Is It Possible to Estimate the Composition of a Cow’s Diet Based on the Parameters of Its Milk?

Ana Villar 1,*, Gregorio Salcedo 2, Iban Vázquez-González 3, Elena Suárez 4 and Juan Busqué 1

Abstract: Understanding the composition of a cow’s diet through the analysis of its milk is very useful in the linking of the product consumed with the systems involved in its production. The aim of this study is to show the diet–milk composition relationship using correspondence analysis and multiple linear regression analysis. This study analyzed 174 tank milk samples taken from 89 commercial farms located in “Green Spain”. Sampling was performed in two different periods: autumn 2016 and spring 2017. The correspondence analysis allowed for study into the general relationships between diet components and their relationship with the composition of milk (chemical composition, fatty acid profile, and fat-soluble antioxidants (FSA)). The model used to estimate the percentage of fresh grass (FG) in the diet had a high predictive power (Radj2 > 0.7), and the explanatory variables included in the model were linolenic acid (C18:3-n3), vaccenic acid (trans11-C18:1), and cis12-C18:1. The regression equation was applied to the 174 tank milk samples individually. To evaluate the equation’s predictive capacity, different thresholds for the dry matter percentage of fresh grass in the ration were marked (15%, 20%, 25%, and 30%), above which milk could be considered “grass-fed milk”, and below which, “not grass-fed milk”. The equation is considered valid when it correctly classifies the sample. The highest percentage of success (89.7%) was obtained by marking a threshold of 25% FG. When analyzing the misclassified milk samples, that is, where the equation did not classify the milk sample well according to its fresh grass composition, it was observed that the majority of cases corresponded to milk samples that came from herds fed with fresh grass above the marked threshold (>25%) but with a high content of concentrate in the ration. The conclusion is that the percentage of concentrate in the diet has a very important influence on the fatty acid profile of milk, particularly with respect to fresh grass. This is in such a way that anywhere above a concentrate content of >30%, the equation’s capacity to estimate the percentage of fresh grass decreases.

Keywords: grass-fed milk; diet authentication; fatty acid profile; carotenoids; multiple linear regression analysis

1. Introduction

In recent years, milk products labeled as “grazing milk” or “grass-fed milk” have been appearing on the market. These products, certified or not, respond to a growing consumer interest in foods that they understand as “natural”, “healthy”, and “respectful of animal welfare and/or environment”. However, the consumer is not aware of the nutritional benefits that come with including grass in the diet of lactating cows. Such benefits are the reduction in the total saturated fatty acid content [1,2], and the increase in the content of polyunsaturated fatty acids, especially vaccenic acid, omega-3 isomers (particularly linolenic acid), and conjugated linoleic acid (CLA) in the milk. These fatty acids are considered to be of great nutritional interest [1,3]. The fatty acid profile of grazing
milk is not always associated with a healthier profile, since it depends on the consumption of fresh herbage in the general diet and not on the fact that the cows are grazing. Thus, Capuano et al. [4], comparing samples of commercial pasteurized milk from the brand *weidemelk* with organically produced milk and with conventional milk, found no difference in the fatty acid (FA) profile between the *weidemelk* milk and conventional milk, both in winter and summer. Therefore, only when grazing cows consume a lot of fresh grass (and conserved forage) is the milk a product with a more heart-healthy fat profile than milk from cows without fresh grass. Likewise, the presence of fresh grass in the diet of lactating cows results in a greater presence of fat-soluble antioxidants, some of which are also of great nutritional interest, such as vitamins A and E and carotenoids [3,5].

If we consider the certification of “grazing milk” to be linked with the allowing of lactating cows to pasture for a number of hours and days per year, what could be referred to as “grass-fed milk” would be milk from cows that consume a certain amount of fresh grass in their diet. This amount should be specified based on the profile of fatty acids, regardless of whether the forage is consumed in grazing or indoors.

In fact, Capuano et al. [6], found that, when only analyzing the fatty acid profile, it is not easy to distinguish milk from grazing cows from milk obtained from cows fed fresh grass indoors.

Some studies have managed to estimate the content of the different fatty acids of interest in milk by using components from the diet [7]. They have also been able to estimate the presence of each component in the diet from the profile of fatty acids [8]. However, the application of these algorithms to tank milk samples, under local conditions, results in a low concordance (unpublished data).

In view of the sudden increase in products labeled as “grass-fed milk”, the objective of this study is to develop a statistical model to estimate the percentage of fresh grass in the diet of lactating cows by using the analysis of FA and fat-soluble antioxidants (FSA) in the tank milk of commercial farms. The final objective is the authentication of milk from cows with a diet consisting of fresh grass, regardless of whether this FG has been consumed in grazing or not. The purpose of this is to increase its added value.

2. Materials and Methods

2.1. Farms

The objective of the study was to include a selection of farms that were representative of the entire spectrum of diets in the north of Spain (Green Spain). However, fundamentally, the study intends to include the greatest possible diversity between diets and does not intend to proportionally be representative of the existing typologies. To this end, 89 commercial farms located in Galicia (25), Asturias (22), Cantabria (25), and Navarra (17) were selected.

2.2. Sample Design

174 bulk tank milk samples coming from two or four milkings were collected from these 89 commercial farms in two periods: autumn 2016 and spring 2017. For each sample, the fatty acid profile (49 FA plus its main relationships) and the content of nine fat-soluble antioxidants (retinol (vitamin-A), all-trans-β-carotene, 9-cis-β-carotene, 13-cis-β-carotene, lutein, zeaxanthin, β-cryptoxanthin, α-tocopherol, and γ-tocopherol (vitamin E)) were determined. For each milk sample, two estimations were made: the mean milk production per cow and the mean composition of the rations fed to lactating cows during the preceding days. Diet composition was expressed in terms of the percentage of dry matter intake (DMI) of each component of the ration in the total DMI per cow, that is: DM of fresh grass (FG); consumed either indoors or in pasture), DM of grass silage (GS), DM of corn silage (CS), DM of dry forage (DF; the sum of hay, straw, alfalfa, etc.) and DM of concentrate (CC).

The following equation estimates the cows’ voluntary intake:

\[
DMI = 12 + 0.372 \times MP_{4FCM}
\]
DMI (Dry Matter Intake) is measured in kg DM day$^{-1}$ per cow and MP$_{4FCM}$ is the 4% fat-corrected milk production (MP$_{4FCM} = (MP \times 0.4) + (MP \times \text{fat}\% \times 0.15)$). The DMI equation is taken from the National Research Council [9] and is used for Holstein lactating cows with 620 kg of live weight and which are in the middle of their lactating period. The daily intake of fresh grass (kg DM) was estimated by subtracting the sum of the dry matter intake from the other ingredients present with respect to the theoretic daily total consumption of every cow.

Each sampling day, 4 bulk tank milk subsamples were taken. One of the samples was delivered to the corresponding dairy interprofessional laboratory (one for each Autonomous Community), in order to perform the physicochemical analysis. The remaining samples were frozen ($< -20^\circ$C) for their later analysis. In particular, two of them were sent, preserving their frozen state, to the Dairy Interprofessional Laboratory of Galicia (LIGAL) to determine their fatty acids (FA) profile and fat-soluble antioxidants (FSA). The last sample was stored as a possible backup.

2.3. Analytic Measurements

2.3.1. Fatty Acids Analysis

Milk fat extraction and fatty acid methyl esters (FAME) preparation (following a basic methylation using potassium hydroxide solution in methanol, 0.5N were according to ISO 14156/IDF172 (2001) [10] and ISO 15884/IDF182 (2002) [11], respectively.

The FAME were analyzed by gas chromatography (GC) using a Varian 3900 GC (Varian, Instruments, Walnut Creek, CA, USA) equipped with a flame ionization detector (FID), a BPX70 capillary column (120 m $\times$ 0.25 mm i.d., 0.25 $\mu$m film thickness; SGE, Trajan Scientific and medical, Melbourne, Australia) and a Varian 8410 autosampler. The injector and detector were kept at a constant temperature of 250 $^\circ$C. The column oven program was according to the methodology described by Kramer et al. [12] (45 $^\circ$C for 4 min, then increased at 13 $^\circ$C/min to 175 $^\circ$C, kept for 27 min, increased at 4 $^\circ$C/min to 215 $^\circ$C and kept for 35 min; with a total run time of 85.62 min). The carrier gas used was Helio at a constant flow of 1.3 mL/min. The split ratio was 20:1 and 1 $\mu$L of the sample was injected.

With this oven temperature program, some trans-18:1 (t-18:1) isomers and CLA isomers were not well resolved. Further separation of trans-18:1 isomers would have required separation with silver-ion thin-layer chromatography or solid-phase extraction, followed by isothermal GC using a relatively low temperature [13]. In the same way, a complete CLA isomer profile would require a Ag+-HPLC analysis according to Cruz-Hernandez et al. [14]. These specific analyses were not carried out as with this column and in a simple run we were able to identify/quantify the main ones in milk fat, allowing separation and identification of the t10-18:1, t11-18:1 and t12-18:1 isomers, as well as the sum of cis-9, trans-11-octadecadienoate (c9,t11 CLA) + t7,c9 CLA + t8c10 CLA.

The FAME peaks were identified using the FAME mix C4-C24 and methyl trans-11-vaccenate standard acquired from Supelco Inc. (Bellefonte, PA, USA), and methyl c9, t11-CLA acquired from Matreya LLC (State College, PA, USA). The GLC-603 and GLC-463 standards from Nu-Check Prep, Inc. (Elyssian, MN, USA) were used to identify the FAME not contained in the FAME mix. Branched-chain FAME were identified using a GC reference standard BC-Mix1 purchased from Applied Science (State College, PA, USA). Additional FAME not present in the standards were identified based on their elution orders and GC/MS identification [14]. Quantification of FAME was based on peak area and calculation using three internal standards: methyl nonanoate (C9:0) and methyl cis-10-heptadecenoate (c10-C17:1) acquired from the Sigma-Aldrich Co. (St. Louis, MO, USA), and methyl trans-10, cis-12-octadecadienoate (t10, c12-CLA) acquired from Matreya LLC (State College, PA, USA) according to the wide range of concentration within the FA profile. Moreover, the methyl nonadecanoate (C19:0) (acquired from Sigma-Aldrich Co., St. Louis, MO, USA) was used as a surrogate.
2.3.2. Fat-Soluble Antioxidants Analysis

For milk antioxidant extraction, the frozen milk samples (−20 °C) were thawed in a refrigerator (4 °C) for 24 h prior to conducting the analysis. They were warmed up in a water bath (20 °C) (Memmert WB/OB 22, Memmert GmbH + Co. KG, Schwabach, Germany) on the day of the simultaneous extraction of carotenoids and vitamins (E and A), brought to 35 °C and then cooled again to 20 ± 2 °C. Subsequently, an aliquot of homogenized sample was taken and the extraction procedure was conducted according to the methodology proposed by Gentili et al. [15].

Final extracts from both matrixes were evaporated to dryness under N2 (Six Port Mini-Vap, Sigma-Aldrich, Merck Life Science S.L.U., Madrid, Spain) and reconstituted in acetonitrile, methanol with ammonium acetate (50 mM) and dichloromethane (75:10:15) before being filtered (GHP, Acrodisc® Syringe Filter, 13 mm, 0.2 µm) and injected on the High Performance Liquid Chromatography (HPLC, Alliance 2695, Waters Corporation, Milford, MA, USA). Then, analyses were carried out on the HPLC equipped with both a photodiode array (PDA) (2998, Waters Corporation, Milford, MA, USA) and a multiwavelength fluorescence (2475, Waters Corporation, Milford, MA, USA) detector. Simultaneous carotenoid and vitamin separation was performed on a RP dC18 Atlantis column (250 mm × 4.6 mm i.d. × 5 µm film thickness, 100 Å) (Waters, Milford, MA, USA). The mobile phase consisted on a quaternary gradient of acetonitrile, methanol with ammonium acetate (50 mM), water and dichloromethane D. Carotenoids were detected at 450 nm using the PDA, whereas detections of vitamin E and vitamin A were accomplished by fluorescence with excitation-emission at 295–330 nm and 330–480 nm, respectively. A mix of individual stock solutions was prepared on mobile phase and external standard calibration curves were built to quantify carotenoids and vitamins from each matrix according to the expected range. Concentrations were also adjusted by the recovery factors of the two surrogates, echinenone and δ-tocopherol.

2.4. Statistical Analysis

From the data available for each farm and the information collected on the day of sampling, and with the values obtained after determining the FA and FSA profiles, a large database was built with 174 data inputs (milk samples) and 172 variables (associated with the farm and sampling).

The statistical software package SPSS (version 17.0; SPSS Inc., Chicago, IL, USA) was used to analyze data.

Initially, an analysis of the data was carried out to study the variables and their relationships: descriptive statistical analysis (mean values and number of cases), correlations between variables, and analysis of variance (ANOVA); the post hoc comparisons among group means were defined by applying a Tukey-b test in case of equal variances and a Dunnett T3 test in case of different variances.

Multivariate analysis (correspondence analysis) was used with the main components of the diet. Its purpose was to value them against each other and with respect to the main orthogonal axes created. In the resulting ordination, the factors of origin of and time of sampling were also represented, as well as the set of quantitative variables related to milk (production, chemical composition, FA, and FSA).

Multiple linear regression was used to predict the percentage of fresh grass in the diet of the lactating cows (response variable (RV)) using the composition of milk and using the productive data associated with the milk sample (explanatory variables (EVs)). As EVs for the construction of the statistical model, variables relating to milk were used (average daily production per cow (prod/day), chemical composition (% fat, % protein, % MSNF (Milk Solids Not Fat), freezing point, and urea), somatic cell count (LgSCC), 52 fatty acids expressed in relative %, 13 groups or relationships between these fatty acids and 9 fat-soluble antioxidants). The response variable was transformed into the arcsine of the square root of its value divided by 100 to improve its distribution. Three stepwise methods (forward, backward and mixed) and a cross-validation with k = 10 were used.
to choose the model with fewer EVs and with a mean square error (MSE) below the minimum MSE threshold $+ 1 \times \text{s.e.}$ (one standard error rule [16]). In the resulting models, it was found that the variance inflation factors were always less than 10, thus avoiding the collinearity of the EVs.

2.5. Equation Validation Indices

For the validation of the fresh grass estimation equation for the lactating cow diet based on the FA and FSA profiles, different statistical indices were used.

The index of agreement ($d$) is an indicator of the relative size of the differences between model-predicted data and observed data [17]:

$$d = 1 - \left[ \sum_{i=1}^{n} (O_i - P_i)^2 / \sum \left( |O'_i| + |P'_i| \right)^2 \right]$$

where $O_i$ and $P_i$ are the observed and model-predicted values for observation $i$, and $|O'_i|$ and $|P'_i|$ are the differences in the mean value from those observed and those model-predicted, respectively, with $n$ being the total number of observations. A value of 1 indicates the maximum accuracy of the model and a value of 0 the minimum accuracy. The root-mean-square error (RMSE), which reflects the differences between the observed and predicted values, and the mean bias error (MBE) of the systematic deviations, were calculated using the equations proposed by Willmott [17].

The Nash–Sutcliffe efficiency index (EF) [18] is another parameter for evaluating the goodness of fit of a model. It can range from minus one to one. A negative value indicates that the mean of the observations is a better estimator than the model’s predictions, and the quality of the estimate increases as this indicator approaches one.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (O_i - P_i)^2}{n}}$$

$$\text{MBE} = \frac{\sum_{i=1}^{n} (P_i - O_i)}{n}$$

$$\text{EF} = 1 - \frac{\sum_{i=1}^{n} (P_i - O_i)^2}{\sum_{i=1}^{n} (O_i - \bar{O}_i)^2}$$

where RMSE is the root-mean-square error, MBE is the mean bias error, and EF is the efficiency index.

Another way to evaluate the predictive value of the equation is to mark a threshold, that is, to establish what percentage of the dry matter of the ration must be fresh grass in order for the sample to be considered within the definition of “grass-fed milk”. Based on the concordance between the sample classification using the equation and the data calculated from the information provided by the farmer about the diet supplied to lactating cows (reference standard), a contingency table was created where the samples were classified as true positive (TP), true negative (TN), false positive (FP), and false negative (FN), with the purpose of calculating the percentage of well-classified samples (“grass-fed milk”/not “grass-fed milk”).

From the classification of the samples as TP, TN, FP, and FN, the different parameters were calculated: sensitivity, specificity, and the positive predictive value and negative predictive value of the equation for each threshold, and in accordance with the following definitions and calculations (Table 1).
Table 1. Definition and calculation of each parameter.

| Indicator       | Definition                                                                 | Calculation     |
|-----------------|---------------------------------------------------------------------------|-----------------|
| Sensitivity     | Percentage of samples that can effectively be classified as “grass-fed milk”, or probability of detecting a “grass-fed milk” sample | TP/TP + FN      |
| Specificity     | Percentage of not “grass-fed milk” samples that are correctly classified, or the probability of detecting a not “grass-fed milk” sample | TN/TN + FP      |
| +Predictive Value | Probability that a sample will be classified as “grass-fed milk” and is in fact “grass-fed milk” | TP/TP + FP      |
| −Predictive Value | Probability that a sample will be classified as not “grass-fed milk” and is in fact not “grass-fed milk” | TN/TN + FN      |

3. Results

3.1. Initial Population

Of the 174 milk samples analyzed, 48.3% (84 samples) came from herds fed with fresh grass. This was either consumed in grazing (27.6%), supplied indoors (12.6%) or was consumed in both ways (8%). With the exception of 2 farms (4 samples), concentrate is included in the diet of all samples and can represent from 7% to 63% of the DM of the diet (from the milk samples coming from animals fed with concentrate). Grass silage and hay are part of the ration for lactating cows in about 70% of the herds (73% and 68.4% of the samples, respectively). On average, the majority component is concentrate, which constitutes 36.7% of the dry matter of the diet, followed by fresh grass, grass silage, and corn silage. The minority component in these northern Spanish farms is dry forage (Table 2).

Table 2. Means and deviations, maximum and minimum values of each component in the diet supplied to lactating cows.

| Composition of the Diet of Lactating Dairy Cows | N = 174 Samples | Farms (1) (2) (%) |
|-----------------------------------------------|----------------|------------------|
| Mean ± std                                    | Maximum       | Minimum          |                 |
| kg DM (2) day⁻¹ per cow                       | 21.4 ± 2.4    | 27.9             | 14.6            | —               |
| Fresh grass                                  | 20.6 ± 27.4   | 100.00           | 0.00            | 47.1            |
| Corn silage                                  | 15.5 ± 15.9   | 53.1             | 0.00            | 54.6            |
| Grass silage                                 | 17.2 ± 14.8   | 59.8             | 0.00            | 72.4            |
| Dry forage                                   | 10.0 ± 10.8   | 53.0             | 0.00            | 67.2            |
| Concentrate                                  | 36.7 ± 13.0   | 63.3             | 0.00            | 97.7            |

(1) 89 livestock farms included in the study. (2) DM = Dry Matter.

Considering the objectives of the project, great variability in the composition of the diet can be observed in the above table. This is in such a way that, for all the components, there are diets which do not include that specific component and diets where the component constitutes >50% of the dry matter of the ration. In the case of the fresh grass, it may be absent or be 100% of the ration.

3.2. Correspondence Analysis

The first three axes of the ordination accounted for 95% of the total variance of the lactating cows’ diets. The first axis (57% of the variance) was strongly correlated with fresh grass (FG), showing that this component explains most of the variability among the collected observations (this axis is related to the presence or absence of fresh grass in the diet). On the opposite side of this axis, the corn silage (CS) was located with the highest absolute value; however, it was not as significant as the FG. Spring (SPR in Figure 1) is well separated from autumn (AUT) in this axis, indicating a higher consumption of fresh grass in the first season.

The fatty acids in the box (Figure 1) had the highest values in the first axis, while the n6/n3 ratio and daily milk production (liters cow⁻¹ day⁻¹) had the most negative values. Spring values are on the positive side. The second axis (19% of variance) compares corn silage (CS) diets with conserved grass forage (DF and GS) diets. Regarding this axis, the highest positive values correspond to carotenes (9-cis-β-carotene, 13-cis-β-carotene,
and all-trans-β-carotene) and, to a lesser extent, to the fatty acids in the box; while the negative values, again, correspond to the n6/n3 ratio and daily milk production. Autumn shows positive values on this axis. The third axis (no figure) explains 18% of the total variance, and separates the diets with more grass silage (positive) from the diets with more dry forage (negative), the former corresponding to higher values of antioxidants in milk (all-β-carotene, β-cryptoxanthin, 9-cis-β-carotene, 13-cis-β-carotene) and the latter to higher values of urea, C17:0, and C15:0.

**Figure 1.** Locations in the axes. 1-Axis (57% variance) and 2-Axis (19%) of a correspondence analysis for the components in the diet of lactating cows (black and arrows). Incorporation in the figure of the quantitative (red) and factorial (blue) explanatory variables. The proximity in the location of variables indicates strong positive correlations. The red rectangle is amplified in the upper right. DM (dry matter), FG (% fresh grass DM, consumed from a trough or in grazing), GS (% grass silage DM), CS (% corn silage DM), DF (% dry forage DM; which includes hay, straw, alfalfa, etc.), CC (% control concentrate DM).
3.3. Multiple Linear Regression Analysis

Table 3 presents the model used to estimate the percentage of fresh grass in the diet. The predictive power of the equation is high ($R_{aj}^2 > 0.7$). In terms of the equation, it is notable that it does not include any fat-soluble antioxidants and, in fact, they did not appear in any of the models used.

Table 3. Model resulting from multiple stepwise linear regressions and with cross-validation for the percentage of fresh grass in the diet of lactating cows, transformed to the arcsine of the square root of the proportion (0% = 0 and 100% = 1.57), according to the profile of fatty acids and fat-soluble antioxidants in milk.

| Model | $R_{aj}^2$ | MSE$_{valid}$ | MSE$_{entr}$ | DF  |
|-------|------------|----------------|-------------|-----|
| RF$_{FG}$ = $-0.13 \pm 0.071 + 0.16 \pm 0.021 \times \text{trans11-C18:1} - 0.76 \pm 0.156 \times \text{cis12-C18:1} + 0.62 \pm 0.102 \times \text{C18:3-n3}$ | 0.72 | 0.052 | 0.047 | 150.4 |

Note: RF$_{FG}$ (result of the formula for fresh grass), t$_{11}$ C18:1 is vaccenic acid; c$_{12}$ C18:1 is an isomer of oleic acid and C18:3 n3 is linolenic acid. Since the dependent variables (% DM of each diet component relative to total DM) were transformed to the arcsine of the square root of their value divided by 100 to improve their distribution, the % FG in the diet is calculated as: 100 $\times$ (sin RF$_{FG}$)$^2$. MSE$_{valid}$: mean square error obtained in the stepwise regression and with cross-validation. MSE$_{entr}$: mean square error considering all observations and the regression formula. DF: Degrees of freedom.

The explanatory variables that appear in the model are linolenic acid (C18:3-n3), vaccenic acid (trans11-C18:1), and cis12-C18:1; fatty acids which are clearly related to the presence of fresh grass in the diet. In fact, linoleic and vaccenic acid have correlations higher than 0.7 with the content of fresh grass in the diet (Pearson $r = 0.757$, $p < 0.01$; Pearson $r = 0.765$, $p < 0.01$, respectively), and the c12-C18:1 has a correlation higher than 0.5 (Pearson $r = -0.547$, $p < 0.01$) [19].

3.4. Use of the Equation as a Diagnostic Test: Evaluation of the Algorithm According to Its Ability to Authenticate “Grass-Fed Milk”

The equation (Table 3) allows for the estimation of the mean percentage of fresh grass dry matter consumed by lactating cows in a dairy farm, by using a sample of tank milk from which we have the fatty acid profile. The distance between the value estimated with the formula and the percentage of FG in the diet, the figure obtained from the information provided by the farmer (hereinafter the reference standard), allows for the validation of the appropriateness of the algorithm.

When applying the equation to the 174 milk samples, it was observed that in 95% of the samples corresponding to diets without fresh grass (FG = 0) the values obtained by the formula were less than 10, that is, less than 10% of fresh grass; but in few cases zero. It was therefore not possible to define “grass-fed milk” only by establishing the presence or absence of fresh grass in the diet without taking into account the proportion of fresh grass in the dry matter. Thus, four levels of fresh grass were proposed: 15%; 20%; 25%, and 30%.

In accordance with the concordance between the value obtained in the equation and the reference standard, a contingency table was created (Table 4). From that table the percentage of well-classified samples (“grass-fed milk” / not “grass-fed milk”) and the validity indicators were calculated from it: sensitivity, specificity, positive predictive value, and negative predictive value, according to different thresholds ($<15\% / \geq 15\%$; $<20\% / \geq 20\%$; $<25\% / \geq 25\%$; $<30\% / \geq 30\%$).

The most effective thresholds correspond to 20% and 25% of FG in the diet since they present a greater number of well-classified samples (88.5% and 89.7%, respectively) (Table 5) and somewhat higher values for the validity indicators (sensitivity, specificity, positive predictive value, and negative predictive value) (Table 6).
Table 4. Contingency table. Classification of milk samples according to the concordance between the sample allocation to a definition of “grass-fed milk” through the equation and the reference standard (1).

| Result of the Formula | Reference Standard (1) | Reference Standard (2) |
|-----------------------|------------------------|------------------------|
| Positive: “grass-fed milk” | Yes (2)               | False Positives (FP)   |
| Negative: not “grass-fed milk” | False Negatives (FN)    | True Negatives (TN)     |

(1) Value obtained from the information provided by the farmer. (2) Concordance (Yes/No); defined according to the chosen threshold: \( \geq 15\% \text{ FG}/<15\% \text{ FG}; \geq 20\% \text{ FG}/<20\% \text{ FG}; \geq 25\% \text{ FG}/<25\% \text{ FG}; \geq 30\% \text{ FG}/<30\% \text{ FG}. 

Table 5. Validation index and percentage of well-classified milk samples with the algorithm based on the fresh grass threshold in the diet.

| Sample Classification | Threshold (Fresh Grass Percentage) above Which the Sample Is Considered “Grass-Fed Milk” | True positives | True negatives | False positives | False negatives | Well-classified |
|----------------------|----------------------------------------------------------------------------------------|----------------|----------------|-----------------|-----------------|-----------------|
|                      | \( \geq 15\%/<15\% \)                                                                      | 28.16\%        | 61.49\%        | 2.87\%          | 7.47\%          | 86.8\%          |
|                      | \( \geq 20\%/<20\% \)                                                                      | 29.31\%        | 59.20\%        | 2.30\%          | 9.20\%          | 88.5\%          |
|                      | \( \geq 25\%/<25\% \)                                                                      | 28.16\%        | 61.49\%        | 2.87\%          | 7.47\%          | 89.7\%          |
|                      | \( \geq 30\%/<30\% \)                                                                      | 20.69\%        | 64.37\%        | 4.60\%          | 10.34\%         | 85.1\%          |

Table 6. Sensitivity, specificity, positive predictive value, and negative predictive value of the equation according to the marked threshold.

| Parameter              | % FG Threshold | \( \geq 15\%/<15\% \) | \( \geq 20\%/<20\% \) | \( \geq 25\%/<25\% \) | \( \geq 30\%/<30\% \) |
|------------------------|----------------|------------------------|------------------------|------------------------|------------------------|
| Sensitivity            | 74\%           | 76\%                   | 79\%                   | 67\%                   |
| Specificity            | 96\%           | 96\%                   | 96\%                   | 93\%                   |
| +predictive value      | 93\%           | 93\%                   | 91\%                   | 82\%                   |
| –predictive value      | 84\%           | 87\%                   | 89\%                   | 86\%                   |

From Table 5, it can be deduced that between 88% and 90% of the samples would be well-classified (TP + TN) with thresholds of 20% and 25% of FG, respectively. However, there are between 10% and 12% of samples that are misclassified according to the equation (FP + FN).

From Table 6, it is deduced that for the positive samples (TP + FN), i.e., samples with \( \geq 20\% \) or \( \geq 25\% \) of FG respectively, the equation only identifies them as such in between 76% and 79% of the cases, meaning the sensitivity of the equation is low. However, the specificity of the algorithm is good, since 96% of the negative samples (TN + FP) are identified as not “grass-fed milk”.

3.5. Analysis of Misclassified Samples

In order to analyze the misclassified samples, the 174 samples were grouped according to the content of concentrate in the diet in accordance with the reference standard, and with 2 levels of percentage for the dry matter of concentrate in the diet (0-30%/\( \geq 30\% \)), or with 3 levels (0-30%/\( \geq 30-40\% /\geq 40\% \)) for the 2 selected thresholds of fresh grass (20%/25%). Table 7 presents the percentage of well-classified samples by strata; that is, the percentage for the samples whose estimated % FG—calculated with the formula—places them in the same stratum as the reference standard.
Table 7. Percentage of well-classified samples based on the fresh grass and concentrate content for the 2 selected thresholds (20%/25%) (1).

| Diet of Lactating Cows (% DM) (1) | % Well-Classified Samples (2) |
|----------------------------------|-------------------------------|
| Fresh Grass                      | Concentrate                   | By Dietary Stratum | Total |
| ≥20                              | 0–30                          | 92.3               | 88.5  |
| <20                              | 0–30                          | 96.3               |       |
|                                  | ≥30                           | 53.6               |       |
|                                  | 0–60                          | 96.3               |       |
| ≥25                              | 0–30                          | 91.9               | 89.7  |
| <25                              | 0–60                          | 95.5               |       |
|                                  | ≥30–40                        | 71.4               |       |
|                                  | 0–40                          | 35.7               |       |
| ≥20                              | ≥30                           | 92.3               |       |
| <20                              | 0–60                          | 96.3               |       |
|                                  | ≥40                           | 76.9               |       |
|                                  | 0–40                          | 41.7               |       |
| ≥25                              | 0–60                          | 91.9               | 89.7  |
| <25                              | 0–60                          | 95.5               |       |
|                                  | ≥30–40                        | 76.9               |       |
|                                  | ≥40                           | 41.7               |       |

(1) % DM: percentage of that component in the total diet. (2) % percentage of samples whose estimated % FG (from the formula) is placed in the same stratum as the reference standard.

The percentage of well-classified samples exceeds 95% for samples without fresh grass, not "grass-fed milk", while only about 70% of the samples with fresh grass are well-classified. The percentage of well-classified samples exceeds 90% for samples with FG and low concentrate content (<30%). However, this percentage drops to 53.6% and 60% for thresholds of 20% FG and 25% FG, respectively, when the percentage of concentrate in the diet exceeds 30%. The percentage of well-classified samples drops to 35.7% and 41.7% for thresholds of 20% FG and 25% FG, respectively, when the concentrate content in the diet exceeds 40% (Table 7).

From this, it is concluded that the algorithm works very well with samples without fresh grass, and that a high content of concentrate in the diet (≥40%/even ≥30%) distorts the fatty acid profile of the rations with fresh grass.

We then apply the other validity indicators (d, EF) to analyse the misclassified samples, choosing the 25% threshold.

Table 8 presents the statistical indices calculated for the analysis of the differences between the estimated values (obtained from the equation) and those observed (reference standard), for a threshold of 25% fresh grass in the diet. The coefficient of determination ($R^2$) is 0.70—a very close value to that obtained with the regression ($R^2_{adj} = 0.72$) (Table 3).

Table 8. Mean error (ME), root-mean-square error (RMSE), mean absolute percentage error (MAPE), $R^2$, index of agreement (Willmott’s index) (d), and efficiency index (Nash–Sutcliffe efficiency) (EF) obtained by comparing the data estimated by the equation and the reference standard values. For the total population (N = 174 data) and eliminating anomalous samples.

| N   | Observed | Estimated | ME    | RMSE | MAPE | $R^2$ | EF   | d    |
|-----|----------|-----------|-------|------|------|-------|------|------|
| 174 | 20.6     | 17.6      | −2.93 | 1.16 | 4.01 | 0.70  | 0.688| 0.90 |
| 149 | 16.7     | 16.23     | −0.47 | 1.13 | 23.32| 0.74  | 0.81 | 0.94 |

The d index, with a value close to one, confirms a high degree of relative concordance between the observed data and the model’s predictions. The EF index indicates that the model is acceptable. The ME is negative, indicating that the model systematically underestimates the fresh grass content (Figures 2 and 3).
If we eliminate the data corresponding to the 25 samples misclassified by the algorithm (diets with more than 25% forage and more than 30% concentrate; FG ≥ 25% and concentrate ≥ 30%), the statistical indices improve notably: \( R^2_{aj} = 0.74 \), \( d = 0.94 \), and \( EF = 0.81 \), respectively.

Figures 2 and 3 graphically compare the values estimated by the equation with the observed data for the percentage of fresh grass. In general, the model below predicts the evaluated variable, but clearly more so with the population of 174 data inputs. This fact is confirmed by comparing the means of the observed and estimated values indicated in Table 8. These figures differ by 14.5% for \( n = 174 \) and only by 3% for \( n = 149 \) (without ≥25% FG and ≥30% concentrate diets), justifying a lower prediction of the estimated values with respect to those observed.

4. Discussion

In terms of the equation, its simplicity stands out, with only 3 fatty acids explaining a large part of the variability for the fresh grass content in the rations of lactating cows. In
the model suggested by Coppa et al. [8], the proportion of fresh grass is predicted with a high confidence level ($R^2 = 0.81$; RMSE = 15.6) but includes 13 fatty acids, many of which are not fatty acids normally identified in chromatographic work.

The algorithm includes linolenic acid (C18:3-n3, the main omega-3 isomer), vaccenic acid (trans11-C18:1) and another isomer of C18:1, cis12-C18:1. Linolenic acid, rumenic acid (cis9-trans11-C18:2, main CLA) and vaccenic acid are the three fatty acids that increase most with the inclusion of fresh grass, and with forage in general, in the diet [1,3,20]. The high correlation between the content of vaccenic acid and rumenic acid in this study ($r = 0.975$; $p < 0.001$) [19] is a result very similar to that observed by Ferlay et al. [21] with an $r = 0.99$, and may explain the absence of the main CLA isomer (cis9, trans11-C18:2) in the model.

cis12-C18:1 is an isomer of oleic acid that presents a negative correlation with the percentage of fresh grass in the diet ($r = -0.547$; $p < 0.01$) [19]. This oleic acid isomer is not included among the FA from the chromatography column of many papers [6,8,22,23]. Coppa’s group does not include it in any of the equations used for predicting the content of all components in the ration using the fatty acid profile [8]. The works in which it was identified also relate it negatively to the content of fresh herbage in the diet [24–26].

The influence of concentrate content in the diet on the expected change to the FA profile in milk is a fact observed in many works. In a study by Stockdale et al. [29], they obtained an equation in which they calculated the CLA content according to the concentrate intake and observed that as the forage in the ration was supplemented with concentrate, the CLA content decreased, except when the amount of concentrate was such that it resulted in a reduction in total fat.

Tudisco et al. [30] found significant differences in the fatty acid profile between a group of cows in organic production, with a diet rich in fresh forage, compared to another group in conventional production, with a diet rich in hay, with the same level of concentrate, inferring that it is exclusively the forage that is responsible for the increase in C18:1 cis9, C18:1 trans11, linoleic acid, alpha-linolenic acid, c9 t11 CLA, t10 c12 CLA, total CLA, MUFA and PUFA.

Some authors explicitly underline the influence of the forage: concentrate ratio in the cows’ diet on the fatty acid profile [24,31,32]. Borreani et al. [24] concluded that high roughage-to-concentrate ratio could explain the high ratio of trans-11 C18:1 to trans-10 C18:1 in milk from farms feeding lactating cows with fresh herbage cut and fed indoors.
and a low roughage-to-concentrate ratio could explain its low value in milk from feeding lactating cows with a medium proportion of concentrate (CCM). Cavaliere et al. (2018) [32] found that milk from cows fed the higher F:C ratio (HFM) had a higher α-linolenic acid value (C18: 3n-3), total n-3, and CLA, and even claim that dietary supplementation with HFM (high forage-to-feed diet) in a rat animal model decreased lipid accumulation in the liver, decreasing inflammation and oxidative stress in the animals.

From this data, it follows that the percentage of concentrate in diets shows inverse correlations to the percentage of fresh forage, and the correlation between the main fatty acids and fresh grass is higher than the concentrate [19]. However, in-depth analysis about the misclassified milk samples with the algorithm led us to the conclusion that concentrate is able to neutralize the benefits of fresh grass on the fatty acid and antioxidant profiles.

5. Conclusions

The application of linear regressions to the relationships between diet, fatty acid profile, and fat-soluble antioxidants has allowed us to obtain a model that estimates the percentage of fresh grass in the diet according to its composition in the context of Green Spain. The algorithm obtained did not include any antioxidants, but it included three fatty acids which were considerably related to the presence of fresh grass in the diet: linolenic acid (C18:3-n3), vaccenic acid (trans11-C18:1), and cis12-C18:1. The predictive power of the equation is high (R^2 ≥ 0.7) and allows the authentication of a milk sample as “grass-fed milk”. For a threshold of 25% fresh grass, the equation is correct for 95.5% of milk samples from farms that contain less than 25% fresh grass. However, the accuracy percentage with milk samples obtained from cows with a ration rich in fresh grass decreases with high contents of concentrate in the diet (≥30%). This means that cows fed a high fresh grass content may have a similar fatty acid profile to cows that have been fed no fresh grass or little fresh grass when a lot of concentrate is included in the diet.

If we try to support a form of production based on fresh grass, the application of the formula is not beneficial. The formula does not identify milk from a cow fed a percentage of fresh grass in the diet above a certain threshold; however, it does guarantee a milk with a good profile of fatty acids compatible with milk from an animal with a high forage and low concentrate diet. So, if we are interested in the nutritional profile of milk, in other words, the profile of fatty acids, the definition of what is or what is not “grass-fed milk” should not only be determined according to the percentage of fresh grass in the diet but should take into consideration the diet as a whole, especially with regard to the content of concentrate in the ration.

Author Contributions: Conceptualization, A.V., J.B.; methodology, J.B., A.V., G.S.; Investigation, A.V., J.B., G.S., I.V.-G.; formal analysis, J.B., A.V., G.S., I.V.-G.; writing—original draft, A.V., J.B.; writing—review, I.V.-G., G.S., E.S. All authors have read and agreed to the published version of the manuscript.

Funding: The work presented is part of a research project funded by the National Institute for Agricultural and Food Research and Technology (INIA), project RTA2014-00086-C03.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors of this work acknowledge that this project is part of a collaboration with the agricultural research centers of Galicia (CIAM), Asturias (SERIDA), and Navarra (INTIA). We also want to give a special mention to the chromatographic work done at INGACAL-CIAM, especially the work done by Laura González. Of course, we would like to thank the collaboration of the farms from the four Autonomous Communities.

Conflicts of Interest: The authors declare no conflict of interest.
25. Borreani, G.; Coppa, M.; Revello-Chion, A.; Comino, L.; Giaccone, D.; Ferlay, A.; Tabacco, E. Effect of different feeding strategies in intensive dairy farming systems on milk fatty acid profiles, and implications on feeding costs in Italy. *J. Dairy Sci.* **2013**, *96*, 6840–6855. [CrossRef]

26. Loor, J.J.; Ferlay, A.; Ollier, A.; Doreau, A.; Chilliard, Y. Relationship among trans and conjugated fatty acids and bovine milk fat yield due to dietary concentrate and linseed oil. *J. Dairy Sci.* **2005**, *88*, 726–740. [CrossRef]

27. Molkentin, J.; Giesemann, A. Differentiation of organically and conventionally produced milk by stable isotope and fatty acid analysis. *Anal. Bioanal. Chem.* **2007**, *388*, 297–305. [CrossRef] [PubMed]

28. Kraft, J.; Collomb, M.; Möckel, P.; Sieber, R.; Jahreis, G. Differences in CLA isomer distribution of cow’s milk lipids. *Lipids* **2003**, *38*, 657–664. [CrossRef] [PubMed]

29. Stockdale, C.R. Levels of pasture substitution when concentrates are fed to grazing dairy cows in northern Victoria. *Aust. J. Exp. Agric.* **2000**, *40*, 913–921. [CrossRef]

30. Tudisco, R.; Cutrignelli, M.I.; Calabrò, S.; Piccolo, G.; Bovera, F.; Guglielmelli, A.; Infascelli, F. Influence of organic systems on milk fatty acid profile and CLA in goats. *Small Rumin. Res.* **2010**, *88*, 151–155. [CrossRef]

31. Chilliard, Y.; Glasser, F.; Ferlay, A.; Bernard, L.; Rouel, J.; Doreau, M. Diet, rumen biohydrogenation and nutritional quality of cow and goat milk fat. *Eur. J. Lipid Sci. Technol.* **2007**, *109*, 828–855. [CrossRef]

32. Cavaliere, G.; Trinchese, G.; Musco, N.; Infascelli, F.; De Filippo, C.; Mastellone, V.; Morittu, V.M.; Lombardi, P.; Tudisco, R.; Grossi, M.; et al. Milk from cows fed a diet with a high forage: Concentrate ratio improves inflammatory state, oxidative stress, and mitochondrial function in rats. *J. Dairy Sci.* **2018**, *101*, 1843–1851. [CrossRef]