A model of graphic diversity in marks on Palaeolithic artefacts based on distance to neighbouring sites

1. Introduction

In a paper from 2013, Boyd, Richerson, and Henrich discusses models of the cultural evolution of technology predicting that smaller and more isolated populations will have fewer and less complex sets of technological artefacts (Boyd et al., 2013). For the Cultural Data Science project, I was interested in testing whether a similar model could be transferred to account for the variation in graphic marks on mobile objects found in the Aurignacian techno-complex (dating from ca. 43,000 to 30,000 bp.) recorded in the *SignBase* (Dutkiewicz et al., 2020). This data set contains information about 531 mobile objects from 65 different archaeological sites, encoding the types of graphic marks or signs observed on each object according to a classification scheme with 31 different types (see Dutkiewicz et al., 2020 for more details on the classification and sign types).

While the data set doesn't contain direct information about either estimated population sizes or inter-site contact, it does contain raw geographical information which can be used to represent the sites as a network. In this network, sites can be more or less close to neighbouring sites, and it is this relative proximity that is used as the predictor for the model in this paper. The central question investigated here is whether a site's distance to its neighbours in the network of sites correlates with the number of distinct types of graphical marks observed in the mobile objects belonging to that site. Based on the theoretical model of Boyd et al. 2013, the prediction would be that the *mark diversity increases as the distance to the neighbouring site decreases*.

2. Methods

2.1 Pre-processing

Two steps of pre-processing were necessary in order to conduct the desired analysis. First, a new column containing the total number of distinct graphic features observed on all objects from each site was computed. The resulting count variable is referred to as the *unique feature count* and the distribution of this variable across sites is visualised in figure 2a.

Second, the distance between each archaeological site and its three nearest neighbouring sites were computed based on the latitude and longitude information in the SignBase, following the so-called *brute force method* described in (Agrawal, 2021). Based on a distance matrix encoding each site's Haversine distance all other sites, a new variable containing the log of the mean distance to the three nearest neighbours was added to the data set. The mean distance was logged to enable an analysis of possible effects of differences in distance magnitudes, not the absolute distance. In addition, the measure was centred so that the average of the logged mean distances would correspond to 0. Figure 2b shows the location of each site, with the colours reflecting its distance to neighbouring sites.
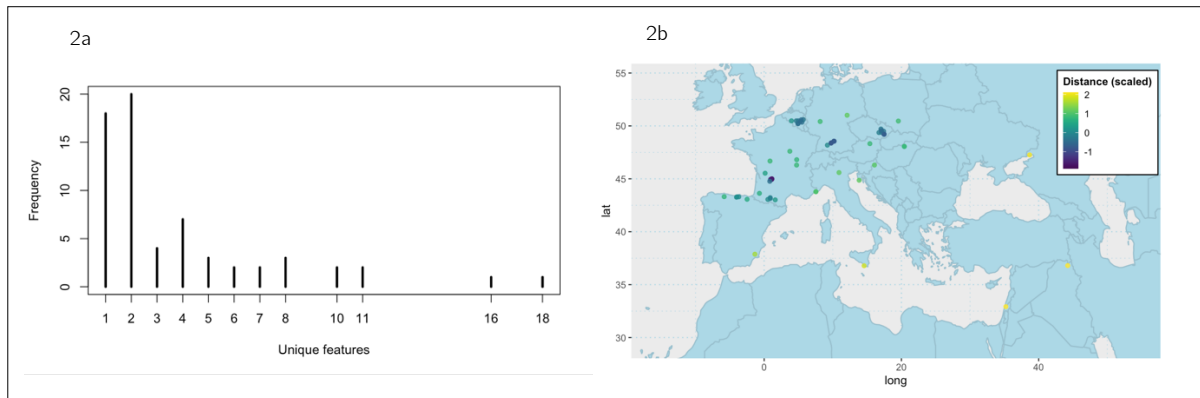
Figure 2a; 2b: Histogram showing the distribution of the unique feature count for each site; All archaeological sites in the SignBase, coloured by the standardised log mean distance to the 3 nearest neighbours.

2.2 Analysis

A Poisson regression model was fitted in R using ULAM to describe the relationship between the unique feature count F and the distance variable D. The model can be formalised as:

$$F_i \sim Poisson(\lambda_i)$$

$$log\ \lambda_i = \alpha + \beta \cdot D_i$$

The primary interest of the analysis is the estimated size and direction of the beta coefficient for the distance variable. The model was informed by moderately restrictive priors for both coefficients, arrived at after prior predictive checks (see the analysis scripts for more detail). Both the model and the modelling process was inspired by the analysis of Kline & Boyd's oceanic foraging toolset data set (Kline & Boyd, 2010) conducted by McElreath in ch. 11 of *Statistical Rethinking* (McElreath, 2020).

3. Results

The posterior distribution of the models' beta estimates has a mean of -0.299, SD = 0.064, suggesting a fairly certain estimate. The bulk of the distribution is between the negative range [-0.407, -0.189], as indicated by the 5.50% and 94.50% intervals. The mean of the posterior distribution of intercept (alpha) estimates is at 1.258, SD = 0.068, with the following 5.50% and 94.50% intervals [1.155, 1.358].

4. Discussion

The model reliably predicts negative relationships between a site's distance to neighbouring sites and the average number of unique features to be found in graphic signs at that site, as reflected in the negative beta estimates. To make the predicted effect of increases in the mean logged distance on the feature count easier to interpret, predicted values are drawn from the model and plotted against the actual data points in figure 4b. While the effect is not huge (decreasing from predicting around 6 features when the mean logged distance = -2 to around 2 features when the

distance = 2), it is still a notable decline in graphic diversity for each unit increase in the distance variable.

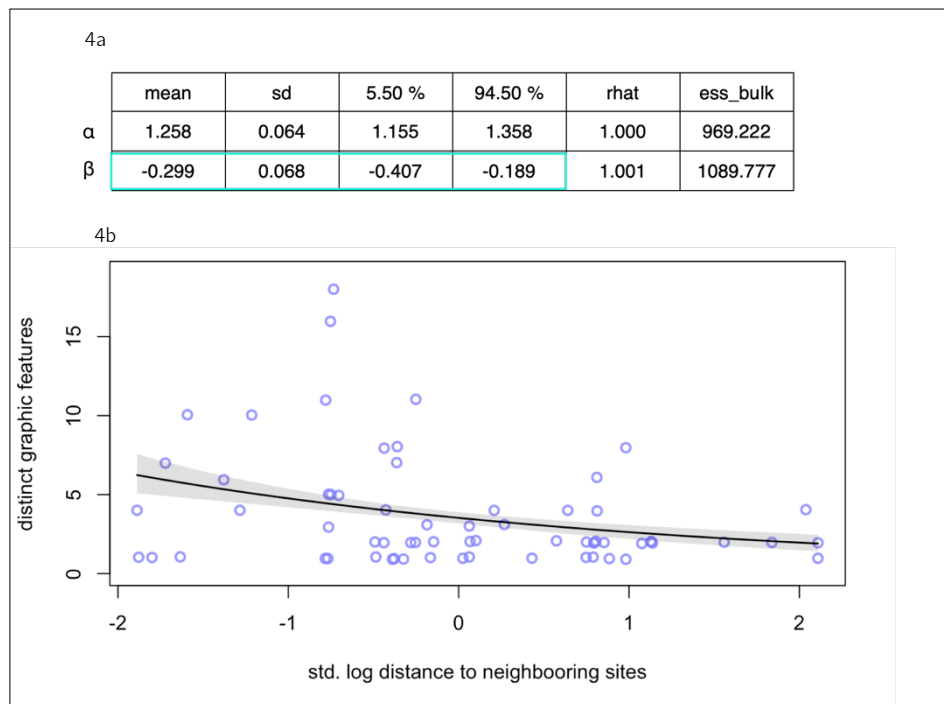|	mean	sd	5.50 %	94.50 %	rhat	ess_bulk
α	1.258	0.064	1.155	1.358	1.000	969.222
β	-0.299	0.068	-0.407	-0.189	1.001	1089.777

Figure 3a; 3b: Model coefficient estimates; Model predictions plotted against actual data.

The certainty of the beta estimates and the direction of the predicted relationship suggests that the model presented by Boyd et al. (2013) described in the introduction, could have explanatory power when it comes to understanding the graphic diversity of signs on objects found at separate archaeological sites. This interpretation would rely on adopting a framework in which signs are considered technologies that are subjected to similar processes of cultural transmission and drift as the more traditional technologies (such as foraging tools) used as examples by Boyd and colleagues.

Yet there are critical differences that makes a direct transfer of the model problematic. Most importantly, there is no guarantee that the proposed measure of the mean distance to neighbouring sites is a reliable proxy for either the population size or the inter-site population. Future iterations of the statistical model would thus benefit from including more direct information about these variables insofar as it is available. A more general question is whether the archaeological sites and the graphic signs on the mobile objects recorded in the SignBase are representative of the actual distribution of sign-creating behaviour in the particular spatiotemporal setting from which they originate. Presently, this is assumed, but further work could investigate the validity of this assumption by consulting the archaeological literature to enable a more critical assessment of the data base used for the analysis.

References

Agrawal, V. (2021, August 4). A Dive into Geospatial Nearest Neighbor Search. *Vibhu's Blog*.

       https://www.vibhuagrawal.com/blog/geospatial-nearest-neighbor-search

Boyd, R., Richerson, P. J., & Henrich, J. (2013). The Cultural Evolution of Technology: Facts

       and Theories. In P. J. Richerson & M. H. Christiansen (Eds.), *Cultural Evolution* (pp.

       119–142). The MIT Press. https://doi.org/10.7551/mitpress/9894.003.0011

Dutkiewicz, E., Russo, G., Lee, S., & Bentz, C. (2020). SignBase, a collection of geometric

       signs on mobile objects in the Paleolithic. *Scientific Data*, *7*(1), 364.

       https://doi.org/10.1038/s41597-020-00704-x

Kline, M. A., & Boyd, R. (2010). Population size predicts technological complexity in Oceania.

       Proceedings of the Royal Society B: Biological Sciences, *277*(1693), 2559–2564.

       https://doi.org/10.1098/rspb.2010.0452

McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*

       (Second edition). CRC Press. https://doi.org/10.1201/9780429029608