

# Chapter 2

## The Use of Similarity Indices in the Analysis of Temporal Distribution of Mammals

M. Belmaker

### 2.1 Introduction

Ecological and paleontological studies look at changes in species composition between spatially and temporally distinct regions. In such cases, data are organized in a numerical  $n \times p$  matrix or data frame, where  $n$  corresponds to different sampling times or sites (in this study assemblages) and  $p$  denotes each of the different variables that describe the locality studied. These may include the biological community (measured in species incidence or relative abundance), or other variables that connote the physical or chemical environment [Le98]. Thus, ecological datasets are multidimensional and represent a geometric hyperspace.

We identify change or stasis in such a system by means of  $\beta$  diversity, which is the variation in species composition between assemblages. It can be used to test for turnover (antonym inertia) and is measured as change in community structure from one assemblage to another along a gradient. If inertia is present in the system, then variability in species incidence or abundance between assemblages may result from sampling bias or other stochastic processes. However, if there is a positive correlation between community structure and a directional change along a vector, this may be related to monotonous changes in the abiotic or biotic environment.

In paleontological assemblages, on the scale of  $10^6 - 10^7$  years, studies have described various patterns of recurring fossil assemblages [Mi93]. The best documented and known pattern is “coordinated stasis,” which describes an empirical pattern of community level stasis coupled with an abrupt change in community structure of fossil assemblages. This inertia is present in spite of independent evidence for climate change. In contrast, in younger assemblages that date to the Quaternary (2.6 mya to present), community structure shifts predictably in response

---

M. Belmaker (✉)  
The University of Tulsa, Tulsa, OK, USA  
e-mail: [miriam-belmaker@utulsa.edu](mailto:miriam-belmaker@utulsa.edu)

to environmental change. The discrepancy in the faunal response between the pre-Quaternary (Phanerozoic) coordinated stasis and the Quaternary pattern [Ho96] was termed the “Pleistocene Paradox” [St01].

Little research has been done on the intermediate time scale ( $10^4 - 10^6$ ). Here, we present a case study to illustrate the use of (dis)similarity indices and the Mantel’s test in the temporal scale, to investigate changes in  $\beta$  diversity among seven mammal communities at the  $4 \times 10^5$  year scale.

## 2.2 The Case Study

The model system used in this study is the paleoanthropological site of ‘Ubeidiya, central Jordan Valley, Israel. The site has been dated to approximately 1.6–1.2 million years ago (ma), and the site exhibits early human remains [Be02], as well as rich lithic and faunal assemblages [Ba93] (Figure 2.1).

A unique method of excavation was used at the site due to the extensive post-depositional tectonic faulting of the sediments. Four trenches, numbered I–IV, were excavated. Within each trench, the archaeological strata were numbered in Arabic numerals from oldest to youngest. For example, stratum III 12 is the 12th geological layer of trench III. The site exhibits over 113 uncovered archaeological strata [Ba93]. However, fossil remains of large enough samples have been found in seven strata only (Figure 2.2).

We studied the large ( $> 10$  kg live weight) mammal communities of the seven strata of ‘Ubeidiya from all seasons of excavations (1959–2001). Specimens identified to other taxa (small mammals, birds, reptiles, turtle, fish, and invertebrates) were not included in this study.

Bones were identified to the lowest taxonomic level possible. We report species and genus level analysis, eliminating bones identified only to family [Be06].

We calculated two dependent variable data matrices. The first was an incident-based matrix with 1 for presence and 0 for absence. The second was a relative (percent) abundance. We transformed abundances at each of the sites by adding one and taking logarithms.

In ‘Ubeidiya, local hydrological conditions suggest a monotonic change (although not linear) in local environment from humid to dryer conditions from strata 7 through 1 [Ma06]. Therefore, an environmental grade was given to each stratum from 1 (driest) through 5 (wettest), as detailed in Table 2.1. For clarity, in lieu of using the original strata names, we labeled the fossil communities from 1 (youngest) through 7 (oldest) (see Table 2.1).

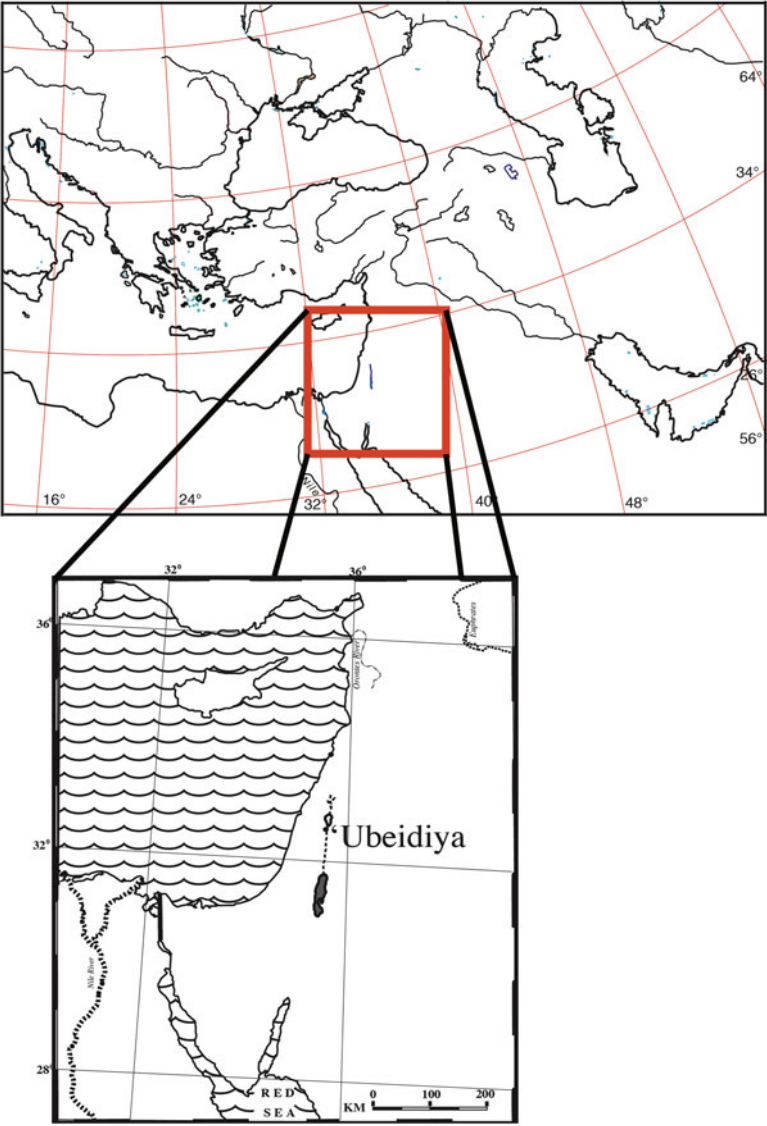


Fig. 2.1 Location of 'Ubeidiya in the Southern Levant

Geological Trenches				
K	III	I	II	Type of Environment
	92	erosion		Fossil, soils, screes and fluvial deposits
	86			
	85	42	51	Marshy to open, turbid like, with some fluvial penetration
	56	~ 33	~ 43	
	48-55	28-32	41-42	Screes in the west and fossil soils
	47	26-27	37-40	Shoreline deposit
		25	36	West: fossil soils and fluvial deposit
		~ 21	33	East: muddy to non-marshy littoral
	26	20	32	Shoreline deposits
29-30	23-25			Wadi beds, gravel laid by floods
		17-19	28-31	Muddy littoral to fossil soils
		top 15-16	26-27	Fine shoreline conglomerate
		main 15		Swampy, muddy littoral
	22	13-14		
	21		22-25	Shoreline deposits
		6-12	21	Swampy, muddy littoral
	20	20		
	19	19	19-20	Quiet, shallow water with water plants
		18	17-18	
		14-17	11-16	Deep water to littoral
		13	9 c,d-10	
		12		Muddy, shallow littoral
		10-11		
		9	9 a,b	
		8	8	Deep water lake
		4-8	2-7	Swampy and littoral to deep water

**Fig. 2.2** Stratigraphic sequence of 'Ubeidiya

**Table 2.1** Environmental gradient in relation to stratigraphic sequence

Strata	Ranked Stratum	Environment
III 11–13	7	5
III 20	6	4
III 21–23	5	4
II 23–25	4	3
II 26–27	3	3
II 36	2	1
II 37	1	1

## 2.3 The Statistical Model

Two criteria are important when we choose the (dis)similarity coefficient [Le98]:

1. The index used should be appropriate for the data. A presence–absence matrix can be transformed to a similarity with a binary coefficient, while for abundance data we need to use quantitative coefficients.
2. Double zeros: If a species is absent from two sites (double zero), we do not know if this is because it is truly absent or because of sampling bias. Thus,

it is preferable to exclude double zeros using an asymmetrical coefficient. The converse is a symmetrical coefficient, in which zeros are treated like any other value. In this study all coefficients used are asymmetrical.

It is beyond the scope of this study to discuss the numerous similarity indices that have been developed. The reader is referred to the work of Legendre [Le98] for an overview of these methods.

Here, we had three dissimilarity matrices: Two dependent variable matrices were calculated from multidimensional community structure, one using presence–absence data and the other using log-transformed relative abundance data. In addition, we had an explanatory (independent variable) vector: the *environment*.

The presence–absence community matrix was transformed into an asymmetrical binary similarity matrix by means of the Sneath and Sokal index [So63]:

$$S_{X_1, X_2} = \frac{a}{a + 2b + 2c}, \quad (2.1)$$

where  $a$  is the number of species common to both assemblages,  $b$  is the number of species present in assemblage one but absent from assemblage two, and  $c$  is the number of species present in assemblage two but absent from assemblage one.

The relative (log-transformed) abundance *community structure* was converted to a similarity matrix by means of the asymmetrical quantitative Gower coefficient [Go71]. This coefficient, used on normalized abundances, is defined by

$$S_{X_1, X_2} = \frac{\sum_{j=1}^p W_{12j} S_{12j}}{\sum_{j=1}^p W_{12j}}, \quad (2.2)$$

where  $S_{12j} = 1 - [|y_{1j} - y_{2j}|/R_j]$ .

The value  $W_j$  is called *Kronecker's data*, and its values are  $W_j = 0$  when  $y_j$  is missing for either one of the objects or both, and  $W_j = 1$  when information is present for both objects.

Both community structure similarity matrices were converted to distance matrices by means of the formula  $d = \sqrt{1 - s}$ , so each similarity index  $s$  was converted to a distance value  $d$  [Le98].

The *environment* vector (see Table 2.1) was converted to a dissimilarity matrix using the Euclidean distance [Le98]

$$D_{X_1, X_2} = \sqrt{\sum_{j=1}^p (Y_{1j} - Y_{2j})^2}. \quad (2.3)$$

Testing if there is a correlation between the similarity of sites structured in space or time, we come across a problem of autocorrelation. Specifically, if there are  $n$  objects that are temporally distinct and the matrix is symmetrical (so the time period spanning from object  $a$  to object  $b$  is the same as the time/distance from  $b$  to  $a$ ), such a matrix contains  $n(n-1)/2$  distances that are not independent, as changing the time of one object would change  $n-1$  of these distances. Thus, we cannot assess the relationship between two matrices using the parametric correlation coefficient.

Mantel's test is a solution to this type of problem [Ma71] as it takes autocorrelation into account. It is a regression in which the variables are (dis)similarity matrices instead of raw data, and so it allows us to test hypotheses regarding the correlation between distances among objects in matrices  $X$  and  $Y$ .

The basic form of the Mantel's statistic is calculated [Le98] as

$$z_M = \sum_{i=1}^{n-1} \sum_{j=i+1}^n x_{ij}y_{ij}, \quad (2.4)$$

where  $i$  and  $j$  are row and column indices. This is based on the non-normalized Pearson product moment correlation coefficient [Di83]. However, the test is conditional on the dis(similarity) index used.

To standardize the Mantel's statistic, it was suggested [Le98] that we should take

$$r_M = \frac{1}{d-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left( \frac{x_{ij} - \bar{x}}{S_x} \right) \left( \frac{y_{ij} - \bar{y}}{S_y} \right), \quad (2.5)$$

where  $d = \left\lceil \frac{n(n-1)}{2} \right\rceil$  is the number of distances in the upper triangle part of each matrix.

Mantel's  $r$  statistic is similar to the coefficient of linear correlation, known as Pearson's  $r$  statistic. To test rank-transformed data, a ranked Mantel's statistic, ranked  $M$ , may be computed by converting within-matrix rank distances into ranks before computing  $rM$ . The correlations are comparable to the nonparametric Spearman correlations  $r_s$ .

In this study, the *environment* vector is ordinal, therefore, we used the ranked  $M$  ( $rM$ ) in lieu of  $r_s$ .

We employed one tailed hypothesis calculated for positive test statistics. To assess the significance of a departure from zero correlation, the rows and columns of one of the matrices are subjected to random permutations 10,000 times, with the statistic recalculated after each permutation. The significance of the observed statistic is the proportion of the permutations that lead to a higher correlation coefficient.

We can formulate two hypotheses:  $H_0 : r_s = 0$  and  $H_1 : r_s > 0$ .

We define  $A$  to be the presence-absence matrix,  $B$  the relative abundance matrix, and  $C$  the environment matrix. We predict that both presence-absence and relative

abundance will change as a function of environmental change so that  $r_s(AC) > 0$  and  $r_s(AB) > 0$ .

We are aware that multiple comparisons may increase the type I error of the statistics of significance for each comparison. In [Fe02] it is suggested that the use of adjusted  $p$ -values should be reconsidered since it increases the chance of making type II errors and requires an increase in sample size. The latter point is of particular importance in paleontological studies. Following the suggestions described in [Fe02], we present unadjusted  $p$ -values and combine the study's statistical significance with the magnitude of the effect, the quality of the study, and findings from other studies instead of adjusted  $p$ -values.

## 2.4 Results

Correlating relative abundance with the local environment change observed in 'Ubeidiya resulted in a significant correlation between the two variables; specifically,  $r_s = 0.543$  and  $p = 0.005$ . This would suggest that the faunal community changed over time due to the environmental change observed by the geomorphological analysis.

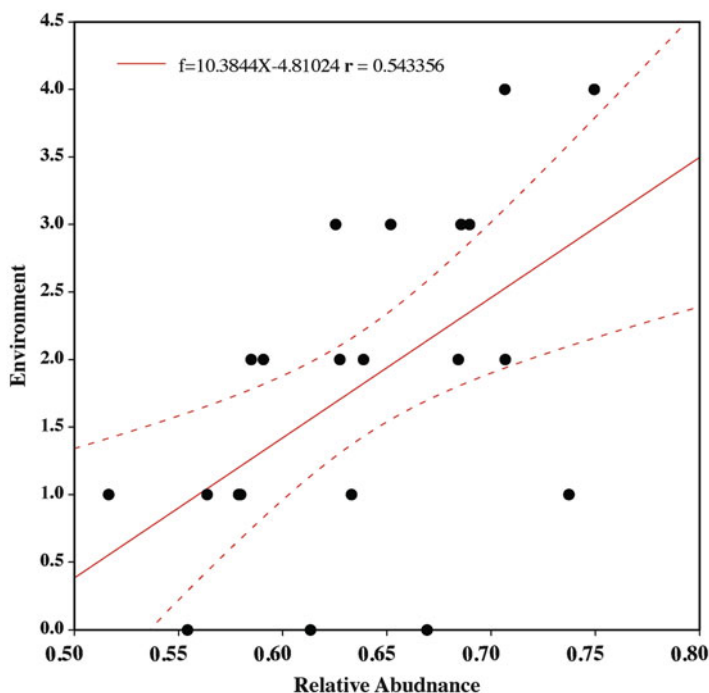
However, contrary to expectations, there is no correlation between *presence-absence community structure* and *environment* ( $r_s = 0.295$ ,  $p = 0.121$ ). This suggests a pattern of inertia in community presence-absence across the  $4 \times 10^5$  years represented by the sequence in 'Ubeidiya (Figure 2.3).

## 2.5 Discussion and Conclusion

The question of identification of stasis or change in community structure has implication for understanding the tempo and mode of ecological and evolutionary processes. The site of 'Ubeidiya is dated within this time period of the "Pleistocene paradox." It is younger than most sites that exhibit coordinated stasis (greater than circa. 100 Ma), yet older than the glacial Pleistocene sites (that is, less than 0.8 Ma).

Applying (dis)similarity indices and the Mantel's test, we concluded that *relative abundance community structure* correlated with *environment*. A detailed observation of species distribution throughout the sequence shows that fallow deer *Dama* sp., roe deer *Capreolus* sp., the large extinct deer *Praemegaceros obscurus*, and the North Africa ass *Equus tabeti* shift their abundance among the strata. Over time, there is a shift from a high to low proportion of woodland taxa (fallow and roe deer) with a concomitant increase in open grassland taxa (the *Praemegaceros* and ass). Thus, changes in the large mammalian fauna in 'Ubeidiya reflect a local environmental shift towards greater aridity [Be06].

In contrast to the pattern of relative abundance, the *presence-absence community structure* did not correlate with environmental change, which implies inertia over the



**Fig. 2.3** Cluster diagram showing the correlation between similarity in large mammal relative abundance and similarity between environmental gradient

entire sequence. This persistence in species incidence across the ‘Ubeidiya sequence is interpreted in accordance with the “recurrent assemblages” model suggested in [Mi93] for paleontological assemblage and supported for modern ecological communities [Ra90].

The response of taxa to climatic shift is dependent on the amplitude and frequency of climatic change [Ra90]. During low and medium amplitudes of environmental shift, taxa may be able to tolerate the change. Thus, despite independent evidence for a climatic shift, no change is observed in the fossil record. In higher amplitudes of environmental shift, taxa will shift their range. This is often observed in the fossil record as a change in abundance. In very high amplitudes of climatic shift, taxa will become extinct. This may be observed in the fossil record as faunal turnover. Relative frequency of species may fluctuate, whereas species presence–absence may remain constant over time [Ra90].

Consequently, the different pattern of turnover and stasis for species presence–absence and relative abundance, which are apparent in ‘Ubeidiya, may be attributed to persistence that occurred during periods of low amplitude environmental change. While species changed in their relative abundance, the amplitude of climate change was not enough to evoke change in species presence–absence pattern. A similar pattern was found in the middle Pleistocene site of Atapuerca, Spain [Ro11], which



may be attributed to medium amplitudes of climatic change, large enough to result in faunal turnover but sufficiently low to maintain a similar ecological structure of the community.

The use of similarity indices and the Mantel's test allows us to illustrate the utility of similarity matrices in the study of paleontological community structure in time.

## References

- [Ba93] Bar-Yosef, O., Goren-Inbar, N.: The Lithic Assemblages of 'Ubeidiya, a Lower Paleolithic Site in the Jordan Valley. The Institute of Archeology/The Hebrew University of Jerusalem, Jerusalem (1993)
- [Be02] Belmaker, M., Tchernov, E., Condemi, S., Bar-Yosef, O.: New evidence for hominid presence in the Lower Pleistocene of the Southern Levant. *J. Hum. Evol.* **43**, 43–56 (2002)
- [Be06] Belmaker, M.: Community structure through time: 'Ubeidiya, a Lower Pleistocene site as a case study. Ph.D. Dissertation, The Hebrew University of Jerusalem (2006)
- [Di83] Diets, E.J.: Permutation tests for association between two distance matrices. *Syst. Zool.* **32**, 21–26 (1983)
- [Fe02] Feise, R.J.: Do multiple outcome measures require p-value adjustment? *BMC Med. Res. Methodol.* **2**, 8–12 (2002)
- [Go71] Gower, J.C.: A general coefficient of similarity and some of its properties. *Biometrics* **27**, 857–871 (1971)
- [Ho96] Holterhoff, P.F.: Crinoid biofacies in Upper Carboniferous cyclothem, midcontinent North America: faunal tracking and the role of regional processes in biofacies recurrence. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **127**, 47–81 (1996)
- [Le98] Legendre, P., Legendre, L.: Numerical Ecology. Elsevier, Amsterdam (1998)
- [Ma06] Mallol, C.: What's in a beach? Soil micromorphology of sediments from the Lower Paleolithic site of 'Ubeidiya, Israel. *J. Hum. Evol.* **51**, 185–206 (2006)
- [Ma71] Mantel, N.: The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220 (1967)
- [Mi93] Miller, W.: Models of recurrent fossil assemblages. *Lethaia* **26**, 182–183 (1993)
- [Ra90] Rahel, F.J.: The hierarchical nature of community persistence: a problem of scale. *Am. Nat.* **136**, 328–344 (1990)
- [Ro11] Rodríguez, J., Burjachs, F., Cuenca-Bescós, G., García, N., Van der Made, J., Pérez González, A., Blain, H.-A.: One million years of cultural evolution in a stable environment at Atapuerca (Burgos, Spain). *Quat. Sci. Rev.* **30**, 1396–1412 (2011)
- [So63] Sokal, R.R., Sneath, P.H.A.: Principles of Numerical Taxonomy. W.H. Freeman, San Francisco (1963)
- [St01] Sterelny, K.: The reality of ecological assemblages: a palaeo-ecological puzzle. *Biol. Philos.* **16**, 437–461 (2001)

Integral Methods in Science and Engineering, Volume 2

Practical Applications

Constanda, C.; Dalla Riva, M.; Lamberti, P.D.; Musolino, P. (Eds.)

2017, XXIV, 312 p. 158 illus., 83 illus. in color.,

Hardcover

ISBN: 978-3-319-59386-9

A product of Birkhäuser Basel