On the Compositional Analysis of Fatty Acids in Pork

ROGER ROS-FREIXEDES and JOAN ESTANY

Departament de Producció Animal, Universitat de Lleida - Agrotecnio Centre, 191 Rovira Roure, 25198 Lleida, Spain

Address correspondence to Joan Estany, Departament de Producció Animal, Universitat de Lleida, 191 Rovira Roure, 25198 Lleida, Spain. Tel.: +34 973702893; Fax: +34 973702874. E-mail: jestany@prodan.udl.cat

Abstract

Fatty acid (FA) composition of pork is an important issue for the pig industry and consumers. Fatty acid composition is commonly described as the percentages of a set of FA relative to total FA and therefore should be statistically treated as compositional data. To our knowledge there is no reference in the literature where specific methods for compositional data analysis have been applied to analyze FA composition in meat quality research. The purposes of this study were (1) to present an overview of compositional data analysis techniques, (2) to apply them to the analysis of the FA composition of muscles and subcutaneous fat from 941 pigs as a case study, and (3) to discuss and interpret the results with respect to those obtained using standard techniques. Results from both approaches indicate that FA composition differed across tissues and muscles but also, for a given muscle, with the intramuscular fat content. It is concluded that FA composition in pork did not display enough variability to become critical for standard statistics, particularly if the individual FA parts remain the same across experiments. However, even in such case, compositional analysis may be useful to correctly interpret the correlation structure among FA.

Key words: compositional data, intramuscular fat, meat quality, subcutaneous fat

1. INTRODUCTION

The quality of fat is a feature becoming increasingly important for both the industry and consumers. Currently, there is enough evidence indicating that fat quantity and quality affect the nutritional, sensory, and technological properties of animal products, particularly pork (Wood et al., 2003; Schmid, 2010). Fat quality is chemically defined in terms of fatty acid (FA) composition, which is commonly presented as a set of percentages corresponding to the relative content of each individual FA (or the sum of some of them) with respect to the total content of the FA that had been determined, i.e., as a vector of positive values whose sum is a constant. Technically, this sort of data is what in statistics is known as compositional data, i.e., multivariate data where the variables represent parts of a whole (Pawlowsky-Glahn and Egozcue, 2006). Compositional data are intrinsically multivariate because each component cannot be interpreted without relating it to any of the other components. They only represent relative information and therefore standard statistical techniques, which were conceived to deal with variables measured on an absolute scale, are inappropriate. Consequently, specific methods for compositional data analysis have been developed since the 1980s (Aitchison, 1982; Aitchison, 1986; Aitchison and Egozcue, 2005; Bacon-Shone, 2011). To our knowledge there is no reference in the literature where compositional data analysis had been applied to meat quality research.

Much research has been undertaken in recent years to assess the effect of influential factors (such as diet, genotype, gender, body weight, age, or fat content, among others) on the FA profile of pork fat and meat, mostly sampled from backfat and loin chops. However, it is also known that the pattern of FA deposition differs not only between the adipose and muscle tissues (Franco et al., 2006; Duran-Montgé et al., 2008; Yang et al., 2010) but also among muscles (Sharma, Gandemer, and Goutefongea, 1987; Leseigneur-Meynier and Gandemer, 1991). The University of Lleida has assembled a biorepository of pig fat and muscle

specimens for conducting research studies on meat quality, including samples from a Duroc genetic line used for producing premium quality pork cuts. Currently, the associated dataset to this line, with around 1700 FA profiles from different muscles and backfat locations (Section 2), provides a valuable resource for revisiting the pattern of FA deposition in pork under a compositional data analysis setting. The purpose of this study was (1) to review the fundamentals of the compositional data analysis techniques (Sections 3-4), and then (2) use this approach to examine the variations in the FA profile of pork meat and fat as a case study (Section 5). The utility of adopting the compositional data approach in the statistical analysis of FA compositions in meat products is discussed in light of the results of the case study.

2. DESCRIBING THE CASE STUDY

The case study comprises data from 971 purebred Duroc barrows used and referenced elsewhere (Bosch et al., 2012; Ros-Freixedes et al., 2012). The pigs were raised at a carcass market weight of around 95-100 kg (Table 1) in 12 commercial batches from 2001 to 2008. All pigs had *ad libitum* access to a commercial feed and were slaughtered at the same abattoir. There, a sample of the muscle *gluteus medius* (GM) was collected from the left ham of all pigs. Moreover, in randomly chosen subgroups of them, additional samples of the muscles *longissimus dorsi* (at the level of the third and fourth last ribs; LD), *semimembranosus* (SM), and *latissimus dorsi* (LT) were also taken, as representative muscles of the loin, ham, and shoulder, respectively. Finally, two samples of the subcutaneous backfat (SF) were obtained at the positions where GM (SFGM) and LD (SFLD) muscle samples were taken. The samples of SM, SFGM, and SFLD were collected immediately after slaughter and frozen in liquid nitrogen until required for analyses. The samples of GM, LD, and LT were collected after chilling for about 24 h at 2ºC, vacuum packaged and stored in deep freeze until analysis. The number of samples per muscle and backfat location by batch is detailed in Table 1.

Once defrosted, a representative aliquot from pulverized freeze-dried samples was used for fat analysis. The intramuscular fat (IMF) content and FA composition were determined in duplicate by quantitative determination of the individual FA by gas chromatography (Bosch et al., 2009). Fatty acid methyl esters were directly obtained by transesterification using a solution of 20% boron trifluoride in methanol (Rule et al., 1997). Methyl esters were determined by gas chromatography using a capillary column SP2330 (30 m × 0.25 mm; Supelco, Bellefonte, PA) and a flame ionization detector with helium as carrier gas at 1 ml/min. The oven temperature program increased from 150 to 225ºC at 7ºC/min and injector and detector temperatures were both 250ºC. The quantification was carried out through area normalization with an external mixture of FA methyl esters (Sigma, Tres Cantos, Madrid). The internal standard was 1,2,3-tripentadecanoylglycerol. The FA composition was expressed as the percentage of each individual FA relative to total FA. The complete profile for each sample included saturated (SFA; C14:0, C16:0, C18:0, and C20:0), monounsaturated (MUFA; C16:1n-7, C18:1n-9, and C20:1n-9), and polyunsaturated (PUFA; 18:2n-6, C18:3n-3, C20:2n-6, and C20:4n-6) FA (Figure 1). The IMF content in the four muscles was calculated as the sum of the individual FA expressed as triglyceride equivalents (AOAC, 1997) on a dry tissue basis. Because C20:0 is present at very low levels, it was not detectable in a few samples. The zero values represent a mathematical challenge for compositional data, which only represent relative magnitudes. To solve this problem several replacement strategies have been proposed (Martín-Fernández and Thió-Henestrosa, 2006; Palarea-Albaladejo, Martín-Fernández, and Gómez-García, 2007). For its simplicity, we followed here the strategy in Sanford, Pierson, and Crovelli (1993) and replaced the zeros by 0.55 times the lowest measured value in each tissue before calculating the FA percentages.

3. SETTING THE PROBLEM

One of the drawbacks of analysing compositional data with conventional methods is that the results can be subcompositionally incoherent (Aitchison, 1986, Chapter 3; Pawlowsky-Glahn and Egozcue, 2006). This becomes particularly evident in correlation analyses, where the correlation coefficient between two given components can differ depending on whether they are expressed relative to a set of components or another. In order to highlight this problem we calculated the correlation between pairs of FA under two different compositional settings. In the first one, the correlation matrix among the complete 11-part FA profile of GM was calculated (Table 2, rows a), while, in the second, the correlation was calculated between each SFA, MUFA, and PUFA expressed relative to the total SFA, MUFA, or PUFA, respectively, in such a way that, for instance, C14:0, C16:0, C18:0, and C20:0 summed up to 100% (i.e., the SFA subcomposition was closed). Then, the correlations among the FA in each subcomposition (SFA, MUFA, and PUFA) were recalculated (Table 2, rows b, c, and d, respectively). As can be seen in Table 2, the two correlations were not consistent, with the discrepancy being particularly relevant for those between C16:0 and C18:0, C16:1 and C18:1, and C18:2 and C20:4, which changed, respectively, from 0.80 to −0.91, 0.11 to −0.98, and 0.59 to −0.89. These changes, both in magnitude and sign, are due to the fact that components in compositional data do not vary independently. It can be proven that for a D-part composition $\mathbf{x} = [x_1, x_2, ..., x_D]$, if $x_1 + x_2 + ... + x_D = \kappa$ (where κ is a constant, often 1 or 100%), then $\mathrm{cov}(x_1, x_2) + \mathrm{cov}(x_1, x_3) + ... + \mathrm{cov}(x_1, x_D) = -\mathrm{var}(x_1)$. Therefore, at least one of the covariances of x_1 with the other components must be negative (Pearson, 1897; Aitchison, 1986, Chapter 3; Filzmoser and Hron, 2009). This negative bias causes that an increase in one of the components results in the decrease in, at least, another one. Hence, the correlations are not free to range over the interval [−1, 1]. The distribution of the bias over the covariance terms, along with the subsequent changes in the correlation matrix among components, depends upon which parts are included in the

composition. As a consequence, the above correlations do not have any neat interpretation. This simple example highlights that the analysis of compositional data using standard techniques may lead to spurious and inconsistent results across subcompositions.

4. OVERVIEW OF COMPOSITIONAL ANALYSIS

Compositional data need to be statistically treated considering that they only carry relative information. Two general approaches have been developed to deal with them. The first is known as staying-in-the-simplex approach. It operates in the so-called simplex space (S^D, for D-part compositions) and uses the Aitchison geometry (Aitchison, 1986, Chapter 2). The second approach resorts to log-ratio transformations (Aitchison, 1986, Chapter 7; Egozcue et al., 2003) to map the simplex to the real space, where the more familiar Euclidean geometry is used and standard statistics methods can be applied. Both approaches can be used complementarily depending on which geometrical framework is preferred. A brief description of both approaches is given below. Some software has been developed to easily process and analyze compositional data, such as the freeware CoDaPack (Thió-Henestrosa and Martín-Fernández, 2005; Comas-Cufí and Thió-Henestrosa, 2011a,b) and the R packages 'compositions' (van den Boogaart, Tolosana, and Bren, 2011) and 'robCompositions' (Templ, Hron, and Filzmoser, 2011).

4.1. Staying-in-the-Simplex

The simplex vector space is defined by the internal simplicial operation of perturbation, the external operation of powering, and the simplicial metric. The operations of perturbation,

$$x \oplus y = [x_1, x_2, \ldots, x_D] \oplus [y_1, y_2, \ldots, y_D] = C[x_1 y_1, x_2 y_2, \ldots, x_D y_D] \tag{4.1}$$

and powering,

$$a \odot x = a \odot \left[x_1, x_2, \ldots, x_D\right] = C\left[x_1^{\,a}, x_2^{\,a}, \ldots, x_D^{\,a}\right] = C\left(x^a\right)$$

(4.2),

where **x** (**y**) is a *D*-part composition, x_i (y_i) are the percentages for each part (i = 1, 2, ..., D), a is a scalar, and C is the closure operator to constant κ (rescaling through division of each part by their total sum), are the equivalent to translation and scalar multiplication in the real space, respectively. The staying-in-the-simplex approach requires an algebra that differs from the one used in standard statistics.

An example of this algebra is found in the calculation of descriptive statistics. The mean and the variance are not suitable statistics for compositional exploratory analyses (Daunis-i-Estadella, Barceló-Vidal, and Buccianti, 2006) and therefore they are replaced in the Aitchison geometry by the centre (**g**) and the variation matrix (**T**), respectively. The centre or geometric mean is defined as:

$$g = C\left[\left(\prod_{j=1}^{n} x_{1j}\right)^{1/n}, \left(\prod_{j=1}^{n} x_{2j}\right)^{1/n}, \ldots, \left(\prod_{j=1}^{n} x_{Dj}\right)^{1/n}\right]$$

(4.3),

where x_{ij} are the percentages for each part (i = 1, 2, ..., D) in sample j, and n is the number of samples. Moreover, the compositions can be centered, i.e., moved to the barycenter of the simplex, using $\mathbf{x} \oplus (-1 \odot \mathbf{g}) = \mathbf{x} \oplus \mathbf{g}^{-1}$ (Pawlowsky-Glahn and Egozcue, 2006). Centering is equivalent to subtracting the arithmetical mean in the Euclidean space. The variation matrix is defined as $\mathbf{T} = [\tau_{ij}]$, with $\tau_{ij} = \mathrm{var}[\ln(X_i/X_j)]$, where X_i and X_j are the data vectors for the parts i and j across samples. Low variance of a log-ratio indicates proportionality between the parts involved. The total variability of the dataset is the sum of the variances of all log-ratios divided by 2D:

$$\text{total-variance} = \frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} var\left[\ln \frac{X_i}{X_j}\right]$$

(4.4).

182

183**4.2. Log-Ratio Transformations**

184 The two first log-ratio transformations were introduced by Aitchison (1986, Chapters 4
185and 6) and the third by Egozcue et al. (2003). These log-ratio transformations make it possible
186to work on compositional data in the real space using Euclidean geometry.

187

188**4.2.1. Additive Log-Ratio**

189 The additive log-ratio (alr) transformation is written in terms of log-ratios of $D-1$
190components relative to an arbitrary D component:

191 $$alr(x)=\left[\ln\frac{x_1}{x_D},\ln\frac{x_2}{x_D},\ldots,\ln\frac{x_{D-1}}{x_D}\right]$$

192 (4.5).

193 This transformation has the obvious disadvantage that the results are dependent on the
194chosen divisor component, which in turn does not have an equivalent for further analyses.
195But, most importantly, the alr-transformation is not isometric, i.e., distances are not preserved
196in the new metric space (Filzmoser and Hron, 2009).

197

198**4.2.2. Centered Log-Ratio**

199 The centered log-ratio (clr) transformation is written in terms of the log-ratio of each
200component relative to the geometric mean of all the components of an individual:

201 $$z=clr(x)=\left[\ln\frac{x_1}{\left(\prod_{i=1}^{D}x_i\right)^{1/D}},\ln\frac{x_2}{\left(\prod_{i=1}^{D}x_i\right)^{1/D}},\ldots,\ln\frac{x_D}{\left(\prod_{i=1}^{D}x_i\right)^{1/D}}\right]$$

202 (4.6).

203 In the $\mathbf{z}=\mathrm{clr}(\mathbf{x})$ transformation all parts of the composition have a direct equivalent, so
204that transformed variables can be easily traced back to the originals. Although the clr

transformation is isometric, it is subcompositionally incoherent. Moreover, the covariance matrix of the clr-transformed variables is singular, which difficults the use of the clr transformation in multivariate statistical analyses requiring the inversion of this matrix. The clr transformation is mostly used in exploratory analysis. The so-called clr-biplots allow for a graphical representation of the distribution of the samples based on their composition. Moreover, the depiction of the links (i.e., the vector connecting the apexes of two variable rays) provides an easy-to-interpret representation of the log-ratios between the two involved components, where its length represents the standard deviation of the log-ratio and the cosine of the angle between the two links the correlation between the two involved log-ratios. A complete description of clr-biplots and their interpretation is given in Aitchison and Greenacre (2002) and Daunis-i-Estadella, Barceló-Vidal, and Buccianti (2006). Conclusions only should be drawn from biplots that explain a large percentage of the total variance. An example is presented in Section 5.1.

4.2.3. Isometric Log-Ratio

The isometric log-ratio (ilr) transforms the raw composition to its coordinates in an orthogonal system based upon an orthonormal basis (Ψ) (Egozcue et al., 2003). If Ψ is chosen following a sequential binary partition (Egozcue and Pawlowsky-Glahn, 2005), the ilr-transformed components are called balances (b_k, where k = 1, 2, ..., D−1). In a sequential binary partition, Ψ is constructed by successive divisions of the set of parts into two mutually exclusive groups (parts in one group are marked with the symbol + while parts in the complementary group with the symbol −) until only one part per group is left (see Table 3 for an example). To be interpretable, partitions should be based on previous knowledge and experience. Then, Ψ is derived replacing the symbols + and − by $\frac{1}{r}\sqrt{\frac{rs}{r+s}}$ and $\frac{-1}{s}\sqrt{\frac{rs}{r+s}}$,

respectively, where r (s) is the number of parts marked with + (−) in each balance, with blanks being zero. Then, the balances $\mathbf{w} = \mathrm{ilr}(\mathbf{x})$ are calculated as $\mathbf{w} = \mathbf{z}\Psi^{\mathrm{T}}$, or directly, in terms of normalized log-ratios between the geometric means of the two groups, as:

$$b_k = \sqrt{\frac{r_k s_k}{r_k + s_k}} \ln ¿¿¿¿ \tag{4.7}$$

where $x_k^{+¿¿}$ and $x_k^{-¿¿}$ represent the subsets of r_k and s_k parts in group + and − of the kth balance, respectively.

Note that, as happens with the alr transformation, there are only $D-1$ balances for a D-part composition, and that the balances may be different for each Ψ. The balances are isometric and subcompositionally coherent and, as a result, they can be analysed using standard statistical techniques. However, because they do not have a one-to-one relation to the original components, their interpretation is not straightforward. This can be overcome by choosing, if it exists, a sequential binary partition leading to interpretable balances or, alternatively, back-transforming them into interpretable D-part compositions lying in the simplex. Because compositions are intrinsically multivariate, estimates on the full set of $D-1$ balances (for instance, either least squares means or regression coefficients) must be jointly back-transformed as $\mathbf{x} = C(e^{w\Psi})$ (Tolosana-Delgado and van den Boogaart, 2011). In Sections 5.2 and 5.4 examples on the application of ilr-transforming and back-transforming are presented. However, it is not possible to back-transform the standard errors associated with least square estimates, but they can be substituted by the corresponding back-transformed confidence intervals. The use of balances is the best choice for correlations (Filzmoser and Hron, 2009), but they cannot be back-transformed either. If the sequential bipartition used does not lead to the desired balances, additional log-ratios can be calculated as linear combinations of the initial $D-1$ set derived from Ψ. For example, apart from the balances

derived from the sequential bipartition in Table 3 (b_1 to b_{10}) we could be interested in the log-ratios of C18:1 and C18:0:

$$\frac{1}{\sqrt{2}}\ln\frac{C18:1}{C18:0}=\frac{\sqrt{3}}{2}b_6-\frac{1}{2}b_7=\frac{\sqrt{3}}{2}\left(\sqrt{\frac{2}{3}}\ln\frac{\sqrt{C18:1\cdot C20:1}}{C18:0}\right)-\frac{1}{2}\left(\frac{1}{\sqrt{2}}\ln\frac{C20:1}{C18:1}\right) \quad (4.8),$$

or, similarly, MUFA and SFA:

$$\sqrt{\frac{12}{7}}\ln\frac{\sqrt[3]{C16:1\cdot C18:1\cdot C20:1}}{\sqrt[4]{C14:0\cdot C16:0\cdot C18:0\cdot C20:0}}=\frac{1}{2\sqrt{2}}b_2-\frac{\sqrt{7}}{6}b_4+\frac{\sqrt{7}}{2\sqrt{6}}b_5+\frac{\sqrt{7}}{3\sqrt{2}}b_6 \quad (4.9).$$

The inclusion of more log-ratios can enrich the interpretation of the results but then it should be noted that the covariance matrix including the new log-ratios will be singular. An example of correlation analysis using balances is given in Section 5.5.

5. ANALYSING THE CASE STUDY

The basics of compositional analysis are illustrated in five examples using the pork FA composition as a case study. The first is an exploratory analysis conducted to examine the differences between IMF and backfat for FA composition (Section 5.1). The second and third introduce the procedures to compare the distinct tissues and muscles in terms of centers (Section 5.2) and variation matrixes (Section 5.3). In Section 5.4 a linear regression is used to assess the effect of IMF content on FA composition. Finally, Section 5.5 illustrates how to interpret correlations among biologically meaningful balances. In Sections 5.2 and 5.4 the compositional and the standard approaches are compared.

5.1. Exploratory Analysis

The distribution of FA composition across muscles and backfat locations was first explored depicting the whole set of observations on a joint biplot (Figure 2). To this purpose

277the dataset **X** was clr-transformed to **Z**, and then singular value decomposed using standard
278procedures (Daunis-i-Estadella, Thió-Henestrosa, and Mateu-Figueras, 2011). The two first
279components accounted for 76% of the total variation. The projection of the samples (Figure
2802a) in the biplot showed that IMF can be clearly discriminated from SF based on FA
281composition. More specifically, the first component, which explained 56% of the total
282variation, was enough to separate IMF from SF samples. The most important FA affecting this
283component was C20:4, whose ray was opposite to those of the other PUFA and formed with
284them a long link along the first component (as an example, the link of ln(C20:4/18:2) is
285represented with a discontinuous line in Figure 2b). The length of these links, which relate to
286the standard deviation of the log-ratio of the two FA involved, indicates that the log-ratio
287between C20:4 and other PUFA (C18:2 and C18:3) displayed a great variation along the
288gradient separating IMF and SF. The SF samples were allocated in a cluster at the left side of
289the biplot and the IMF samples were clustered at the right side, indicating that the ratios
290C20:4/C18:2 and C20:4/C18:3 were greater in IMF than in SF. Despite some overlapping, the
291samples from each muscle can also be singled out (Figure 2a), especially within batch (Figure
2922b). In doing so, SM samples were mostly found in the upper region of the IMF cluster
293whereas those from GM (left), LT (middle), and LD (right) were in the lower. This could not
294be done for SF, where only one backfat location was analysed per batch. The distribution
295pattern of the batch centers suggested that the effect of the batch on the FA composition of
296IMF could be, at least partially, explained by differences in the age at slaughter (Table 1).
297Because IMF increases with age and saturation with IMF, pigs slaughtered at later ages are
298expected to have more saturated fat (Bosch et al., 2012). Accordingly, within muscle, the
299samples from pigs slaughtered at later ages (Table 1; batches 5-7 and 10-11) should tend to
300show greater SFA/PUFA ratios and therefore appear preferentially lower-left in the biplot
301relative to those from pigs slaughtered at earlier ages (Table 1; batches 1-4, 8-9, and 12).

A biplot for each muscle was also set up. The effect of batch was removed centering the data by batch (which is the equivalent in the simplex to subtract the mean of the batch) before they were clr-transformed and singular value decomposed. The IMF content was included in the biplots as a supplementary variable (Daunis-i-Estadella, Thió-Henestrosa, and Mateu-Figueras, 2011) to assess the relationship between IMF content and composition. The loading plots of the two first components by muscle are given in Figure 3. The two first components explained from 67% (GM) to 74% (SM) of the total variance. The loading plots showed a similar pattern among muscles, with SM being the most different. In all muscles, SFA and MUFA were in the opposite side to PUFA for the first component. The cosine of the angle between two links refers to the correlation between their log-ratios. In general, the angles between links involving two SFA (C16:0, C18:0), two MUFA (C16:1, C18:1), or a SFA with a MUFA, were small, indicating high correlations among them. Because C18:0 can be synthesized from precursor C16:0 by an elongase, and both C16:1 and C18:1 are synthesized by the same Δ^9 desaturase from C16:0 and C18:0, respectively (Cook and McMaster, 2002; Figure 1), the product/substrate ratio C18:0/C16:0 is frequently used as an indicator of the elongase activity, and ratios C16:1/C16:0 and C18:1/C18:0 of the Δ^9 desaturase activity. Thus, the high correlations among ratios of these four FA are biologically consistent and in line with the correlations found by other authors (Ntawubizi et al., 2010). The links involving C14:0, in all the muscles, and C20:0, in SM, had much greater angles, and thus lower correlations, with the other links. This might be because C14:0, unlike other SFA, is mainly of dietary origin (Wood et al., 2008; Figure 1) and because C20:0 is subjected to relatively larger instrumental error and greater number of zeros. Small angles, and thus high correlations, were also found between links corresponding to log-ratios of PUFA. However, in all the muscles, the links involving two SFA, two MUFA, or a SFA with a MUFA, on one side, and the links involving PUFA, on the other side, were almost perpendicular to each other. This indicates low

correlations between these two groups of log-ratios, in accordance with the low association of PUFA with SFA and MUFA reported in literature (Cameron and Enser, 1991; Zhang et al., 2007; Ntawubizi et al., 2010; Yang et al., 2010). Overall, the results indicate that SFA and MUFA behave similarly to each other but differently from PUFA, in line with their different deposition patterns. Fat depots, IMF and SF, can be divided into two fractions: phospholipids and neutral lipids. Phospholipids have structural functions and have abundant PUFA, particularly C20:4, which is the major PUFA in cell membranes (Larsson et al., 2004), whereas neutral lipids, mainly composed of SFA and MUFA, have storage functions. It means that IMF increases with neutral lipids while phospholipids remain relatively constant (Cameron and Enser, 1991; De Smet, Raes, and Demeyer, 2004), which is the reason for the positive relationship of IMF with SFA and MUFA, but negative with PUFA (Cameron and Enser, 1991; Zhang et al., 2007; Yang et al., 2010). The IMF content displayed a negative collinearity with C20:4 in all the muscles, supporting that increased IMF is associated with decreased C20:4, namely phospholipids, and PUFA, as well as to increased SFA and MUFA (Cameron and Enser, 1991; De Smet, Raes, and Demeyer, 2004; Bosch et al., 2012).

5.2. Differences among Tissues and Muscles

The centers of the FA composition of IMF and SF (Eq. 4.3) established that the most abundant FA were C18:1 (44.0-46.1%), C16:0 (21.2-24.3%), C18:2 (9.2-16.2%), and C18:0 (10.6-12.1%) in all the studied muscles and backfat locations, in agreement with the general knowledge on meat FA composition (Valsta, Tapanainen, and Männistö, 2005). The centers revealed differences of FA composition among the muscles and backfat locations. These differences were estimated and tested using the balances described in Table 3. The balances were analyzed using a linear mixed model, in which fixed effects included the batch (1 to 12), tissue (the four muscles and the two backfat locations), and carcass weight as a covariate. The

352pig and the residual were the random effects. Variances were estimated by restricted 353maximum likelihood and fixed effects were tested following a Kenward-Roger approach. The 354differences between tissues were contrasted with the Tukey HSD test at a significance level of 3550.05. The analyses were performed using JMP 8 software (SAS Institute Inc., Cary, NC). The 356least squares means and confidence intervals for the balances were back-transformed as 357indicated in Section 4.2.3. Results were compared with those obtained using the same model 358for raw FA percentages instead of balances.

359 The centers adjusted for batch and carcass weight are given in Table 4. The ordinary 360least squares means differed on average only by 0.1% (SD 0.1), with a maximum of 0.8% 361(C18:1). Significant differences among muscles and backfat locations were found, with 362compositional and standard approaches leading to similar conclusions. The two backfat 363locations showed greater contents of the PUFA C18:2, C18:3, and C20:2 than IMF in all 364muscles, but lower of C20:4. By contrast, IMF was more saturated and monounsaturated, 365although for some FA the differences between IMF and SF were not significant. These 366findings were in line with the well-known result that essential PUFA, C18:2 and C18:3, which 367are from dietary origin (Figure 1), are preferentially deposited in SF (Kloareg, Noblet, and 368van Milgen, 2007; Duran-Montgé et al., 2008). That the C20:4 displays an opposite trend to 369other PUFA (see Figure 2b) could be explained by the much greater fraction of phospholipids 370in IMF as compared to SF. Among muscles, SM had higher concentrations of C18:2 and 371C20:4 than GM, LD, and LT, and lower of the main SFA and MUFA. The observed 372differences in muscle composition can be partly attributed to IMF content (Table 4).

373

374**5.3. Variation within Tissue and Muscle**

375 The variation arrays and the total-variances (Eq. 4.4) were calculated for each muscle 376and backfat location. The total-variance of the composition of IMF in GM was 0.57. After

adjusting for batch (i.e., centering by batch), the total-variance decreased to 0.32. This indicates that around one half of the variability of the muscle FA composition is due to common environmental effects in a batch. The adjusted total-variance was higher for IMF in SM (0.97) than in GM, LD, and LT, which were very similar to each other (0.27-0.32) and to SFLD (0.37). The total-variance for SFGM was much lower (0.10). In general, the log-ratios involving C18:1 were the ones displaying the lowest variances (0.01-0.33) in all cases. Interestingly, the log-ratios involving C20:4 showed the highest relative variability in all cases (0.02-0.73), except for IMF in SM and SFLD, where C20:0 was the most variable FA. Nonetheless, the high variability of C20:0 could be due, because of its low content, to the relatively large analytical errors and replaced zeros. The variability of C20:4 is partly due to the variance of the phospholipids fraction in the IMF content, which, as it will be shown in Section 5.4, is not neutral with respect to IMF content. Overall, the variation of FA composition in pork is low. The largest element of the variation matrix of IMF in GM was 0.48 and the maximum across tissues was 1.13 for SM. These values are, for example, 10-fold and 4-fold lower than those reported by Daunis-i-Estadella, Barceló-Vidal, and Buccianti (2006) for geological compositional data, the area of expertise where compositional data techniques have been mostly applied.

5.4. Regression on Intramuscular Fat Content

Results in Section 5.2 support that fat content influences fat composition (Wood et al., 2008; Bosch et al., 2012). This relationship can be assessed by performing a compositional regression analysis of FA composition on IMF content (Aitchison, 1986, Chapter 7; Egozcue and Pawlowsky-Glahn, 2011; Egozcue et al., 2012). The 109 samples of GM in batch 1 were used for this purpose. The 10 balances described in Table 3 were compositionally regressed on IMF content (JMP 8 software, SAS Institute Inc., Cary, NC) and then the results were

compared with the simple regression of the raw FA percentages on IMF content. The vectors of estimated intercepts (**i**) and slopes (**s**) in the ilr-setting were back-transformed to the simplex as **i'** = $C(e^{i\Psi})$ and **s'** = $C(e^{s\Psi})$. Then, the FA composition at a given IMF content (**x**) can be predicted operating either in the simplex, with **x** = **i'** $\oplus$(IMF $\odot$ **s'**), or in the real space, with **w** = ilr(**i'**) + IMF × ilr(**s'**) = **i** + IMF × **s** and then back-transforming **w** to **x** = $C(e^{w\Psi})$.

The balances more influenced by IMF content were balances 1 and 8 (R^2 = 0.23 and 0.20, respectively). The R^2 associated to the other balances was lower than 0.08. The balance 1 was built to represent the ratio PUFA vs. SFA + MUFA, while balance 8 was associated to the ratio n-6 vs. n-3 PUFA (i.e., C18:2 + C20:2 + C20:4 vs. 18:3). This is consistent with results discussed in Section 5.1, where PUFA and, particularly, C20:4, more abundant in phospholipids, decrease as IMF content increases. Similar results were found for raw percentages, with C18:2 and C20:4 showing the highest R^2 (0.34 and 0.14). The relationship between FA and IMF content is displayed in Figure 4. For simplicity, only three FA are displayed, although the analyses were done using the whole 11-FA composition. A relevant difference between compositional and standard regression is that in this latter case, at extreme values of the covariate, the predicted values can be non-sense. Thus, at high IMF contents negative percentages are predicted for C18:2 (IMF > 65%) and C20:4 (IMF > 35%). This does not happen in the compositional analysis. The back-transformed regressions of the 10 balances on IMF content were non-linear and asymptotically bounded, with predicted values always lying within the [0, 100] range. However, within the expected range of values for IMF, from 5% to 30% on dry matter basis (equivalent to approximately 1% to 10% of fresh meat), the compositional regression is almost linear, overlapping with the standard regression. Predicted values, even using validation samples from other batches, were almost identical under the two approaches. In the expected range of values for IMF the standard regression led to similar results to the compositional analysis. A similar conclusion is reached in models

other than the regression used here, which is deliberately simple for illustrative purposes.

5.5. Correlations among Enzymatic Indices

The correlations between balances for GM are given in Table 5. The balances described in Table 3 were established in accordance with known metabolic pathways for FA synthesis in pigs (Figure 1). Because they are regulated by specific enzymes the balances can be thought in terms of enzymatic activity. The first balance can be interpreted as a polyunsaturation index (PUFA vs. SFA + MUFA), which separates the PUFA and the SFA and MUFA pathways. Balances 2 to 7 are associated to SFA and MUFA metabolism, where balances 2, 3, 4, and 7 can be interpreted as indexes of elongase activity, and balances 5 and 6 of Δ^9 desaturase activity. Note that although they are aimed at representing different elongation or desaturation steps, in general they are not ratios between single products and substrates. For instance, balance 3 accounts not only for the elongation of C16:0 to C18:0, but also for the amount of C16:0 that has alternatively been desaturated to C16:1 and the amount of C18:0 further transformed into C20:0, C18:1, and C20:1. The balances can be an interesting alternative to elementary indexes between only two FA because they also include further or alternative products derived from the same substrate (Figure 1). However, because they are designed based upon a sequential bipartition, some balances cannot include all the desired FA (e.g., balance 6 does not include C20:0, which can be elongated from C18:0). As expected, all the elongase balances were positively correlated among them, as well as the two desaturation indexes. However, interestingly, the correlation among the desaturation and the elongase indexes was negative. The polyunsaturation index was negatively correlated to the elongase activity but positively to the Δ^9 desaturase activity. Balances 8, 9, and 10 are associated with PUFA metabolism. Balance 8 is the ratio between n-6 and n-3 FA, which is known to play a crucial role in the nutritional quality of fat (Schmid, 2010). The positive correlation between

balance 1 and balance 8 indicates that the n-6/n-3 ratio increased with polyunsaturation. Balance 9 reflects the total efficiency of biosynthesizing C20:4 from any of the two pathways using C18:2 as a precursor, while balance 10 only accounts for the intermediate elongation step from C18:2 to C20:2 carried out in one of the two pathways (Figure 1). The positive correlation between balances 9 with balances 1 and 8 confirmed that the percentage of C20:4 increases with PUFA and with the n-6/n-3 ratio. A correct interpretation of the balances may help to gain new insight into FA metabolism. Note that in this example we used only the $D-1$ balances described in Table 3, which derived from a unique sequential bipartition. More log-ratios could be calculated and added as discussed in Section 4.2.3. For example, the correlation between the log-ratios of C18:1/C18:0 and MUFA/SFA (Eq. 4.8 and 4.9) was 0.70.

6. CONCLUSIONS

Fatty acid compositions, which by nature are compositional data, should be statistically treated as such. There are two complementary approaches to analyse compositional data: either operate in the simplex space or make use of log-ratios to operate in the real space. The ilr transformation allows for a straightforward handling of geometric elements in the simplex using standard statistical procedures. Nonetheless, for the case study considered here we found that the inferences drawn from compositional analysis did not substantively differ from those obtained using standard statistics techniques on raw data. The low variability of FA composition across fat pork depots may explain why the standard approach, although methodologically inconsistent, is robust enough for practical purposes. This is likely to happen to other unprocessed raw food products, where natural variability is subjected to homeostatic biological constraints. Results evidenced that IMF and SF behave differently in terms of FA composition, with IMF showing more SFA, MUFA, and C20:4, and that FA

composition differs among muscles, with SFA and MUFA increasing with IMF. Compositional analysis proved to be useful in correctly interpreting the correlation structure among FA components. Choosing an appropriate set of balances may help not only to avoid spurious results but also to better address the biological mechanisms involved in FA deposition. Careful attention is recommended in cases of higher expected variability, such as when comparing differentiated processed products, where a compositional analysis may lead to more dramatic changes.

REFERENCES

Aitchison, J. (1982), "The Statistical Analysis of Compositional Data" (with discussion), *Journal of the Royal Statistical Society. Series B (Methodological)*, 44, 139-177.

Aitchison, J. (1986), "The Statistical Analysis of Compositional Data". Chapman and Hall Ltd. (reprinted 2003 with additional material by The Blackburn Press), London, UK.

Aitchison, J., and Egozcue, J. J. (2005), "Compositional Data Analysis: Where Are We and Where Should Be Heading?," *Mathematical Geology*, 37, 829-850.

Aitchison, J., and Greenacre, M. (2002), "Biplots of Compositional Data," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 51, 375-392.

AOAC (1997), Supplement to "AOAC Official Method 996.06: Fat (Total, Saturated, and Monounsaturated) in Foods Hydrolytic Extraction Gas Chromatographic Method," in *Official Methods of Analysis*, 16th ed. (Supplement, March 1997), ed. Association of Official Analytical Chemists, Arlington, VA.

Bacon-Shone, J. (2011), "A Short History of Compositional Data Analysis," in *Compositional Data Analysis: Theory and Applications*, eds. V. Pawlowsky-Glahn, and A. Buccianti, John Wiley & Sons Ltd., West Sussex, UK.

Bosch, L., Tor, M., Reixach, J., and Estany, J. (2009), "Estimating Intramuscular Fat Content

502 and Fatty Acid Composition in Live and Post-Mortem Samples in Pigs," *Meat Science*,
503 82, 432-437.
504Bosch, L., Tor, M., Reixach, J., and Estany, J. (2012), "Age-Related Changes in Intramuscular
505 and Subcutaneous Fat Content and Composition in Growing Pigs Using Longitudinal
506 Data," *Meat Science*, 98: 358-363.
507Cameron, N. D., and Enser, M. B. (1991), "Fatty Acid Composition of Lipid in *Longissimus*
508 *Dorsi* Muscle of Duroc and British Landrace Pigs and its Relationships with Eating
509 Quality," *Meat Science*, 29, 295-307.
510Comas-Cufí, M., and Thió-Henestrosa, S. (2011a), "CoDaPack 2.0: A Stand-Alone, Multi-
511 Platform Compositional Software," in *CoDaWork'11: 4th International Workshop on*
512 *Compositional Data Analysis*, eds. J. J. Egozcue, R. Tolosana-Delgado, R., and M. I.
513 Ortego, Sant Feliu de Guíxols, Spain.
514Comas-Cufí, M., and Thió-Henestrosa, S. (2011b), "CoDaPack 2.0. Manual and Software,
515 Version 2.01", http://ima.udg.edu/CoDaPack/
516Cook, H. W., and McMaster, C. R. (2002), "Fatty Acid Desaturation and Chain Elongation in
517 Eukaryotes," in *Biochemistry of Lipids, Lipoproteins and Membranes* (4th ed.). eds. D.
518 E. Vance, D. E., and J. E. Vance, Elsevier, New York.
519Daunis-i-Estadella, J., Barceló-Vidal, C., and Buccianti, A. (2006), "Exploratory
520 Compositional Data Analysis," in *Compositional Data Analysis in the Geosciences:*
521 *From Theory to Practice*, eds. A. Buccianti, G. Mateu-Figueras, and V. Pawlowsky-
522 Glahn, Geological Society Special Publications 264, 161-174, London, UK.
523Daunis-i-Estadella, J., Thió-Henestrosa, S., and Mateu-Figueras, G. (2011), "Including
524 Supplementary Elements in a Compositional Biplot," *Computers & Geosciences*, 37,
525 696-701.
526De Smet, S., Raes, K., and Demeyer, D. (2004), "Meat Fatty Acid Composition as Affected

527 by Fatness and Genetic Factors: A Review," *Animal Research*, 53, 81-98.

528Duran-Montgé, P., Realini, C. E., Barroeta, A. C., Lizardo, R., and Esteve-Garcia, E. (2008),
529 "Tissue Fatty Acid Composition of Pigs Fed Different Fat Sources," *Animal*, 2, 1753-
530 1762.

531Egozcue, J. J., Daunis-i-Estadella, J., Pawlowsky-Glahn, V., Hron, K., and Filzmoser, P.
532 (2012), "Simplicial Regression. The Normal Model," *Journal of Applied Probability*
533 *and Statistics*, 6, 87-108.

534Egozcue, J. J., and Pawlowsky-Glahn, V. (2005), "Groups of Parts and their Balances in
535 Compositional Data Analysis," *Mathematical Geology*, 37, 795-828.

536Egozcue, J. J., and Pawlowsky-Glahn, V. (2011), "Basic Concepts and Procedures," in
537 *Compositional Data Analysis: Theory and Applications*, eds. V. Pawlowsky-Glahn, and
538 A. Buccianti, John Wiley & Sons Ltd., West Sussex, UK.

539Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., and Barceló-Vidal, C. (2003),
540 "Isometric Logratio Transformations for Compositional Data Analysis," *Mathematical*
541 *Geology*, 35, 279-300.

542Filzmoser, P., and Hron, K. (2009), "Correlation Analysis of Compositional Data,"
543 *Mathematical Geosciences*, 41, 905-919.

544Franco, I., Escamilla, M. C., García, J., Fontán, M. C. G., and Carballo, J. (2006), "Fatty Acid
545 Profile of the Fat from Celta Pig Breed Fattened using a Traditional Feed: Effect of the
546 Location in the Carcass," *Journal of Food Composition and Analysis*, 19, 792-799.

547Kloareg, M., Noblet, J., and van Milgen, J. (2007), "Deposition of Dietary Fatty Acids, *De*
548 *Novo* Synthesis and Anatomical Partitioning of Fatty Acids in Finishing Pigs," *British*
549 *Journal of Nutrition*, 97, 35-44.

550Larsson, S. C., Kumlin, M., Ingelman-Sundberg, M., and Wolk, A. (2004), "Dietary Long-
551 Chain n-3 Fatty Acids for the Prevention of Cancer: A Review of Potential

552 Mechanisms," *American Journal of Clinical Nutrition*, 79, 935-945.

553Leseigneur-Meynier, A., and Gandemer, G. (1991), "Lipid Composition of Pork Muscle in
554 Relation to the Metabolic Type of the Fibres," *Meat Science*, 29, 229-241.

555Martín-Fernández, J. A., and Thió-Henestrosa, S. (2006), "Rounded zeros: some practical
556 aspects for compositional data," in *Compositional Data Analysis in the Geosciences:*
557 *From Theory to Practice*, eds. A. Buccianti, G. Mateu-Figueras, and V. Pawlowsky-
558 Glahn, Geological Society Special Publications 264, 191-201, London, UK.

559Ntawubizi, M., Colman, E., Janssens, S., Raes, K., Buys, N., and De Smet, S. (2010),
560 "Genetic Parameters for Intramuscular Fatty Acid Composition and Metabolism in
561 Pigs," *Journal of Animal Science*, 88, 1286-1294.

562Palarea-Albaladejo, J., Martín-Fernández, J. A., and Gómez-García, J. (2007), "A Parametric
563 Approach for Dealing with Compositional Rounded Zeros," *Mathematical Geology*, 39,
564 625-645.

565Pawlowsky-Glahn, V., and Egozcue, J. J. (2006), "Compositional Data and their Analysis: An
566 Introduction," in *Compositional Data Analysis in the Geosciences: From Theory to*
567 *Practice*, eds. A. Buccianti, G. Mateu-Figueras, and V. Pawlowsky-Glahn, Geological
568 Society Special Publications 264, 1-10, London, UK.

569Pearson, K. (1987), "Mathematical Contributions to the Theory of Evolution: On a Form of
570 Spurious Correlation Which May Arise when Indices are used in the Measurement of
571 Organs," *Proceedings of the Royal Society of London*, 60, 489-498.

572R Development Core Team (2010). "R: A Language and Environment for Statistical
573 Computing," R Foundation for Statistical Computing, Vienna, Austria. http://www.R-
574 project.org.

575Ros- Freixedes, R., Reixach, J., Tor, M., and Estany, J. (2012), "Expected Genetic Response
576 for Oleic Acid Content in Pork," *Journal of Animal Science*, 90, 4230-4238.

577Rule, D. C. (1997), "Direct Transesterification of Total Fatty Acids of Adipose Tissue, and of
578 Freeze-Dried Muscle and Liver with Boron-Trifluoride in Methanol," *Meat Science*, 46,
579 23-32.
580Sanford, R. F., Pierson, C. T., and Crovelli, R. A. (1993), "An Objective Replacement Method
581 for Censored Geochemical Data," *Mathematical Geology*, 25, 59-80.
582Schmid, A. (2010), "The Role of Meat Fat in the Human Diet," *Critical Reviews in Food*
583 *Science and Nutrition*, 51, 50-66.
584Sharma, N., Gandemer, G., and Goutefongea, R. (1987), "Comparative Lipid Composition of
585 Porcine Muscles at Different Anatomical Locations," *Meat Science*, 19, 121-128.
586Templ, M., Hron, K., and Filzmoser, P. (2011), "robCompositions: Robust Estimation for
587 Compositional Data. Manual and Package, Version 1.5.0," http://cran.r-project.org/web/
588 packages/robCompositions
589Thió-Henestrosa, S., and Martín-Fernández, J. A. (2005), "Dealing with Compositional Data:
590 The Freeware CoDaPack," *Mathematical Geology*, 37, 773-793.
591Tolosana-Delgado, R., and van der Boogaart, K. G. (2011), "Linear Models with
592 Compositions in R," in *Compositional Data Analysis: Theory and Applications*, eds. V.
593 Pawlowsky-Glahn, and A. Buccianti, John Wiley & Sons Ltd., West Sussex, UK.
594Valsta, L. M., Tapanainen, H., and Männistö, S. (2005), "Meat Fats in Nutrition," *Meat*
595 *Science*, 70, 525-530.
596van den Boogaart, K. G., Tolosana, R., and Bren, M. (2011), "compositions: Compositional
597 Data Analysis. Manual and Package, Version 1.10-2,"
598 http://www.stat.boogaart.de/compositions
599Wood, J. D., Enser, M., Fisher, A. V., Nute, G. R., Sheard, P. R., Richardson, R. I., Hughes,
600 S. I., and Whittington, F. M. (2008), "Fat Deposition, Fatty Acid Composition and Meat
601 Quality: A Review," *Meat Science*, 78, 343-358.

602Wood, J. D., Richardson, R. I., Nute, G. R., Fisher, A. V., Campo, M. M., Kasapidou, E.,
603 Sheard, P.R., and Enser, M. (2003), "Effects of Fatty Acids on Meat Quality: A
604 Review," *Meat Science*, 66, 21-32.
605Yang, K. X., Ma, J. W., Guo, Y. M., Guo, T. F., Zhao, Y. G., Ding, N. S., Betti, M., Plastow,
606 G. S., and Huang, L. S. (2010), "Correlations between Fat Depot Traits and Fatty Acid
607 Composition in Abdominal Subcutaneous Adipose Tissue and Longissimus Muscle:
608 Results from a White Duroc × Erhualian Intercross F2 Population," *Journal of Animal*
609 *Science*, 88, 3538-3545.
610Zhang, S., Knight, T. J., Stalder, K. J., Goodwin, R. N., Lonergan, S. M., and Beitz, D. C.
611 (2007), "Effects of Breed, Sex, and Halothane Genotype on Fatty Acid Composition of
612 Pork Longissimus Muscle," *Journal of Animal Science*, 85, 583-591.