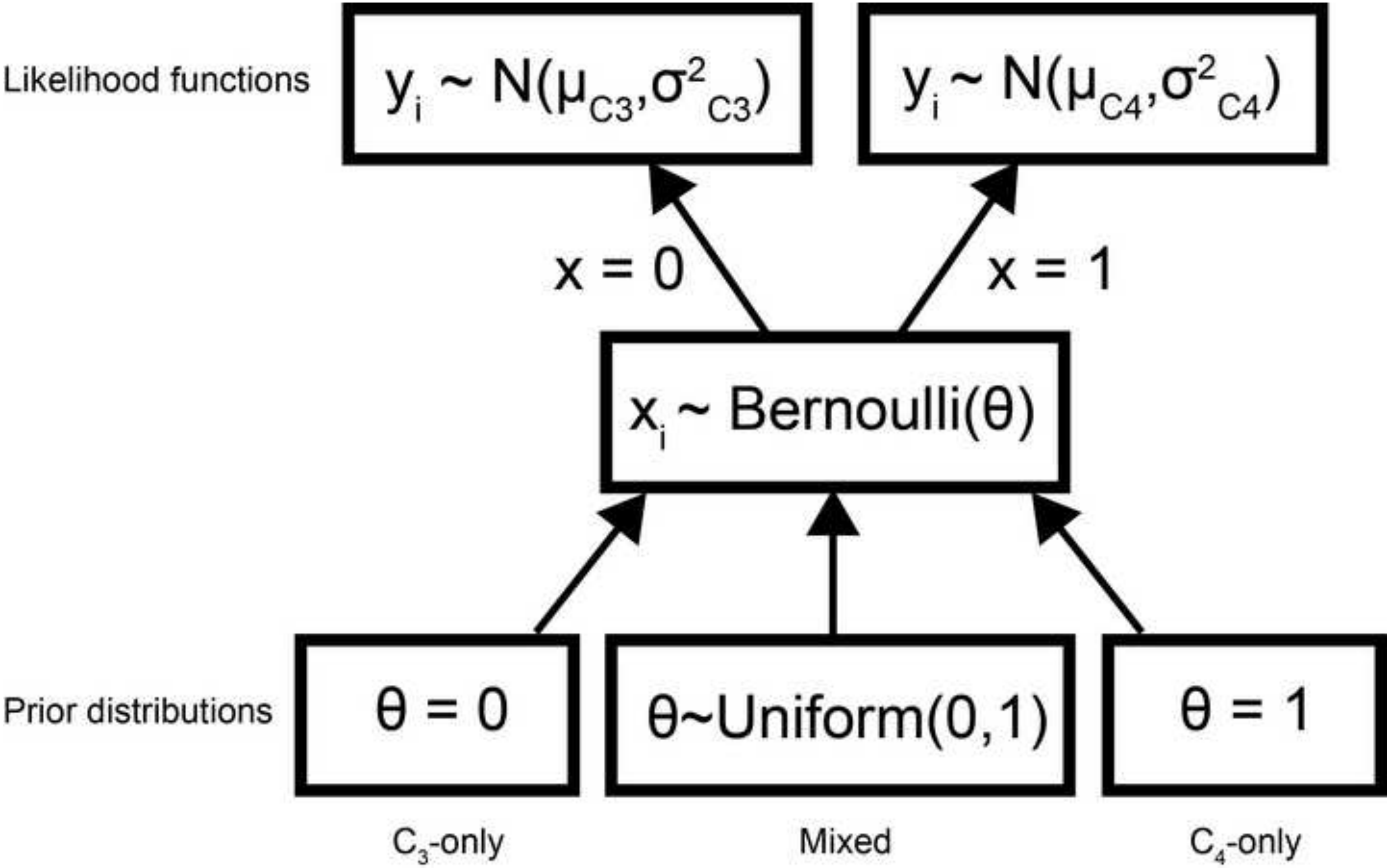

Likelihood functions
$y_i \sim N(\mu_{C3},\sigma^2_{C3})$
$y_i \sim N(\mu_{C4},\sigma^2_{C4})$
$x = 0$
$x = 1$
$x_i \sim \text{Bernoulli}(\theta)$
Prior distributions
$\theta = 0$
$\theta \sim \text{Uniform}(0,1)$
$\theta = 1$
C_3-only
Mixed
C_4-only

Figure 2

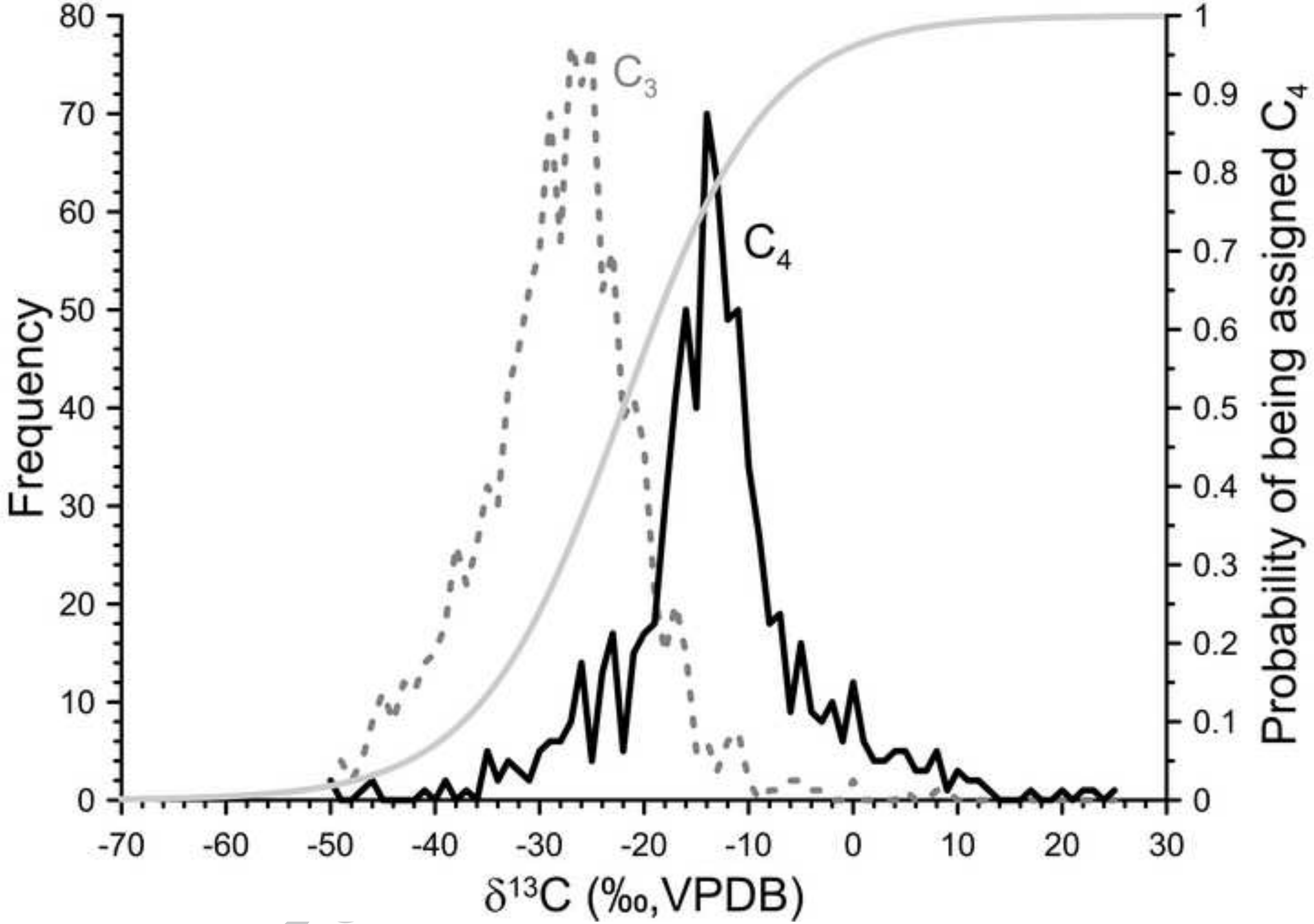

Frequency
80
70
60
50
40
30
20
10
0
-70
-60
-50
-40
-30
-20
-10
0
10
20
30
$\delta^{13}C$ (‰,VPDB)
C_3
C_4
Probability of being assigned C_4
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

Figure 3

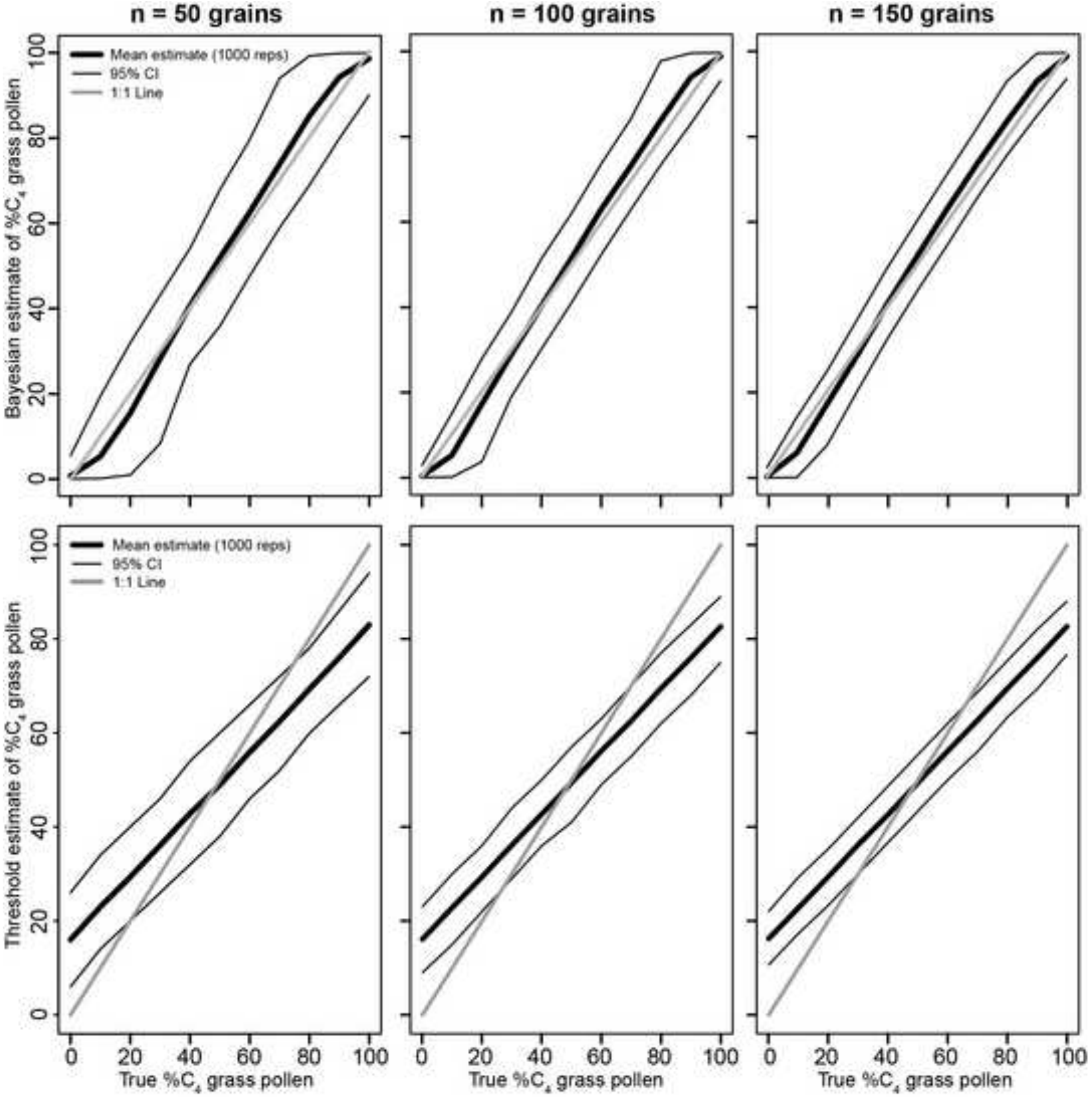

n = 50 grains
n = 100 grains
n = 150 grains
Mean estimate (1000 reps)
95% CI
1:1 Line
Bayesian estimate of %C_4 grass pollen
Threshold estimate of %C_4 grass pollen
0
20
40
60
80
100
True %C_4 grass pollen

Figure 4

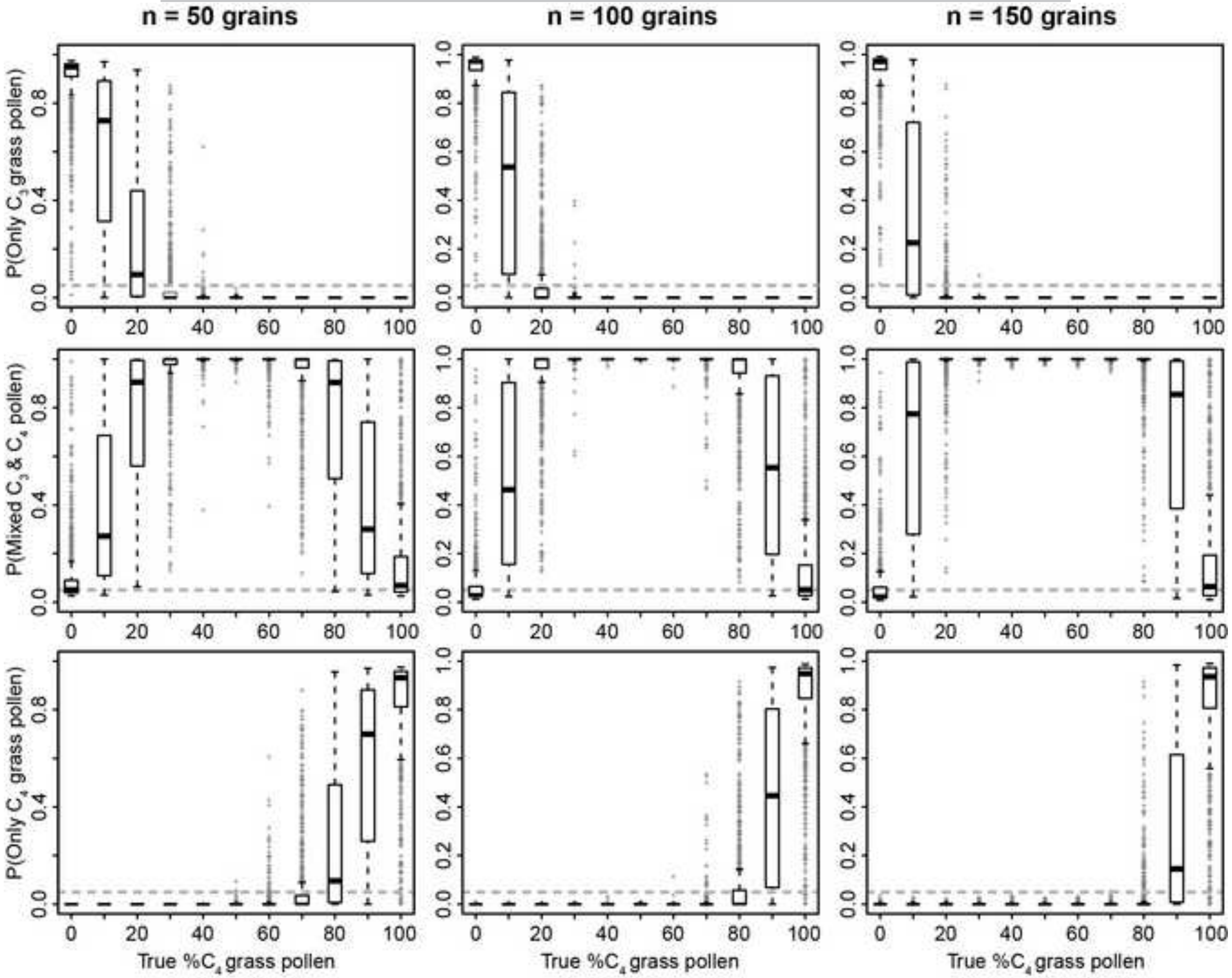
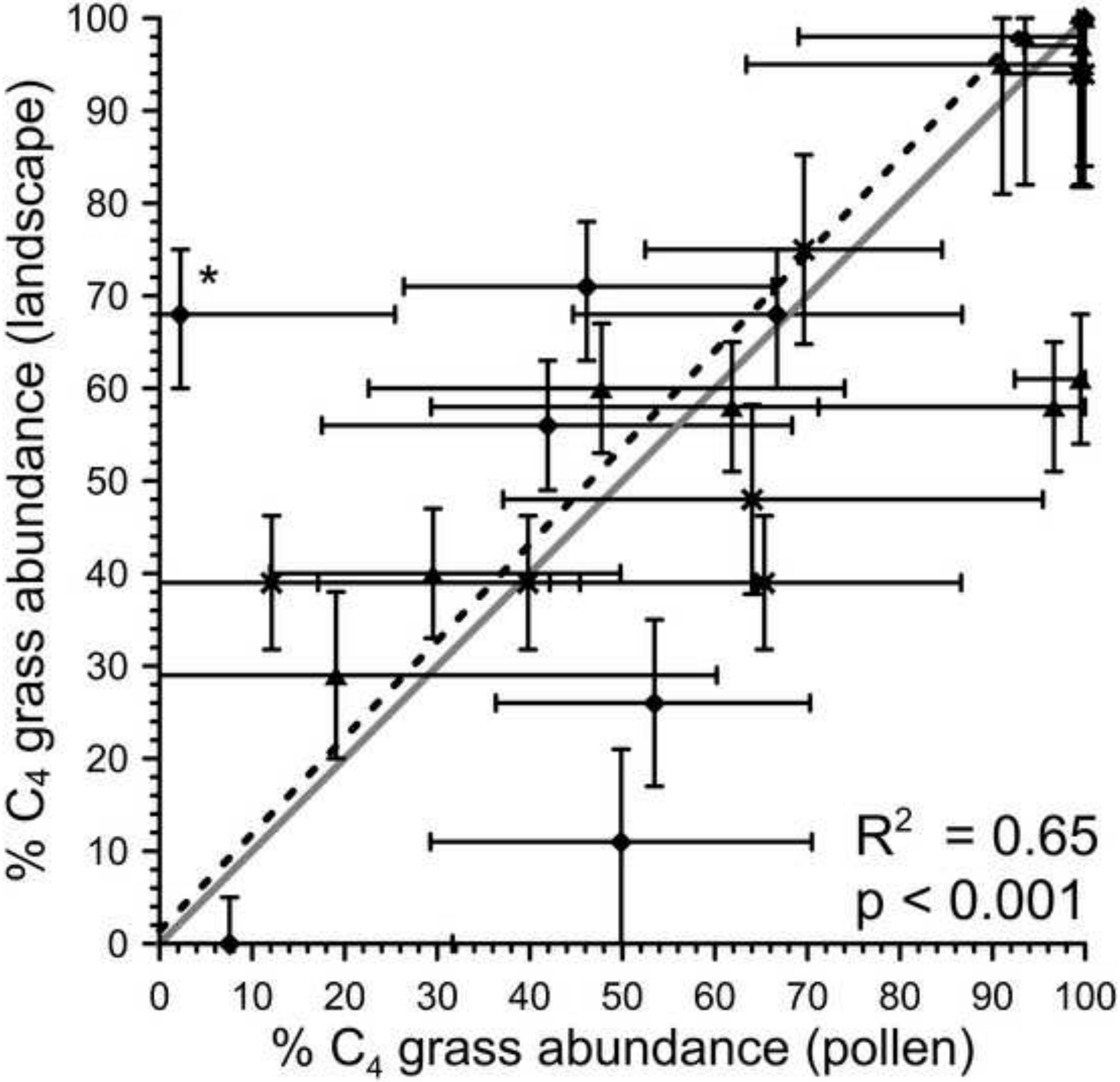

n = 50 grains
n = 100 grains
n = 150 grains
P(Only C_3 grass pollen)
P(Mixed C_3 & C_4 pollen)
P(Only C_4 grass pollen)
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
True %C_4 grass pollen

Figure 5


% C$_4$ grass abundance (landscape)
% C$_4$ grass abundance (pollen)
0
10
20
30
40
50
60
70
80
90
100
*
R^2 = 0.65
p < 0.001