Journal of Archaeological Science 173 (2025) 106106

Contents lists available at ScienceDirect

Journal of Archaeological Science

journal homepage: www.elsevier.com/locate/jas

# Beyond baselines of performance: Beta regression models of compositional variability in craft production studies

Jasmine Vieri [a,*], Enrico R. Crema [a,b], María Alicia Uribe Villegas [c], Juanita Sáenz Samper [c], Marcos Martinón-Torres [a,b]

[a] McDonald Institute for Archaeological Research, University of Cambridge, Downing Street, Cambridge, CB2 3ER, United Kingdom
[b] Department of Archaeology, University of Cambridge, Downing Street, Cambridge, CB2 3DZ, United Kingdom
[c] Museo del Oro, Banco de la República, Calle 16 #5-41, Bogotá D.C., Colombia

ARTICLE INFO

Dataset link:
https://github.com/jmkvieri/BBLoP

*Keywords:*
Beta regression
Bayesian statistics
Modelling
Chemical compositions
Craft production
Muisca
Gold alloys

ABSTRACT

Chemical analyses of archaeological artefacts are often used for provenance studies and for assessing whether specific performance characteristics were targeted by craftspeople in the past. Traditionally, the answers to these questions were sought by identifying compositional averages and by studying their correlations with either the geochemical signatures of candidate raw material sources or the corresponding physical or chemical properties of the studied materials. However useful, this approach only exploits part of the potential information locked inside the chemical compositions of archaeological artefacts. We argue that different levels of compositional dispersion observed within and across archaeological assemblages, and in particular changes in them as a function of behaviourally meaningful factors (such as the size, function, or recovery location of the objects), are sources of information in themselves. To gain probabilistic insights into both types of variability (averages and dispersions) simultaneously, we introduce variable dispersion beta regression models for the archaeological sciences. In doing so, we show how adopting the beta distribution provides a significantly improved alternative to previous solutions to modelling compositional data within the field — namely, those involving simple linear regression on log-transformed data. These approaches often result in numerically impossible predictions, whilst beta regression restricts the model predictions between the upper and lower compositional bounds, accounts for the inherently inconsistent variances of compositional data, and explicitly permits the modelling of compositional dispersions as a function of covariates. Finally, we expand upon this toolset by showing how using a hierarchical model specification within the framework accounts for both local variation and more widely shared practices of material processing and procurement concurrently, and alleviates issues to do with sampling uncertainty. We demonstrate the proposed approach with a study of Muisca gold procurement practices (AD 600–1600) in the Eastern Highlands of Colombia, based on a dataset of 243 elemental analyses. The results allow us to argue for intra-regional movements of fresh geological gold imported from a variety of distant sources. We suggest these movements could result from contributions of gold by people converging into the same location for festivities. The approaches taken to modelling compositional data are readily applicable to other sub-disciplines of the archaeological sciences, such as compositional studies of ceramics and glass, or modelling the variability of diets in isotopic studies (see Supplementary Material S0 for an extended summary in Spanish).

## 1. Introduction

Technological studies have long recognised that the desired performance characteristics of materials and technologies are culturally construed (Schiffer and Skibo, 1997). However, considerations of social and cultural meaning have remained more successful in compositional studies focused on local scales of analysis, than on regional or cross-regional ones. This is largely down to how archaeometric evidence and its potential have been viewed in the past. For instance, in their seminal paper, Sillar and Tite (2000, 17) highlighted how archaeological materials science "allows us to assess the extent to which physical and chemical performance characteristics have influenced past technological choices and, thus, provides a baseline against which the role of cultural factors can be considered". This search for culturally preferred performance characteristics – or, equally, the cultural processes of raw

* Corresponding author.
*E-mail address:* jmkv2@cam.ac.uk (J. Vieri).

https://doi.org/10.1016/j.jas.2024.106106
Received 31 August 2024; Accepted 24 October 2024

material procurement – has typically manifested in the calculation and comparison of compositional averages, including across large spans of space and time. Focusing on the averages alone, however, fails to properly account for individual variation, in turn limiting the insight into the intentionality of any one craftsperson in terms of the target performance properties, or in terms of their practices of procuring and processing raw materials.

At the same time, while seemingly easier to reconstruct, neither can local-scale decision-making processes be fully understood unless they are situated within their broader social and technological systems. In being informed by their society's cultural values and ideological concepts (Lechtman, 1977; Sillar and Tite, 2000), people often make technological choices on a subconscious level. Behavioural proxies may then only become meaningful in a relational network, rather than reflecting specific design decisions made on individual artefacts. In other words, the signatures observed for any one context may only have meaning within the broader context of similar or dissimilar ones. This represents a dilemma, in both necessitating that the field accounts for how artisans and consumers operate within their social and historical contexts, whilst also implying that such contexts are methodologically more challenging to reconstruct.

For a more nuanced interpretation of compositional proxies within the archaeological sciences, this paper presents two new ways of thinking about and modelling compositional variability in large-scale archaeological datasets. The new tools put forward intend to explicitly address both (A) how studying compositional dispersions complements insights from their typical tendencies, even where the intentionality of artisans is beyond our reach, and (B) how local and more widely shared cultural practices of material processing and procuring vary concurrently.

To provide a robust framework for doing so, we also present a new approach to modelling compositional data within the field. Theoretical discussions around how to derive behaviourally meaningful models from compositional data have long been accompanied by methodological disagreement over how to treat compositional data in statistical applications (Aitchison et al., 2002; Baxter, 1995; Baxter and Freestone, 2006; Tangri and Wright, 1993), with log-transformations being used as the standard go-to approach. This paper offers a new, improved modelling solution based on the beta distribution, which properly accounts for the constrained nature of compositional data (i.e., the data being bounded between 0 and 100 wt%).

Utilising the beta regression framework, we then expand upon new ways of modelling compositional variability, introducing variable dispersion submodels and using archaeometallurgical datasets as an example. We propose to treat changes in compositional dispersions as an opportunity to identify processes that have led to their creation, and thus provide new tools that allow us to move beyond "baselines of performance" in the chemical analysis of archaeological materials. We show that *regardless of the underlying drivers of individual decision-making processes*, assessing these changes and their drivers at different scales of analysis is an important source of information in itself, which complements or even alters archaeological interpretations based on compositional averages. In particular, such an approach allows us to make inferences about the broader systemic contexts of craft production, beyond the level of individual production episodes. Thus far, such changes in compositional dispersions have not fully been exploited as an archaeological source of information. In particular, excess emphasis on compositional means has typically resulted in researchers considering the dispersion of observations around their sought-after averages mainly as a source of "noise" in the data, or as something undesirable and largely uninformative. Even within broader archaeological modelling applications, model variance has mainly been treated as an indication of how strong the association between the predictors (e.g. distance to ore sources) and the modelled outcome (e.g. compositions) is. This is typical even if this "noise" is modelled as being influenced by the modelled predictors, to allow for more statistically robust inferences.

In archaeometallurgy, at the local scales, some have explored descriptive statistics such as coefficients of variation (CVs) in slag composition to compare standardisation in metallurgical engineering (Humphris et al., 2009; Pryce et al., 2010). Overall, while some have also focused attention towards the shape of different compositional distributions or cumulative frequencies and what these tell us about primary/secondary alloying practices (Pollard et al., 2018), no explicit attempts have been made to examine compositional dispersions in probabilistic terms. Rather, archaeometallurgical research at regional scales is still focused on assigning objects into groups based on compositional signatures and mapping these fixed assignments (Perucchetti et al., 2020), and/or making inferences on the basis of comparing point estimates, such as average compositions, across different cultural groups or time periods (Bray and Pollard, 2012; Pollard and Bray, 2014). Alternative approaches have included, for instance, grouping metals according to presence/absence classifications of certain elements and conducting ubiquity analysis of the different groups, providing potential insights on recycling (Bray et al., 2015).

These approaches, although contributing to our understanding of the bigger picture, neglect important information locked inside the chemical compositions of archaeological artefacts. Whilst we agree that processes of recycling provide insights into the changing value of metals and their contexts of production, we argue that a fundamental difference of interpretation arises if we carefully examine how compositional variability is structured. For example, if we observe an increased frequency in a highly mixed alloy type (see for example Bray et al., 2015) accompanied by high levels of dispersion of compositions within the mixed alloy group, we might interpret the evidence at hand as increased improvisation and lack of centralised control over metal procurement and processing practices. However, if the overall dispersion of compositions were shown to be low against this surge in the use of a highly mixed alloy type, the observed pattern might suggest that all the metal in circulation at some point became pooled and recycled at a few centralised locales. Focusing on averages or the ubiquity of different compositional groups fails to account for such differences, whilst simultaneously modelling both averages and dispersions can do so.

At the same time, to fully appreciate the continuity and discontinuity of human engagements with technological knowledge and their environments, craft production studies also need to find ways of assessing patterns of both local variation and supra-local shared practices concurrently. This applies even within the context of the more traditional discussions focused on the typical compositions and their drivers. If the compositional signatures of different local groups are aggregated into a single homogeneous entity for analysis, the patterns reflecting any local choices become easily obscured. In addition, this artificial aggregation simultaneously obscures the overarching drivers of craft production practices at the broader scales by placing excess emphasis on the patterns observed for the more intensively sampled locales, and by taking their average to be representative of the whole. As such, not only have archaeologists placed excessive emphasis on average compositions, but also, in ignoring sample interdependence, they have also risked focusing on *artificially constructed averages*. This is despite the availability of statistical approaches, known as multilevel models (also known as hierarchical models), that can explicitly account for sample interdependence, whilst simultaneously modelling variation at the global scales (for other archaeological applications, see: Banks et al., 2019; Perri et al., 2019; Wolfhagen, 2020; Fernée and Trimmis, 2021; Crema et al., 2024).

Statistical modelling, generally speaking, has been underexploited in archaeological craft production studies. Modelling changes not only in the typical compositions, but also in their dispersions, can drastically alter any archaeological interpretations based on the data. Assessing how much compositional variability is explained by shared practices at the regional level and how much by differences across sites, in turn, provides insight into, e.g., the strength of regional standardisation, or

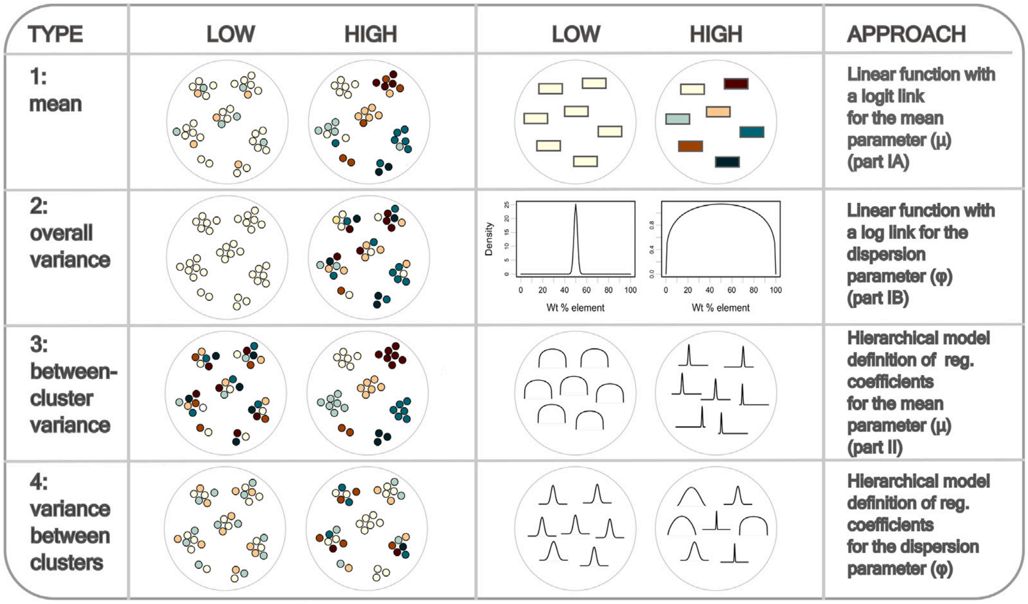


**Fig. 1.** Hypothetical examples of the four main types of variability encountered in compositional archaeological datasets, alongside the suggested approaches for modelling them within the beta regression framework. Circles represent individual artefacts, with the colours depicting different hypothetical chemical compositions. Each cluster of circles, in turn, reflects a different sub-grouping in the data, e.g., different archaeological contexts of discovery. *Type 1* variability captures variation in compositional averages across such sub-groupings. *Type 2* variability, in turn, captures the overall dispersion of compositions for each dataset. *Type 3* variability captures how the chemical compositions vary at the local and supra-local levels concurrently. For instance, the example with low variability is one where the structure in the compositional variability cannot be explained by differences across the sub-groupings. In contrast, the example with high variability suggests that all of the compositional variability in the data is explained by differences across these sub-groupings. Finally, although not explored in this paper, *Type 4* variability can be used to model how compositional dispersion varies at the local and supra-local levels concurrently, shedding light on processes of standardisation at different scales of analysis, and could be done by extending hierarchical model definitions to the variable dispersion sub-model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

lack thereof, in craft production practices. More broadly, discussions within archaeology are often centred around the variability of cultural practices and what this may imply about past human behaviour – for instance, in relation to non-specialised and specialised craft production settings (Costin, 1991; DeMarrais, 2013). Similarly, others have proposed that periods of innovation are associated with higher degrees of variability than periods of technological stagnation (Eerkens and Lipo, 2005). However, while this concept of "variability" is frequently referenced, its precise implications for archaeological data are often not clearly defined.

Fig. 1 conceptualises the four main different sources of compositional variability, in particular, that can be encountered in archaeological datasets. To introduce the reader to each type in more detail, this paper will be divided into three main parts. The first introduces beta regression as the new standard for archaeological compositional data analysis and how it can be used to model compositional averages (IA), followed by the incorporation of variable dispersion submodels into this framework (IB). It is targeted at the general archaeological science audience, to encourage the widespread application of the approaches within the field. The second part then builds upon these tools with the adoption of hierarchical model definitions in compositional craft production studies (II). This second part does not go into detail about the underlying mathematical principles, but, rather, expects some knowledge of the relevant statistical approaches from the reader, given that multilevel modelling approaches have previously been discussed by other archaeologists (see citations above; for more generic and thorough introductions to the mathematical principles behind multilevel modelling, see, e.g., Goldstein (1987), McElreath (2020), with the latter also providing extensive examples of practical applications in R and Stan). Finally, the third part puts the tools introduced in both of these sections into practice, by presenting an archaeological case study focused on Muisca metal procurement and processing practices from pre-Hispanic Colombia (III).

## 2. Part I: Modelling compositional data

### 2.1. Part IA: Working with the compositional constraint

Modelling compositional data presents challenges because they exhibit features that are not tractable by most conventional statistical approaches. The first challenge relates to the skewness or asymmetry of compositional data, which occurs as the result of the compositional constraint (i.e., their constituents summing up to a constant – in the case of analytical chemical data on object compositions, 100wt% after normalisation). This means that compositions are likely to violate the normality assumption that is inherent to many statistical approaches, such as simple linear models that form the basis of many standard tests commonly used by archaeologists, including t-tests and ANOVA, as well as more complex modelling approaches. A closely related issue is that compositional variables are typically heteroskedastic, whereby their variance tends to be higher when the central tendency falls closer to half the maximum value, e.g. 0.5, than when it falls near the lower and upper bounds, e.g., 0 or 1 (Cribari-Neto and Zeileis, 2010) — whereas most statistical tests and models assume homoskedasticity. Compositional data, therefore, require either the use of an alternative statistical distribution to the Gaussian, or the use of data transformations.

Until now, a common solution within archaeology has been to use logarithmic data transformations, to make the data suitable for more conventional statistical approaches, given that they can often reduce skewness and help the data approximate a normal distribution. They do, however, also routinely fail to do so, especially but not only when the data have a left-skewed distribution. Log-ratio transformations are further sometimes used when dealing with more elements of interest, following Aitchison's proposal in 1982 (Aitchison, 1982), given the need to further account for the constant sum constraint in such cases (Aitchison et al., 2002; López-García et al., 2018; Pawlowsky-Glahn and Egozcue, 2006). Others have opted to apply conventional statistical approaches to non-transformed data (Tangri and Wright,

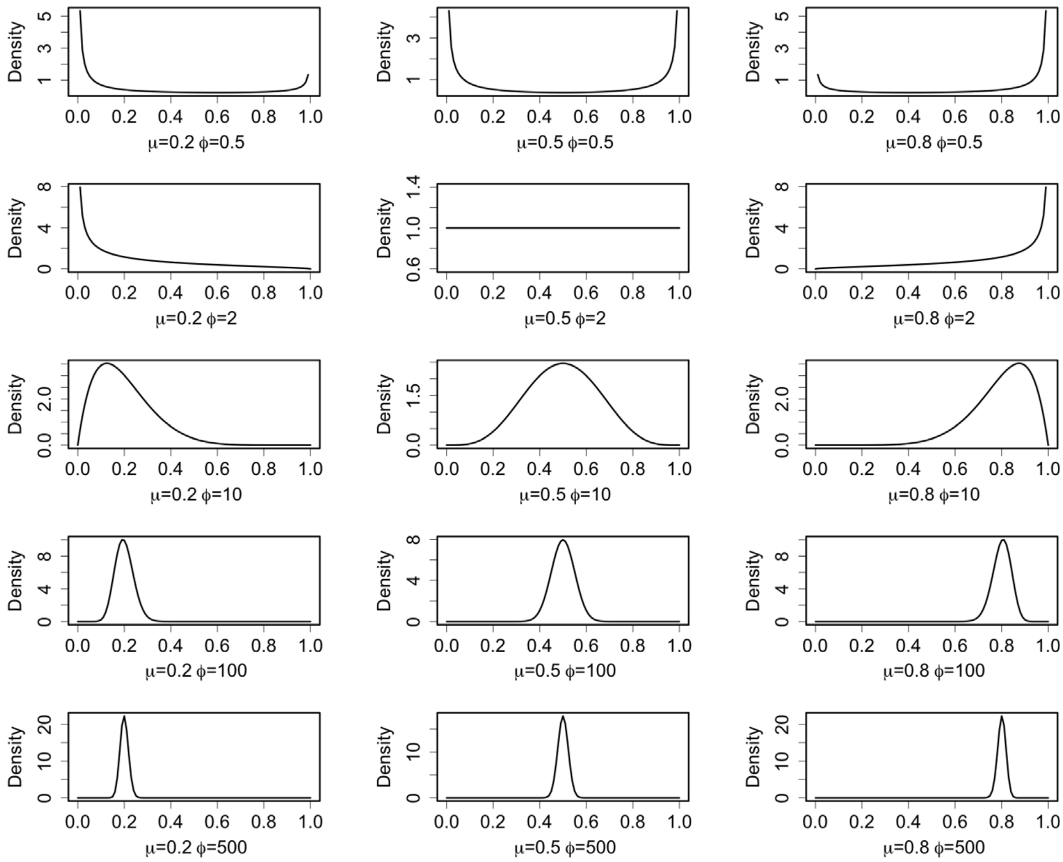


**Fig. 2.** Beta densities at different values of $\mu$ (mean) and $\phi$ (dispersion). Note the flexibility of the beta distribution in allowing for a multitude of shapes for the compositional outcome, depending on the corresponding parameter values. Note, also, how the overall dispersion is higher when the central tendency falls towards the middle of the interval (0, 1) than towards the two compositional extremes, even where the dispersion parameter is fixed at the same value, e.g., on the third row where $\phi = 10$.

1993). Some have pointed out that while the elemental composition of raw materials may be expected to follow a normal or a log-normal distribution for a specific geological source, archaeological materials are often the product of mixing various sources of raw material, resulting in more complex signatures (Baxter, 1995, 515). Log-transformation may thus mask archaeologically important information.

We will show that neither approach is ideal, given that traditional statistical models of compositional data based on both non-transformed and log-transformed data are prone to making predictions outside of the sample outcome space (e.g., 120wt%), potentially biasing inferences on the strength and nature of the relationships with covariates.

**A new solution: the beta regression**

We present an alternative solution for analysing compositional data whereby the constrained nature of the data is directly modelled, as opposed to being dealt with data transformations. We do so by adopting the beta regression structure, first proposed by Ferrari and Cribari-Neto (2004), which works on a similar basis to other Generalized Linear Models (GLMs). The key idea is to model the compositional variable of interest as being beta-distributed and therefore restricted to the standard unit interval (0, 1).[1] This will accommodate for skewness in the data. In fact, the beta distribution is highly flexible in allowing for a multitude of different shapes (Ferrari and Cribari-Neto, 2004, 801-802), including skewed, symmetric, U-shaped, or even flat distributions, depending on the corresponding parameter values, as shown in Fig. 2. Therefore, it can also accommodate several compositional signatures, for instance, those of assemblages including either recycled or freshly alloyed objects (Pollard et al., 2018).

We thus substitute the more commonly used Gaussian probability distribution with the beta distribution in our basic model definition. We use the alternative parametrisation of the distribution introduced by Ferrari and Cribari-Neto (2004), with the parameters $\mu$ and $\phi$ representing the mean and dispersion of our distribution. As the authors point out, these are more straightforward to interpret in terms of modelling outcomes compared to the more traditionally used shape parameters, $p$ and $q$ (see Supplementary Material S1.1 for details).

A logit link function $\mathrm{logit}(\cdot)$ is used to map the non-linear relationship between the response and the predictors onto a linear one, following the same principle used in other GLMs. Thus, given an observed proportion $y_i$ for the sample $i$ (e.g. percentage of Cu for a given artefact), we can describe our model using the following probabilistic notation: For $i = 1 \ldots n$ :

$$y_i \sim \mathrm{Beta}(\mu_i, \phi) \tag{1}$$

$$\mathrm{logit}(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$$

where $x_{ik}$ are the corresponding predictor values for the given sample $i$ (e.g., the object volume modelled using a continuous variable, or the object type modelled using a categorical indicator variable), with a total number of predictors $k$, and with $\beta_0$ being the intercept, and $\beta_{1\ldots k}$ the corresponding regression coefficients. Fig. 1 shows hypothetical examples of assemblages with low and high variability of this type (*Type 1*), pertaining to the average expected compositions according to some predictor.

Here, it is further important to note that $\phi$, which describes the dispersion of values around the predicted means, is a precision parameter, and therefore higher values imply less and lower values more

[1] For fitting a beta regression model, the compositional response variable, therefore, needs to be converted from percentages to proportions.

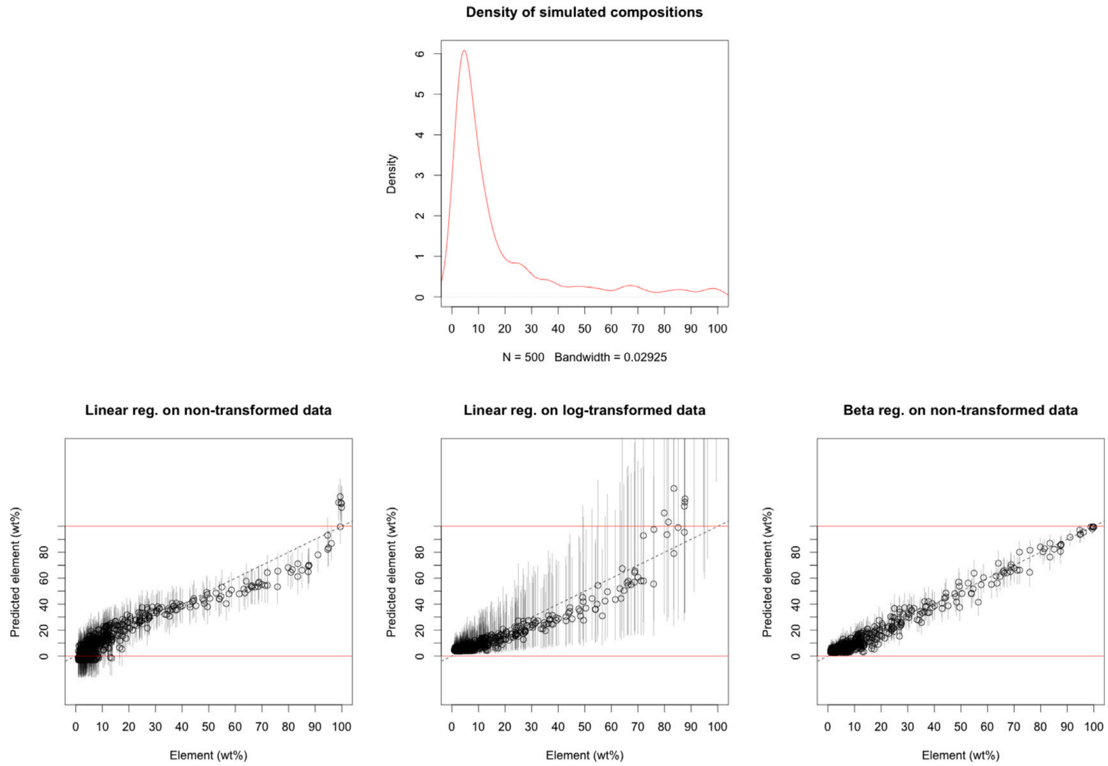


**Fig. 3.** A comparison of model performance on the same set of simulated data bounded between 0wt% and 100wt %, the density of which is shown on the top row. Left, simple linear model on non-transformed data; middle, a model on log-transformed data; right, a beta regression. Each point reflects the correspondence between an individual compositional observation in the dataset and the estimated composition returned by the model, with predicted values on the $y$-axis and original values on the $x$-axis. All models were fitted within a Bayesian framework with the vertical lines representing the 95% prediction intervals for each observation.

dispersion overall. This is in contrast to most scale parameters such as $\sigma$ in a normal distribution, for which the opposite is true. The variance of a beta regression model, in turn interpreted in the more traditional sense of higher values implying more dispersion, is calculated as:

$$\mathrm{var}(y_i) = \frac{\mu_i(1-\mu_i)}{1+\phi} \tag{2}$$

As seen here, the model variance is dependent on the values of $\mu$, and subsequently, on the values of the predictors (Ferrari and Cribari-Neto, 2004, 803). This is what allows for natural heteroskedasticity in the model predictions. In other words, the model will expect less variance when the central tendency of compositions falls towards either 0wt% or 100wt% than towards 50wt%.

Fig. 3 demonstrates how these two key features – the bounded nature of the beta distribution and the naturally heteroskedastic parametrisation of the model variance –, benefit the modelling of compositional data. It contrasts the predictive performance of a Bayesian simple linear regression on non-transformed (left) and log-transformed (middle) data to the one offered by a Bayesian beta regression (right) using the same simulated dataset. The simple linear model on non-transformed data fails to constrain the mean predicted response below the lower and above the upper compositional constraint, while the log-transformed model makes numerically impossible predictions above 100wt%. While the latter does constrain the response above 0, it simultaneously fails to account for the heteroskedastic nature of compositional data, by assuming that the variance is constant at the log-transformed scale. This results in increasingly wider prediction intervals towards higher values of the simulated elemental compositions, when transformed back to the original compositional scale. It is further worth noting that this is *despite* log-transformation having alleviated skewness in the data in this particular case (Figure S1). In contrast, the beta regression example in Fig. 3C shows that no predictions are made outside of the standard unit interval (0, 1), while the prediction intervals for each individual observation are fairly consistent throughout different compositional values. In sum, adopting the beta regression structure provides an excellent way of dealing with univariate compositional data in the archaeological sciences and beyond, owing to the following reasons:

(a) Ensures that no predictions are made outside of the interval (0, 1).

(b) Accommodates for skewed, U-shaped, uniform and quasi-normal distributions for the outcome.

(c) Allows for natural heteroskedasticity in the model, even where the dispersion is not explicitly modelled as a function of covariates.

(d) Does not require data transformations, whereby the model outputs are readily interpretable at the original compositional scale (Ferrari and Cribari-Neto, 2004).

One limitation of the beta regression approach is that any observation for the response variable cannot take values of either exactly 0 or 1. The analyst will, therefore, need to decide on how to deal with such values. In cases where the detection limit for a given element in the whole dataset is known, i.e. if dealing with an identical instrument and analytical procedures, a pre-determined proportion of this (e.g. 2/3) could be used to replace the 0s. In collated legacy data sets, one possibility is to use an artificial detection limit for the given element instead, taken to be the lowest value detected for said element. Values at the upper compositional bound can be fixed at values that are nominally close to it, e.g. 0.999. It is also worth noting that log-transformation is similarly incapable of dealing with 0s and, therefore, provides no superior solution to this problem.

### 2.2. Part IB: Model variance — an archaeologically insightful source of information

As stated above, adopting the beta distribution offers a straightforward solution to account for the inherently heteroskedastic nature of bounded data. However, in many cases, the dispersion of values around the mean can also vary as a function of covariates. These patterns of variation can provide insights on behaviourally meaningful changes in, e.g., the range of alloy compositions or geological sources of metal being exploited. Ignoring this type of structure, where it is present, also leads to biased estimates of the other model parameters. This, in turn, also affects, e.g., the predictions for the often sought-after compositional averages, even in such cases where the changes in dispersion were to have little behavioural interpretability.

Simas et al. (2010) introduced the varying dispersion beta regression model, where the dispersion parameter is also linked to a linear function with covariates through a link function $g(\cdot)$ (with a common choice being $\log(\cdot)$, which ensures that the dispersion parameter remains positive and assumes that changes in the parameter are multiplicative according to the predictors), so that for $i = 1 \ldots n$ :

$$y_i \sim \mathrm{Beta}(\mu_i, \phi_i) \tag{3}$$

$$\mathrm{logit}(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$$

$$\log(\phi_i) = \gamma_0 + \gamma_1 z_{i1} + \gamma_2 z_{i2} + \cdots + \gamma_p z_{ip}$$

Here, $z_i p$ are the new corresponding predictor values used to model the dispersion $\phi_i$, and $\gamma_{1\ldots p}$ are the corresponding regression coefficients.

At its simplest, we can adopt this new tool to examine how different cultural groups vary in terms of the standardisation of their raw material procurement and processing practices. Fig. 1 shows hypothetical examples of assemblages with low and high levels of compositional variability of this type (*Type 2*). Were the two examples relate to two time periods, for instance, the signatures could imply strong centralised control over these practices for one period, and significantly reduced levels of such standardisation or centralisation for the other. Importantly, utilising the beta regression framework also allows for assessing whether such changes are supported by the data in probabilistic terms, unlike exploratory methods such as CVs. The latter additionally risks providing biased estimates in the first place, given, once more, their inability to account for the compositional constraint.

Therefore, simply modelling the dispersion of compositions across different chronological or geographical groupings is a starting point for investigating the standardisation of craft production activities. However, not only does this framework readily lend itself to systematically comparing the variability of compositions across different metallurgical horizons or cultural groups. We can similarly extend the principle of modelling compositional dispersions as a function of any other categorical predictor (e.g. object function) and/or of a continuous predictor (e.g. the volume of the objects or the distance from their recovery location to the nearest ore sources).

Fig. 4 shows the posterior predictions of a variable dispersion beta regression model based on three different simulated datasets, where the average proportion of an element is expected to stay the same (in this case, ~50wt%), but with its variance being a function of a hypothetical continuous predictor. Such predictor takes on values between 0 and 5, resulting in compositions with the dispersion remaining constant (Fig. 4A), positively correlated (Fig. 4B), or negatively correlated (Fig. 4C), to the predictor. The top row shows the 95% posterior prediction intervals of expected object compositions at different values of the predictor variable. The bottom row shows the regression lines for the dispersion submodels.

These simulated examples demonstrate how focusing on average compositions in archaeometallurgical research projects may be misleading. In all three simulated datasets, the average compositions stay the same. However, the first dataset consistently has fluctuations of about ±9wt% around this average. In the second one, the fluctuations are within a few wt% at low values of the continuous predictor and increase to c. 25 wt% at its highest values. The opposite applies to the third example.

Assuming that these are data for alloy composition in metal objects, claiming that alloying or metal procurement practices consistently stayed the same across all three datasets based on the average compositions would result in an erroneous, or at best partial, archaeological interpretation. For instance, if the continuous predictor is the distance to the nearest ore sources (e.g., in hundreds of km), scenario B could imply that craft producers located in close vicinity to ore sources had a strong preference for specific alloy compositions, or that they practised more recycling than craft producers located further away from primary sources of metal. Scenario C could imply the opposite, for instance, with the metals having already undergone several recycling episodes before reaching consumers further away. Since we evidently do not have direct access to information on the data-generating processes such as how much recycling, and of which metals, took place at a given place and time, modelling the dispersion of compositions in this manner offers an indirect interpretative framework for their understanding.

Ultimately, the hypothetical differences in interpretation are notable and of concern because of the high prevalence of studies within the field that still rely on interpreting observed evidence exclusively in terms of average compositions, disregarding the information contained in their dispersion. The same applies to discussing compositional groupings, whether based on elemental ranges or presence/absence classifications, given their inherent inability to consider different levels of variability observed within each compositional grouping. As such, the variable dispersion approach allows us to move beyond such simplistic models in enabling us to consider how the repertoire of choices made by craft producers is impacted by broader societal processes. As already stated, models that ignore changes in compositional dispersion as a function of covariates, as in the scenarios in Figs. 4B and 4C, are also fundamentally misspecified, resulting in biased estimates of the other model parameters. Accounting for any variable dispersion structure thus simultaneously results in statistically more robust inferences.

## 3. Part II: Multilevel models of compositional data

Researchers working with compositional data at larger scales also need to account for the fact that human practices were not static through time and space. In other words, they should not generalise to broader regions or time periods based on patterns artificially induced by data aggregation, given the inherent presence of naturally occurring 'clusters' in archaeological data. Examples of these clusters include, for instance, groups of finds found in the same stratigraphic layer, from the same archaeological site, or across the same geographical region. Importantly, samples from the same cluster might be conditioned by some unmodelled variables, leading to interdependencies in the data that can bias inferences. For example, the inhabitants of one site may have less access to copper than those of another, but still make technological decisions about the desired copper contents for the specific object types they manufactured.

This problem is exacerbated when combined with sample imbalance. Non-hierarchical models assume that the relationship between the response and the covariate remains globally consistent (McElreath, 2020), therefore the presence of a single intensively sampled location with some unusual characteristics might skew the inferred global relationship to a covariate. Even where the different locales were to be represented by fairly equal and large sample sizes, models that do not account for sample interdependencies will simply take the average of all these locales, and assign this new, artificially constructed average to all of them. Where highly different alloy compositions were used at the

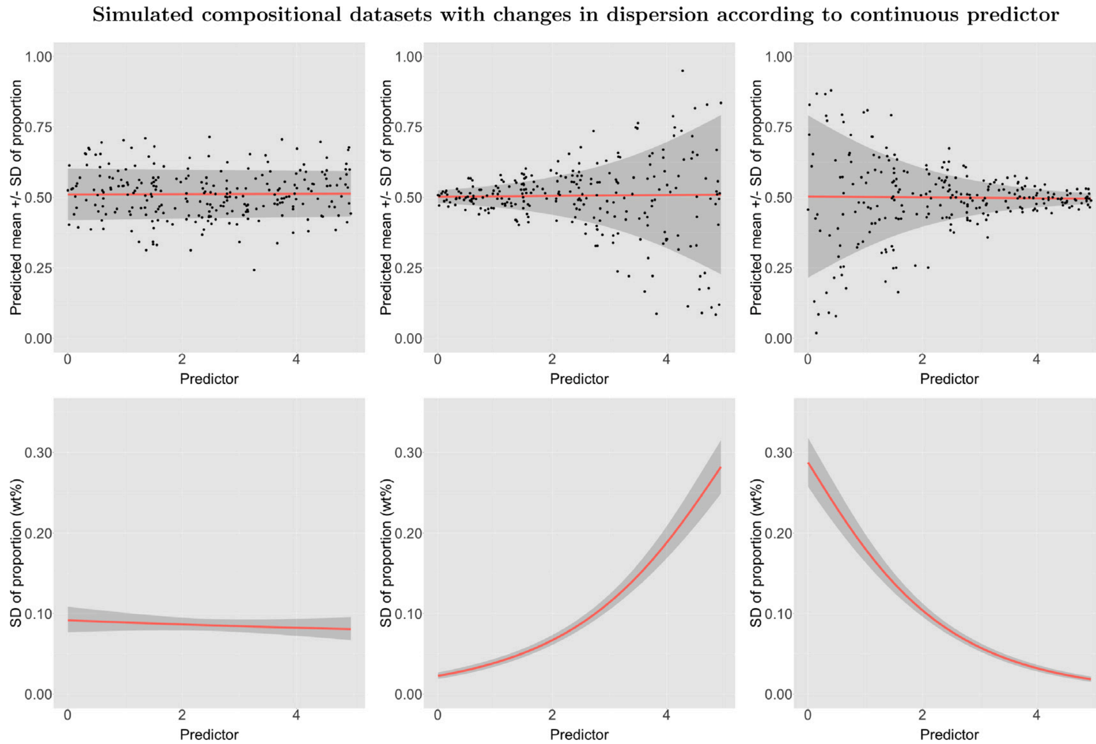


**Fig. 4.** Examples of three different simulated datasets and their respective variable dispersion beta regression model outputs. The top row shows the original simulated data points, alongside the posterior predictions for the mean predicted response shown as the red lines, with the grey envelopes showing the expected dispersion of values around the means as standard deviations, derived from the predictions for the model variance as calculated based on the values of $\mu$ and $\phi$ using the formula in Eq. (2). The bottom row, in turn, shows the regression lines for said standard deviations, with the grey envelope reflecting the 95% PI intervals. The continuous predictor was standardised before running the regression analysis and then back-transformed to the original scale, with full code used to simulate the data provided in the GitHub repository. All models are specified using a Bayesian framework. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

different settlements, the resulting average may not be representative of the alloying practices of *any* the studied sites.

Moreover, it is often of interest to explicitly model the drivers of craft production practices across different sub-groupings in the data, to appreciate any patterns of local variation. One approach could be to run a model where the compositional responses and their drivers are modelled separately for each cluster, e.g., where a categorical predictor represents each site and the other predictors are allowed to interact with it. At the same time, regional research projects tend to rely on archaeological legacy data which, in addition to representing a palimpsest of past human activities, are also inevitably the product of different sampling regimes. If one archaeological site has only produced a few objects with highly unusual compositions, this may reflect sampling error or a bias, rather than the true underlying practices of metal use at the site. Modelling each site's compositions using the categorical predictor approach will ignore any uncertainty arising from varying sampling intensities and tends to return overly confident parameter estimates for poorly sampled locales.

Finally, it is also pertinent to simultaneously understand *both* whether broadly shared practices explain most of the variability in the data, or if differences across more local groupings drive more structure in the dataset (cf. example of high levels of Type 3 variability in Fig. 1). This can provide important insights into, e.g., high levels of regional standardisation, where variation in the data is explained by the same changes in the covariates across all different local sub-groupings, or – at the other end of the spectrum, – high degrees of freedom for individual choice/expediency of raw material procurement and processing practices, with these relationships changing between locales.

Here, we propose to account for, and explicitly model, sample interdependence arising from shared influences in craft production practices by integrating hierarchical levels into our beta regression models of compositional data, thus turning them into multilevel models. As noted by Fernée and Trimmis (2021), multilevel modelling offers the ideal tool for assessing variability at different levels in nested archaeological data. Within the context of modelling compositional archaeometallurgical data, it allows for assessing how alloying and metal procurement practices operated on both local and supra-local scales simultaneously. By learning from both the compositional patterns observed globally and locally, the approach also yields results that account for sampling uncertainty by shrinking the predictions for poorly sampled locales or clusters towards the more typical compositions on the global scale (McElreath, 2020).

Hierarchical levels can be used to model only baseline variation in object compositions, in which case they are used to model the intercept, or also/only to model the impact of covariates on the compositional response, in which case they are also introduced to the slope parameters. Fig. 5 shows simulated examples of hierarchical beta regression models with (A) baseline variation only, (B) variation according to covariates only, and (C) variation in both. Note how adopting the beta distribution, once more, successfully restricts the predicted responses to the interval (0, 1).

Whilst the multilevel approach has a long history of use in other social sciences (Goldstein, 1987), its potential has been underexploited in archaeology, and in particular, it has not been previously applied to investigate compositional data. It has, however, been previously applied to the chronological modelling of radiocarbon dates (Banks et al., 2019); in stable isotope analysis, to the estimation of dietary variation (Perri et al., 2019); within zooarchaeology for biometric analyses (Wolfhagen, 2020); to investigate rates of diffusion of subsistence and cultural practices (Crema et al., 2024); and to the study of intra-site variation in stone sphere metrics, as well as to model whether the expected proportions of ware types vary more between contexts or trenches (Fernée and Trimmis, 2021). As such, the mathematical principles behind hierarchical modelling have been previously discussed

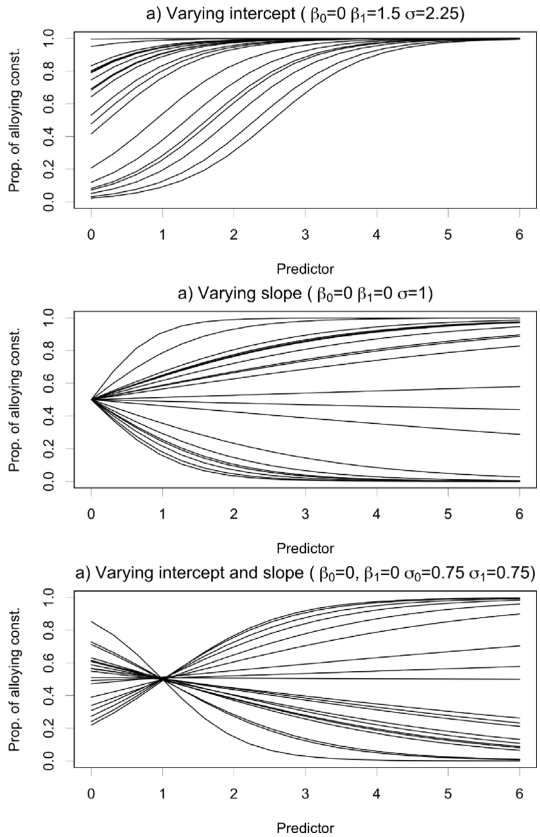


**Fig. 5.** An example of (a) varying intercepts (b) varying slopes (c) varying intercepts and slopes in a beta regression model based on twenty different clusters. The $\beta$ parameters refer to the global mean of each parameter for all of the clusters combined ($\beta_0$ = intercept, $\beta_1$ = slope). $\sigma$ refer to the standard deviation of the cluster-specific coefficients from this mean ($\sigma_0$ = intercept, $\sigma_1$ = slope).

by other archaeologists. In short, new levels can be introduced into the model parameters by having a global parameter which represents the "average cluster" (also known as a fixed effect), with additional parameters for each individual cluster (also known as varying effects) representing their deviation from this average. Importantly, the model is able to share information across these two levels as the result of their hierarchical relationship (see McElreath, 2020, Chapters 13 and 14 for further details).

Overall, multilevel models benefit large-scale regional research of compositional data in the archaeological sciences as follows:

(a) They allow for the explicit modelling of variation across different sub-groupings, i.e. clusters, in the data — such as recovery locations, periods, regions, or stratigraphic layers.

(b) They simultaneously improve estimates of global parameters by accounting for sample interdependence.

(c) They account for the lack of representativeness in small sample sizes, by pooling estimates for clusters with fewer samples, according to (i) the actual sample sizes, (ii) the overall variability between clusters, (iii) and how far the observed, non-adjusted mean of each cluster is from the global mean across all clusters. The approach, therefore, allows us to directly incorporate uncertainty arising from uneven sample sizes into the statistical model (McElreath, 2020).

## 4. Part III: Muisca goldwork as a case study

To demonstrate the potential of applying these new methods to archaeometallurgical datasets in practice, we present the results of a hierarchical Bayesian variable dispersion beta regression model on metal compositions from the Eastern Cordillera of Colombia (AD 600-1600) (Fig. 6). Muisca (*muexcas/moxcas*) is the term attributed by the Spanish to a variety of Chibcha-speaking groups inhabiting this region in the 16th century (Simón, 1982). At the time of European arrival, these comprised numerous fragmented polities, with some organised under larger confederations and others being independent (Langebaek, 2000, 2019, 155-157). Their inhabitants practised the cultivation of maize, potatoes, and other taxa, alongside hunting, mining for salt and emeralds, and engaging in textile, ceramic, lithic, and metal craft production.

Metal use in Muisca societies was closely intertwined with social, political, and religious structures. Groups of votive gold-alloy figures (Fig. 7A), often combined with other materials, were deposited throughout time and space as part of a widely shared cultural practice of making religious offerings; access to these appears not to have been restricted to any one sector of society (Lleras Pérez, 1999; Langebaek, 2003; Uribe et al., 2013). In the 16th century, during the Early Colonial period, votive offerings were made at the end of festivities led by chiefs and spiritual leaders, which could last days and also served the purpose of redistributing goods (Langebaek, 1987, 50), in addition to coinciding with days of market exchange (Langebaek, 1987, 139). Metal adornments (Fig. 7B) were also often worn in such ceremonial contexts, although reportedly by individuals of important social and political status (Londoño Laverde, 1996). These adornments were often handed down from generation to generation (Langebaek, 2003, 264), and archaeologically, they have been found with mummified individuals deposited in sacred places, and sometimes buried in cemeteries. As such, they appear to contrast with votive figures in having a more limited consumer pool. This is regardless of how the individuals wearing them achieved their distinct social and political status, and of what exactly this entailed.

In terms of alloy composition, the vast majority of Muisca metalwork has gold, silver and copper in detectable levels, although their actual concentrations vary across a wide range, from native (argentiferous) gold with negligible copper content through more mixed *tumbaga* to almost pure copper. The silver in these alloys was an unintentional element naturally present as an impurity in the gold that was not manipulated for, bar the possible selection of native gold nuggets based on their colour. Ranges in-between 0–37 wt% for silver content have been found in compositional analyses of native Colombian gold (Uribe Villegas and Martinón-Torres, 2012a), and there is no evidence of parting technologies to refine the gold, or of silver extraction (Lleras Pérez, 2015, 106). As gold does not naturally occur in the region inhabited by the Muisca, this metal is thought to have been obtained through war or barter from the peoples inhabiting the surrounds of the Magdalena river (Lleras Pérez, 1999; Sáenz-Samper, 2021, 73) (located in the metallogenic provinces 3b and 3c in Fig. 6A), or from the Guane territory in the north (metallogenic provinces of 4a and 4b). Gold was exchanged for hallucinogens, textiles, salt, and emeralds. Once within the Muisca territory, it was brought to markets by intermediaries and exchanged, for instance, for coca leaves or food (Langebaek, 1987, 88-92). Copper, on the other hand, is readily available as mineral deposits in the Eastern Cordillera (Fig. 6B), although, with the potential exception of native copper occurrences, the actual smelting of metal from these deposits likely would have required more investment of technology, time, and labour in comparison to gold (Lleras Pérez, 1999, 73).

While the social contexts of metalworking and consumption undoubtedly changed throughout their millennia of pre-Hispanic use, radiocarbon dating has revealed some continuity of these practices

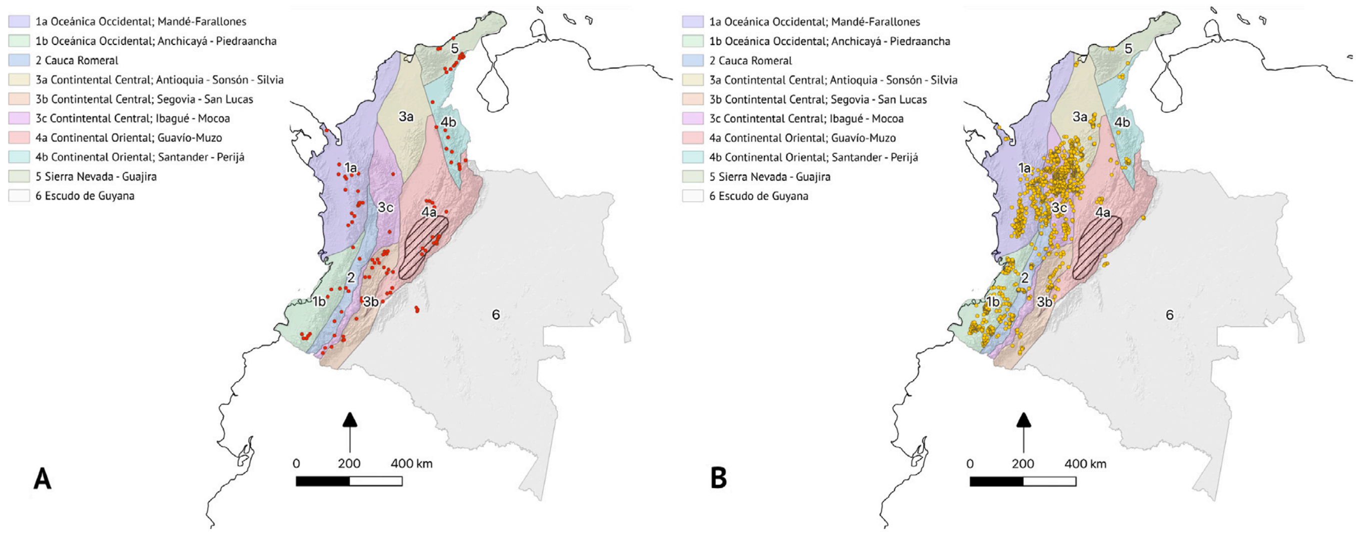


**Fig. 6.** The locations of (A) gold ore sources and (B) copper ore sources throughout present-day Colombia, with the location of the Muisca region highlighted in the Eastern Cordillera of Colombia. Metallogenic provinces from Arias Restrepo (2005b), originally taken from Mapa Metalogénico de Colombia escala 1:500.000 (INGEOMINAS, 2002); gold ore sources digitised from Arias Restrepo (2005c) (who ranked them for probability of pre-Hispanic exploitation on a scale of 1–4 (low probability; medium probability; high probability; very high probability); only those with medium probability or above are shown here, given that those ranked as low probability include, for instance, ore sources of which there is no evidence on the earth's surface, which need to be mined subterraneously, or which need chemical analysis to be detected, e.g., disseminated gold.), and copper ore sources from a map courtesy of the Museo del Oro, Banco de la República, based on Arias Restrepo (2005a), Arias Restrepo (2005b), INGEOMINAS (2002), Lobo Guerrero Arenas (2005). All of the digitisation of maps/data, as well as the drawing of maps itself, was done in QGIS v. 3.4.4 (QGIS Development Team, 2022) with the projection Magna-Sirgas EPSG 4686. The hillshades on the maps are based on STRM data produced by NASA. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

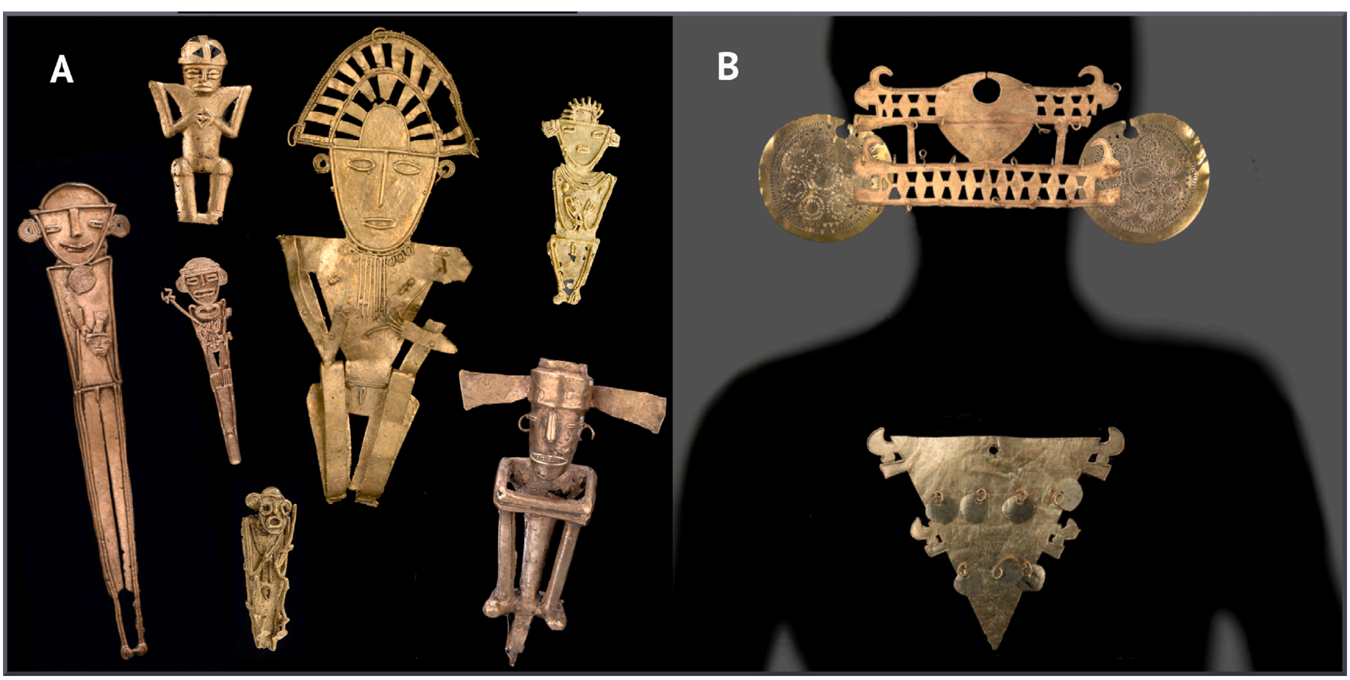


**Fig. 7.** (A) Selected examples of Muisca votive figures from different recovery locations. Votive offerings were typically deposited in sacred places such as water courses or mountain tops, and at sanctuaries near settlements (Plazas and Falchetti, 1985; Londoño Laverde, 1996, 57). The "coffee-bean" style eyes are characteristic of the region's votive metalwork as well as in anthropomorphic representations in ceramic figures or vessels. Shown approximately to scale, artefact height varies from c. 16 to 5.4 cm. (B) Selected Muisca adornments from different recovery locations shown as they may have been worn on an individual. Shown approximately to scale, artefact height varies from c. 15.4 to 8.5 cm. Compiled from individual images, image copyrights © Museo del Oro, Banco de la República.

(see Uribe Villegas and Martinón-Torres (2012b) for a list of dates associated with the metallurgy). Both classes of metalwork were adopted early on into the Muisca period, and the votive figures or *tunjos*, in particular, represent a unique and fairly internally coherent style throughout their history of use. In comparison, the adornments share more stylistic similarities with other metalworking traditions to the west and the north (Lleras Pérez, 1999; Plazas, 1998). Despite differences in how it was subsequently worked with, however, it is clear that the use of imported gold played an important role for both classes of metalwork. In 1587, the Spanish differentiated between two groups of goldworkers who respectively specialised in the production of votive figures and adornments (Rojas, 1965 in Langebaek, 1987, 102-103), which may or may not have corresponded to different approaches to metal procurement. It is difficult to tell, however, whether or not such differentiation existed before the conquest or if the religious persecution by the Spanish could have driven such divisions. This is given that votive offerings were considered idolatry by the catholic invaders (Uribe Villegas and Martinón-Torres, 2021, 163), whereas the use of adornments was not.

Previous research has also tentatively revealed that the votive figures that comprised a single offering were typically made at the same workshop, and even by the same artisans, using the same raw materials and invariably cast using the lost-wax technique. It also appears they were deposited soon after production, often still bearing dirt from the mould, casting imperfections and/or feeders, with no finishing work conducted on them, suggesting they were not meant for long-term social display (Falchetti, 1989; Lleras Pérez, 1999; Martinón-Torres and Uribe-Villegas, 2015a; Uribe Villegas and Martinón-Torres, 2012a; Uribe Villegas, 2012). The process of making the figurines, and particularly the use of wax, is also argued to have played a highly important symbolic role for the region's inhabitants, perhaps being part of the ritual itself (Martinón-Torres and Uribe-Villegas, 2015b). In addition, sometimes raw materials or intermediate products, such as gold nuggets, or melting ingots, were intentionally deposited in these offerings (Martinón-Torres and Uribe-Villegas, 2015a). The votive figures are typically fairly small in size (Falchetti, 1989, 4), although a number of larger figures do exist, such as the internationally famous Muisca raft (Uribe Villegas et al., 2021). While votive offerings were thus likely practised by a large proportion of the society, variations in the size of individual objects and composite offerings may suggest that some people could amass larger quantities of metal for making particular offerings intended to fulfil their specific needs to communicate with deities.

In contrast, adornments were often finished post-manufacture, with casting sprues and funnels removed (Falchetti, 1989). Occasional repairs and polished surfaces have also been detected (Lleras Pérez, 1999, 43; Martinón-Torres and Uribe-Villegas, 2015b, 382). Depletion gilding was sometimes used to enhance the golden appearance of the metal, in contrast to votive figures that only show evidence for enriched surface layers as the result of occasional exposure to fire, corrosion, or modern cleaning (Martinón-Torres and Uribe-Villegas, 2015a; Uribe Villegas, 2007; Vieri et al., 2020). In terms of size, the vast majority of the adornments are small, with, for instance, the standardised lost-wax cast beads that are highly typical of the region's ornamentation generally being less than three centimetres by maximum dimension (although many such individual beads would have comprised a whole necklace); object types other than beads typically being anywhere between 0–10 cm in size. A smaller number of larger adornments, such as the types of breastplates, ear and nose adornments shown in Fig. 7B are known.

Overall, little is still known about the range and types of gold sources exploited for either, or about how gold circulated and exchanged hands throughout the region. Here, we show how our new modelling tools allow us to consider the variability of gold sources employed in metal manufacture at different scales of analysis, in turn allowing us to consider the broader contexts of Muisca gold manufacture and consumption.

### 4.1. Data

The data used for modelling comprised both legacy data (n = 106) and new analytical data (n = 132), adding up to a total of 243 object compositions on both adornments and votive figures from 46 different municipalities of recovery (Table S3). The compositionally analysed objects include both archaeologically excavated ones, as well as those obtained by the Museo del Oro and other museums through legal purchases, which often lack precise site location data, but were often sold together with information on their geographical place of origin at the level of municipalities — see Ethics Statement, Supplementary Material 2.1. The majority of the legacy data has been previously published in Uribe Villegas and Martinón-Torres (2012b) and largely consists of chemical analyses conducted on items of archaeological metalwork by researchers and their collaborators at the Museo del Oro over the last decades,[2] with additional legacy data obtained from La Niece (1998) and Rovira (1994). Additional data, published more recently, was also included from Martinón-Torres and Uribe-Villegas (2015a). A full list of legacy data sources is provided in Supplementary Material S2.2.1. Only legacy data obtained by X-ray Fluorescence (XRF) or portable X-ray Fluorescence (pXRF) was included in the modelling to allow for better instrument comparability than if we were to include analyses by other analytical methods such as the Fire Assay (FA) analyses that are prevalent in earlier literature.

The other half of the dataset consists of new pXRF analyses collected in January-February of 2022 at the premises of the Museo del Oro in Bogotá. All of the analyses were carried out with a portable Niton™ XL3t Gold XRF Analyser, equipped with a gold X-ray tube with an excitation potential of 50 kV. The analyses were conducted at 50 kV, with an acquisition time of 60 s, using a bespoke calibration, designed and tested for the analysis of precious metal alloys in line with the research objectives of the Museo del Oro by Lina María Campos Quintero. An important benefit is that the analyses produced by this study will have been carried out with the same instrument and calibration as many of the more recent pXRF analyses undertaken at the museum and included from the legacy data, allowing for excellent cross-comparability between them. Full details of the sampling and analytical procedures are reported in Supplementary Material S2.2.2.

The compositional dataset relates to the major alloying constituents of the goldwork, given that the availability of trace element data for this region is scarce at the present moment (but see: Martinón-Torres and Uribe-Villegas, 2015a; Vieri et al., 2020). Nevertheless, the adoption of the new tools that we have presented here allows the identification of general trends in pre-Hispanic gold procurement practices, even if based on major alloying constituents. This is because, as noted above, the silver content of alloys was not artificially manipulated by the Muisca. As such, the proportion of silver in gold (Ag-in-Au) is a broad indicator of the geological composition of the native argentiferous gold deposits used, accounting for the dilution effect on silver introduced by any artificial copper additions. This ratio can be calculated as:

$$\text{Ag } in \text{ Au}(wt\%) = \frac{\text{Ag}(wt\%)}{\text{Ag}(wt\%) + \text{Au}(wt\%)} \tag{4}$$

Since gold nuggets or ingots from several different geological sources could be melted together, this proportion may also reflect contributions from the variety of different sources used. In short, Ag-in-Au contents are a broad indicator of either gold sourcing or the practices of mixing different geological sources of gold, therefore opening up interesting avenues for the discussion of metal procurement, movement, and processing throughout the pre-Hispanic Eastern Cordillera.

### 4.2. Methods: model building

Following exploratory data analysis (see Supplementary Material S2.3), the new methods were put into use in a beta regression model with the Ag-in-Au ratios as the compositional response variable. We modelled changes in these ratios as a function of both object type and volume. This allowed us to assess whether different practices of metal sourcing appear to have been adopted by their manufacturers, given the potential existence of different groups of metalsmiths involved in their production (Langebaek, 1987, 103). Volume, in turn, was included to assess how raw material procurement needs were met or changed when

[2] Museo del Oro is comprised of several branches, with its main premises in Bogotá, and with other regional museums including; Museo del Oro Quimbaya, Museo del Oro Nariño, Museo del Oro Calima, Museo del Oro Tairona, and Museo del Oro Zenú. All of the objects in these different branches have a consolidated inventory managed from the main branch in Bogotá.

manufacturing larger objects. Given that all the gold used by the Muisca had to be imported (Lleras Pérez, 1999; Sáenz-Samper, 2021, 73), working with higher quantities of metal could have potentially involved the recycling of other objects, combining gold from several different geological sources, or the sourcing of gold from specific sources with more abundant or voluminous occurrences of the native metal.

These predictors were introduced to both the linear model definition of the mean response $\mu$ and the model dispersion $\phi$, to respectively account for changes in both *Type 1* and *2* variability (cf. with Fig. 1 and the methods introduced in Part I of this paper). Table S5 summarises details of the chosen covariates. We then introduced a hierarchical level into both the intercept and slope parameters in the linear definition of $\mu$, to assess the regional heterogeneity or homogeneity of the gold procurement practices, as well as to account for any biases arising from varying sampling intensities for different artefact recovery locations in the dataset (cf. with discussion of *Type 3* variability in Part II and the corresponding examples in Fig. 1). Given the scarcity of higher-resolution spatial data, the basis for the hierarchical clusters was taken to be at the level of municipality centroids, which are modern administrative divisions within present-day Colombia, intermediate in size between localities and departments.

Supplementary Material 2.2.3 provides a detailed description of the steps involved in data pre-processing prior to the model being run. Against the background of navigating the bias–variance trade-off in modelling applications, Supplementary Material S2.5 also provides details of formal model comparison with simpler models where neither variable dispersion nor hierarchical structure in the data is accounted for. This demonstrated that the newly developed methods provided improved estimates compared to more traditional regression approaches.

In adopting a Bayesian framework to modelling, we aimed to use *weakly informative priors*, which allow for all possibilities that can be deemed plausible on the outcome scale, whilst being more cautious about extreme, unrealistic parameter values. The final mathematical model definition, including details of these priors and summaries, is provided in Supplementary Material 2.4.

#### 4.2.1. Software and model code

The model was fitted using *Stan* (Stan Development Team, 2021b), using the *RStan* package version 2.26.22 (Stan Development Team, 2021a) as an interface to communicate between *Stan* and *R*, with all post-sampling analyses and graphs conducted in *R* v. 4.3.1 (R Core Team, 2023), and using *RStudio* v. 2023.02.2 (RStudio Team, 2023). A number of other R packages were used during the post-processing stage, including *shinystan* (Gabry and Veen, 2022) and *xtable* (Dahl et al., 2019) to create tables of posterior summary statistics. The source codes corresponding to these model definitions are available on the GitHub repository, and different parts of the model codes took inspiration from Clark (2022), Hipson (2022), McElreath (2020), Stan Development Team (2021a). We ran four chains and 3000 iterations per chain, out of which 1000 each were dedicated to warmup, meaning a total of 8000 iterations after warmup across all four chains. Good convergence between the chains was reached ($Rhat < 1.01$), with bulk and tail effective sample sizes deemed sufficient for the resolution of this study (Bulk ESS and Tail ESS min. $> 200 \times$ no. of chains) per parameter. Test runs with more iterations were found not to change the variance of the model estimates to a notable degree.

### 4.3. Results

The model outputs do not show clear evidence of a correlation between the average expected silver-in-gold ratios and object volume, nor do they suggest a differentiation between adornments and votive figures ($\beta_2$ and $\beta_3$, Table 1; Fig. 8A). This suggests either that (A) the sources of gold were fairly similar for different-sized objects of both classes on average, or that (B) the sample size is not sufficient to tell apart any differences in the overall tendencies. Hence, the structure in the dataset cannot presently be explained by *Type 1* variability (cf. Fig. 1).

The 95% Posterior Interval (PI) for the linear submodel for $\phi$, however, does not overlap with 0, showing a clear signal for reduced dispersion (*Type 2* variability) of the Ag-in-Au ratios as a function of object volume ($\gamma_2$, Table 1, Fig. 8B). Increased dispersion of votive figures at baseline volume when compared to adornments was also observed ($\gamma_3$, Table 1).

Importantly, the predicted ranges of Ag-in-Au ratios further show notable variations between municipalities, particularly in the case of the votive figures. This is evidenced by the posteriors of the parameters $\sigma_{m1}$ and $\sigma_{m3}$, in Table 1, which capture *Type 2* variability across different recovery locations for the adornments (the baseline category) and for the votive figures, respectively. While the prediction intervals remain wide for many of the individual recovery locations, the votive figures deviate towards more unusual values more, ranging from c. 11% at its lowest, to c. 28wt% at its highest, on average (Fig. 9).

In contrast, there is more consistency in the compositional ratios of adornments across the different municipalities.[3] The municipality of Colombia deviates the most from the typical Ag-in-Au ratios of ~16wt%.[4] Other than this, the predicted Ag-in-Au ratios of adornments consistently take on values in between the whole range of variability observed in votive figures, although there are still some fluctuations across a more restricted range of compositions.

Artefact volume, in contrast, was found to have a limited impact on compositional variation across municipalities ($\sigma_{m2}$ in Table 1), suggesting that the patterns detected globally also apply at the local level, with object size thus being a poor predictor of the average Ag-in-Au ratio both in terms of *Type 1* as well as *Type 3* variability.

#### 4.3.1. Muisca gold and internal exchange networks

As shown by both the globally predicted Ag-in-Au ratios (Fig. 8A), and by the expected compositions at the municipal level (Fig. 9), it seems likely that similar sources of gold were often exploited for both object classes. The high levels of compositional variability observed for votive figures across municipalities (Fig. 9B), in turn, can be used to suggest that gold from several different geological sources were used throughout the region, which highlights the extent of Muisca exchange networks. This also suggests that, in some cases at least, different geological deposits of gold were not mixed, preserving the more extreme compositional signatures, and potentially pointing to the sourcing of fresh rather than mixed or recycled gold in such cases. The signature observed for overall increased variability of votive figures on the global scale ($\gamma_3$ in Table 1) further provides support to this hypothesis, which should be verified with further data collection and analyses. The geological sources that are likely to have contributed to this variability include those to the Magdalena River valley to the west, and potentially towards the other peripheries of the Muisca region (Fig. 6). In the future, geological compositional data will be able to aid in understanding which of these sources are likely to have been exploited in the past.

Having said that, many of the recovery locations for votive figures also showed compositions that fall more towards the regional average (Fig. 9). While this may simply reflect that more gold sources are likely to fall in that compositional range, it could suggest the dilution of

[3] This applies even when considering the smaller sample size of the votive figures when compared to the adornments (Table S3). While in a non-hierarchical model, smaller sample sizes are often associated with more variability, the hierarchical modelling approach adopted here has allowed for accounting for sample imbalance as a source of uncertainty through the process of adaptive shrinking, as per the discussion in Part II in this paper.

[4] This municipality is represented by objects pertaining to the Complex Western Style rather than the core Muisca style. This makes them unusual on the global scale by their contextual and stylistic traits.

**Table 1**
Posterior summaries of the parameters returned by the hierarchical variable dispersion beta regression used for modelling Muisca goldwork Ag-in-Au ratios, including mean estimates (*mean*), standard deviations (*sd*), the 95% prediction intervals, the number of effective samples (*n_eff*), as well as *Rhat* values reflecting the convergence of the chains). The $\beta$ parameters capture changes in the average silver-in-gold ratios (*Type 1* variability); the $\gamma$ parameters variation in the compositional dispersions (*Type 2* variability); and the $\sigma_m$ parameters capture between-cluster variance (*Type 3* variability), i.e., in this case, how much variation can be explained across different municipalities of recovery as a function of each of the predictors.

| Predictor | Parameter | mean | sd | 2.5% | 25% | 50% | 75% | 97.5% | n_eff | Rhat |
|---|---|---|---|---|---|---|---|---|---|---|
| Intercept *(adornment)* | $\beta_1$ | −0.42 | 0.06 | −0.54 | −0.46 | −0.42 | −0.38 | −0.30 | 3498 | 1.00 |
| Volume (*per 1.24* cm$^3$) | $\beta_2$ | −0.07 | 0.06 | −0.19 | −0.11 | −0.07 | −0.03 | 0.04 | 3378 | 1.00 |
| Object type *(votive)* | $\beta_3$ | 0.18 | 0.13 | −0.06 | 0.10 | 0.18 | 0.26 | 0.43 | 5060 | 1.00 |
| Intercept *(adornment)* | $\gamma_1$ | 3.27 | 0.13 | 3.02 | 3.19 | 3.27 | 3.36 | 3.51 | 3497 | 1.00 |
| Volume (*per 1.24* cm$^3$) | $\gamma_2$ | 0.35 | 0.14 | 0.07 | 0.26 | 0.35 | 0.45 | 0.63 | 4413 | 1.00 |
| Object type *(votive)* | $\gamma_3$ | −0.64 | 0.26 | −1.15 | −0.82 | −0.63 | −0.46 | −0.14 | 4054 | 1.00 |
| Intercept *(adornment)* | $\sigma_{mun1}$ | 0.22 | 0.07 | 0.09 | 0.17 | 0.22 | 0.26 | 0.36 | 1684 | 1.00 |
| Volume (*per 1.24* cm$^3$) | $\sigma_{mun2}$ | 0.22 | 0.09 | 0.04 | 0.16 | 0.21 | 0.28 | 0.42 | 1071 | 1.00 |
| Object type *(votive)* | $\sigma_{mun3}$ | 0.47 | 0.17 | 0.14 | 0.36 | 0.47 | 0.58 | 0.83 | 1147 | 1.00 |

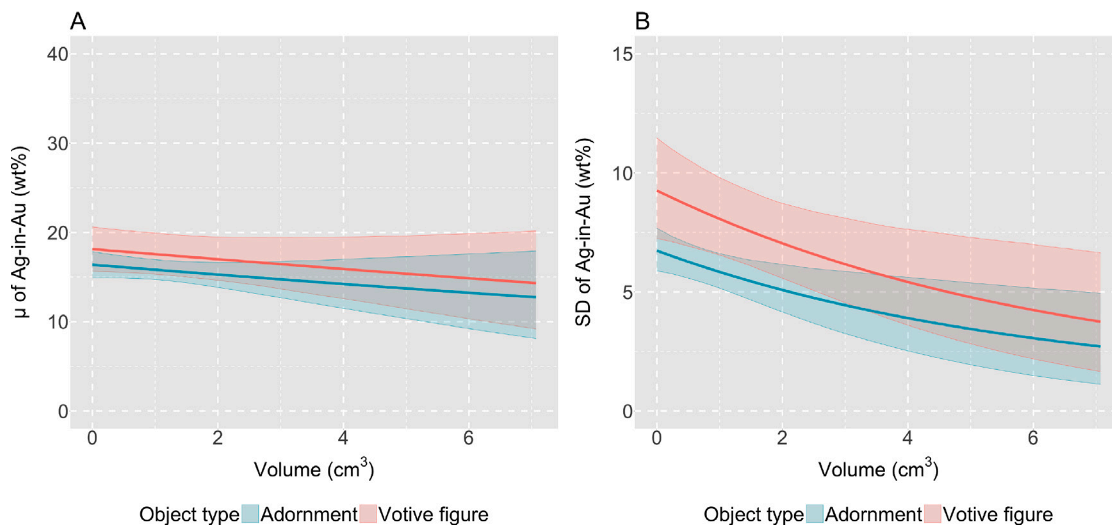


**Fig. 8.** (A) Posterior predictions for the average Ag-in-Au ratios ($\mu$) at different object volumes, with the green line referring to the mean predictions for adornments, and the red line to the mean predictions for votive figures. Corresponding 95% PI intervals are also shown as shaded. (B) Posterior predictions for the standard deviations of the Ag-in-Au ratios (*SD*) at different object volumes, with the green line referring to the mean predictions for adornments, and the red line to the mean predictions for votive figures. Corresponding 95% PI intervals are also shown as shaded. We report the standard deviations (calculated based on the model variance (Eq. (2)), with both $\phi$ and $\mu$ indexed according to each sample $i$ as in Eq. (3)), instead of the changes in the dispersion parameter directly, to allow for a more intuitive interpretation of the results on the compositional scale. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

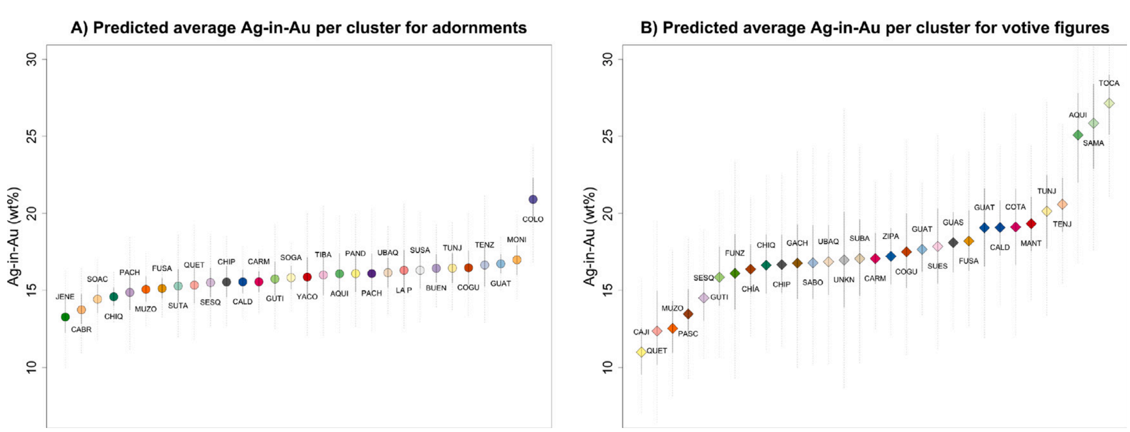


**Fig. 9.** Predicted Ag-in-Au ratios for different municipalities for (A) adornments and (B) votive figures. The colours reflect the municipality of recovery, and the $y$-axis shows the Ag-in-Au predictions as proportions. The dark lines represent the 50% PI intervals, with the dashed lighter lines corresponding to 95% intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

extreme Ag-in-Au contents through mixing, in some cases at least. Such dilution can be expected to reflect either (a) the combination of raw gold from different sources to make a conglomerate batch, or (b) the mixing through recycling of other objects. In adornment manufacture, while some municipal variations were detected in the gold sourcing or mixing practices (Fig. 9), these are typically less extreme compared to votive figures, with the Ag-in-Au ratios of adornments consistently taking on values in a narrower bracket contained within the whole range of variability observed in votive figures (except for Colombia — see footnote 4). As such, either the range of gold sources employed for adornments was more constrained, or the different sources were subsequently mixed more often, resulting in the dilution of the more extreme compositional signatures. Although these differences could also be explained by temporal differences in the raw material procurement practices, rather than differences across the two object classes, contextual evidence suggests that votive figures, at least, represent a fairly unique and consistent tradition throughout their history of use.

Finally, the proposed practices of gold mixing were found to be even more extensive for the larger objects represented by the dataset on the global scale. This is given that the compositions of larger objects *both* concentrate around similar mean predictions as for smaller-volume objects (Fig. 8A), *and* that there is significantly reduced compositional dispersion of Muisca goldwork towards higher volumes on the global scale (Fig. 8B). The manufacture of larger objects would have required more raw material in comparison to most of the objects, which are typically less than 1 $cm^3$ in size (~71% for votive figures, and ~73% for adornments, in this dataset).

For the adornments, these dilutional patterns may be explained by recycling. The adornments did not play the same ritual and religious role as the votive ones. They also appear to also have had longer life histories, with evidence for finishing post-manufacture and repairs (Lleras Pérez, 1999, 43; Martinón-Torres and Uribe-Villegas, 2015b, 382), as well as for generational gifting (Langebaek, 2003, 264).

However, recycling is less likely to have been less socially permissible in the case of the votive figures. First, these were deposited soon after production in sacred places such as water courses of mountain tops (Plazas and Falchetti, 1985, 57), or at sanctuaries (Londoño Laverde, 1996), where they are fairly unlikely to have been disturbed by the recollection of the metal. Additionally, raw materials and processes of transformation have been argued to be ritually important in votive figure manufacture (Martinón-Torres and Uribe-Villegas, 2015a,b), which again makes the use of recycled gold less likely. Rather, the fact that the Ag-in-Au for larger objects tends to converge around the Muisca average could imply that gold from different geological sources was brought to the goldworkers at the time the offering was made, or obtained as part of a ritual economy network, and their mixing during manufacture may have been part of the ritual itself.

The more sizeable Muisca goldwork offerings were likely commissioned and mediated by chiefs, religious specialists or other individuals who were able to amass larger quantities of metal in order to communicate with the deities on behalf of the community; their deposition may have taken place as part of larger festivities attended by a multitude of people, sometimes including chiefs or others from different parts of the region. This ritual mobility and exchange of gold could have, subsequently, resulted in even more extensive dilution of the whole range of compositional sources available throughout the region towards the middle.

Interestingly, if we now revisit how gold was introduced into the Muisca markets, Langebaek (1987, 151; 2019) has argued that subsistence goods played a limited role in market exchange and that economic activities were highly intertwined with religious and political ones. Marketplaces were used to host festivities that could last for days, during which leaders redistributed goods to their community. Markets are also argued to have served the role of "ethnic integration", as they could be attended by people from different polities. It is then conceivable that different sources of gold were sometimes brought into these communal gatherings, because the people who attended the festivities came from polities that had access to different exchange networks, bringing the gold with them. These different sources were subsequently used in making the offerings that were to be deposited at the end of the festivities.

The above scenario is supported by the evidence from Tocancipá - one of the few votive goldwork offerings for which trace element data are available. Here, all the objects are small, but they represent a range of types: natural nuggets, small ingots, and anthropomorphic figurines. The elemental analyses show distinct groupings that are consistent with a variety of sources or metal batches (Martinón-Torres and Uribe-Villegas, 2015a), in what might represent a concrete demonstration of the broader pattern inferred from the data modelling approach.

## 5. Discussion

We have shown that statistical models based on the beta distribution provide more robust and accurate means to analyse compositional data in archaeology. In contrast to the more established reliance on log-transformations, our approach restricts predictions to the standard unit interval (0, 1), allowing at the same time to account or event model heteroscedasticity (Ferrari and Cribari-Neto, 2004). While log-transformation approaches may be still viable where the data follow a log-normal or normal distribution — mainly likely to occur when the metal in the assemblage was obtained from an individual geological source (Baxter, 1995, 515) —, it is frequently an insufficient approach to dealing with the challenges introduced by the compositional constraint. In particular, within archaeometallurgy, artificially manipulated alloying constituents often have highly skewed profiles. Similarly, while working with Gaussian models on non-transformed data can be successful where the element of concern displays limited variance relative to its central tendency (therefore not even necessitating log-transformations), this often does not apply to manipulated constituents that can take on values over a wide range.

Indeed, as was shown here, traditional models based on the Gaussian distribution both on non-transformed and log-transformed data risk yielding predictions that exceed either both (e.g., 0wt% and 100wt%), or the upper compositional constraint, respectively. Moreover, even where log-transformation successfully alleviates skewness in the data, it was shown to incorrectly assume that the model variance is constant on the log-transformed scale. This can result in highly biased estimates of model variance when back-transformed to the compositional scale (Fig. 3). It can similarly be insufficient when working with compositional signatures that reflect unintentional impurities where, e.g., the recycling of metals with highly different impurity profiles has occurred at varying rates. The validity of the log-transformation approach needs to be assessed on a case-by-casis (Baxter, 1995, 515). The beta regression approach, in turn, was shown to readily apply to both cases of skewed and non-skewed compositional data, given it accommodates a multitude of distributional shapes (Fig. 2).

We thus propose that the beta regression approach should be adopted as the standard approach to working with univariate compositional responses within the field, such as in our case study of Muisca Ag-in-Au ratios. Examples of future applications in craft production studies beyond archaeometallurgy could include understanding the use of a specific colourant in vitreous materials, or the study of processes involving the manipulation of clay recipes in ceramic technological studies, with a focus on a single compound that is likely to mainly represent contributions from a particular temper, or lack thereof. The approach is similarly applicable to other sub-disciplines of the archaeological sciences making use of compositional data, such as stable isotope analyses.

In many archaeological applications, of course, it is more accurate to consider multiple compositional constituents simultaneously. For instance, the use of a single raw material may need to be assessed in relation to the importance of more than one other one. The same applies to provenance studies concerned with trace element data, which are typically informative only when concerned with more than one element. While others have argued for the use of log-ratios (Aitchison, 1982), the challenges arising from such transformations, e.g., incorrect estimates of the model variance, also apply to such multivariate applications. They further result in the loss of the original data scale, which cannot be readily reconstructed, unlike in the case of the univariate log-transformation. We propose that developing multivariate Dirichlet-distributed responses – of which, in fact, the beta distribution is a special univariate case of – is more appropriate (Maier, 2014). Their adoption would allow for discussions of the underlying drivers of different choices made by craftproducers, where the impact of such drivers is simultaneously assessed against more than two compositional constituents. Such applications could have immense potential, e.g., in the systematic examination of different impurity profiles in different archaeological assemblages, as part of more in-depth procurement studies.

We have further added to the toolkit of the data analyst within the archaeological sciences, not only in terms of providing a statistically more robust solution to working with compositional data, but also by providing two new ways of thinking about and modelling compositional variability in craft production studies. One of these (variable dispersion submodels) allows us to consider how broader social contexts operate to result in more or less constrained craft production practices at the assemblage level (*Type 2* variability in Fig. 1). The other (multilevel modelling) accounts for the local nuance of human practices and the uncertainty arising from varying sampling intensities in large-scale craft production studies (*Type 3* variability). Although not explored in this paper, by drawing upon both of the approaches, researchers could additionally model how compositional dispersions vary at the local and supra-local levels concurrently, therefore shedding light on processes of standardisation at different scales of analysis (*Type 4* variability). Future researchers can choose to adopt each of the newly introduced approaches independently or combine them in the same model as appropriate for their research.

Both hypothetical examples and the Muisca case study were used to demonstrate how modelling these sources of variability can be readily incorporated into the beta regression framework. For instance, the significantly reduced dispersion of Muisca goldwork Ag-in-Au ratios towards higher object volumes (Fig. 8B), combined with the inferences regarding relatively consistent object compositions on the global scale (Fig. 8A), allowed us to postulate that these reflect dilutional patterns. We suggested this could result from the combination of multiple sources of gold in votive figure manufacture and plausibly either that or the recycling of materials in the case of the adornments. In contrast, had we focused on compositional averages alone, the results would have implied a false sense of continuity in raw material procurement practices across all object volumes, in addition to providing biased estimates of the other model parameters. From a theoretical point of view, it would naturally still be preferable to access information on the actual data-generating processes contributing to the variation in the compositional averages alone — e.g., within the context of this case study, which geological sources were exploited throughout the Muisca region in the first place, which of the constituent sources were subsequently mixed, for objects of which size, and at which points through time. This would allow us to eliminate or minimise the compositional dispersions in the model outputs. In the inevitable absence of such information within archaeology, however, the variable dispersion approach has been shown to provide a tool that indirectly provides cues on what can possibly explain the structure of the variability in the observed data.

The general principle of directly modelling variance as an archaeological source of information is also similarly applicable to multivariate Dirichlet-distributed responses of compositional data, and to non-compositional data, e.g. in linear models making use of the Gaussian distribution or in other classes of GLMs. For instance, it could be used to test whether the intra-assemblage variability of projectile point metrics varies over time, relating to, e.g., different levels of prestige-based or conformist cultural transmission, as in the case studies explored through CVs by Eerkens and Lipo (2005).

Notably, had we further ignored shared influences across different recovery locations, we would not have been able to infer that Muisca gold procurement practices varied more between different recovery locations for the object class of votive figures than they did for adornments (Figs. 9). We would have thus not been able to hypothesise that gold circulated from several different sources to different parts of the Muisca region and were possibly sometimes employed in votive figure manufacture in its raw geological form; and to hypothesise that, in the case of the votive figures, the diluted metal employed in larger object manufacture (as detected based on changes in the compositional dispersions discussed above) was likely mixed shortly before deposition. The multilevel modelling was key for providing support to this hypothesis, which is also backed by the contextual and historical evidence on their transient nature of production (Falchetti, 1989; Lleras Pérez, 1999; Martinón-Torres and Uribe-Villegas, 2015a). While a non-hierarchical variable dispersion model would have still suggested higher dispersion of votive figure compositions compared to adornments, it would have had no means of differentiating whether such variability arose from people consistently using varied sources of gold for votive figure manufacture when depositing offerings, or whether specific locations had different tendencies in this regard.

Ultimately, therefore, multilevel modelling explicitly allowed us to infer how Muisca metal procurement practices operated on local and supra-local scales concurrently, providing grounds to potentially highlight local nuance in metalworking practices. Simultaneously, the bias introduced by sampling imbalance was explicitly addressed through the process of partial pooling, which assessed the probability that any extreme observations at poorly sampled locations were so unusual as to warrant us to be cautious of the observed signatures, resulting in less extreme compositional predictions than those empirically observed.

In the future, the beta regression toolkit could further be improved with, for instance, methods that further account for sample interdependence across dimensions such as time and space. This is given that the processes that lead to the formation of the archaeological record also operate spatially and temporally, i.e. they are likely to share more similarities when closer in time or space to one another. In the case of the spatial dimension, we propose this could be achieved by adopting either *Intrinsic Auto-Regressive Models for Areal Data* (ICAR) (Besag, 1974; Besag and Kooperberg, 1995) model specifications, in particular when working with polygon data, or by Gaussian Process models (Rasmussen and Williams, 2005), when working with spatial data at the scale of localities. The latter would also be readily applicable to modelling continuous sample interdependence through time.

## 6. Conclusion

Beta regression should be adopted as the new standard approach to treating compositional data within the field of the archaeological sciences. It provides probabilistic insights into the drivers of the four main different types of variability expected in compositional archaeological datasets, at different scales of analysis, ranging from the local to the global. Compared to established approaches such as log-transformations, it simultaneously provides more statistically robust inferences. Modelling variable dispersion, i.e., modelling the amount of unexplained variation in our datasets as a function of covariates, was further shown to provide unprecedented insights into the broader social contexts of metallurgical production. Introducing hierarchical levels into the beta regression framework, in turn, has been demonstrated to re-introduce local variation into regional-scale reconstructions, whilst accounting for the inherent sampling uncertainties that are present within the archaeological record.

Importantly, the application of these two novel approaches has further highlighted that the chemical compositions of archaeological artefacts do, in fact, have more to offer than only the desired performance characteristics targeted by metalsmiths. By solely focusing compositional data analysis either on provenance or on what constitutes desirable performance characteristics, archaeological craft production studies fail to account for a wealth of information on how broader societal processes shape aspects of past craft production activities. In particular, Muisca metal procurement practices had little to do with individual decision-making processes on optimising the performance characteristics of their metalwork. Neither did our most important inferences regarding the social contexts of metallurgical production arise from pinpointing the exact locations of the geological ore sources employed in metal manufacture. Rather, the varying levels of dispersion witnessed in metal compositions can be a source of information in itself, regardless of which exact environmental, technological, or cultural factors drove the decision-making of craftspeople in the first place.

## CRediT authorship contribution statement

**Jasmine Vieri:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project management, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Enrico R. Crema:** Conceptualization, Methodology, Resources, Supervision, Validation, Writing – review & editing. **María Alicia Uribe Villegas:** Conceptualization, Resources, Writing – review & editing. **Juanita Sáenz Samper:** Conceptualization, Resources, Writing – review & editing. **Marcos Martinón-Torres:** Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Validation, Writing – review & editing.

## Declaration of competing interest

The authors declare no competing interests.

## Acknowledgements

Primary funding was obtained from the Arts and Humanities Research Council UK (AHRC), which funded the lead author's Cambridge AHRC-Doctoral Training Partnership Scholarship (2112128). Additional funding was obtained from the Osk. Huttunen Foundation and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101021480, REVERSEACTION project). In addition, we would like to thank St John's College, University of Cambridge, and the Department of Archaeology, University of Cambridge, for conference funding to disseminate the results of this research. We would like to thank Lina María Campos Quintero for supporting the analytical pXRF work undertaken at the Museo del Oro, for kindly sharing her calibration procedures, and for insightful discussions on the archaeological case study; Clark M. Rodríguez for the object photographs; as well as Jessica Pérez Fonseca and Orlando Castillo Vaca for further facilitating the analytical work. We are grateful to Dr. Elizabeth DeMarrais and Prof. Andrew Bevan for their comments and suggestions that helped improve this work. Finally, we thank the anonymous reviewers and the journal editor for their comments on this paper.

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jas.2024.106106.

## Data availability

All data and code used in this article are publicly available in a dedicated GitHub repository (https://github.com/jmkvieri/BBLoP) and archived in Zenodo (https://doi.org/10.5281/zenodo.13600304), with the Supplementary Material providing further details of relevant metadata and the final model code.

## References

Aitchison, J., 1982. The statistical analysis of compositional data. J. R. Stat. Soc. Ser. B Stat. Methodol. 40 (2), 139–177.

Aitchison, J., Barceló-Vidal, C., Pawlowsky-Glahn, V., 2002. Some comments on compositional data analysis in archaeometry, in particular the fallacies in Tangri and Wright's dismissal of logratio analysis. Archaeometry 44 (2), 295–304. http://dx.doi.org/10.1111/1475-4754.t01-1-00061.

Arias Restrepo, J.A., 2005a. Caracterización del nivel tecnológico minero precolombino y zonas con posibilidad de explotación minera precolombina. Memoria Explicativa. Subdirección Técnica Museo del Oro, Banco de la República, Bogotá, D.C.

Arias Restrepo, J.A., 2005b. Mapas compilativos de manifestaciones y depósitos de cobre, oro, plata y platino en Colombia escala 1:500.000. Memoria Explicativa. Subdirección Técnica Museo del Oro, Banco de la República, Bogotá, D.C.

Arias Restrepo, J.A., 2005c. Posibilidades tecnológicas de explotación precolombina de cobre, oro, plata y platino. Memoria Explicativa. Subdirección Técnica Museo del Oro, Banco de la República, Bogotá, D.C.

Banks, W.E., Bertran, P., Ducasse, S., Klaric, L., Lanos, P., Renard, C., Mesa, M., 2019. An application of hierarchical Bayesian modeling to better constrain the chronologies of Upper Paleolithic archaeological cultures in France between ca. 32,000–21,000 calibrated years before present. Quat. Sci. Rev. 220, 188–214. http://dx.doi.org/10.1016/j.quascirev.2019.07.025.

Baxter, M.J., 1995. Standardization and transformation in principal component analysis, with applications to archaeometry. Appl. Stat. 44 (4), 513. http://dx.doi.org/10.2307/2986142.

Baxter, M.J., Freestone, I.C., 2006. Log-ratio compositional data analysis in archaeometry. Archaeometry 48 (3), 511–531.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. J. R. Stat. Soc. Ser. B Stat. Methodol. 36 (2), 192–236, URL https://www.jstor.org/stable/2984812.

Besag, J., Kooperberg, C., 1995. On conditional and intrinsic autoregression. Biometrika 82 (4), 733–746. http://dx.doi.org/10.2307/2337341, URL https://www.jstor.org/stable/2337341.

Bray, P., Cuénod, A., Gosden, C., Hommel, P., Liu, R., Pollard, A., 2015. Form and flow: The 'Karmic cycle' of copper. J. Archaeol. Sci. 56, 202–209. http://dx.doi.org/10.1016/j.jas.2014.12.013.

Bray, P., Pollard, A., 2012. A new interpretative approach to the chemistry of copper-alloy objects: Source, recycling and technology. Antiquity 86 (333), 853–867. http://dx.doi.org/10.1017/S0003598X00047967.

Clark, M., 2022. Bayesian Beta Regression | Model Estimation by Example. URL https://m-clark.github.io/models-by-example/.

Costin, C.L., 1991. Craft specialization: Issues in defining, documenting, and explaining the organization of production. Archaeol. Method Theory 3, 1–56, URL https://www.jstor.org/stable/20170212.

Crema, E.R., Bloxam, A., Stevens, C.J., Vander Linden, M., 2024. Modelling diffusion of innovation curves using radiocarbon data. J. Archaeol. Sci. 165, 105962. http://dx.doi.org/10.1016/j.jas.2024.105962.

Cribari-Neto, F., Zeileis, A., 2010. Beta regression in R. J. Stat. Softw. 34, 1–24. http://dx.doi.org/10.18637/jss.v034.i02.

Dahl, D.B., Scott, D., Roosen, C., Magnusson, A., Swinton, J., 2019. Xtable: Export Tables to LaTeX or HTML (v. 1.8-4). URL https://CRAN.R-project.org/package=xtable.

DeMarrais, E., 2013. Understanding heterarchy: Crafting and social projects in pre-Hispanic northwest Argentina. Camb. Archaeol. J. 23 (03), 345–362. http://dx.doi.org/10.1017/S0959774313000474.

Eerkens, J.W., Lipo, C.P., 2005. Cultural transmission, copying errors, and the generation of variation in material culture and the archaeological record. J. Anthropol. Archaeol. 24 (4), 316–334. http://dx.doi.org/10.1016/j.jaa.2005.08.001.

Falchetti, A.M., 1989. Orfebrería prehispánica en el altiplano central colombiano. Bol. Museo Oro 25, 2–41.

Fernée, C.L., Trimmis, K.P., 2021. Detecting variability: A study on the application of Bayesian multilevel modelling to archaeological data. Evidence from the Neolithic Adriatic and the Bronze Age Aegean. J. Archaeol. Sci. 128, 105346.

Ferrari, S., Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. J. Appl. Stat. 31 (7), 799–815.

Gabry, J., Veen, D., 2022. Shinystan: Interactive Visual and Numerical Diagnostics and Posterior Analysis for Bayesian Models (v. 2.6.0). URL https://CRAN.R-project.org/package=shinystan.

Goldstein, H., 1987. Multilevel Models in Education and Social Research. Oxford University Press, Oxford.

Hipson, W., 2022. Bayesian Varying Effects Models in R and Stan, Will Hipson, URL https://willhipson.netlify.app/post/stan-random-slopes/varying_effects_stan/ (accessed 9.12.2022).

Humphris, J., Martinón-Torres, M., Rehren, T., Reid, A., 2009. Variability in single smelting episodes – a pilot study using iron slag from Uganda. J. Archaeol. Sci. 36 (2), 359–369. http://dx.doi.org/10.1016/j.jas.2008.09.020.

INGEOMINAS, 2002. Atlas Colombiano de Información Geológico Minera Para La Inversión. ACIGEMI. (Versión Digital).

La Niece, S., 1998. Metallurgical case studies from the British Museum's collections of pre-Hispanic gold. Bol. Museo Oro 44–45, 138–157.

Langebaek, C.H., 1987. Mercados, poblamiento e integración étnica entre los muiscas: siglo XVI. Banco de la República.

Langebaek, C.H., 2000. Arqueología regional en el Valle de Leiva: Procesos de ocupación humana en una región de los Andes orientales de Colombia. Number 2 in Informes Arqueológicos del Instituto Colombiano de Antropología e Historia. Instituto Colombiano de Antropología e Historia.

Langebaek, C.H., 2003. The political economy of pre-Columbian gold work: Four examples from northern South America. In: Gold and Power in Ancient Costa Rica, Panama, and Colombia. Dumbarton Oaks, Washington, D.C, pp. 245–278.

Langebaek, C.H., 2019. Los muiscas: la historia milenaria de un pueblo chibcha. Penguin Random House Grupo Editorial, Bogotá, D.C.

Lechtman, H., 1977. Style in technology - some early thoughts. In: Lechtman, H., Merrill, R.S. (Eds.), Material Culture: Styles, Organization, and Dynamics. American Ethnological Society, St Paul, pp. 3–20.

Lleras Pérez, R., 1999. Prehispanic Metallurgy and Votive Offerings in the Eastern Cordillera Colombia. Number 778 in BAR International Series. Archaeo Press, Oxford.

Lleras Pérez, R., 2015. La metalurgia prehispánica en el norte de Suramérica: una visión de conjunto. In: Lleras Pérez, R. (Ed.), Metalurgia en la América antigua: Teoría, arqueología, simbología y tecnología de los metales prehispánicos. Institut français d'études andines, Lima, pp. 94–113, URL http://books.openedition.org/ifea/5853.

Lobo Guerrero Arenas, J., 2005. Fuentes escritas para una historia de la minería prehispánica en Colombia. Manuscrito. Subdirección Técnica, Museo del Oro, Banco de la Republica, Bogotá, D.C.

Londoño Laverde, E., 1996. El lugar de la religión en la organización social muisca. Bol. Museo Oro 40, 63–87.

López-García, P., Argote-Espino, D., Fačevicová, K., 2018. Statistical processing of compositional data. The case of ceramic samples from the archaeological site of Xalasco, Tlaxcala, Mexico. J. Archaeol. Sci. Rep. 19, 100–114. http://dx.doi.org/10.1016/j.jasrep.2018.02.023.

Maier, M.J., 2014. DirichletReg: Dirichlet Regression for Compositional Data in R. Technical Report 125, Institute for Statistics and Mathematics, Wirtschafts Universität Wien, Vienna, pp. 1–26.

Martinón-Torres, M., Uribe-Villegas, M.A., 2015a. The prehistoric individual, connoisseurship and archaeological science: The Muisca goldwork of Colombia. J. Archaeol. Sci. 63, 136–155. http://dx.doi.org/10.1016/j.jas.2015.08.014.

Martinón-Torres, M., Uribe-Villegas, M.A., 2015b. Technology and culture in the invention of lost-wax casting in South America: An archaeometric and ethnoarchaeological perspective. Camb. Archaeol. J. 25 (1), 377–390. http://dx.doi.org/10.1017/S0959774314001164.

McElreath, R., 2020. Statistical Rethinking: A Bayesian Course with Examples in R and Stan, second ed. Chapman and Hall/CRC.

Pawlowsky-Glahn, V., Egozcue, J.J., 2006. Compositional data and their analysis: An introduction. Geol. Soc. Lond. Special Publ. 264 (1), 1–10. http://dx.doi.org/10.1144/GSL.SP.2006.264.01.01.

Perri, A.R., Koster, J.M., Otárola-Castillo, E., Burns, J.L., Cooper, C.G., 2019. Dietary variation among indigenous Nicaraguan horticulturalists and their dogs: An ethnoarchaeological application of the Canine Surrogacy Approach. J. Anthropol. Archaeol. 55, 101066. http://dx.doi.org/10.1016/j.jaa.2019.05.002.

Perucchetti, L., Montero-Ruiz, I., Bray, P., 2020. Mapping archaeometallurgical data of the Iberian Copper Age: Different ways to look at a big picture. J. Archaeol. Sci. 119, 105165.

Plazas, C., 1998. Cronología de la metalurgia colombiana. Bol. Museo Oro 44–45, 3–77, URL https://publicaciones.banrepcultural.org/index.php/bmo/article/view/6894.

Plazas, C., Falchetti, A.M., 1985. Cultural patterns in the prehispanic goldwork of Colombia. In: Benson, E.P. (Ed.), The Art of Precolumbian Gold: The Jan Mitchell Collection, first U.S. ed In: A New York Graphic Society Book, Little, Brown and Co., Boston, pp. 47–59.

Pollard, A.M., Bray, P., 2014. Chemical and isotopic studies of ancient metals. In: Roberts, B.W., Thornton, C.P. (Eds.), Archaeometallurgy in Global Perspective. Springer, pp. 217–238.

Pollard, A.M., Liu, R., Rawson, J., Tang, X., 2018. From alloy composition to practice: Chinese bronzes. Archaeometry http://dx.doi.org/10.1111/arcm.12415.

Pryce, T.O., Pigott, V.C., Martinón-Torres, M., Rehren, T., 2010. Prehistoric copper production and technological reproduction in the Khao Wong Prachan Valley of Central Thailand. Archaeol. Anthropol. Sci. 2 (4), 237–264. http://dx.doi.org/10.1007/s12520-010-0043-y.

QGIS Development Team, 2022. QGIS Geographic Information System. QGIS Association, URL https://www.qgis.org.

R Core Team, 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.

Rasmussen, C.E., Williams, C.K.I., 2005. Gaussian Processes for Machine Learning. The MIT Press, Cambridge, Massachusetts, http://dx.doi.org/10.7551/mitpress/3206.001.0001.

Rojas, U., 1965. El cacique de Turmequé y su época. Imprenta Departamental, Tunja.

Rovira, S., 1994. Pre-Hispanic goldwork from the Museo de America, Madrid: A new set of analyses. In: Scott, D.A., Meyers, P. (Eds.), Archaeometry of Pre-Columbian Sites and Artifacts: Proceedings of a Symposium Organized By the UCLA Institute of Archaeology and the Getty Conservation Institute, Los Angeles, California, March 23-27, 1992. Getty Conservation Institute, Marina Del Rey, CA, pp. 323–350.

RStudio Team, 2023. RStudio: Integrated Development Environment for R (v. 2023.06.1+524). RStudio, PBC, Boston, MA, URL http://www.rstudio.com/.

Sáenz-Samper, J., 2021. Vestigios arqueológicos de mineros, metalurgistas y orfebres prehispánicos en Colombia. In: Vetter Parodi, L., Lleras Pérez, R. (Eds.), Los Metales En Nuestra Historia. Instituto Francés de Estudios Andinos, Lima, pp. 231–260.

Schiffer, M.B., Skibo, J.M., 1997. The explanation artifact variability. Am. Antiq. 62 (1), 27–50. http://dx.doi.org/10.2307/282378.

Sillar, B., Tite, M.S., 2000. The challenge of 'technological choices' for materials science approaches in archaeology. Archaeometry 42 (1), 2–20. http://dx.doi.org/10.1111/j.1475-4754.2000.tb00863.x.

Simas, A.B., Barreto-Souza, W., Rocha, A.V., 2010. Improved estimators for a general class of beta regression models. Comput. Statist. Data Anal. 54 (2), 348–366. http://dx.doi.org/10.1016/j.csda.2009.08.017, URL http://arxiv.org/abs/0809.1878.

Simón, F.P., 1982. Noticias historiales de las conquistas de Tierra Firme en las Indias Occidentales. Bogota. Biblioteco Banco Popular, Bogotá, Originally written in 1627.

Stan Development Team, 2021a. RStan: The R interface to Stan (v. 2.26.22). URL https://mc-stan.org/.

Stan Development Team, 2021b. Stan Modeling Language User's Guide. URL https://mc-stan.org.

Tangri, D., Wright, R.V.S., 1993. Multivariate analysis of compositional data: Applied comparisons favour standard principal components analysis over Aitchison's loglinear contrast method. Archaeometry 35 (1), 103–112. http://dx.doi.org/10.1111/j.1475-4754.1993.tb01026.x.

Uribe, M.A., Londoño, E., Quintero, J.P., Martinón-Torres, M., 2013. Historias de ofrendas muiscas (exposición temporal, Museo Del Oro - Bogotá D.C., 30 de mayo - 20 octubre 2013). Museo del Oro, Banco de la República, Bogotá, URL https://www.cervantesvirtual.com/obra/historias-de-ofrendas-muiscas-1072561.

Uribe Villegas, M.A., 2007. From Chemical Analyses to Meaning: A Study of Materials and Choices in Muisca Metalwork from Pre-Hispanic Colombia (MA dissertation). University College London, London.

Uribe Villegas, M.A., 2012. Contexto, significado y color en la selección de materiales en la orfebrería muisca. Un estudio analítico e interpretativo de la composición química de artefactos. Bol. Arqueología 23.

Uribe Villegas, M.A., Martinón-Torres, M., 2012a. Composition, colour and context in Muisca votive metalwork (Colombia, AD 600–1800). Antiquity 86 (333), 772–791. http://dx.doi.org/10.1017/S0003598X00047918.

Uribe Villegas, M.A., Martinón-Torres, M., 2012b. Typology, technology, composition and context of Muisca metalwork (Colombia, AD 600-1800): A Database. J. Open Archaeol. Data 1 (e4), 1–3.

Uribe Villegas, M.A., Martinón-Torres, M., 2021. Los orfebres, las ofrendas y el oro entre los muiscas. Resultados de una investigación arqueometalúrgica en las figuras votivas de metal. In: Meisel Roca, A., Uribe Villegas, M.A., Langebaek, C.H. (Eds.), Economías Prehispánicas de Colombia. Banco de la República, Bogotá, pp. 127–170.

Uribe Villegas, M.A., Martinón-Torres, M., Quintero Guzman, J.P., 2021. The Muisca raft: Context, materiality and technology. In: McEwan, C., Hoopes, J.W. (Eds.), Pre-Columbian Central America, Colombia, and Ecuador: Toward an Integrated Approach. Dumbarton Oaks Research Library and Collection, Washington D.C., pp. 275–303.

Vieri, J., Uribe-Villegas, M.A., Martinón-Torres, M., 2020. Of forming, gilding and intentionality in pre-Columbian goldwork: Analytical characterisation of artefacts from the Museo del Oro, Bogotá. J. Archaeol. Sci. Rep. 34, 102626. http://dx.doi.org/10.1016/j.jasrep.2020.102626.

Wolfhagen, J., 2020. Re-examining the use of the LSI technique in Zooarchaeology. J. Archaeol. Sci. 123, 105254. http://dx.doi.org/10.1016/j.jas.2020.105254.