

More than three million years of community evolution. The temporal and geographical resolution of the Plio-Pleistocene Western Eurasia mammal faunas

P. Raia^{a,*}, F. Carotenuto^a, C. Meloro^{a,c}, P. Piras^b, C. Barbera^a, T. Kotsakis^b

^a Dipartimento Scienze della Terra, Università degli Studi di Napoli Federico II, L.go San Marcellino, 10, 80138, Napoli, Italy

^b Dipartimento di Geologia, Università Di Roma 3, L.go San Leonardo Murialdo 1, 00146, Roma, Italy

^c The University of Hull, HYMS, Loxley Building, Cottingham Road, Hull HU6 7RX, UK

ARTICLE INFO

Article history:

Received 3 September 2008

Received in revised form 26 January 2009

Accepted 7 February 2009

Keywords:

Biogeography
Community evolution
Quaternary
Large mammals
PCOM

ABSTRACT

In this study we collected, in a database, faunal lists for more than 800 Eurasian fossil localities yielding large mammal remains so as to produce continental-level ecologic-evolutionary units (Eurasian Paleocommunities, EA PCOMs) via bootstrapped cluster analysis. EA PCOMs are meant to represent taxonomically distinct assemblages of species. EA PCOMs allow the evolution of large mammal assemblages to be traced both in time and space. This latter attribute (spatial resolution) represents the most important innovation here and contrasts with classic biochronologic schemes, from which EA PCOMs clearly depart. The merits and limitations of this innovation are discussed in detail.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

In 2005 we published a paper in *Palaeogeography Palaeoclimatology Palaeoecology* presenting a new method to define extinct large mammal paleocommunities of the Italian Plio-Pleistocene (Raia et al., 2005; Raia et al., 2006a). Largely biochronological in essence, the method applies cluster analysis on faunal lists of fossil sites (or single stratigraphic horizons within sites, when appropriate). Then, it selects among the different partition levels the resulting dendrogram offers via a statistical (bootstrap randomization) procedure, hence its name bootstrapped cluster analysis (BCA). BCA bears advantages over simple cluster analysis, as it allows discriminating among partition levels (that is the number of paleocommunities that actually where there), and on classic biochronology, for it avoids using subjective criteria such as experts' choices of specific taxa to separate biochrons, which could be severely misleading (Walsh, 1998).

The paleocommunities we recognized by BCA, named PCOMs, proved exceptionally useful to our goal, which was to get ecologically sound, distinct in time, ensembles of species fully amenable to evolutionary-ecological investigations. PCOMs replicate living mammal communities in occupancy frequency distribution (Raia et al., 2006b), body mass distribution (Raia et al., 2006b), predator/prey ratios (Raia et al., 2007) and abundances (Meloro et al., 2007). Thereby, they allow

studying how occupancy, predation, and body mass changed through time and influenced species survival, interacting with each other and with presumed climate change. Climatic effects were also shown to control taxonomic turnover across communities (Raia et al., 2005) in agreement with the turnover pulse hypothesis (Vrba, 1995) and diversity patterns (Meloro et al., 2008).

Here we extend the geographic coverage of that study to Western Eurasia, and increase the time span by including faunas down to the early Holocene (Italian faunal assemblages we drew PCOMs from were not younger than 150 ky). Our aim is to provide continental-level paleocommunities, which we name EA PCOMs (Eurasian PCOMs), as a tool for studying large-scale community evolution in Western Eurasia, including, in perspective, the study of interactions of regional and local phenomena on species dispersal, survival, and macroecology (cf. Gaston and Blackburn, 2000).

At larger spatial and temporal scales, we could not anticipate if BCA would work as well as with the Italian fossil faunas in our previous studies. As a consequence, we tried to improve our methodology. A potential limitation of BCA is that time is not a built-in variable; thereby biochronologic reasoning should be relied upon a posteriori to arrange the paleocommunities (or, for that matters, EA PCOMs) in time. This task is easy if fossil faunas are arranged in a consistent biochronologic scheme as was the case with Italian fossil faunas we dealt with. At larger spatial or time scales, and with less familiar faunas and species, time ordering is less trivial, especially because faunas are diachronic. Alroy (1998, 2000) pointed out that the North American Land Mammal Ages (NALMA) system, which is a continental biochronologic scheme, is fraught with limitations deriving from diachrony. Hence, biochronologic

* Corresponding author. Tel.: +39 081 253 8 331; fax: +39 081 552 09 71.

E-mail addresses: pasquale.raia@libero.it (P. Raia), carotenuto_francesco@yahoo.it (F. Carotenuto), carlo.meloro@hyms.ac.uk (C. Meloro), ppiras@uniroma3.it (P. Piras), carmela.barbera@unina.it (C. Barbera), kotsakis@uniroma3.it (T. Kotsakis).

information could not help if you are to draw PCOMs at large spatial/time scales. Consequently, we choose to include a time-ordering, numerical variable as an additional BCA entry factor by relying on Fortelius et al.'s (2006) spectral ordering procedure, after controlling for its consistency with another independent age estimator, that is Alroy's maximum likelihood appearance event ordination (ML AEO, Alroy, 2000) and absolute dating (radiocarbon, K–Ar, U–Th and paleomagnetic estimates, where available). After EA PCOMs have been obtained, we present them in terms of their taxonomic composition and age, compare them with classic biochronological schemes, and illustrate their spatial structure. As discussed at some length below, this latter point is completely novel and makes conspicuous the difference between bio-chrons, which occupy distinct periods along the time scale, and EA PCOMs, which are, instead, subjected to the additional twist of geographical separation among themselves, and could be, as a consequence, partially overlapping in time.

2. Materials and methods

2.1. The database

We collected either from web sources or literature (see Appendix A) faunal lists of 811 distinctive fossil horizons of Western Eurasia including at least four large mammals species (we knew empirically that horizons yielding less than four species could not be placed consistently in any cluster applying BCA, see Raia et al., 2005). The lists (local faunal assemblages, LFA) span in time from middle Pliocene to early Holocene. In analogy with PCOMs, the computation of EA PCOMs was limited to species having broadly similar preservation potential and ecologically linked via conceivable trophic interactions. Accordingly, we included taxa belonging to ungulates, carnivores of the families Ursidae, Canidae, Hyaenidae and Felidae, and proboscideans, with a minimum estimated body mass of 7 kg. Rodents, lagomorphs, smaller carnivores such as mustelids and viverrids, primates, bats, soricomorphs, erinaceomorphs, and marine species were excluded because either their small body size prevents fossilization of most remains (Damuth, 1982), taphonomic factors provide idiosyncratic preservation potential (think of rodent bone accumulation by birds of prey and wood rats) or ecological interactions with the selected species are marginal at best (with the exception of canids which could have been feeding on rodents as well as smaller felids [e.g. lynxes] on lagomorphs). By the term Western Eurasia we mean LFAs not greater than 60° E in longitude. East and most of Central Asia were excluded from our database because of inconsistent taxonomy (e.g. Chinese LFAs) or under-sampling (e.g. central Asian countries such as Iran, Afghanistan, Azerbaijan and so on). When independent faunal lists for the same LFA were available in the literature, we selected the youngest reference. Nonetheless, we present all sources in Appendix A. For each LFA we recorded occurring species, latitude, longitude, and, when available, chronological age estimate. LFAs coming from the same fossil sites (different horizons within a section) were treated separately, but nested LFAs were excluded to avoid redundant information. We tagged as “nested” each LFA whose list is just a subset of that of another horizon of the same stratigraphic section. For instance, suppose that in a fossil site three different fossiliferous horizons are recognised. The lowest yields species A, B, C, D, and E, the middle yields species A, B, C, and F and the topmost yielding species B, C, D and E. We designated the topmost level as nested within the lowest and therefore excluded it. Data were arranged in a presence/absence matrix and then synonymised by relying on the latest works on species taxonomy and our own personal opinions in a minority of cases we are familiar with (the list of synonyms is presented in Appendix B). Synonymy is a major problem in computing similarity among LFAs because any measure of resemblance between assemblages could be severely underestimated by inconsistent

taxonomy. Even worse, this inconsistency could have a geographical basis because taxonomic discrepancies could derive from national-based traditions in nomenclature. For instance, most Italian palaeontologists refer to the straight-tusked elephant as *Elephas antiquus* instead of *Palaeoloxodon antiquus*. By maintaining them as separate entities we would get the wrong consequence of two contemporary elephant species geographically separated, while there was clearly only one, with easily imaginable consequences on similarity calculation and its spatial structure. The same applies to Villafranchian deer *Axis* (= *Pseudodama*), megacerine deer (which were commonly ascribed to numerous genera such as *Megaloceros*, *Megaceroides*, *Praemegaceros*, *Dolichodoryceros*, *Allocaenelaphus*, *Eucladoceros*) caballine horses (which have been given lots of both specific and subspecific names such as *germanicus*, *soloutrensis*, *gallicus* and so on) and many others. These same sources were used to update faunal lists if appropriate. In its emended (in synonymy and uncertain taxon entries) form, the database includes 781 LFAs and 220 taxonomic entities (Appendix C).

As for LFA ages, we used data as provided by specific papers and online databases (see Appendix A). When available, we used absolute age estimates for most of younger localities. In keeping with Fortelius et al. (2006) when faced with data providing maximum and minimum age estimates, we computed the arithmetic mean, as for most of old localities (for instance those provided by NOW Database, <http://www.helsinki.fi/science/now/>). We did not apply more stringent criteria, such as relying on numerical (absolute) estimates only, because we were interested in obtaining the correct chronologic ordering, not in assigning to each locality a numerical age. Consequently, we sought to maximize the number of data points and to have evenly-distributed number of data over time.

2.2. Time ordering techniques, spectral ordering

Of the several time-ordering procedures available in the literature, we applied the most common, Alroy's maximum likelihood event ordination (Alroy, 2000) and the recently-developed spectral ordering of Fortelius et al. (2006). We introduce this latter technique first, and devote more space to explain it, for it is less familiar in the palaeontological literature.

Spectral ordering is the ordination of samples (here LFAs) according to their similarity. When applying spectral ordering one faces the classical Consecutive Ones Problem (C1P) (Booth and Lueker, 1976). In a (0, 1) matrix C , the C1P is seeking the permutation matrix IIC where for each column all the ones are consecutive. Fiedler eigenvector was proposed as the solution to C1P in Chung (1997) and in Atkins et al. (1999). In matrix algebra, Fiedler eigenvector, denoted as $\mathbf{v}_{n-1} = (\mathbf{v}_1, \dots, \mathbf{v}_n)$, is the vector with the second smallest eigenvalue of the Laplacian matrix. Fortelius et al. (2006) used spectral ordering to put LFAs in sequence by extracting Fiedler eigenvector. The eigenvector \mathbf{v}_{n-1} has the property of minimizing the value:

$$\sum_i s(i,j) (\mathbf{v}_i - \mathbf{v}_j)^2.$$

Thus it minimizes the distance between coordinates (position along the vector) of two localities.

For the seriation problem solving we first started with the calculation of a locality–locality similarity index by computing:

$$s(x_i - x_j) = \frac{c(x_i - x_j)}{|t(x_i)|^{1/2} |t(x_j)|}$$

Where i and j denote the localities, x_i is the faunal list of locality i and x_j is the list for locality j , $|t(x_i)|$ and $|t(x_j)|$ are the number of taxa in locality i and j , respectively, $c(x_i, x_j)$ is the number of taxa that i and j share. With these indices we computed the locality–locality similarity matrix S

which is an $n \times n$ symmetric matrix. Then in Matlab we calculated the diagonal matrix \mathbf{D} of \mathbf{S} and the Laplacian Matrix, as given by the formula:

$$\mathbf{L} = \mathbf{D} - \mathbf{S}$$

Hence the eigenvectors of the Laplacian matrix were calculated and ordered from the largest eigenvalue λ to the smallest. As $\lambda_n = 0$, which implies $\mathbf{v}_n = \mathbf{0}$, we chose the λ_{n-1} , which is the value of the Fiedler eigenvector. We assigned a spectral index to each locality according to its coordinate on the vector. A complete source code to calculate Fiedler eigenvector in Matlab environment from presence/absence data is presented in Fortelius et al. (2006).

2.3. Time ordering techniques, maximum likelihood appearance event ordination

Alroy's ML AEO (Alroy, 2000) is the final version of a stepwise ordination method to arrange LFAs in their most probable temporal sequence and calibrate them with the absolute time-scale (see also Alroy, 1994, 1996). This method received great attention by biochronologists for its intrinsic precision and great robustness (Alberdi et al., 1997; Bowen et al., 2002; Lindsay, 2003; Hernández Fernández et al., 2004; Tsubamoto et al., 2004; Domingo et al., 2007). It draws on taxon conjunctions in faunal lists. The ML AEO provides an ordination of first appearance events FAE and last appearance events LAE (F/L statements in Alroy's terminology). After F/L statements have been computed, it accounts for disjunctive geographical ranges by applying a square graph algorithm to infer F/L statements for those species that could not have occurred in the same faunal list. Then, a maximum likelihood optimization criterion is applied swapping events in a sequence to minimize "nuisance" parameters for each given event in the sequence. The position of each faunal list is ultimately defined by its youngest FAE and oldest LAE in the event sequence.

The next step is to hinge the event sequence to geochronological age estimates. With ML AEO, an interpolation method is applied for calibration, to maximize the number of hinge points. A full explanation of the method is available in Alroy (2000). ML AEO analysis was performed with the PAST software (Hammer et al., 2001).

Once obtained, Fiedler and ML AEO scores were tested for correlation with age estimates separately, in order to verify the hypothesis that these scores are good proxies of the LFAs age. Then, both vectors were regressed against geochronological age estimates in order to calculate an age (in yr) for each LFA we had no absolute age estimate for. These latter estimates were meant to collate EA PCOMs into the time scale, via the LFA they include. Our goal here was not to find the best method for calculating presumed ages for LFAs, and we do not discuss virtues of either time-ordering method. Instead, once the two age estimates were obtained for each LFA, we averaged them, and took this mean value as to represent that LFAs age. When radiometric and absolute dating were available, we used these latter data and discarded the age estimate we had calculated via regression.

For BCAs, we used Fiedler scores as a time-ordering variable as it is based on similarity between entire faunal lists, while ML AEO orders species appearances and disappearances. Therefore, Fiedler vector is consistent with the clustering procedure (in cluster analysis each column, here a LFA, represents a vector being compared to other vectors as a unit).

2.4. BCA

Bootstrapped cluster analysis was presented in detail elsewhere (see Raia et al., 2005; Raia et al., 2006a) but we briefly review it here. We present a general description first, and then discuss the statistical details.

The method is based on successive clustering sessions, each performed on the groups discriminated by the previous analyses. If a

clustering session recognizes two distinct groups, the succeeding includes two separate cluster analyses (one for each group) and so on. At first, it could sound puzzling that we resorted on BCA to seek after taxonomically clear-cut groups of species and then opted to perform nested experiments, thereby apparently rejecting the groups themselves. Indeed, the rationale to perform nested computations (that is BCAs on the subgroups resulting from a previous BCA) stands in the hierarchical similarity structure in the data. This hierarchy exists because the matrix as such includes both evolutionary and ecological processes acting at different time and spatial scales, as will be apparent below. As a matter of fact, the very proof that this hierarchy is real is that BCA found nested clusters of significantly different composition, in keeping with classic biochronology that recognizes separate units within larger ones (such as Faunal Units within Mammal Ages). Clustering sessions were stopped when no statistically significant group difference was found or groups were just subsets of species nested within a larger group. In this latter case, the cluster could well be a sampling of a geographically limited area (whose diversity is necessarily smaller than the regional pool) or the result of sampling of taphonomically similar conditions (for instance, hyena dens would repeatedly sample the hyena's preferred prey, which forms a subset of the whole community). Non-divisible groups at the end of clustering sessions were given the status of EA PCOM. At each session, outlying LFAs were removed from the data set and the BCA re-computed.

The first step was to reduce the—781 LFAs \times 220 species—presence/absence matrix (plus one column containing the Fiedler vector, see Appendix C) to a 781 \times 105 genera matrix. We initiated our analyses with genera because in our previous experience with BCA clustering (Raia et al., 2005) we discovered the first BCA run just split the matrix in large groups based on genera-level taxonomic turnover. Genera live longer than species, thereby clustering genera best captures grand-scale, evolutionary trends. At this stage, we had not included Fiedler scores yet, as they were calculated on similarity matrix among lists of species, not genera. Similarity among genus-level lists was calculated by the Jaccard index (appropriate given binary presence/absence data and insensitive to LFA richness, number of taxa). In successive BCAs, based on species list, we calculated resemblance between LFAs by using the Gower index of similarity. Gower index was selected because treats all variables equally, irrespective of their dimensionality. Hence its usage was set by the inclusion of Fiedler scores. All cluster analyses were performed with the UPGMA algorithm. After each cluster analysis, we tested the robustness of partitioning among branches of each dendrogram by bootstrap resampling. Presence–absence data were shuffled (with replacement) among sampling units and the clustering procedure repeated each time. This was run n times up to 1000. At each n th repetition, a distance statistic G^* between the bootstrap sample and the reference sample for a given partition level k was computed. G^* was compared with the distance statistics G^0 generated under the assumption that the k partition level is in fact sharp. The probability $P(G^0 < G^*)$ calculated over n pairwise comparisons should be higher than the significance level for k to be sharp (Pillar, 1999). We chose the first k level to set at P values higher than 0.05 (if three groups are sharp— $k = 3$ —then two will be sharp as well, but we opted for $k = 3$ to obtain the finest resolution). Nested analyses continued after this stage so far as distinctive (in taxonomic composition) groups could be recognized. BCA were performed with Multiv 2.1.2 software (Pillar, 2001).

2.5. Issues concerning the record

The fossil record is notoriously discontinuous. Therefore, dealing with such a problem is mandatory in most, if not all, paleoecological studies. In this study, we had to tackle the issues of—possibly—both temporal and geographical discontinuities. The first matter of concern (time discontinuity) would be relevant if BCAs are sensitive to sample

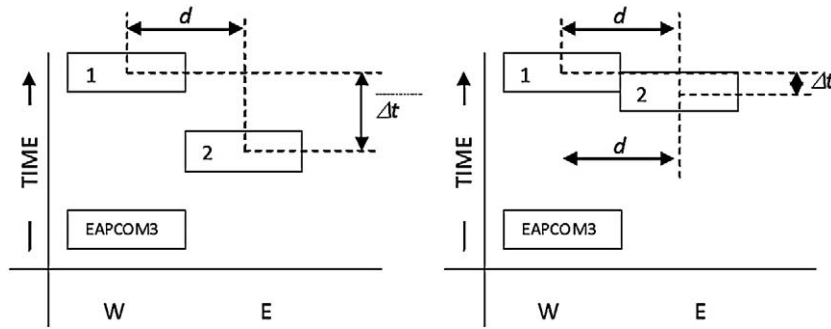


Fig. 1. Hypothetical occurrence of three consecutive EA PCOMs along the West-East axis. On the left, the non-overlapping time placement of these EA PCOMs avoid geographical interpretation to be put forward. On the contrary, the cartoon on the right shows overlapping (in time) EA PCOMs whose geographical separation is most probably genuine.

size. In this case, BCA might over split LFAs—dense time intervals leading to an excessive number of EA PCOMs. Our data set is quite rich in of late Pleistocene localities, both because they have higher preservation potential (being younger) and because part of our data comes from a very detailed collection of archaeological sites dating to the MIS 3 stage. Therefore, we expect a number of EA PCOMs in the late Pleistocene as a potential effect of over splitting.

A more serious threat is the geographical distribution of LFAs per time interval. Consider the cartoons in Fig. 1; in this simplified scheme, an EA PCOM (1) could occupy the westernmost stretch of Western Eurasia (marked with “W” along the x-axis). Then, the successive EA PCOM (2) could occur in the East (marked with “E”). A third EA PCOM (3) occurs in the West, again. For sure, some species of (1) survive into (2). And some of (2) (plus a few of (1)) survive into (3). In this case, we could have interpreted the succession of these EA PCOMs as a geographical pattern through time, with, say, species appearing in EA PCOM 2 going westward through time to colonize the “W” region to make EA PCOM 3 up along with species surviving in situ. Yet this pattern cannot be excluded as an artefact for these two paleocommunities (1 and 2) are not overlapping in time and could just represent the effect of paucity of LFAs in the East during the time spans when EA PCOMs 1 and 3 formed in the West. Now, consider the example on the right side of the figure. EA PCOMs taxonomic composition and geographical position are the same, and the interpretation we put forward above may be proposed in this case, too. Yet, EA PCOMs 1 and 2 are overlapping in time. Thereby, their distinctiveness might not depend on the vagaries of preservation, but on true taxonomic differentiation. In proposing true EA PCOMs, we paid attention to time overlap before putting any “geographical” interpretation forward. In addition, we tested randomness in LFA occurrence in time and space. First, we partitioned East and West LFAs putting a divide at 20° 42' East (the longitude of the centroid of our LFA distribution). We chose this West–East split because we observed a posteriori that geographical differentiation among and within EA PCOMs occurs along the West–East axis. Then, we divided the time scale (as represented by age estimates) in equal 500 ky intervals and counted the number of LFAs in each interval. Finally, we compared the distribution of LFAs per time interval for Eastern and Western localities with a χ -square test. If significant deviations do occur, it is critical to ascertain if these LFA-dense time bins correspond (both temporally and geographically) with a given EA PCOM. A significant association would cast doubts on the reliability of that EA PCOM's geographical distribution.

We did not perform tests for the whole sample because we were interested in the reliability of each single EA PCOM only.

3. Results

The correlation between the Fiedler scores and the geochronologic ages is high (Fig. 2, $R^2 = 0.968$; $p = 1.591 \cdot 10^{-294}$). The regression of Fiedler scores on age is highly significant as well ($F = 11,930$, $b = 11,636,613.222$, $n = 392$). We got qualitatively the same results

by using ML AEO dating. With Alroy's method, Pearson product moment is highly significant as well (Fig. 2, $r = 0.918$, $p = 1.735 \cdot 10^{-214}$). Regressing ages against ML AEO scores gives very reliable results, again ($F = 4360.04$, $b = 12,992.599$, $n = 394$). Despite these convincing results, we emphasize that LFA ordering is not as powerful for very young faunas. Correlations become non-significant for faunas younger than 50 ky for both Fiedler vector scores ($r = 0.048$, $p = 0.469$, $n = 230$) and ML AEO scores ($r = 0.014$, $p = 0.860$,

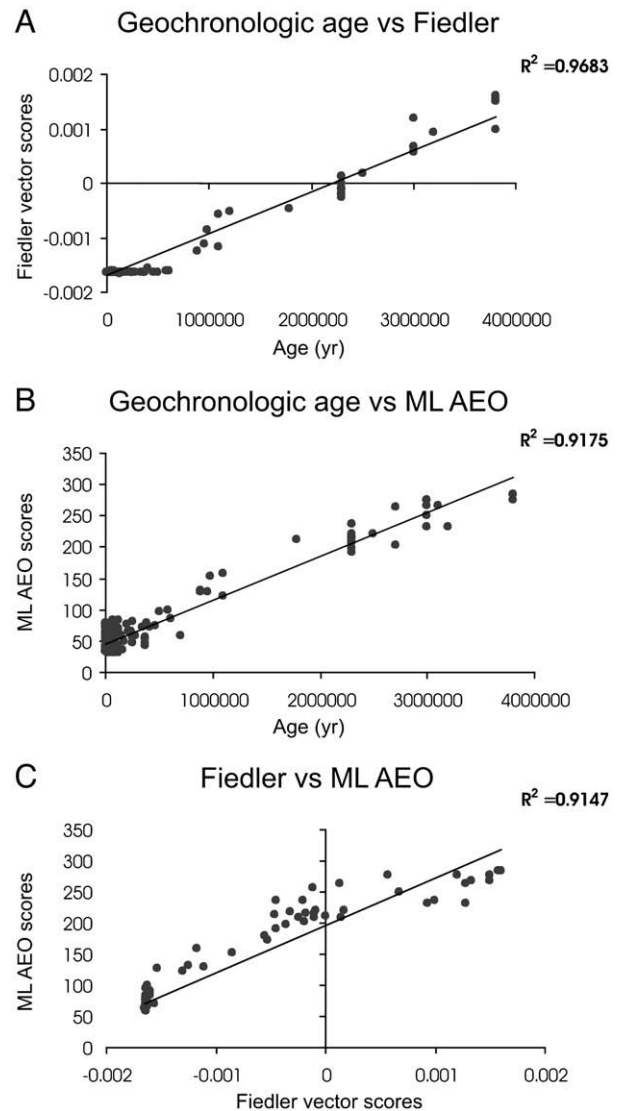


Fig. 2. Plots showing the correlation between time scores (ML AEO, A; Fiedler Vector, B) and geochronologic ages. In C, the two scores are plotted against each other.

$n=230$). Of course, this is hardly a problem for the vast majority of these faunas have been dated with radiometric methods.

3.1. The EA PCOMs and their age

The bootstrapped cluster analysis, as applied at genus level to detect EA PCOMs, has shown a starting partitioning level of two large clusters that we named G1 and G2. We considered valid the first partitioning level with a p value just above the alpha level (0.05) just to ensure fine partitioning resolution (as in Raia et al., 2005). Then, we computed the BCA on these clusters for further partitioning.

The G2 group was divided into two clusters named G2.1 and G2.2 (Fig. 3A). Then the BCA on G2.2 also detected two other clusters named G2.2.1 and G2.2.2 (Fig. 3B). A further partition on G2.2.1 finally detected the three groups G2.2.1.1, G2.2.1.2 and G2.2.1.3 (Fig. 3C).

The G1 group was divided into two clusters named G1.1 and G1.2 (Fig. 3D). Then the last was divided into G1.2.1 and G1.2.2 (Fig. 3E). No further partitioning was statistically advisable.

Eventually, we got 8 EA PCOMs, listed as follows (from the oldest to the youngest). EAPCOM 1 (G2.1; age span: 3.7–3 My) roughly corresponds to “Early Villafranchian”, supporting classic biochronologic attribution for most of its localities (e.g. Triversa, Capeni, Tulucesti, Les Etouaires). “Middle Villafranchian” appears split in two EA PCOMs. The older, EA PCOM 2 (G2.2.1.1; age span 2.5–1.9 My), includes localities such as Saint Vidal, Coupet, and Norwich Crag being 2 to 2.5 My old. The younger, EA PCOM 3 (G2.2.1.2; age span 2.2–1.5 My), includes both classic Middle and Late “Villafranchian” localities (such as Saint Vallier, Montopoli, Dmanisi, and Tegelen). EA PCOM 4 (G2.2.1.3; age span: 1.9–1.3 My) covers “early Late Villafranchian” to “middle Late Villafranchian” including localities such as Psekups, Fonelas P1, Poggio Rosso and

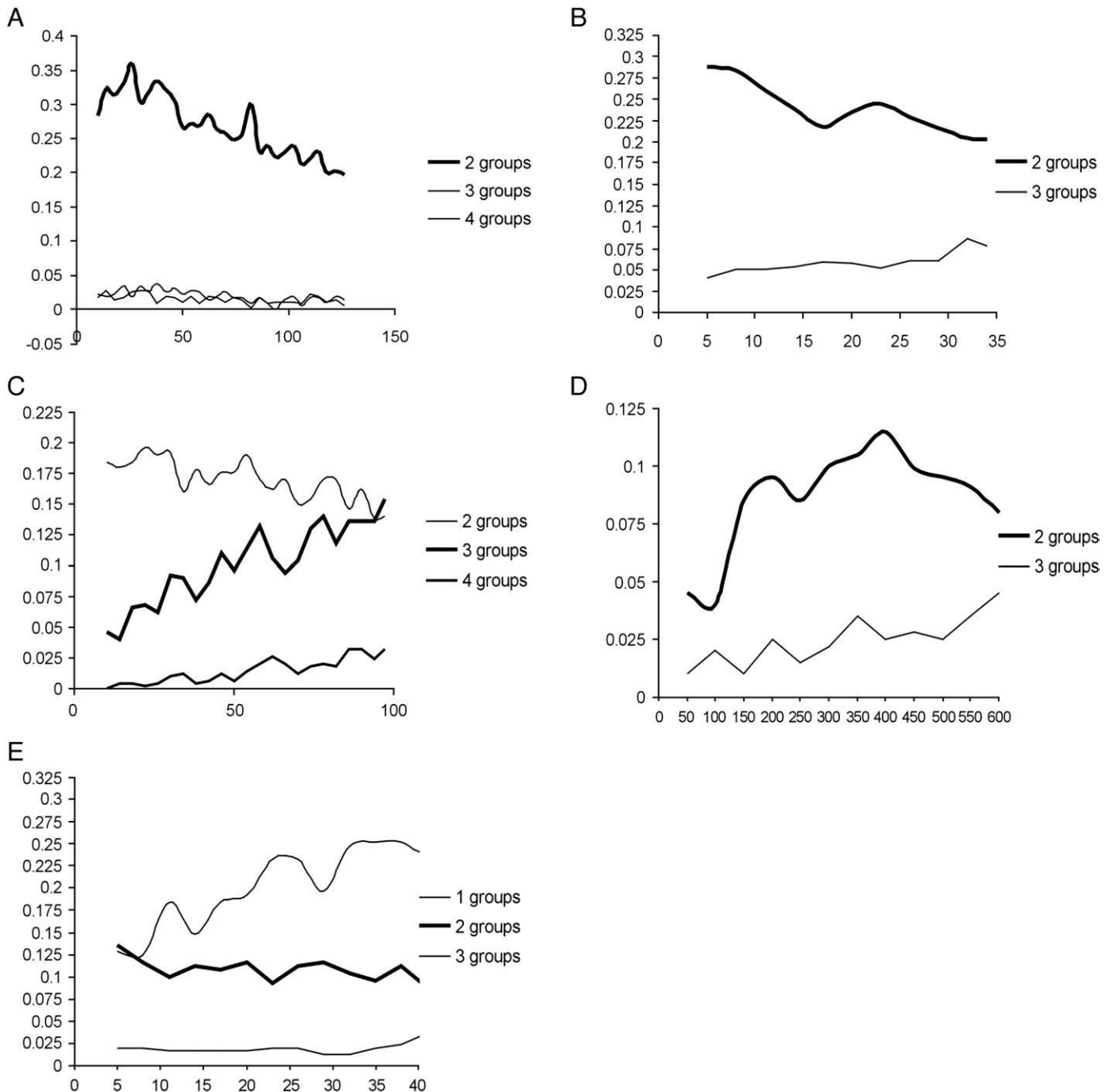


Fig. 3. Results of BCA as performed on group G2 (A), on group G2.2 (B), on group G2.2.1 (C), on group G1 (D), and performed on group G1.2 (E).

Pantalla. EA PCOM 5 (G2.2.2; age span: 1.8–1 My) covers “middle Late Villafranchian” to “late Late Villafranchian”. It includes localities such as Ceyssaguet, Colle Curti, Venta Micena, Pirro Nord, and Sainzelles.

EA PCOM 6 (G1.2.1; age span: 1–0.3 My) includes “late Late Villafranchian” to “Early Galerian” localities such as Voigstedt, Suessenborn, Tiraspol, and Pakefield Rootlet bed. EA PCOM 7 (age span: 0.6–0.068 My) spans over most of “Galerian” including, for example, Orgnac 3, Lunel Viel, Isernia, and Mauer. Lastly, EA PCOM 8 collapses latest Galerian to younger (Aurelian, Eemian, Weichselian up to Holocene) localities, the oldest being radiometrically dated at 458 ky. A synoptic view of EA PCOMs age distribution is presented in Fig. 4. Through BCA sessions, 88 LFAs were removed from the data set as they were outliers, the vast majority of them pertain to the latest EA PCOM (8).

3.2. Geographical distribution and composition of EA PCOMs

Age estimates can reveal spatial patterns in the occurrences of species assemblages. For instance, if a faunal assemblage maps on the East then successively on the West, it may be proposed that part of that fauna had migrated from the East to the West (a pattern we expect since most components of European large mammal biotas are of Asian origin) provided sampling factors are accounted for. This could occur both within an EA PCOM, or between successive EA PCOMs, for intense taxonomic turnover over space might actually split faunas in separate EA PCOMs.

EA PCOM 1 spans most of Europe. It includes 24 localities and 49 taxa (Fig. 5). Although the oldest LFAs occur in the Eastern part of the range, no spatial pattern seems to be apparent. The most abundant species are the gomphothere *Anancus arvernensis* and the mastodon *Mammot borsoni* the tapir *Tapirus arvernensis*, the rhino *Stephanorhinus jeanvireti* and the ancient deer *Rusa rhenana*. EA PCOM 2 includes only 12 localities and 27 taxa. No spatial pattern in estimated age is apparent (Fig. 5). The fauna is characterized by commonness of *A. arvernensis*, yet it differs from the previous assemblage by the presence and the rising commonness of the “southern” mammoth *Mammuthus meridionalis* and the rhino *S. etruscus*. It should be noted that a major temporal gap intervenes between EA PCOMs 1 and 2. Apparently, we had no locality whose age was comprised between 2.5 and 3 My. The onset of EA PCOM 2 coincides with the so-called “Elephant-Equus event” (Azzaroli et al., 1988; Koenigswald and Werdelin, 1992). And its demise just precedes the beginning of the Pleistocene.

EA PCOM 3 occurs mostly over Western and Southern parts of Europe (Fig. 5). Youngest localities occur in the East (Dmanisi A) and the North (Tegelein, and the even younger Blassac La Gironde). The fauna is dominated by the deer *Eucladoceros ctenoides*, the equid *Equus stenonis*, the hyaenid *Pliocrocuta perrieri*, plus *S. etruscus* and *M. meridionalis*. Twenty-three localities and 49 taxa are included in

this EA PCOM. The succeeding EA PCOM 4 partially overlaps, in age, with the previous assemblage. It includes 27 LFAs, mostly distributed over Southern Europe, and 45 taxa (Fig. 5). The most abundant species is *E. stenonis*, followed by *S. etruscus*, *M. meridionalis* and the giant hyaenid *Pachycrocuta brevirostris*. Other abundant species, differ with the previous assemblage, are the canid *C. etruscus*, the fallow deer *Axis nestii* and the bovid *Leptobos etruscus*. EA PCOM 5 occurs over South and Western Europe. Oldest localities seem to occur in the central-Eastern part of its geographical range (Fig. 5). The most common species is *M. meridionalis*, while distinctive elements are the abundant occurrences of the hippo *Hippopotamus antiquus*, the deer *Axis farnetensis*, the rhino *S. hundsheimensis*, and the equid *E. altidens*. The latter two species represent clade-level evolution from older representatives. EA PCOM 5 includes 17 LFAs and 44 taxa. This EA PCOM plainly marks the latest Villafranchian. Although localities occur mostly in Western Europe, lack of faunas of comparable age in the East seems to be a problem of sampling (see below). That is, the current, western, spread of this EA PCOM as is, it's most probably a sampling artefact. The two next EA PCOMs greatly overlap in age (see Figs. 4, and 5), but are clearly separated geographically. Thereby, we argue these two EA PCOMs represent a case of progressive taxonomic turnover between EA PCOMs starting from the North-East. EA PCOM 6 includes 23 LFAs mostly distributed over North-Western Europe, and a total of 43 taxa (Fig. 5). The most common species are the mammoth *M. trogontherii*, the rhino *S. etruscus*, and the deer *Praemegaceros verticornis*, *Cervalces latifrons* and *Cervus elaphus*. This paleocommunity is, as for its composition, clearly “Galerian”, in stark contrast with the preceding fauna. This conclusion is further borne out by the appearance of advanced form of *Bison* wolf *Canis lupus*, cave hyena *Crocuta crocuta*, roe deer *Capreolus capreolus*, straight-tusked elephant *Elephas antiquus*, horse *E. ferus*, wild boar *S. scrofa* and ancient cave bear *Ursus deningeri*. The slightly younger EA PCOM 7 includes 48 taxa and 32 LFAs located in South-Western Europe (Fig. 5). Red deer, horse, wild boar and straight-tusked elephant are the most common species. Novel elements are the fallow deer *D. dama*, and the hydruntine horse *E. hydruntinus*, thars (genus *Hemitragus*) woolly mammoth *M. primigenius*, cave lion *Panthera leo*, Irish elk *Megaloceros giganteus*, chamois *Rupicapra rupicapra*, cave bear *Ursus spelaeus*, and rhinos *Stephanorhinus hemitoechus* and *S. kirchbergensis*. It is easily argued that this EA PCOM represents the complete setting of Galerian fauna coming from North-Eastern EA PCOM 6, a notion further supported by the northern geographical position of its oldest localities.

The latest EA PCOM 8 is, not surprisingly, full of LFAs (537). Albeit it includes twenty times as many localities as older EA PCOMs, species richness is strikingly similar (49 species), a plain indication that all EA PCOMs are fully comparable to each other. It includes the well-known late Pleistocene megafauna. Dominant species (as for their commonness in the record) are red deer, wolf, horse, reindeer, cave hyena, wild boar, roe deer, woolly mammoth, cave bear, and cave lion, in this order. Taxonomic turnover from the previous turnover is marked by disappearance of old elements such as the sabre toothed cat *Homotherium latidens*. And by appearance of ibex *Capra ibex*, woolly rhino *Coelodonta antiquitatis* (both of them very common) plus rarer elements such as moose *Alces alces*, the saiga antelope *Saiga tatarica*, European lynx *Lynx lynx*, and khulan *Equus hemionus*. Owing to abundant data, EA PCOM 8 shows a very strong geographical structure, with a clear South-West to North-East age gradient (Fig. 5). This gradient most probably testifies recolonization of post glacial habitats for the youngest localities are Holocene LFAs located in Scandinavia, North-Eastern Europe and Western Siberia.

The overall distribution of LFAs is greatly skewed to the West, where LFAs are some 3 times as numerous. Deviations from expected number of LFAs per time period per geographic category (“East” or “West”) are highly significant (see Table 1). Most important deviations (marked with an asterisk) occur in the West, with evidence for

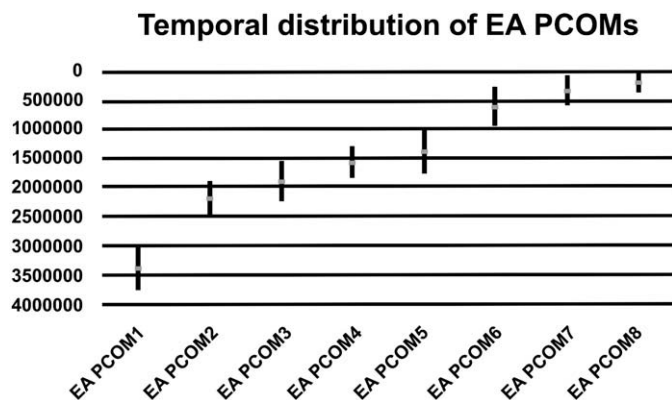


Fig. 4. Temporal distribution of EA PCOMs. Overlaps are apparent in younger EA PCOMs.

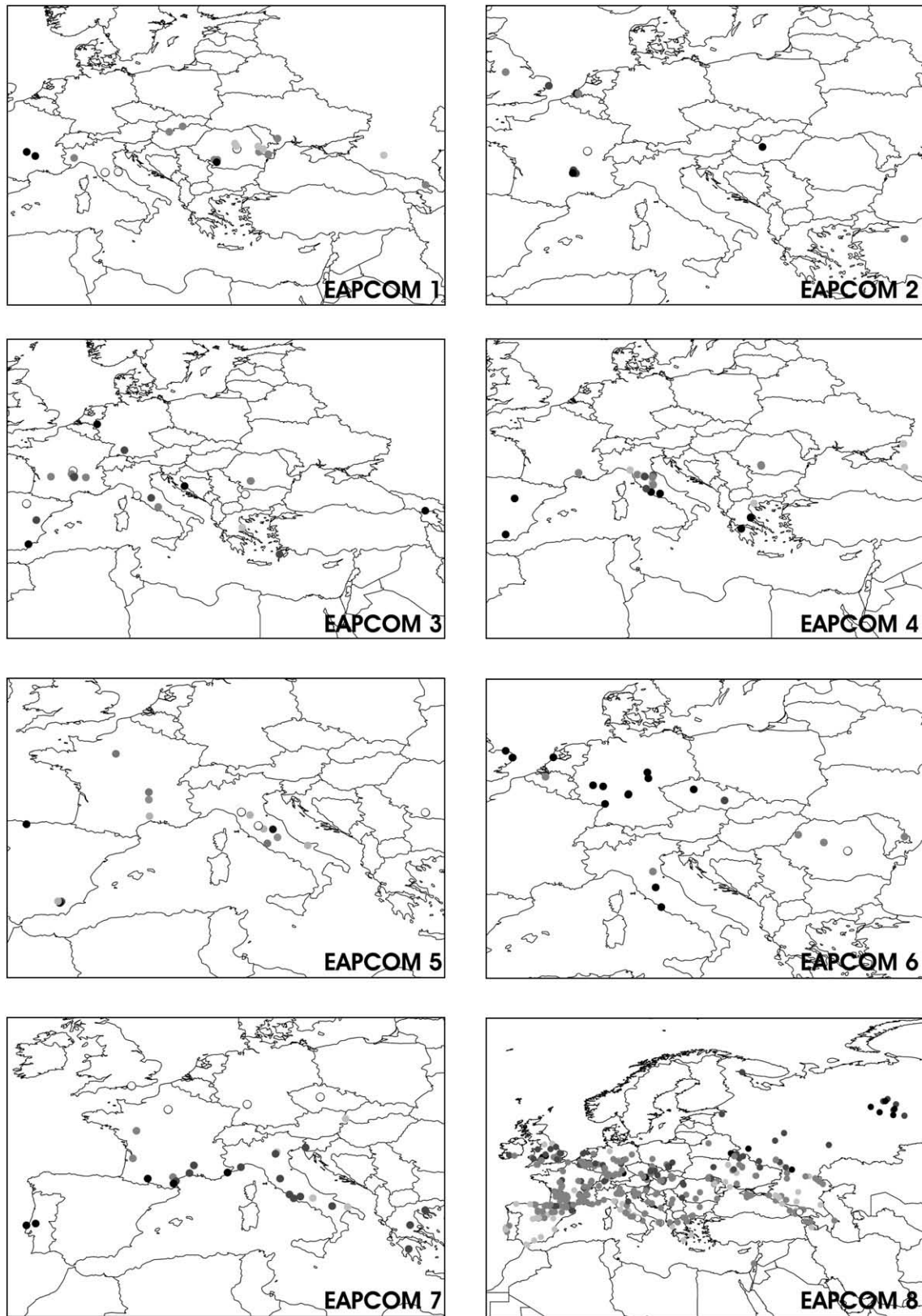


Fig. 5. Spatial distributions of LFAs included in EA PCOM 1 to 8. Black dots are younger LFAs as based on Fiedler vector scores, lighter colors represent older LFAs.

undersampling in the period from 4 to 3.5 My ago, and evidence of oversampling in the period between 1 and 1.5 My. For the oldest time intervals, only one EA PCOM (1) has been recognized. Despite the scarcity of LFAs in the West, its distribution is pan-European. The

period between 3 and 2.5 My is surprisingly devoid of LFAs. As a consequence, it is not possible to ascertain the real duration of EA PCOM 1. From the beginning of the Pleistocene, Western localities became more and more common. Their number peaked exceptionally

Table 1

Number of LFAs per time period, partitioned into “West” and “East”. The expected number of LFAs is much higher than the real data in the West for time periods from 4 to 3 My; and much lower there for the period from 1.5 to 1 My.

EA PCOM	Time interval	E	W	Expected E	Expected W	Total
6, 7, and 8	0–500 ky	128	441	128.2	408.8	537
6 and 7	500 ky–1 My	2	12	1.9	6.1	8
4 and 5	1–1.5 My	3	24	3.1	9.9*	13
2, 3, 4 and 5	1.5–2 My	12	24	11.9	38.1	50
2 and 3	2–2.5 My	2	8	1.9	6.1	8
1	3–3.5 My	11	4	11.0	35.0*	46
1	3.5–4 My	3	1	3.1	9.9*	13
Totals		161	514	161.0979	513.9021	675

Asterisks indicate significant differences.

high (in proportion to the expected value) during the time slice from 1.5 to 1 My. Two EA PCOMs (4 and 5) span this interval, and EA PCOM 5 appears to be mainly western, but this notion should be discarded in the light of sampling bias.

4. Discussion

The evolution of faunal assemblages over time is a major topic in palaeobiology. It drew considerable interest from mammalian biochronologists who sought after methods to divide assemblages in consistent, successive faunal units. Well-developed biochronological schemes, such as European MN zones (Mein, 1975) and North American NALMAs (Wood et al., 1941; Woodburne and Swisher, 1995) served this interest successfully for decades (Bernor et al., 1992). Yet, mammalian biochrons are less than perfect as a concept (Lindsay, 2003). Experts should supply reference species to mark biochron boundaries, and this could be misleading (see discussion in Walsh, 1998). Even worse, diachrony puts severe limitation on these boundaries, as they usually vary (in time) over space (e.g. Alroy, 1998). The emergence of robust statistical methods for time ordering of faunas renewed debates over mammalian biochronology, provided a solution for diachrony and proposed a Gordian-knot-like answer to the problematic usage of subjective criteria to separate biochrons (there are no biochrons with such techniques). One obvious limitation of time-ordering techniques is that they do not admit, by default, any spatial structure in the distribution of faunas (for instance, ML AEO explicitly overcomes the effect of species disjunctive distributions by the “square graph” method). This is far from problematic if time ordering is the only aim to pursue. In addition, one must assume community evolution to be continuous and that successive faunas occupy equal time bins. Yet, constant community evolution is unlikely under models advocating either the effect of abiotic factors on the evolution of communities, such as Vrba's turnover pulse hypothesis (Vrba, 1995) or compositional resilience (Brett and Baird, 1995). EA PCOMs do not require any such assumptions, and address directly the problem of drawing boundaries between communities relying on the taxonomic composition of LFAs and (with the improvement of the method proposed here) numerical age estimates. In their nature, EA PCOMs were conceived to provide an ecologically sound and reliable depiction of past communities, with both temporal and spatial limits. To our knowledge, there is no published study showing this spatial structuring in extinct communities. Not even our own past experience with Italian PCOMs showed any spatial resolution, for the obvious reason that large mammals had geographical ranges much larger than the Italian peninsula. Thus, in essence, EA PCOMs are not true biochronologic units, but evolutionary-ecologic units of consistent taxonomic composition, with defined geographic and time ranges and, thereby, possibly overlapping time spans. We found strong overlap between some EA PCOMs, especially at the onset of Galerian Mammal Age, when two distinct communities inhabited Eurasia for some time. EA PCOM 6 was characterized by typical Galerian species, mainly new Asian immigrants.

This North-East to South-West pattern in large mammal taxonomic turnover is a recurrent pattern for most but the oldest EA PCOMs (see Fortelius et al. 1996, for a similar contention), and reverses only at the end of Pleistocene, when North-Western, colder habitats lost their ice caps. It is unclear, to our knowledge, if the more even spatial distribution of older EA PCOMs and their lack of clear, within-EA PCOM, spatial patterns in LFA's age, reflects a poor fossil record or more even habitat conditions in warmer, Villafranchian climates (Imbrie et al., 1993b).

The quality of the record is, plainly, a major issue. In our analysis, the vast majority of localities falls in a single EA PCOM, not surprisingly, the latest. On the one hand, this fact testifies to the robustness of BCAs, which are clearly insensitive to redundant data and uneven sample sizes. On the other, a larger sample size increases statistical sensitivity (Hair et al., 2001). Therefore, possibly the spatial patterns we observed *within* EA PCOM 8 are the same we observe *between* older EA PCOMs. Yet, it is clear that these spatial patterns exist and they seem to correlate with dispersal events (immigration from Asia) for EA PCOMs 2, 4, 6, and 7; and major climatic events, for EA PCOMs 4–8.

5. Conclusion

The paleocommunities we present here are clearly different from formal biochrons so far proposed. We argue that these two units, plus time-ordered sequence of LFAs, serve different aims. Of course, all of them have pros and cons. Biochrons are clearly useful if the objective is looking at large-scale patterns in community turnover. And are the only units of reference permitting correlation to stratigraphic units, whichever difficult this task happens to be (Lindsay, 2003). Further, biochrons provide referential lists of taxa to which is often easy to address new data without performing analyses anew. A time-ordered sequence is a very robust tool to look at instantaneous turnover rates, clade-level patterns of taxonomic evolution and fine-grained biochronologic resolution (Alroy, 2000). An often-neglected advantage is that a sequence is a continuous variable, which is much more powerful than categories (such as biochrons and PCOMs) to regress patterns against time. Finally, EA PCOMs offer the best depiction of past communities (in the ecological sense) and is the only way to look at spatial patterns in their distribution, provided the geographical scale of observation is large enough.

Acknowledgments

Although they are not directly involved in this study, we are grateful to generations of biochronologists, biostratigraphers, and taxonomists who provided the data used here. On the same line of reasoning, we are grateful to the big effort our colleagues managing online databases are sustaining to make data available via the web. Diana Pushkina kindly let us including additional data on Eastern localities she and one of us (PR) used for another study. Most of these data were originally published in Russian and we could have not even thought to look at these papers. Maria Teresa Alberdi, Fred Kop and an anonymous reviewer gave us important advices that let us improving the quality and readability of the manuscript. Daniela Fusco kindly helped us managing GIS procedures.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.palaeo.2009.02.005.

References

- Alberdi, M.T., Azanza, B., Cerdeño, E., Prado, J.L., 1997. Similarity relationship between mammal faunas and biochronology from latest Miocene to Pleistocene in the Western Mediterranean area. *Eclogae Geol. Helv.* 90, 115–132.
- Alroy, J., 1996. Conjunction among taxonomic distributions and the Miocene mammalian biochronology of the Great Plains. *Paleobiology* 18, 326–343.

- Alroy, J., 1994. Appearance event ordination: a new biochronologic method. *Paleobiology* 20, 191–207.
- Alroy, J., 1998. Diachrony of mammalian appearance events: implications for biochronology. *Geology* 26, 23–26.
- Alroy, J., 2000. New methods for quantifying macroevolutionary patterns and processes. *Paleobiology* 26, 707–733.
- Atkins, J.E., Boman, E.G., Hendrickson, B., 1999. A spectral algorithm for seriation and the consecutive ones problem. *SIAM J. Comput.* 28, 297–310.
- Azzaroli, A., De Giuli, C., Ficarelli, G., Torre, D., 1988. Late Pliocene to Early Mid-Pleistocene mammals in Eurasia: faunal succession and dispersal events. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 66, 77–100.
- Bernor, R.L., Fahlbusch, V., Mittmann, H.V. (Eds.), 1992. *The Evolution of Western Eurasian Neogene Mammal Faunas*. Columbia University Press, New York.
- Bowen, G.J., Clyde, W.C., Koch, P.L., Ting, S., Alroy, J., Tsubamoto, T., Wang, Y., Wang, Y., 2002. Mammalian dispersal at the Paleocene/Eocene boundary. *Science* 15 (295), 2062–2065.
- Booth, K.S., Lueker, G.S., 1976. Testing for the consecutive ones property, interval graphs and graph planarity using PQ-tree algorithms. *J. Comput. Syst. Sci.* 13, 335–379.
- Brett, C.E., Baird, G.C., 1995. Coordinated stasis and evolutionary ecology of Silurian to middle Devonian faunas in the Appalachian basin. In: Erwin, D.H., Anstey, R.L. (Eds.), *New Approaches to Speciation in the Fossil Record*. Columbia University Press, New York, N.Y., pp. 285–315.
- Chung, F.R.K., 1997. *Spectral Graph Theory*. CBMS Regional Conference Series in Mathematics 92. American Mathematical Society, Providence, R.I.
- Damuth, J., 1982. Analysis of the preservation of community structure in assemblages of fossil mammals. *Paleobiology* 8, 434–446.
- Domingo, M.S., Alberdi, M.T., Avanza, B., 2007. A new quantitative biochronological ordination for the Upper Neogene mammalian localities of Spain. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 255, 361–376.
- Fortelius, M., Werdelin, L., Andrews, P., Bernor, R.L., Gentry, A., Humphrey, L., Mittmann, W., Viranta, S., 1996. Provinciality, diversity, turnover and paleoecology in land mammal faunas of the later Miocene of western Eurasia. In: Bernor, R.L., Fahlbusch, V., Mittmann, W. (Eds.), *The Evolution of Western Eurasian Neogene Mammal Faunas*. Columbia University Press, pp. 414–448.
- Fortelius, M., Gionis, A., Jernvall, J., Mannila, H., 2006. Spectral ordering and biochronology of European fossil mammals. *Paleobiology* 32, 206–214.
- Gaston, K.J., Blackburn, T.M., 2000. *Patterns and processes in macroecology*. Blackwell Sci.
- Hair, J.F., Black, B., Babin, B., Anderson, R.E., Tatham, R.L., 2001. *Multivariate Data Analysis*. Prentice Hall.
- Hammer, Ø., Harper, D.A.T., Ryan, P.D., 2001. PAST: Paleontological statistical package for education and data analysis. *Paleontologia Electronica* 4, 9 Version 1.42.
- Hernández Fernández, M., Azanza, B., Álvarez Sierra, M.A., 2004. Iberian Plio-Pleistocene biochronology: micromammalian evidence for MNS and ELMAS calibration in southwestern Europe. *J. Quat. Sci.* 19, 605–616.
- Imbrie, J., Berger, A., Shackleton, N.J., 1993. Role of orbital forcing: a two-million year perspective. In: Eddy, J.A., Oeschger, H. (Eds.), *Global Changes in the Perspective of the Past*. John Wiley & Sons, New York, pp. 263–267.
- Koenigswald von, W., Werdelin, L., 1992. Mammalian migration and dispersal events in the European Quaternary. *Courier Forschung Senckenberg* 153.
- Lindsay, E., 2003. Chronostratigraphy, biochronology, datum events, land mammal ages, stage of evolution, and appearance event ordination. *Bull. Am. Mus. Nat. Hist.* 279, 212–230.
- Mein, P., 1975. Résultats du groupe de travail des vertébrés: bizonation du Néogène méditerranéen à partir des Mammifères. Report Activity RCMNS working groups, Bratislava, 78–81.
- Meloro, C., Raia, P., Barbera, C., 2007. Effect of predation on prey abundance and survival in Plio-Pleistocene mammalian communities. *Evol. Ecol. Res.* 9, 505–525.
- Meloro, C., Raia, P., Carotenuto, F., Barbera, C., 2008. Diversity and turnover of Plio-Pleistocene large mammal fauna from the Italian Peninsula. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 268, 58–64.
- Pillar, V.D., 1999. How sharp are classifications? *Ecology* 80, 2508–2516.
- Pillar, V.D., 2001. MULTIV 2.1.1. Program, Porto Alegre.
- Raia, P., Piras, P., Kotsakis, T., 2005. Turnover pulse or Red Queen? Evidence from the large mammal communities during the Plio-Pleistocene of Italy. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 221, 293–312.
- Raia, P., Piras, P., Kotsakis, T., 2006a. Detection of Plio-Quaternary large mammal communities of Italy: integration to biochronology. *Quat. Sci. Rev.* 25, 846–854.
- Raia, P., Meloro, C., Loy, A., Barbera, C., 2006b. Species occupancy and its course in the past: macroecological patterns in extinct communities. *Evol. Ecol. Res.* 8, 181–194.
- Raia, P., Meloro, C., Barbera, C., 2007. Inconstancy in predator/prey ratios in Quaternary large mammal communities of Italy, with an appraisal of mechanisms. *Quat. Res.* 67, 255–263.
- Tsubamoto, T., Takai, M., Egi, N., 2004. Quantitative analyses of biogeography and faunal evolution of middle to late Eocene mammals in East Asia. *J. Vertebr. Paleontol.* 24, 657–667.
- Vrba, E.S., 1995. On the connections between paleoclimate and evolution. In: Vrba, E.S., Denton, G.H., Partridge, T.C., Burckle, L.H. (Eds.), *Paleoclimate and Evolution with Emphasis on Human Origins*. Yale Univ. Press, New Haven and London, pp. 24–48.
- Walsh, S.L., 1998. Fossil datum terms, paleobiological event terms, paleontostratigraphy, chronostratigraphy, and the definition of land-mammal “age” boundaries. *J. Vertebr. Paleontol.* 18, 150–179.
- Wood, H.E., Chaney, R.W., Clark, J., Colbert, E.H., Jepson, G.L., Reedsides, J.B., Stock, C., 1941. Nomenclature and correlation of the North American continental Tertiary. *Geol. Soc. Amer. Bull.* 52, 1–48.
- Woodburne, M.O., Swisher III, C.C., 1995. Land mammal high resolution geochronology, intercontinental overland dispersals, sea-level, climate, and vicariance. In: Berggren, W.A., Kent, D.V., Hardenbol, J. (Eds.), *Geochronology, Time-scales and Global Stratigraphic Correlation: Framework for an Historical Geology*. Soc. Stratigr. Geol., Special Publication, vol. 54, pp. 335–364.