GEOHORIZONS

Probabilistic correlation of single stratigraphic samples: A generalized approach for biostratigraphic data

Surangi W. Punyasena, Carlos Jaramillo, Felipe de la Parra, and Yuelin Du

ABSTRACT

Existing quantitative methods for biostratigraphic dating and correlation commonly ignore one of the key strengths of the microfossil record—relative abundance data. In this study, we present a maximum likelihood-based biostratigraphic method that demonstrates how microfossil abundance can be used in the stratigraphic placement of isolated samples. Precise correlation and dating of isolated paleontological samples is not possible with current methods, which are primarily intended for the alignment of longer stratigraphic sequences. In contrast, the probabilistic approach provided by likelihood analysis results in sample age estimates with defined confidence intervals. Therefore, all the uncertainties inherent in our age assessment (resulting from small sample sizes, incomplete sampling, imperfect knowledge of stratigraphic distributions, lack of taxonomic resolution of biostratigraphic data, and underlying environmental, paleogeographic, and sedimentologic processes) are explicit in our results. We conclude with a field test of the method, from data collected from an oil well from the Catatumbo Basin, Colombia, illustrating the use of our approach in a real-world case study and highlighting how our method could be generalized to a wide range of stratigraphic problems.

Manuscript received February 16, 2011; provisional acceptance May 5, 2011; revised manuscript received May 31, 2011; final acceptance June 20, 2011.
DOI:10.1306/06201111026

AUTHORS

Surangi W. Punyasena ~ *Department of Plant Biology, University of Illinois, 505 S. Goodwin Avenue, Urbana, Illinois 61801-3750; punyasena@life.illinois.edu*

Surangi W. Punyasena is an assistant professor in plant biology at the University of Illinois, Urbana-Champaign. She is a palynologist and paleoecologist with expertise in neotropical pollen identification and analysis. Her laboratory specializes in the development of statistical and computational tools for the evaluation of palynological data to test long-standing hypotheses in paleoecology, paleoclimate, and biostratigraphy.

Carlos Jaramillo ~ *Smithsonian Tropical Research Institute, Panama, Republic of Panama, P.O. Box 0843-03092, Panama; jaramilloc@si.edu*

Carlos Jaramillo is a staff scientist with the Smithsonian Tropical Research Institute in Panama. His research investigates the causes, patterns, and processes of tropical biodiversity at diverse scales of time and space. He is also interested in Cretaceous–Cenozoic biostratigraphy of low latitudes, developing methods for high-resolution biostratigraphy, and the paleobiogeography of Tethys.

Felipe de la Parra ~ *Instituto Colombiano del Petróleo, Km 7 via Piedecuesta, Santander, Colombia; felipe.delaparra@ecopetrol.com.co*

Felipe de la Parra is the head of the biostratigraphy team of Ecopetrol S.A., the state oil company of Colombia. His research focuses on using the tropical palynological record and geologic information to answer paleobiological questions. He is also interested in developing strong biostratigraphic frameworks and in finding novel ways of using paleontological and geologic information to improve oil exploration.

Yuelin Du ~ *Department of Electrical and Computer Engineering, University of Illinois, 1406 W. Green Street, Urbana, Illinois 61801-2918; du6@illinois.edu*

Yuelin Du is working toward his Ph.D. in electrical and computer engineering at the University of Illinois, Urbana-Champaign. He received his B.S. degree in microelectronics from Beijing University, Beijing, China and his M.S. degree

from the University of Illinois at Urbana-Champaign. His research interest is in computer-aided design.

ACKNOWLEDGEMENTS

Support for this research came from the University of Illinois School of Integrative Biology, the Colombian Petroleum Institute-Ecopetrol S.A., and the Smithsonian Institute.
The AAPG Editor thanks the following reviewers for their work on this paper: Joachim E. Amthor, Nick Fryer, and Robert W. Scott.

INTRODUCTION

Stratigraphic correlation is a fundamental component of sedimentary geology research. Consistent and reproducible correlations provide the linkage between disparate stratigraphic horizons that are critical to understanding relationships among geologic strata and to the testing of geologic hypotheses. Together with radiometric isotopes, microfossils with broad geographic distributions and narrow temporal ranges provide the largest data source for absolute and relative dating and correlation. The presence and absence of fossil species reflect the geologic history of evolution, extinction, and environmental change, creating a paleobiological clock that records the passing of time with changing paleontological assemblages.

Current methods of biostratigraphic correlation were primarily developed to compare stratigraphic series long enough to capture the turnover of taxa from these evolutionary and environmental processes. Biostratigraphers, however, are also commonly called on to provide an age or stratigraphic position in a sedimentary sequence of isolated samples. Isolated samples represent a single point in time, so therefore do not contain an extended sequence of first and last occurrences of multiple taxa. Assessments of such samples are normally reported as ages based on taxonomic composition (e.g., middle Eocene) with no corresponding evaluation of the certainty of the age estimate. In this study, we present a practical solution to the biostratigraphic dating of individual localities and an open-source downloadable program for duplicating our methodology. The approach presented differs from current biostratigraphic assessments by providing not only an age estimate for single isolated samples, but also a confidence interval of the strength of the age determination.

The concept is straightforward. We take advantage of the one feature of microfossil data that has not been used to a significant degree in quantitative biostratigraphy—the relative abundance of microfossil species. The prevailing wisdom is that local abundances of paleontological taxa are too noisy, too locally influenced to contain biostratigraphically useful information (Sadler, 2004). Truth to this assumption exists because local environmental conditions can be a large factor in the relative abundance of species, as demonstrated by modern ecological (e.g., Gentry, 1982, 1990; Phillips and Miller, 2002; Punyasena et al., 2008a) and Quaternary paleoecological (Birks and Birks, 1980; Faegri et al., 1989) studies.

However, at a regional scale, changing abundance patterns may also reflect the changing evolutionary dominance of species, from origin to extinction. Although a species may persist

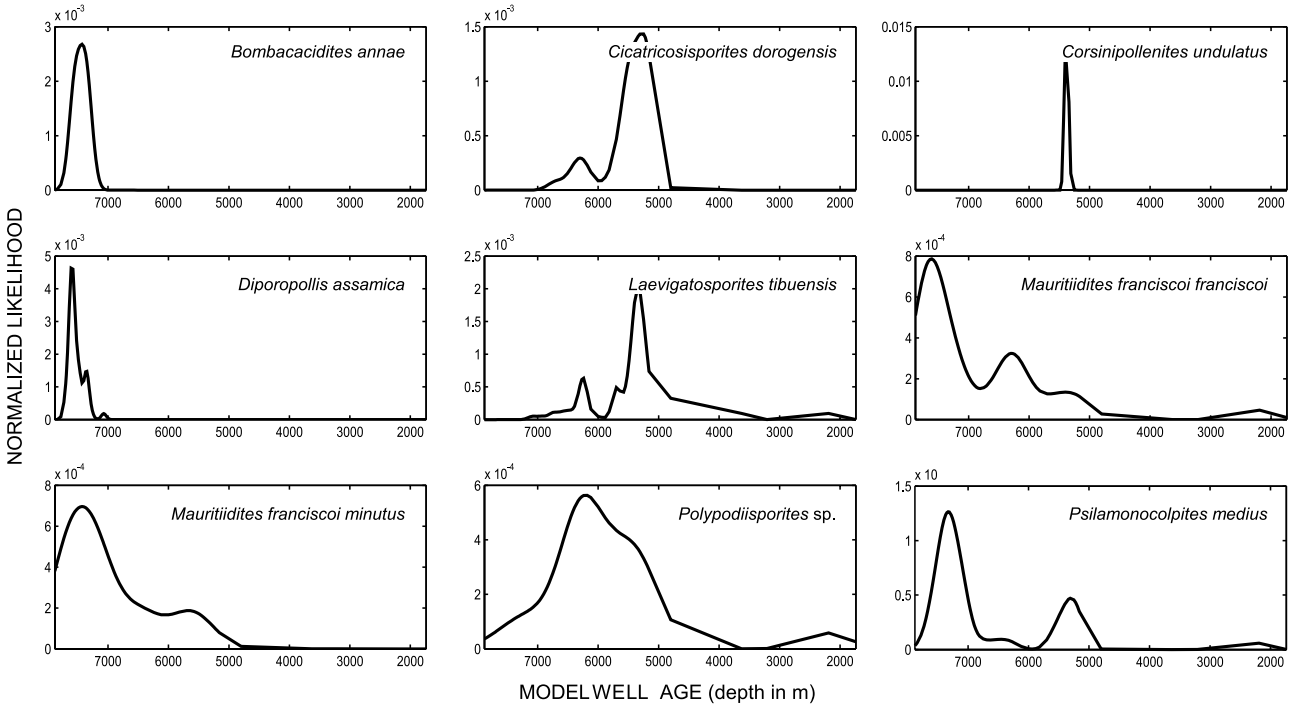

Figure 1. Examples of age-abundance distributions of nine taxa from the Catatumbo Basin, Colombia. Note that older taxa were overrepresented in the model well data set.

for an extended period, it may only be common for a much shorter time frame. This is particularly true for many plant species, which remain as marginal components of a community following the evolution and spread of more successful competitors (Knoll, 1984; Traverse, 1988). Ignoring this evolutionary trajectory in abundance discards potentially significant stratigraphic information. Abundance also can be potentially very useful when working at the reservoir scale. Changes in abundances of diverse taxa or groups of taxa are commonly useful in identifying flooding surfaces or surfaces of great correlation value.

Given the time-averaged nature of biostratigraphic samples, changes in the relative abundance of microfossils with large sample sizes are proportional to the changing regional dominance of a species, as well as local environmental conditions. Conservatively, abundances cannot be taken at face value. The potential for local environmental and preservational differences makes it inappropriate to align peaks and troughs in species abundances as in a standard time series analysis. However, in the method we present—weighted maximum likelihood of nonparametric probability densities—abundance serves instead as a weighting mechanism. Abundance data are used in two ways: (1) to provide shape to the known stratigraphic ranges of taxa based on the changing relative abundance of a taxon through time (Figure 1) and (2) to weigh the relative influence of a taxon in an isolated sample (equation 5). We avoid the assumption that absence is absolute by modeling the age-taxon relationship using a kernel density function and by only including taxa present in our isolated stratigraphic samples in our calculations. The method is derived from an approach used successfully in the paleoclimatic modeling of Quaternary palynological data (Punyasena, 2008; Punyasena et al., 2008b).

Because abundance data modify the influence of total stratigraphic ranges in our estimate, first and last occurrences have less impact on the age determination than periods when the taxon is most abundant. Because we also weight the relative influence of a taxon within a sample by abundance, the age estimate derived by our method also is not constrained by the range of a single occurrence of any given taxon. The final result is a likelihood distribution that describes the age of a sample as a range of probabilities. These values define the confidence

limits of our most likely age estimate. This provides an estimate of the inherent uncertainty of the data sample than current methods that provide a single statement of age. As in traditional biostratigraphy, our results ultimately rely on the quality and completeness of the collective knowledge of regional taxon distributions that define our age-taxon relationships, and this method is most appropriate for samples where the regional biostratigraphy is well known.

CURRENT METHODS FOR BIOSTRATIGRAPHIC CORRELATION

The most traditional, and still among the most common, approach to biostratigraphic zonation is manual correlation. The experienced biostratigrapher, relying on a combination of experience and field data, aligns sequences by hand, incorporating fossil absences as much as their occurrence, and mentally bridges gaps in stratigraphic data to produce coherent qualitative correlations (Gradstein, 2005). When records are relatively short and complete, this process of alignment is straightforward. However, the task becomes exponentially more difficult as the number of sections and taxa to be compared increases and when using temporally long-ranging taxa.

Most of the biostratigraphic sections are far from complete, and regional differences in species distributions exist because local first and last occurrences of fossil species do not reliably replicate global speciation and extinction events (Sadler, 2004). Conflicts in the ordering of paleontological data are therefore the norm in biostratigraphic alignments. Stratigraphic correlations need to account for the imperfections of the fossil record. Quantitative, but nonstatistical, approaches such as graphic correlation and zonation (Shaw, 1964; later refined by Edwards, 1978), allow the comparison of large data sets quickly and efficiently and the means of identifying conflicting biostratigraphic events. Graphic correlation works by simultaneously comparing the first and last occurrences of multiple species within two sections using lines of correlation. The solution—a final composite sequence—is reached by iteratively comparing the first and last occurrences of taxa in local sections with a hypothesized sequence that is refined with each iteration (Harper and Crowley, 1985).

Graphic correlation still works best with relatively complete paleontological records, such as those represented by microfossils (Gradstein, 2005). However, increasingly sophisticated statistical algorithms have been developed to use more fragmented and contradictory sequences (Alroy, 1992, 1994, 2000; Agterberg and Gradstein, 1999; Sadler and Cooper, 2003; Sadler, 2004). These methods have improved on manual methods of ordering the occurrences of paleontological species by providing algorithmic solutions to the alignment of first and last appearances. These statistical solutions have allowed the incorporation of more challenging paleobiological sequences in stratigraphic correlations, such as those represented by vertebrate (Alroy, 1992, 1994, 2000) and macroinvertebrate data (Sadler and Cooper, 2003; Sadler, 2004).

The two most widely adopted probabilistic methods, CONOP (constrained optimization) (Sadler, 2004) and RASC (ranking and scaling) (Agterberg and Gradstein, 1999), seek to minimize contradictions in the alignment of biostratigraphic events, although through very different means. The CONOP method uses simulated annealing algorithms to find the sequences, with contradictions penalized. The RASC method aligns the average stratigraphic position of up to five correlative events: first occurrence, first consistent occurrence, peak occurrence, last consistent occurrence, and last occurrence. These methods are most successful in identifying solutions when the stratigraphic sections used have sufficient temporal length to capture multiple instances of taxonomic turnover in fossil composition. The greater the number of biostratigraphic events, such as the number of first and last occurrences, and the more complete the fossil representation, the more likely a stable alignment will be found.

All three methods (graphic correlation, CONOP, and RASC) were developed for correlating stratigraphic sequences that represent relatively long time series. Correlations are based on the ordering and alignment of stratigraphic events, such as the

first and last occurrences of a species. When attempting to correlate a single sample with a longer stratigraphic column, this approach may be inconclusive because no first and last species occurrence data are available in an isolated sample, only relative abundances. Using only first and last occurrences to correlate biostratigraphic data recovered from a single locality means that samples could only be dated to a range of time in which species found in a sample co-occur. Unless the sample captures the presence of a taxon with a singularly limited temporal range, the estimated age will range broadly, providing little stratigraphic precision.

APPLYING MAXIMUM LIKELIHOOD TO STRATIGRAPHIC CORRELATION

Placing a single isolated microfossil sample in a stratigraphic column starts with an accurate representation of known distribution of microfossil abundances. The composite representation of the entire column provides the baseline model of changes in fossil assemblages over time to which a single sample can be compared. This composite column, which can also be described as the biostratigraphic zonation, represents our entire a priori knowledge of paleontological ranges and diversity. It includes all known taxa and abundance data for all time slices. The assumption is that abundance changes are caused by changing taxonomic abundances through time. Therefore, comparisons should be limited to strata and samples that share similar depositional environments. The composite column can be derived using any of the correlation methods used above, if vetted appropriately. The accuracy and completeness of the composite column will affect the accuracy of the resulting sample estimates.

Changes in taxon abundances in the composite column can be described in absolute measures of time (e.g., years) when radiometric dates and an accurate time-depth model are available, as relative measures of time (e.g., older or younger), or when age is related to stratigraphic position. In our demonstration example, our composite column forms a model well where time is represented as depth.

The taxon abundances of the composite column provide the comparative data against which all single sample estimates are matched. First, each taxon in the composite column is modeled as a probability density function using a nonparametric kernel density estimator (Bowman and Azzalini, 1997). This probability density represents the likelihood of finding a particular taxon at a particular time. The shape of this probability density curve is calculated directly from the empirical abundance data available in the composite core. Kernel density estimators do not have a fixed structure and depend on all data points to reach an estimate; therefore, the resulting density curve is a nonparametric distribution.

The probability density of a given age or depth, represented as x, is calculated as

$$f(x) = \frac{1}{n}\sum K\left(\frac{x - x(i)}{h}\right), i = 1, \ldots, n \qquad (1)$$

where n is the total number of observations, K is the kernel function, and h is the bandwidth or window used. The contribution of all other data points, $x(i)$, to the density estimate at point x depends on the distance between $x(i)$ and x, the kernel function used, and the bandwidth across which the kernel function is calculated.

Density estimates can be fitted to several kernel shapes. We used the Gaussian kernel in this analysis, which fits the data to a normal distribution and is defined as

$$K(t) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \qquad (2)$$

where n is the total number of observations, and t is the parenthetical value described in equation 1, the difference between $x(i)$ and x, divided by the bandwidth. Although a Gaussian kernel is used to weigh the relative contribution of the abundance data to the density function, this does not mean that a normal probability distribution will be the result. In fact, the shape of the density function is affected more by the size of the bandwidth (h) than by the kernel used. No single best approach for determining the optimal bandwidth for a given

sample exists. Instead, we used a default method known as "Silverman's rule of thumb" (Silverman, 1986), which derives a conservative bandwidth size based on the standard deviation (σ) and the number of observations (n) of the data:

$$h = \sigma\left(\frac{4}{3}n\right)^{-\frac{1}{5}} \tag{3}$$

Figure 1 provides examples of probability densities of taxa from our model system generated by the Gaussian kernel density function.

Once the probability density functions for all the taxa found in the composite core are derived, the age estimation of an unknown sample (θ) is fairly straightforward; the likelihood distribution of the age estimate, $L(\theta)$, is the product of the density functions of all the taxa (F_1, F_2, … F_n) found in the unknown single sample:

$$L(\theta) = \Pr(F_1, F_2 \ldots F_n|\theta) \tag{4}$$

This is the equivalent of the sum of the log likelihoods of these taxa. To give the most abundant taxa, the greatest influence on the final age estimate, and conversely, minimize the effect of rare and single-occurrence taxa, our density functions are weighed by the normalized relative abundance of each taxon (c_i):

$$\ln L(\theta) = \sum_{i=1}^{n} c_i[\ln \Pr(F_i \mid \theta)] \tag{5}$$

The result is a vector of probability estimates for each time slice, which can be compared when normalized by a uniform prior probability.

The maximum likelihood value, L(θ) from equation 5, represents the most likely age. Like the initial probability distributions for each taxon, the resulting likelihood vector can be multimodal, with competing age estimates. This reflects the underlying uncertainty and incompleteness of the microfossil data in the isolated sample and the composite core. The integral value of the area underneath the likelihood function describes the certainty of an estimate. A broad flat distribution indicates a large degree of uncertainty, and a narrow peaked distribution describes a high degree of certainty.

FIELD APPLICATION OF THE PROBABILISTIC CORRELATION METHOD

We applied our method to an exploratory well recently drilled by the Colombian petroleum company Ecopetrol in the Eocene-age Catatumbo Basin of northern Colombia. Because of the structural complexities of the underlying geologic formations, with many faults and folds with high-dipping flanks, palynology was a critical component in determining the stratigraphic position of the drilling bit during the first phase of exploratory drilling. Our biostratigraphic model (the composite stratigraphic column) was developed from 79 regional cores (Jaramillo et al., 2011). The cores included abundance data for 179 palynomorph species and span six formations, from oldest to youngest: Colon-Mito Juan, Catatumbo, Barco, Los Cuervos, Mirador, and Carbonera. Graphic correlation was used to align this composite (Jaramillo et al., 2011).

A total of 24 individual samples were analyzed on-site during the drilling of an exploratory well (Table 1). Given the regional geology, the exploratory well was expected to drill through the Mirador and Los Cuervos formations.

Composite abundance data were used to calculate taxon-age probability density functions (equation 1) for each palynomorph species across the entire column. This produced a "model well" of known taxon-age distributions. Because of the lack of an age-depth model, we used stratigraphic position, instead of absolute ages, to locate samples in time. Depth was also used because one of the purposes of biostratigraphy during the drilling was determining the key stratigraphic positions that engineer drillers needed to establish for several reasons (e.g., casing points, drilling-bit changes, fluid-pressure variations). These taxon-age probability densities were then used to calculate the weighted likelihood (equation 4) of each of the 24 single samples occurring at a given depth. Based on the results, we place the Mirador–Los Cuervos

Table 1. The 24 Samples Illustrated in Figure 2 and the Palynological Species Used in Our Age Determinations

Sample	Sample Count*	Species Used in Age Estimate
OST1-6570	6	*Laevigatosporites tibuensis, Psilatriletes* sp. 25–50, *Retistephanoporites minutiporus, Spirosyncolpites spiralis*
OST1-6630	4	*Laevigatosporites tibuensis, Psilatriletes* sp. 25–50, *Mauritiidites franciscoi minutus*
OST1-6690	7	*Mauritiidites franciscoi franciscoi, Psilatriletes* sp. 25–50, *Cicatricosisporites dorogensis, Polypodiisporites* sp., *Retitrescolpites magnus*
OST1-6710	2	*Psilatriletes* sp. 25–50, *Cicatricosisporites dorogensis*
OST1-6750	10	*Psilamonocolpites medius, Laevigatosporites tibuensis, Psilatriletes* sp. 25–50, *Mauritiidites franciscoi minutus, Psilatriletes* <25
OST1-7050	5	*Psilatriletes* sp. 25–50, *Mauritiidites franciscoi minutus, Psilatriletes* <25
OST1-7070	7	*Diporopollis assamica, Mauritiidites franciscoi minutus, Psilamonocolpites medius, Psilatriletes* sp. 25–50
OST1-7080	9	*Psilamonocolpites medius, Psilatriletes* sp. 25–50, *Polypodiisporites* sp., *Retitricolpites* sp.
OST1-7090	13	*Mauritiidites franciscoi franciscoi, Proxapertites psilatus, Psilamonocolpites medius, Laevigatosporites tibuensis, Psilatriletes* sp. 25–50, *Polypodiisporites* sp.
OST1-7130	6	*Proxapertites operculatus, Psilamonocolpites medius, Psilatriletes* sp. 25–50, *Verrutriletes* sp.
OST1-7140	15	*Proxapertites operculatus, Proxapertites psilatus, Psilamonocolpites medius, Psilatriletes* sp. 25–50, *Mauritiidites franciscoi minutus*
OST1-7150	7	*Mauritiidites franciscoi franciscoi, Proxapertites operculatus, Proxapertite spsilatus, Psilatriletes* sp. 25–50, *Polypodiisporites* sp.
OST1-7220	4	*Proxapertites psilatus, Psilamonocolpites medius, Psilatriletes* sp. 25–50
OST1-7280	9	*Diporopollis assamica, Mauritiidites franciscoi franciscoi, Proxapertites psilatus, Psilamonocolpites medius, Polypodiisporites* sp., *Bombacacidites annae*
OST1-7350	76	*Diporopollis assamica, Proxapertites operculatus, Proxapertites psilatus, Psilamonocolpites medius, Psilatriletes* sp. 25–50, *Mauritiidites franciscoi minutus, Psilatriletes* <25, *Bombacacidites annae*
OST1-7370	44	*Diporopollis assamica, Mauritiidites franciscoi franciscoi, Proxapertites operculatus, Proxapertites psilatus, Psilamonocolpites medius, Psilatriletes* sp. 25–50, *Psilatriletes* <25, *Bombacacidites annae, Foveotricolpites perforatus*
OST1-7400	8	*Proxapertites operculatus, Psilatriletes* sp. 25–50, *Bombacacidites annae*
OST1-7460	61	*Diporopollis assamica, Mauritiidites franciscoi franciscoi, Proxapertites operculatus, Proxapertite spsilatus, Psilamonocolpites medius, Psilatriletes* sp. 25–50, *Retistephanoporites minutiporus, Psilatriletes* <25, *Bombacacidites annae, Proxapertites minutihumbertoides*
OST1-7490	35	*Diporopollis assamica, Proxapertites operculatus, Proxapertites psilatus, Psilamonocolpites medius, Psilatriletes* sp. 25–50, *Mauritiidites franciscoi minutus, Psilatriletes* <25
OST1-7520	14	*Diporopollis assamica, Proxapertites operculatus, Proxapertites psilatus, Psilamonocolpites medius, Polypodiisporites* sp.
OST1-7550	48	*Diporopollis assamica, Proxapertites operculatus, Proxapertites psilatus, Psilatriletes* sp. 25–50, *Mauritiidites franciscoi minutus, Psilatriletes* <25, *Bombacacidites annae, Monocolpopollenites ovatus, Proxapertites cursus*
OST1-7580	52	*Diporopollis assamica, Proxapertites operculatus, Proxapertites psilatus, Psilamonocolpites medius, Psilatriletes* sp. 25–50, *Retidiporites magdalenensis*
OST1-7610	56	*Diporopollis assamica, Proxapertites operculatus, Proxapertites psilatus, Psilamonocolpites medius, Psilatriletes* sp. 25–50, *Verrutriletes* sp., *Bombacacidites annae, Proxapertites cursus*
OST1-7640	42	*Diporopollis assamica, Mauritiidites franciscoi franciscoi, Proxapertites operculatus, Proxapertites psilatus, Psilamonocolpites medius*

*Sample count represents the total number of palynomorphs used in the age determination of each sample.

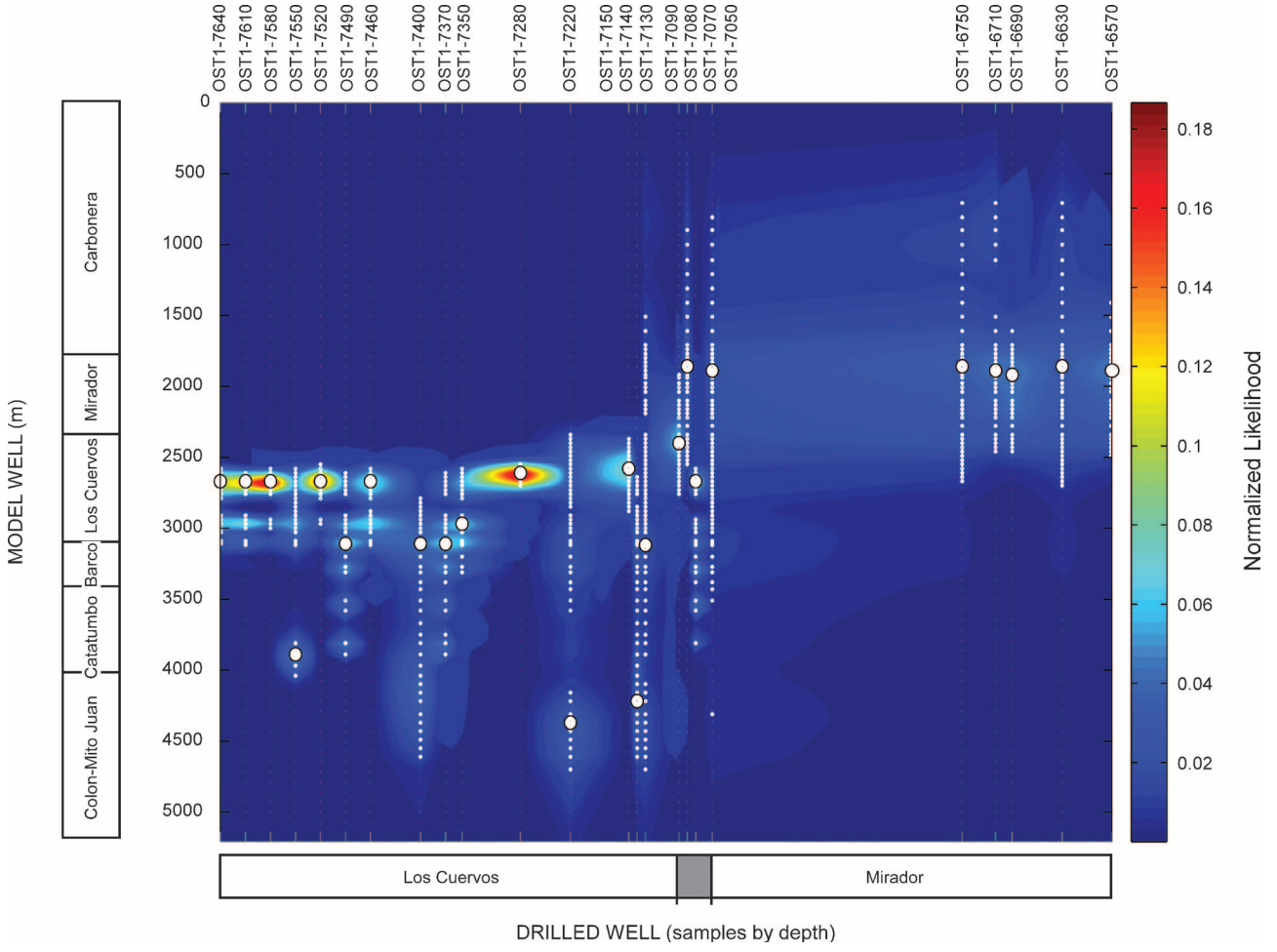

Figure 2. Stratigraphic age estimates for 24 isolated samples recovered from exploratory drilling of the Catatumbo Basin, Colombia. Normalized likelihood values are represented by color, with higher and lower likelihood values represented by red and blue, respectively. Maximum likelihood estimates (MLEs) are denoted by filled white circles. The 95% confidence limits of the MLEs are empirically defined by the white dots.

boundary between samples OST1-7090 and OST1-7080 (Figure 2).

The stratigraphic positions produced by the analysis and, in particular, the delimitation of the Mirador–Los Cuervos boundary agreed with the assessment of the drilling operations geologist. One sample result (OST1-7070) was out of sequence; this sample was identified as falling within the older Los Cuervos model well age range when the sample sequentially fell within the younger Mirador samples (Figure 2). One explanation for this discrepancy is that this particular sample had a very low recovery of palynomorphs (sample size = 7). Only four species were found, and all were taxa with stratigraphic ranges that span the Cenozoic (*Diporopollis assamica*, *Mauritiidites franciscoi minutus*, *Psilamonocolpites medius*, *Psilatriletes* sp. 25-50) (Table 1). *Psilatriletes* is first found in the Lower Cretaceous, *P. medius* appears in the Maastrichtian, and *M. franciscoi minutus* is first found in the lower Paleocene. All three taxa are present through the Holocene. The age determination comes solely from the fungi *D. assamica* that is common in the Los Cuervos Formation but tends to be more abundant toward the middle of the formation, not the top.

Younger estimates from the Mirador Formation tended to have larger confidence intervals than those of the older samples taken from the Los Cuervos Formation. The lower certainty in the younger sample results primarily reflects a bias in

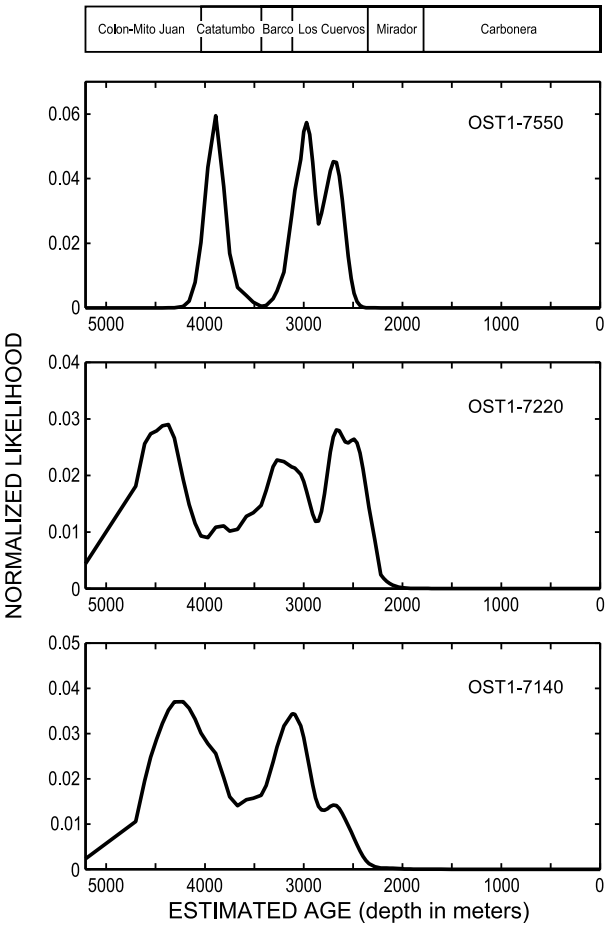

Figure 3. Normalized likelihood age estimates for the three Catatumbo Basin samples (OST1-7550, OST1-7220, and OST1-7140) with older than expected age estimates. Note the multimodal distribution of the age estimates.

the model well taxa. Most of the 68 model well taxa available in this analysis were most abundant in the Los Cuervos and older formations. This made older age estimates more likely than younger ones. Three samples (OST1-7550, OST1-7220, and OST1-7140) produced age estimates that were much older than would be expected based on the geology of the drilled well. These samples were estimated to fall within the Colon-Mito Juan and Catatumbo formations based on our analysis. The results, however, were multimodal (Figure 3). Alternate peaks placed these samples in the Los Cuervos Formation—meaning that the second-best age estimates are what would be expected given the sample sequence.

Despite the caveats of low palynomorph sample size, low species diversity, and biases in the model well data, our analysis demonstrates in a real-life example, with data taken on-site during drilling (e.g., with not enough time to do high counts per sample), that the age estimates provided by a likelihood approach to biostratigraphic age determination are reliable and can be used in real-life situations.

CONCLUSIONS

The Catatumbo Basin palynomorphs demonstrate how a fairly simple application of maximum likelihood analysis can be a powerful correlation and dating tool of isolated samples in biostratigraphy. Our results indicate that this method can provide a quantitative means of correlating single samples, incorporating all known biostratigraphic information on taxon-age distributions and providing confidence limits on an age assessment. The accuracy of the age estimates is directly related to the quality of the regional biostratigraphic information available. The results will improve as the composite data for the biostratigraphic age–abundance model becomes more comprehensive with increased exploration.

Notably, the described method can also be applied to problems other than pure biostratigraphy. Any variable that changes along a gradient (e.g., depth of a well) can be analyzed, including well-logging data (gamma ray, resistivity, density, among others) and geochemical and petrophysical properties of a rock. For example, we have combined lithologic data from the formation evaluation log with biostratigraphic information to calculate the probability of being at any given stratigraphic position during the drilling of structurally complex wells. This approach has proven to be very successful because not only do we produce a quantitative estimation of the stratigraphic position, but also we can quantify the level of certainty in our estimation. We anticipate that many applications will be developed using the simple, but powerful, analytical tool presented here.

A downloadable version of the software used to complete our analysis and generate Figure 2 is available from Punyasena and Du (2011).

REFERENCES CITED

Agterberg, F. P., and F. M. Gradstein, 1999, The RASC method for ranking and scaling of biostratigraphic events: Earth Science Reviews, v. 46, p. 1–25, doi:10.1016/S0012-8252(99)00007-0.

Alroy, J., 1992, Conjunction among taxonomic distributions and the Miocene mammalian biochronology of the Great Plains: Paleobiology, v. 18, p. 326–343.

Alroy, J., 1994, Appearance event ordination: A new biochronologic method: Paleobiology, v. 20, p. 191–207.

Alroy, J., 2000, New methods for quantifying macroevolutionary patterns and processes: Paleobiology, v. 26, p. 707–733, doi:10.1666/0094-8373(2000)026<0707:NMFQMP>2.0.CO;2.

Birks, H. J. B., and H. H. Birks, 1980, Quaternary paleoecology: London, Edward Arnold, v. viii, 289 p.

Bowman, A. W., and A. Azzalini, 1997, Applied smoothing techniques for data analysis: The kernel approach with S-Plus illustrations—Oxford statistical science series: Oxford, Oxford University Press, v. xi, 208 p.

Edwards, L. E., 1978, Range charts and no-space graphs: Computers and Geosciences, v. 4, p. 247–255, doi:10.1016/0098-3004(78)90057-2.

Faegri, K., P. E. Kaland, and K. Krzywinski, 1989, Textbook of pollen analysis: Chichester, Wiley, v. x, 328 p.

Gentry, A. H., 1982, Patterns of neotropical plant species diversity: Evolutionary Biology, v. 15, p. 1–85.

Gentry, A. H., 1990, Four neotropical rainforests: New Haven, Yale University Press, v. xiii, 627 p.

Gradstein, F. M., 2005, Quantitative methods for applied microfossil biostratigraphy, *in* E. A. M. Koutsoukos, ed., Applied stratigraphy: Dordrecht, Springer, p. 361–388.

Harper, C. W., and K. D. Crowley, 1985, Insights on why graphic correlation (Shaw method) works: A discussion: Journal of Geology, v. 93, p. 503–506, doi:10.1086/628969.

Jaramillo, C., M. Rueda, and V. Torres, 2011, A palynological zonation for the Cenozoic of the Llanos foothills of Colombia: Palynology, v. 35, p. 46–84, doi:10.1080/01916122.2010.515069.

Knoll, A. H., 1984, Patterns of extinction in the fossil record of vascular plants, *in* M. H. Nitecki, ed., Extinctions: Chicago, University of Chicago Press, p. 21–68.

Phillips, O., and J. S. Miller, 2002, Global patterns of diversity: Alwyn H. Gentry's forest transect data set: St. Louis, Missouri Botanical Garden Press, v. 89, 319 p.

Punyasena, S. W., 2008, Estimating neotropical paleotemperature and paleoprecipitation using plant family C*limatic optima*: Palaeogeography, Palaeoclimatology, Palaeoecology, v. 265, p. 226–237, doi:10.1016/j.palaeo.2008.04.025.

Punyasena, S. W., and Y. Du, 2011, Biostratigraphy files download, https://www.life.illinois.edu/punyasena/download/ (accessed June 2, 2011).

Punyasena, S. W., G. Eshel, and J. C. McElwain, 2008a, The influence of climate on the spatial patterning of neotropical plant families: Journal of Biogeography, v. 35, p. 117–130, doi:10.1111/j.1365-2699.2007.01773.x.

Punyasena, S. W., F. E. Mayle, and J. C. McElwain, 2008b, Quantitative estimates of glacial and Holocene temperature and precipitation change in lowland Amazonian Bolivia: Geology, v. 36, p. 667–670, doi:10.1130/G24784A.1.

Sadler, P. M., 2004, Quantitative biostratigraphy: Achieving finer resolution in global correlation: Annual Review of Earth and Planetary Sciences, v. 32, p. 187–213, doi:10.1146/annurev.earth.32.101802.120428.

Sadler, P. M., and R. A. Cooper, 2003, Best-fit intervals and consensus sequences: Comparison of the resolving power of traditional biostratigraphy and computer-assisted correlation, *in* P. J. Harries, ed., High-resolution approaches in stratigraphic paleontology: Dordrecht, Kluwer Academic Publishers, p. 49–94.

Shaw, A. B., 1964, Time in stratigraphy: New York, McGraw-Hill, v. xix, 365 p.

Silverman, B. W., 1986, Density estimation for statistics and data analysis: Monographs on statistics and applied probability: London, Chapman and Hall, 175 p.

Traverse, A., 1988, Plant evolution dances to a different beat: Historical Biology, v. 1, p. 277–301, doi:10.1080/08912968809386480.