J. Dairy Sci. 95:7346–7354
http://dx.doi.org/10.3168/jds.2012-5546

Use of multivariate factor analysis to define new indicator variables for milk composition and coagulation properties in Brown Swiss cows

N. P. P. Macciotta,*[1] A. Cecchinato,† M. Mele,‡ and G. Bittante†
*Dipartimento di Agraria, Sezione Scienze Zootecniche, Università di Sassari, Via de Nicola 9, 07100 Sassari, Italy
†Dipartimento di Agronomia, Animali, Alimenti, Risorse Naturali e Ambiente, Università di Padova, Viale dell'Università 16, 35020 Legnaro, Padova, Italy
‡Dipartimento di Agronomia e Gestione dell'Agroecosistema, Sezione di Scienze Zootecniche, Università di Pisa, Via S. Michele degli Scalzi, 2, 56124 Pisa, Italy

ABSTRACT

The aim of this study was to elucidate the structure of relationships between milk yield, composition, and coagulation properties of Brown Swiss cattle. Multivariate factor analysis was used to derive new synthetic variables that can be used for selection purposes. For this reason, genetic parameters of these new variables were estimated. Individual records on milk yield, fat and protein percentages, casein content, lactose percentage, somatic cell count, titratable acidity, and pH were taken on 1,200 Italian Brown Swiss cows located in 38 herds. Factor analysis was able to extract 4 latent variables with an associated communality equal to 70% of the total original variance. The 4 latent factors were interpreted as indicators of milk composition, coagulation, acidity, and mammary gland health, respectively. Factor scores calculated for each animal exhibited coherent patterns along the lactation and across different parities. Estimation of genetic parameters of factor scores carried out with a multiple-trait Bayesian hierarchical model showed moderate to low heritabilities (raging from 0.10 to 0.23) and genetic correlations (from −0.15 to 0.46). Results of the present study support the hypothesis of a simpler structure that controls, at least in part, the covariance of milk composition and coagulation properties. Moreover, extracted variables may be useful for both breeding and management purposes, being able to represent, with a single value for each animal, complex traits such as milk coagulation properties or health status of the mammary gland.
Key words: milk coagulation property, milk composition, factor analysis

INTRODUCTION

Breeding goals currently included in selection programs for dairy cattle are related to milk yield and quality, milkability, type, longevity, health, and reproduction. The aggregation of all of these traits in economic selection indexes is often complex. In particular, the estimation of a coherent (co)variance matrix is hampered by collinearity between some traits and by the need for different models (linear, censored, threshold). A further complication is difficulty in estimating selection index weights and in understanding their proper meaning.

An example of the situation described above is represented by quality traits related to milk nutritional value and technological properties. Fatty acid composition (Soyeurt et al., 2006), protein composition (Rutten et al., 2011), and milk coagulation properties (**MCP**; Dal Zotto et al., 2008; Cecchinato et al., 2009; De Marchi et al., 2009) are becoming increasingly important because of consumer concerns. Actually, milk nutritional and technological properties are defined by several variables, among which exist a complicated correlation structure that is rather difficult to interpret (Ikonen et al., 2004). Studies on correlations between milk production traits and coagulation properties carried out with standard multiple-trait genetic models have reported conflicting results in terms of both magnitude and sign of the estimated values (Ikonen et al., 2004; Cassandro et al., 2008; Cecchinato et al., 2011; Vallas et al., 2010). Apart from possible effects of sampling and differences in the statistical approaches used, such variability in the results could be because the nature of genetic relationships among these traits is only partially known. In particular, with common statistical genetic approaches, such as univariate or multivariate animal models, it is rather difficult to assess the covariance because of the direct linkage between 2 variables from the quota that could be ascribed to an indirect relationship because of other variables defining the system or due to environmental factors. An approach able to identify a possible simpler underlying structure would be advisable.

Received March 19, 2012.
Accepted September 1, 2012.
[1]Corresponding author: macciott@uniss.it

On the contrary, the choice of a single variable for representing milk technological properties may not be the optimal strategy, given that each of the MCP is related to specific aspects of the coagulation process and that some technical issues in their measurement still exist (Bittante, 2011). In addition, from a practical standpoint, farmers often prefer to deal with a combined index instead of several single traits (Glantz et al., 2011). An alternative approach could be aimed at exploiting the correlation structure between original variables to derive synthetic indexes without any a priori definition.

Multivariate statistics offer a set of efficient tools for analyzing complex correlation patterns. A suitable technique for investigating the correlation structures of variables defining milk quality is the factor analysis. This is because its basic theoretical assumption is that the (co)variance of a complex multivariate system can be partitioned into 2 quotas (Morrison, 1976). The first is shared by all variables and is called communality, whereas the second is peculiar to each variable and is termed uniqueness. The common covariance is generated by a set of one or more latent variables, termed common factors. Thus, factor analysis can be used to investigate relationships between the original variables, and values of common factors (scores) can be treated as new phenotypes and considered for further analyses. A further feature of interest for factors is that they are uncorrelated or, in the case of rotations, correlated weekly. For example, Todaro et al. (2005) used factor analysis to study the yield, chemical composition, and coagulation properties of goat milk. They derived 3 latent factors from 11 original variables. Factors were related to coagulation speed, milk yield, and curd firmness, respectively.

Previously, factor analysis has been used in animal science to find indicators of management and production levels for dairy cattle herds (Enevoldsen et al., 1996), to evaluate relationships between longevity and type traits (Vukasinovic et al., 1997), and to model the shape of the lactation curve (Wilmink et al., 1987; Macciotta et al., 2004; Macciotta et al., 2006; Aspilcueta-Borquis et al., 2011). However, no studies dealing with factor analysis applied to MCP and milk quality traits in dairy cattle are available in the literature.

The aim of this paper was to study the structure of relationships between milk yield, milk composition, and MCP of Brown Swiss cattle farmed in Italy. In particular, multivariate factor analysis was used to derive new synthetic variables with a possible technical meaning that could be used for selection purposes. For this reason, the genetic parameters of these new variables were also estimated.

MATERIALS AND METHODS

Sample Collection and Analysis of MCP and Milk Quality Traits

A total of 1,200 Italian Brown Swiss cows were milk-sampled during the period from June 2006 to July 2007. Cows were progeny of 50 AI sires and were reared in 30 herds located in northern Italy. Individual milk samples (1 per animal) were collected during the morning milking of a test day. After collection and with no preservative added, milk samples were stored in portable refrigerators (4°C) and transferred to the milk quality laboratory of the Veneto Agricoltura Institute (Thiene, Italy). Measures of MCP were obtained by using a computerized renneting meter (Polo Trade, Monselice, Italy). This measuring device has been widely used to investigate MCP (Ikonen et al., 1999; Cassandro et al., 2008; Bonfatti et al., 2011). The principle of the computerized renneting meter is based on control of the oscillation, which is driven by an electromagnetic field created by a swinging pendulum immersed in the milk container. A survey system measures differences in the electromagnetic field caused by milk coagulation: the greater the extent of coagulation, the smaller the pendulum swing. The analysis produces a diagram, as reported by Dal Zotto et al. (2008). Milk samples (10 mL) were heated to 35°C, and 200 μL of rennet (Hansen standard 190 with 63% of chymosin and 37% of pepsin; Pacovis Amrein AG, Bern, Switzerland), diluted to 1.6% in distilled water, was added to the milk. Measurement of MCP ended within 31 min after addition of the clotting enzyme. This analysis provided measurements of rennet coagulation time (the time interval, in minutes, from addition of the clotting enzyme to the beginning of the coagulation process) and a_{30} (the width, in millimeters, of the diagram at 31 min after the addition of rennet, which is a measure of curd firmness). Samples that did not coagulate within 31 min were classified as noncoagulating milk.

In addition to MCP, measures of milk fat percentage, milk protein percentage, milk CN content, lactose percentage, titratable acidity (expressed in Soxhlet-Henkel degrees), milk pH, and SCC were available. Values of SCC were log-transformed to SCS (Ali and Shook, 1980).

Pedigree information was supplied by the Italian Brown Swiss Cattle Breeders Association (ANARB, Verona, Italy) and included cows with phenotypic records for the investigated traits and all their known ancestors. Each animal with phenotypic records had at least 4 known ancestors in the pedigree data. Means and standard deviations of the considered traits are reported in Table 1.

Table 1. Mean and standard deviation of milk yield, composition, and coagulation traits

Trait	Mean	SD
Milk yield (kg)	29.16	7.75
TS (%)	9.32	0.47
Fat percentage	3.93	0.83
Protein percentage	3.66	0.35
CN percentage	2.85	0.27
Lactose percentage	4.96	0.22
SCC (×1,000/mL)	188	432
Titratable acidity	3.30	0.44
pH	6.67	0.15
Rennet coagulation time (min)	14.91	4.70
Rate of firming (min)	3.60	3.24
Curd firmness (mm)	40.05	11.38

Statistical Analysis

Theory of Multivariate Factor Analysis. Multivariate factor analysis aims at synthesizing information contained in a set of n observed variables (y_1, ..., y_n) by seeking a new set of p ($p < n$) variables (X_1,... X_p), termed common latent factors. Factor analysis assumes that the variance of each original variable can be decomposed into its common and unique components, termed communality and uniqueness, respectively. In matrix notation, the partitioning of the correlation (or covariance) matrix of the original variables ($\mathbf{S}$) can be written as

$$\mathbf{S} = \mathbf{BB}' + \mathbf{\Psi}, \qquad [1]$$

where $\mathbf{BB}'$ is the communality and $\mathbf{\Psi}$ is the residual (co)variance matrix (Morrison, 1976; McDonald, 1985). As a consequence of (co)variance modeling, each of the n original variables can be represented as a linear combination of p common factors that generates covariances between variables plus a residual specific variable (Morrison, 1976):

$$y_1 = b_{11}X_l + b_{12}X_2 + ... + b_{1p}X_p + e_1,$$

$$y_m = b_{m1}X_l + b_{m2}X_2 + ... + b_{mp}X_p + e, \qquad [2]$$

where X_j is the jth common factor, b_{ij} are factor coefficients (or loadings, i.e., correlations between the jth common factor), and e_i is the ith residual specific variable. Loadings are the elements of the $\mathbf{B}$ matrix used in model [1]. Common factors generate covariances between original variables, whereas the residual specifically contributes only to the individual variation. Factor analysis was performed on all composition traits and MCP using a varimax rotation (SAS Institute, 2008). The analysis was carried out on raw data. No precorrection was performed to evaluate the relationship between latent variables and some systematic factors.

The first step in assessing the suitability of a data set for factor analysis is based on a comparison between phenotypic Pearson and partial correlations values, with the latter measuring relationships among each pair of variables conditional on all possible effects of the other variables (Macciotta et al., 2004). This difference is measured by the Kaiser measure of sampling adequacy, which quantifies the difference between the off-diagonal elements of $\mathbf{S}$ and the anti-image correlation matrix $\mathbf{Q} = \mathbf{PS}^{-1}\mathbf{P}$, where $\mathbf{P} = [\text{diag}(\mathbf{S}^{-1})]^{-1}$ (Cerny and Kaiser, 1977).

No reference criterion is available for assessing the number of factors that should be retained (Schmitt, 2011). Thus, in the present work, the number of latent variables was chosen according to the amount of variance explained by the extracted factors, according to their readability in terms of biological meaning and relationships with the original variables (Morrison, 1976), and according to the outcome of the Bartlett chi-squared test, which compares $\mathbf{S}$ with $\mathbf{BB}'$ correlation matrices (Lawley and Maxwell, 1971).

Factor scores calculated for each cow were treated as a new variable and analyzed with the following univariate linear model (SAS Institute, 2008):

$$y_{ijkl} = HERD_i + PAR_j + DIM_k + e_{ijkl}, \qquad [3]$$

where y is the score of the pth factor of the lth cow; HERD is the fixed effect of the ith herd (38 levels); PAR is the fixed effect of the jth parity (5 levels: 1, 2, 3, 4, >4); DIM is the fixed effect of lactation stage k (10 intervals of DIM of 30 d, each starting from parturition); and e is the random residual, $\sim\mathbf{N}(0, \sigma^2_e)$. Least squares means of the DIM effect allow the reconstruction of the average lactation curve of the considered trait corrected for other effects included in the model (Stanton et al., 1992).

Genetic Analysis. Statistical inference was based on a set of bivariate analyses, which considered pairs of traits. These traits were factor scores and individual milk yield. Each bivariate analysis was based on the following linear mixed model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{h} + \mathbf{Z}_2\mathbf{a} + \mathbf{e},$$

where $\mathbf{y}$ was a vector of records for traits 1 and 2; $\mathbf{X}$, $\mathbf{Z}_1$, and $\mathbf{Z}_2$ were the appropriate incidence matrices for systematic effects in $\mathbf{b}$, herd effects in $\mathbf{h}$, and animal additive genetic effects in $\mathbf{a}$, respectively; and $\mathbf{e}$ was a vector of random residuals.

Bayesian Inference. (Co)variance components and related parameters were estimated using a Bayesian approach and Markov-chain Monte Carlo methods (Sorensen and Gianola, 2002). All traits were continuous variables, and their values were assumed to be sampled from the following multivariate normal distribution:

$$p(\mathbf{y}|\mathbf{b}, \mathbf{h}, \mathbf{a}, \mathbf{R}) \sim \mathbf{MVN}(\mathbf{Xb} + \mathbf{Z}_1\mathbf{h} + \mathbf{Z}_2\mathbf{a}, \mathbf{I} \otimes \mathbf{R}),$$

where $\mathbf{y}$, $\mathbf{b}$, $\mathbf{h}$, $\mathbf{a}$, $\mathbf{X}$, $\mathbf{Z}_1$, and $\mathbf{Z}_2$ are defined above, $\mathbf{R}$ is a 2×2 matrix of residual (co)variances, and $\mathbf{I}$ is a 2×2 identity matrix. The data were properly ordered within the vectors and the vectors $\mathbf{a}$ and $\mathbf{h}$ contained the effects for both traits, individual by individual.

In a Bayesian setting, we assumed

$$p(\mathbf{a}|\mathbf{G}) \sim \mathbf{MVN}(\mathbf{0}, \mathbf{A} \otimes \mathbf{G}), \text{ and}$$

$$p(\mathbf{h}|\mathbf{H}) \sim \mathbf{MVN}(\mathbf{0}, \mathbf{I} \otimes \mathbf{H}),$$

where $\mathbf{G}$ is a 2×2 matrix of additive-genetic (co) variances, $\mathbf{A}$ is the numerator of Wright's relationship matrix between individuals, $\mathbf{H}$ is a 2×2 (co)variance matrix for herd effects, and $\mathbf{I}$ is a 2×2 identity matrix. Flat priors were assumed for effects in $\mathbf{b}$, as well as for $\mathbf{G}$, $\mathbf{H}$, and $\mathbf{R}$.

Gibbs Sampler. Marginal posterior distributions of unknown parameters were estimated by performing numerical integration using the Gibbs sampler (Gelfand and Smith, 1990). This was used to obtain autocorrelated samples from the joint posterior distributions and subsequently from the marginal posterior distributions of all unknowns in the model. The lengths of the chain and of the burn-in period were assessed by visual inspection of trace plots, as well as by the diagnostic tests of Geweke (1992) and Gelman and Rubin (1992). After a preliminary run, we decided to construct a single chain consisting of 850,000 iterations and to discard the first 50,000 iterations as a very conservative burn-in. Subsequently, 1 in every 200 successive samples was retained, to store draws that were more loosely correlated. Thus, 4,000 samples were used to determine the posterior distributions of unknown parameters. The lower and upper bounds of the highest 95% probability density regions for parameters of concern were obtained from the estimated marginal densities. Moreover, the posterior probability for heritability >0.1 was computed using the estimated marginal densities. The posterior median was used as the point for all parameters. Autocorrelations between samples and estimates of Monte Carlo standard errors (Geyer, 1992) were calculated.

RESULTS AND DISCUSSION

Several studies have shown the difficulty in managing and interpreting relationships between the large number of variables used to define milk nutritional and technological quality (Ikonen et al., 2004; Vallas et al., 2010). This fact represents, together with the costs of measuring the traits, a constraint for the large-scale implementation of selection for milk technological quality. Moreover, the analysis of correlated traits with sampling errors that tend to be correlated may add further difficulty in the interpretation of results (Bolormaa et al., 2010). Thus, a reduction in the dimensionality of the system by factor analysis may be of great applicative interest. A single variable or a few variables, well defined in terms of technical and biological meaning, and uncorrelated with each other or with a low level of correlation, may be desirable indicators of the different aspects of milk quality.

In the present study, the marked difference between Pearson and partial correlations (Table 2) indicated that a latent correlation structure exists. Overall, the Kaiser measure of sampling adequacy was equal to 0.62. This value is not too far from the empirical threshold of 0.8, which is the optimal value to consider a data set suitable for factor analysis, and is higher than the lower limit of 0.5 proposed for accepting data for factor analysis (Cerny and Kaiser, 1977).

Four latent common factors with an associated communality equal to 70% of the total original variance of the system were extracted. Rotated factor patterns and communalities of original variables are reported in Table 3. The structure was quite easy to interpret: few variables with large loadings and many variables with small loadings could be observed within each factor, respectively. This result suggested the hypothesis of a simpler structure that controls, at least in part, the covariance of milk quality traits. The partitioning of explained variance between extracted factors was quite balanced, with an expected slight predominance of the factor 1 (eigenvalue = 0.29), whereas the eigenvalues of the other 3 factors ranged between 0.1 and 0.19 (Table 3). This is a peculiarity of factor analysis in comparison with principal components analysis, another multivariate dimension reduction technique. In principal components analysis, the first component is usually associated with a larger amount of variance in comparison with successive variables (Jombart et al., 2009).

The unexplained variance consists largely of random, unpredictable variations and systematic error (Enevoldsen et al., 1996). In the present study, the larger uniqueness (i.e., 1 − communality) was found for traits such as SCC, fat percentage, and curd firming time

Table 2. Pearson (above the diagonal) and partial (under the diagonal) phenotypic correlations among milk composition traits and milk coagulation properties[1]

Item	TS	FP	PP	CAS	LAC	SCC	SH	pH	RCT	k_{20}	a_{30}
TS	*	0.20	0.71	0.77	0.41	−0.12	0.40	−0.10	0.08	−0.09	0.17
FP	−0.01	*	0.36	0.40	−0.15	0.04	0.04	0.05	−0.04	0.02	0.12
PP	0.34	−0.23	*	0.94	−0.09	0.02	0.35	0.02	0.13	0.01	0.08
CAS	0.03	0.34	0.89	*	0.47	−0.04	0.39	−0.01	0.11	−0.05	0.14
LAC	0.54	−0.24	−0.65	0.12	*	−0.30	0.28	−0.21	−0.05	−0.22	0.26
SCC	0.02	−0.001	−0.03	0.03	−0.19	*	−0.16	0.11	0.07	0.06	−0.16
SH	−0.01	−0.09	0.09	0.05	0.12	−0.06	*	−0.51	−0.05	0.02	0.11
pH	−0.02	−0.02	0.02	0.06	−0.02	0.005	−0.51	*	−0.001	0.04	−0.06
RCT	−0.01	−0.07	0.04	0.02	0.05	0.03	−0.07	0.09	*	0.25	−0.32
k_{20}	0.01	0.06	0.05	−0.06	−0.07	−0.04	−0.13	0.13	0.16	*	−0.36
a_{30}	−0.03	0.13	0.04	−0.02	0.13	−0.07	0.28	−0.21	−0.25	−0.26	*

[1]Standard errors were in the ranges of 0.010 to 0.029 and 0.013 to 0.029 for Pearson correlations and partial correlations, respectively. FP = fat percentage; PP = protein percentage; CAS = CN content; LAC = lactose percentage; SH = titratable acidity; RCT = rennet coagulation time; k_{20} = curd firming time; a_{30} = curd firmness.

(Table 3), that is, variables that are known to show a large amount of random variation.

The first latent variable, which explained about 30% of the original variance, was positively associated with TS, fat, protein, and CN contents. It was therefore considered an indicator of milk composition (Composition). The second common factor, explaining about 20% of the variance, was positively associated with renneting and curd firming time and negatively associated with curd firmness, respectively. It was considered a Coagulation factor. Its scores were therefore negatively correlated with the overall technological quality of milk. In Girgentana breed goats, Todaro et al. (2005) found 2 different factors related to MCP, one associated with renneting time and another with curd firmness, respectively. Such a difference could be related to the faming system (extensive in goats vs. semi-intensive in cows), but also to differences in the genetic control of these traits in the 2 species. Finally, it should be remembered that in this study, the curd firming time was not found to be related to any of the extracted factors. Actually, this variable is considered the most problematic among MCP in terms of the feasibility of measurement and the repeatability of results (Bittante, 2011). The third extracted latent variable was associated with measures of acidity. It could be termed Acidity. Finally, the fourth common factor was positively correlated with lactose content and negatively with SCC. Larger scores were therefore associated with higher lactose and lower SCC, that is, with a better health status of the mammary gland, with milk lactose content and SCC being 2 indicators of the integrity of mammary gland cells (Dohoo and Meek, 1982; Hamann and Kromker, 1997). The factor was termed Udder_Health. Similar results were obtained when SCS was used instead of SCC.

Table 3. Rotated factor pattern and variable communality

Item	Factor 1	Factor 2	Factor 3	Factor 4	Communality
Original variable[1]					
TS	0.814[2]	0.006	−0.136	0.378	0.82
FP	0.526[2]	−0.241	0.120	−0.403	0.51
PP	0.940[2]	0.075	−0.080	−0.093	0.90
CAS	0.962[2]	0.009	−0.090	0.047	0.93
LAC	0.074	−0.159	−0.169	0.818[2]	0.73
SCC	−0.011	0.101	0.008	−0.651[2]	0.43
SH	0.347	−0.106	−0.779[2]	0.171	0.77
pH	0.082	0.110	0.900[2]	−0.045	0.83
RCT	0.138	0.743[2]	0.034	0.095	0.58
k_{20}	−0.021	0.667[2]	0.020	−0.208	0.49
a_{30}	0.148	−0.729[2]	−0.193	0.166	0.62
Eigenvalue	0.29	0.19	0.12[2]	0.10	

[1]FP = fat percentage; PP = protein percentage; CAS = CN content; LAC = lactose percentage; SH = titratable acidity; RCT = rennet coagulation time; k_{20} = curd firming time; a_{30} = curd firmness.

[2]Values indicate loadings greater than 0.60 that have been considered to be significant for the interpretation of the factor pattern.

Table 4. Least squares means ± SE of the 4 extracted factors for different levels of parity

Parity	Composition	Coagulation	Acidity	Udder_Health
First	0.00 ± 0.06^{A}	0.022 ± 0.08	-0.245 ± 0.06^{A}	0.535 ± 0.07^{A}
Second	-0.132 ± 0.04^{A}	0.064 ± 0.06	-0.047 ± 0.04^{AB}	0.078 ± 0.05^{B}
Third	-0.206 ± 0.05^{A}	-0.103 ± 0.07	0.065 ± 0.05^{B}	-0.137 ± 0.06^{B}
Fourth	-0.274 ± 0.06^{A}	-0.039 ± 0.08	0.109 ± 0.06^{B}	-0.261 ± 0.08^{BC}
Fifth or greater	-0.656 ± 0.08^{B}	-0.135 ± 0.10	0.162 ± 0.08^{B}	-0.764 ± 0.10^{BCD}

[A–D]Means within columns with different superscripts differ ($P < 0.001$).

The highest communalities were observed for CN and protein contents, whereas the lowest was for SCC (Table 2). Results for protein agreed with the report of Todaro et al. (2005).

Individual scores for Composition were affected by all the main effects included in model [3] ($P < 0.001$). The lowest values were for cows in the fifth or higher parity (Table 4) and were significantly different ($P < 0.001$) from those of younger cows. The pattern of Composition along the lactation (Figure 1, upper left panel), obtained by plotting least squares means of DIM against the month of lactation, was quite similar to the lactation curve for milk composition traits.

Coagulation scores were affected by herd and the DIM interval ($P < 0.001$), but not by parity ($P = 0.27$). The absence of a relationship between lactation number and MCP has been reported in several studies (Lindström et al., 1984; Davoli et al., 1990; Ikonen et al., 1999). The effect of herd was probably related to differences in management and feeding practices. Tyrisevä et al. (2004) reported a slight improvement in MCP as well as in milk, fat, and protein yields associated with an increased frequency of concentrate feeding. However, Ojala et al. (2005) reported that herd had markedly lower effects on the MCP variation than on milk yield traits in dairy cows. The pattern of Coagulation along the lactation was characterized by an increase, even though it was together with relevant variability (Figure 1, upper right panel). Given that this factor is negatively related to the technological properties of milk, the pattern indicates a worsening of MCP properties along the lactation. The variation in MCP across DIM could be related to changes in physical and chemical characteristics of the milk during lactation, particularly macrocomponents, the CN micelle structure, and the salt equilibrium.

The factor Acidity was affected by all the main effects included in model [3] ($P \leq 0.01$). As expected,

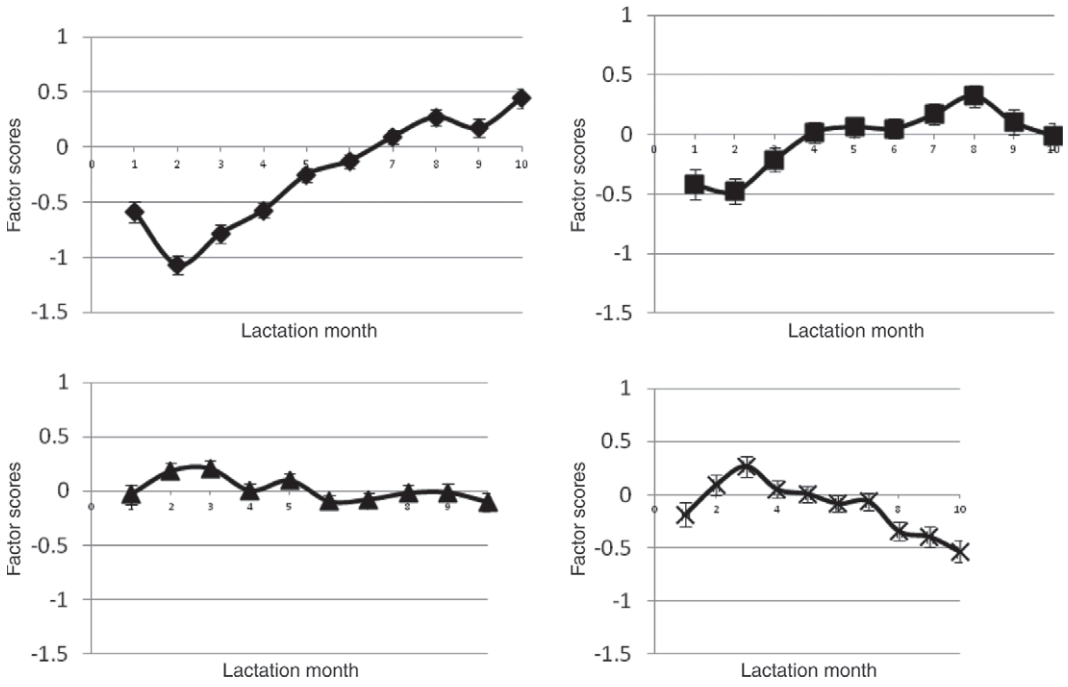

Figure 1. Pattern of the extracted common factor scores along the lactation (upper left: Composition; upper right: Coagulation; lower left: Acidity; lower right: Udder_Health).

Table 5. Features of the marginal posterior distribution of heritability for the investigated traits[1]

	Heritability		
Trait	Median	HPD95	$P(h^2 > 0.1)$
Composition factor	0.203	0.081; 0.398	0.94
Coagulation factor	0.227	0.106; 0.413	0.98
Acidity factor	0.102	0.050; 0.190	0.52
Udder_Health factor	0.137	0.041; 0.288	0.74
Milk yield (kg)	0.085	0.010; 0.231	0.40

[1]Median = median of the marginal posterior density of the parameter; HPD95 = lower and upper bounds of the 95% highest posterior density region; $P(h^2 > 0.1)$ = posterior probability of heritability being greater than 0.1.

given its positive correlation with pH, the average score for Acidity increased significantly ($P < 0.001$) from the first to later parities (Table 4). The pattern along the lactation can be observed in the lower left panel of Figure 1: the values of the scores show a slight increase in the first part of lactation and a subsequent smooth decrease.

Finally, scores for Udder_Health were also affected by all the effects included in model [3] ($P < 0.001$). A clear pattern across parities could be observed (Table 4), with the highest value (indicating the best condition of the mammary gland) for first-calving cows and a progressive decrease for older parities. In addition, the progressive decrease in Udder_Health scores along the lactation, after an initial increase, (Figure 1, lower right panel) was expected. It mimics quite closely the standard shape of the lactation curve for milk yield.

Posterior medians of heritability for the 4 latent factors and milk yield were low to moderate (Table 5). In particular, the heritability for Coagulation was similar to values reported for MCP for the Italian Simmental, Brown Swiss, and Holstein breeds (Bonfatti et al., 2011; Cecchinato et al., 2011), whereas it was lower than those reported by other authors for Estonian Holstein (Vallas et al., 2010) and Finnish Ayrshire cows (Ikonen et al., 2004). Actually, methodologies for MCP assessment sometimes vary markedly across different studies. Estimates for Composition are in agreement with previous reports obtained for fat and protein percentages (Vallas et al., 2010) in data sets of comparable size. The median heritability for Acidity was of the same order of magnitude as values reported for pH and titratable acidity in Italian Holsteins (Cecchinato et al., 2011). The heritability of the Udder_Health factor was similar to values reported for the incidence of mastitis (Appuhamy et al., 2009) and for SCS (Shook and Schutz, 1994).

In addition, genetic correlations between extracted factors and milk yield exhibited low to moderate values (Table 6). The Coagulation factor showed a moderate positive correlation with Acidity, confirming previous reports (Ikonen et al., 2004; Cassandro et al., 2008), and a low and negative correlation with Composition and milk yield. No genetic relationships were detected between Coagulation and Udder_Health. The Composition factor showed a moderate negative correlation with milk yield and Acidity, as expected, and a low relationship with Udder_Health.

When latent factors were extracted, their scores were uncorrelated. The low genetic correlations observed between latent variable scores are a consequence of the rotation procedure adopted, which, from a geometric standpoint, corresponds to a rotation of the axis of the factor space to obtain the pattern of loadings that could be most easily interpreted (Krzanowsky, 2003). In any case, the bounds of the 95% probability density region of the marginal posterior densities of the genetic correlations always included zero. Although this result does not add further elements to the assessment of correlations between different aspects of milk quality, it actually highlights a positive feature of the new variables. The low correlation allows a largely independent selection of different characteristics of milk when extracted latent variables are used as breeding goals in a selection program.

CONCLUSIONS

The use of a few variables with a low degree of relationship between them and a clear technical meaning may represent a valuable option for both management and breeding purposes. Compared with the original traits, latent factors are able to represent some specific aspects of milk quality more efficiently; that is,

Table 6. Estimates of genetic correlations between milk yield, composition, coagulation, health, and acidity factors[1]

Trait	Composition	Coagulation	Acidity	Udder_Health
Milk yield (kg)	−0.297 (−0.84, 0.58)	−0.147 (−0.74, 0.63)	0.345 (−0.39, 0.94)	0.132 (−0.72, 0.89)
Composition		−0.116 (−0.59, 0.52)	−0.281 (−0.71, 0.20)	0.126 (−0.51, 0.68)
Coagulation			0.458 (−0.008, 0.81)	0.008 (−0.57, 0.54)
Acidity				0.168 (−0.43, 0.63)

[1]Estimates are the medians (lower and upper bound of the 95% probability density region in parentheses) of the marginal posterior densities of the genetic correlations.

the number of traits is drastically reduced. It is often easier, in practice, to focus on aggregate indexes instead of evaluating different variables that do not have well-defined estimates of relationships between them. The combination of original traits in factor scores is not carried out by fixing a priori weights but is based on loadings that are derived from the original correlation matrix. Multivariate factor analysis allows calculations to be simplified by using a reduced variable space but without excluding any trait. In general, the extracted variables exhibited a coherent behavior across levels of fixed factors considered in this research, such as parity and lactation stage. As far as the meaning of the extracted variables is concerned, of particular interest are the factors Coagulation and Udder_Health, which are able to represent each complex animal trait with a single value.All new variables exhibited genetic variation, almost in agreement with results for original variables. Thus, their inclusion as breeding goals in selection programs for improving milk quality could be implemented. In any case, relationships with other traits, such as reproduction or type, have to be carefully assessed. The derivation of their economic weights should be based on the milk market scenarios, farming systems, feed supply and costs, and industry goals (Shook, 2006). A reduced number of variables, compared with the use of simple traits, should simplify the calculation of weights. Moreover, the low degree of correlation between factors could provide the advantage of an independent selection for these variables. It could therefore be more feasible and simpler, for example, to place different economic emphasis on milk composition and on its coagulation properties. This may be of great interest for cattle breeds whose milk is primarily processed into typical cheeses. The breeding goals of dairy cattle are in continuous development, and several new phenotypes yielded by the increasing implementation of precision farming may be available in the near future (Boichard and Brochard, 2012). Technologies such as mid-infrared spectroscopy may automate, and therefore reduce, the measurement costs of traits such as MCP and milk FA composition. However, such a huge amount of information will require adequate statistical tools to make inferences about the causes of variation and to study the pattern of relationships. Factor analysis could represent a valuable option in this sense.

ACKNOWLEDGMENTS

The authors acknowledge the Italian Ministry of University and Research for financial support (PRIN2005-prot. 2005074889_002). The authors also gratefully acknowledge the Provincial Breeders Associations of Treviso, Venezia, Vicenza, and Padova and the Italian Brown Swiss Cattle Breeders Association (ANARB) for providing the milk recording and pedigree data, and the laboratory of Veneto Agricoltura (Thiene, Italy) for milk analyses. The support from Trento Province and the Superbrown Consortium of Bolzano and Trento is also acknowledged.

REFERENCES

Appuhamy, J. A. D. R. N., B. G. Cassell, and J. B. Cole. 2009. Phenotypic and genetic relationships of common health disorders with milk and fat yield persistencies from producer-recorded health data and test-day yields. J. Dairy Sci. 92:1785–1795.

Aspilcueta-Borquis, R. R., F. Baldi, F. R. Araujo Neto, and L. G. Albuquerque. 2011. Genetic parameters of total milk yield and factors describing the shape of lactation curve in dairy buffaloes. J. Dairy Res. 79:60–65. http://dx.doi.org/10.1017/S0022029911000823.

Bittante, G. 2011. Modeling rennet coagulation time and curd firmness of milk. J. Dairy Sci. 94:5821–5832.

Boichard, D., and M. Brochard. 2012. New phenotypes for new breeding goals in dairy cattle. Animal 6:544–550.

Bolormaa, S., J. E. Pryce, B. J. Hayes, and M. E. Goddard. 2010. Multivariate analysis of genome-wide association study in cattle. J. Dairy Sci. 93:3818–3833.

Bonfatti, V., A. Cecchinato, L. Gallo, A. Blasco, and P. Carnier. 2011. Genetic analysis of detailed milk protein composition and coagulation properties in Simmental cattle. J. Dairy Sci. 94:5183–5193.

Cassandro, M., A. Comin, M. Ojala, R. Dal Zotto, M. De Marchi, L. Gallo, P. Carnier, and G. Bittante. 2008. Genetic parameters of milk coagulation properties and their relationships with milk yield and quality traits in Italian Holstein cows. J. Dairy Sci. 91:371–376.

Cecchinato, A., M. De Marchi, L. Gallo, G. Bittante, and P. Carnier. 2009. Mid-infrared spectroscopy predictions as indicator traits in breeding programs for enhanced coagulation properties of milk. J. Dairy Sci. 92:5304–5313.

Cecchinato, A., M. Penasa, M. De Marchi, L. Gallo, G. Bittante, and P. Carnier. 2011. Genetic parameters of coagulation properties, milk yield, quality, and acidity estimated using coagulating and noncoagulating milk information in Brown Swiss and Holstein-Friesian cows. J. Dairy Sci. 94:4205–4213.

Cerny, B. A., and H. F. Kaiser. 1977. A study of a measure of sampling adequacy for factor-analytic correlation matrices. Multivariate Behav. Res. 12:43–47.

Dal Zotto, R., M. De Marchi, A. Cecchinato, M. Penasa, M. Cassandro, P. Carnier, L. Gallo, and G. Bittante. 2008. Reproducibility and repeatability of measures of milk coagulation properties and predictive ability of mid-infrared reflectance spectroscopy. J. Dairy Sci. 91:4103–4112.

Davoli, R., S. Dall'Olio, and V. Russo. 1990. Effect of κ-casein genotype on the coagulation properties of milk. J. Anim. Breed. Genet. 107:458–464.

De Marchi, M., C. C. Fagan, C. P. O'Donnell, A. Cecchinato, R. Dal Zotto, M. Cassandro, M. Penasa, and G. Bittante. 2009. Prediction of coagulation properties, titratable acidity, and pH of bovine milk using mid-infrared spectroscopy. J. Dairy Sci. 92:423–432.

Dohoo, I. R., and A. H. Meek. 1982. Somatic cell counts in bovine milk. Can. Vet. J. 23:119–125.

Enevoldsen, C., J. Hindede, and T. Kristensen. 1996. Dairy herd management types assessed from indicators of health, reproduction, replacement, and milk production. J. Dairy Sci. 79:1221–1236.

Gelfand, A., and A. F. M. Smith. 1990. Sampling based approaches to calculating marginal densities. J. Am. Stat. Assoc. 85:398–409.

Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7:457–511.

Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion). Pages 164–193 in Bayesian Statistics. J. O. Berger, J. M. Bernardo, A. P.

Dawid, and A. F. M. Smith, ed. Oxford University Press, Oxford, UK.
Geyer, C. J. 1992. Practical Markov chain Monte Carlo. Stat. Sci. 7:473–483.
Glantz, M., H. Lindmark Månsson, M. Paulsson, and H. Stålhammer. 2011. Genomic selection in relation to bovine milk composition and processability. J. Dairy Res. 79:53–59.
Hamann, J., and V. Kromker. 1997. Potential of specific milk composition variables for cow health management. Livest. Prod. Sci. 48:201–208.
Ikonen, T., K. Ahlfors, R. Kempe, M. Ojala, and O. Ruottinen. 1999. Genetic parameters for the milk coagulation properties and prevalence of noncoagulating milk in Finnish dairy cows. J. Dairy Sci. 82:205–214.
Ikonen, T., S. Morri, A. M. Tyriseva, O. Ruottinen, and M. Ojala. 2004. Genetic and phenotypic correlations between milk coagulation properties, milk production traits, somatic cell count, casein content, and pH of milk. J. Dairy Sci. 87:458–467.
Jombart, T., D. Pontier, and A. B. Dufour. 2009. Genetic markers in the playground of multivariate analysis. Heredity 102:330–341.
Krzanowsky, W. J. 2003. Principles of Multivariate Analysis. Oxford University Press Inc., New York, NY.
Lawley, D. N., and A. E. Maxwell. 1971. Factor Analysis as a Statistical Method. 2nd ed. Butterworths, London, UK.
Lindström, U. B., V. Antila, and J. Syväjärvi. 1984. A note on some genetic and non-genetic factors affecting clotting time of Ayrshire milk. Acta Agric. Scand. 34:349–355.
Macciotta, N. P. P., D. Vicario, and A. Cappio-Borlino. 2006. Use of multivariate analysis to extract latent variables related to level of production and lactation persistency in dairy cattle. J. Dairy Sci. 89:3188–3194.
Macciotta, N. P. P., D. Vicario, C. Dimauro, and A. Cappio-Borlino. 2004. A multivariate approach to modelling shapes of individual lactation curves in cattle. J. Dairy Sci. 87:1092–1098.
McDonald, R. P. 1985. Factor Analysis and Related Methods. Lawrence Erlbaum Associates Publishers, Mahwah, NJ.
Morrison, F. 1976. Multivariate Statistical Methods. McGraw-Hill, New York, NY.
Ojala, M., A. M. Tyrisevä, and T. Ikonen. 2005. Genetic improvement of milk quality traits for cheese production. Pages 307–311 in Indicators of Milk and Beef Quality. J. F. Hocquette and S. Gigli, ed. Wageningen Academic Publishers, Wageningen, the Netherlands.
Rutten, M. J. M., H. Bovenhuis, J. M. L. Heck, and J. A. M. van Arendonk. 2011. Predicting bovine milk protein composition based on Fourier transform infrared spectra. J. Dairy Sci. 94:5683–5690.
SAS Institute. 2008. SAS/STAT Version 9.2. SAS Inst. Inc., Cary, NC.
Schmitt, T. 2011. Current methodological considerations in exploratory and confirmatory factor analysis. J. Psychoed. Assess. 29:304–321.
Shook, G. E. 2006. Major advances in determining appropriate selection goals. J. Dairy Sci. 89:1349–1361.
Shook, G. E., and M. M. Schutz. 1994. Selection on somatic cell score to improve resistance to mastitis in the United States. J. Dairy Sci. 77:648–658.
Sorensen, D., and D. Gianola. 2002. Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics. Springer-Verlag, New York, NY.
Soyeurt, H., P. Dardenne, A. Gillon, C. Croquet, S. Vanderick, P. Mayeres, C. Bertozzi, and N. Gengler. 2006. Variation in fatty acid contents of milk and milk fat within and across breeds. J. Dairy Sci. 89:4858–4865.
Stanton, T. L., L. R. Jones, R. W. Everett, and S. D. Kachman. 1992. Estimating milk, fat, and protein lactation curves with a test day model. J. Dairy Sci. 75:1691–1700.
Todaro, M., M. L. Scatassa, and P. Giaccone. 2005. Multivariate factor analysis of Girgentana goat milk composition. Ital. J. Anim. Sci. 4:403–410.
Tyrisevä, A.-M., T. Vahlsten, O. Ruottinen, and M. Ojala. 2004. Noncoagulation of milk in Finnish Ayrshire and Holstein-Friesian cows and effect of herds on milk coagulation ability. J. Dairy Sci. 87:3958–3966.
Vallas, M., H. Bovenhuis, T. Kaart, K. Pärna, H. Kiiman, and E. Pärna. 2010. Genetic parameters for milk coagulation properties in Estonian Holstein cows. J. Dairy Sci. 93:3789–3796.
Vukasinovic, N., J. Moll, and N. Kiinzi. 1997. Factor analysis for evaluating relationships between herd life and type traits in Swiss Brown cattle. Livest. Prod. Sci. 49:227–234.