**Developing a predictive model for the energy content of goat milk as the basis for a Functional Unit formulation to be used in the LCA of dairy goat production systems**

P. P. Danieli and B. Ronchi

**Supplementary Table S1** *List of keyword and bibliographic databases used in the computer-based search for data on GE content and composition of goat milk and published predictive models (PPMs) for the GE content of goat or cow milk.*

|  |  |
| --- | --- |
| Keywords | Bibliographic databases |
| Milk | USDA National Agriculture Library |
| Energy content | ISI Web of ScienceTM |
| Goat milk | Google Scholar© |
| Gross energy | MEDLINETM |
| Calorimetry | Scifinder® |
| Milk composition | Scopus® |
| Nutritional values |  |
| Milk quality |  |

**Supplementary Table S2** *List of data sources, ordered by publication date, from which data on the energy content and compositional traits of goat milk have been obtained*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Reference1 | Breed | N. of animals | Parity | Lactation stage | Groupingfactor(s) | N. data | Data type | Data codes |
| Peterson and Turner (1939) | ND | 4 | ND | ND | ND | 13 | id | 2A2.3 |
| Perrin (1958) | ND | 1 | ND | Early | DIM | 5 | id | 23A3 |
| Aguilera *et al.* (1990) | Granadina | 12 | 2 | Mid | Diet, DIM | 4 | ad | 20A, 20D |
| SanzSampelayo *et al.* (1988) | Granadina | ND | ND | ND | - | 1 | ad 4 | 29A |
| SanzSampelayo *et al.* (1998) | Granadina | 5 | 2 | Mid | Diet | 2 | ad | 16A, 16B |
| SanzSampelayo *et al.* (1999) | Granadina | 20 | 2 | Mid | Diet | 4 | ad | 25A - 25D |
| SanzSampelayo *et al.* (2002) | Granadina | 6 | 2 | Mid | Diet | 3 | ad | 11A - 11C |
| Molina-Alcaide *et al.* (2010) | Granadina | 18 | 3 | Mid | Diet | 3 | ad | 17A - 17C |
| Tovar-Luna *et al.* (2010) | Alpine | 16 | 1-3 | Early-late | Diet, DIM | 6 | ad | 19A - 19F |
| Magistrelli *et al.* (2011) | ND | 8 | ND | ND | - | 1 | ad4 | 31A |

ND = not declared; DIM = days in milk; id = individual (raw) data; ad = aggregated data (mean value or referring to bulk samples).

1 See the references below.

2 The data source provided 18 records reporting the GE content, but only 13 were found completed with all the compositional data.

3 All the individual data were aggregated to calculate an (arithmetic) mean value.

4Sample of bulk milk, no details about the flock have been retrieved.

**Supplementary Table S3** *List of published predictive models (PPMs) for the GE content of goat or cowmilk*

|  |  |  |
| --- | --- | --- |
| PPM | PPM codes | Reference2 |
| GE = 92.3 × F% + 57.1 × P% + 39,5 × Lac% | Abd | Abderhalden (1908) |
| GE = 239 × (0.0406 × 10 × F% + 1.509) | AFRC | AFRC (1998) |
| GE (kcal/lb) = 4132 × F(lb) + 2658 × P(lb) + 1792 × Lac(lb) | And | Andersen (1926) |
| GE = 392,4 + 103,3 × F% | E&L | Economides and Louca (1981) |
| GE (kcal/lb) = 51 × (2+2/3) + 51 × F% | Ga | Gaines (1928) |
| GE = 93,18 × F% + 58,6 × P% + 39,5 × Lac%; GE = 96,50 × TS | M&H1, M&H2 | Malcolm and Hall (1907) |
| GE = 292 + 104,74 × F%; GE = 9,3 × F% + 4,1 × SNF% | M&P1, M&P2 | Mavrogenis and Papachristoforou (1988) |
| GE = 312,9 + 117,7 × F% | MF&S | Morand-Fehr and Sauvant (1978) |
| GE (kcal/qt) = 52,78 × F% + 16,41 × P% + 37,87 × TS% + 46,91 × Lac% -2,75 × d(20°C)-57,70; GE (kcal/qt) = 113,7334 × (F% + 2,4404)  GE (kcal/qt) = 105,287 × (F% + 2,4185); GE (kcal/qt) = 90,67 × F% + 54,27 × P% + 26,73 × Lac% + 55,44 | O&S1, O&S2, O&S3, O&S4 | Overman and Sanmann (1926) |
| GE (kcal/100g) = 9,11 × F% + 37,4 × TN% + 3,95 × Lac%  GE (kcal/100g) = 9,11 × F% + 37,4 × (TN%-NPN%) + 3,95 × Lac% | Per1, Per2 | Perrin (1958) |
| GE (kcal/100ml) = 39,618+9,564 × F%; GE (kcal/100ml) = 21,682+7,085TS% | P&T1, P&T2 | Peterson and Turner (1939) |
| GE (kcal/lb) = 4220 × F(lb) + 1860 × SNF(lb) | S&B | Stocking and Brew (1920) |
| GE (kcal/lb) = 41,84 × F% + 22,29 × SNF% - 25,58 | T&R | Tyrrell and Reid (1965) |
| GE =1000 × (0,0929 × F%+0,0547 × P%+0,192) | NRC | NRC (2001) |
| GE = (1.4694 + 0.4025 × F%) × 239 | IGR | Nsahlai *et al.* (2004) |
| GE = (0,4 + 0,0075 × (F% − 35)) × 1700 | INRA | INRA (2007) |

1 Expressed as kilocalories per kilogram, unless otherwise indicated: kilocalories per 100 millilitres (100ml), per 100 grams (100g), per quart (qt), per pound (lb).

2 See the references listed below.

**Supplementary Table S4** *List of statistics used in the multi-step process for testing the developed regression models for the gross energy (GE) prediction in goat milk*

| Object of evaluation | Statistics | Concise description of rationale and methods used |
| --- | --- | --- |
| Outliers | Weisberg “t” statistic  Mahalanobis distance (D2) | Regarding the occurrence outlying data in the space of dependent variable (GE), the “t” statistic was calculated according to Weisberg (1980)1 for the standardised residuals (z residuals) within each model. The calculated values were compared with the critical thresholds (α = 0.05) reported in Weisberg (1980): 3.50, 3.52 and 3.53 for regressions based on 1, 2 or 3 predictors, respectively.  The Mahalanobis distances (D2) was used to assess the occurrence of outlying data in the predictors’ space testing against the critical values at the 0.05 level of significance of 8.24, 9.94 and 11.48 for a sample size of 25 observations and 1, 2 or 3 predictors, respectively (Barnet and Lewis, 1978). |
| Infuential data | Leverage (L)  Cook’s distance (D) | Because the occurrence of influential point data, even in the absence of outliers, may affect the reliability of a regression model, thereby reducing its predictive potential (Stevens, 1978), specific residual statistics were performed: the leverage (L) and Cook’s distance (D) statistic. The leverage gives a measure of the impact of a single observed response on the respective predicted value by a model. In this study, cut off values for L of 0.22, 0.33 and 0.44 were adopted as suggested by Cohen *et al.* (2003) for small samples and 1, 2 and 3 regressors, respectively.  The Cook’s distance is related to the shift of the standardised regression parameters (*βi*s) within their confidence region due to one or more influential data points (Cook, 1977). As a rule of thumb, data points exhibiting a D value greater than one should be regarded as having a high degree of influence on the regression (Cook, 1977). However, we used a precautionary cut off value as low as 0.17 that allows, at the worst, for the *i-th* data point to move the regression parameters only within the 5% (α = 0.05) of the corresponding confidence region (Cook, 1977). |
| Multi-collinearity / over-fitting | Variance Inflation Factor (VIF)  Akaike’s Information Criterion (AIC) | Multi-collinearity was assessed through the Variance Inflation Factor (VIF) against a threshold of 10 (Kutner, 2004) but viewing at VIF values ≥ 3 and < 10 as potential cases of unrevealed collinearity.  Over-fitting may be avoided by minimizing the Akaike’s Information Criterion (AIC) (Akaike, 1974). In this work, the AIC of each model was scaled to the lowest one recorded and the Loss of Information (∆*i*) experienced if one considers the fitted model *i* rather than the best one and was then estimated according to Burnham and Anderson (2004). The ∆*i* values allow for a meaningful interpretation without the unknown scaling constants and sample size that enter into the AIC values (Burnham and Anderson, 2004). |
| Variable selection | Mallows’s *Cp* statistic | For the variable selection as a further step in the DPM evaluation process, one of the useful tools is the Mallows’s *Cp* statistic. The procedure requires fitting the regressions by comparing the results to the respective *Cp* statistic (Mallows, 1973). For selection purposes, the *Cp* values obtained were compared with the thresholds of 3.1, 3.9 and 4.8 calculated for 1, 2 and 3 regressors according to Gilmour (1996). |

1 See the references listed below

**Supplementary Table S5** *Results of the Principal Component Analysis performed on the complete dataset (26 records) and partial dataset, including the goat milk yield (22 records): factor loading values of the original variables on the three principal components (PCs)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Whole dataset (n = 26) | | |  | Partial dataset (n = 22) | | |
|  | PC 1st | PC 2nd | PC 3rd |  | PC 1st | PC 2nd | PC 3rd |
| MY (kg/d) | - | - | - |  | 0.874 | 0.136 | -0.324 |
| F (%) | 0.965 | 0.055 | -0.029 |  | -0.961 | -0.061 | 0.024 |
| P (%) | 0.791 | -0.207 | -0.491 |  | -0.796 | 0.196 | 0.488 |
| Lac (%) | 0.738 | 0.088 | 0.649 |  | -0.734 | -0.088 | -0.656 |
| TS (%) | 0.994 | 0.041 | 0.033 |  | -0.992 | -0.048 | -0.039 |
| SNF (%) | 0.965 | 0.018 | 0.126 |  | -0.965 | -0.025 | -0.133 |
| ASH (%) | -0.258 | 0.952 | -0.127 |  | 0.265 | -0.950 | 0.120 |
| GE (MJ/kg) | 0.978 | 0.086 | 0.007 |  | -0.978 | -0.093 | -0.014 |

MY = milk yield; F = fat; P = protein; Lac = lactose; TS = total solids; SNF = solids non-fat; ASH = ash; GE = gross energy.

**Supplementary Table S6** *Newly developed predictive model (DPMs) selection: evaluation of outliers and influential data, standard error of the estimate, collinearity, information criterion and the Mallows’s Cp statistic*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Eq. | Predictors | *t*max | *D2*max | *Lmax* | *D*max | VIF1 | Δ*i*1 | *Cp*1 | *Retained models* |
| 1.1 | TS | 2.22 | 4.17 | 0.20 | 0.12 | - | 11.0 u | 36.8 u | NO |
| 1.2 | F | 2.04 | 3.51 | 0.17 | 0.17 | - | 4.5bl | 35.0 u | YES |
| 2.1 | F, TS | 2.64 | 5.12 | 0.32 | 0.14 | 12.5/12.5 u | 0 s | 22.0 s | YES |
| 2.2 | F, SNF | 2.64 | 7.39 | 0.32 | 0.14 | 2.4/2.4 s | 0 s | 22.0 s | YES |
| 2.3 | F, Lac | 2.43 | 7.39 | 0.26 | 0.20 | 1.8/1.8 s | 11.6 u | 35.7 u | NO |
| 2.4 | F, P | 1.98 | 5.90 | 0.23 | 0.14 | 1.9/1.9 s | 10.3 u | 36.9 u | NO |
| 3.1 | F, TS, ASH | 2.52 | 8.48 | 0.36 | 0.13 | 1.1/13.7 u | 0.8 s | 21.5 s | YES |
| 3.2 | F, SNF, ASH | 2.53 | 8.47 | 0.36 | 0.13 | 1.1/2.6 s | 0.8 s | 21.5 s | YES |
| 3.3 | F, Lac, ASH | 2.63 | 7.51 | 0.33 | 0.32(17A, 17C) | 1.1/1.9 s | 8.4 u | 27.2 bl | NO |
| 3q.1 | F, Lac, Lac2 | 2.83 | 8.97 | 0.38 | 0.24(2A, 17C) | 1.1/2.0 s | 6 bl | 23.2 bl | NO |
| 3q.2 | F, Lac, ASH2 | 1.86 | 8.76 | 0.37 | 0.29(17C) | 1.1/2.6 s | 6.5 bl | 24.3 bl | NO |

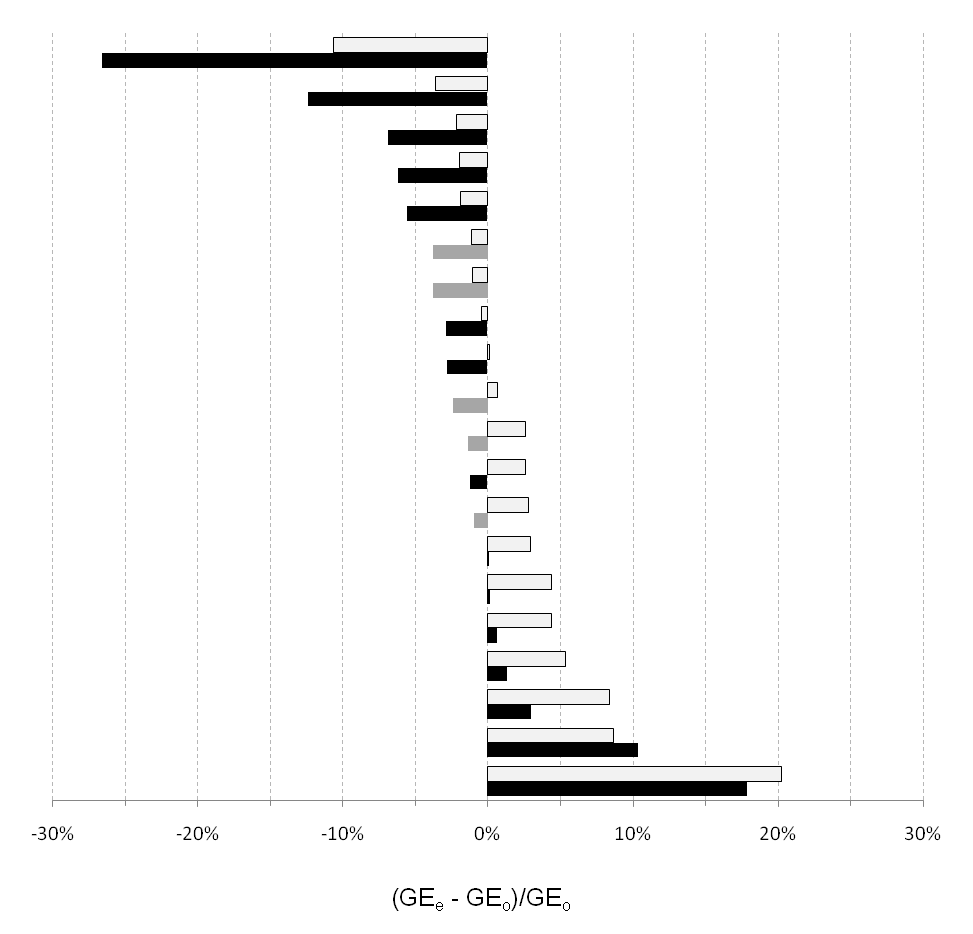
*tmax*= highest value of the Weisberg’s statistic “t” recorded for each DPM; *D2max* = highest value of the Mahalanobis distance recorded for each DPM; *Lmax*= highest value of the leverage registered for each DPM; *Dmax* = highest value of the Cook’s distance recorded for each DPM (within brackets the codes of records for which the D values were greater than the threshold); VIF = Variance Inflation Factor (min/max); Δ*i*: information loss (*i.e.,* the difference between an AIC value and the minimum value recorded for the set of regression models); *Cp* = Mallows’s *Cp* statistic.

1Superscripts indicate that results may be considered satisfying (s), borderline (bl) or unsatisfying (u) the selection conditions (see the text for further details).

**Supplementary Table S7** *Results of regression analysis with permutation tests applied to the collected dataset of goat milk traits (mean values)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Eq. | Regression coefficients | Torig | c | *P* = c/N | 95% conf.  interval of *P* | |
| 1.2 | βF = 0.259 | TF = 18.492 | 0 | < 0.0001 | 0 | 0.0004 |
| 2.1 | βF = 0.209  βTS = - 0.133 | TF = 3.065  TTS = 3.103 | 52  57 | 0.0052  0.0057 | 0.0039  0.0043 | 0.0068  0.0074 |
| 2.2 | βF = 0.342  βSNF = - 0.133 | TF = 11.530  TSNF = 3.103 | 0  48 | < 0.0001  0.0048 | 0  0.0035 | 0.0004  0.0064 |
| 3.1 | βF = 0.233  βTS = 0.119  βASH = 0.195 | TF = 3.303  TTS = 2.710  TASH = 1.188 | 30  122  2442 | 0.0030  0.0122  0.2442 | 0.0020  0.0101  0.2358 | 0.0043  0.0145  0.2527 |
| 3.2 | βF = 0.352  βSNF = 0.119  βASH = 0.195 | TF = 11.467  TTS = 2.710  TASH = 1.188 | 0  133  2482 | < 0.0001  0.0133  0.2482 | 0  0.0111  0.2398 | 0.0004  0.0157  0.2567 |

βi = regression coefficient for the *i*-th predictor included in a linear regression model. Torig = Wald statistic calculated by the estimated regression coefficients and the relative standard errors for each predictor variable included during the regression model development (Ti= βi/SEi). c = number of Ti value (absolute value) obtained during the permutation test equal or higher than the respective Ti-orig value; N= number of permutations for each test = 20,000.

**Supplementary Figure S1** Bar plot of the errors of the gross energy estimate (GEe) to the GEo ratio of the individual goat milk samples from Peterson and Turner (1939) (black bars) and Perrin (1958) (gray bars) estimated through the M&P1 model or Eq. 2.2 (all white bars) developed in this work. The data are ranked in ascending order as separate series, according to the predictive model used.

References

Akaike H 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control AC-19, 716-723.

Abderhalden E 1908. Text-book of physiological chemistry in thirty lectures. John Wiley & Sons, New York, USA (Cited by Perrin, 1958).

Agricultural and Food Research Council (AFRC) 1998. The nutrition of goats. Technical Committee on Response to Nutrients, Report No. 10. CAB International, Wallingford, UK.

Aguilera JF, Prieto C and Fonollá J 1990. Protein and energy metabolism of lactating Granadina goats. British Journal of Nutrition 63, 165–175.

Andersen AC 1926. Undersogelser over komaelkens sammensaetning. (Investigations on the composition of cow's milk.) Beretning Nordisk Jordbrugsforskning 4/7, 133-145. (Cited by Gaines, 1928).

Barnett V and Lewis T 1978. Outliers in statistical data. Wiley Series in Probability and Mathematical Satitstics. John Wiley and Sons Ltd., New York, USA.

Burnham KP and Anderson AR 2004. Multimodel inference. Understanding AIC and BIC in model selection. Sociological Methods & Research 33, 261-304.

Cohen J, Cohen P, West SG and Aiken LS 2003. Applied multiple regression/correlation analysis for the behavioral sciences. Lawrence Erlbaum Associates, Publishers, Mahwah, NJ, USA.

Cook RD 1977. Detection of influential observation in linear regression. Technometrics 19, 15-19.

Economides S and Louca A 1981. The effect of the quality and quantity of feed on the performance of pregnant and lactating goats. In Nutrition and Systems of goat feeding, International Symposium 12-15 May 1981, Tours, France. pp. 286-291.

Gaines WL 1928. The energy basis of measuring milk yield in dairy cows. Agricultural Experiment Station Bulletin 308, 403-438.

Gilmour SG 1996. The Interpretation of Mallows'sCp-statistic. The Statistician 45, 49-56.

Higgins JPT and Thompson SG 2004. Controlling the risk of spurious findings from meta-regression. Statistics in Medicine 23, 1663-1682.

Institutnational de la recherché agronomique (INRA) 2007. Alimentation des bovins, ovins e tcaprins. Besoins animaux. Valeurs des aliments. Editions Quae, Versailles, France.

Kutner MH, Nachtsheim CJ and Neter J 2004. Applied Linear Regression Models. McGraw-Hill, New York, NY, USA.

Magistrelli D, Pinotti L, Rapetti L and Rosi F 2011. Ghrelin, insulin and pancreatic activity in the peri-weaning period of goat kids. Journal of Animal Physiology and Animal Nutrition 95, 40–46.

Malcolm J and Hall AA 1907. The heat value of milk as a test of its quality. The Journal of Agricultural Science 2, 89–95.

Mallows CL 1973. Some comments on Cp. Technometrics 15, 661-675.

Mavrogenis AP and Papachristoforou C 1988. Estimation of the energy value of milk and prediction of fat-corrected milk yield in sheep and goats. Small Ruminant Research 1, 229–236.

Molina-Alcaide E, Morales-García EY, Martín-García AI, Ben Salem H, Nefzaoui A and Sanz-Sampelayo MR 2010. Effects of partial replacement of concentrate with feed blocks on nutrient utilization, microbial N flow, and milk yield and composition in goats. Journal of Dairy Science 93, 2076–2087.

Morand-Fehr P and Sauvant D 1978. Nutrition and optimum performance of dairy goats. Livestock Production Science 5, 203–213.

National Research Council (NRC) 2001. Nutrient requirements of dairy cattle. National Research Council, National Academy of Sciences, Washington DC, USA.

Nsahlai IV, Goetsch AL, Luo J, Johnson ZB, Moore JE, Sahlu T, Ferrell CL, Galyean ML and Owens FN 2004. Metabolizable energy requirements of lactating goats Small Ruminant Research 53, 253–273.

Overman OR and Sanmann FP 1926. The energy value of milk related to composition. Agricultural Experiment Station Bulletin 282, 206-218.

Perrin DR 1958. The calorific value of milk of different species. Journal of Dairy Research 25, 215–220.

Peterson VE and Turner CW 1939. The energy content of goat milk. Journal of Nutrition 17, 293–301.

SanzSampelayo MR, Munoz FJ, Guerrero JE, Gil Extremera F and Boza J 1988. Energy metabolism of the Granadina breed goat kid. Use of goat milk and milk replacer. Journal of Animal Physiology and Animal Nutrition 59, 1-9.

Sanz-Sampelayo MR, Perez L, Boza J and Amigo L 1998. Forage of different physical forms in the diets of lactating Granadina goats: nutrient digestibility and milk production and composition. Journal of Dairy Science 81, 492–498.

SanzSampelayo MR, Pérez L, Gil Extremera F, Boza JJ and Boza J 1999. Use of different dietary protein sources for lactating goats: milk production and composition as functions of protein degradability and amino acid composition. Journal of Dairy Science 82, 555–565.

SanzSampelayo MR, Pérez L, Martı́nAlonso JJ, Amigo L and Boza J 2002. Effects of concentrates with different contents of protected fat rich in PUFAs on the performance lactating Granadina goats: Part II. Milk production and composition. Small Ruminant Research 4, 141–148.

Shapiro SS and Wilk MB 1965. An analysis of variance test for normality (complete samples). Biometrika 52, 591-611.

Stata 2013. Permute – Montecarlo permutation tests. In Stata base reference manual. Release 14, pp. 1573-1582. Stata Press, College Station, TX, USA,

Stevens JP 1984. Outliers and influential data point in regression analysis. Psychological Bulletin 95, 334-344.

Stocking WA and Brew JD 1920. Milk – the essential food. The Dairyman's League News 3, 10 (cited by Gaines, 1928).

Tovar-Luna I, Puchala R, Sahlu T, Freetly HC and Goetsch AL 2010. Effects of stage of lactation and dietary concentrate level on energy utilization by Alpine dairy goats. Journal of Dairy Science 93, 4818–4828.

Tyrrell HF and Reid JT 1965. Prediction of the energy value of cow’s milk. Journal of Dairy Science 48, 1215–1223.

Weisberg S 1980. Applied linear regression. John Wiley and Sons Ltd., New York, NY, USA.