



# Interpreting Past Human Diets Using Stable Isotope Mixing Models

Christina Cheung<sup>1,2</sup> · Paul Szpak<sup>3</sup>

Accepted: 26 October 2020/Published online: 16 November 2020  
© Springer Science+Business Media, LLC, part of Springer Nature 2020

## Abstract

Palaeodietary reconstruction using stable isotope analysis is becoming increasingly common, as is the practice of using mixing models to quantify ancient dietary compositions. However, many archaeologists may be unaware of the complexities and pitfalls of stable isotope mixing models (SIMMs). This study serves to provide an overview of the basic principles of SIMMs, evaluates the performances of several of the most commonly used SIMM software packages, and offers some field-specific guidelines for the application of SIMMs in archaeological contexts. We present a series of simulated and published archaeological data to demonstrate and evaluate the different types of SIMMs. We compared the outputs of linear mixing models, simple probabilistic models (IsoSource), and conditional probabilistic models (FRUITS and MixSIAR). Our results show that each mixing model has its pros and cons, and archaeologists should select the best model based on a number of factors, including familiarity with coding languages, sample characteristics (*i.e.* sample size and normality) of the consumer groups, and research questions.

**Keywords** Stable isotopes · Palaeodietary reconstruction · Mixing models

---

✉ Christina Cheung  
christina.t.t.cheung@gmail.com

<sup>1</sup> EA – Eco-anthropologie (UMR 7206), Muséum National d'Histoire Naturelle, CNRS, Université Paris Diderot, Paris, France

<sup>2</sup> UMR 7269, CNRS, Minist Culture, LAMPEA, Aix Marseille Université, Aix-en-Provence, France

<sup>3</sup> Department of Anthropology, Trent University, 1600 West Bank Drive, Peterborough, Ontario K9L 0G2, Canada

## Introduction

The practice of measuring the ratios of different stable isotopes of the same element in body tissues to answer various physiological questions, including nutrition-related ones, began as early as the 1930s (Krogh 1937; Schoenheimer and Rittenberg 1940). The use of stable isotope analysis in archaeology to reconstruct palaeodietary practices, however, did not start until the 1960s (Hall 1967), and by the late 1990s, it had become routine in archaeological research (reviewed by Katzenberg 2008; Schoeninger and Moore 1992). A recent study conducted by Szpak *et al.* (2017) inventoried all the published archaeological studies involving original bulk stable carbon ( $\delta^{13}\text{C}$ ) and nitrogen ( $\delta^{15}\text{N}$ ) isotope analysis between 1980 and 2015. Their survey revealed that the number of publications has exponentially increased from only 15 between 1980 and 1989 to 255 between 2010 and 2015.

Early studies chiefly utilized stable carbon isotope analysis to detect the consumption of certain isotopically distinctive resources, such as maize or marine foods, in past diets (Bender *et al.* 1981; Chisholm *et al.* 1983; Hobson and Collier 1984; Lynott *et al.* 1986). This approach was revolutionary because isotopic analyses provided a means of assessing the relative importance of different foods in the diet based on what was actually ingested and incorporated into the body tissues. This technique avoided many of the problems associated with assessing past human diets based on faunal abundance, macrobotanical remains, material culture, representations in artwork, and dental microwear. It also opened up the possibility of looking at the diet of specific individuals or segments of a population (*e.g.* dietary differences with sex, age, or status). An important question that has loomed large in the field of isotopic anthropology has been, how quantitative can we be in our estimation of past human diet? Are we limited to asking “what were they eating” in the broadest sense, or can we really ask “how much of this were they eating”, and even “how confident are we in this estimation”?

Using isotopic data from mixtures (consumer tissues) and sources (foods), mixing models produce estimates of the relative contribution of these foods to the consumer’s diet. The approach of using stable isotopic measurements to quantify dietary compositions was first developed and applied in ecological studies using the basic concepts of isotopic mass balance with a single isotope and two distinct end members (Haines 1976; McCaunaghey and McRoy 1979). Despite being used as early as 1977 in archaeology (Vogel and van der Merwe 1977), these models have been largely underutilized by bioarchaeologists until relatively recently. Stable isotope mixing models (SIMMs) use measured stable isotope values of both consumers (humans in the context of this paper) and food sources to estimate the contribution of each dietary source to the consumer’s diet. The nature of samples obtained from the archaeological record creates some unique issues with the application of these methods, with some suggesting that these models are “simply not appropriate for archaeological material” (Richards *et al.* 2006:443). Based on anecdotal evidence, many bioarchaeologists either fall into this camp (mixing models should be avoided in archaeology) or are undeterred by the “black box” aspect of the mixing models in their search for quantitative reconstructions of diet.

The main objectives of this paper are as follows:

- 1) Provide an overview of several of the most commonly used mixing models.

- 2) Use four case studies (both simulated and archaeological) to illustrate the strengths and weaknesses of each of these models.
- 3) Outline some features that archaeologists can use—especially for those who are not familiar with Bayesian statistics.

## Isotopic Mixing Models

An isotopic mixing model, in the context of palaeodietary reconstruction, is an analytical model that allows bioarchaeologists to mathematically determine the relative contribution of different food sources to the diets of a particular organism, based on the isotopic compositions of both the consumers and the sources. In archaeology, bone collagen stable carbon ( $\delta^{13}\text{C}$ ) and nitrogen ( $\delta^{15}\text{N}$ ) isotope analyses are the most widely used method for this purpose. Therefore, this study will focus on SIMMs applied to archaeological bone collagen, although the basic principles of isotopic mixing could be applied to other elements (e.g.  $\delta\text{D}$ ,  $\delta^{18}\text{O}$ ,  $\delta^{34}\text{S}$ , or  $^{87}\text{Sr}/^{86}\text{Sr}$ ) or components of body tissues (e.g. bioapatite, keratin, or even individual amino acids).

There are many types of mixing models, although the underlying principles are largely similar. We provide a brief description of each of the four major mixing models that have been used in archaeology. The purpose of this study is to compare the different mixing models' performances from the perspective of end users, and therefore, we will not attempt to comment on, or criticize the underlying mathematical principles of these models. For the sake of consistency, we adopt the mathematical notations used by Parnell *et al.* (2013): for number of consumers  $n_a$ ,  $a = 1, 2, \dots, N$ , for number of isotope tracers  $j_b$ ,  $b = 1, 2, \dots, J$ , and for number of dietary sources  $k_c$ ,  $c = 1, 2, \dots, K$ .

All models utilized to reconstruct diet require some adjustment to account for the trophic discrimination factor (TDF)—which is the isotopic offset between food and consumer. TDF is conventionally accepted as approximately + 1.0‰ for collagen<sub>source</sub> to collagen<sub>consumer</sub> comparisons (Bocherens and Drucker 2003; DeNiro and Epstein 1978; Schoeninger and DeNiro 1984) and + 4.5 to + 6.0‰ for consumed tissue<sub>source</sub> to collagen<sub>consumer</sub> for  $\delta^{13}\text{C}$  (Ambrose and Norr 1993; Krueger and Sullivan 1984; van der Merwe 1982). For  $\delta^{15}\text{N}$  values, an apparent discrimination of + 3.0 to + 6.0‰ per trophic level is generally used (DeNiro and Epstein 1981; Hedges and Reynard 2007; O'Connell *et al.* 2012), with a value of + 3.4‰ being the most common across a range of tissue types and taxa (Post 2002). A controlled feeding study examining TDFs in human bone collagen is logistically and ethically impossible. Thus, for demonstrative purpose, this study will use the generally accepted TDF values for mammals of + 1.0‰ for  $\Delta^{13}\text{C}_{\text{collagen-collagen}}$  (+ 5‰ for  $\Delta^{13}\text{C}_{\text{collagen-edible tissue}}$ ) and + 4‰ for  $\Delta^{15}\text{N}_{\text{collagen-collagen}}$  and  $\Delta^{15}\text{N}_{\text{collagen-edible tissue}}$  throughout (Table 1). The post-trophic discrimination adjusted food sources are represented by the notations ' $k_1$ ', ' $k_2$ ', ..., ' $k_K$ '.

## Linear Mixing Models (LMMs)

The most basic type of all mixing models, a simple linear mixing model can be run with one or two isotope tracers (e.g.  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$ ). It uses a geometric approach to calculate the Euclidean distances between the means of the sources and the consumers. The distance between the consumers and the respective dietary source is

**Table 1** Trophic discrimination factor (TDF) adopted throughout this study

Comparison	$\Delta^{13}\text{C}$ (‰)	SD	$\Delta^{15}\text{N}$ (‰)	SD
Edible tissue to consumer bone collagen	+5.00	0.63	+4.00	0.74
Animal bone collagen to consumer bone collagen	+1.00	0.63	+4.00	0.74

assumed to be inversely proportional to its relative contribution. Equation 1 is a modified version of the formula provided by Ben-David *et al.* (1997); (see also Kline Jr *et al.* 1993):

$$\%k_c \text{ in diet} = \left( \frac{\frac{1}{N'k_c}}{\frac{1}{N'k_1} + \frac{1}{N'k_2} + \frac{1}{N'k_3}} \right) \times 100 \quad (1)$$

in which  $k_c$  represents one of the sources  $k_1$ ,  $k_2$ , or  $k_3$ .

While this model is mathematically sound, it obviously cannot account for the complexity of not only the feeding behaviours of complex organisms, but also the biochemical processes in which foods are incorporated into a consumer's tissues. One key concern is that in a dual-isotope system/model, the incorporation of the two isotopes might not be a linear process. As early as 1991, an improved approach of using LMMs to reconstruct dietary contributions was proposed by Schwarcz (1991), where the contributions of each source ( $K$ ) to each isotope tracer ( $J$ ) are considered separately. For a dual isotope mixing system, the proportion  $P_{kK}$  can be determined with Equation 2:

$$\begin{bmatrix} \delta_{j1}^N = P_{k1} \cdot \delta_{j1}^{k_1} + P_{k2} \cdot \delta_{j1}^{k_2} + P_{k3} \cdot \delta_{j1}^{k_3} \\ \delta_{j2}^N = P_{k1} \cdot \delta_{j2}^{k_1} + P_{k2} \cdot \delta_{j2}^{k_2} + P_{k3} \cdot \delta_{j2}^{k_3} \\ P_{K1} + P_{K2} + P_{K3} = 1 \end{bmatrix} \quad (2)$$

in which  $\delta_{jJ}^N$  represents each of the isotopic measurements of the consumer  $N$  with the isotope tracer  $J$ ,  $\delta_{j1}^{k_1}$  represents the isotopic measurement of the TDF—corrected food source  $'k_K$ , and  $P_{kK}$  represents the proportion of each respective source in consumer  $N$ 's diet.

This model still proves to be ineffective in accounting for all the complexity that surrounds the incorporation of isotopes into body tissues. Furthermore, both of these models are limited to providing a single solution for  $J + 1$  sources, where  $J$  represents the number of isotope tracers measured. In situations where the diet of the consumers would consist of more than three isotopically distinct source groups, these models cannot provide a solution. Additionally, when these models do provide a solution, they have the disadvantage of providing only a *single* solution, which fails to account for variation in source isotopic compositions and TDFs. Phillips and Koch (2002) put forth a more complex model, taking into consideration the variations in elemental concentrations among different food resources. This is particularly important when it comes to reconstructing the diets of omnivorous organisms with complex diets, as different types

of foods can contain vastly different proportions of elemental components. While the ratios between carbon and nitrogen in animal proteins are largely comparable, there can be large differences in the percentage of protein in animal and plant products and the relative proportion of protein and lipids in animal products (Newsome *et al.* 2004). Thus, in a dual isotope system using  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values as tracers, simple LLMs can only provide an estimation for the proportion of animal vs. plant *protein* in a consumer's diet, but not an estimation of the proportion of animal vs. plant *biomass*. In order to account for the different elemental concentrations in food sources, Phillips and Koch (2002) introduced an improved concentration-weighted model (Eq. (3)). Building upon Equation (2), this model allowed the input of the different concentrations of each respective element (isotopic tracer) in each source group. With this model, it is possible to calculate the proportion of a source ( $K$ ) as biomass ( $\theta$ ) for each dietary source group ( $k_K$ ), provided the food contributed some elements of the isotope tracer(s) ( $j$ , e.g. contained carbon or nitrogen) being analysed in the consumer's tissue:

$$P_{kK} = \frac{k_1^{\alpha(j,j)} \times k_1^{\theta}}{\left(k_1^{\alpha(j,j)} \times k_1^{\theta}\right) \times \left(k_2^{\alpha(j,j)} \times k_2^{\theta}\right) \times \left(k_3^{\alpha(j,j)} \times k_3^{\theta}\right)} \quad (3)$$

The main advantage of LLMs is that they are very straightforward. Vogel and van der Merwe (1977) were among the first to demonstrate how LLMs can help to detect, and subsequently quantify, the consumption of  $\text{C}_4$  plants by ancient populations. Following Schwarcz, archaeological studies started to employ different versions of LLMs to quantify palaeodietary proportions (Bocherens *et al.* 2006; Finucane 2009; Pate 1998; Schwarcz 1991), though the overall palaeodietary interpretations tended to be qualitative. Because of the simplicity of this approach, it is still commonly used, for example, to quantify the relative inputs of two very distinctive sources such as  $\text{C}_3$  vs.  $\text{C}_4$  sources (Guo *et al.* 2011; Hu *et al.* 2006) or plant vs. animal protein (Tomczyk *et al.* 2020).

Limited by the mathematical rules of linear equations, LLMs are only able to work with  $K \leq (J + 1)$  sources, which means that in a dual isotope system, LLMs can only handle three sources. Thus, LLMs are good for simple systems, but not powerful enough to deal with more complex (biologically and socially) systems. LLMs also provide quantification without any indication of the uncertainty on these outputs, and a single, discrete numerical solution presents a false sense of certainty with respect to the interpretations of diet composition.

### Simple Probabilistic Models

Building upon the mathematical concept of LLMs, simple probabilistic modelling provides a more sophisticated approach to the reporting of the proportional contribution of dietary sources in a consumer's diet. This type of model is able to provide a finite number of solutions to the mixing system by providing constraints on the  $J + 1$  model and running iterations through all possible combination of source contributions for a specified increment (e.g. 0.00, 0.01, 0.02, ..., 1.00) (Phillips and Gregg 2003). There are several advantages of simple probabilistic models compared to LLMs: (i) with a probabilistic model, it is possible to consider more than  $J + 1$  sources (*i.e.* it is possible

to work with an underdetermined system); (ii) potential source proportions are given as ranges, instead of single values, thus providing a more realistic estimation. An example of a program that uses such a model to help quantify proportions of source groups is IsoSource (<https://www.epa.gov/eco-research/stable-isotope-mixing-models-estimating-source-proportions>), which has been used in numerous archaeological studies (Bocherens *et al.* 2005; Newsome *et al.* 2004; Yoneda *et al.* 2004). However, this type of simple probabilistic model does not consider source concentration, and therefore, when modelling with bone collagen isotopic compositions, they can only provide dietary reconstruction in terms of dietary protein.

### Conditional Probabilistic Models—Bayesian Mixing Models (BMMs)

BMMs are complex mathematical models that compute/address the conditional probabilities of an event. Instead of giving a density distribution of all the possible solutions, BMMs assess the *posterior probabilities* of source proportions in a mixing system. The *posterior probability* is a key Bayesian concept that describes the probability of a certain condition/pattern under prescribed parameters, which are composed of two elements: observation (*prior distribution*) and other additional information (*likelihood*):

$$\text{Posterior} \propto \text{Prior distribution} \times \text{Likelihood} \quad (4)$$

In the case of dietary reconstruction using stable isotope measurements, the *prior distribution* typically refers to the three basic premises of mixing models: isotopic values of the consumer(s), isotopic values of the source groups, and TDFs. The *likelihood* can refer to any other relevant biological or social information and relationships, as long as they can be codified in mathematical terms, such as values, formulae, or ratios. What sets BMMs apart from LMMs is the inclusion of the component “likelihood”. By also considering these factors, BMMs allow the user to produce better-informed inferences about an organism’s diet.

In BMMs, users may manipulate the *likelihood* component by modifying different parameters, also known as *priors* in some literature (but not to be confused with *prior distribution*). *Priors* are optional in that their absence still allows BMMs to generate estimations for dietary compositions. The incorporation of these priors, however, allows us to acknowledge the complex nature of human subsistence practices and the uncertainties associated with inferences made about them using stable isotope analysis. In turn, these models provide more realistic outputs. These complications may include sampling error, natural variability in inter-individual metabolic rate and/or physiology, and various anthropogenic factors (Stock *et al.* 2018; Stock and Semmens 2016b). Some of the more basic *priors* allow the input of variances or uncertainties for the various model parameters, such as source variances ( $\omega$ ) and trophic discrimination factor variances. Others are slightly more specific, such as hyperparameters ( $\alpha$ ). The hyperparameter<sup>1</sup> is an important notion in BMMs and therefore merits some discussion. It allows the user to incorporate other lines of evidence into the mixing model, by informing the model of the presumed contribution of each source group to the

<sup>1</sup> Hyperparameters ( $\alpha$ ) are also called *priors* in FRUITS and in the literature more generally. In order to differentiate this from other *priors*,  $\alpha$  will be referred to as hyperparameters throughout this paper.

consumers' diets. For example, a well-established theory in archaeology is that most post-Neolithic human populations subsisted on at least 50% plant foods (Larsen 2003; Nestle 1999). Given this supposition, we could instruct a BMM to consider the input of plant foods more heavily, for example, that they must contribute > 50% to the diet. In archaeology, hyperparameters could be informed by zooarchaeological evidence, palaeobotanical evidence, written sources, and coprolite contents.

Currently there are several major software packages that employ BMMs for dietary reconstruction. The two most widely used in archaeology are Food Reconstruction Using Isotopic Transferred Signals (FRUITS) (Fernandes *et al.* 2014) and MixSIAR (formerly MixSIR and SIAR) (Parnell *et al.* 2010; Stock *et al.* 2018; Stock and Semmens 2016a). Archaeological examples using FRUITS include a Neolithic site in Carding Mill Bay, Scotland (Bownes *et al.* 2017), and a Neolithic site in Ostorf, Germany (Fernandes *et al.* 2015). Examples involving SIAR (one of the predecessors to MixSIAR) include an early Bronze Age site in Anyang, China (Cheung *et al.* 2017a), and a Pre-contact site in Matanzas, Cuba (Chinique de Armas *et al.* 2015).

Both FRUITS and MixSIAR use a Markov chain Monte Carlo (MCMC) method to derive the most likely logistic model from the probability distribution. The major difference between the two tools is that being a package in R, MixSIAR is by nature more flexible, programmable, and compatible with other statistical packages than FRUITS, which is a self-contained software package. While FRUITS is less customizable, the more constrained environment means that even the less statistically inclined users with no experience in coding or programming can navigate the system relatively easily, with fewer opportunities for user error.

Below, we provide a brief description of some of the more advanced features available in FRUITS and MixSIAR, respectively. Note that as MixSIAR was not initially designed for archaeological applications, many features are not necessarily applicable to archaeological contexts. Therefore, if used inappropriately, these features can add unnecessary complications to the results. Archaeologists will have to use their own discretion to decide whether or not a certain feature is suitable for their case studies.

### Accounting for Different Proportions of Macronutrients in Food

One feature that FRUITS offers is the ability to account for the proportions of macronutrients (*i.e.* lipids, carbohydrates, protein) in different foods (Fernandes *et al.* 2014). While bone collagen isotopic compositions mostly reflect the isotopic signature of dietary proteins, the isotopic signature of bioapatite should reflect those of the bulk diet (Ambrose and Norr 1993; Fernandes *et al.* 2012; Krueger and Sullivan 1984; Lee-Thorp *et al.* 1989). Certain non-essential amino acids from collagen could also be routed from sources other than dietary proteins; however, the conditions under which this happens are complicated (Jim *et al.* 2006). Thus, this feature might be of particular interest to those working with bioapatite or compound specific isotope analysis.

There are several caveats concerning this feature. First, the isotopic compositions of the sources' various macronutrients need to be known. This is not always possible in archaeological contexts, as components such as lipids do not preserve well in bones and teeth (Scott 2020). While fatty acids do preserve in ceramic residues, the isotopic composition of these samples cannot be associated with a particular animal taxon. Of



course, values can be approximated from modern proxies; this approach, however, is not without problems (see critique in (Szpak 2014)). Secondly, the formulae used to calculate the proportion of each macronutrient in diet are based on the assumption that the examined diet is a “healthy” one—one that satisfy the minimum acceptable intake of protein to maintain basic body functions (Fernandes *et al.* 2014). Thus, if the users have any doubt about the health of the studied individuals (*e.g.* apparent signs of nutritional stress), or that there is reason to suspect that the studied population may have different nutritional requirements than the average modern human, they would need to use this feature with caution. Nevertheless, if all the precautions are considered and the isotopic measurements of macronutrients are available, this feature can provide added dimension to the reconstruction of palaeodiets, especially to groups with specialized diets (*e.g.* lipid or carbohydrate-rich diets).

## Error Structure

Error structure is an important aspect of MixSIAR and relates to a manner in which the model accounts for a certain level of inherent uncertainty that is independent of variation in source proportions. MixSIAR users can choose to consider a mixing model under different error structures: process error only, residual error only, or both (*i.e.* process\*residual). These different error structures impact consumer isotopic values in different ways, even if the “same” diet is being consumed (Fig. 1). Figure 1a presents a scenario where no error structure is considered. In this scenario, there is no significant variation within the source groups, and the consumers are physiologically similar. Thus, the differences in the isotopic compositions of the consumers can be entirely explained by the differences of their diets (*i.e.* source proportions).

Process errors are the more “emergent” and “principal mechanistic” factors contributing to variance in the isotopic values of the consumer(s) (Stock and Semmens 2016b). In this scenario, variance in isotopic compositions of the consumers occurs because there is isotopic variation in the sources, and individual consumers may sample from one side or another of those distributions (Fig. 1b). Despite the fact that the three consumers displayed in Fig. 1b consume 50% of source 1 and 50% of source 2, there is some variation in their isotopic compositions.

Residual errors ( $\sigma$ ) are derived from factors other than process errors (*i.e.* from secondary mechanistic factors) that contribute to the variance in the isotopic values of the consumer(s) ( $\delta_{j,j}^N$ ) (Hopkins III and Ferguson 2012; Stock and Semmens 2016b). These residual errors occur at the level of the consumer (Fig. 1c). Again, even though all three consumers in Fig. 1c have the same diet composition *and* all three sample from the middle of the distribution for each source, there is still variation among the consumers caused by individual physiological states that are independent of the diet.

Lastly, Fig. 1d depicts a scenario where both process and residual errors are considered and connected by a multiplicative error term ( $\xi$ ) (Stock and Semmens 2016b). In this scenario, both process and residual errors contribute to the differences in the isotopic compositions of the consumers.

From Fig. 1, it is clear that the “no error” option would be inappropriate for humans. In archaeological contexts, process errors should be more or less minimized because of the long turnover time of bone collagen, which should, more than most tissues,



effectively integrate the true mean isotopic compositions for all sources consumed. However, this may be contingent upon other factors, such as sampling strategies. For example, in a study where only individuals from a certain age and sex profile are sampled, to highlight the fact that most of the differences in the isotopic compositions in this group should stem from differences in dietary practices, a “process error only” structure may be more appropriate. Therefore, depending on each case study and research design, users should experiment with using these three different error structures (*i.e.* “process error only”, “residual error only”, and “process\*residual error”), to decide which is best for their specific case study (see case study 4 for a demonstration). For demonstrative purposes, all MixSIAR examples involved in this study have adopted the third error structure (residual only), unless stated otherwise.

### Intragroup Comparison

MixSIAR allows users to compare the estimated dietary compositions of two or more groups, provided all groups are directly comparable (*i.e.* these groups had access to the *exact* same set of resources, have similar variances and correlations). For example, archaeologists can compare the intra-population variability in dietary compositions between sexes or those buried with different types of grave goods, orientations, and any other applicable groupings. In order to make the comparisons, it is required to first define the categorization of these different groups as either “fixed” or “random” effects, and it is up to the users to decide which categorization better describes the nature of a particular grouping. The “fixed” category generally refers to well-defined groupings, such as sex (male vs. female), continuous effects (*e.g.* changes in diets over time, discussed in more detail below), and any other categorization with only two groups. The “random” category generally refers to more fluid groupings. Instead of treating the subgroups as discrete populations, the “random” grouping considers the subgroups as parts of the larger population, thus allowing the model to share a general parameter between these subgroups. Models run under a “random” categorization will return results showing an overall estimation of the whole population, as well as estimations of each subgroup, respectively. Examples of random effect groupings include social-economic status (*e.g.* number of grave goods), individuals buried in different orientations, and individuals buried in different localities or clusters. Random effect groupings also allow users to specify the hierarchical relationships (if any) among these different groupings (Semmens *et al.* 2009; Semmens *et al.* 2013). For example, in a certain site, burial location may be a stronger determinant of diet than the number of burials goods. As there is no fixed rule as to how different groupings should be defined, caution should be taken when using this feature. Case study 4 provides a demonstration of the different effects of these categorizations.

The continuous effect feature allows the user to examine changes in dietary compositions in consumers responding to a continuous effect, whether intra-individual (over the lifetime of an individual or individuals) or intra-population (over a longer period of time). In the example provided by the developer, the continuous effect in MixSIAR is demonstrated by comparing the dietary compositions of alligators of different lengths, among other criteria (Nifong *et al.* 2015; Stock and Semmens 2016a). This particular approach is obviously not feasible in most archaeological case studies, where remains are usually too fragmented for this kind of categorization.

**Fig. 1** Different error structures used in mixing models: **a** “no error”; **b** “process error only”; **c** “residual error only”; **d** “process\*residual error”. The no error structure assumes that the isotopic variations in the consumer can be entirely explained by variation in source proportions; the process error only structure takes isotopic variations in the sources into consideration; the residual error only structure takes isotopic variations in the consumers into consideration; and in the process\*residual error structure, isotopic variations in both the sources and the consumers are considered

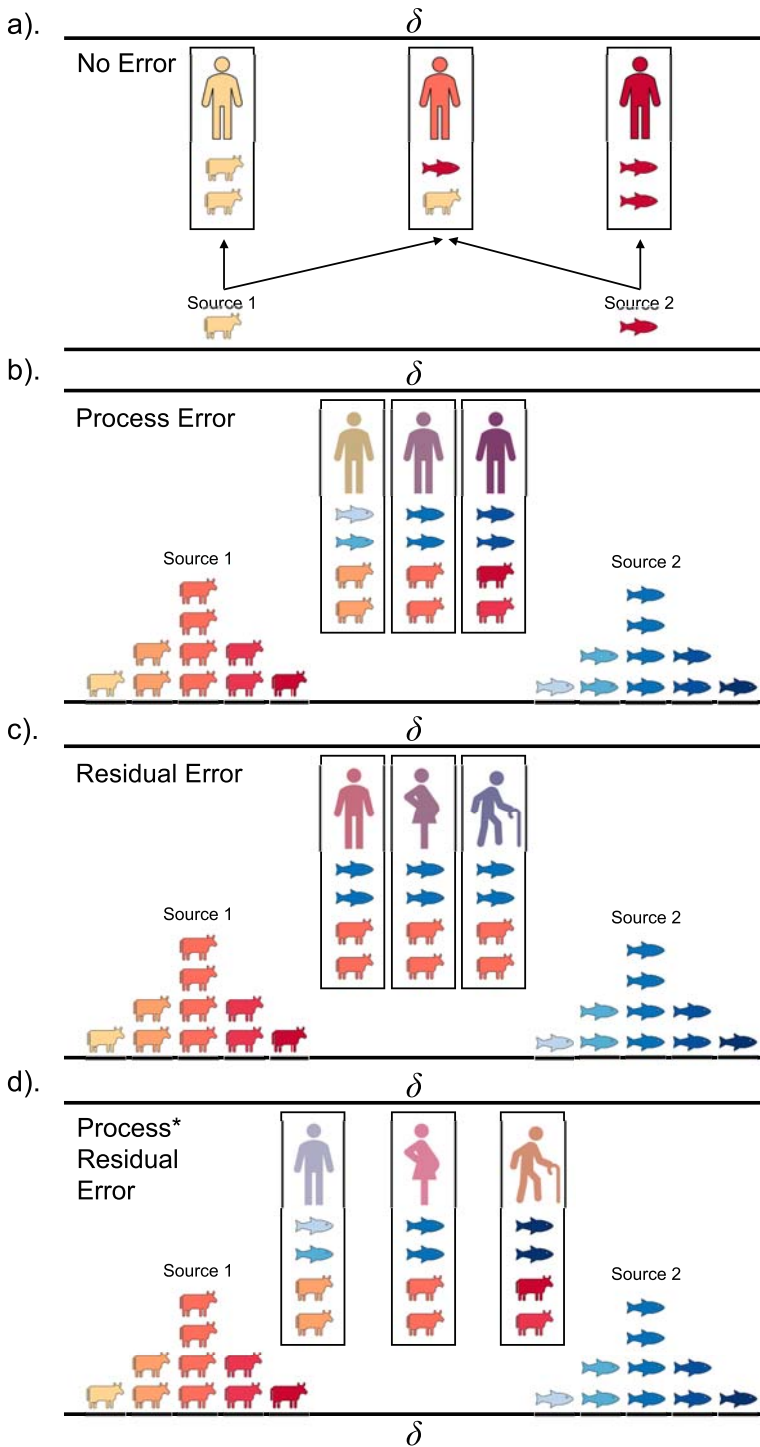
However, this feature may still be useful in other circumstances and could be used to answer questions pertaining to changes in diets through various biological, environmental, and socio-cultural processes such as weaning, seasonal changes, migratory activities, and cultural changes. For example, certain materials such as hair or tooth (especially for animals with larger or continuously growing teeth) may permit a fine-resolution intra-individual time series analysis (Guiry *et al.* 2016). Other examples include studies that have gathered data over a long temporal sequence (Cheung *et al.* 2019; Müldner and Richards 2007). To date, no bioarchaeological study has employed this feature. We have tried to apply this feature on the three aforementioned studies. Unfortunately, none of the attempts yielded particularly informative results in that the model outputs do not add any new insights to these datasets. Figure 2b demonstrates the output of the continuous effect by MixSIAR on Guiry *et al.*'s (2016) pig dentine data, in comparison to a different visualization technique generated by ggplot2 (Wickham 2009) (Figure 2a). Whether MixSIAR's output is more informative or not may vary among studies as well as the particular research questions, but we have attempted to apply this feature to case studies that should be ideally suited and have not found it to be useful. This is not necessarily a comment on the capability of the feature, but it does show that in order to apply this feature to archaeological studies in a meaningful way, extensive experimentation with the programming conditions might be required.

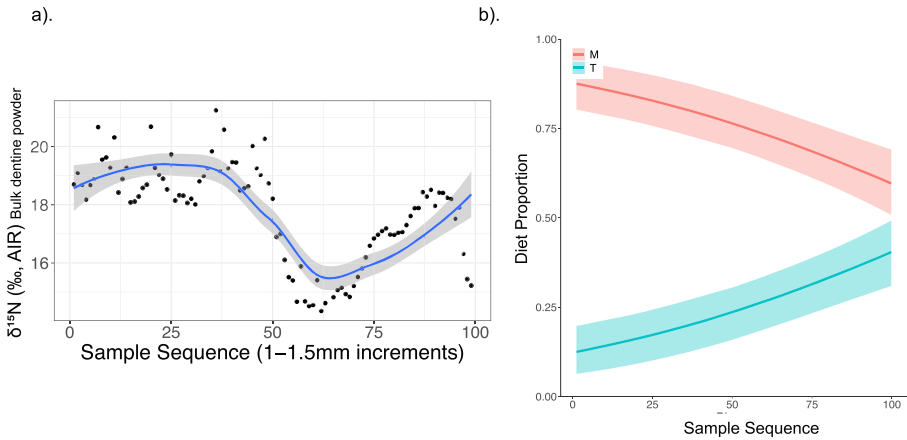
## Source Aggregation

Because all SIMMs lose power with increasing numbers of sources, it is necessary to aggregate or combine multiple sources into a smaller number of source groups. In terms of source aggregation, currently there are two main types: *a priori* or *a posteriori*.

### A Priori

Most archaeologists conduct *a priori* source combining to a certain degree, by comparing species with similar feeding behaviours (*e.g.* seals and porpoises or deer and elk) or by simply “eye-balling” the spatial distribution of species in an isotope scatterplot. An early ecological study recommended using the supervised machine learning “k-nearest neighbour” (k-NN) algorithm to combine sources *a priori* (Rosing *et al.* 1998). However, for the majority of archaeological case studies, there are simply not enough source data to conduct machine learning meaningfully. In view of this, we suggest an alternative approach using a cluster analysis to help determine source groups. This approach involves an initial k means test followed by manual source groupings based on the output of the test. First, the optimal number of clusters can be determined by the function *NbClust()* within the R packages “NbClust” (Charrad *et al.* 2014), “factoextra” (Kassambara and Mundt 2017), and “tidyverse” (Wickham 2017). This function uses up to 30 indices to





**Fig. 2** Continuous effect as seen on the serial measurements of dentine  $\delta^{15}\text{N}$  values from a pig that had a significant dietary shift over its lifetime. **a** A modified figure (ggplot2 output) using raw data from the original published report (Guiry *et al.* 2016). A non-parametric regression analytical tool (LOESS, in the R package ggplot2) is used to create the smoothed regression line for the incremental data, confidence level = 0.95. **b** Plot produced by MixSIAR estimating the proportional input of the two source groups, *M* refers to marine-based protein, *T* refers to terrestrial-based protein. Baseline values of source groups were obtained from another published report (Guiry *et al.* 2012)

identify the best clustering scheme within a given population. After an optimal number of clusters have been selected, users can manually determine the source groups based on an assessment of this output. An example of this process is shown in case study 3 below.

## A Posteriori

This is a less common approach, where a mixing model is run initially with a higher number of aggregated sources, and then the distributions of certain sources can be combined depending on observations and other contextual factors. The sources can be combined manually or in MixSIAR, as the package provides a function “combine\_sources” that helps aggregate the contributing proportions from multiple sources. Note that with this option the user still needs to select manually which sources to combine, with the options to input hyperparameters ( $\alpha$ ). One advantage of using *a posteriori* source aggregation is that it allows users to aggregate contributions of sources with different concentrations, such as those from plants and animals. For example, this would allow users to combine source groups such as  $\text{C}_4$  plants +  $\text{C}_4$ -consuming animals into a larger group “ $\text{C}_4$  sources”, and the software package would still be able to account for the differences in source concentration among the different species within this group.

## Materials and Methods

To illustrate the strengths and weaknesses of various SIMMs, we have tested the three major types of SIMMs: LMMs, simple probabilistic models, and BMMs using four case studies.

## Methods: Selection of Case Studies and Evaluation Methods

The first case study is a hypothetical one, with simulated data designed to illustrate a “best-case scenario” when it comes to reconstructing diets with bone collagen. This case study looks at a hypothetical coastal hunter-gatherer population living in a high-latitude region. There are a few reasons why we chose to use a hypothetical population over a real example as the first case study. The main reason being that there is no systematic way to monitor how different mixing models perform in real-life situations, since the true diets of humans from archaeological contexts are never known. Accordingly, the accuracy of the models cannot be assessed. A coastal, high-latitude setting was specifically chosen as these populations subsist on mostly animal products with very little plant-based food, especially with respect to dietary protein (Nestle 1999), and that the food sources are typically isotopically distinctive.

The second and third case studies are based on actual archaeological data. We selected two sites, *Yinxu* and *Missignac*, to represent an ideal and a less-than-ideal scenario for SIMM applications in archaeology, respectively. Other than abundant isotopic data from both humans and fauna ( $n > 50$ ), some baseline knowledge about the dietary practices at these sites has been established from other archaeological evidence, such as extensive zooarchaeological research, as well as historical records. Even with the multiple lines of evidence, it is not possible to arrive at a “true” diet of these two groups. Thus, the reconstructed proportions calculated by the different methods will be compared to check to what extent the different estimations converge. The purpose of this comparison is not to assess the relative performance of each of these methods (*i.e.* faunal analysis, SIMMs, or historical records). A “good” reconstruction will be judged by how well the different lines of evidence agree with one another, whereas a “bad” reconstruction is one where the different lines of evidence give highly contradictory estimations.

The last case study aims to test the effects of different error structures and hyperparameters on BMMs. The same hypothetical population from the first case study is used. Using simulated data allows us to expand the sample size freely while keeping different parameters under control.

In order to evaluate the “accuracy” of each SIMM estimations with the simulated data, two additional metrics were used: total difference (TD) and average difference (AD). TD refers to the total absolute differences between the “true” contribution and the projected contribution (represented by the mean % of estimation) across all sources. AD refers to the average of the total absolute differences between the “true” contribution and the projected contribution (represented by the mean % of estimation) across all sources. These two different metrics serve to highlight different elements of the model performance: while AD provides a fairer assessment of the overall performance of the SIMMs, TD allows us to visualize the magnitude of total deviation. We used a scoring system to avoid judging the performances based on minute (and rather meaningless) differences between the “true” and projected contributions. Models that reported a TD > 50% and an AD > 25% were regarded acceptable and were given a score between 1 and 5 (Table 2), with better performing models scoring higher.

All LMMs and MixSIAR analyses were conducted using R version 3.6.0 (RStudio Team 2019) with RStudio (RStudio Team 2018). The scripts involved are provided in Online Resource 1. FRUITS analyses were conducted using the version FRUITS 3.0.

**Table 2** Scoring scheme for assessing estimation accuracy of source proportions in SIMMs

Score	TD	AD
5	0–10%	0–5%
4	11–20%	6–10%
3	21–30%	11–15%
2	31–40%	16–20%
1	41–50%	21–25%
0	> 50%	> 25%

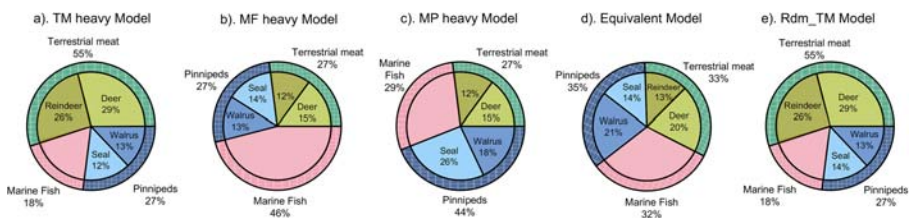
## Materials: Case Studies

### Hypothetical Coastal Population

In this case study, the actual variety of foods consumed by the hypothetical coastal population were reduced to five “categories” to simplify the model. The broader source groups consumed in a high-latitude hunter-gatherer community are all represented, and all isotopic data are informed by actual isotopic measurements of similar taxa. We calculated the isotopic compositions of the consumers by multiplying a proportion (0–1) by the isotopic composition of each source and then summing these values (Eq. (2)). The total proportions of all sources was 1 in each case (Eq. (2)). The source isotopic compositions were kept constant.

In order to examine the performances of all the models, we present five different scenarios in this particular case study (Fig. 3):

1. A terrestrial-meat heavy model (TM): where terrestrial meat (*i.e.* deer and reindeer) contributes around 50% of the group’s diet.
2. A marine fish heavy model (MF): where marine fish contributes around 50% of the group’s diet.
3. A pinniped heavy model (MP): where pinnipeds (*i.e.* seal and walrus) contribute around 50% of the group’s diet.
4. An equivalent model (EQ): where all three major source groups (*i.e.* terrestrial meat, marine fish, and pinnipeds) are consumed in roughly equal amounts.
5. A terrestrial-meat heavy model with random correction: in order to mimic natural variability (*i.e.* process error), a random effect (between – 5% and + 5%) is applied to every individual in the TM model.



**Fig. 3** Pie charts illustrating the dietary proportions of the five hypothetical groups: **a** Terrestrial-meat heavy model; **b** marine fish heavy model; **c** pinniped heavy model; **d** equivalent model; and **e** terrestrial meat heavy model with random correction

As two-tracer LMMs can only handle three sources, a 3-source model was also tested under these five scenarios, where the sources are divided into three groups: terrestrial meats (deer and reindeer), pinnipeds (seal and walrus), and marine fish. We ran all mixing models (*i.e.* simple LMM, isotope-specific LMM, IsoSource, FRUITS, and MixSIAR) with these five scenarios, under two source aggregation schemes (3-source and 5-source), and compared the analytical results with the known proportions. Obviously, this hypothetical case study undermines many physiological, social, and environmental complexities of human dietary habits, and cannot be used to assess the true mathematical power of each mixing model. Thus, the actual proportions from each source group are only used as a reference to establish grounds for discussion. Note that in this case study, carbon and nitrogen source concentrations are not taken into consideration since all sources are animals. All associated data and analytical results associated with this case study are provided in Online Resource 2.

### Yinxu, Shang Dynasty

*Yinxu* (c. 1200–1046 BC) is an early Bronze Age site located in the modern city of Anyang in Henan Province, China. We chose this site because it is well known that people living in the region cultivated millet, a C<sub>4</sub> crop, in a predominately C<sub>3</sub> environment (Bettinger *et al.* 2010; Crawford 2009; Yang *et al.* 2012; Zhou *et al.* 2016). Therefore, the isotopic signature of the different source groups, especially in terms of  $\delta^{13}\text{C}$  values, should be very distinctive.

The consumer data ( $n = 39$ ) in this case study come from one of the many residential neighbourhoods, *Xin'an Zhuang*, in *Yinxu* (Cheung *et al.* 2017b). In terms of sources, the faunal data ( $n = 127$ ) come from a nearby bone workshop at *Tiesanlu*, another residential neighbourhood at *Xiaomintun*, and a sacrificial pit at *Anyang*, all within *Yinxu* (Cheung *et al.* 2017a; Si 2013; Yan 2010). Unfortunately, there is no known published plant isotope data from the site, so the mean isotopic values for C<sub>4</sub> plants are derived from millet grains ( $n = 49$ ) from seven Neolithic sites in Shaanxi province (Wang *et al.* 2018), whereas those for C<sub>3</sub> plants are inferred from global averages. In addition to isotopic values, faunal abundance data are also included to help gauge the performance of various mixing models. The zooarchaeological data come from two collections, one from a bone workshop at *Tiesanlu* (Campbell *et al.* 2011) and another from a domestic context at *Guandimiao*, a Shang period village in Henan province (Hou *et al.* 2019).

Archaeologists have found at least 29 species of mammal, and many different species of fish and bird, at *Yinxu* (Zhu 2005). Dog, horse, pig, deer, cattle, and sheep are the most numerous, among which dog and horse were likely not primarily considered as food animals (Yang and Ma 2010). Campbell *et al.* (2011) observed that the animal assemblage from section 5 of the *Tiesanlu* bone workshop (TSL) was predominately cattle, followed by pig and deer. While the nature of TSL meant that faunal remains discovered there are likely biased towards animals with larger and more workable bones (*e.g.* cattle), the proportion generally agrees well with written records and faunal assemblages from other parts of the site (Chen 1985; Yang and Ma 2010). In order to counter the bias, we compared the frequency of the six most common animal taxa found in TSL with those from domestic contexts in *Guandimiao* (GDM), a contemporaneous and likely subsidiary settlement to *Yinxu* (Hou *et al.* 2019). Cattle



**Table 3** Weight (%) of fauna taxa recovered from TSL and GDM, respectively (Campbell *et al.* 2011; Hou *et al.* 2019)

		Tiesanlu (TSL)	Guandimiao (GDM)
Source group	Species	% by weight	% by weight
C <sub>4</sub> -consuming herbivores	Cattle	64.5%	59.5%
C <sub>4</sub> -consuming omnivores	Pig	13.3%	32.3%
	Dog	--	3.5%
Mixed consumers	Sheep/goat	--	1.6%
	Fish + shellfish	--	Negligible
C <sub>3</sub> -consuming herbivores	Deer	2.8%	3.1%
Others/ indeterminate		19.4%	0.1%
Total		100%	100%

dominated the faunal assemblages from both contexts, followed by pig (Table 3). These proportions were compared with the estimations generated by the SIMMs.

Note that the % weight of dry bones does not necessarily equate to the % of edible tissues available for consumption; the size of the animal is also a factor. Also, it is likely that other than meat, the ancient *Yinxu* residents also consumed bone marrow, offal, and blood, a practice that has continued in modern Chinese cuisine. As it would be quite difficult to extrapolate the edible tissue weight from dried archaeological bone weight, for the purpose of this study, we will consider them roughly equal. In addition, an abundance of contemporary and near-contemporary historical records mentioned that the subsistence economy of *Yinxu* was based heavily on horticultural products. Other than millets, an assortment of cereals such as sorghum, rice, barley, wheat, as well as other fruits and vegetables were also grown (Chang 1976; Yang and Ma 2010). This is further supported by other archaeological evidence such as palaeobotanical remains (Lee *et al.* 2007; Xu *et al.* 2010) as well as farming tools (Yang and Ma 2010). Thus, it is important to also consider the contribution of these plants to the diets of the *Yinxu* inhabitants.

All associated data and analytical results associated with this case study are provided in Online Resource 3.

For the purpose of this study, plants and animals available as food for the *Yinxu* inhabitants were divided into six main source groups: C<sub>4</sub>-consuming herbivores (mostly cattle), C<sub>4</sub>-consuming omnivores (mostly pigs and dogs), C<sub>4</sub> plants (mostly millet), mixed consumers (mostly sheep/goat, some fish), C<sub>3</sub>-consuming herbivores (mostly deer), and C<sub>3</sub> plants (various cereals). In terms of hyperparameters, while there is no published estimate of the proportion of animal vs. plant food consumed at this site, since *Yinxu* is known to have been an agrarian society, we will assume plant foods contributed at least 50% of the total food consumed at *Yinxu*.

### Missignac-Saint Gilles le Vieux, Medieval Period

The third case study serves to highlight the limits of stable isotope mixing models. Missignac-Saint Gilles le Vieux (675-1175 AD) is a medieval rural site located close to the modern town of Aimargues in Gard, Southern France. This site was chosen

specifically because the residents relied heavily on C<sub>3</sub> terrestrial resources, which is typical for most of continental Europe. These dietary habits meant that all the source groups had very similar isotopic signatures, and thus, their relative contributions to the diet should be difficult to differentiate using SIMMs. We have isotopic data from 152 humans and 63 fauna (animals unlikely to have been used for food are excluded) from the site (Mion *et al.* 2018). Similar to the *Yinxu* case study, there is no isotopic data from plant remains from the site. Moreover, historically, plant-based food played a prominent role in Medieval European diets. Even though no direct historical record survived from this site, a review of late Medieval period diets from the region suggested that carbohydrates accounted for a large proportion of a peasant's daily diet (Campech 1996). One study estimated that up to 84.6% of the daily caloric intake of an average peasant came from bread, whereas only 4.6% came from meat (Ladurie 1976). Therefore, plant-based foods such as various cereals and their products (*i.e.* bread) needed to be considered in our SIMMs. In order to provide an isotopic baseline for plant-based food available to the Missignac residents, the mean isotope values of C<sub>3</sub> plants are derived from cereal grains ( $n = 17$ ) from a Bronze Age Greek site (Fraser *et al.* 2013). We verified the appropriateness of using plant data from this site by comparing the cereal values from Greece with those of the domesticated herbivores found from Missignac. Despite the temporal and geographic distances between the two sites, the mean isotope values of the plants ( $\delta^{13}\text{C} = -23.5\text{‰}$  and  $\delta^{15}\text{N} = +3.4\text{‰}$ ) compared reasonably well with those of the cattle ( $\delta^{13}\text{C} = -19.5\text{‰}$  and  $\delta^{15}\text{N} = +7.6\text{‰}$ ) considering the known TDFs for bone collagen (Table 1).

In French archaeology, cattle, pig, and sheep/goat are often collectively referred to as the *triade domestique*, as these are usually the most abundantly recovered domesticated animals. This is the case at Missignac, where sheep/goat dominated the assemblage, and cattle and pig follow closely (Maufras *et al.* 2018). Terrestrial faunal remains come from both the residential (*habitat*) and storage (*aire d'ensilage*) areas and are quantified using NISP (number of identified specimens), where every taxonomically identifiable specimen found is counted. Only species amounting for more than 1% of the total assemblage are listed individually, those less than 1% are placed in the "Others" category, together with animals unlikely to have been exploited for their meat, such as horse and dog (Table 4). Aquatic faunal remains come from the sieved sediments from two pits, one in the residential area and one in the storage area (Maufras *et al.* 2018). As the zooarchaeological evidence from this site is presented using a different set of metrics than the previous case study, it is not possible to directly compare the consumption patterns between these two sites. Table 4 will only serve to provide a rough overview of faunal abundance at Missignac.

Another challenge with this case study is that there is no obvious way to group the sources, since most have a strong C<sub>3</sub> signal. Of the major sources of animal protein, most species have very similar, if not completely overlapping ranges of  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values (Table 5). While the differences in the means of  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values between cattle and sheep/goat/pig are statistically significant, the differences are too small to be informative for mixing models. All associated data and analytical results associated with this case study are provided in Online Resource 4.

In order to determine how best to divide the sources, we used the previously mentioned cluster-analysis assisted source aggregation method using the function *NbClust()* in R (the script for performing *NbClust()* is provided in Online Resource

**Table 4** Faunal remains from the medieval site of Missignac

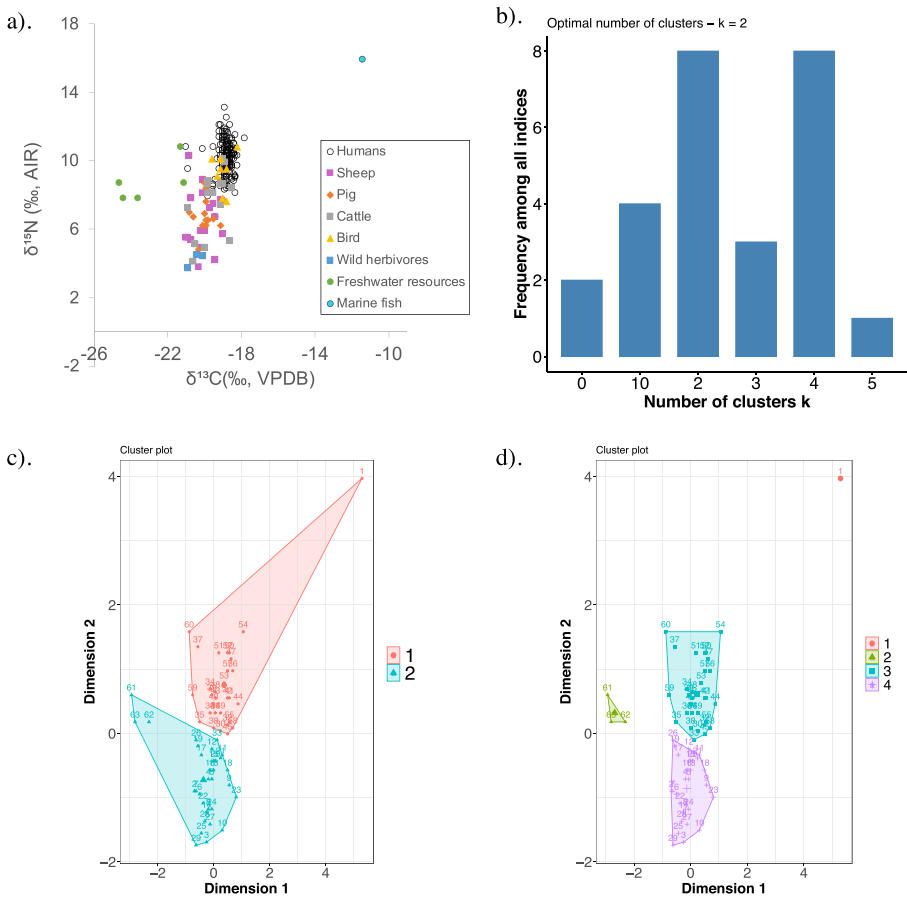
		Total NISP	Percentage
Terrestrial	Sheep/goat	1064	36.2%
	Pigs	638	21.7%
	Cattle	646	22.0%
	Chicken	157	5.3%
	Duck	71	2.4%
	Amphibians	45	1.5%
	Others	322	10.9%
	Total	2943	100%
Aquatic	Freshwater fish	640	12.6%
	Marine fish	32	> 0.1%
	Indeterminate	4391	86.8%
	Total	5063	100%

1). In this case, “NbClust()” suggests that our data would be best represented by a 2- or 4-cluster model (Fig. 4). We believe the 4-cluster model is optimal, from a functional perspective. Thus, following this model, we divided the fauna into 4 main groups: marine resources, freshwater resources, higher-trophic terrestrial meat (consists of birds and possibly some suckling domesticates), and lower-trophic terrestrial meat (consists of mostly adult domesticates and some wild animals), plus a fifth group consists of  $C_3$  plants that is not depicted in the figure.

FRUITS only allows up to 100 consumer datapoints per model. For sample sets with more than 100 consumers, users can either use the means and standard deviations of the group or run multiple models to incorporate all individual data. In this study, means and standard deviations of the consumers were used in the FRUITS modelling. In terms of hyperparameters, we adopted the same assumption as in the last case study, that plant food would contribute to at least 50% of the total calories consumed at Missignac. No other assumption was made about the proportion of other sources.

**Table 5** Means and standard deviations of  $\delta^{13}C$  and  $\delta^{15}N$  values of faunal samples from Missignac (Mion, et al. 2018)

Species	N	Mean $\delta^{13}C$ (‰)	SD	Mean $\delta^{15}N$ (‰)	SD
Sheep	14	− 20.2	0.5	+ 6.6	1.5
Goat	4	− 19.7	0.8	+ 6.7	2.6
Pig	13	− 20.0	0.4	+ 6.8	1.0
Cattle	15	− 19.5	0.7	+ 7.6	1.9
Bird	8	− 19.0	0.4	+ 9.3	1.1
Wild animals	3	− 20.5	0.4	+ 4.2	0.4
Marine fish	1	− 11.4	-	+ 15.9	-
Freshwater resources	5	− 23.0	1.7	+ 8.8	1.2



**Fig. 4** Determining the optimal number of source groups for Missignac: **a** scatter plot showing bone collagen  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values of fauna and humans from Missignac (data from Mion *et al.* 2018); **b** histogram depicting the best clustering schemes for this study as determined by various cluster suitability measuring criteria; **c** the faunal data in a 2-cluster scheme; **d** the faunal data in a 4-cluster scheme

## Error Structures and Hyperparameters

Informative priors can influence the outcome of a model, especially if the data set is small. Here, our last case study aims to test:

- How different error structures and intra-population groups compare with each other
- How much could hyperparameters sway the SIMMs estimation
- What is considered a sufficient sample size that is robust enough to stand against poorly informed hyperparameters

To test for (a), we chose to use the simulated dataset from the first case study to illustrate the effects of different error structures and intra-population group settings on the dietary reconstruction of various sub-population groups. Assuming among a high-

**Table 6** Summary statistics of the 6 simulated populations with increasingly larger  $n$ 

	N=	Mean $\delta^{13}\text{C}$ (‰)	SD	Mean $\delta^{15}\text{N}$ (‰)	SD
TM1	10	− 15.9	0.8	+ 10.9	0.7
TM2	20	− 16.1	0.7	+ 10.8	0.7
TM3	30	− 16.1	0.6	+ 10.7	0.6
TM4	60	− 16.0	0.5	+ 10.8	0.5
TM5	100	− 16.0	0.4	+ 10.9	0.5
TM6	150	− 15.9	0.4	+ 10.8	0.4

latitude coastal hunter-gatherer population, there are two subgroups (*e.g.* two different sites, two different time periods, two different sexes): one relied more heavily on terrestrial resources (*i.e.* TM heavy) and another relied more heavily on pinnipeds (*i.e.* MP). We tested the following:

- i) TM and MP as fixed effect, residual error only
- ii) TM and MP as fixed effect, process error only
- iii) TM and MP as fixed effect, process\*residual effect
- iv) TM and MP as random effect, residual error only
- v) TM and MP as random effect, process error only
- vi) TM and MP as random effect, process\*residual effect

Furthermore, we also tested the two groups as separate populations and compared the performance of each of the error structures (process only, residual only, process\*residual) with those generated by IsoSource and FRUITS.

To test for (b) and (c), we used the terrestrial-heavy model (TM) group. We ran the mixing models under three different conditions: uninformative hyperparameters ( $\alpha = c(1,1,1,1,1)$ ); informative and accurate hyperparameters ( $\alpha = c(1.4,0.9,1.3,0.7,0.7)$ ); and informative but inaccurate hyperparameters ( $\alpha = c(0.8,0.5,1.6,1.5,0.6)$ ). In the third scenario, the consumption of marine fish is underestimated, consumption of seal is overestimated, and some deer has been misidentified as reindeer. All of these errors represent common scenarios encountered in archaeological contexts. We ran these models on populations of six different sizes (Table 6). We stopped at  $n = 150$  consumers as few archaeological studies include more samples than this. All associated data and analytical results are provided in Online Resource 5.

## Results

### Hypothetical High-Latitude Coastal Hunter-Gatherer Population

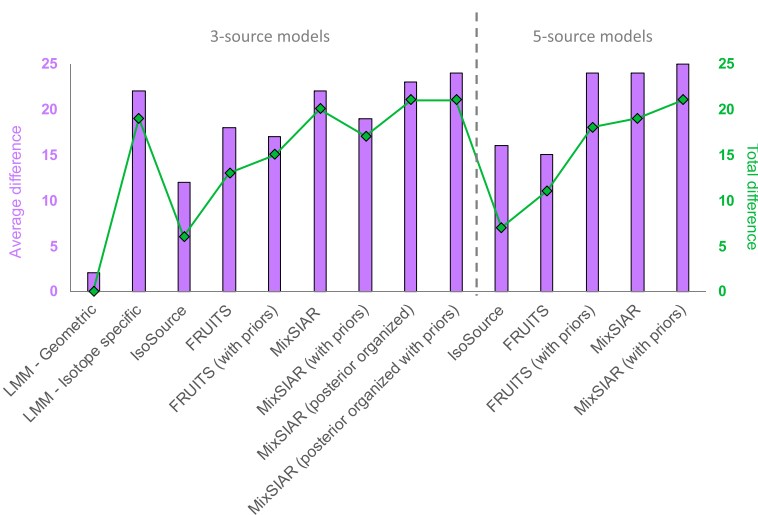
Table 7 summarizes the performance of each SIMM in each scenario with case study 1. As LMMs do not provide any margin of error, their results were excluded from this table. Note that for the “Equivalent” scenario, no 3-source BMM was run with

**Table 7** Summary of whether estimations generated by different SIMMS covered the “true” contributions of all sources: *M* refers to the means and standard deviations ranges; %ile refers to the 2.5th–97.5th percentile ranges (1st–99th percentile ranges for IsoSource). *Y* yes; *N* no; *n/a* no model is run; \ model failed. The total tallies the success rate of each method

SIMMs		TM heavy		MP heavy		MF heavy		Equivalent		Rdm effect		Total	
		M	%ile	M	%ile	M	%ile	M	%ile	M	%ile	M	%ile
3-source models	IsoSource	N	N	N	N	N	N	N	N	N	N	0/5	0/5
	FRUITS	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	5/5	5/5
	FRUITS (with $\alpha$ )	Y	Y	Y	Y	N	Y	n/a	n/a	Y	Y	3/4	4/4
	MixSIAR	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	5/5	5/5
	MixSIAR (with $\alpha$ )	Y	Y	Y	Y	Y	Y	n/a	n/a	Y	Y	4/4	4/4
	MixSIAR (posterior organized)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	5/5	5/5
	MixSIAR (posterior organized with $\alpha$ )	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	5/5	5/5
5-source models	IsoSource	N	Y	N	Y	Y	Y	N	N	N	Y	1/5	4/5
	FRUITS	Y	Y	Y	Y	\	\	Y	Y	\	\	3/3	3/3
	FRUITS (with $\alpha$ )	Y	Y	N	Y	Y	Y	N	N	Y	Y	3/5	4/5
	MixSIAR	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	4/5	5/5
	MixSIAR (with $\alpha$ )	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	5/5	5/5

hyperparameters, as the default prior setting is to assume equal contribution among all sources. The standard summary statistics generated by these SIMMs are also slightly different: IsoSource reports output in 1st, 50th, 99th percentile, minimum, maximum, mean, and standard deviation; FRUITS reports the mean, standard deviation, 2.5th, 50th, and 97.5th percentile; and MixSIAR reports the DIC (deviance information criterion), mean, standard deviation, 2.5th, 5th, 25th, 50th, 75th, 95th, and 97.5th percentile. With this particular case study, in terms of accuracy, a 3-source model with IsoSource is the least reliable (Table 7); this is largely due to the extremely small standard deviations provided by the model. In most cases, a 2.5th–97.5th percentile range (1st–99th percentile for IsoSource) provides a much larger, and hence, more accurate, estimation than the means and standard deviations do. All models appear to struggle with the “Equivalent” scenario, suggesting the mathematics behind these models is less effective in handling certain distribution shapes, potentially when diets consist of approximately equal proportions of four or more isotopically distinct sources. Among all the BMMs, MixSIAR performed slightly better than FRUITS, whether considering the means and standard deviation ranges or the 2.5th–97.5th percentile ranges (also known as the 95th credible interval). It is also interesting to note that the models with hyperparameters did not always provide estimations that were more accurate.

As shown in Fig. 5, while the geometric LMM had the lowest accuracy score, the isotope-specific approach performed drastically better and even outperformed some of the more sophisticated SIMMs. All 5-source models performed consistently better than their 3-source counterparts. If source aggregation is required, *a posteriori* aggregations using “combine\_sources” in MixSIAR provided better accuracy than *a priori* aggregation. The models with the highest scores for TD were the 5-source models with MixSIAR (with hyperparameters) and the two 3-source models with MixSIAR (posteriorly organized). The model with the highest score for AD was the 5-source model with MixSIAR (with hyperparameters). Note that without the input of hyperparameters,



**Fig. 5** Scores assessing the performance of each SIMM using average difference (AD) and total difference (TD) between the “real” source proportions and the estimations generated by the different SIMMs



FRUITS was unable to generate an estimate for two of the scenarios (MF-heavy and Rdm effect).

### **Yinxu, Shang Dynasty.**

Table 8 shows the different estimations of the dietary compositions of the *Yinxu* inhabitants. For the MixSIAR results, the proportions for the total animal protein in the diet are calculated using *a posteriori* source aggregation. For FRUITS, these are calculated from the estimated means. The “% animal protein” values are only used to loosely demonstrate the proportional relationships between the sources and should not be considered in an absolute sense. Some key observations from this scenario:

- Most models are able to capture the differential proportions of C<sub>4</sub> herbivores and C<sub>4</sub> omnivores that are reflected in zooarchaeological evidence.
- Without hyperparameters, FRUITS appears to have overestimated the contribution of animal protein to the diet and did not reflect the same ratio of C<sub>4</sub> herbivores/C<sub>4</sub> omnivores as other models do;
- All models suggested that C<sub>3</sub>/C<sub>4</sub> mixed and C<sub>3</sub> herbivores may be underrepresented in the zooarchaeological record.

### **Missignac-Saint Gilles le Vieux, Medieval Period**

Table 9 shows the different estimations of the dietary compositions of the Missignac inhabitants. Similar to the second case study, the values of this “% animal protein” were calculated by either *a posteriori* source aggregation (MixSIAR), or using the estimated means (FRUITS), and should only be used as a loose reference for understanding the proportional relationships between the sources. Some key observations from this scenario:

- Without hyperparameters, all models overestimated the importance of animal protein.
- All models suggested low-trophic level terrestrial food was consumed in higher quantities than other source groups, which is consistent with the zooarchaeological evidence
- While the true proportions of aquatic and terrestrial resources are unknown, both SIMMs and zooarchaeological evidence suggest that the Missignac inhabitants consumed more freshwater than marine resources.

### **Error Structures and Hyperparameters**

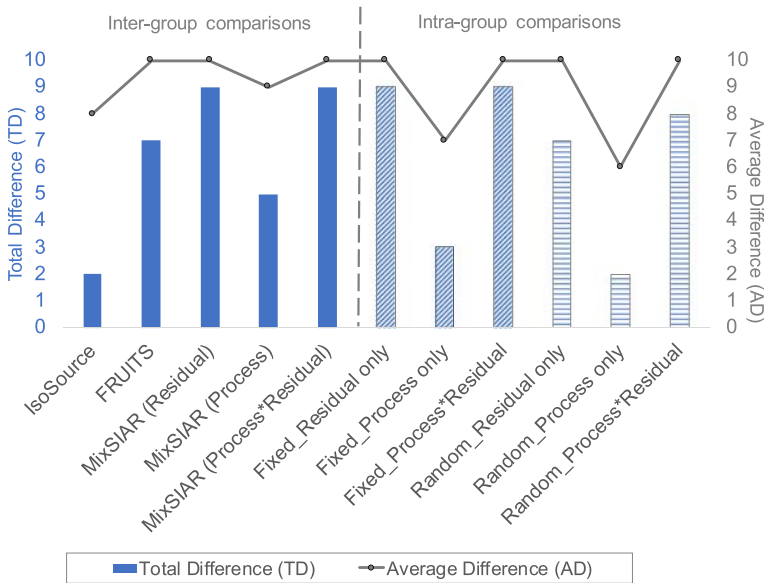
Figure 6 summarizes the performance of each model under different conditions. Our tests revealed that (1) models treating the two groups as separate populations perform

**Table 8** Summary of the estimations made by different mixing modelling tools for *Yinxu*, expressed as means and standard deviations (SD). % of animal protein was calculated as follows:  $1 - (\% C_3 \text{ plants} + \% C_4 \text{ plants})$ .  $C_4$ -consuming herbivores refers to mostly cattle;  $C_4$ -consuming omnivores refers to mostly pigs;  $C_4$  plants refers to mostly millet; mixed consumers refers to sheep/goat, fish, and any other animals that have intermediate  $\delta^{13}C$  values;  $C_3$  plants refers to wheat, barley, and possibly other wild plant resources;  $C_4$ -consuming herbivores refers to mostly deer

	$C_4$ -consuming herbivores (%)	$C_4$ -consuming omnivores (%)	$C_4$ plants (%)	Mixed consumers (%)	$C_3$ -consuming herbivores (%)	$C_3$ plants (%)	% animal protein (%)
IsoSource	21.1	12.0	41.6	9.9	7.6	7.8	50.6
SD	15.6	7.9	7.6	7.9	5.4	5.3	
% animal protein	41.7	23.7		19.6	15.0		
FRUITS	23.6	22.9	24.9	9.4	6.8	12.4	62.7
SD	17.3	16.2	18.1	8.4	5.6	10.4	
% animal protein	37.6	36.6		15.0	10.8		
FRUITS (with $\alpha$ )	20.5	8.5	44.8	9.1	5.9	11.3	43.9
SD	8.2	5.9	12.5	7.8	4.9	9.5	
% animal protein	46.6	19.3		20.7	13.4		
MixSIAR	24.9	4.6	53.0	3.2	4.3	10.0	36.9
SD	19.4	3.2	17.3	2.7	2.6	6.7	18.2
% animal protein	67.5	12.5		8.7	11.7		
MixSIAR (with $\alpha$ )	32.6	4.8	48.9	1.6	5.7	6.4	44.7
SD	19.1	3.3	17.0	2.6	3.6	8.4	17.7
% animal protein	72.9	10.7		3.6	12.8		

**Table 9** Summary of the estimations made by different mixing modelling tools for Missignac, expressed in means and standard deviations (SD). % of animal protein is calculated by excluding the contributions from plant sources. Marine resources refers to mostly marine fish; low trophic terrestrial sources refers to mostly adult domesticates (cattle, sheep/goat/pig) and some wild animals; high trophic terrestrial sources refers to birds (chicken and goose) and some suckling domesticates (cattle, sheep/goat/ pig); freshwater resources refers to mostly freshwater fish and ducks

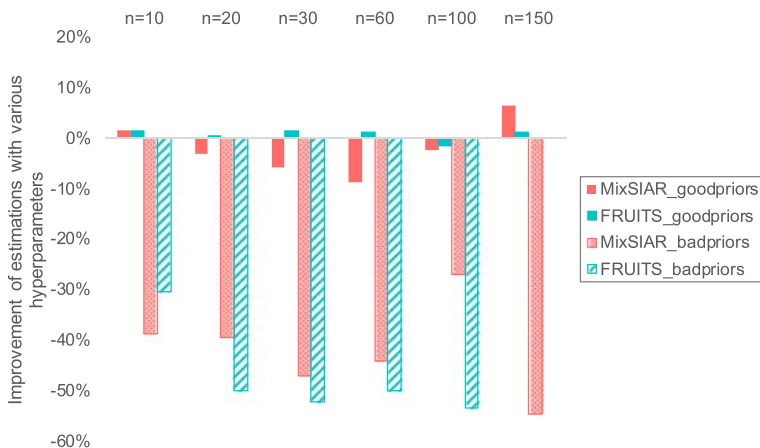
	Marine resources (%)	Low trophic terrestrial (%)	High trophic terrestrial (%)	Freshwater resources (%)	C <sub>3</sub> plants	% animal protein
IsoSource	3.6	36.5	16.8	11.9	31.8	68.2
SD	2.3	2.4	11.2	7.1	1.5	
% animal protein	5.3	53.5	24.6	17.5		
FRUITS	5.9	32.5	20.2	16.3	25.1	74.9
SD	4.8	19.6	16.3	13.3	18.9	
% animal protein	7.9	43.4	26.9	21.8		
FRUITS (with $\alpha$ )	5.4	15.1	11.5	16.5	51.5	48.6
SD	4.3	9.8	8.9	13.1	13.3	
% animal protein	11.1	31.1	23.8	34.1		
MixSIAR	2.6	49.3	8.5	5.1	34.5	65.5
SD	1.6	21.7	6.4	3.2	22.9	22.9
% animal protein	4.9	75.3	13.0	7.8		
MixSIAR (with $\alpha$ )	1.3	33.8	11.5	3.3%	50.1	49.9
SD	1.2	19.4	6.9	2.9	19.7	19.7
% animal protein	2.6	67.7	23.1	6.6		



**Fig. 6** Total difference and average difference between the “real” source proportions (the simulated TM and MP models from the first case study) and the estimations generated under different error structures and intra-population group settings. All intra-group comparisons are conducted using MixSIAR. Solid bars refer to outputs generated from two separate models; hatched bars refer to outputs generated from one model; diagonally hatched bars refer to models run under a fixed effect; horizontally hatched bars refer to models run under a random effect

better than models treating the two groups as subgroups of a population, and (2) models using residual error only and process\*residual error performed better than those using process error only.

To evaluate the effects of “good” and “bad” hyperparameters, respectively, we set the default models (without hyperparameters) as the “neutral” models. Among each



**Fig. 7** Results from the hyperparameters test: positive values indicate that the estimations were improved by the use of a hyperparameter; and negative values indicate that the estimations were worsened by the use of a hyperparameter.

population, the total difference (not the score TD, but the actual difference in percentage) between each model and the “true” values were compared. Figure 7 shows the differences between the default models and models with good and bad hyperparameters, respectively. Positive values mean the hyperparameters improved the model, while negative values mean that the hyperparameters worsened the accuracy of the model. With MixSIAR, good hyperparameters only improved the accuracy of the reconstructions 2 of 6 times, while with FRUITS, good hyperparameters improved the accuracy 5 of 6 times. However, the input of bad hyperparameters invariably decreased the accuracy, in much greater magnitude than the “improved” estimations ascribed to the good hyperparameters.

As discussed previously, in all the preceding examples (except for Missignac), all consumer data were entered as discrete data points in all BMMs, while source data were entered as means and standard deviations. This approach was meant to ensure the results generated by these different models were comparable. In this case study, however, FRUITS was unable to generate results with discrete consumer data points in 11 out of 15 instances (excluding the  $n = 150$  model, as the upper limit for consumer data is  $n = 100$  for FRUITS). Therefore, means and standard deviations of the consumers were used instead.

## Discussion

### Hypothetical Coastal Hunter-Gatherer Population

While our results show some variability in the performances of the SIMMs, most of the estimations were reasonably accurate. This is especially true with the BMMs, where both FRUITS and MixSIAR produced models that were both accurate and precise. Furthermore, with this particular case study, where source concentration was not an issue, the isotope-specific LMM’s performance was comparable to the BMMs, and even better than IsoSource. However, this is also possibly due to the fact that our consumer values were derived from linear equations (see Online Resource 2).

In terms of the output of the results, LMMs are obviously disadvantaged, as they do not provide an error range. While IsoSource does provide ranges in the form of minima, maxima, 1st percentile, 99th percentile, and standard deviations, these ranges are often too small to be helpful. This was especially true in the 3-source models, where the standard deviations for two of the source groups were as small as 0.7% for the 3-source MP heavy-model (Online Resource 2). For the BMMs, the 95% CI were more inclusive than the standard deviations and therefore provided better accuracy; however, some of the ranges were so large that it begs the question of whether the model was useful at all. For example, in the 5-source TM model, the 95% CI for deer was estimated to be 6.6–49.8% by MixSIAR and 3.32–51.8% by FRUITS (Online Resource 2). These estimations are likely correct, but too broad to be informative. The ranges provided by the standard deviations or the 50% CI, on the other hand, are more reasonable (MixSIAR,  $29.3 \pm 11\%$ ; FRUITS,  $28.36 \pm 13.08\%$ ). Fry (2013) suggested that reconstructed solutions should be reported as a range, or as he called a “minmax solution”, instead of means and standard deviations, to avoid leading to an assumption that certain solutions (*i.e.* the mean) are more feasible than others (*i.e.* either ends of the range).

Bayesian statistics is a method that evaluates the probability of a certain hypothesis (in this case, the contributions of different dietary components) by considering both observed data and other evidence. Thus, the results generated are expressed as *posterior* probability, and the “means” reported here are the *posterior* means. The *posterior* mean should reflect the centre of the mass of the distribution curve, which is useful for describing the *posterior* likelihood distribution (especially if both means and medians are close together). While we generally disagree with Fry’s characterization of “means” in Bayesian statistics, we agree that it is a good idea to report the probability as a range. Using the same example (5-source TM model, with the MixSIAR output), we could either report the mean and standard deviation, that deer likely contributed  $29.3 \pm 11.0\%$  (18.3–40.3%) to the group’s diet, or as 95% credible intervals [0.03, 0.52]: the probability that deer contributed to between 3–52% of this group is 95%. There are pros and cons to either approach. While the 95% CIs are larger with lower explanatory power, the uncertainty is itself an important indicator of how confident the estimation is. For example, in a study on the dietary compositions of camelid fibre textiles in Prehispanic Peru, with large sample sizes coupled with simple diets and isotopically distinctive source groups ( $C_3$  and  $C_4$  plants), Szpak *et al.* (2020) demonstrated that 95% CIs can be very well defined, in one instance even as narrow as 6% [0.22, 0.28]. This tight range is obviously impossible to obtain in more complex omnivores, such as humans, that utilize many distinct sources; however, this example shows that the size of the CI is a good reflection of the robustness of the model, where estimations with exceptionally large CIs should be taken with a grain of salt. That said, *posterior* means are not completely without value. As demonstrated in case studies 2 and 3, the means can be used to determine the proportional relationship between source groups, which cannot be done with a 95% CI. Using the *posterior* means as point estimates is not wrong *per se*, but in order to avoid creating misleading summaries, it is essential to also include the uncertainties in the data description.

Using summary statistics such as CI (50% or 95%) or means and standard deviations do not always provide a comprehensive picture of the results. For BMMs, Semmens *et al.* (2013) strongly recommended that users should also check the graphic outputs produced by the programs, which depict posterior densities of proportional source contributions, as well as relationships between source contributions. This may sometimes reveal important information such as multi-modality in the distribution or strong correlations between certain sources. Ultimately, the decision of how to report the model outputs is highly dependent on the case study and the specific research questions, and bioarchaeologists need to consider a host of other information such as the amount of variation in the sources, sample sizes, and other relevant archaeological evidence, to decide which is the best way to present the estimations generated by SIMMs. A summary of all the graphical outputs produce by IsoSource, FRUITS, and MixSIAR are provided in Online Resource 6.

### Yinxu, Shang Dynasty.

Even though it is not possible to evaluate the “accuracy” of the SIMMs estimations with this case study, most of the SIMMs’ estimations agree fairly well with zooarchaeological evidence; thus, we consider it likely that the results are accurate. The MixSIAR models were particularly good at capturing the proportional relationship

between  $C_4$ -consuming herbivores and  $C_4$ -consuming omnivores, with or without considering hyperparameters. Even though the model does not take source concentration into consideration, IsoSource performed reasonably well here. This is possibly because the source groups in this case study are sufficiently isotopically distinctive. For FRUITS, the default model produced an estimation that likely exaggerated the contribution of animal protein, which was then “corrected” by the hyperparameters. In this case, based on archaeological evidence as well as historical records, we are fairly confident that as an agrarian society, the *Yinxu* inhabitants likely consumed a high proportion (> 50%) of plant food. However, judging by how much the proportions can change after hyperparameters were applied (Fig. 7), users should be extremely cautious when applying hyperparameters to their models.

### Missignac-Saint Gilles le Vieux, Medieval Period

Similar to the last case study, it is not possible to evaluate the “accuracy” of the SIMMs estimation with this case study. From a mathematical point of view, SIMMs performed adequately. Other than likely underestimating the total input of plant food, the proportions of the animal sources were generally consistent among all the models and the zooarchaeological evidence. From an archaeological point of view, however, SIMMs failed to answer more in-depth questions about animal consumption patterns in this site. For example, we could not possibly distinguish whether sheep/goat, cattle, or pig were the most consumed animal among the *triade domestique* at this site because of the similarity of the isotopic compositions of the sources, a problem that cannot be subverted with mixing models. Depending on the investigator’s research interests, the difference in the consumption of “low trophic terrestrial meat” vs. “high trophic terrestrial meat” may be inconsequential. This case study shows that despite having a reasonably sized isotopic faunal baseline, SIMMs are ultimately constrained by the isotopic variation of the source groups. In this case, unfortunately, bone collagen  $\delta^{13}C$  and  $\delta^{15}N$  values are not particularly effective in elucidating dietary patterns at this site. Researchers working in this region, or in similar situations should consider additional approaches, such as incorporating more isotope systems such as sulphur ( $\delta^{34}S$ ), isotopic measurements of bioapatite that represent the whole diet, or compound specific isotope analysis, all of which may provide more information to help distinguish among source groups.

### Error Structures and Hyperparameters

Our tests with the simulated dataset on the effects of different error structures and intra-population group settings have not provided clear guidelines for bioarchaeologists using MixSIAR. With this particular case study, to make a comparison between the dietary compositions of the two subgroups, treating the two groups as separate populations appears to be the best option (Fig. 6). This may be a better approach in archaeology anyway, as one of the conditions to treat the different grouping as “subgroups” is that all concerned individuals have to obtain all their food from the same pool of resources. This is easy to set up in a hypothetical population, but in actual archaeological case studies, it is not always obvious. For example, in the Roman *colonia* Glevum (modern-day Gloucester, UK), it was suggested that some men had



consumed more C<sub>4</sub> resources (likely millet) than women (Cheung *et al.* 2012). A separate study on the same individuals suggested that these men could have come from continental Europe, as shown by strontium ( $^{87}\text{Sr}/^{86}\text{Sr}$ ) and stable oxygen ( $\delta^{18}\text{O}$ ) isotope analyses (Chenery *et al.* 2010). As there is no evidence of millet farming in the British Isles during that period (Müldner *et al.* 2011), based on the background of the site, it is quite possible that the slightly elevated  $\delta^{13}\text{C}$  values could have resulted from differential access to C<sub>4</sub> food stuffs as imported ration, or residual dietary signals from before their arrival at Glevum. In this case, access to C<sub>4</sub> resources is exclusive to only a certain part of the population. Using the same pool of local resources to reconstruct the dietary practices of the men and women would be inappropriate in this particular example.

Technically, we could add C<sub>4</sub> resources as one of the potential source groups when conducting SIMMs in a case study such as the one described for Glevum. However, an additional isotopically distinctive source may sway the output significantly, especially with LMMs. In the example discussed here, all SIMMs would likely suggest that women could have consumed some amount of C<sub>4</sub> plants, despite none being available. Even with the more sophisticated BMMs, if a source group is provided, no matter how improbable, the estimated proportion will never truly equal to 0%. Thus, this example shows that bioarchaeologists need to be cautious when using fixed/random effects to conduct intrapopulation comparisons in MixSIAR. Two key questions must first be asked: is there other historical or archaeological evidence that suggests all subgroups had access to all resources? Can the isotopic compositions of all individuals be entirely explained by local resources? If the answer is no to both questions, we suggest that it is best to treat these groups as separate populations.

Regarding hyperparameters, our results have shown that these should be applied with great caution. As shown in Fig. 7, the risks of using a “bad” hyperparameter are much greater than the rewards of using a “good” one. Having said that, one hyperparameter archaeologists could adopt concerns plant protein contribution. Without *priors*, SIMMs provide estimations of dietary proportions strictly based on the data entered (*i.e.* consumer isotopic values, source isotopic values, TDF, source concentration). Thus, the models will assume equal accessibility to all the sources presented, but do not consider factors such as opportunistic foraging, subsistence economy structures, as well as natural environmental conditions, which may vary greatly between sites. As a result, if other lines of evidence (such as historical texts) indicated that a group depended heavily on agricultural crops, using a hyperparameter to inform the model of this “bias” could help redress the estimations. However, note that in our second case study, the default MixSIAR model already gave a low estimate of total animal protein contribution, the hyperparameter (set as plant > 50%) was therefore unnecessary. In this case, the inclusion of this hyperparameter “backfired” and further raised the estimation of total animal protein contribution.

Even though hyperparameters can be potentially misleading, in theory, with a large enough sample size, the model should still be able to produce a *posterior* estimation that is mostly informed by the data. Our initial intention was to test what sample size is considered “large enough” to withstand potentially ill-informed hyperparameters. Unfortunately, as shown in Fig. 7, even when the size of the consumer dataset is 150, an ill-informed hyperparameter still strongly sways the results. Limited by the inherently incomplete nature of archaeological records, very few bioarchaeological studies can

involve more than 100 consumer measurements from a single site (Missignac being one such exception). Thus, within the scale of archaeological studies, unfortunately, sample size will likely never be sufficient to counter ill-informed priors.

FRUITS was unable to return estimations for 11 of 15 instances using discrete consumer data, especially for the populations with larger sample sizes. This was surprising as the dataset was simulated using simple linear equations, and as shown in Table 6, the variation decreases as population size increases. This means that the groups with larger sample sizes are becoming increasingly homogenous. We were forced to use means and standard deviations of these six simulated populations to generate the results shown in Fig. 7. One problem with using means and standard deviations to represent a population in FRUITS is that the program assumes that the dataset is normally distributed (Fernandes *et al.* 2014). In this case study, our datasets were generated with help from the function “rnorm” in R (see Online Resource 1), thus are mostly normally distributed. Ultimately, it is not a just comparison between FRUITS and MixSIAR in this particular instance, as one uses means and standard deviations of a group (FRUITS), while the other uses discrete data points (MixSIAR). Nevertheless, we have demonstrated that the amount of variance (reflected by the standard deviations) has little influence on the effect of hyperparameter inputs. Additionally, for bioarchaeologists interested in using BMMs to reconstruct past dietary compositions, this case study shows that it is important to first check for normality of the data before choosing a model to use—especially if the consumer sample size is larger than 100.

## Conclusion

This study has shown that while SIMMs have the potential to be a powerful tool for bioarchaeologists interested in research questions concerning past foodways, it is important to be acutely aware of the limitations and constraints of these tools. Table 10 provides a summary of the major features in IsoSource, FRUITS, and MixSIAR discussed in this study.

In the following, we have provided a checklist of issues to consider when choosing a modelling tool for archaeological case studies:

- Be realistic—no SIMM can reliably estimate the proportions of isotopically indistinguishable source groups (*e.g.* the *triade domestique* at Missignac, case study 3) in one’s diet. Research questions should focus on the kinds of differences that can actually be detected by SIMMs and stable isotope analysis in general.
- When organizing source groups, regrouping sources *a posteriori* appeared to be more reliable than doing so *a priori*, and this is especially true if the source groups have vastly different source concentrations.
- When using MixSIAR, it may not be easy to determine which error structure (*i.e.* residual only vs. process\*residual) and/or which intra-population group setting (*i.e.* fixed, random, or as separate population) is best for a site. Therefore, users are encouraged to experiment with multiple settings and decide which best reflects the archaeological realities of the site. If in doubt, treating intra-population groups as separate populations is likely a safer option.

**Table 10** A summary of all the basic available features in the three major mixing models discussed in this study (IsoSource, FRUITS, and MixSIAR) that may be applicable to bioarchaeological studies.

	<i>IsoSource</i>	<i>FRUITS</i>	<i>MixSIAR</i>
<b>Operating platform</b>	Windows	Windows	All operating platforms
<b>No. of biotracers</b>	up to 5	up to 50	Limit not specified, but in the manual, up to 22 fatty acids were used as tracers in one analysis.
<b>Consumer data</b>			
<i>Data input</i>	As means	As discrete datapoints, or as means + standard deviations. (If entered as means + standard deviations, the model will assume a normal distribution for the dataset)	As discrete datapoints
<i>Sample size</i>	$n = 1$	$n \leq 100$	No apparent upper limit
<b>Source data</b>			
<i>No. of source groups</i>	Up to 10	Up to 50	Not specified, but in the manual, an example script demonstrated a SIMM using 22 fatty acids. As discrete datapoint or as means + standard deviations. (If entered as means + standard deviations, the model will assume a normal distribution for the dataset)
<i>Data input</i>	As means	As means + standard deviations (the model assume a normal distribution for the dataset)	As discrete datapoint or as means + standard deviations. (If entered as means + standard deviations, the model will assume a normal distribution for the dataset)
<i>Source variance (o)</i>	Not considered	Considered. Source data inputted as means and standard deviations (o).	Considered. Source data inputted as discrete data points, or as means + standard deviations (o) + sample sizes.
<i>Trophic discrimination factor (TDF or <math>\tau</math>)</i>	Not built in, need to manually adjust and thus will not consider variance of TDF.	Built in, can consider variance, but all sources must share the same TDF. • If sources have significantly different TDF for each biotracer (e.g. plants vs. animal), may need to manually adjust source data to reflect that.	Built in, can consider variance and apply different TDF to each source.
<i>Source concentration</i>	Not built in	Optional: can consider different concentration + variance for each source.	Optional: can consider different concentration (no variance) for each source.
<b>Customizability</b>			
<i>Hyperparameters (a)</i>	Not built in	Optional • Can be vague (e.g. source A > source B) or specific (i.e. as mathematical equation)	Optional • Relatively specific, the relationship needs to be expressed mathematically as a ratio (e.g. source A: source B = 1:3) • All components have to be positive. • The larger the $\alpha$ , the more informative it is. • There are three major ways to arrange the $\alpha$ in MixSIAR. In a $k$ -source model, where $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$ : 1. $\sum \alpha = K$ 2. $\sum \alpha = 1$ 3. At random: e.g., 20 cattle bones, 10 pig bones, and 5 fish bones are found at the site, $\alpha = (20, 10, 5)$ . • All examples in this study arrange the $\alpha$ with the first approach
<i>Error structure</i>	Not built in	Not built in	Can choose between 3 configurations (Figure 1): 1. Process error only 2. Residual error only 3. Residual*Process errors
<i>Intra-population comparison</i>	Not built in	Not built in	Optional: can quantify and compare the diets of more than two subgroups in one single model • Not applicable to all situations. Only when the groups are directly comparable. • Can categorize groups as fixed or random, or whether categories are hierarchical or not (Stock <i>et al.</i> , 2018).
<i>Continuous factor</i>	Not built in	Not built in	Optional: can monitor changes in dietary compositions within a group over a continuous factor. Factors that could be relevant to archaeologists include time and age.
<i>Proportional contribution from macronutrients</i>	Not built in	Optional: • Need to know the isotopic compositions of the concerned macronutrients in sources.	Not built in.
<b>Result output</b>			
<i>Format</i>	Text summary (mean, minimum, maximum, 1st percentile, 50th percentile, 99th percentile, and standard deviation), graph, etc.	Text summary (mean, standard deviation, 2.5 <sup>th</sup> percentile, median, 97.5 <sup>th</sup> percentile), figure....	Default setting: Text summary (mean, standard deviations, 2.5%, 5%, 25%, 50%, 75%, 95%, 97.5%, DIC, Scaled posterior density, matrix plots) Optional: xplot for scaled food groups + consumers, hyperparameters plots, and other statistical tests.
<i>Posterior source combining</i>	Not built in	Not built in	Optional: with function "combine_sources".
<i>Solution</i>	May fail to generate any output if consumer data are outside the source polygon	May fail to generate any output if consumer data are outside the source polygon	Will always provide a solution even though consumer data may be well outside the source polygon • Note that this is not necessarily an advantage. Users are strongly advised to check the bivariate plot first to determine whether the outputs are reasonable.
<b>Coding knowledge</b>	Not required	Not required	Required, especially when using the more advanced features. • Users can seek help and support from many online forums such as GitHub and Stack Overflow.

- Hyperparameters should not be applied lightly. One must be very certain about a particular dietary preference or bias; otherwise, hyperparameters have the potential to provide erroneous results.

- In all cases, it is best to run the model without hyperparameters first, and to compare the two models (with and without hyperparameters) to make sure the estimations from the model with hyperparameters are not too unrealistic.
- If possible, cross-check SIMM estimations with other lines of evidence such as the zooarchaeological record.
- Always report statistical results with uncertainties, whether as CIs (50%, 90%, or 95%) or means and standard deviations.
- It is important to make sure all data used are reliable, passing all relevant quality control criteria. No matter how good a model is, the result will only be as good as the data inputted into the model.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10816-020-09492-5>.

**Acknowledgements** We are grateful to Drs. Eric Guiry and Leïa Mion for allowing us to use their data to evaluate the performances of different SIMMs. We also thank Dr. Estelle Herrscher and the four anonymous reviewers for their helpful and constructive comments on this manuscript.

**Availability of Data and Material** All archaeological data are obtained from published reports. All simulated data are provided in the Supplementary files.

**Funding** CC is supported by the Agence Nationale de la Recherche (Project-ANR-17-CE27-0023 “NEOGENRE”). PS is supported by the Canada Research Chairs program.

## Compliance with Ethical Standards

**Conflicts of Interest** The authors declare that they have no conflicts of interest.

**Code Availability** All codes used in this study are provided in the Supplementary files.

## References

- Ambrose, S., & Norr, L. (1993). Experimental evidence for the relationship of the carbon isotope ratios of whole diet and dietary protein to those of bone collagen and carbonate. In J. Lambert & G. Grupe (Eds.), *Prehistoric Human Bone: Archaeology at the Molecular Level* (pp. 1–37). Heidelberg: Springer.
- Ben-David, M., F. R. W., & Schell, D. (1997). Annual and seasonal changes in diets of Martens: evidence from stable isotope analysis. *Oecologia*, 111(2), 280–291.
- Bender, M., Baerreis, D., & Steventon, R. (1981). Further light on carbon isotopes and Hopewell Agriculture. *American Antiquity*, 43, 346–353.
- Bettinger, R., Barton, L., & Morgan, C. (2010). The origins of food production in North China: a different kind of agricultural revolution. *Evolutionary Anthropology*, 19(1), 9–21.
- Bocherens, H., & Drucker, D. (2003). Trophic level isotopic enrichment of carbon and nitrogen in bone collagen: case studies from recent and ancient terrestrial ecosystems. *International Journal of Osteoarchaeology*, 13(1–2), 46–53.
- Bocherens, H., Drucker, D., Billiou, D., Patou-Mathis, M., & Vandermeersch, B. (2005). Isotopic evidence for diet and subsistence pattern of the Saint-Césaire I Neanderthal: review and use of a multi-source mixing model. *Journal of Human Evolution*, 49(1), 71–87.

- Bocherens, H., Mashkour, M., Drucker, D., Moussa, I., & Billiou, D. (2006). Stable isotope evidence for palaeodiets in Southern Turkmenistan during Historical Period and Iron Age. *Journal of Archaeological Science*, 33(2), 253–264.
- Bownes, J. M., Ascough, P. L., Cook, G. T., Murray, I., & Bonsall, C. (2017). Using stable isotopes and a Bayesian mixing model (FRUITS) to investigate diet at the early neolithic site of Carding Mill Bay, Scotland. *Radiocarbon*, 59(5), 1275–1294.
- Campbell, R., Li, Z., He, Y., & Yuan, J. (2011). Consumption, exchange and production at the great settlement Shang: bone-working at Tiesanlu, Anyang. *Antiquity*, 85(330), 1279–1297.
- Campech, S. (1996). *Usages et goûts culinaires au Moyen-Age en Languedoc et en Aquitaine, Archéologie du Midi médiéval*. Carcassonne and Villeneuve-Termenès: Centre d'archéologie médiévale du Languedoc, Carcassonne.
- Chang, K. (1976). Food and food vessels in Ancient China. *Early Chinese Civilization: Anthropological Perspectives*, (pp. 115–148). Cambridge: Harvard University Press.
- Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: an {R} package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, 61, 1–36.
- Chen, Z. 陳. (1985). 商代晚期的家畜和家禽(Domesticated animals of the Late Shang). 農業考古, 288–295.
- Chenery, C., Müldner, G., Evans, J., Eckardt, H., & Lewis, M. (2010). Strontium and stable isotope evidence for diet and mobility in Roman Gloucester, UK. *Journal of Archaeological Science*, 37, 150–163.
- Cheung, C., Jing, Z., Tang, J., Weston, D., & Richards, M. (2017a). Diets, social roles, and geographical origins of sacrificial victims at the Royal Cemetery at Yinxu, Shang China: new evidence from stable carbon, nitrogen, and sulfur isotope analysis. *Journal of Anthropological Archaeology*, 48, 28–45.
- Cheung, C., Jing, Z., Tang, J., Yue, Z., & Richards, M. (2017b). Examining social and cultural differentiation in early Bronze Age China using stable isotope analysis and mortuary patterning of human remains at Xin'an Zhuang, Yinxu. *Archaeological and Anthropological Sciences*, 9(5), 799–816.
- Cheung, C., Schroeder, H., & Hedges, R. (2012). Diet, social differentiation and cultural changes in Roman Britain: new isotopic evidence from Gloucestershire. *Archaeological and Anthropological Sciences*, 4(1), 61–73.
- Cheung, C., Zhang, H., Hepburn, J., Yang, D., & Richards, M. (2019). Stable isotope and dental caries data reveal rapid adoption of wheat in Ancient China. *PLoS One*, 14(7), e0218943.
- Chinique de Armas, Y., Buhay, W., Rodríguez Suárez, R., Bestel, S., Smith, D., Mowat, S., & Roksandic, M. (2015). Starch analysis and isotopic evidence of consumption of cultigens among fisher–gatherers in Cuba: the archaeological site of Canimar Abajo, Matanzas. *Journal of Archaeological Science*, 58, 121–132.
- Chisholm, B., Nelson, D., & Schwarcz, H. (1983). Marine and terrestrial protein in prehistoric diets on the British Columbia Coast. *Current Anthropology*, 24(3), 396–398.
- Crawford, G. (2009). Agricultural origins in North China pushed back to the Pleistocene-Holocene boundary. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 7271–7272.
- DeNiro, M., & Epstein, S. (1978). Influence of diet on the distribution of carbon isotopes in animals. *Geochimica et Cosmochimica Acta*, 42(5), 495–506.
- DeNiro, M., & Epstein, S. (1981). Influence of diet on the distribution of nitrogen isotopes in animals. *Geochimica et Cosmochimica Acta*, 45(3), 341–351.
- Fernandes, R., Grootes, P., Nadeau, M.-J., & Nehlich, O. (2015). Quantitative diet reconstruction of a Neolithic population using a Bayesian mixing model (FRUITS): The case study of Ostorf (Germany). *American Journal of Physical Anthropology*, 158(2), 325–340.
- Fernandes, R., Millard, A., Brabec, M., Nadeau, M., & Grootes, P. (2014). Food reconstruction using isotopic transferred signals (FRUITS): a Bayesian model for diet reconstruction. *PLoS One*, 9(2), e87436.
- Fernandes, R., Nadeau, M., & Grootes, P. (2012). Macronutrient-based model for dietary carbon routing in bone collagen and bioapatite. *Archaeological and Anthropological Sciences*, 4(4), 291–301.
- Finucane, B. (2009). Maize and sociopolitical complexity in the Ayacucho Valley, Peru. *Current Anthropology*, 50(4), 535–545.
- Fraser, R., Bogaard, A., Charles, M., Styring, A., Wallace, M., Jones, G., Ditchfield, P., & Heaton, T. (2013). Assessing natural variation and the effects of charring, burial and pre-treatment on the stable carbon and nitrogen isotope values of Archaeobotanical cereals and pulses. *Journal of Archaeological Science*, 40(12), 4754–4766.
- Fry, B. (2013). Minmax solutions for underdetermined isotope mixing problems: reply to Semmens et al. (2013). *Marine Ecology Progress Series*, 490, 291–294.
- Guiry, E., Hepburn, J., & Richards, M. (2016). High-resolution serial sampling for nitrogen stable isotope analysis of Archaeological mammal teeth. *Journal of Archaeological Science*, 69, 21–28.

- Guiry, E., Noël, S., Tourigny, E., & Grimes, V. (2012). A stable isotope method for identifying transatlantic origin of pig (*Sus scrofa*) remains at French and English fishing stations in Newfoundland. *Journal of Archaeological Science*, 39(7), 2012–2022.
- Guo, Y., Hu, Y., Zhu, J., Zhou, M., Wang, C., & Richards, M. (2011). Stable carbon and nitrogen isotope evidence of human and pig diets at the Qinglongquan site, China. *Science China Earth Sciences*, 54(4), 519–527.
- Haines, E. B. (1976). Relation between the stable carbon isotope composition of fiddler crabs, plants, and soils in a salt marsh. *Limnology and Oceanography*, 21(6), 880–883.
- Hall, R. (1967). Those late corn dates: isotopic fractionation as a source of error in carbon-14 dates. *Michigan Archaeologist*, 13, 171–180.
- Hedges, R., & Reynard, L. (2007). Nitrogen isotopes and the trophic level of humans in archaeology. *Journal of Archaeological Science*, 34(8), 1240–1251.
- Hobson, K., & Collier, S. (1984). Marine and terrestrial protein in Australian Aboriginal diets. *Current Anthropology*, 25(2), 238–240.
- Hopkins III, J., & Ferguson, J. (2012). Estimating the diets of animals using stable isotopes and a comprehensive Bayesian mixing model. *PLoS One*, 7, 1–13.
- Hou, Y., Campbell, R., Zhang, Y., & Li, S. (2019). Animal use in a Shang Village: the Guandimiao zooarchaeological assemblage. *International Journal of Osteoarchaeology*, 29(2), 335–345.
- Hu, Y., Ambrose, S., & Wang, C. (2006). Stable isotopic analysis of human bones from Jiahu site, Henan, China: implications for the transition to agriculture. *Journal of Archaeological Science*, 33(9), 1319–1330.
- Jim, S., Jones, V., Ambrose, S., & Evershed, R. (2006). Quantifying dietary macronutrient sources of carbon for bone collagen biosynthesis using natural abundance stable carbon isotope analysis. *British Journal of Nutrition*, 95(6), 1055–1062.
- Kassambara, A., & Mundt, F. (2017). factoextra: extract and visualize the results of multivariate data analyses.
- Katzberg, M. (2008). Stable isotope analysis: a tool of studying past diet, demography, and life history. In M. Katzberg & S. Saunders (Eds.), *Biological Anthropology of the Human Skeleton Second* (Edition ed., pp. 413–441). Hoboken, NJ: John Wiley & Sons, Inc..
- Kline Jr., T., Goering, J., Mathisen, O., Poe, P., Parker, P., & Scanlan, R. (1993). Recycling of elements transported upstream by runs of Pacific salmon: II.  $\delta^{15}\text{N}$  and  $\delta^{13}\text{C}$  Evidence in the Kvichak River Watershed, Bristol Bay, Southwestern Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, 50(11), 2350–2365.
- Krogh, A. (1937). The use of isotopes as indicators in biological research. *Science*, 85(2199), 187–191.
- Krueger, H., & Sullivan, C. (1984). Models for carbon isotope fractionation between diet and bone. *Stable Isotopes in Nutrition*, (pp. 205–220) American Chemical Society.
- Ladurie, M. (1976). *The Peasants of Languedoc*. Urbana and Chicago: University of Illinois Press.
- Larsen, C. (2003). Animal source foods and human health during evolution. *The Journal of Nutrition*, 133(11), 3893S–3897S.
- Lee-Thorp, J., Sealy, J., & van der Merwe, N. (1989). Stable carbon isotope ratio differences between bone collagen and bone apatite, and their relationship to diet. *Journal of Archaeological Science*, 16(6), 585–599.
- Lee, G., Crawford, G., Liu, L., & Chen, X. (2007). Plants and people from the Early Neolithic to Shang Periods in North China. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 1087–1092.
- Lynott, M., Boutton, T., Price, J., & Nelson, D. (1986). Stable carbon isotopic evidence for maize agriculture in Southeast Missouri and Northeast Arkansas. *American Antiquity*, 51(1), 51–65.
- Maufras, O., Hernandez, J., Rochette, M., & Thomas, B. (2018). Aimargues - Madame - Saint-Gilles le Vieux: Missignac, villa médiévale et ses abords (Ve - Xiiiie s.).
- McConnaughey, T., & McRoy, C. P. (1979). Food-web structure and the fractionation of carbon isotopes in the Bering Sea. *Marine Biology*, 53(3), 257–262.
- Mion, L., Herscher, E., André, G., Hernandez, J., Donat, R., Fabre, M., Forest, V., & Salazar-García, D. (2018). The influence of religious identity and socio-economic status on diet over time, an example from Medieval France. *Archaeological and Anthropological Sciences*.
- Müldner, G., Chenery, C., & Eckardt, H. (2011). The ‘headless Romans’: multi-isotope investigations of an unusual burial ground from Roman Britain. *Journal of Archaeological Science*, 38(2), 280–290.
- Müldner, G., & Richards, M. (2007). stable isotope evidence for 1500 years of human diet at the City of York, UK. *American Journal of Physical Anthropology*, 133(1), 682–697.
- Nestle, M. (1999). Animal v. plant foods in human diets and health: is the historical record unequivocal? *Proceedings of the Nutrition Society*, 58, 211–218.



- Newsome, S., Phillips, D., Culleton, B., Guilderson, T., & Koch, P. (2004). Dietary reconstruction of an Early to Middle Holocene human population from the Central California Coast: insights from advanced stable isotope mixing models. *Journal of Archaeological Science*, 31(8), 1101–1115.
- Nifong, J., Layman, C., & Silliman, B. (2015). Size, sex and individual-level behaviour drive intrapopulation variation in cross-ecosystem foraging of a top-predator. *Journal of Animal Ecology*, 84(1), 35–48.
- O'Connell, T., Kneale, C., Tasevska, N., & Kuhnle, G. (2012). The diet-body offset in human nitrogen isotopic values: a controlled dietary study. *American Journal of Physical Anthropology*, 149(3), 426–434.
- Parnell, A., Phillips, D., Bearhop, S., Semmens, B., Ward, E., Moore, J., Jackson, A., & Inger, R. (2013). Bayesian stable isotope mixing models. *Environmetrics*, 24, 387–399.
- Parnell, A. C., Inger, R., Bearhop, S., & Jackson, A. L. (2010). Source partitioning using stable isotopes: coping with too much variation. *PLoS One*, 5(3), e9672.
- Pate, D. (1998). Bone collagen stable nitrogen and carbon isotopes as indicators of past human diet and landscape use in Southeastern South Australia. *Australian Archaeology*, 46(1), 23–29.
- Phillips, D., & Gregg, J. (2003). Source partitioning using stable isotopes: coping with too many sources. *Oecologia*, 136(2), 261–269.
- Phillips, D., & Koch, P. (2002). Incorporating concentration dependence in stable isotope mixing models. *Oecologia*, 130(1), 114–125.
- Post, D. M. (2002). Using stable isotopes to estimate trophic position: models, methods, and assumptions. *Ecology*, 83(3), 703–718.
- Richards, M. P., Jacobi, R., Stringer, C., Pettitt, P. B., & Cook, J. (2006). Marine diets in the European late Upper Paleolithic: a reply to Bocherens and Drucker (2006). *Journal of Human Evolution*, 51(4), 443–444.
- Rosing, M., Ben-David, M., & Barry, R. (1998). Analysis of stable isotope data: a K nearest-neighbors randomization test. *The Journal of Wildlife Management*, 62(1), 380–388.
- RStudio Team. (2018). *RStudio: Integrated Development Environment for R*. Boston, USA: RStudio, Inc..
- Schoenheimer, R., & Rittenberg, D. (1940). The study of intermediary metabolisms of animals with the aid of isotopes. *Physiological Reviews*, 20(2), 218–248.
- Schoeninger, M., & DeNiro, M. (1984). Nitrogen and carbon isotopic composition of bone collagen from marine and terrestrial animals. *Geochimica et Cosmochimica Acta*, 48(4), 625–639.
- Schoeninger, M., & Moore, K. (1992). Bone stable isotope studies in archaeology. *Journal of World Prehistory*, 6, 246–296.
- Schwarcz, H. (1991). Some theoretical aspects of isotope paleodiet studies. *Journal of Archaeological Science*, 18(3), 261–275.
- Scott, M. (2020). *Chew the fat: an examination of the preservation of fatty acids in archaeological bone*. Peterborough: Trent University.
- Semmens, B., Ward, E., Moore, J., & Darimont, C. (2009). Quantifying inter- and intra-population niche variability using hierarchical bayesian stable isotope mixing models. *PLoS One*, 4(7), e6187.
- Semmens, B., Ward, E., Parnell, A., Phillips, D., Bearhop, S., Inger, R., Jackson, A., & Moore, J. (2013). Statistical basis and outputs of stable isotope mixing models: comment on fry (2013). *Marine Ecology Progress Series*, 490, 285–289.
- Si, Y. 司. (2013). 2500 BC–1000 BC 中原地區家畜飼養策略與先民肉食資源消費 *Feeding practices of domestic animals and meat consumption of ancients in the central plain of china: 2500 BC – 1000 BC*. University of Chinese Academy of Sciences.
- Stock, B., Jackson, A., Ward, E., Parnell, A., Phillips, D., & Semmens, B. (2018). Analyzing mixing systems using a new generation of bayesian tracer mixing models. *PeerJ*, 6, e5096.
- Stock, B., & Semmens, B. (2016a). MixSIAR GUI User Manual. Version 3.1. <https://github.com/brianstock/MixSIAR>.
- Stock, B., & Semmens, B. (2016b). Unifying error structures in commonly used biotracer mixing models. *Ecology*, 97(10), 2562–2569.
- Szpak, P. (2014). Complexities of nitrogen isotope biogeochemistry in plant-soil systems: implications for the study of ancient agricultural and animal management practices. *Frontiers in Plant Science*, 5.
- Szpak, P., Metcalfe, J., & Macdonald, R. (2017). Best practices for calibrating and reporting stable isotope measurements in archaeology. *Journal of Archaeological Science: Reports*, 13, 609–616.
- Szpak, P., Millaire, J.-F., Chapdelaine, C., White, C., & Longstaffe, F. (2020). An integrated isotopic study of Early Intermediate Period Camelid Husbandry in the Santa Valley, Perú. *Environmental Archaeology*, 25(3), 279–295.
- RStudio Team. (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.



- Tomczyk, J., Regulski, P., Lisowska-Gaczorek, A., & Szostek, K. (2020). Dental caries and stable isotopes analyses in the reconstruction of diet in Mesolithic (6815–5900 BC) Individuals from Northeastern Poland. *Journal of Archaeological Science: Reports*, 29, 102141.
- van der Merwe, N. (1982). Carbon isotopes, photosynthesis, and archaeology: different pathways of photosynthesis cause characteristic changes in carbon isotope ratios that make possible the study of prehistoric human diets. *American Scientist*, 70, 596–606.
- Vogel, J., & van der Merwe, N. (1977). Isotopic evidence for early maize cultivation in New York State. *American Antiquity*, 42(2), 238–242.
- Wang, X., Fuller, B., Zhang, P., Hu, S., Hu, Y., & Shang, X. (2018). Millet manuring as a driving force for the Late Neolithic agricultural expansion of North China. *Scientific Reports*, 8(1), 5552.
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
- Wickham, H. (2017). tidyverse: Easily Install and Load the ‘Tidyverse’.
- Xu, Q. 許., Cao, X. 曹., Wang, X. 王., Li, Y. 李., Jing, Z. 蔣., & Tang, J. 唐. (2010). 殷墟文化發生的環境背景及人類活動的影響 Generation of Yinxu culture: environmental background and impacts of human activities. 第四紀研究 *Quaternary Sciences*, 30, 273–286.
- Yan, L. 閻. (2010). 穩定同位素陸生動物礦化組織研究中的應用 *The stable isotopes analysis of biomineral tissues from terrestrial animal*. University of Chinese Academy of Sciences.
- Yang, S. 楊., & Ma, J. 馬. (2010). 商代史. 六 – 商代經濟與科技 (Economy and Technology in Shang Dynasty). Beijing: China Social Sciences Press.
- Yang, X., Wan, Z., Perry, L., Lu, H., Wang, Q., Zhao, C., Li, J., Xie, F., Yu, J., Cui, T., Wang, T., Li, M., & Ge, Q. (2012). Early millet use in Northern China. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 3726–3730.
- Yoneda, M., Suzuki, R., Shibata, Y., Morita, M., Sukegawa, T., Shigehara, N., & Akazawa, T. (2004). Isotopic evidence of inland-water fishing by a Jomon population excavated from the Boji Site, Nagano, Japan. *Journal of Archaeological Science*, 31(1), 97–107.
- Zhou, X., Li, X., Dodson, J., & Zhao, K. (2016). Rapid agricultural transformation in the Prehistoric Hexi Corridor, China. *Quaternary International*, 426, 33–41.
- Zhu, Y. 朱. (2005). 關於商代中原地區野生動物諸問題的考察 The investigation on the wild animals in Zhong Yuan District in Shang Period. 殷都學刊 (*Yindu Journal*), 3, 1–9.