

Scotland's Rural College

Compositional mixed modelling of methane emissions and ruminal volatile fatty acids from individual cattle and multiple experiments

Palarea-Albaladejo, J; Rooke, JA; Nevison, IM; Dewhurst, RJ

Published in:
Journal of Animal Science

DOI:
[10.2527/jas.2016.1339](https://doi.org/10.2527/jas.2016.1339)

First published: 08/06/2017

Document Version
Peer reviewed version

[Link to publication](#)

Citation for published version (APA):

Palarea-Albaladejo, J., Rooke, J.A., Nevison, I.M., & Dewhurst, R.J. (2017). Compositional mixed modelling of methane emissions and ruminal volatile fatty acids from individual cattle and multiple experiments. *Journal of Animal Science*, 95(6), 2467 - 2480. <https://doi.org/10.2527/jas.2016.1339>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

1 Running head: Methane emissions and volatile fatty acids

2

3

4

5

6

7

8

9

10

11

12

**Compositional mixed modelling of methane emissions
and ruminal volatile fatty acids from individual cattle and multiple experiments¹**

J. Palarea-Albaladejo,^{*2} J. A. Rooke,[†] I. M. Nevison,^{*} and R. J. Dewhurst[†]

^{*}Biomathematics and Statistics Scotland, Edinburgh, Scotland, United Kingdom, EH9 3FD

[†]Future Farming Systems, SRUC, West Mains Road, Edinburgh. United Kingdom, EH9 3JG

¹This research was funded by the Scottish Government's Rural and Environment Science and Analytical Services Division (RESAS), AHDB Beef and Lamb, and by DEFRA and the devolved administrations through the UK Agricultural Greenhouse Gas Inventory Research Platform. J. Palarea-Albaladejo was also supported by the Spanish Ministry of Economy and Competitiveness through the project "CODA-RETOS" (Ref. MTM2015-65016-C2-1-R; MINECO/FEDER, UE).

²Corresponding author: javier.palarea@bioss.ac.uk

13 **ABSTRACT:** The aim of the study was to investigate the association of methane (CH₄)
14 yields (g/kg DMI) with rumen VFA molar proportions and animal and diet-related covariates
15 from individual animals and multiple experiments. The dataset available consisted of 284
16 measurements of CH₄ yields for beef cattle from 6 experiments measured in indirect
17 respiration chambers. A compositional modelling approach was employed where VFA
18 measurements were considered as a whole, instead of in isolation, emphasizing their
19 multivariate relative scale. The analysis revealed expected close groupings of acetate and
20 butyrate; propionate and valerate; iso-butyrate and iso-valerate. Linear mixed models were
21 then fitted to examine relationships between CH₄ yield and VFA, represented by meaningful
22 log-contrasts of components called compositional balances, while accounting for other
23 animal and diet-related covariates and random variability between experiments. A
24 compositional balance representing (acetate · butyrate)/propionate best explained the
25 contribution of VFA to variation in CH₄ yield. The covariates DMI, forage:concentrate
26 proportion (expressed as a categorical variable diet type: high concentrate, mixed
27 forage:concentrate or high forage), and diet ME were also statistically significant. These
28 results provided new insights into the relative inter-relationships amongst VFA
29 measurements and also between VFA and CH₄ yield. In conclusion, VFA molar proportions
30 as represented by compositional balances were a significant contributor to explaining
31 variation in CH₄ yields from individual cattle.

32 **Key words:** methane production, volatile fatty acids, compositional data, mixed models.

33

INTRODUCTION

34
35 Methane (CH₄), derived almost entirely (90%) from enteric fermentation, is a major
36 contributor to greenhouse gas emissions from the livestock sector, and cattle are responsible
37 for most (77%) of CH₄ emissions (Gerber et al., 2013). While modifying the diet of cattle is
38 the most effective short-term method for mitigating CH₄ emissions, longer-term the wide
39 variation in CH₄ yield (g CH₄/kg DMI) between individual animals (up to 2-fold when fed
40 the same diet, Rooke et al., 2014) must be exploited. A limitation to this is the relatively slow
41 output achievable using indirect respiration chambers to measure CH₄ yield and the cost of
42 such measurements. The amount of CH₄ produced from a specific diet by rumen archaea
43 depends largely upon the amount of hydrogen (H₂) produced as an end-product of
44 fermentation of feed carbohydrates by other organisms in the rumen microbiome. There are
45 well established stoichiometric relationships between the pattern of VFA and H₂ produced by
46 rumen fermentation and resulting CH₄ formation (Wolin, 1960; Murphy et al., 1982; Alemu
47 et al., 2011), but these stoichiometric relationships have usually been modelled using data at
48 the diet level. In the current study, we use CH₄ and VFA measurements for individual
49 animals from 6 experiments to address the hypothesis that VFA pattern in an individual
50 animal could be used as an explanatory variable in accounting for variation in CH₄ yield in
51 addition to other diet and animal characteristics. A key methodological novelty is that we
52 consider VFA (expressed as molar proportions) as a composition of intrinsically co-
53 dependent amounts carrying only relative information. This was embedded into a linear
54 mixed modelling framework to account for the variability originating from the multiple study
55 structure of the data.

56

57

MATERIALS AND METHODS

58

59

60

61

62

The experiments which were included in the database were conducted at Scotland's Rural College (SRUC) Beef and Sheep Research Centre in Edinburgh. Each individual experimental protocol was approved by SRUC's Animal Welfare and Ethical Review Body, the Animal Experiments Committee and was conducted in accordance with the requirements of the UK Animals (Scientific Procedures) Act, 1986.

63

Description of the Data

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

The data analyzed here were obtained from 6 experiments carried out at SRUC between 2011 and 2014. The cattle used were either steers (Rooke et al., 2014, Troy et al., 2015, 2016, Duthie et al., 2017; exp. 1–4) or beef cows (Duthie et al., 2015; exp. 5 (unpublished) and of varying breed types (Aberdeen Angus x Limousin; Limousin x Aberdeen Angus; Luining; Charolais cross bred; exp. 6, Aberdeen Angus x Limousin and Limousin x Aberdeen Angus). Diets fed were of 3 types which were used as a categorical variable, diet type, in analysis. High concentrate diets (Concentrate; < 100 g forage/kg DM) were based on ground barley and barley straw with either rapeseed meal or distillers dark grains (similar to distillers grains with solubles but low in sulfur). Mixed forage:concentrate diets (Mixed; 400 – 600 g forage/kg DM) were based on grass silage, whole crop barley silage, barley, and either rapeseed meal or distillers dark grains. High forage diets (Forage; > 700 g forage/kg DM) consisted of barley straw and either grass silage or brewers' grains (Duthie et al., 2015) or a mixture of grass silage (696 g/kg DM) and whole crop barley silage (293 g/kg DM) in exp. 5. All diets were offered ad libitum as total mixed rations (fed once daily) and had been fed for at least 3 weeks before measurements of CH₄ yield. Methane output was measured (48 h) using indirect respiration chambers. As it was not possible to take rumen samples whilst animals were in the chambers, a single sample of rumen fluid for VFA analysis was taken by stomach tube within 2 h of cattle leaving the chambers

82 (approximately 25 h after feed bins were last refilled). Full details of methodology can be
83 found in Rooke et al. (2014); Duthie et al. (2015) and Troy et al. (2015). The constituents of
84 the diets are summarized in Table 1 and mean cattle BW, DMI, and CH₄ production in Table
85 2. It should be noted that only animals which had complete records of DMI, CH₄, and VFA
86 were used in the current analysis; thus, there may be small discrepancies from the original
87 references.

88 TABLE 1 (DIET CONSTITUENTS)

89 TABLE 2 (SUMMARY OF VARIABLES)

90 ***VFA Measurements as Compositional Data***

91 Data accounting for relative parts of a whole are known as compositional data (see
92 e.g. Aitchison, 1986; Pawlowsky-Glahn et al., 2015 for discussion, formal properties and
93 principles). This is the case for chemical mixtures such as VFA composition when expressed
94 as portions of the whole in either weight or volume units. Thus, the components of the VFA
95 mixture are intrinsically co-dependent positive amounts carrying only relative information.
96 Changes in one or several components affect the remaining ones and, consequently, an
97 equivalent symmetric overall change should be measured on the latter as well. The relative
98 and symmetric scale is commonly recognized in practice by re-expressing the data in
99 proportions adding up to 1 or similar units like mmol/mol used here. It is important to note
100 that the data do not necessarily have to add up to the same constant total. Given a D -part
101 composition $\mathbf{x} = [x_1, \dots, x_D]$, the statistical analysis focuses on the log-ratios $\ln(x_i/x_j)$
102 between components. Using this type of transformation, the results do not depend on the
103 units of measurement and the relative scale of the data is considered. It does not matter either
104 whether the original full mixture or a subset (subcomposition) of components, which may or
105 may not add to total VFA, are used as here.

106 ***Basic Compositional Statistics for the VFA Data Set***

107 Following Pawlowsky-Glahn and Egozcue (2002), the composition best representing
108 the center of a data set $\mathbf{X} = [x_{ij}]_{n \times D}$ consisting of n compositional samples \mathbf{x}_i of D
109 components is given by the so-called compositional geometric mean (**CGM**), or
110 compositional center, as $\text{CGM}(\mathbf{X}) = C(g_1, \dots, g_D)$, where $g_j = (\prod_{i=1}^n x_{ij})^{1/n}$, for $j =$
111 $1, \dots, D$, is the geometric mean of the the j th column of \mathbf{X} . The closure operator C normalizes
112 the resulting vector of geometric means to be expressed in the chosen scale. For example, if
113 working with proportions, C would mean multiplying each component by $1/\sum_{j=1}^D g_j$ so that
114 the total sum of each composition is 1. Moreover, instead of using ordinary correlations, the
115 relative variability structure of \mathbf{X} is given by the matrix of log-ratio variances $\mathbf{T} = [\tau_{ij}]_{D \times D}$,
116 where $\tau_{ij} = \text{var}(\ln(x_i/x_j))$, for $i, j = 1, \dots, D$, with var referring to the ordinary variance
117 measure. Relationships between components are then understood in terms of proportionality.
118 A log-ratio variance which is close to 0 indicates that 2 components x_i and x_j are nearly
119 proportional (highly co-dependent); that is, their log-ratio is nearly constant. A measure of
120 global dispersion is provided by the total variance $\text{totvar}(\mathbf{X}) = 1/2D \sum_{i,j=1}^D \tau_{ij}$. Estimators
121 of these measures from data are obtained by the standard maximum likelihood procedure.

122 ***Log-ratio Coordinate Representation***

123 Recent advances in the area (Egozcue et al., 2003) allow the definition of isometric
124 (metric-preserving) mappings between the original composition \mathbf{x} and log-ratio coordinates
125 living in the ordinary real space, which facilitates the use of standard statistical methods.
126 These isometric log-ratios (**ilr**) can be constructed in infinitely many different ways, although
127 they are essentially orthogonal rotations of one to another. This means that results from any
128 set of ilr coordinates transform back into the same results in terms of the original
129 composition. Note that compositions of size D correspond to vectors of isometric log-ratio
130 coordinates of size $D - 1$ (the actual degrees of freedom of the composition). A procedure
131 known as sequential binary partition (**SBP**; Egozcue and Pawlowsky-Glahn, 2005) allows the

132 construction of tailored ilr coordinates, usually called compositional balances (b_i , where
 133 $i = 1, \dots, D - 1$), representing log-contrasts between subsets of components of \mathbf{x} . This is of
 134 great practical relevance because balances can be defined according to insights from
 135 exploratory data analysis (see Model A1 below) or using biological knowledge.

136 ***Building Balances for the VFA Composition***

137 Balances are obtained using SBP by successive splits of the components of the VFA
 138 composition \mathbf{x} into 2 mutually exclusive groups until only groups of 1 component are left (see
 139 left-hand side of Table 4). These two groups are coded by the signs + and – respectively. The
 140 collection of $D - 1$ balances b_i , for $i = 1, \dots, D - 1$, is obtained as

$$b_i = \sqrt{\frac{r_i s_i}{r_i + s_i}} \ln \frac{(\prod_{k=1}^{r_i} x_{ik}^+)^{1/r_i}}{(\prod_{k=1}^{s_i} x_{ik}^-)^{1/s_i}}, \quad [1]$$

141 where x_{ik}^+ and x_{ik}^- refer to the subsets of r_i and s_i components going, respectively, into the +
 142 (numerator) and – (denominator) groups. The $D - 1$ balances fully represent the information
 143 in the composition \mathbf{x} and, as previously said, they are appropriate to be used in standard
 144 statistical modelling. Note that the log-ratio term in Eq. [1] is computed as the ratio
 145 between the geometric means of the corresponding + and – components. It is multiplied by a
 146 normalizing constant to give the b_i that allow balances to be compared and confers on them
 147 desirable geometric properties. A balance then measures the relative importance, in geometric
 148 mean, of one group against the other by means of a log-contrast between them. Table 4 (left)
 149 depicts two alternative but equivalent balance representations of the VFA composition used
 150 in this work to produce a linear mixed model for CH₄ emissions. For example, the balance b_1
 151 in Model A1 was computed as

$$b_1 = \sqrt{\frac{8}{6}} \ln \frac{(\text{iso-butyrate} \cdot \text{iso-valerate} \cdot \text{butyrate} \cdot \text{acetate})^{1/4}}{(\text{valerate} \cdot \text{propionate})^{1/2}}. \quad [2]$$

152 In Model A2, the SBP is modified to isolate a balance of particular biological interest

153 (balance b_2 in Model A2) so that its significance can be statistically tested. Note that,
 154 regardless of the balance representation chosen, it holds that $\text{totvar}(\mathbf{X}) = \sum_{i=1}^{D-1} \text{var}(b_i)$ for a
 155 given compositional data set \mathbf{X} . Thus, the balances can be ranked according to their
 156 contribution to explaining the total variability within the data set.

157 *A Compositional Linear Mixed Model for CH₄ Emissions*

158 A linear mixed model (LMM) approach was adopted to integrate quantitative
 159 findings from the 6 different studies. The linear association between CH₄ emissions
 160 expressed as g/d or g/kg DMI (see Table 2) was moderately high and positive (Pearson's
 161 $r = 0.75$). Accordingly, similar estimates of the effects of the given explanatory variables
 162 would be obtained. For the purpose of this work we used CH₄ yields expressed in g/kg DMI
 163 as the response variable. Methane yield was log transformed to more closely satisfy the
 164 normality assumption of the model residuals. Formally, the response vector \mathbf{y}_i from the i th
 165 experiment was modelled as

$$\begin{aligned} \mathbf{y}_i &= \mathbf{B}_i \cdot \beta_1 + \mathbf{C}_i \cdot \beta_2 + \mathbf{Z}_i \cdot f_i + \varepsilon_i, \quad i = 1, \dots, 6, \\ f_i &\sim N(0, \sigma_f^2), \\ \varepsilon_i &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \end{aligned} \quad [3]$$

166 where β_1 and β_2 were the coefficients of the fixed effects associated with, respectively, (a)
 167 the VFA balances (\mathbf{B}_i matrix) as obtained from Eq. [1] given a SBP, and (b) the other
 168 explanatory covariates (\mathbf{C}_i matrix). The BW and DMI summarized the information about
 169 animal covariates in the model. For diet-related covariates, diet type (Concentrate, Mixed or
 170 Forage) was included as an explanatory factor representing the diet contents (forage, starch,
 171 and neutral detergent fiber) across experiments, along with metabolizable energy (ME). All
 172 the covariates were log transformed to be introduced in the model. The term $\mathbf{Z}_i \cdot f_i$ was the
 173 experiment random effects term, with f_i assumed to be normally distributed with mean 0 and
 174 variance σ_f^2 . The within-group random errors ε_i were assumed to be normally distributed

175 with means 0 and variances σ^2 . Random variability of model intercepts between experiments
176 was assumed, which implied that the design matrix of the random effects \mathbf{Z}_i equaled a unit
177 vector. The random effects f_i and the random errors ε_i were assumed to be independent for
178 different experiments and independent of each other for the same experiment.

179 Model fitting was conducted by restricted maximum likelihood estimation. The
180 marginal statistical significance of the fixed effect coefficients was assessed by conditional t -
181 tests (Pinheiro and Bates, 2000). Conditional F -tests were applied to jointly test for the
182 significance of the VFA balance coefficients and, hence, of the VFA composition. Statistical
183 significance was concluded when associated P -values were < 0.05 . An approximate model
184 goodness-of-fit measure for mixed models was provided by using the marginal and
185 conditional R^2 coefficients for mixed models (R_m^2 , % variance explained by fixed terms; and
186 R_c^2 , % variance explained by both fixed and random terms) proposed in Nakagawa and
187 Schielzeth (2013). Note however that, due to the complications added by the random effects
188 structure in mixed models, these are pseudo- R^2 coefficients. Hence, interpretation and
189 comparison of model explained variances based on these coefficients must be conducted with
190 extreme caution. Comparison of models with nested fixed effect structures and the same
191 random effect structure was conducted using the Akaike information criterion (AIC) and
192 likelihood ratio test (LRT) as obtained from maximum likelihood estimation of the models.
193 Statistical analyses, including compositional analyses, and modelling were conducted in the
194 R system for statistical computing v3.2 (R Core Team, 2016).

195

196

RESULTS

197 *Exploratory Analysis of the VFA Composition*

198 Table 3 shows compositional summary statistics for the VFA composition across all
199 the experiments (other ordinary statistics for individual diets are supplied in Appendix 1).

200 The overall CGM reveals that, as expected, acetate was the most abundant VFA (mean, 656
201 mmol/mol), whereas iso-butyrate, iso-valerate, and valerate were all present at < 20
202 mmol/mol. The results per diet type illustrate the differences in mean VFA profiles between
203 them. The variation matrix indicates that acetate and butyrate held the strongest
204 proportionality association ($\tau = 0.10$). Contrarily, propionate and butyrate or iso-valerate
205 were the least proportionally associated components ($\tau = 0.29$).

206

TABLE 3 (VFA SUMMARY)

207 The relative variation structure of the data and the relationships between samples
208 (points) and VFA components (rays) were approximately represented in Fig. 1 using a
209 compositional biplot (Aitchison and Greenacre, 2002) explaining 68% of the data variability.
210 The lengths of the links between arrowheads approximate the log-ratio variances (Table 3) of
211 the corresponding components. Thus, propionate showed in general the greatest log-ratio
212 variances (lowest proportionality) with all the others, particularly with iso-valerate and
213 butyrate (relationship highlighted using dashed lines). The samples were distinguished by
214 diet type. Concentrate diet type mostly associated with greater relative amounts of valerate
215 and propionate, Forage diet type mostly linked to acetate and the Mixed diet type
216 intermediate between Concentrate and Forage diet types but also linked to greater relative
217 amounts of butyrate and iso-valerate.

218

FIGURE 1 (BIPLOT)

219 The variation matrix in Table 3 was used to arrange the VFA components into
220 homogenous groups by the Ward's clustering method (Ward, 1963). This allowed for a
221 hierarchical representation of the structure of proportionality relationships between VFA
222 components in a dendrogram (Fig. 2).

223

FIGURE 2 (DENDROGRAM)

224 Three groups of VFA components can be clearly distinguished: propionate and
225 valerate, acetate and butyrate, and iso-butyrate and iso-valerate. This configuration is
226 coherent with the biplot analysis above and can be used to define a set of compositional
227 balances between VFA components based only on the proportionality structure inferred from
228 the data. In particular, the balance contrasting propionate and valerate against the remaining
229 VFA components (balance b_1 in Eq. [2]), corresponding to the top split in Fig. 2, explained
230 most of the total variability (totvar) in the data set (34.42%) and was used as starting point to
231 generate a collection of balances through SBP (Table 4, SBP for Model A1). Note that the
232 subsequent balances b_2, b_3, b_4 , and b_5 corresponded to log-contrasts between the components
233 located at each of the two branches of the nodes of the dendrogram as indicated in Fig 2.

234 *Associations between CH₄ Emissions, VFA Composition, and Diet and Animal Covariates*
235 *Model A1.*

236 Table 4 (top, Model A1) shows parameter estimates and statistical significances from
237 the model for CH₄ yield based on VFA balances according to the relative variation structure
238 of the data as detailed above. All the balances in Model A1 but b_5 ($P = 0.832$) and b_4
239 ($P = 0.338$) were statistically significant. A joint test for the coefficients of the balances
240 confirmed a statistically significant effect of the VFA composition as a whole ($F =$
241 9.931 ; $P < 0.001$). The Ln BW was the only covariate not statistically significant ($P =$
242 0.129). The marginal and conditional R^2 coefficients (64.65% and 68.60% respectively)
243 reflected an acceptable model fit. Ordinary checks for normality and homogeneity of
244 variances of the model residuals were satisfactory (not shown).

245 TABLE 4 (CLMM results)

246 *Model A2.*

247 The fact that some balances were not statistically significant in Model A1 raised the
248 question of whether some VFA components were not relevant to explain CH₄ yield. Some

249 further exploratory analysis suggested that the SBP used for Model A1 could be refined by
250 using a balance representing the main VFA responsible for H₂ production (acetate and
251 butyrate) and consumption (propionate) during carbohydrate fermentation. This was achieved
252 by defining the SBP shown in Table 4 (bottom) for Model A2, with a new balance b_2 given
253 by

$$b_2 = \sqrt{\frac{2}{3}} \ln \frac{(\text{acetate} \cdot \text{butyrate})^{1/2}}{\text{propionate}}. \quad [4]$$

254 Under this arrangement, which is simply a re-parametrization of Model A1, only the
255 new balance b_2 was statistically significant. This result clearly picks out the log-contrast
256 acetate-butyrate versus propionate (Eq. [4]) as the main driver of the relationship between
257 CH₄ yield and VFA composition. It is important to stress at this point that Model A2 is
258 entirely equivalent to Model A1, except for the use of a different set of ilr coordinates which
259 allowed to test for the significance of a relationship of biological interest and enhanced
260 interpretability. Overall measures, R^2 coefficients, AIC, model intercepts, estimates for the
261 remaining covariates and random effects estimates were all identical. The technical reason for
262 this is that different ilr transformations are orthogonal rotations of each other, and those
263 estimates are invariant under such rotations. The AIC measure was used to rank the
264 explanatory variables according to their relative importance by the sum of Akaike weights
265 (Burnham and Anderson, 2002) over all possible models from Model A2 in which the
266 variable was included. Note that all the statistically significant variables (diet type, Ln DMI,
267 Ln ME, and b_2 ; see Table 4) were given the same Akaike weight of 1 (values range between
268 0 and 1) and, hence, they were all considered of analogous importance in the model.

269 **Model B.**

270 Instead of considering the entire VFA composition, we fitted here a simplified LMM
271 for CH₄ yield based only on balance b_2 from Model A2 (Eq. [4]; summary statistics in

272 Appendix 1) and the statistically significant covariates (see parameter estimates in Table 5).
273 A LRT to compare Model A2 with Model B provided a statistically non-significant result
274 (LR statistic = 6.36; $P = 0.273$) and, hence, supported the use of the simplified model
275 following the principle of parsimony. The estimated β coefficients for the covariates were
276 essentially the same as obtained before. Note that a model not including b_2 as explanatory
277 variable but including the other covariates produced overall estimates $R_m^2 = 57.11\%$,
278 $R_c^2 = 62.33\%$ and AIC = -170.28. The corresponding LRT (LR statistic = 42.87; $P < 0.001$)
279 supported the inclusion of the VFA compositional balance b_2 in the model.

280 TABLE 5 (MODEL B RESULTS)

281 The relationships between CH₄ yield and the subset of VFA components involved in
282 Eq. [4] are depicted in Fig. 3. Propionate was entered independently whereas acetate and
283 butyrate concentrations, on the one hand, and the remaining VFA components, on the other
284 hand, were entered together by geometric mean to obtain the [propionate, acetate · butyrate,
285 others] subcomposition. The observed values were displayed on a ternary diagram, with the
286 diet types distinguished by color and shape. Model B was used to produce expected Ln CH₄
287 yields from randomly generated samples of this subcomposition. An interpolated surface was
288 fitted to them and used to fill the ternary diagram with colors according to the values.

289 FIGURE 3 (TERNARY DIAGRAM)

290 Each vertex of the ternary diagram corresponds with corner VFA subcompositions
291 consisting of 100% of the component represented in there and 0% of any other. Maximum
292 data variability occurred along the propionate to (acetate · butyrate) direction, with little
293 variability observed in the direction pointing to the other vertex. The lowest expected CH₄
294 yields were associated with the greatest propionate concentrations, whereas the greatest
295 expected CH₄ yields were associated with the greatest (acetate · butyrate) concentrations.

296 Using Model B as basis, separate LMMs were fitted to check whether the relationship
297 between the disclosed VFA balance (Eq. [4]) and CH₄ yield was reproduced within a diet
298 type. For both Concentrate and Mixed diet types we obtained a similar statistically highly
299 significant positive estimated β coefficient for the balance representing the (acetate
300 butyrate) / propionate ratio ($P < 0.001$), thus supporting the association once the effects of
301 DMI and ME were accounted for within these diet types (see summary tables in Appendix 2).
302 As expected the amount of unexplained random variability increased in relation to the overall
303 Model B. The effect of the VFA balance was however negligible within the Forage diet type
304 ($\hat{\beta} = 0.073$; $P = 0.704$; summary table not shown). The number of animals on the Forage
305 diet type was lower (45 animals versus 88 and 151 on the Concentrate and Mixed diet types
306 respectively) and the forage constituents of the 2 trials which contributed to the Forage diet
307 type were very different, which resulted in a sparse data set for this case.

308

309

DISCUSSION

310 The use of compositional methods within the natural sciences is rapidly increasing,
311 with applications found in the study of environmental pollution (Howel, 2007), aroma
312 volatile compounds (Korhonová et al., 2009), meat fatty acid composition (Ros-Freixedes
313 and Estany, 2014) and free-ranging animal diets (Stewart et al., 2014), among others. It has
314 been shown that treating the components in isolation, without relating them to each other,
315 may lead to misleading or paradoxical conclusions. For example, the standard linear
316 correlation measured between same 2 components can dramatically vary depending on the
317 other components considered in the composition (Palarea-Albaladejo and Martín-Fernández,
318 2013). As an illustration of this, using our own VFA data, the correlation between iso-
319 butyrate and iso-valerate is 0.62 when the entire 6-component VFA composition is
320 considered, whereas it is -0.87 when working with only the [iso-butyrate, iso-valerate,

321 valerate] subcomposition. Other common inconsistencies include statistical confidence
322 intervals covering nonsensical negative values or singularity and collinearity problems in
323 linear models (Fox, 1997; Hron et al., 2012). In this work, we introduced compositional
324 linear mixed models to investigate the association between rumen VFA and CH₄ yield from
325 individual cattle, while accounting for the effects of other animal and diet-related covariates
326 and the random variation originating from multiple experiments. Meaningful normalized log-
327 contrasts (compositional balances) were defined between subsets of VFA components and
328 their contribution to variability in CH₄ yields investigated.

329 Exploratory analysis revealed interesting proportionality associations between the
330 VFA components and their connections with different diet types (Fig. 1). The VFA
331 composition as a whole had a statistically significant association with CH₄ yield. The links
332 between VFA components depicted in Fig. 2, and used in Model A1 (Table 4), described well
333 the underlying stoichiometry of carbohydrate fermentation (Wolin, 1960). Thus, VFA
334 associated with H₂ production, acetate and butyrate, were closely related and well separated
335 from propionate and valerate which are associated with H₂ consumption. The branched chain
336 VFA, iso-butyrate and iso-valerate were also closely related as might be expected given that
337 these VFA are products of the catabolism of branched-chain AA. The closer alignment of the
338 branched chain VFA with acetate and butyrate than with propionate and valerate is consistent
339 with the requirement for these iso-acids by structural carbohydrate fermenting bacteria such
340 as *Ruminococcus albus* (Allison and Bryant, 1963; Liu et al., 2014; Wang et al., 2015).
341 Acetate, propionate, and butyrate are quantitatively the most important VFA in the rumen
342 environment. When a log-contrast representing these 3 VFA (Eq. [4]) was used in Model A2
343 (Table 4), it was the only compositional balance having a statistically significant effect on
344 CH₄ yield. Note that it is analogous to the so-called glucogenic ratio [(acetate + butyrate) /
345 propionate]; the β coefficient was positive as would be expected because H₂ available for

346 CH₄ formation would be expected to be positively correlated with this ratio. Indeed, this is
347 the best empirical representation of the mechanistic relationships proposed by Wolin (1960)
348 relating CH₄ production to VFA. The relationship was consistent across **concentrate and**
349 **mixed** diet types. Thus, compositional analysis produced relationships between VFA which
350 described accurately the underlying biology. Note that using an alternative SBP to isolate a
351 balance between butyrate only and propionate (in the numerator and denominator
352 respectively of Eq. [1]) provided results very similar to those using the balance in Eq. [4]. We
353 also implemented a SBP to test the compositional balance representing the exchange between
354 acetate and propionate only, which has been advocated in Janssen (2010) and Sauvart et al.
355 (2011). Its effect on CH₄ yield was highly statistically significant as well ($P = 0.005$),
356 however in this case it was not the only statistically significant balance in the model and,
357 hence, the results were not so neat. This would then confer butyrate concentrations relative to
358 propionate a leading role in the balance given by Eq. [4].

359 In developing models, other animal (DMI, BW) and nutritional (diet type, ME)
360 covariates were tested and DMI, diet type, and ME were included and their influence on the
361 model was in the direction expected. Thus, DMI was associated with a negative β coefficient
362 recognizing that increased DMI is associated with increased rumen outflow and decreased
363 extent of fermentation and therefore CH₄ production. Diet type was included as the diets fed
364 were grouped into three distinct forage to concentrate ratios. The use of starch and NDF
365 concentrations and the log-ratio NDF/starch were tested directly as covariates. However, they
366 were highly related to each other and the information about the variation in chemical
367 composition of diets was robustly represented by ME and diet type. From a statistical point of
368 view, using diet type also provided greater consistency and numerical stability of the
369 estimation process. Diet type was associated with a positive β coefficient as it changed from
370 the Concentrate to the Mixed and then to the Forage diet type as expected from the mean

371 values for CH₄ yield for the Concentrate (14.7 g/kg DMI) and Mixed (22.2 g/kg DMI) diet
372 types (derived from Table 2). At first sight the positive β coefficient for the Forage diet type
373 appears anomalous as mean CH₄ yield (17.6 g/kg DMI) was less than for the Mixed diet type.
374 High forage diets are normally associated with high CH₄ yields because the high structural
375 fiber content of these diets produces an acetate-dominated fermentation which was indeed
376 observed (Table 3). The explanation for this apparent anomaly is that 2 diets in the Forage
377 diet type had low ME concentrations (Table 1). Thus, the positive β coefficient associated
378 with ME in the model likely adjusts responses for the digestibilities of these diets.

379 Many empirical models relating animal and diet variables to CH₄ have been produced
380 (recent examples include Hristov et al., 2013a; Storlien et al., 2014; Ricci et al., 2013; Ramin
381 and Huhtanen, 2013). Where CH₄ production is scaled either as g/kg DMI (as in the present
382 study) or as kJ/MJ GE intake (Hristov et al., 2013a; Ramin and Huhtanen, 2013), terms
383 related to DMI, diet digestibility and diet composition have been included in models. The
384 inclusion of VFA in empirical models has been less common, largely because of scarcity of
385 data. In some models diet composition has been used. Alemu et al., (2011) used VFA molar
386 proportions predicted from the stoichiometry of fermentation to explain CH₄ and found that
387 goodness of fit was model dependent. Ramin and Huhtanen (2013) compared the goodness of
388 fit for CH₄ from models which including observed VFA proportions and concluded as here
389 that models which included terms relating to combinations of VFA (acetate, propionate, and
390 butyrate) gave superior explanations of CH₄ yield than individual VFA. However, all models
391 noted above did not consider the compositional aspect of the data, that is, their natural
392 relative and symmetric scale, and were based on treatment means for diets and did not use
393 individual animal data in their analysis. Indeed, use of individual animal data in analysis of
394 CH₄ production is not common. Mills et al. (2001) compared observed CH₄ production with
395 that predicted from the mechanistic models of Dijkstra et al. (1992) and found relationships

396 were superior at the treatment than individual animal level; however, data for VFA
397 proportions was not reported. Robinson et al. (2010) generated a range of VFA
398 concentrations by varying DMI of sheep ($n = 10$) fed a single diet of lucerne chaff and found
399 that VFA concentrations only accounted for 25 – 30% of the variance. The present study was
400 not comparable with Robinson et al. (2010) because of the narrow range of VFA molar
401 proportions (700 – 730 mmol/mol acetate) used in that study. In the current study, a much
402 greater variation in CH_4 yields and range of VFA proportions was available.

403 In reviewing mitigation options for reducing enteric CH_4 emissions, Hristov et al. (2013b)
404 classified strategies into manure management practices and animal husbandry (which
405 included genetics). Hristov et al. (2013b) noted that from genetic variation a reduction in
406 predicted CH_4 production in the order of 11 to 26 percent was theoretically possible and that
407 genomic selection tools could further increase the reduction in CH_4 production. However,
408 effective application of genomic selection required significant international effort and
409 collaboration to bring together relevant data because of the large datasets required. Because
410 phenotypic CH_4 measurements are produced mainly using indirect respiration chambers,
411 genetic progress is limited by slow throughput and cost. **Therefore, there is a need for indirect
412 proxy measurements for CH_4 emissions which are capable of rapid throughput and lower cost
413 than chambers and these have been recently reviewed by Negussie et al. (2017) including
414 critical appraisal of their limitations.** The rumen samples which were used for measurement
415 of VFA in the current study were of necessity single spot samples and, therefore, would not
416 have captured changes in response to feed intake and fermentation. However, as cattle had
417 not been given access to fresh feed for 25 h when sampled, variation due to short term feed
418 intake would have been minimized and variability in VFA molar proportions between
419 animals and experiments due to differences in feeding pattern reduced. Further cattle are

420 often fasted before slaughter and therefore results from the current study could potentially
421 apply to samples taken at slaughter.

422 Although we found the association between VFA composition and CH₄ yield highly
423 statistically significant ($P < 0.001$), particularly through the (acetate · butyrate) / propionate
424 compositional balance, there were other terms involved and the dataset included only beef
425 cattle. If the use of VFA were implemented in practice, standardized protocols, such as those
426 for determining residual feed intake (Basarub et al., 2003) may be appropriate. Other factors
427 including rumen pH, protozoal population, and substrate utilization by methanogens, related
428 to rumen fermentation of individual animals not captured by the VFA balances are probably
429 responsible for unexplained variation and have been reviewed by Ellis et al. (2008). More
430 recently, the genetic background of the animal has been shown to be important (King et al.,
431 2011; Hernandez-Sanabria et al., 2013) as are inter-related phenotypic factors such as rumen
432 size (Goopy et al., 2013), feed intake pattern (Carberry et al., 2014) and colonization of the
433 rumen after birth (Yanez-Ruiz et al., 2015) which determine the host-specificity of the rumen
434 microbiome (Weimer et al., 2010; Wallace et al., 2015). Indeed, in one of the experiments
435 which contributed to the present data set (Rooke et al., 2016) the VFA pattern present in
436 individual animals before imposition of experimental treatments was a significant covariate
437 for subsequent samples. Apart from animal factors, differences in diet characteristics may
438 influence the fate of H₂ in the rumen and variability in VFA pattern. Where nitrate was used
439 as a mitigation strategy (e.g. Troy et al., 2015) reduction of nitrate to ammonia diverts H₂
440 away from VFA formation and increases the ratio of acetate to propionate. This is
441 contradictory to balance b_2 (Table 5) and may have contributed to the unexplained variation
442 in the overall model.

443

444

CONCLUSIONS

445 This work demonstrated the use and benefits of a novel statistical approach to the
446 analysis of VFA compositions from individual animals and multiple experiments. The results
447 were coherent with biological knowledge and emphasized the contribution of rumen VFA to
448 explain cattle CH₄ yield at an individual animal level. Further research is needed to determine
449 other possible contributing factors and investigate the scope for setting up more specialized
450 empirical models within the same compositional framework to improve predictive capacity
451 based on VFA measurements.

452

453

CONFLICT OF INTEREST

454 The authors state that there is no conflict of interest in relation to this work.

455

456

REFERENCES

457 Aitchison, J. 1986. *The Statistical Analysis of Compositional Data*. Chapman and Hall,
458 London.

459 Aitchison, J., and M. Greenacre. 2002. Biplots of compositional data. *J. R. Stat. Soc. Ser. C*
460 (Applied Stat.) 51:375–392. doi: 10.1111/1467-9876.00275.

461 Alemu, A. W., J. Dijkstra, A. Bannink, J. France, and E. Kebreab. 2011. Rumen
462 stoichiometric models and their contribution and challenges in predicting enteric
463 methane production. *Anim. Feed Sci. Technol.* 166–167:761–778. doi:
464 10.1016/j.anifeedsci.2011.04.054.

465 Allison, M. J. and M. P. Bryant. 1963. Biosynthesis of branched-chain amino acids from
466 branched-chain fatty acids by rumen bacteria. *Arch. Biochem. Biophys.* 101:269–277.
467 doi: 10.1016/S0003-9861(63)80012-0.

468 Basarab, J. A., M. A. Price, J. L. Aalhus, E. K. Okine, W. M. Snelling, and K. L. Lyle. 2003.
469 Residual feed intake and body composition in young growing cattle. *Can. J. Anim.*

- 470 Sci. 83:189–204. doi: 10.4141/A02-065.
- 471 Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a
472 practical information-theoretic approach. Springer, New York.
- 473 Carberry, C. A., D. A. Kenny, A. K. Kelly, and S. M. Waters. 2014. Quantitative analysis of
474 ruminal methanogenic microbial populations in beef cattle divergent in phenotypic
475 residual feed intake (RFI) offered contrasting diets. *J. Anim. Sci. Biotechnol.* 5(1):41.
476 doi: 10.1186/2049-1891-5-41.
- 477 Carberry, C. A., S. M. Waters, D. A. Kenny, and C. J. Creevey. 2014. Rumen methanogenic
478 genotypes differ in abundance according to host residual feed intake phenotype and
479 diet type. *Appl. Environ. Microbiol.* 80:586–594. doi: 10.1128/AEM.03131-13.
- 480 Casellas, J., and N. Ibáñez-Escriche. 2012. Bayesian recursive mixed linear model for gene
481 expression analyses with continuous covariates. *J. Anim. Sci.* 90:67–75. doi:
482 10.2527/jas.2010-3750.
- 483 Dijkstra, J., H. D. S. C. Neal, D. E. Beever, and J. France. 1992. Simulation of nutrient
484 digestion, absorption and outflow in the rumen - model description. *J. Nutr.*
485 122:2239–2256.
- 486 Duthie, C. A., J. A. Rooke, J. J. Hyslop, and A. Waterhouse. 2015. Methane emissions from
487 two breeds of beef cows offered diets containing barley straw with either grass silage
488 or brewers' grains. *Animal* 9:1680–1687. doi: 10.1017/S1751731115001251.
- 489 Duthie, C. A., J. A. Rooke, S. Troy, J. J. Hyslop, D. W. Ross, A. Waterhouse, and R. Roehe.
490 2016. Impact of adding nitrate or increasing the lipid content of two contrasting diets
491 on blood methaemoglobin and performance of two breeds of finishing beef steers.
492 *Animal* 10:786–795. doi: 10.1017/S1751731115002657.
- 493 Duthie, C. A., M. Haskell, J. J. Hyslop, A. Waterhouse, R. J. Wallace, R. Roehe, and J. A.
494 Rooke. 2017. The impact of divergent breed types and diets on methane emissions,

- 495 rumen characteristics and performance of finishing beef cattle. *Animal* (in press). doi:
496 10.1017/S1751731117000301.
- 497 Egozcue, J. J., V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal. 2003.
498 Isometric logratio transformations for compositional data analysis. *Math. Geol.*
499 35:279–300. doi: 10.1023/A:1023818214614.
- 500 Egozcue, J. J., and V. Pawlowsky-Glahn. 2005. Groups of parts and their balances in
501 compositional data analysis. *Math. Geol.* 37:795–828. doi: 10.1007/s11004-005-7381-
502 9.
- 503 Ellis, J. L., J. Dijkstra, E. Kebreab, A. Bannink, N. E. Odongo, B. W. McBride, and J. France.
504 2008. Aspects of rumen microbiology central to mechanistic modelling of methane
505 production in cattle. *J. Agric. Sci.* 146:213–233. doi: 10.1017/S0021859608007752.
- 506 Fox, J. 1997. *Applied regression analysis, linear models, and related methods*. SAGE
507 Publications, Inc.
- 508 Gerber, P.J., H. Steinfeld, B. Henderson, A. Mottet, C. Opio, J. Dijkman, A. Falcucci, and G.
509 Tempio 2013. *Tackling climate change through livestock – A global assessment of*
510 *emissions and mitigation opportunities*. Food and Agriculture Organization of the
511 United Nations, Rome, Italy.
- 512 Goopy, J. P., A. Donaldson, R. Hegarty, P. E. Vercoe, F. Haynes, M. Barnett, and V. H.
513 Oddy. 2014. Low-methane yield sheep have smaller rumens and shorter rumen
514 retention time. *Br. J. Nutr.* 111:578–585. doi: 10.1017/S0007114513002936.
- 515 Hernandez-Sanabria, E., L. A. Goonewardene, Z. Q. Wang, M. Zhou, S. S. Moore, and L. L.
516 Guan. 2013. Influence of sire breed on the interplay among rumen microbial
517 populations inhabiting the rumen liquid of the progeny in beef cattle. *PLoS One* 8(3):
518 e58461. doi: 10.1371/journal.pone.0058461.
- 519 Howel, D. 2007. *Multivariate data analysis of pollutant profiles: PCB levels across Europe*.

- 520 Chemosphere 67:1300–1307. doi: 10.1016/j.chemosphere.2006.11.025.
- 521 Hristov, A. N., J. Oh, J. L. Firkins, J. Dijkstra, E. Kebreab, G. Waghorn, H. P. S. Makkar, A.
522 T. Adesogan, W. Yang, C. Lee, P. J. Gerber, B. Henderson, and J. M. Tricarico.
523 2013a. SPECIAL TOPICS-Mitigation of methane and nitrous oxide emissions from
524 animal operations: I. A review of enteric methane mitigation options. *J. Anim. Sci.*
525 91:5045–5069. doi: 10.2527/jas.2013-6583
- 526 Hristov, A. N., J. Oh, C. Lee, R. Meinen, F. Montes, T. Ott, J. Firkins, A. Rotz, C. Dell, A.
527 Adesogan, W. Yang, J. Tricarico, E. Kebreab, G. Waghorn, J. Dijkstra, and S.
528 Oosting. 2013b. Mitigation of greenhouse gas emissions in livestock production – A
529 review of technical options for non-CO₂ emissions. In: P. Gerber, B. Henderson and
530 H. P. S. Makkar, editors *FAO Animal Production and Health Paper No. 177. Food*
531 *and Agricultural Organisation, Rome, Italy.*
- 532 Hron, K., P. Filzmoser, and K. Thompson. 2012. Linear regression with compositional
533 explanatory variables. *J. Appl. Stat.* 39:1115–1128. doi:
534 10.1080/02664763.2011.644268.
- 535 Janssen, P. H. 2010. Influence of hydrogen on rumen methane formation and fermentation
536 balances through microbial growth kinetics and fermentation thermodynamics. *Anim.*
537 *Feed Sci. Technol.* 160, 1–22. doi: <http://dx.doi.org/10.1016/j.anifeedsci.2010.07.002>.
- 538 King, E. E., R. P. Smith, B. St-Pierre, and A. D. G. Wright. 2011. Differences in the rumen
539 methanogen populations of lactating jersey and holstein dairy cows under the same
540 diet regimen. *Appl. Environ. Microbiol.* 77:5682–5687. doi: 10.1128/AEM.05130-
541 11.
- 542 Korhonová, M., K. Hron, D. Klimčíková, L. Müller, P. Bednár, and P. Barták. 2009. Coffee
543 aroma - statistical analysis of compositional data. *Talanta* 80:710–715. doi:
544 10.1016/j.talanta.2009.07.054.

- 545 Liu, Q., C. Wang, C. X. Pei, H. Y. Li, Y. X. Wang, S. L. Zhang, Y. L. Zhang, J. P. He, H.
546 Wang, W. Z. Yang, Y. S. Bai, Z. G. Shi, and X. N. Liu. 2014. Effects of isovalerate
547 supplementation on microbial status and rumen enzyme profile in steers fed on corn
548 stover based diet. *Livest. Sci.* 161:60–68. doi: 10.1016/j.livsci.2013.12.034.
- 549 Murphy, M. R., R. L. Baldwin, and L. J. Koong. 1982. Estimation of stoichiometric
550 parameters for rumen fermentation of roughage and concentrate diets. *J. Anim Sci.*
551 55:411–421. doi: 10.3168/jds.2012-6042.
- 552 Nakagawa, S., and H. Schielzeth. 2013. A general and simple method for obtaining R^2 from
553 generalized linear mixed-effects models. *Methods Ecol. Evol.* 4:133–142. doi:
554 10.1111/j.2041-210x.2012.00261.x.
- 555 Negussie, E., de Haas, Y., Dehareng, F., Dewhurst, R. J., Dijkstra, J., Gengier, N., Morgavi,
556 D. P., Soyeurt, H., van Gastelen, S., Yan, T. and Biscarini, F. 2017. Invited review:
557 Large-scale indirect measurements for enteric methane emissions in dairy cattle: A
558 review of proxies and their potential for use in management and breeding strategies. *J.*
559 *Dairy Sci.*, 100: 2433–2453. doi: 10.3168/jds.2016-12030.
- 560 Palarea-Albaladejo, J., and J. A. Martín-Fernández. 2013. Values below detection limit in
561 compositional chemical data. *Anal. Chim. Acta* 764:32–43. doi:
562 10.1016/j.aca.2012.12.029.
- 563 Pawlowsky-Glahn, V., and J. Egozcue. 2002. BLU estimators and compositional data. *Math.*
564 *Geol.* 34:259–274. doi: 10.1023/A:1014890722372.
- 565 Pawlowsky-Glahn, V., J. J. Egozcue, and R. Tolosana-Delgado. 2015. Modelling and
566 analysis of compositional data. John Wiley & Sons, Ltd.
- 567 Pinheiro, J. C., and D. M. Bates. 2000. Mixed effects models in S and S-Plus. Springer-
568 Verlag.
- 569 Poulsen, M., C. Schwab, B. B. Jensen, R. M. Engberg, A. Spang, N. Canibe, O. Hojberg, G.

- 570 Milinovich, L. Fragner, C. Schleper, W. Weckwerth, P. Lund, A. Schramm, and T.
571 Urich. 2013. Methylophilic methanogenic Thermoplasmata implicated in reduced
572 methane emissions from bovine rumen. *Nature Commun.* 4:1428. doi:
573 10.1038/ncomms2432.
- 574 Ramin, M. and P. Huhtanen. 2013. Development of equations for predicting methane
575 emissions from ruminants. *J. Dairy Sci.* 96:2476–2493. doi: 10.3168/jds.2012-6095.
- 576 R Core Team. 2016. R: A Language and Environment for Statistical Computing. R
577 Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- 578 Ricci, P., J. A. Rooke, I. Nevison, and A. Waterhouse. 2013. Methane emissions from beef
579 and dairy cattle: Quantifying the effect of physiological stage and diet characteristics.
580 *J. Anim Sci.* 91:5379–5389. doi: 10.2527/jas.2013-6544.
- 581 Robinson, D. L., J. Goopy, and R. S. Hegarty 2010. Can rumen methane production be
582 predicted from volatile fatty acid concentrations? *Anim. Prod. Sci.* 50: 630–636. doi:
583 10.1071/AN09214.
- 584 Rooke, J. A., R. J. Wallace, C. A. Duthie, N. McKain, S. M. de Souza, J. J. Hyslop, D. W.
585 Ross, T. Waterhouse, and R. Roehe. 2014. Hydrogen and methane emissions from
586 beef cattle and their rumen microbial community vary with diet, time after feeding
587 and genotype. *Br. J. Nutr.* 112:398–407. doi: 10.1017/S0007114514000932.
- 588 Rooke, J. A., S. M. Troy, C.-A. Duthie, J. J. Hyslop, D. W. Ross, T. Waterhouse, R. Roehe,
589 and R. J. Wallace 2016. Long-term changes in rumen volatile fatty acids (VFA) and
590 archaea to bacteria (AB) ratios in cattle fed either high concentrate or mixed
591 forage:concentrate diets with or without the addition of nitrate (NO₃). *Adv. Anim.*
592 *Biosci.* 7:53.
- 593 Ros-Freixedes, R., and J. Estany. 2014. On the compositional analysis of fatty acids in pork.
594 *J. Agric. Biol. Environ. Stat.* 19:136–155. Doi: 10.1007/s13253-013-0162-x.

- 595 Rymer, C., and Agnew, R. E. 2004. Summary of characterisation methods. In: C. Thomas,
596 editor, *Feed into Milk*. Nottingham University Press, Nottingham, UK.
- 597 Sauvant, D., S. Giger-Reverdin, A. Serment and L. Broudicou. 2011. Influences of diet and
598 rumen fermentation on methane production by ruminants. *INRA Prod. Anim.* 24:433–
599 446.
- 600 St-Pierre, N. R. 2001. Integrating quantitative findings from multiple studies using mixed
601 model methodology. *J. Dairy Sci.* 84:741–755. doi: 10.3168/jds.S0022-
602 0302(01)74530-4.
- 603 Stewart, C., S. Iverson, and C. Field. 2014. Testing for a change in diet using fatty acid
604 signatures. *Environ. Ecol. Stat.* 21:775–792. doi: 10.1007/s10651-014-0280-9.
- 605 Storlien, T. M., H. Volden, T. Almoy, K. A. Beauchemin, T. A. McAllister, and O. M.
606 Harstad. 2014. Prediction of enteric methane production from dairy cows. *Acta Agric.*
607 *Scand. Sect. A-Anim. Sci.* 64:98–109. doi: 10.1080/09064702.2014.959553.
- 608 Troy, S. M., C. A. Duthie, J. J. Hyslop, R. Roehe, D. W. Ross, R. J. Wallace, A. Waterhouse,
609 and J. A. Rooke. 2015. Effectiveness of nitrate addition and increased oil content as
610 methane mitigation strategies for beef cattle fed two contrasting basal diets. *J. Anim.*
611 *Sci.* 93:1815–1823. doi: 10.2527/jas.2014-8688.
- 612 Troy, S. M., C. A. Duthie, J. J. Hyslop, D. W. Ross, R. Roehe, and J. A. Rooke. 2016. The
613 effects of dietary nitrate addition and increased lipid concentration on methane (CH₄)
614 and hydrogen (H₂) emissions from beef cattle are independent. *Adv. Anim. Biosci.* 7:
615 50.
- 616 Wallace, R. J., J. A. Rooke, N. McKain, C. A. Duthie, J. J. Hyslop, D. W. Ross, A.
617 Waterhouse, M. Watson, and R. Roehe. 2015. The rumen microbial metagenome
618 associated with high methane production in cattle. *BMC Genomics* 16. doi:
619 10.1186/s12864-015-2032-0.

- 620 Wang, C., Q. Liu, Y. L. Zhang, C. X. Pei, S. L. Zhang, Y. X. Wang, W. Z. Yang, Y. S. Bai,
621 Z. G. Shi, and X. N. Liu. 2015. Effects of isobutyrate supplementation on ruminal
622 microflora, rumen enzyme activities and methane emissions in Simmental steers. *J.*
623 *Anim. Physiol. Anim. Nutr.* 99:123–131. doi: 10.1111/jpn.12191.
- 624 Ward, J. H. 1963. Hierarchical grouping to optimize an objective function. *J. Amer. Statist.*
625 *Assoc.* 58:236–244. doi: 10.1080/01621459.1963.10500845.
- 626 Weimer, P. J., D. M. Stevenson, H. C. Mantovani, and S. L. C. Man. 2010. Host specificity
627 of the ruminal bacterial community in the dairy cow following near-total exchange of
628 ruminal contents. *J. Dairy Sci.* 93:5902–5912. doi: 10.3168/jds.2010-3500.
- 629 Wolin, M. J. 1960. A theoretical rumen fermentation balance. *J. Dairy Sci.* 43:1452–1459.
630 doi: 10.3168/jds.S0022-0302(60)90348-9.
- 631 Yanez-Ruiz, D. R., L. Abecia, and C. J. Newbold. 2015. Manipulating rumen microbiome
632 and fermentation through interventions during early life: a review. *Frontiers*
633 *Microbiol.* 6:1133. doi: 10.3389/fmicb.2015.01133.
- 634

635 APPENDIX 1: SUMMARY STATISTICS OF THE VFA MOLAR PROPORTIONS AND
 636 THE COMPOSITIONAL BALANCE BASED ON THE (ACETATE · BUTYRATE) /
 637 PROPIONATE RATIO (b_2 IN EQ. [4], MODELS A2 AND B) PER STUDY AND DIET

Study ^{1,3}	Diet		Acetate	Propionate	Butyrate	Iso- butyrate	Iso- valerate	Valerate	b_2
Concentrate diet-type (forage less than 100 g/kg DM)									
1	1	Minimum	477	163	64	8	6	12	-0.71
		Q1 ²	554	240	77	10	10	14	-0.38
		Median	560	310	96	11	13	15	-0.24
		Q3 ²	566	329	122	14	23	17	0.10
		Maximum	645	417	155	17	33	27	0.54
2	3	Minimum	482	152	45	9	4	12	-0.78
		Q1	544	235	80	11	10	14	-0.36
		Median	562	284	95	12	16	16	-0.13
		Q3	590	318	131	16	28	20	0.03
		Maximum	668	411	221	20	72	31	0.47
4	7	Minimum	401	183	46	7	7	8	-0.93
		Q1	500	264	65	8	10	17	-0.57
		Median	535	335	80	14	15	18	-0.46
		Q3	561	402	123	17	21	22	-0.03
		Maximum	635	453	132	21	23	30	0.37
4	8	Minimum	527	126	59	9	10	6	-0.55
		Q1	561	181	66	11	12	10	-0.43
		Median	596	274	83	12	16	16	-0.21
		Q3	629	327	123	13	20	21	0.42
		Maximum	694	359	192	29	28	23	0.84
4	9	Minimum	495	177	50	10	10	10	-0.76
		Q1	527	239	65	15	16	16	-0.62
		Median	539	312	83	23	23	18	-0.30

		Q3	586	364	106	23	26	25	0.05
		Maximum	627	415	176	26	42	28	0.45
Mixed diet-type (400 – 600 g forage / kg DM)									
1	2	Minimum	541	105	107	10	11	10	0.16
		Q1	650	128	111	11	13	12	0.28
		Median	665	170	121	12	14	13	0.46
		Q3	705	190	144	12	15	13	0.64
		Maximum	725	229	152	32	36	17	0.92
2	4	Minimum	596	90	76	11	12	10	-0.18
		Q1	634	158	103	13	16	13	0.17
		Median	648	174	125	15	18	14	0.41
		Q3	665	207	141	16	21	15	0.54
		Maximum	707	269	177	21	27	17	1.07
4	10	Minimum	548	137	75	9	11	8	-0.42
		Q1	614	170	85	13	12	12	0.03
		Median	653	196	110	14	16	16	0.20
		Q3	661	235	119	16	17	20	0.46
		Maximum	686	341	161	17	19	26	0.60
4	11	Minimum	624	118	82	10	6	11	0.03
		Q1	659	144	116	12	8	12	0.33
		Median	668	164	119	13	12	16	0.45
		Q3	682	182	140	14	14	18	0.65
		Maximum	707	229	164	16	15	22	0.79
4	12	Minimum	603	136	84	10	12	6	-0.01
		Q1	623	160	99	12	14	13	0.08
		Median	635	206	107	14	16	16	0.23
		Q3	670	231	113	16	18	18	0.40
		Maximum	717	248	186	19	32	23	0.70
6	14	Minimum	597	120	82	4	8	8	-0.06
		Q1	645	146	115	8	13	11	0.26

		Median	665	180	123	10	15	12	0.38
		Q3	683	200	136	11	16	13	0.63
		Maximum	723	244	154	15	20	24	0.75
6	15	Minimum	650	118	104	7	4	7	0.34
		Q1	664	134	113	8	11	12	0.42
		Median	690	154	122	10	14	12	0.50
		Q3	697	170	138	10	17	14	0.68
		Maximum	720	189	163	11	23	17	0.84
6	16	Minimum	623	140	97	6	10	9	0.23
		Q1	662	150	107	7	12	10	0.38
		Median	678	158	124	9	14	11	0.48
		Q3	695	174	141	10	17	14	0.55
		Maximum	715	197	161	11	21	21	0.65
6	17	Minimum	675	121	93	6	8	1	0.36
		Q1	684	144	113	8	13	9	0.48
		Median	690	147	123	8	13	11	0.56
		Q3	708	154	134	9	15	12	0.59
		Maximum	736	177	143	12	17	14	0.79
Forage diet-type (>700 g forage / kg DM)									
3	5	Minimum	726	106	53	4	5	4	0.04
		Q1	751	129	60	6	6	5	0.23
		Median	767	144	65	6	7	6	0.35
		Q3	792	165	70	8	8	6	0.46
		Maximum	804	189	78	11	14	8	0.69
3	6	Minimum	694	125	63	2	5	4	0.19
		Q1	718	155	69	6	7	5	0.29
		Median	734	163	82	7	8	6	0.31
		Q3	748	170	86	8	9	7	0.37
		Maximum	790	196	99	10	11	11	0.51
5	13	Minimum	662	114	72	0	9	6	0.13

Q1	669	186	84	10	10	10	0.18
Median	676	202	91	10	11	10	0.19
Q3	698	203	98	12	12	11	0.22
Maximum	750	209	110	17	15	12	0.72

¹1, Rooke et al., 2014; 2, Duthie et al., 2017; 3, Duthie et al., 2015; 4, Troy et al., 2015; 5, unpublished observations; 6, Troy et al., 2016.

²Q1, Q3; respectively first and third quartiles.

³To obtain individual data please contact richard.dewhurst@sruc.ac.uk.

638 APPENDIX 2: ESTIMATES FOR THE CONCENTRATE AND MIXED DIET TYPES OF
 639 THE SIMPLIFIED MIXED MODEL FOR METHANE EMISSIONS (G/KG DMI IN LOG
 640 SCALE) USING THE COMPOSITIONAL BALANCE BASED ON THE (ACETATE ·
 641 BUTYRATE) / PROPIONATE RATIO (b_2 IN EQ. [4], MODELS A2 AND B) AND THE
 642 STATISTICALLY SIGNIFICANT COVARIATES FROM THE OVERALL MODEL B

Concentrate diet type				
Fixed effects	$\hat{\beta}$	SE	<i>t</i> -value	<i>P</i> -value
Intercept	0.744	2.106	0.35	0.725
b_2	0.297	0.049	6.03	< 0.001
Ln ME, MJ/kg DM	1.026	0.834	1.23	0.222
Ln DMI, g/kg BW ^{0.75} /d	-0.279	0.098	-2.85	0.006
$R_m^2 = 32.33\%^1, R_c^2 = 38.69\%^1, \hat{\sigma}_f = 0.06^2, \hat{\sigma} = 0.17^2$				

643

Mixed diet type				
Fixed effects	$\hat{\beta}$	SE	<i>t</i> -value	<i>P</i> -value
Intercept	-0.805	1.520	-0.53	0.597
b_2	0.206	0.050	4.10	<0.001
Ln ME, MJ/kg DM	1.806	0.614	2.94	0.004
Ln DMI, g/kg BW ^{0.75} /d	-0.289	0.060	-4.82	< 0.001
$R_m^2 = 21.46\%, R_c^2 = 41.53\%, \hat{\sigma}_f = 0.07, \hat{\sigma} = 0.13$				

¹ R_m^2, R_c^2 ; respectively % variance explained by fixed terms (marginal) and by both fixed and random terms (conditional).

² $\hat{\sigma}_f, \hat{\sigma}$; respectively estimated standard deviations of the random effects and random error terms.

Table 1. Sources of data and nutritional characteristics of the diets used in the study

Study ¹	Diet	Forage, g/kg DM	Starch, g/kg DM	NDF, g/kg DM	ME ² , MJ/kg DM
Concentrate diet type (forage less than 100 g/kg DM)					
1	1	80	412	254	12.3
2	3	79	415	248	12.8
4	7	84	439	227	12.2
4	8	80	476	204	12.0
4	9	78	416	211	12.9
Mixed diet type (400 – 600 g forage / kg DM)					
1	2	484	234	388	11.9
2	4	505	284	374	12.0
4	10	490	298	289	11.6
4	11	499	318	272	11.4
4	12	497	262	280	12.2
6	14	557	281	308	11.6
6	15	558	308	295	11.4
6	16	555	264	317	11.9
6	17	556	247	313	11.6
Forage diet type (>700 g forage / kg DM)					
3	5	774	65	771	7.4
3	6	1000	0	693	8.1
5	13	1000	36	473	10.7

¹1, Rooke et al., 2014; 2, Duthie et al., 2017; 3, Duthie et al., 2015; 4, Troy et al., 2015; 5, unpublished observations; 6, Troy et al., 2016a.

²ME estimated from feed composition (Rymer and Agnew, 2004).

1 Table 2. Methane emissions, DMI, and BW of cattle included in database

2

Study ¹	Diet	<i>n</i>	BW, kg		DMI, kg/d		Methane, g/d		Methane, g/kg DMI	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD
Concentrate diet type (forage less than 100 g/kg DM)										
1	1	15	667	45.1	11.2	2.12	153	44.6	13.9	4.07
2	3	35	635	55.4	10.6	1.74	147	31.2	13.9	2.32
4	7	12	675	41.0	10.0	1.55	148	48.1	14.6	3.31
4	8	13	675	54.3	9.1	2.09	136	27.4	15.3	3.39
4	9	13	687	55.7	9.7	2.04	149	32.4	15.8	3.58
Mixed diet type (400 – 600 g forage / kg DM)										
1	2	13	652	37.4	9.4	1.09	218	46.6	23.4	4.99
2	4	33	605	54.5	9.2	1.59	189	40.2	20.5	3.05
4	10	12	703	40.6	9.8	1.55	235	36.8	24.3	5.06
4	11	12	707	30.2	10.4	1.56	212	25.4	20.6	2.60
4	12	12	705	25.7	10.5	1.53	242	35.2	23.2	2.10
6	14	17	673	26.1	10.4	1.80	245	46.9	23.8	3.58
6	15	16	649	52.5	9.8	2.18	214	42.8	22.1	2.68
6	16	18	652	24.1	10.2	1.41	238	38.6	23.4	2.79
6	17	18	652	33.9	10.2	1.86	210	27.5	20.9	2.48
Forage diet type (>700 g forage / kg DM)										
3	5	16	639	78.5	9.4	1.84	126	24.0	13.8	3.36
3	6	17	607	97.1	10.1	1.65	160	33.3	16.5	5.55
5	13	12	704	58.3	13.7	1.77	308	36.9	22.6	2.12

3 ¹1, Rooke et al., 2014; 2, Duthie et al., 2017; 3, Duthie et al., 2015; 4, Troy et al., 2015; 5, unpublished observations; 6, Troy et al., 2016.

4

5 Table 3. Compositional geometric mean (CGM), overall and by diet group, and relative
6 variation matrix for the VFA composition across experiments

		Acetate	Propionate	Butyrate	Iso- butyrate	Iso- valerate	Valerate
	Overall	656	199	106	11	14	13
CGM ¹ , mmol/mol	Concentrate	574	282	96	14	17	17
	Mixed	670	170	121	11	15	13
	Forage	739	162	77	8	8	7
Variation matrix ²	Acetate		0.19	0.10	0.18	0.25	0.28
	Propionate	0.19		0.29	0.18	0.29	0.16
	Butyrate	0.10	0.29		0.18	0.21	0.23
	Iso-butyrate	0.18	0.18	0.18		0.13	0.15
	Iso-valerate	0.25	0.29	0.21	0.13		0.22
	Valerate	0.28	0.16	0.23	0.15	0.22	

7 ¹Normalized vector of geometric means of the VFA composition.

8 ²Matrix of log-ratio variances between pairs of VFA components.

9

10 Table 4. Sequential binary partitions (SBP) providing alternative balance coordinate representations ($b_i, i = 1, \dots, 5$) of the VFA composition and estimates of
 11 the associated compositional mixed model for methane emissions (g/kg DMI) in log scale, Models A1 (top) and A2 (bottom)

	SBP ¹	b_1	b_2	b_3	b_4	b_5	Fixed effects	$\hat{\beta}$	SE	t -value	P -value	
Model A1	Iso-valerate	+	+		+		Intercept	-2.134	1.041	-2.05	0.041	
	Iso-butyrate	+	+		-		b_1	0.163	0.035	4.66	<0.001	
	Butyrate	+	-			+	b_2	-0.122	0.043	-2.82	0.005	
	Acetate	+	-			-	b_3	0.158	0.047	3.36	0.001	
	Valerate	-		+			b_4	0.043	0.044	0.96	0.338	
	Propionate	-		-			b_5	-0.017	0.078	-0.21	0.832	
	r		4	2	1	1	1	Mixed diet	0.374	0.034	11.13	<0.001
	s		2	2	1	1	1	Forage diet	0.732	0.101	7.23	0.002
								Ln ME, MJ/kg DM	1.753	0.250	7.01	<0.001
								Ln BW, kg	0.195	0.128	1.52	0.129
							Ln DMI, g/kg BW ^{0.75} /d	-0.400	0.059	-6.84	0.000	
	$R_m^2 = 64.65\%^2$	$AIC = -207.51^3$										
	$R_c^2 = 68.60\%^2$	$\hat{\sigma}_f = 0.06^4$	$\hat{\sigma} = 0.16^4$									
Model A2	Acetate	+	+	+			Intercept	-2.134	1.041	-2.05	0.041	
	Butyrate	+	+	-			b_1	0.008	0.039	0.22	0.829	
	Propionate	+	-				b_2	0.256	0.050	5.11	<0.001	
	Valerate	-			+		b_3	0.017	0.078	0.21	0.832	
	Iso-butyrate	-			-	+	b_4	0.026	0.036	0.72	0.470	
	Iso-valerate	-			-	-	b_5	-0.043	0.045	-0.96	0.338	
	r		3	2	1	1	1	Mixed diet	0.374	0.034	11.13	<0.001
	s		3	1	1	2	1	Forage diet	0.732	0.101	7.23	0.002
								Ln ME, MJ/kg DM	1.753	0.250	7.01	<0.001
								Ln BW, kg	0.195	0.128	1.52	0.129
							Ln DMI, g/kg BW ^{0.75} /d	-0.400	0.059	-6.84	<0.001	
	$R_m^2 = 64.65\%$	$AIC = -207.51$										
	$R_c^2 = 68.60\%$	$\hat{\sigma}_f = 0.06$	$\hat{\sigma} = 0.16$									

¹ Symbol + means that a VFA component is allocated to the numerator of the corresponding compositional balance b_i , whereas - means that it is allocated to the denominator. Letters r and s refer to the number of VFA components in numerator and denominator respectively.

² R_m^2 , R_c^2 ; respectively % variance explained by fixed terms (marginal) and by both fixed and random terms (conditional).

³ Akaike information criterion measure of the relative quality of the model for the full data set.

⁴ $\hat{\sigma}_f$, $\hat{\sigma}$; respectively estimated standard deviations of the random effects and random error terms.

- 12 Table 5. Estimates of the simplified mixed model (Model B) for methane emissions (g/kg
 13 DMI in log scale) using the compositional balance based on the (acetate · butyrate) /
 14 propionate ratio (b_2 in Eq. [4], Model A2) and the statistically significant covariates

Fixed effects	$\hat{\beta}$	SE	t -value	P -value
Intercept	-0.837	0.613	-1.36	0.174
b_2	0.243	0.035	6.97	< 0.001
Mixed diet	0.374	0.033	11.47	< 0.001
Forage diet	0.720	0.105	6.83	0.002
Ln ME, MJ/kg DM	1.725	0.246	6.99	< 0.001
Ln DMI, g/kg BW ^{0.75} /d	-0.356	0.051	-6.96	< 0.001

$$R_m^2 = 63.51\%{}^1 \quad \text{AIC} = -211.15{}^2$$

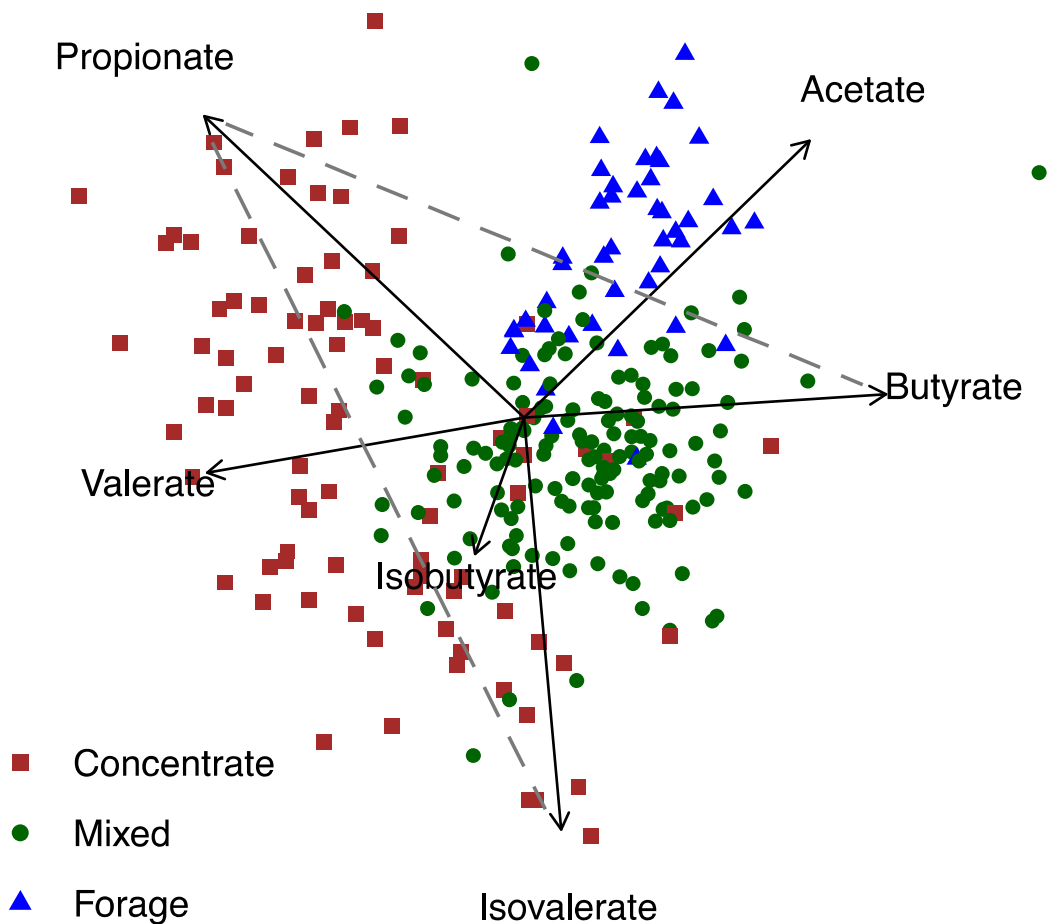
$$R_c^2 = 68.72\%{}^1 \quad \hat{\sigma}_f = 0.06{}^3 \quad \hat{\sigma} = 0.16{}^3$$

¹ R_m^2 , R_c^2 ; respectively % variance explained by fixed terms (marginal) and by both fixed and random terms (conditional).

²Akaike information criterion measure of the relative quality of the model for the full data set.

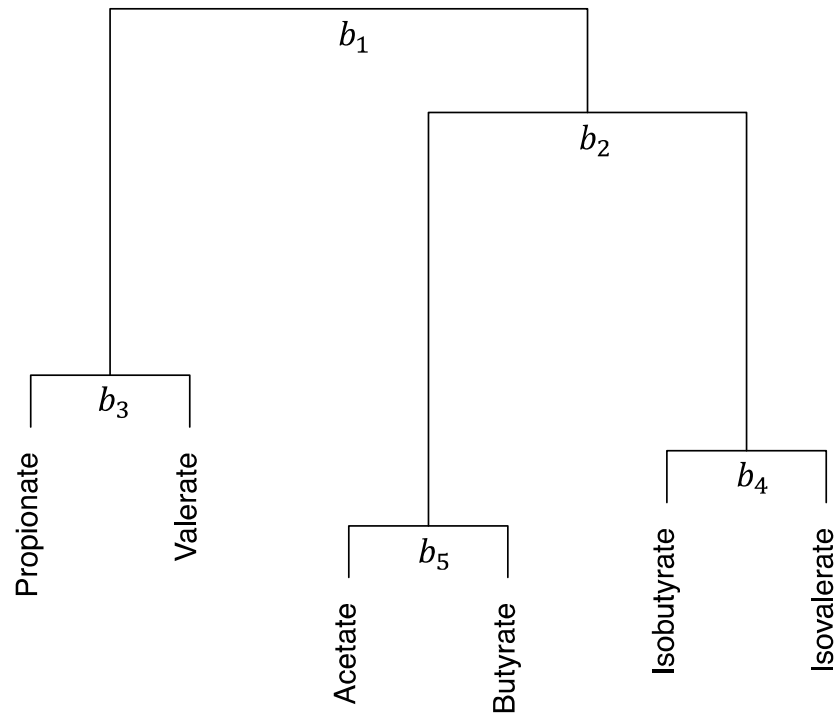
³ $\hat{\sigma}_f$, $\hat{\sigma}$; respectively estimated standard deviations of the random effects and random error terms.

15 Figure 1. Compositional biplot of the VFA (mmol/mol) data set with the components
 16 represented by rays and the collected samples represented by points according to concentrate,
 17 mixed, and forage diet types. The links (dashed lines) between arrowheads are proportional to
 18 the log-ratio variances between the corresponding VFA components.



19
 20
 21

- 22 Figure 2. Groupings of VFA (mmol/mol) components according to proportionality
23 relationships from the variation matrix and associated balances ($b_i, i = 1, \dots, 5$) for Model A1.

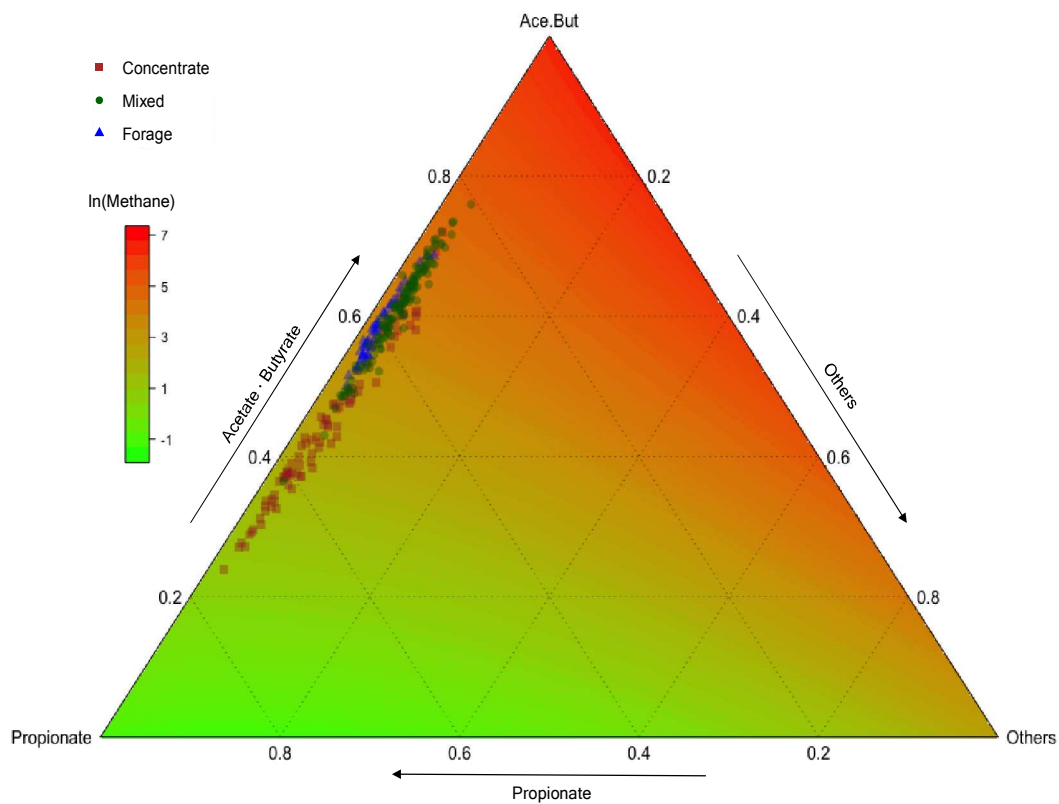


24

25

26 Figure 3. Ternary diagram of the [acetate · butyrate (Ace.But), propionate, others] VFA
 27 subcomposition for concentrate, mixed, and forage diet types and expected methane
 28 emissions (g/kg DMI in log scale) from Model B.

29



30

31

Propionate

Acetate

Butyrate

Valerate

Isobutyrate

Isovalerate

■ Concentrate

● Mixed

▲ Forage

