



Rear shape in 3 dimensions summarized by principal component analysis is a good predictor of body condition score in Holstein dairy cows

A. Fischer,*†‡ T. Luginbühl,§ L. Delattre,§ J. M. Delouard,§ and P. Faverdin*†¹

*INRA, UMR 1348 PEGASE, F-35590 St-Gilles, France

†Agrocampus-Ouest, UMR1348 PEGASE, F-35000 Rennes, France

‡Institut de l'élevage, F-35650 Le Rheu, France

§DIOuest Lannion, F-22300 Lannion, France

ABSTRACT

Body condition is an indirect estimation of the level of body reserves, and its variation reflects cumulative variation in energy balance. It interacts with reproductive and health performance, which are important to consider in dairy production but not easy to monitor. The commonly used body condition score (BCS) is time consuming, subjective, and not very sensitive. The aim was therefore to develop and validate a method assessing BCS with 3-dimensional (3D) surfaces of the cow's rear. A camera captured 3D shapes 2 m from the floor in a weigh station at the milking parlor exit. The BCS was scored by 3 experts on the same day as 3D imaging. Four anatomical landmarks had to be identified manually on each 3D surface to define a space centered on the cow's rear. A set of 57 3D surfaces from 56 Holstein dairy cows was selected to cover a large BCS range (from 0.5 to 4.75 on a 0 to 5 scale) to calibrate 3D surfaces on BCS. After performing a principal component analysis on this data set, multiple linear regression was fitted on the coordinates of these surfaces in the principal components' space to assess BCS. The validation was performed on 2 external data sets: one with cows used for calibration, but at a different lactation stage, and one with cows not used for calibration. Additionally, 6 cows were scanned once and their surfaces processed 8 times each for repeatability and then these cows were scanned 8 times each the same day for reproducibility. The selected model showed perfect calibration and a good but weaker validation (root mean square error = 0.31 for the data set with cows used for calibration; 0.32 for the data set with cows not used for calibration). Assessing BCS with 3D surfaces was 3 times more repeatable (standard error = 0.075 versus 0.210 for BCS) and 2.8 times more reproducible than manually scored BCS (standard error

= 0.103 versus 0.280 for BCS). The prediction error was similar for both validation data sets, indicating that the method is not less efficient for cows not used for calibration. The major part of reproducibility error incorporates repeatability error. An automation of the anatomical landmarks identification is required, first to allow broadband measures of body condition and second to improve repeatability and consequently reproducibility. Assessing BCS using 3D imaging coupled with principal component analysis appears to be a very promising means of improving precision and feasibility of this trait measurement.

Key words: body condition score, 3-dimensional imaging, principal component analysis, precision livestock farming

INTRODUCTION

Body condition assesses body reserves and is often used as an indirect indicator of reproduction and health status in dairy cattle management. Thin or fat cows are commonly known to be less efficient in reproduction with reduced success at first AI, longer calving-to-calving interval, and earlier return to heat cycles (Dechow et al., 2002; Berry et al., 2003). In the same way, body condition is correlated with health status (Ruegg and Milton, 1995), but the strength of this association depends on the disease (Roche and Berry, 2006). Genetic selection enhances the genetic production potential of the dairy herd but weakens its reproductive and health performance. Improving the reproductive and health status of dairy cows while maintaining production is a central issue in dairy husbandry and justifies an increasing interest in body condition phenotyping (Coffey et al., 2003; Pryce and Harris, 2006).

Major concern for selection is the difficulty in achieving accurate, objective, and high-throughput measurement of body condition in dairy cows. Body reserves can be recorded either directly by measuring the quantity of body lipids after slaughtering animals, or indirectly by measuring traits which are highly cor-

Received October 13, 2014.

Accepted March 29, 2015.

¹Corresponding author: philippe.faverdin@rennes.inra.fr

related with lipid levels. Whole-body dissection is time-consuming, cumbersome, expensive, and irreversible for broadband use (Szabo et al., 1999). Assessing body reserves indirectly has been largely analyzed: methods can be precise but time-consuming, expensive, and invasive, such as measuring adipocyte diameter or diffusion space of deuterated water (Waltner et al., 1994). An imaging technique using ultrasonography offers new perspectives for achieving repeatable and noninvasive measures of body reserves, though they are not high-throughput methods (Schröder and Staufenbiel, 2006). On the farm, body condition is usually based on scoring, visually or by palpation, specific anatomic areas according to a chart. Body condition score appears as the cheapest and most practical method, though it suffers from its subjectivity and low reproducibility for an individual monitoring. Small and rapid variations of body condition occur during the first half of lactation in dairy cows. However, these variations are hardly detected because scores for the same cow vary between scorers (Kristensen et al., 2006) and are not sufficiently reproducible (Pryce et al., 2014).

Imaging technologies have recently become more affordable and their image quality and precision justify potential on-farm application. Therefore, few research groups (Ferguson et al., 2006; Bewley et al., 2008; Halachmi et al., 2008; Negretti et al., 2008; Azzaro et al., 2011; Bercovich et al., 2013; Weber et al., 2014) have attempted to automate BCS to achieve a more objective and less time-consuming method. Directly scoring body condition on 2-dimensional (**2D**) images is as efficient as standard BCS, but is still as subjective and labor consuming as the latter (Ferguson et al., 2006). Subsequently developed indirect methods aimed at reducing time and labor consumption of body condition monitoring. The first step was to build an acquisition system capable of acquiring high-quality images at an affordable price and not too sensitive to environmental changes. The second step was to define the information to be extracted from images to be used to assess BCS. Methods developed by Bewley et al. (2008), Halachmi et al. (2008), and Negretti et al. (2008) did not use whole information kept in 2D images but extracted indicators they assumed to be sensitive to BCS variation, such as angles, areas, or 2D shape of the rear.

Instead of using partial characteristics of the shape and keeping the rear shape in the common 2D space, Azzaro et al. (2011) and Bercovich et al. (2013) dealt with whole information kept in the rear shape. Azzaro et al. (2011) used principal component analysis (**PCA**) and Bercovich et al. (2013) compared partial least square regression (**PLSR**) and Fourier descriptors (**FD**). These 3 methods are efficient tools commonly used in shape processing (Vranic and Saupe, 2001; Allen et al.,

2003; Zion et al., 2007). Bercovich et al. (2013) concluded that the best method was the model predicting BCS linearly from a few FD. The PCA learning method proposed by Azzaro et al. (2011) performed better on external validation than did methods using partial 2D information (Bewley et al., 2008; Halachmi et al., 2008) and PLSR or FD learning methods (Bercovich et al., 2013). The main reason according to Bercovich et al. (2013) was that they could only focus on the tailhead area, whereas the hooks are important too (Edmonson et al., 1989). These results reflect that it is important to focus on the area going from the hook bones to the pin bones and to work with whole information previously compressed with factor extraction techniques (PCA, PLSR, and FD) rather than using partial indicators.

Dealing with 2D images implies a loss of information that is kept in the third dimension. More recent work assessing body condition with 3D surfaces showed a level of calibration similar to the best calibration observed with 2D methods (Weber et al., 2014).

To enhance the prediction quality of an assumed shape-correlated indicator, using 3D appeared more relevant than using 2D because 3D depicts the most complete information available to analyze the shape's variability. The idea in this project was to work with whole information available to depict a 3D surface to identify the traits of the variation in shapes, which are associated with body condition variability. Therefore, the present study aimed at working closely with imaging experts from 3DOuest (Lannion, France) to develop a method combining the use of 3D shapes of the rear and the reduction of the number of 3D variables using PCA to assess BCS with greater objectivity and higher precision. Moreover, because only a few studies analyzed their method validation, we assessed external validation, repeatability, and reproducibility of the method.

MATERIALS AND METHODS

Experimental System Overview

Data. Data were collected at the INRA-UMR PEGASE experimental dairy station in Méjusse, France, between March and July 2013. Cows are milked twice a day and weighed individually and automatically at the milking parlor's exit on a weighing static station (DeLaval France, Elancourt, France).

Surface Acquisition System. The 3D acquisition system was an Xtion PRO Live Motion Sensor (ASUSTek Computer Inc., Taiwan). Ninety pictures are captured in 3 s and stacked to build a 3D surface. The sensor was attached 2 m up from the soil level at weigh station entry and connected to a mechanical

sensor detecting the opening and closing of entry doors. The door opening reset the camera to its initial position, and the door closing automatically started the scan. The reference area was located between the top of the hook bones and the pin bones. Indeed, most of the BCS charts focus their scoring on 2 areas: the flank and the area between the hook bones and pin bones. Edmonson et al. (1989) observed that overall BCS was reliably correlated with hooks and pins prominence, with the depression between the hook bones and with the depression between the hook bones and the pin bones. The 3D surfaces were automatically saved to a computer with date and hour of scan.

Manual Body Condition Scoring. Cows were scored for body condition once per month, on the same day as 3D scans, by 3 technicians using the French BCS chart defined by Bazin et al. (1984). This system involves palpation of the tail head and of the last rib. Scores were scaled every 0.25 points, ranging from 0 for a thin cow to 5 for a fat cow. Each technician assigned one BCS resulting from the mean of the BCS per area, and the mean of the BCS assigned by the 3 technicians was taken as the final BCS.

Methodology: From 3D Surface to BCS Estimation

The method involves 2 main steps: first, the surfaces are normalized using several transformations to render them comparable, followed by a calibration step that adjusts and selects the best equation to assess BCS from the 3D information set previously summarized by PCA (Figure 1).

3D Surface Normalization: Defining Shared Information to Be Analyzed. Because position in the weigh station varies for each cow, the scanned anatomical area varied between acquisitions. The first step consisted of extracting common 3D information to ensure that the information being analyzed referred exactly to the same anatomical area for each 3D surface and to the same number of 3D points per 3D surface, independent of the size of the cow. The idea was to align the surfaces and then to superimpose them to find the surface common to the set of 3D surfaces used for the calibration (Figure 2). To assess BCS independently of the anatomical size of the cow, the 3D surfaces were standardized on a common rear size.

The 3D surfaces were aligned to compensate for location differences in the weigh station. To align the 3D surfaces, the coordinate space, which is originally centered on the camera, was transformed in a space centered on the cow's rear, called rear-centered space. Therefore, 4 anatomical landmarks were manually identified: top of the left hook bone (**HBL**) and top of the right hook bone (**HBR**), as well as with 2 points at the

base of the sacrum, one on the left side (**SBL**) and one on the right side (**SBR**). These 4 landmarks were used to define the X- and Y-axes of the rear-centered space, and the Z-axis was defined as orthogonal to the X- and Y-axes (Figure 3). The 3D surfaces were then aligned by superimposing the rear-centered space defined for each 3D surface.

The 3D surfaces were standardized on a common rear size to delete any variability in 3D surfaces that could be related to size and not to body condition. The mean of the coordinates was estimated for each landmark in the calibration set. The 3D surfaces used for the calibration were resized such that the 4 landmarks had these calculated means as coordinates (Figure 3).

From this point on, 3D surfaces were standardized on a common rear size, though they did not necessarily have the same number of 3D points. To analyze comparable information between 3D surfaces, the aim was to extract the set of 3D points shared by the set of 3D surfaces used for the calibration. These 3D surfaces were orthogonally projected onto the X-Y plane of the rear-centered space. The intersection of these projections defines the area shared by the whole calibration set. This intersection, called the mask, was then split into a 150×150 grid (i.e., 22,500 pixels). Each 3D surface was orthogonally projected onto this mask. Because a 3D surface contains more than 22,500 points, each pixel refers to several 3D point projections. To ensure having exactly the same number of points per pixel, only one 3D point per pixel was kept: the one with the highest Z coordinate, and the others were removed from the pixel.

Calibration Step: Assessing BCS from 3D Information. After normalization, the next step consisted in summarizing the variability of 3D surfaces by performing PCA. The calibration set was defined with 3D surfaces characterized at the most with 22,500 pixels described with 3 coordinates each; pixels without any 3D-points were not kept in the data. Each 3D surface was the result of a combination of 67,500 descriptors. A PCA was performed to define a space characterizing as much variability of the data set as possible with the least number of dimensions. In this way, 3D surfaces used for calibration were used in a PCA as the statistical individuals and their 67,500 3D references as the variables. Each 3D surface was projected onto this PCA space and became a linear combination of its coordinates in this PCA space.

The last step of the calibration process consisted of predicting BCS using the summarized information of the 3D shape by PCA. The calibration aimed at mathematically defining the link between the manually scored BCS and the coordinates of the 3D surface in the PCA space. The coordinates on the eigenvectors

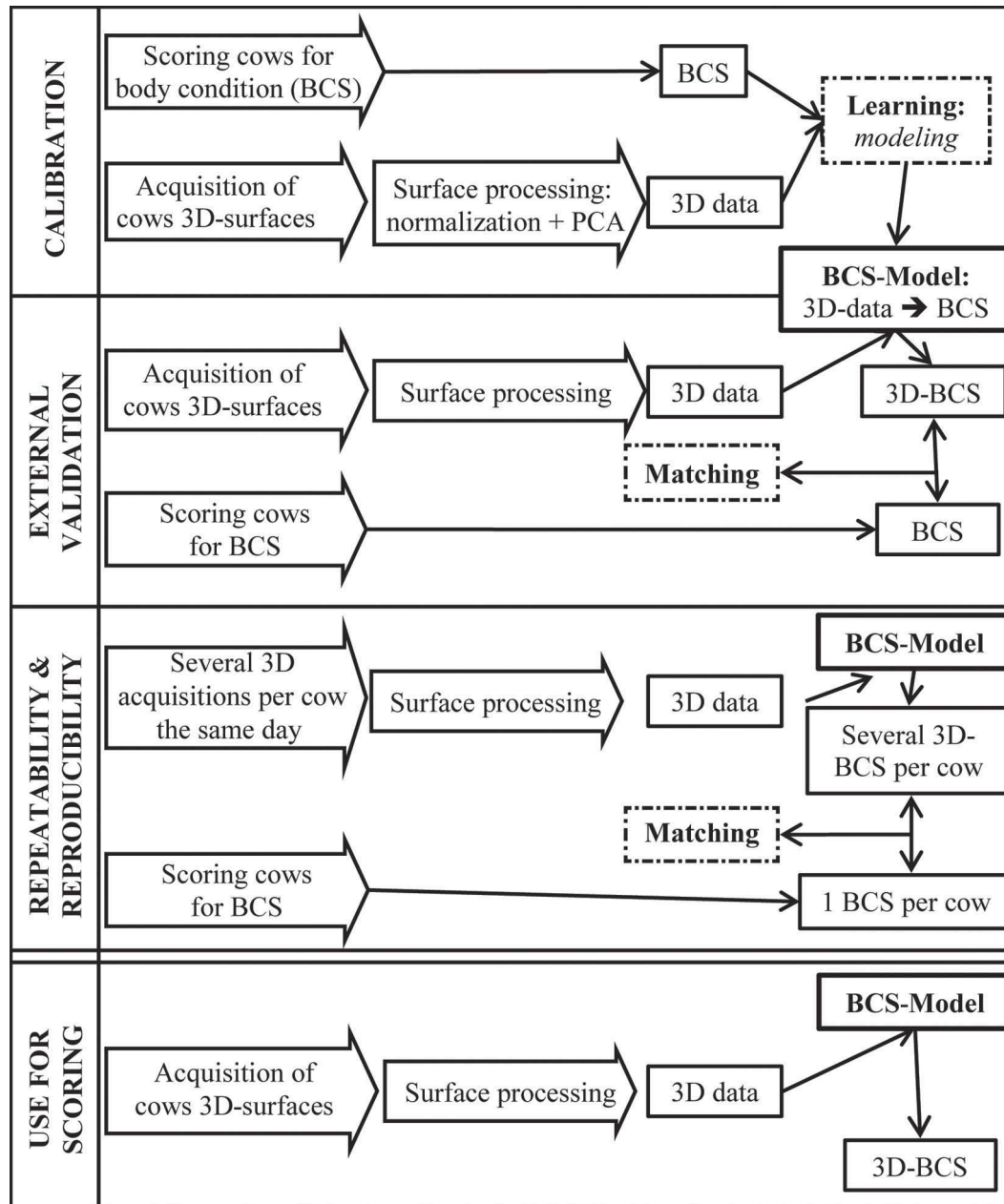


Figure 1. Methodology to estimate BCS with 3-dimensional surfaces: calibration, external validation, repeatability, and reproducibility and use for scoring.

were regressed on BCS with a multiple linear regression with the `lm` function of R (R Core Team, 2013). This model is defined as follows:

$$\text{BCS}_i = \alpha + \beta_1 \times \text{eig}_{i1} + \beta_2 \times \text{eig}_{i2} + \dots + \beta_k \times \text{eig}_{ik} + \dots + \beta_n \times \text{eig}_{in} + \varepsilon_i \quad [1]$$

where BCS_i is the BCS estimated by technicians of the i th cow; eig_{ik} is the coordinate of the 3D surface as-

sociated with cow i on the k th principal component of the PCA; n is the maximum number of principal components allowed in the model according to degrees of freedom; α , β_1 , β_2 , \dots , β_n are the regression parameters, and ε_i is the residual error. The best model was selected by stepwise regression thanks to the step function in R. The stepwise regression initiates with the null model and adds at each step the more explanatory variable at 5% significance and deletes the variable that was previ-

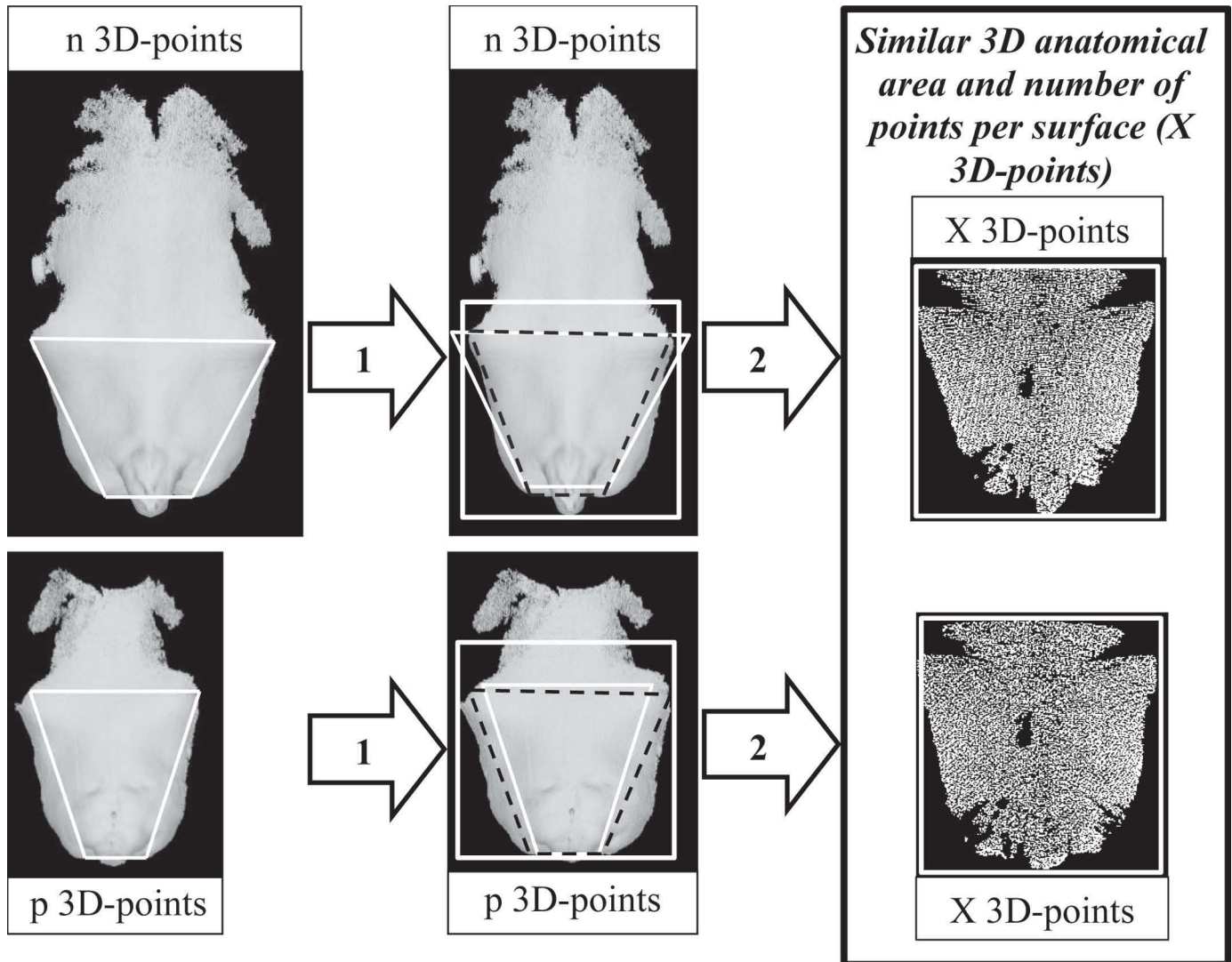


Figure 2. Three-dimensional (3D) surface processing to extract the 3D information common to the set of 3D surfaces used for the calibration: example of 2 3D surfaces, one initially composed of n 3D points and one with p 3D-points ($n > p$). The first step (1) rescales the individual format of each 3D surface (white trapezoid) on a common rear format (dashed black trapezoid), and second step (2) superimposes the 3D surfaces used for the calibration to find the set of 3D points common to the 3D surfaces used for the calibration (white rectangle with X common 3D points; $n > p > X$).

ously added if this variable is not as explanatory as the freshly added variable. The best model is the one with the least variables and the smallest Akaike information criterion. To look over model robustness, the quality of the model selected by stepwise and those built at each step of the stepwise were compared according to their adjusted coefficient of determination (R^2) and the error of prediction when performed on external data sets. To assess the BCS of a 3D surface out of calibration set, the normalized surface was projected onto the mask, then onto the PCA space, and its coordinates on the PCA space were replaced in the model.

Procedure to Validate the Method

To develop a method for a high-throughput use, it is important to characterize the properties of this new technology in terms of validation and repeatability. Figure 1 describes the steps used to validate the method.

External Validation Procedure. Validation qualifies the capacity of the selected model to accurately estimate the predicted variable (here, BCS scored by 3 technicians) when testing the model on individuals not used for calibration. Estimated BCS is called 3D BCS. The mean standard error of prediction (**MSEP**) was

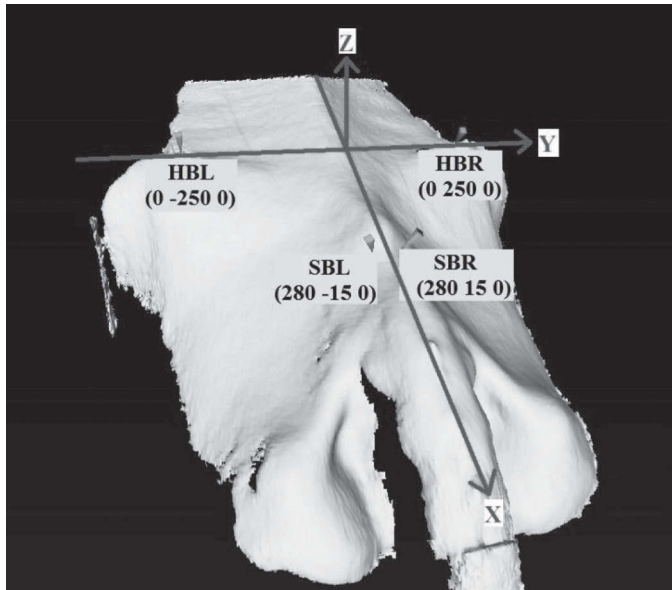


Figure 3. Localization of the 4 anatomical landmarks hook bone left (HBL), hook bone right (HBR), sacrum basis left (SBL), and sacrum basis right (SBR) used to normalize the 3-dimensional surfaces and their coordinates after size standardization.

used to qualify validation and was defined according to Wallach (2006) as the sum of the following 3 errors to better understand MSEP variability: the biased squared error, defined as the squared difference between average measured BCS and average estimated BCS; the slope error, depending on how closely correlated the slope of the regression of BCS on 3D BCS was to 1; and the unexplained error, depending on how variations in BCS and 3D BCS were correlated. Validation reliability was tested on 2 sets of cows: on the one hand on 3D surfaces from cows selected for calibration, but with different stage in lactation (**validationset_stage** data set), and on the other hand on 3D surfaces from cows not used for calibration (**validationset_cows** data set).

Repeatability and Reproducibility. Repeatability assesses the error generated when estimating an indicator several times on the same sample with the same methodology in the same environment in a short period of time. Reproducibility assesses the same error but under variable environmental conditions. In this way, repeatability was estimated by extracting the 4 landmarks 8 times on the same day, from the same 3D scan, and reproducibility was estimated with cows scanned 8 times each on the same day, with the 4 landmarks extracted once per 3D surface. The 3D BCS variation was corrected for the effect of the chosen cows in extracting the residuals of the model of ANOVA, which explains 3D BCS with the cow's identity factor. Coefficients of

variation for repeatability (CV_r) and for reproducibility (CV_R) were assessed:

$$CV_r = \frac{\sigma_r}{M_r} \times 100, \quad CV_R = \frac{\sigma_R}{M_R} \times 100, \quad [2]$$

where σ_r and σ_R are SD of the corrected 3D BCS, respectively, for the repeatability and reproducibility data sets; and M_r and M_R are, respectively, the mean 3D BCS in the repeatability and reproducibility data sets. In the same way the repeatability and reproducibility of BCS were estimated by correcting BCS variability from the effect of the chosen cows and from the effect of the experts dealing with body condition scoring: 1 model is an ANOVA including "cow's identity" as a factor and 1 ANOVA model includes cow's identity and BCS expert's identity as factors to explain BCS for repeatability and reproducibility, respectively. The more repeatable (reproducible) the 3D BCS is, the smaller the CV_r (CV_R).

Data Sets Used. Once per month from March to July 2013, all lactating cows were scanned on the same day as body condition scoring at evening milking. Dry cows with extremely high BCS were scanned twice within 3 mo and added to the sample to build a calibration data set representing a wide range of animal factors including BCS, BW, parity, and DIM. All data sets were built from this perspective: 57 3D surfaces associated with 56 cows (1 cow used twice at a different stage in lactation) built a calibration data set with BCS ranging from 0.5 to 4.8 and a mean of 2.2, and included all BCS extreme cows. Validation data sets included 25 3D surfaces each, with BCS between 0.8 and 4.8 and a mean of 2.3 for validationset_stage, and between 0.8 and 2.9 with a mean BCS of 1.8 for validationset_cows. Method repeatability and reproducibility were analyzed with 6 cows associated with 6 BCS between 1.1 and 4.8 (Figure 4).

RESULTS

A Functional Acquisition System Without Major Problems

The same prototype was used throughout the whole data collection period and gathered 1,433 scans from March to July 2013. One acquisition and its recording lasted 4 s. Despite the high performance of the 3D acquisition system, an average of 20% of 3D surfaces per acquisition series were rejected due to large holes and intermittent areas probably associated with cow movement while scanning.

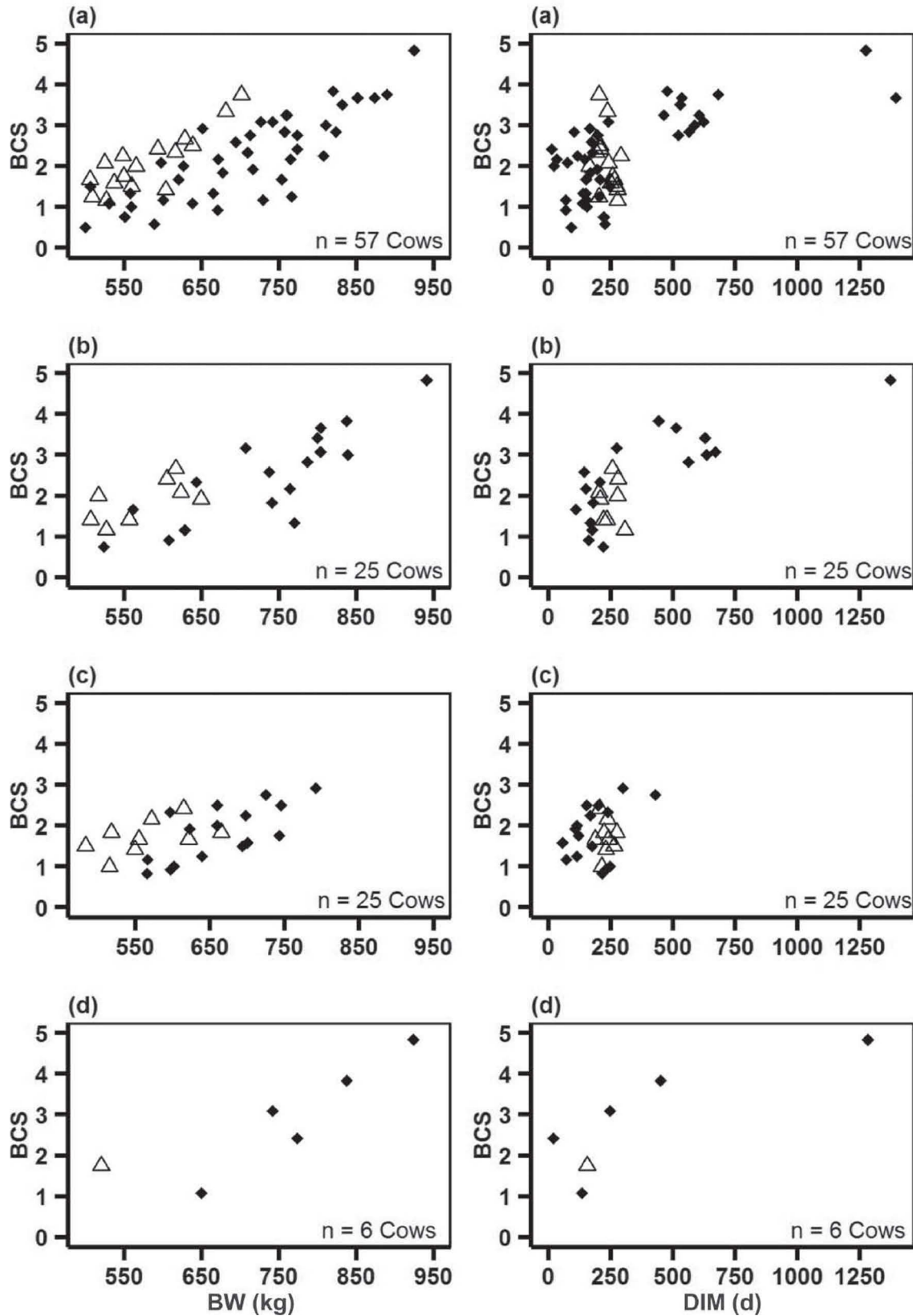


Figure 4. Description of data sets used for (a) the calibration, (b) the validation for validationset_stage and (c) the validation for validationset_cows and (d) the repeatability-reproducibility according to their BCS, BW, and DIM for primiparous (Δ) and multiparous cows (\blacklozenge).

The final data set used to develop the method contained 180 manually selected 3D surfaces and involved 81 cows: 56 cows used for the calibration and 25 used at a different lactation stage for the validationset_stage data set, and 25 other cows used for the validationset_cows data set; 6 of these 81 cows were used to estimate repeatability and reproducibility.

The Most Complete Model Provides a Perfect Calibration

The 57 3D surfaces defining the calibration data set were used in a PCA to compute the eigenvectors describing those surfaces. Their coordinates on the eigenvectors were regressed linearly on their corresponding BCS. The PCA resulted in 57 eigenvectors that were a combination of the set of 3D points referenced in each 3D surface before PCA computing.

The best multiple linear regression model selected by stepwise assessed BCS with 55 coordinates, associated with 55 eigenvectors. This model was able to explain almost 100% of BCS variability ($R^2 = 1$) in the calibration set, with an error close to 0.

Considering that the model had 56 degrees of freedom, a maximum of 55 variables could have been selected. Therefore the most complete model was also the best model selected by stepwise. The selection discarded the coordinates associated with the 57th and the 38th eigenvectors. Each principal component is described by its inertia, which is the percentage of the variability of the 3D surfaces it explains. The PCA orders eigenvectors according to inertia, considering that the first is the eigenvector with the highest inertia. Eigenvector ranking differed from their order of appearance in the selected linear model. However, the first 2 eigenvectors in PCA were also the first 2 eigenvectors selected by stepwise. They explained 68.5% of 3D surface variability and the model including these 2 eigenvectors explained 71% of the variability of the BCS. In the end the 13 first eigenvectors were useful to describe 95% of 3D surface variability, whereas 19 had to be implemented in the model by stepwise to describe 95% of BCS variability.

External Validation Shows Similar Results for Both Validation Sets

The BCS was estimated on 2 data sets: validationset_stage, which contained cows previously used for calibration, but at different stages in lactation and validationset_cows, which contained new cows. Predicted BCS were strongly correlated with manually scored BCS when using the model on both validation

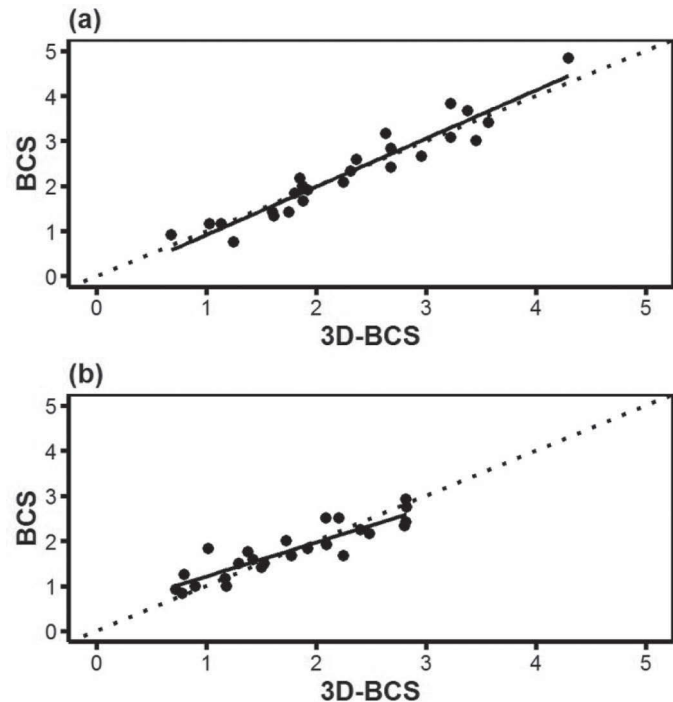


Figure 5. Validation of 3D BCS for both data sets (a) validationset_stage and (b) validationset_cows regarding manual BCS. The line of best fit is the solid line and the line of perfect concordance (slope of 1 and intercept of 0) is the dashed line.

sets ($r = 0.89$ for validationset_cows and 0.96 for validationset_stage, $P\text{-value} < 10^{-16}$, Figure 5). The MSEP for both data sets were 0.09 and 0.10 point of BCS for validationset_stage and validationset_cows, respectively. The MSEP was mostly unexplained with a higher rate being unexplained for validationset_stage compared with validationset_cows: 95.4 and 70.3% of MSEP variability were unexplained, respectively; the remaining 29% were explained by slope error.

Running the set of models, selected at each step throughout the stepwise, on validation sets highlighted that prediction error decreased as stepwise continued. The MSEP was minimal at the 39th and 32nd step for validationset_stage and validationset_cows, respectively. The model selected by stepwise is not the one with the highest validation quality.

A Repeatable and Reproducible Method

The particularity of this step was to assess error associated with the method independently of error associated with the reference method, in this case BCS. Processing each 3D surface 8 times meant that the 8 processes had the same BCS but that landmarking was repeated 8 times. Similarly, scanning each cow 8 times

the same day meant that the error associated with BCS was the same for each scan, but the error associated with acquisition and processing differed for each scan. In this way, reproducibility error included repeatability error in its variability. The σ_r was equal to 0.075 ($CV_r = 2.91\%$) and σ_R was equal to 0.103 ($CV_R = 3.95\%$) BCS point. The method's acquisition system and surface processing are associated with an error characterizing 4% of the variation of assessed BCS. Ultimately, BCS assessed with 3D surfaces was 2.8 times more repeatable and reproducible than manually scored BCS (Table 1).

The step-by-step evolution of σ_r during stepwise selection was curvilinear with a minimum at the 28th step with σ_r of 0.068.

DISCUSSION

Several studies have investigated development of new methods to automatically assess the BCS of dairy cows with greater accuracy using 2D (Bewley et al., 2008; Halachmi et al., 2008, 2013; Negretti et al., 2008; Azzaro et al., 2011; Bercovich et al., 2013) or 3D information (Weber et al., 2014) relating to the surface of the cow's rear. The method proposes a new development in the use of pictures to assess BCS learning 3D-information with PCA.

An Efficient Acquisition System

Even though one more dimension was sampled, the developed acquisition system was as efficient as previous systems developed to collect 2D surfaces: the 3D camera was neither light sensitive, contrary to the camera used by Bewley et al. (2008), nor sensitive to changes in ambient humidity, contrary to the camera used by Azzaro et al. (2011). Consequently, fewer images had to be removed: 80% of the 3D surfaces were of good quality, whereas for Bewley et al. (2008) and Azzaro et al. (2011) only 23 and 2.75% of the entire pictures set were of sufficient quality, respectively.

Landmark Extraction Automation: An Imperative Step Before Monitoring

Identification of anatomical landmarks is essential in image processing because it defines a shape-centered space, which is the basis for subsequent analysis. This step was performed manually to assess the proof of concept and, therefore, was the most time-consuming step with on average 3 min spent per 3D surface. In comparison, Bewley et al. (2008) and Azzaro et al.

Table 1. Repeatability and reproducibility of the measures of body condition scored by experts and of BCS assessed by the 3-dimensional (3D) method¹

Method	σ_r	σ_R	CV_r	CV_R
3D BCS ²	0.075	0.103	2.91	3.95
BCS	0.210	0.280	10.05	13.44

¹ σ_r = standard deviation of repeatability; σ_R = standard deviation of reproducibility; CV_r = coefficient of variation of repeatability; CV_R = coefficient of variation of reproducibility.

²3D BCS = BCS assessed with the 3D method.

(2011) identified 23 anatomical points to describe the shape. Dividing the number of extracted landmarks by 8 significantly reduces the time consumed during image processing, but also facilitates the potential automation of the landmark extraction procedure. Nevertheless, extracting the 4 landmarks manually is too time and labor consuming to cope with high-throughput use on farm. Moreover, the results showed that method reproducibility was limited by method repeatability. In fact method repeatability is defined by the repeatability of landmark extraction. The subjectivity of manual extraction limits the repeatability of the method. To improve 3D BCS repeatability and consequently 3D BCS reproducibility, a solution would be to automate landmark extraction. Moreover, the high frequency of measurements enabled by automation will statistically increase the precision of 3D BCS estimates.

A Promising Methodology: 3D Rear Information Reflects BCS Variability Well

The model developed to assess BCS adjusted perfectly on observed BCS. With an R^2 equal to 1 and a SD equal to 0, the model explained the whole BCS variability in calibration set. The selected model contained 55 variables, which is the maximum the model can contain in light of the degrees of freedom. The equation containing 55 variables showed the best results for the calibration set, but the changes to the SD of the model when predicting BCS on validation sets during the stepwise was minimal around the 32nd step. We assume that the selected model probably includes coordinates associated with eigenvectors, which may explain specific shapes in the calibration set, without any generic meaning. This over-parameterization of the model could explain that external validation shows a smaller error when using a model with fewer variables. An influence analysis of the 3D surfaces on the parameters would analyze the relevance of each parameter in the model and a ridge regression would analyze the robustness of the model.

Good Validation and Reproducibility Limited by BCS Reproducibility

The correlation between 3D predicted BCS and observed BCS ($r = 0.89$ – 0.96) was similar to that observed by Halachmi et al. (2013; $r = 0.94$), who worked with 2D images of the cow's rear. Nevertheless, this comparison is not scientifically valid because experimental designs (BCS chart, BCS scorers) differed between both studies; the different methods should be used under the same conditions.

Both validation sets, validationset_stage and validationset_cows, displayed similar errors. Consequently, we assume that the model is no more sensitive to surfaces associated with cows used at a different lactation stage in the calibration set than surfaces never used before in method development. Azzaro et al. (2011) developed a method based on polynomial kernel PCA to learn 2D shapes and used linear regression of the eigenvalues to assess BCS. These researchers used the methods of Halachmi et al. (2008) and Bewley et al. (2008) and compared the validation of the 3 methods. Models developed by Bewley et al. (2008) and Azzaro et al. (2011) demonstrated the best validation (SD = 0.33 and 0.31 unit of BCS, respectively). Bercovich et al. (2013) used Fourier descriptors combined with a linear regression to assess BCS from the 2D rear contour. Their model validation was estimated with the RMSEP calculation and was worse than the current model: RMSEP = 0.34 and 0.31–0.32 units of BCS, respectively. To correctly compare the validation of these various methods, we should test them in the same experimental conditions and calculate the same indicators. Nevertheless, we may highlight here that validation errors are similar to the reproducibility error of the reference method, which was 0.28 BCS units in our study. The error generated when assessing BCS with 3D surfaces naturally includes the error of the reference method, here BCS. Calibrating 3D surfaces on a more precise and repeatable measure such as ultrasonography (Schröder and Staufenbiel, 2006) should reduce the body energy reserve measurement error. Weber et al. (2014) calibrated a set of distances, depths, and volumes extracted from 3D surfaces of the rear of dairy cows to assess backfat thickness measured by ultrasonography. Their method calibrated well with backfat thickness ($r = 0.96$), but neither validation nor repeatability was evaluated.

The method developed was 2.8 times more repeatable and reproducible than manually scored BCS. When assessing BCS with 3D surfaces, the measurement includes an error of 0.1 BCS units, independent of the error due to the reference method used. The majority of reproducibility error seemed to overlap with

repeatability error. Repeatability error translates the error associated with manual landmark extraction. In this way, the next step would be to automate landmark extraction to completely eliminate human subjectivity.

In conclusion, the model performed well for calibration and validation. Nevertheless, one essential step is the automation of landmark extraction, which is underway: similar to Bercovich et al. (2013), an algorithm extracts automatically the 4 landmarks in identifying the local extreme points. After automation validation, the next step will be to validate the model on more dairy herds before using the method on a large scale. In practice, other experimental farms and later commercial farms will be equipped with the prototype to collect more 3D surfaces and validate the model on those surfaces.

Nonbiased Learning of Global 3D Information with PCA: An Evolving Approach of Imaging

Principal component analysis is more flexible than using predetermined criteria thanks to its learning capacity. In fact, restricting trait analysis to a small number of specific criteria is risky as a first approach because these criteria are not necessarily the best ones for predicting this trait. The space defined by PCA allows as much variability kept in 3D surfaces as possible to be expressed, with the least number of dimensions. In this way, PCA-based learning on the global information contained in a shape should minimize such a bias.

Azzaro et al. (2011) and Bercovich et al. (2013) used factor extraction tools too: PCA, PLSR, and FD, respectively. Their average error rate on external validation was 0.31 unit of BCS for Azzaro et al. (2011) and 0.34 unit of BCS for Bercovich et al. (2013). The model developed in the current work performed as good as the model developed by Azzaro et al. (2011). The rear size of the 3D shapes has been standardized on a common size to ensure having variables with exactly the same definition on each 3D surface and to avoid problems due to body size variability as Bercovich et al. (2013) observed with PLSR method.

In terms of reproducibility, Bercovich et al. (2013) had an average error of 0.11 unit of BCS with the FD method, similar to our model error of 0.10 unit of BCS. Nevertheless, the BCS range they used to assess reproducibility was smaller than ours: from 2.2 to 3.6 (1 to 5 scale), from 1.1 to 4.8 (0 to 5 scale), respectively. It is important to characterize the method on the whole scale, even for the extremes. This comparison of studies shows that methods based on global shape analysis summarized by a statistical tool (PCA, PLSR, or FD) are the most efficient, whatever the statistical tool.

Moreover, as our approach focuses on the whole information contained in the shape and not on a local characteristic, this method could be easily adapted to other breeds or to other body condition charts, even if the more meaningful criteria differ. Method adaptation needs first to define a calibration set reflecting BCS variability as closely as possible, then to define the PCA space with this new calibration set, and finally to assess the coefficients of the model predicting BCS from the coordinates of the eigenvectors. Such a method is not only adaptable to several body condition charts, but also, thanks to its impartiality, should calibrate well on other traits if the anatomical region is meaningful. All these evolutions can be carried out very rapidly, without any change in the software.

CONCLUSIONS

The cow's rear was described in 3D in the PCA space defined by the calibration data set. The PCA space was seen as the space expressing as much variability as available in 3D surfaces. The goal was to develop an automated system able to assess the BCS of dairy cows with more objectivity and lower costs than manually scored BCS and to use it on a large scale. Validation analysis showed good performance and consolidated the developed approach. For the moment, only 4 anatomical landmarks are manually extracted. Automation is underway and is essential to enable the method to be used on a large scale. A further improvement would be to calibrate the method on a more precise and objective method, such as ultrasonography, to reduce the error associated with the reference method. The efficacy of this methodology confirms that the use of 3D surfaces is a promising tool for phenotyping body condition, and shows that it will be possible to use 3D surfaces to assess several phenotypes, such as BW, which for the moment are not easy to measure accurately on a low budget.

ACKNOWLEDGMENTS

The authors express their gratitude to the staff of the Méjusseaux experimental farm, and more particularly thanks are due to J. Lassalas, M. Fargetton, G. Théaud, A. Cozien, M. Guilloux, P. Pichot, and J. Esplan for their involvement and efforts in collecting and transmitting BCS, BW, and herd information. Thanks are also due to the Join Technical Unit of Research and Ingeneering on dairy farming (UMT RIEL, Rennes, France) for the constructive and instructive exchanges. The authors thank the French national institute for agricultural research (INRA), Division of Animal Physiol-

ogy and Livestock Systems (PHASE Nouzilly, France) division for its financing.

REFERENCES

- Allen, B., B. Curless, and Z. Popovic. 2003. The space of human body shapes: Reconstruction and parameterization from range scans. *ACM Trans. Graph.* 22:587–594.
- Azzaro, G., M. Caccamo, J. D. Ferguson, S. Battiato, G. M. Farinella, G. C. Guarnera, G. Puglisi, R. Petriglieri, and G. Licitra. 2011. Objective estimation of body condition score by modeling cow body shape from digital images. *J. Dairy Sci.* 94:2126–2137. <http://dx.doi.org/10.3168/jds.2010-3467>.
- Bazin, S., P. Augeard, M. Carteau, H. Champion, Y. Chilliard, G. Cuyllé, C. Disenhaus, G. Durand, R. Espinasse, A. Gascoin, M. Godineau, D. Jouanne, O. Ollivier, and B. Remond. 1984. Grille de notation de l'état d'engraissement des vaches pie-noires. Institut Technique de l'Elevage Bovin, Paris, France.
- Bercovich, A., Y. Edan, V. Alchanatis, U. Moallem, Y. Parmet, H. Honig, E. Maltz, A. Antler, and I. Halachmi. 2013. Development of an automatic cow body condition scoring using body shape signature and Fourier descriptors. *J. Dairy Sci.* 96:8047–8059. <http://dx.doi.org/10.3168/jds.2013-6568>.
- Berry, D. P., F. Buckley, P. Dillon, R. D. Evans, M. Rath, and R. F. Veerkamp. 2003. Genetic relationships among body condition score, body weight, milk yield, and fertility in dairy cows. *J. Dairy Sci.* 86:2193–2204.
- Bewley, J. M., A. M. Peacock, O. Lewis, R. E. Boyce, D. J. Roberts, M. P. Coffey, S. J. Kenyon, and M. M. Schutz. 2008. Potential for estimation of body condition scores in dairy cattle from digital images. *J. Dairy Sci.* 91:3439–3453. <http://dx.doi.org/10.3168/jds.2007-0836>.
- Coffey, M. P., G. Simm, W. G. Hill, and S. Brotherstone. 2003. Selection for energy balance in dairy cows. *Interbull Bull.* 31:107. Accessed Apr. 23, 2015. <https://journal.interbull.org/index.php/ib/article/download/808/799>.
- Dechow, C. D., G. W. Rogers, and J. S. Clay. 2002. Heritability and correlations among body condition score loss, body condition score, production and reproductive performance. *J. Dairy Sci.* 85:3062–3070.
- Edmonson, A. J., I. J. Lean, L. D. Weaver, T. Farver, and G. Webster. 1989. A body condition scoring chart for Holstein dairy cows. *J. Dairy Sci.* 72:68–78.
- Ferguson, J. D., G. Azzaro, and G. Licitra. 2006. Body condition assessment using digital images. *J. Dairy Sci.* 89:3833–3841.
- Halachmi, I., M. Klopčic, P. Polak, D. J. Roberts, and J. M. Bewley. 2013. Automatic assessment of dairy cattle body condition score using thermal imaging. *Comput. Electron. Agric.* 99:35–40. <http://dx.doi.org/10.1016/j.compag.2013.08.012>.
- Halachmi, I., P. Polak, D. J. Roberts, and M. Klopčic. 2008. Cow body shape and automation of condition scoring. *J. Dairy Sci.* 91:4444–4451. <http://dx.doi.org/10.3168/jds.2007-0785>.
- Kristensen, E., L. Dueholm, D. Vink, J. E. Andersen, E. B. Jakobsen, S. Illum-Nielsen, F. A. Petersen, and C. Enevoldsen. 2006. Within- and across-person uniformity of body condition scoring in Danish Holstein cattle. *J. Dairy Sci.* 89:3721–3728.
- Negretti, P., G. Bianconi, S. Bartocci, S. Terramocchia, and M. Verna. 2008. Determination of live weight and body condition score in lactating Mediterranean buffalo by visual image analysis. *Livest. Sci.* 113:1–7. <http://dx.doi.org/10.1016/j.livsci.2007.05.018>.
- Pryce, J. E., and B. L. Harris. 2006. Genetics of body condition score in New Zealand dairy cows. *J. Dairy Sci.* 89:4424–4432.
- Pryce, J. E., W. J. Wales, Y. de Haas, R. F. Veerkamp, and B. J. Hayes. 2014. Genomic selection for feed efficiency in dairy cattle. *Animal* 8:1–10. <http://dx.doi.org/10.1017/s1751731113001687>.
- R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Accessed Apr. 23, 2015. <http://www.R-project.org/>.
- Roche, J. R., and D. P. Berry. 2006. Periparturient climatic, animal, and management factors influencing the incidence of milk fever

- in grazing systems. *J. Dairy Sci.* 89:2775–2783. [http://dx.doi.org/10.3168/jds.S0022-0302\(06\)72354-2](http://dx.doi.org/10.3168/jds.S0022-0302(06)72354-2).
- Ruegg, P. L., and R. L. Milton. 1995. Body condition scores of Holstein cows on Prince Edward Island, Canada: Relationships with yield, reproductive performance, and disease. *J. Dairy Sci.* 78:552–564.
- Schröder, U. J., and R. Staufenbiel. 2006. Invited Review: Methods to determine body fat reserves in the dairy cow with special regard to ultrasonographic measurement of backfat thickness. *J. Dairy Sci.* 89:1–14.
- Szabo, C., L. Babinszky, M. W. A. Verstegen, O. Vangen, A. J. M. Jansman, and E. Kanis. 1999. The application of digital imaging techniques in the in vivo estimation of the body composition of pigs: A review. *Livest. Prod. Sci.* 60:1–11.
- Vranic, D. V., and D. Saupe. 2001. 3D Shape Descriptor Based on 3D Fourier Transform. Proceedings of ECMCS-2001, the 3rd EURASIP Conference on Digital Signal Processing for Multimedia Communications and Services, Budapest, Hungary. <http://nbn-resolving.de/urn:nbn:de:bsz:352-231461>.
- Wallach, D. 2006. Evaluating crop models. Pages 17–25 in *Working with Dynamic Crop Models*. D. Wallach, D. Makowski, and J. W. Jones, ed. Elsevier, London, UK.
- Waltner, S. S., J. P. McNamara, J. K. Hillers, and D. L. Brown. 1994. Validation of indirect measures of body fat in lactating cows. *J. Dairy Sci.* 77:2570–2579.
- Weber, A., J. Salau, J. H. Haas, W. Junge, U. Bauer, J. Harms, O. Suhr, K. Schönrock, H. Rothfuß, S. Bielecki, and G. Thaller. 2014. Estimation of backfat thickness using extracted traits from an automatic 3D optical system in lactating Holstein-Friesian cows. *Livest. Sci.* 165:129–137. <http://dx.doi.org/10.1016/j.livsci.2014.03.022i>.
- Zion, B., V. Alchanatis, V. Ostrovsky, A. Barki, and I. Karplus. 2007. Real-time underwater sorting of edible fish species. *Comput. Electron. Agric.* 56:34–45. <http://dx.doi.org/10.1016/j.compag.2006.12.007>.