ARTICLE

Check for updates

Resolution and the detection of cultural dispersals: development and application of spatiotemporal methods in Lowland South America

Philip Riris [1 ✉] & Fabio Silva[1]

Inferring episodes of expansion, admixture, diffusion, and/or migration in prehistory is undergoing a resurgence in macro-scale archaeological interpretation. In parallel to this renewed popularity, access to computational tools among archaeologists has seen the use of aggregated radiocarbon datasets for the study of dispersals also increasing. This paper advocates for developing reflexive practice in the application of radiocarbon dates to pre-historic dispersals, by reflecting on the qualities of the underlying data, particularly chrono-metric uncertainty, and framing dispersals explicitly in terms of hypothesis testing. This paper draws on cultural expansions within South America and employs two emblematic examples, the Arauquinoid and Tupiguarani traditions, to develop an analytical solution that not only incorporates chronometric uncertainty in bivariate regression but, importantly, tests whether the datasets provide statistically significant evidence for a dispersal process. The analysis, which the paper provides the means to replicate, identifies fundamental issues with reso-lution and data quality that impede identification of pre-Columbian cultural dispersals through simple spatial gradients of radiocarbon data. The results suggest that reflexivity must be fed back into theoretical frameworks of prehistoric mobility for the study of dispersals, in turn informing the construction of more critical statistical null models, and alternative models of cultural expansion should be formally considered alongside demographic models.

[1] Institute for Modelling Socio-Environmental Transitions, Bournemouth University, Poole, UK. ✉email: priris@bournemouth.ac.uk

Introduction

The first regional appearances of distinctive cultural phenomena have long captivated the attention of archaeologists in lowland South America. Although archaeological interest in dispersals here has antecedents in the 19th century, dispersal-focused research was taken up as a programmatic interest as the cumulative weight of linguistic, ethnographic, and archaeological evidence was brought to bear during the 20th century (Métraux 1928; Steward 1948; Taylor and Rouse 1955; Hurault 1972). Understanding dispersals has remained a central fixture of archaeological research in lowland South America since. Dispersals, employed here as an umbrella term, comprise diverse mechanisms ranging from migratory population movements and demic expansions, to the diffusion of crops and other cultural traits (Evans and Meggers 1960; Lathrap 1970; Brochado 1984; Zucchi 1985; Oliver 1989). Recent research has enhanced and built upon the paradigmatic models of previous generations of scholars (Noelli 1998; Zucchi 2002; Hornborg 2005; Araujo 2007; Almeida and Neves 2015; Bonomo et al., 2015; Iriarte et al., 2017a; Antczak et al., 2017; De Souza et al., 2020), while ancient DNA evidence is also enabling past movements of human and plant populations to be traced and compared in new ways (Posth et al., 2018; Zarrillo et al., 2018; Kistler et al., 2018; Mühlen et al., 2019). Understanding dispersal dynamics is crucial for accurately situating archaeological information in relation to independent climatic, historical linguistic, and palaeoecological data. Facilitating cross-disciplinary comparisons of this nature is therefore positioned to inform on pre-Columbian ecodynamics and resilience, the relative importance of push/pull factors in driving dispersals, and improve on existing estimates for languages branching from ancestral stock (Rostain 2013; Bonomo et al., 2015; Iriarte et al., 2017a; Antczak et al., 2017). Rather than evaluating the degree of support for different dispersal models via ethnographic, linguistic, or archaeological (material culture) information, or debating the nature of underlying drivers, this paper focuses on inferring dispersal dynamics on the basis of geolocated pre-Columbian radiocarbon dates. Dispersal-focused research in our domain of interest remains largely isolated from this literature methodologically, with one notable recent exception (De Souza et al., 2020). We aim to close the gap by performing quantitative analyses of chronometric data and present a new method for achieving this.

Archaeological evidence for pre-Columbian cultural dispersals is to a large degree underpinned by the use of ever-expanding spatiotemporal datasets, that is, georeferenced chronometric dates (Anderson and Gillam 2000; Bonomo et al., 2015; De Souza et al., 2020). Taking a broader view, archaeology has experienced rapid methodological developments in the aggregate analysis of radiometric data (Williams, 2012; Bronk Ramsey, 2017), including methods that formally account for both the spatial and temporal components of dispersal processes (e.g. Davison et al., 2006; Russell et al., 2014; Silva and Steele 2014; Henderson et al., 2014; Silva et al., 2015; Isern et al., 2017; Silva and van der Linden 2017). Quantitative approaches to dispersals typically search for characteristic spatiotemporal trends, namely, decreasing age with increasing distance from source. A canonical version of this type of analysis that many archaeologists are familiar with is the spread of prehistoric farming economies in western Eurasia (Ammerman and Cavalli-Sforza, 1971). When investigating such a trend, linear regression can be used to infer valuable information on the dispersal process. Continuing the Eurasian Neolithic example, recent research has typically employed values for the calendar date of first appearance of the dispersing archaeological entity at each site (e.g. pottery, crops), and the distance to each site from a putative origin (Pinhasi et al., 2005; Gkiasta et al., 2003; Fort et al., 2012; Silva and Steele 2014; Silva et al., 2015).

These techniques can be used to analyse innovation-adoption diffusion as well as population expansion. In a given case, all that is required is that the dispersal dynamic was spatially homogeneous and influenced by distance from the centre(s) of origination, that the rate of spread was slow enough to be resolvable by radiometric dating, and that a sufficient sample of dated occurrences is available for analysis. We return to these points below.

Our goals are to develop analytical tools to analyse pre-Columbian cultural dispersals by: (1) formally taking into account the full extent of chronometric uncertainty in radiocarbon data and (2) assessing whether the datasets present *statistically significant* evidence for a dispersal process. We argue that these two points, largely ignored by the methodological literature, are crucial first steps in the analysis of any archaeological dispersal. This is because they address the underlying qualities of archaeological datasets as well as provide a test for whether subsequent analysis and modelling are warranted or, conversely, whether efforts should focus on seeking more and better data prior to further analysis. With reference to the broader theoretical goals of identifying archaeological dispersals, our innovative method enables an exploration into the limits of statistical inference from radiometric dates. Our treatment of this issue is generally applicable across cultural contexts and is therefore relevant to archaeological practice beyond our case studies.

Pre-Columbian dispersals in lowland South America

Tracing dispersals in space and time is deeply entrenched in lowland South American archaeology and therefore remains strongly influenced by the dominant theoretical trends of mid-century North American and European archaeological practice (Rouse 1958; Meggers and Evans 1958; Denevan 1966; Lathrap 1970). Amazonian and circum-Amazonian archaeological cultures that have featured in substantial dispersal-focused research are summarised in Table 1, to which we also add the human colonisation of the continent. Crop dispersals also form an important part of the cultural geography of pre-Columbian South America (Piperno 2011; Clement et al., 2015; Shepard et al., 2020) and would in theory be amenable to spatiotemporal modelling with appropriate data. The analysis we develop focuses specifically on two illustrative cultural dispersals from this list and outlines the associated archaeological data (Fig. 1). First, our data on the Tupiguarani expansion builds on two existing compilations, with our amendments (Bonomo et al., 2015; De Souza et al., 2020). Second, the Arauquinoid expansion data (Zucchi 1985; Versteeg 2008) forms part of a broader long-term effort by the lead author to compile the available archaeological radiocarbon dates in the Orinoco basin and surrounding regions. We only employ dates whose georeferenced location, original publication, and lab code could be ascertained, and exclude dates with errors of greater than ±200 ^{14}C years.

Tupiguarani expansion. Tupiguarani is one of the most widely distributed archaeological cultures of Late Holocene lowland South America. The area occupied by speakers of Tupi-Guarani languages before the European Conquest comprised, conservatively, an area from the mouth of the Amazon to the La Plata Delta, bracketed by the tributaries of the Madeira in the west and the Atlantic Ocean in the east (Noelli 1998; O'Hagan et al., 2019). Although neither wholly nor exclusively occupied by people who spoke Tupi-Guarani languages, this area nonetheless represents almost a third of the total surface of South America. Based on preceding work by Lathrap (1970), Brochado (1984) famously proposed that population pressure due to intensified exploitation

Table 1 Inferred spatiotemporal boundaries of dispersals in South America.

Dispersal	Period	Origin	Spatial domain	References
Palaeoindian colonisation	Starting before 14,000 cal BP	Panama isthmus	Entire continent	Meltzer, 2009; Rothammer and Dillehay, 2009
Amazonian polychrome ceramics	1500–500 cal BP	Amazon delta	Amazon basin	Brochado, 1989; Almeida, 2015
Tupiguarani tradition	3000–500 cal BP	South-Central Amazon	Amazon & La Plata basins, Atlantic coast	Bonomo et al., 2015; Iriarte et al., 2017a
Arauquinoid series	1400–300 cal BP	Middle Orinoco	Coastal Guianas & Lower Amazon	Oliver, 1989; Rostain, 2013
Koriabo ceramics	1200–500 cal BP	Guiana Shield	Guiana Shield & Lower Amazon	Boomert, 2004; Duin, 2014
Southern proto-Jê culture	2200–500 cal BP	Central Brazil	Southern Brazilian Highlands	Araujo, 2007; Iriarte et al., 2017b
Barrancoid series	2500–1700 cal BP	Lower Orinoco	Orinoco & Amazon rivers	Lathrap, 1970; Boomert, 2000

of aquatic and floodplain environments in Amazonia led to a wave of migration of Tupi-Guarani speakers outwards. Supported by the 29 radiocarbon determinations available at the time, this work provided the first synthetic explanation for the expansion in terms of demography, linguistics, material culture, and cultural ecology, as well as a template for much subsequent research (Heckenberger et al., 1998; Noelli 2008; Iriarte et al., 2017a; Silva and Noelli 2017; Noelli et al., 2018).

The southwards dispersal of Tupiguarani people was estimated to have begun before 3000 cal BP (Heckenberger et al., 1998; Rodrigues and Cabral, 2012), noting that the earliest Tupiguarani-associated radiocarbon date (discounting possible alternative ancestral Tupian sites in southwest Amazonia) calibrates to 2698–2158 cal BP (Beta-227301: 2410 ± 40 at 2σ with the SHCal20 curve). Historically, the covariance of Tupi-Guarani languages, Tupiguarani material culture, and ethnic identity was assumed to be more or less static, as archaeological ceramics are highly diagnostic with polychrome and corrugated decoration, while the languages within this family are more closely related than any other branch of broader Tupian stock. Current scholarship is more cautious in connecting change in one category directly to change in another, both in archaeology (Noelli 2008; Neves 2011; Iriarte et al., 2017a) and in linguistics (Mello and Kneip 2017; O'Hagan et al., 2019). Recent works have proposed new phasings for the archaeological correlates of the expansion (Bonomo et al., 2015; De Souza et al., 2020), while efforts to systematise connections between the archaeological and linguistic record are also ongoing (Silva and Noelli 2017; Noelli et al., 2018).

Arauquinoid expansion. The Arauquinoid series, in the nomenclature of Venezuelan and Caribbean archaeology, is estimated to have emerged around 1400 cal BP (AD ~600) near the Middle Orinoco–Apure confluence in present-day Venezuela (Rouse and Cruxent 1963; Zucchi 1985). Arauquinoid ceramics spread along the Orinoco and its major tributaries, reaching across the coastal Guianas as far east as present-day Cayenne before 700 cal BP (Rostain 2013). Although regional variation is evident, particularly in land use patterns (Roosevelt 1980; Navarrete,2008; Versteeg 2008), ceramic decoration is diagnostic and consistent enough for sub-styles to be considered regional expressions of a single interlinked archaeological culture (Rostain, 2013, p. 114). The distribution of Arauquinoid cultures overlaps considerably with that of recorded Cariban languages, leading many archaeologists to consider a connection plausible (Tarble and Zucchi 1984; Zucchi 1985; Boomert 2000; Antczak et al., 2017). Demographic expansion leading to sustained population growth has been suggested as the main driver behind the coeval rapid spread of languages and culture, the latter evidenced by the relatively long-lived nature and high archaeological visibility of Arauquinoid sites (Roosevelt 1980; Gassón 2002). Incised Punctate tradition ceramics in the Lower Amazon share some Arauquinoid characteristics, and its makers are suggested to have spoken a Cariban language too (Gomes 2008; Heckenberger and Neves 2009).

The spatial distribution of Arauquinoid cultures contrasts with the comparatively short time scale of their expansion. Basal Arauquinoid dates for occupations in the Middle Orinoco (GX-8981: 1692–1065 cal BP at 2σ with IntCal20, 1455 ± 140) overlap considerably with the first appearances of Arauquinoid pottery in Suriname (GrN-8955: 1305–1177 cal BP, 1335 ± 38) and French Guiana (GrN-9361: 1260–1059 cal BP, 1210 ± 30). Taking the extremes of these assays at face value, the distribution of Arauquinoid sites would appear to support a broad west-to-east trend. Given that the geodesic distance between the first and last
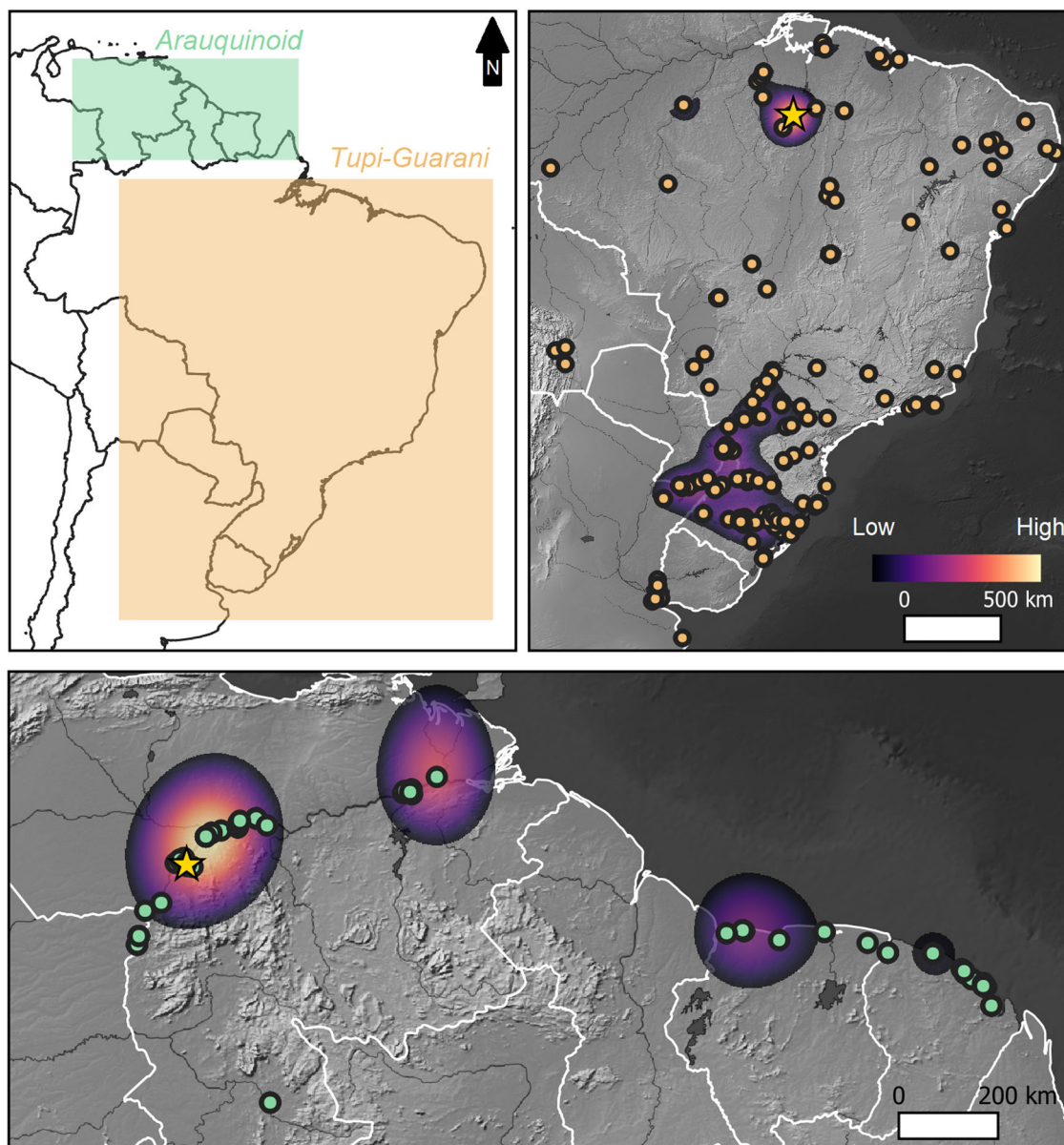
Fig. 1 Domains of the Arauquinoid (green) and Tupiguarani (orange) radiocarbon datasets in South America, with earliest dated sites (yellow stars). Insets show kernel density estimates ($\sigma = 100$ km) of georeferenced radiocarbon dates, illustrating relatively high-density areas, filtered with an arbitrary cut-off of one standard deviation above the mean density.

determination is on the order of 1200 km, such a degree of overlap in calibrated probability distributions raises the question of how fast the expansion realistically occurred. As the main mechanism for the spread of Arauquinoid cultures is said to be migration underpinned by population expansion (Antczak et al., 2017, p. 134), the medians of the calibrated dates (1229 and 1128 cal BP) imply positively explosive growth and migration rates, were this to be the case.

Archaeological dispersals and chronometric uncertainty. A typical dispersal process is such that the time of arrival will increase with distance from the source, which results in a bivariate spatio-temporal trend that can be quantified using regression analysis. The latter was first applied in archaeology to radiocarbon dates for the spread of the Neolithic in Europe (Ammerman and Cavali-Sforza 1971) and have since been used in countless other quantitative studies (e.g. Pinhasi et al., 2005; Fort et al., 2012; Silva and Steele 2014; Silva et al., 2015; Isern

et al., 2017; De Souza et al., 2020). Regression analysis calculates the line that best-fits the empirical set of dates and distances from a pre-selected source. This involves estimating values for the two regression parameters—the intercept and the slope—resulting in a linear equation which, for a given distance, predicts the arrival time (Eq. (1)).

$$\text{time}_{\text{arrival}} = \text{intercept} + \text{slope} \times \text{distance} \qquad (1)$$

A rate-of-spread, or dispersal speed, can be calculated from the inverse of the slope and, following Ammerman and Cavali-Sforza (1971), this has become the de facto parameter when dealing with archaeological dispersals (Fort, 2012), despite masking considerable geographic variation (e.g. Bocquet-Appel et al., 2012; Silva and Steele 2014; Fort 2015) and ignoring other, arguably more important, dynamics (Silva and Vander Linden 2017). Conversely, the intercept can be thought of as the origination time of the dispersal process, even though this parameter is not usually given the same attention as its counterpart.

Presently available methods of linear regression require as input a single date value (a single year) from each chronometric observation—usually the calibrated median date (e.g. Pinhasi et al., 2005). This uncritical use of median values is not only at odds with the use of radiometric dates elsewhere in archaeology but, more importantly, fails to capture the chronological uncertainty that is a key element of a radiometric determination. This uncertainty, especially that of calibrated radiocarbon dates, is not trivial; uncertainty is non-normally distributed due to calibration curve effects. Consequently, any method that does not take the full radiometric probability distribution into account does so at the risk of arriving at the wrong conclusions. Hazelwood and Steele (2004) were the first to discuss this problem and showed the significant impact chronometric uncertainty has on the determination of the slope, and hence on the determination of the rate-of-spread. We illustrate this point with two of the earliest calibrated dates in our Arauquinoid dataset Fig. 2 (top), one in the Middle Orinoco and the other in the eastern Coastal Guianas and hence separated by more than 1300 km (see dataset details in next section). A linear regression performed on their median values yields a very small slope of 0.0197, which would correspond to a dispersal speed of 50.7 km/yr over this spatiotemporal domain. If one were to take the calibrated 95% range of these dates into account, as Hazelwood and Steele (2004) pointed out, one would be forced to acknowledge that the real slope may be any value between −0.279 and 0.339, thus implying that the rate-of-spread may be anywhere between ~3.0 km/yr and infinity, with either direction of spread being possible. These ranges are not only considerably broader than the value obtained for the median regression indicates, but are also unrealistic given the available archaeological information. A slope near zero implies that the trait or people dispersing would have appeared simultaneously across the entire region under study (due to an infinite rate-of-spread), which is quite clearly nonsensical. Conversely, a positive value for the slope would mean that the Arauquinoid dispersal happened *from* the Atlantic coast of the Guianas *toward* the Middle Orinoco. The implication is that the uncertainty in the calibrated date leads to an uncertainty (i.e. a range) in the inferred dispersal parameters that, as in the example just given, can be so extreme as to cast doubt over any one interpretation of this dataset. These observations cast considerable doubt over dispersal studies that are based on bivariate regressions of median values in general, a point that we return to later.

As a corollary to the recognition of chronometric uncertainty, Hazelwood and Steele (2004) highlighted what is essentially a problem of resolution: high chronometric uncertainty is only problematic when the dates corresponding to first arrival across the dispersal domain cannot be resolved, i.e. when they cannot be chronologically differentiated. When first arrival dates on opposite ends of the dispersal domain are essentially the same, the data does not allow detection of a dispersal. A dispersal may very well have occurred, but the chronometric data does not have enough resolution to allow us to observe its dynamics. Using a thought-experiment consisting of two dates corresponding to first arrival on two different locations, Hazelwood and Steele (2004) derived an algebraic rule for *minimum chronometric resolution* which essentially states that, in order to detect a dispersal signal, the sum of the calibrated uncertainties of the two dates must necessarily be much smaller than the time elapsed between arrivals at the extremes of the domain, illustrated in Fig. 2 with two contrasting cases.

Hazelwood and Steele's (2004) algebraic rule, however, is only meaningful in a situation where each date truly corresponds to the first arrival of the dispersing element at two sites on opposite ends of the domain. A realistic case study will have more than two
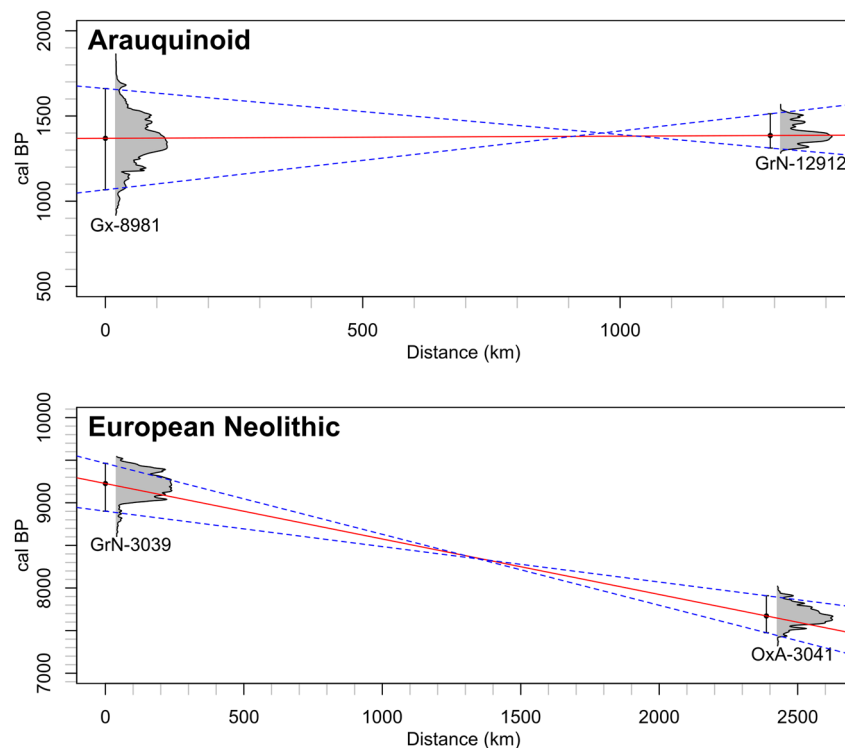

Fig. 2 Example of two radiocarbon dates in each dataset that are among the earliest at opposite ends of the dispersal domain. *Top:* Middle Orinoco and Suriname for the Arauquinoid dispersal. *Bottom:* Greece and Great Britain for the European Neolithic. Shown are the median dates (dot), 95% calibrated range (error bars) and full calibrated distributions (grey-shaded), as well as the results of linear regressions using the median dates (solid red line) and the cross ends of the 95% ranges (dashed blue lines).

dates, several of which will not correspond to first arrival but to later instances of the dispersing cultural or material trait, originating from multiple sites. For such cases, a quick and easy algebraic test is not forthcoming, making it difficult to estimate whether the dataset has enough chronological resolution. A numerical approach that should also take into account non-normal calibrated distributions, however, can be developed and it is a key aim of this paper to do so.

To achieve this, we reformulate Hazelwood and Steele's question of minimum chronological resolution as a question of the statistical significance of the slope. If there is not enough resolution to differentiate two dates at both ends of the domain, then a linear regression conducted on those two dates will yield a slope close to zero (Fig. 2, top). If, however, the elapsed time is sufficiently larger than the uncertainty of those dates, then the slope obtained from linear regression should constitute statistically significant evidence against the null hypothesis of zero slope (a flat line implying instantaneous dispersal). The question of resolution then becomes one of looking at and reporting the significance of the slope obtained from a linear regression, as quantified by a *p*-value. To realise Hazelwood and Steele (2004)'s full vision, a significance test that takes the calibrated uncertainty into account is necessary. Before detailing how we achieved this, it is important to highlight that failure to reject the null hypothesis does not demand that we accept the data as evidence for the absence of a dispersal process. Rather, it simply means that the data is insufficient to allow one to infer a dispersal due to lack of chronometric resolution. A dispersal may still have occurred in the past, but the lack of statistical significance means that the radiometric dataset is such that we cannot use it to infer the *dynamics* of dispersal, such as direction, rate-of-spread, preferred routes, or mode.

Materials and methods

Datasets of radiocarbon determinations, aggregated from the literature, form the basis for our regression analyses. We make use of two datasets of radiocarbon determinations to explore the Tupiguarani and Arauquinoid expansions. These contain only dates whose affiliation to Tupiguarani and Arauquinoid contexts, location, and lab code could be cross-referenced with a publication. We have chosen not to include thermoluminescence dates, as local measurements of annual dose rates are systematically underreported, and their absence impedes accurate calibration to calendric ages. In total, the databases comprise 375 Tupiguarani-affiliated dates from 234 sites, and 178 Arauquinoid-affiliated dates from 47 sites. Due to a lack of demonstrative archaeological work on the topic, we do not consider either pre-Arauquinoid Corozal (Roosevelt 1997) or Koriabo (Rostain 2008) ceramics to be part of the Arauquinoid series.

To comment briefly on the spatial distribution of the data, notable voids are apparent, such as the entire territory of modern Guyana, the interior of the Guiana Shield, and the Orinoco Delta for the Arauquinoid (Fig. 1a), and for the Tupiguarani, the country of Paraguay, as well as large tracts of the Lower Paraná and Uruguay rivers (Fig. 1b). More sites affiliated to these cultures are known, however, in most cases narrowing the geographical gaps, if not the temporal ones (Rostain 2013; Corrêa 2017). The superficial limitations of the data notwithstanding, we are confident that our data collection efforts have managed to capture the majority of relevant chronometric determinations in both domains. We consider that the relatively small number of dated sites increases the need for systematic analyses to critically assess the underlying cultural dispersal processes. All dates were calibrated with the *rcarbon* package for R (Crema and Bevan 2020), using the method described by Ramsey (2008), as well as the

SHCal20 calibration curve (Hogg et al. 2020) on the Tupiguarani dataset and the IntCal20 curve (Reimer et al., 2020) on the Arauquinoid.

Since a radiometric dataset is bound to include dates which do not correspond to the first arrival of the dispersing element, regressing to the conditional mean—as is done by ordinary least squares—will not yield the most useful or valid relationship (Steele 2010; Silva and Steele 2014; Silva et al., 2015). What is needed is a method that can approximate not to the central tendency of the dataset but to the earliest dates throughout the domain considered, without having to subset or filter the data a priori, as done, for example, by Russell et al. (2014) and de Souza et al. (2020). Quantile regression (Koenker 2005) is uniquely suited for this purpose as it infers parameters that approximate to a given quantile edge, which can be chosen so as to to capture the top percentile(s) of the dataset, thus regressing to the earliest dates. This approach is more robust than ordinary least-squares since, among other things, it does not require the data to be homoscedastic, that is, normally distributed around the central trendline with constant variance. In fact, quantile regression is particularly suited for dataset in the presence of unequal variance (Cade and Noon 2003), which makes it ideal for radiometric datasets of the type considered here. Furthermore, quantile regression does not make assumptions about the distribution of regression residuals, meaning it can handle multimodal distributions. This is advantageous within archaeology to accurately take into account instances of local discontinuities (e.g. arrival of a particular cultural element, followed by abandonment, followed by re-adoption).

The choice of quantile to regress to is expected to impact results since this parameter effectively measures the weight of importance being given to very early dates. We have chosen to regress to the 99th-percentile of the dataset, allowing the top 1st-percentile to include outliers which are therefore neither arbitrarily excluded nor overemphasised by this analysis. This can be seen as a very liberal approach since it minimises the number of datapoints that could be considered very early outliers. To ensure that our results are not affected by this choice we have rerun our analyses using more conservative regressions to the 90th and 95th-percentiles which show that our conclusions are not sensitive to this parameter (Table S1). Quantile regressions were conducted in R using the *quantreg* package (Koenker 2020), which uses a modified Barrodale and Roberts algorithm for regression (Koenker and d'Orrey 1987) and where parameter standard errors were obtained using the *xy*-pair method which was found to perform well under a variety of circumstances (Koenker 1994, 2005).

Linear regression techniques require two values for each observation: the arrival time and the distance from the source of dispersal. With regards to the latter we have used geodesic ('crow-flies') distance between each archaeological site and the one chosen as the putative source of dispersal. We opted to use as origin the site with the *actual* oldest date, rather than extrapolate to fictional ones using best-fitting approaches (e.g. Pinhasi et al., 2005; De Souza et al., 2020). For the Tupiguarani we used the site of Abraham (De Souza et al., 2020), whereas for the Arauquinoid dispersal we have chosen El Valle as the site with the oldest calibrated radiocarbon date in Venezuela (see Fig. 1), since the nearby type site of Arauquin lacks radiocarbon dates (Cruxent and Rouse 1961).

Currently available methods of linear regression require a single value (a single year) per observation (each archaeological site). Typically, the calibrated median date is used for this. As discussed above, this fails to capture the intrinsic chronometric uncertainty that is encapsulated in a radiometric determination and which in other use cases is regarded as important by

archaeologists. To assess, in a quantitative manner, the impact of this uncertainty in the outcomes of linear regression, one can take a large number of random samples of the calibrated radiometric distributions for use in a regression algorithm, recording its outputs and thereby assessing the uncertainty they accumulate. We have done this Monte Carlo sampling 1000 times, performed the quantile regressions on these 1000 datasets, and saved the outputted regression parameters for analysis, using R code especially designed for this purpose.

Quantifying the uncertainty in the regression parameters due to dating errors is, however, insufficient; it remains necessary to test whether or not the results constitute statistically significant evidence against a null slope. This is perhaps the most important quantitative step, since a linear regression algorithm will always produce a line, *regardless of whether or not there is sufficient resolution to detect the signal of a dispersal*. For traditional statistical data sources, a significance test can be easily done on the slope, testing whether this value is significantly different from zero. For chronometric data sources with non-normal calibrated distributions, this is not as straightforward, and no currently available methods can be used.

The first stage in developing a test for statistical significance is identifying what the null hypothesis is, and how best to model it quantitatively. For present purposes, the key is to consider what a dispersal with zero slope (an infinitely fast dispersal, as mentioned previously) could look like in a given region with the available dataset. Naturally, this phenomenon will still occur for slopes close to zero, but the inherent chronometric uncertainty of calibrated radiocarbon dates means that there will be a broader range of slopes around the value of zero that nevertheless still fall under the expectations of the null hypothesis. A significance test should therefore estimate this range and derive a *p*-value from it. We have developed a procedure in R inspired by similar work done for radiocarbon summed probability densities (Shennan et al., 2013; Timpson et al., 2014; Crema et al., 2016) and structural orientations in skyscape archaeology (Silva 2020). The null model is constructed by randomly permuting the spatial and temporal components of the dataset, thereby effectively and efficiently removing any spatio-temporal trends that may be present, whilst retaining the other important characteristics of the dataset, such as its spatial and temporal resolution and extent, which are key to identifying the envelope of the null hypothesis. Simulated datasets can then be analysed in the same way as the empirical dataset, namely through random chronometric sampling and quantile regressions.

The resulting regression parameters from the 1000 regressions are aggregated by summing their individual probability distributions, which are modelled as normal curves using the parameter values and standard errors. This approach is valid for quantile regression since the asymptotic normality of parameters inferred by quantile regression has been demonstrated previously (Koenker 2005). This process is repeated a large number of times, 1000 in our case, to ensure that the range of random possibilities was properly sampled and included in the analysis. The aggregated summed probability distributions (SPD) for each of the regression parameters are then used to calculate the 95% confidence envelope of the null hypothesis, i.e. the envelope within which the regression parameter SPDs are expected to fall in if the dataset conforms to the null hypothesis. Direct comparison of the empirically derived SPD with the confidence envelope is used to calculate a two-tailed *p*-value, following the approach of Timpson et al. (2014), with the correction for Monte Carlo algorithms of North et al. (2002). For present purposes we have set the *a*-level of this significance test to 0.05.

Ideally, to ensure that a new algorithm works as it is expected it is good practice to test it on simulated data that mimics the nuances of empirical data. However, in our specific case, the nature of radiometric archaeological datasets is such that to construct a suitable model to simulate all of its idiosyncrasies would demand a paper on its own. To bypass this problem, we have decided to test it using a dataset that has a well-established east-to-west dispersal: the Pinhasi et al. (2005) radiocarbon dataset for the spread of the Neolithic in Europe. As a control check, we have tested the method on a randomised version of the same database where the radiocarbon dates were randomly assigned to the archaeological sites, thereby removing the spatio-temporal dispersal trend. The results, shown in Figs. S1 and S2, were as expected in indicating that only the non-randomised dataset constitutes statistically significant evidence for a dispersal: slope *p*-value « 0.001 (***) versus 0.999 (ns) for the randomised dataset, thereby validating our innovative methodology. The source files for our method, the datasets and the scripts used to produce the figures and tables in this paper can be found in the following repository: https://doi.org/10.5281/zenodo.4434387.

Results
The application of the above methodology to the Arauquinoid dataset produced the results summarised in Fig. 3 and Table 2. The most obvious point is that both the regression using median dates (red line) and the range obtained when performing our innovative regression using the full chronometric uncertainty (blue-shaded region) result in regression lines that are very close to horizontally flat, and hence very close to the null hypothesis.

The regression parameters (Table 2) make it clear that the best estimate for the slope, whether one is using median dates or taking the full chronometric uncertainty into account, is very close to zero (around 0.03) and hence would correspond to a very fast speed (33.33 km/yr) over this domain. Curiously enough, this slope value is also positive, which would have to be interpreted as a dispersal in the opposite direction. In other words, the available radiometric data for the Arauquinoid is not inconsistent with a dispersal *from* modern French Guiana *toward* the Middle Orinoco, the reverse of archaeological expectations. With respect to the intercept, we get a broad range of values spanning four to seven centuries and roughly corresponding to the calendar range of the earliest dates in this dataset, further emphasising that the only possible regression on this dataset will resemble a flat line. Even if a statistically significant figure were obtained, the range of origination dates is disproportionately large relative to the timespan of the Arauquinoid expansion itself.

In terms of statistical significance, the *p*-values obtained for the regression using median dates are somewhat ambiguous. On the one hand, the obtained intercept appears to be statistically significant, whereas on the other hand the slope is not. Using our novel method, the results are unequivocal that this dataset does not constitute enough evidence to reject the null hypothesis. This is rather clear in Fig. 3 where, in all three panels, the blue lines corresponding to the results of the empirical dataset fall within the envelopes of the null hypothesis (grey-shaded regions).

Furthermore, Fig. 4 and Table 3 summarise the results for the Tupiguarani dataset which, at first glance, produced qualitatively different results. A cursory look at the bivariate plot shows that the output of the regression using median dates (red line) and those using full uncertainty (blue-shaded range) have a definite negative slope, tracking earlier dates closer to the source of dispersal and later dates further away from it. Such an uncritical look at the results of the linear regression would result in an erroneous interpretation, as the significance tests show that the obtained regression line and parameter values are, nevertheless, still within the confidence envelope of the null hypothesis (grey-shaded areas). Contrary to initially promising appearances, therefore, this
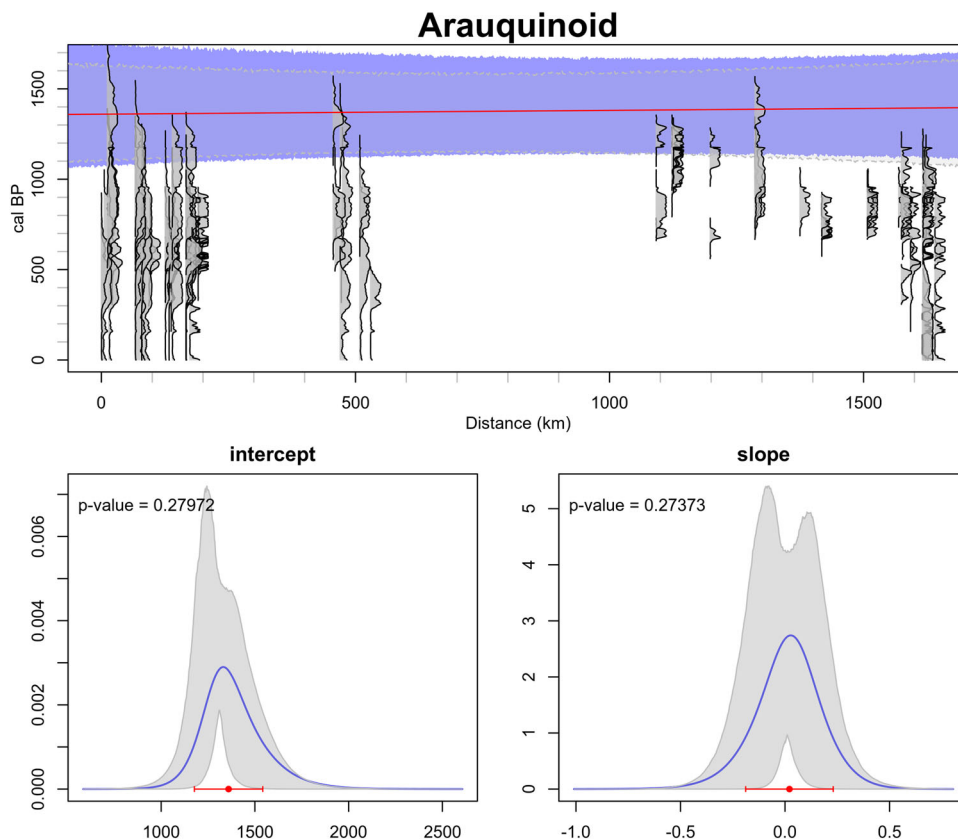
Fig. 3 Results of the radiometric regression analysis on the Arauquinoid dataset. *Top:* Bivariate plot of age versus distance from chosen origin, showing the calibrated radiometric densities (grey curves), the result of the quantile regression to 99th-percentile when using median dates (red line), the spread of regressed lines when taking the full radiometric uncertainty into account (blue-shaded region), and the confidence envelope of the null hypothesis of no-dispersal (grey-shaded region with dashed border). *Bottom Left:* Summed probability density of obtained y-intercept regression parameter for the empirical dataset (blue curve), confidence envelope of the null hypothesis of nodispersal (grey-shaded region), best estimate and corresponding 95% range when doing regression with median dates (red dot and bar). *Bottom Right:* Summed probability density of obtained slope regression parameter with the same key as the previous panel.

Table 2 Regression results for the arauquinoid dataset.			
Arauquinoid	**Peak value**	**95% range**	***p*-value**
Regression using median dates			
Intercept	1360 cal BP	1173–1547 cal BP	«0.001 (***)
Slope	0.021	−0.201–0.244	0.853 (ns)
Regression using chronometric uncertainty			
Intercept	1332 cal BP	1086–1735 cal BP	0.274 (ns)
Slope	0.021	−0.343–0.343	0.285 (ns)

dataset also does not constitute sufficient evidence to discard the null hypothesis.

The *p*-value for the slope obtained in the regression using median dates already indicates no significance (although its value close to 0.05 could, in some circles, be said to be "approaching significance", a term we reject). Interestingly, this result suggests that even when using a less robust approach, for example employing calibrated date medians, one may still be able to identify when the data is insufficient to reject the null hypothesis, provided that one actually examines the significance of the inferred slope.

Discussion
Methodological implications. The results validate the use of our methodology, and especially highlight the need for the

significance testing component to become a staple in the analysis of archaeological datasets for dispersals. Without it, one may assume that the data quality is high enough to obtain trustworthy regression results, and hence infer rates-of-spread and/or origination times, when in fact the dataset simply cannot be used for such purposes.

The lack of statistical significance of the Arauquinoid dataset, being composed of a small number of radiocarbon dates, spread across a small domain and with the dispersal occurring over a relatively short time-span, was not unexpected. What came as a surprise was the lack of significance for the Tupiguarani data, considering that this dataset comprises a spatial domain covering ~7 million km² — an area of similar order of magnitude to that of the spread of the European Neolithic (~12 million km²), and contrasting that of the Arauquinoid (~660 thousand km²). This case study therefore illustrates that the nature of the problem is not one of scale—i.e. of whether the dispersal occurred over a large enough area—but rather one of resolution—i.e. whether the chronometric data available has enough resolution to observe the dispersal.

Furthermore, the Tupiguarani case study illustrates the pitfalls of overconfidence when it comes to regressions using only median dates from radiocarbon distributions. Using the latter, one would find a slope that is sufficiently away from zero not to warrant closer inspection, even though its *p*-value would already indicate lack of significance for this slope. Our approach, on the other hand, shows that the empirical SPDs are squarely inside the
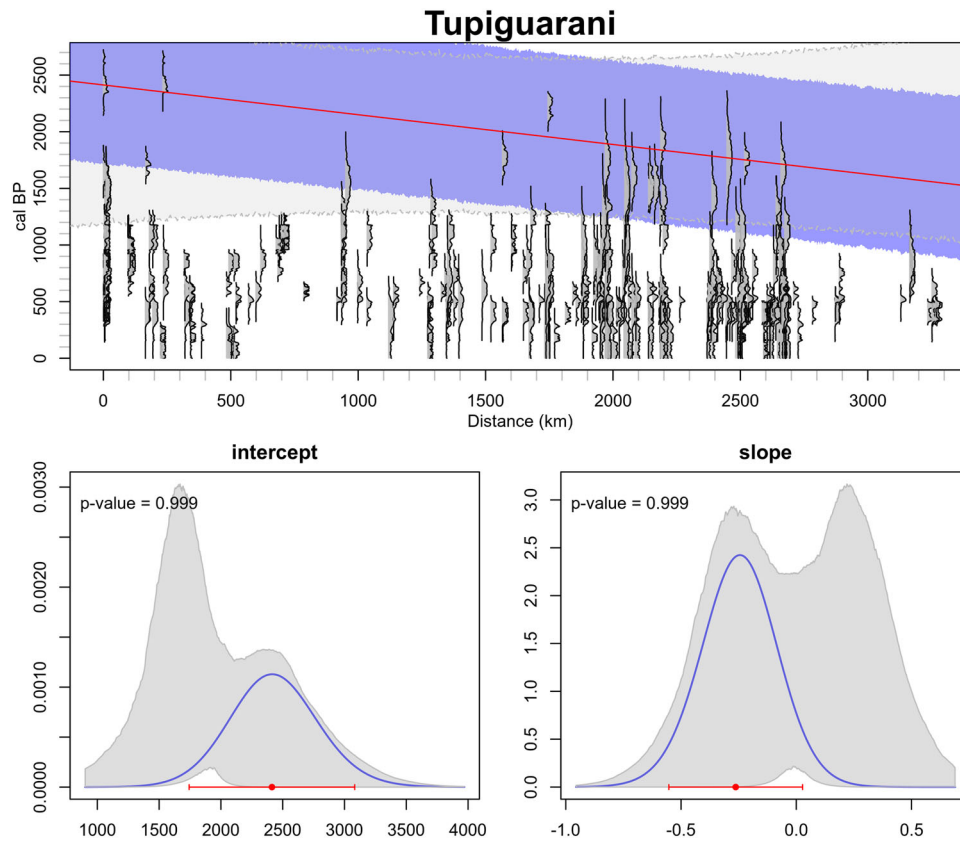
Fig. 4 Results of the radiometric regression analysis on the Tupiguarani dataset. *Top:* Bivariate plot of age versus distance from chosen origin, showing the calibrated radiometric densities (grey curves), the result of the quantile regression to 99th-percentile when using median dates (red line), the spread of regressed lines when taking the full radiometric uncertainty into account (blue-shaded region), and the confidence envelope of the null hypothesis of no-dispersal (grey-shaded region with dashed border). *Bottom Left:* Summed probability density of obtained *y*-intercept regression parameter for the empirical dataset (blue curve), confidence envelope of the null hypothesis of nodispersal (grey-shaded region), best estimate and corresponding 95% range when doing regression with median dates (red dot and bar). *Bottom Right:* Summed probability density of obtained slope regression parameter with the same key as the previous panel.

Table 3 Regression results for the tupiguarani dataset.			
Tupiguarani	**Peak value**	**95% range**	**_p_-value**
Regression using median dates			
Intercept	2413 cal BP	1729–3097 cal BP	« 0.001 (***)
Slope	−0.263	−0.572–0.047	0.097 (ns)
Regression using chronometric uncertainty			
Intercept	2414 cal BP	1710–3159 cal BP	0.999 (ns)
Slope	−0.245	−0.574–0.073	0.999 (ns)

confidence interval of the null hypothesis and that, therefore, this dataset cannot be used to resolve a dispersal.

With regards to methodological limitations, our results rely on choices such as the origin of the dispersal; and the use of geodesic distance over alternatives. We believe, however, that these choices would not significantly impact our results or alter the conclusions. Regarding the first choice, we have used the best possible putative origins, which ensure that the earliest empirical dates appear at or close to them; as opposed to, for example allowing a best-fitting algorithm to extrapolate and choose the origin (e.g. De Souza et al., 2020). Furthermore, the significance testing methodology developed here diminishes the impact of the choice of origin since the model for the null hypothesis decouples the spatial and temporal trends in the dataset. Using alternative origins for the Tupiguarani, such as southwestern Amazonia (Iriarte et al., 2017a), is therefore unlikely to impact the

quantitative results, although non-significant qualitative differences can of course be expected.

With respect to the choice of proxy for travelled distance we acknowledge that other options, such as shortest-path-on-land distance (e.g. Pinhasi et al., 2005) or cost distance (e.g. Silva and Steele 2014), may be more realistic. However, we equally believe they would not significantly affect the results. The domains of the two dispersal processes analysed are composed entirely of land surface; thus shortest-path-on-land is the same as geodesic distance. The addition of cost distances, on the other hand, would involve the addition of extra parameters to the model which could, theoretically, produce better fits. However, costs would have to be such as to introduce perhaps unrealistically high rates-of-spread and, even so, this may not lead to improved significance since our method uses those very distances regardless of how they were obtained. Nevertheless, we acknowledge that further study could shed light on this topic, but this will be left for a future publication.

Another limitation is that this method does not distinguish between a scenario wherein there is a single site at a given distance from the source with, say, 100 dates from an archaeological very different scenario where there are 100 sites, equally distanced from the source, each with a single date. This, however, is not unique to our methodology but rather transversal to the majority of regression approaches that conflate a dataset that is intrinsically spatial (and hence two-dimensional) into a unidimensional parameter—the distance from the source.

Furthermore, the regression framework that we have employed is one that assumes a linear relationship between distance and time elapsed, and hence spatial homogeneity and isotropy. This assumption is intrinsic to the majority of dispersal studies in archaeology, with notable exceptions that either: (a) started from a spatially heterogeneous modelling approach (e.g. Davison et al., 2006; Silva and Steele 2014; Henderson et al., 2014; Isern et al., 2017); or (b) have observed intrinsic non-linearity or spatial heterogeneity in the chronometric data (e.g. Bocquet Appel et al., 2012; Silva and Stele 2014; Silva et al., 2015; Fort 2015; Silva and Vander Linden 2017). To move past these limitations, one would have to use more complex regression models, such as generalised additive models, hierarchical models, multidimensional or multi-level regression, which are beyond the scope of this paper. Nevertheless, the theoretical framework and methodological approach to significance testing laid out here should be easily transferable to a variety, if not the entirety, of regression methods, regardless of their complexity and intrinsic assumptions. Therefore, our contribution goes well beyond the sole usage of linear quantile regression on the case studies discussed above.

Archaeological implications. We underline that archaeological, ethnographic, and linguistic data overwhelmingly support the inference that important cultural dispersals occurred in lowland South America in the pre-Columbian period. We also suggest that the fact of whether a given dispersal has taken place (indexed, for example, by widespread similarities in material culture) is less interesting than understanding the cultural processes and parameters behind it. Our analysis was precipitated by the recognition that in order to sufficiently distinguish between candidate processes, such as migration, demic diffusion, acculturation, range expansions, or others, a useful starting point is inferring dispersal parameters from the available radiometric data. As this research developed, this task became problematic to achieve rigorously. The results comprehensively demonstrate that all the available radiocarbon evidence that can be marshalled in favour of two widely recognised Late Holocene pre-Columbian cultural dispersals is not fit for this purpose. Any proposed sub-phasings contained by our modelling domains, for example on the scale of individual watersheds (Bonomo et al., 2015), are likely to be spurious too. The empirical radiocarbon data cannot be distinguished from artificial data that lack any spatiotemporal structure. Radiocarbon records of analogous dispersals (Table 1) may produce similar results, assuming analogous spatiotemporal resolutions and domains. Our findings also demonstrate the need for testing beyond archaeological cultures, for example dispersals of crops, such as maize or manioc whose footprint in the archaeological record is often slight and radiometrically elusive, largely due to preservation and taphonomy (Kistler et al., 2018; Shepard et al., 2020).

In a historical perspective, demography has long featured among the pre-eminent explanations for culture change in the tropical lowlands of South America, especially models of demic diffusion and population pressure (Lathrap 1970; Meggers 1971; Denevan 1996; Zucchi 2002; Moraes and Neves 2012; Arroyo-Kalin 2018; Arroyo-Kalin and Riris 2020). Demography is currently a vigorous area of research among the historical sciences in general, due to a range of recent methodological innovations within and beyond archaeology (French et al., 2020). As noted, the results of our analysis of the Arauquinoid- and Tupiguarani-affiliated radiocarbon records indicate that, in these two domains, pre-Columbian expansions cannot be distinguished from noise. Consequently, we suggest that recourse to demographic models of cultural expansion should be re-evaluated. In this regard, Gasson (2002) and Van den Bel (2015) have

previously highlighted the apparent speed at which the Arauquinoid series appeared across the Orinoco and Guianas, and suggest alternative modes of dispersal such as exchange and copying to account for this phenomenon.

Among the requirements for successfully detecting dispersals we identified in the introduction to this paper is that the rate of spread (slope) be slow enough to resolve, taking into account the inherent uncertainty of radiocarbon dating. We consider our non-detections worthy of further attention in the context of apparently explosive Indigenous growth rates during the first millennium of the Common Era (Arroyo-Kalin and Riris 2020). It is possible that the dispersals occurred too fast to be resolved, although we lack the necessary information to distinguish this from the lack of a sufficient sample of dates. Cultural traits can emerge and spread far faster than humans can reproduce (Perreault 2012), and for this reason we highlight the need for alternative hypotheses to test chronometric evidence against. As a starting point, we suggest that large-scale systematisation of archaeological data, with supporting linguistic or ethnographic datasets where possible, may offer such an alternative. From a theoretical and methodological perspective, innovations in quantitative methods for examining cultural evolution with large, complex datasets are a promising area (e.g. Riede et al., 2019). The vast amounts of archaeological information on South American cultures, especially late pre-Columbian ceramic traditions with numerous traits amenable to codification, is awaiting such a treatment. Improved models of Indigenous language change (Mello 2000; Gildea 2003; Meira and Franchetto 2005; Hill and Santos-Granero 2010; Eriksen 2011; Eriksen and Galucio 2014) may also prove instrumental to test for covariance alongside archaeological data (e.g. Creanza et al., 2015).

More broadly, we advocate for critical reflection on the appropriateness of archaeological radiocarbon date assemblages for the study of prehistoric dispersals globally. As a first step, as illustrated by our approach, fundamental issues of surrounding data quality and resolution must be considered by archaeologists. The issues discussed here are likely ubiquitous beyond the three case studies presented here. Next, this reflexivity must be fed back into theoretical frameworks surrounding prehistoric mobility and dispersal studies. For example, alternative models of cultural expansion should be formally considered alongside demographic/demic models. Ultimately, we hope our analysis, which we provide the means to replicate, informs the construction of more critical statistical null models in archaeology.

Conclusions
Chronological uncertainty cannot be ignored in the analysis of spatial radiometric datasets for the purposes of inferring the dynamics of archaeological dispersals, a point first highlighted by Hazelwood and Steele (2004). When the amount of chronological uncertainty is of the order of the time elapsed during the dispersal process, the dataset does not offer sufficient resolution to be able to identify a dispersal signal, and hence the dataset cannot be used to infer dispersal parameters, routeways, modes, or any other dynamics. In this paper, we have reformulated this problem as a test for statistical significance against the null hypothesis that the dataset is too noisy to reveal a dispersal signal and developed bespoke algorithms to perform such analyses.

We applied our methodology to two illustrative and iconic pre-Columbian lowland South American dispersals: the Arauquinoid series and the Tupiguarani tradition. The results were unanimous in showing that, in both instances, the available data cannot resolve dates at opposite ends of the spatial domain, and hence the datasets do not provide statistically significant evidence for dispersals. We want to stress that this is not to say that these

cultural expansions are false, but rather that they occurred so quickly that, with the presently available data, they cannot be discerned in the record. This casts serious doubts over studies reliant on this, or similar datasets, that have estimated rates-of-spread, dispersal routes and timings, as our results show that, when chronometric uncertainty is fully considered, the datasets simply cannot be used in this way.

More broadly, our work tries to move the quantitative study of archaeological dispersals forward in two ways. Firstly, it recognises and acknowledges the inherent qualities and uncertainties of the archaeological datasets in routine use, as opposed to treating them as mathematical quantities that can be simplified in ways that ignore (rather than condense) those qualities and uncertainties. And secondly, it develops more robust methodologies that specifically incorporate, both in principle and technically, those qualities and uncertainties. We argue that, in future, the innovative methodology developed herein should form a new baseline test for the study of dispersals in archaeology. Our significance test provides a way to check whether further analysis and modelling of radiometric dates is warranted. Taking widely separated median dates and drawing a line between them is insufficient and clearly no longer adequate.

Data availability

Datasets related to this study are available as Supplementary Information and in the following Zenodo repository: https://doi.org/10.5281/zenodo.4434387.

References

Almeida FO, Neves EG (2015) Evidências arqueológicas para a origen dos Tupi-Guarani no leste da Amazônia. Mana 21:499–525

Almeida FO (2015) A arqueologia dos fermentados: a etílica história dos Tupi-Guarani. Estud av 29:87–118

Ammerman AJ, Cavalli-Sforza LL (1971) Measuring the rate of spread of early farming in Europe. Man 6:674–688

Anderson DG, Gillam JC (2000) Paleoindian colonization of the Americas: implications from an examination of physiography demography and artifact distribution. Am Antiq 65:43–66

Antczak A, Urbani B, Antczak MM (2017) Re-thinking the migration of Cariban-speakers from the Middle Orinoco River to North-Central Venezuela (AD 800). J World Prehist 30:131–175

Araujo AGM (2007) A tradição cerâmica Itararé-Taquara: características área de ocorrência e algunas hipóteses sobre a expansão dos grupos Jê no sudeste do Brasil. Rev Arqueol 20:9–38

Arroyo-Kalin M (2018) Human niche construction and population growth in Pre-Columbian Amazonia. Archaeol Int 20:122–136

Arroyo-Kalin M, Riris P (2020) Did pre-Columbian populations of the Amazonian biome reach carrying capacity during the Late Holocene? Philos Trans R Soc B. https://doi.org/10.1098/rstb.2019.0715

Bocquet-Appel J-P, Naji S, Vander Linden M et al. (2012) Understanding the rates of expansion of the farming system in Europe. J Archaeol Sci 39:531–546

Bonomo M, Angrizani RC, Apolinaire E et al. (2015) A model for the Guarani expansion in the La Plata Basin and littoral zone of southern. Brazil Quat Int 356:54–73

Boomert A (2000) Trinidad Tobago and the Lower Orinoco interaction sphere: an archaeological/ethnohistorical study. Alkmaar, Cairi Publications.

Boomert A (2004) Koriabo and the polychrome tradition: the late-prehistoric era between the Orinoco and Amazon mouths. In: Delpuech A, Hofman C (eds) Late Ceramic Age societies in the Eastern Caribbean. BAR International Series, Oxford, pp. 251–268

Brochado JPP (1984) An ecological model of the spread of pottery and agriculture into Eastern South America. PhD dissertation, University of Illinois Urbana-Champaign.

Brochado JPP (1989) A Expansão dos tupi e da cerâmica da Tradicão Policrômica Amazônica. Dédalo 9:41–47

Bronk Ramsey C (2008) Radiocarbon dating: revolutions in understanding. Archaeometry 50:249–275. https://doi.org/10.1111/j.1475-4754.2008.00394.x

Bronk Ramsey C (2017) Methods for summarizing radiocarbon datasets. Radiocarbon 59:1809–1833

Cade BS, Noon BR (2003) A gentle introduction to quantile regression for ecologists. Front Ecol Environ 1(8):412–420

Clement CR, Denevan WM, Heckenberger MJ et al. (2015) The domestication of Amazonia before European conquest. Proc R Soc B 282:20150813. https://doi.org/10.1098/rspb20150813

Corrêa AA (2017) Datações na bibliografia arqueológica brasileira a partir dos sítios Tupi. Cad LEPAARQ 27:380–406

Creanza N, Ruhlen M, Pemberton TJ, Rosenberg NA, Feldman MW, Ramachandran S (2015) A comparison of worldwide phonemic and genetic variation in human populations. Proc Natl Acad Sci USA 112:1265–1272.

Crema E, Habu J, Kobayashi K et al. (2016) Summed probability distribution of 14C dates suggests regional divergences in the population dynamics of the Jomon Period in Eastern Japan. PLoS ONE 11(4):e0154809. https://doi.org/10.1371/journal.pone.0154809

Crema ER, Bevan A (2020) Inference from large sets of radiocarbon dates: software and methods. Radiocarbon 1−17. https://doi.org/10.1017/RDC.2020.95

Cruxent JM, Rouse I (1961) Arqueología Cronológica de Venezuela, vol 1. Union Panamericana, Washington

Davison K, Dolukhanov P, Sarson GR et al. (2006) The role of waterways in the spread of the Neolithic. J Archaeol Sci 33:641–652

Denevan WM (1966) A cultural-ecological view of the former aboriginal settlement in the Amazon basin. Prof Geogr 18:346–351

Denevan WM (1996) A bluff model of riverine settlement in prehistoric Amazonia. Ann Assoc Am Geogr 86:654–681

de Souza JG, Mateos JA, Madella M (2020) Archaeological expansions in tropical South America during the late Holocene: assessing the role of demic diffusion. PLoS ONE 15(4):e0232367. https://doi.org/10.1371/journalpone0232367

Duin RS (2014) Ethnographic and Archaeological "Cultures" in Guiana Northern Amazonia. In: Rostain S (ed) Antes de Orellana: Actas del 3er Encuentro Internacional de Arqueología Amazónica. Instituto Francés de Estudios Andinos, Arequipa, pp. 89–96

Eriksen L (2011) Nature and culture in prehistoric Amazonia: using GIS to reconstruct ancient ethnogenetic processes from archaeology linguistics geography and ethnohistory. Lund Studies in Human Ecology, vol 12. Lund University.

Eriksen L, Galucio AV (2014) The Tupí expansion. In: O'Connor L, Muysken P (eds) The native languages of South America: origins, development typology. CUP, Cambridge, pp. 177–202

Evans C, Meggers BJ (1960) Archeological investigations in British Guiana, vol. 177. Smithsonian Institution, Washington

Fort J (2012) Synthesis between demic and cultural diffusion in the Neolithic transition in Europe. Proc Natl Acad Sci USA 109:18669–18673. https://doi.org/10.1073/pnas1200662109

Fort J (2015) Demic and cultural diffusion propagated the Neolithic transition across different regions of Europe. J R Soc Interface 12:20150166

Fort J, Pujol T, Vander Linden M (2012) Modelling the Neolithic transition in the near east and Europe. American Antiquity 77:203–219

French J, Riris P, Fernandez Lopez de Pablo J, Lozano S, Silva F (2020) A manifesto for palaeodemography in the 21st century. Proc R Soc B https://doi.org/10.1098/rstb.2019.0707

Gassón R (2002) Orinoquía: the archaeology of the Orinoco River basin. J World Prehist 16:237–311

Gildea S (2003) Proposing a new branch for the Cariban language family. Amerindia 28:1–26

Gkiasta M, Thembi R, Shennan S et al. (2003) Neolithic transition in Europe: the radiocarbon record. revisited. Antiquity 77:45–62

Gomes DMC (2008) Cotidiano e Poder na Amazônia Pré-Colonial. Edusp, São Paulo

Hazelwood L, Steele J (2004) Spatial dynamics of human dispersals: constraints on modelling and archaeological validation. J Archaeol Sci 31:669–679

Heckenberger MJ, Neves EG, Petersen J (1998) De onde surgem os modelos? As origens e expansões Tupi na Amazônia Central. Rev Antropol 41:69–96

Heckenberger MJ, Neves EG (2009) Amazonian archaeology. Annu Rev Anthropol 38:251–266

Henderson DA, Baggaley AW, Shukurov A et al. (2014) Regional variations in the European Neolithic dispersal: the role of the coastlines. Antiquity 88:1291–1302

Hill JD, Santos-Granero F (eds) (2010) Comparative Arawakan histories: rethinking language family and culture area in Amazonia. University of Illinois Press, Champaign

Hogg AG, Heaton TJ, Hua Q et al. (2020) SHCal20 Southern Hemisphere calibration, 0−55,000 years cal BP. Radiocarbon 62:1–20

Hornborg A (2005) Ethnogenesis regional integration and ecology in prehistoric Amazonia. Curr Anthropol 46:589–620

Hurault J-M (1972) Francais et Indiens en Guyane 1604–1972. Union Générale d'Éditions, Paris

Iriarte J, Smith RJ, Gregorio de Souza J et al. (2017a) Out of Amazonia: Late-Holocene climate change and the Tupi–Guarani trans-continental expansion. Holocene 27:967–975

Iriarte J, DeBlasis P, De Souza JG et al. (2017b) Emergent complexity changing landscapes and spheres of interaction in southeastern South America during the middle and late Holocene. J Archaeol Res 25:251–313

Isern N, Zilhão J, Fort J et al. (2017) Modeling the role of voyaging in the coastal spread of the Early Neolithic in the West Mediterranean. Proc Natl Acad Sci USA 114:897–902

Kistler L, Maezumi SY, De Souza JG et al. (2018) Multiproxy evidence highlights a complex evolutionary legacy of maize in South America. Science 362:1309–1313

Koenker R (1994) Confidence intervals for regression quantiles. In: Mandl P, Huskova M (eds) Asymptotic statistics. Springer-Verlag, New York, pp. 349–359

Koenker R (2005) Quantile regression. Cambridge University Press, Cambridge

Koenker R (2020) quantreg: quantile regression R package version 5.67. https://CRANR-projectorg/package=quantreg

Koenker RW, D'Orrey V (1987) Computing regression quantiles. Appl Stat 36:383–393

Lathrap DW (1970) The upper Amazon. Thames and Hudson, London

Meltzer DJ (2009) First peoples in a new world: colonizing Ice Age America. University of California Press, Berkeley

Meggers BJ (1971) Amazonia: man and culture in a counterfeit paradise. University of Texas Press, Atherton

Meggers BJ, Evans C (1958) Archaeological evidence for a prehistoric migration from the Río Napo to the mouth of the Amazon. University of Arizona Press, Tucson

Meira S, Franchetto B (2005) The southern Cariban languages and the Cariban family. Int J Am Linguist 71:127–192

Mello AAS (2000) Estudo histórico da família linguística tupi-guarani: aspectos fonológicos e lexicais. PhD thesis, Universidade Federal de Santa Catarina.

Mello AAS, Kneip A (2017) Novas evidências linguísticas (e algumas arqueológicas) que apontam para a origem dos povos tupi-guarani no leste amazônico. Lit Lingüíst 36:299–312

Métraux A (1928) La civilisation matérielle des tribus Tupí-Guaraní. Paul Geuthner, Paris

Moraes C, Neves EG (2012) O ano 1000: adensamento populacional interação e conflito na Amazônia Central. Amaz-Rev Antropol 4:122–148

Mühlen GS, Alves-Pereira A, Carvalho CRL et al. (2019) Genetic diversity and population structure show different patterns of diffusion for bitter and sweet manioc in Brazil. Genet Resour Crop Evol 66:1773–1790

Navarrete R (2008) The prehistory of Venezuela—not necessarily an intermediate area. In: Silverman H, Isbell WH (eds) Handbook of South American archaeology. Springer, New York, pp. 429–458

Neves EG (2011) Archaeological cultures and past identities in the pre-Colonial Central Amazon. In: Hornborg A, Hill JD (eds) Ethnicity in ancient Amazonia: reconstructing past identities from archaeology, linguistics, and ethnohistory. University of Colorado Press, Boulder, pp. 31–56

Noelli FS (1998) The Tupi: explaining origin and expansions in terms of archaeology and historical linguistics. Antiquity 72:648–63

Noelli FS (2008) The Tupi expansion. In: Silverman H, Isbell WH (eds) Handbook of South American archaeology. Springer, New York, pp. 659–670

Noelli FS, Brochado JPP, Corrêa AA (2018) A linguagem da cerâmica Guaraní: sobre a persistência das práticas e materialidade (parte 1). Rev Bras Linguíst Antropol 10:167–200

North BV, Curtis D, Sham PC (2002) A note on the calculation of empirical P values from Monte Carlo procedures. Am J Hum Genet 71:439–441. https://doi.org/10.1086/341527

O'Hagan Z, Chousou-Polydouri N, Michael L (2019) Phylogenetic classification supports a Northeastern Amazonian Proto-Tupí-Guaraní homeland. Liames 19:1–29

Oliver JR (1989) The archaeological linguistic and ethnohistorical evidence for the expansion of Arawakan into Northwestern Venezuela and Northeasthern Colombia. PhD thesis, University of Illinois Illinois

Perreault C (2012) The pace of cultural evolution. PLoS ONE 7:e45150. https://doi.org/10.1371/journal.pone.0045150

Pinhasi R, Fort J, Ammerman AJ (2005) Tracing the origin and spread of agriculture in Europe. PLoS Biol 3:e410

Piperno DR (2011) The origins of plant cultivation and domestication in the New World tropics: patterns process and new developments. Curr Anthropol 52: S453–S470

Posth C, Nakatsuka N, Lazaridis I et al. (2018) Reconstructing the deep population history of Central and South America. Cell 175:1185–1197

Reimer PJ, Austin WE, Bard E et al. (2020) The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). Radiocarbon 62:725–757

Riede F, Hoggard C, Shennan S (2019) Reconciling material cultures in archaeology with genetic data requires robust cultural evolutionary taxonomies. Palgrave Commun 5:55. https://doi.org/10.1057/s41599-019-0260-7

Rodrigues A, Cabral AS (2012) Tupían. In: Campbell L, Grondona V (eds) The indigenous languages of South America. Mouton de Gruyter, Berlin, p 495–574

Roosevelt AC (1980) Parmana: Prehistoric Maize and manioc subsistence along the Amazon and Orinoco. Academic Press, New York

Roosevelt AC (1997) The excavations at Corozal Venezuela: stratigraphy and ceramic seriation. Department of Anthropology and Peabody Museum Yale University, New Haven

Rostain S (2008) The archaeology of the Guianas: an overview. In: Silverman H, Isbell WH (eds) Handbook of South American archaeology. Springer, New York, pp. 279–302

Rostain S (2013) Islands in the rainforest. Left Coast Press, Walnut Creek

Rothhammer F, Dillehay TD (2009) The late Pleistocene colonization of South America: an interdisciplinary perspective. Ann Hum Genet 73:540–549

Rouse I (1958) The inference of migrations from anthropological evidence. University of Arizona Press, Tucson

Rouse I, Cruxent JM (1963) Venezuelan archaeology. Yale University Press, New Haven

Russell T, Silva F, Steele J (2014) Modelling the spread of farming in the Bantu-speaking regions of Africa: an archaeology-based phylogeography. PLoS ONE 9:e87854. https://doi.org/10.1371/journalpone0087854 .

Shennan S, Downey SS, Timpson A et al. (2013) Regional population collapse followed initial agriculture booms in mid-Holocene Europe. Nat Commun 4:2486. https://doi.org/10.1038/ncomms3486

Shepard GH Jr, Neves E, Clement CR, Lima H, Moraes C, dos Santos GM (2020) Ancient and traditional agriculture in South America: tropical lowlands. In: Shugart HH (ed) Oxford research encyclopedia of environmental science. OUP, Oxford

Silva F (2020) A probabilistic framework and significance test for the analysis of structural orientations in skyscape archaeology. J Archaeol Sci 118:105138. https://doi.org/10.1016/jjas2020105138

Silva F, Steele J (2014) New methods for reconstructing geographical effects on dispersal rates and routes from large-scale radiocarbon databases. J Archaeol Sci 52:609–620. https://doi.org/10.1016/jjas201404021

Silva F, Stevens CJ, Weisskopf A et al. (2015) Modelling the geographical origin of rice cultivation in Asia using the rice archaeological database. PLoS ONE 10: e0137024. https://doi.org/10.1371/journalpone0137024

Silva F, Vander Linden M (2017) Amplitude of travelling front as inferred from 14C predicts levels of genetic admixture among European early farmers. Sci Rep 7:11985

Silva FA, Noelli FS (2017) Arqueologia e linguistica: construindo as trajetórias históricos-culturais dos povos. Tupí Crítica Soc 7:55–87

Steele J (2010) Radiocarbon dates as data: quantitative strategies for estimating colonization front speeds and event densities. J Archaeol Sci 37:2017e2030

Steward JH (1948) Culture areas of the tropical forests. In: Steward JH (ed) Handbook of South American Indians. Bureau of American ethnology bulletin 143. Smithsonian Institution, Washington, pp. 883–899

Tarble K, Zucchi A (1984) Nuevos datos sobre la Arqueología Tardía del Orinoco: La Serie Valloide. Acta Cient Venez 35:434–445

Taylor DM, Rouse I (1955) Linguistic and archaeological time depth in the West Indies. Int J Am Linguist 21:105–115

Timpson A, Colledge S, Crema E et al. (2014) Reconstructing regional population fluctuations in the European Neolithic using radiocarbon dates: a new case-study using an improved method. J Archaeol Science 52:549–557

van den Bel MM (2015) Archaeological investigations between Cayenne Island and the Maroni river: a cultural sequence of western coastal French Guiana from 5000 BP to present. Sidestone Press, Leiden

Versteeg A (2008) Barrancoid and Arauquinoid Mound Builders in Coastal Suriname./. In: Silverman H, Isbell WH (eds) Handbook of South American archaeology. Springer, New York, pp. 303–318

Williams AN (2012) The use of summed radiocarbon probability distributions in archaeology: a review of methods. J Archaeol Sci 39:578–589

Zarrillo S, Gaikwad N, Lanaud C et al. (2018) The use and domestication of Theobroma cacao during the mid-Holocene in the upper Amazon. Nat Ecol Evol 2:1879–1888

Zucchi A (1985) Evidencias arqueológicas sobre grupos de posible lengua Caribe. Antropológica 63-64:23–44

Zucchi A (2002) A new model of the Northern Arawak expansion. In: Hill JD, Santos-Granero F (eds) Comparative Arawakan histories: rethinking language family and culture area in Amazonia. University of Illinois Press, Urbana, pp. 199–221

Acknowledgements

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-021-00717-w.

Correspondence and requests for materials should be addressed to P.R.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.