Performance monitoring in the cattle sector innovates with 3D imagery

T. Dechaux[1], A. Lebreton[1], Y. Do[1], M. El Jabri[1], L. Delattre[2], C. Gillé-Perrie[4], C. Allain[1] and M. Bruyas[3]

[1]Institut de l'Elevage, 149 Rue de Bercy, F-75595 Paris, France
[2]3D Ouest, 5 Rue Louis de Broglie, 22300 Lannion, France
[3]Eliance, 149 Rue de Bercy, F-75595 Paris, France
[4]Races de France, 149 Rue de Bercy, F-75595 Paris, France

Abstract

Performance recording and phenotyping of beef cattle are on the verge of a new era. Indeed, Bruyas *et al*. (2023) have recently shown that it is possible to collect three-dimensional images of beef calves at weaning using a 3D device suitable for high-throughput phenotyping and to automatically extract morphological parameters (heights, widths, volumes, surfaces, etc.). The aim of this new study, which is part of the PHENO3D project, was to develop models based on artificial intelligence to estimate Body Weight (BW) and morphological linear scores based on previously estimated body measurements. To achieve these objectives, 1194 Charolais calves aged 4 to 12 months and weighing from 90 to 620 kg were scanned on 14 commercial farms. Most of them were scanned twice, allowing a total of 2210 3D images to be acquired. Reference measurements were collected on these same animals: each calf was weighed on an electronic scale (BW) and scored by 3 experienced technicians. Scoring resulted in the estimation of 10 elementary scores, which then allowed for the calculation of 2 synthetic scores (ratings from 1 to 100) used ultimately for genetic selection: 1. the muscular conformation (MUS), relating to the musculature of the animal and 2. the size (SKE) of the animal relating to skeletal development. To predict BW, MUS, and SKE, various Machine Learning (ML) algorithms such as Extreme Gradient Boosting, Random Forest, and Elastic Net Regression were trained using 70% of the images and tested on the remaining 30%. The models were evaluated using Mean Absolute Error (MAE) and Spearman's correlation (rs). The repeatability of predictions was also assessed by Spearman's correlation between estimates made for the 1st image and the 2nd (when available). For BW, 1462 images were used for the learning model and 356 images for testing. For the best model, BW was predicted with a rs of 0.97 and an MAE of 12.1 kg (4.2%). The repeatability rs was 0.98 between the two images. For MUS and SKE, 1267 images were used to train the model and 308 images for testing. For the best model, MUS and SKE were predicted with respective rs values of 0.78 and 0.75, and MAEs of 7.1 (14.5%) and 6.5 (11.9%). The repeatability rs for these predictions for MUS and SKE were respectively 0.81 and 0.87. The Spearman's correlation for prediction and repeatability of MUS and SKE were higher than the average results obtained by experienced scorers during annual certification sessions. These results show that automating the scoring process using a 3D scanner combined with ML models is possible and allows for more accurate and repeatable estimates than those obtained by long-term scorers. The performances achieved on the Charolaise breed allow us to consider multiplying our models on the 9 other beef cattle breeds scored today (Limousine, Blonde d'Aquitaine, Salers, Aubrac, Parthenaise, Rouge des Prés, Blanc Bleu, Gasconne des Pyrénées, and Bazadaise) and to project towards the industrialization of the PHENO3D solution.

Keywords: Phenotyping, Calves, Weaning, 3D imaging, Artificial intelligence, pheno3d.
Presented at the ICAR Anual Conference 2024 in Bled at the Session 10: New approaches in the field of functional traits for management and breeding

Introduction

In the French beef cattle sector, genetic selection heavily relies on the monitoring and phenotyping of a diverse animal population (Griffon *et al*., 2017). This crucial process for breeding organizations, is predominantly conducted through a network of affiliated farmers and involves initial phenotyping usually conducted around the calves weaning. Technicians from either the Eliance network or breeding organizations of the Races de France network undertake on-farm data collection, encompassing animal weighing and morphological traits assessment. The morphological evaluation encompasses 19 scores, evaluating both muscular and skeletal development, as well as functional traits. Trained technicians visually perform this linear scoring, following the detailed methodology outlined by Lajudie *et al*. (2014) (Section 3 - ICAR Guidelines for Beef Cattle Production Recording). Despite the effectiveness of visual scoring, it requires extensive training and is susceptible to subjective biases. Hence, there is a pressing need in the beef sector to automate scoring processes to reduce training costs and minimize the impact of human biases on measurements.

To tackle these challenges and modernize the phenotyping process, the PHENO3D project was launched, representing a collaboration between Eliance (the French federation of breeding advising and service companies), Races de France (French federation of breeding organizations), and Idele (the French Livestock Institute). The project aims to harness 3D imaging technology and artificial intelligence to streamline phenotyping by automating weight measurement and morphological scoring of beef calves (Bruyas *et al*., 2022). An initial milestone of PHENO3D involved the development of a 3D scanning device capable of accurately capturing the three-dimensional profiles of weaning-age beef calves and extracting relevant morphological data from these images. The validation of this technology, following a methodology similar to that described by Le Cozler *et al*. (2019), compared live animal measurements with those derived from 3D images, yielding promising results (Bruyas *et al*., 2023). This successful validation marked a significant advancement, reinforcing the project’s trajectory and paving the way for subsequent phases of development and implementation.

Material and methods

Technology and image processing

The 3D scanner utilized in this investigation was previously detailed by Bruyas *et al*. (2022). It comprises a modular gantry with dimensions of 3 x 2.5 x 0.7 meters and incorporates ten depth sensors (see figure 1a). These sensors synchronize their data acquisition processes to produce comprehensive 3D images of the entire body of beef calves. Animals are scanned while in motion, passing beneath the device by walking or trotting, thereby enabling high throughput phenotyping. Integrated algorithms automatically enhance the images, streamlining the process for immediate image analysis. Through preprocessing and new feature extraction techniques, hundreds of indicators are automatically extracted to estimate body traits from 3D images. The developed methodology facilitates the automatic extraction of key body traits (Do *et al*., 2024), such as hip width (HW), chest depth (CD), wither height (WH), sacrum height (SH), body volume (BV), body surface (BS), and other measurements across numerous body slices (see figure 1b). All these body measurements were subsequently utilized to construct the prediction models developed in this investigation.

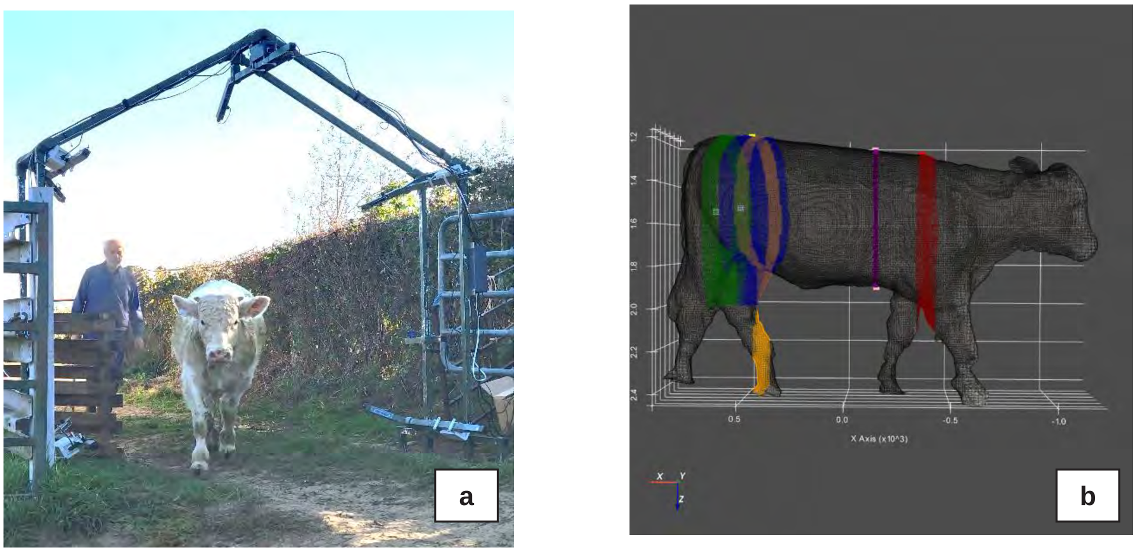

Figure 1. 3D scanner used for the trial (a) and automated image analysis (b).

Animals and reference data

To conduct our study, we scanned a total of 1194 Charolais calves, aged 4 to 12 months, and ranging in weight from 90 to 778 kg, across 14 commercial farms. Most of these calves underwent two scans, resulting in a total of 2210 3D images. All captured images were securely stored in the Microsoft Azure cloud platform.

Concurrently, reference measurements were obtained from these same animals: each calf was individually weighed on an electronic scale to determine its body weight (BW) and assessed by three experienced technicians. The visual scoring process led to the estimation of 10 elementary scores, each rated on a scale from 1 to 10, where a lower score indicates lower values and a higher score indicates higher values, based on frame and muscularity traits. The assessed traits are detailed in figure 2 below.

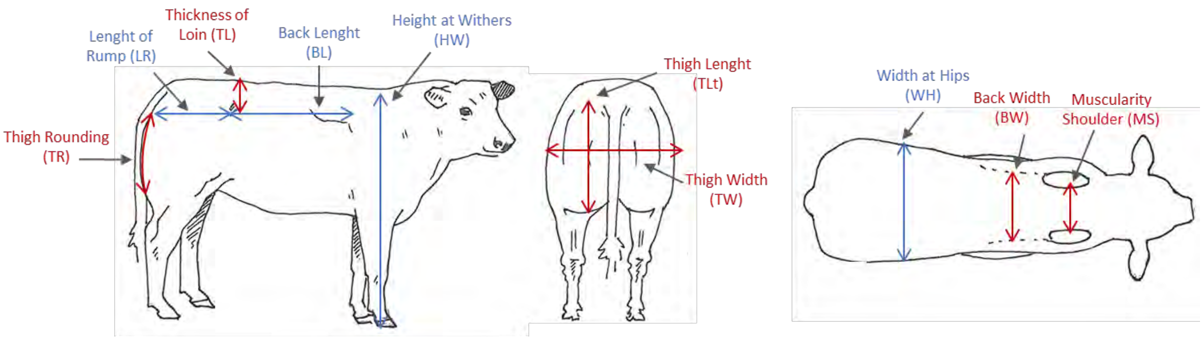

Figure 2. Frame traits (blue) and muscularity traits (red) used for linear scoring.

The 10 elementary scores were subsequently utilized to compute 2 synthetic scores, each rated on a scale from 1 to 100, which are ultimately employed for genetic selection purposes:

1. Muscle development (MUS), which pertains to the overall musculature of the animal.
2. Skeletal development (SKE), which relates to the body frame of the animal.

Figure 3 below displays images of calves exhibiting extreme morphologies for MUS and SKE. The four images depict calves of roughly the same age but with notable variations in size and muscularity.

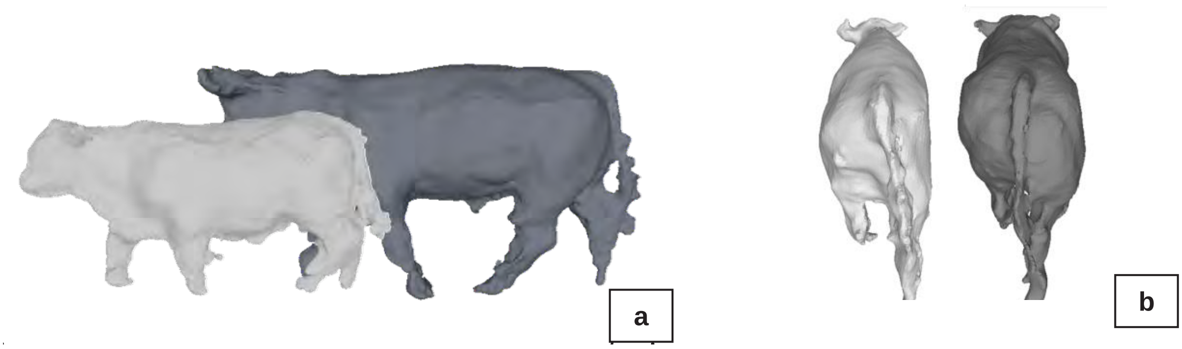

Figure 3. 3D image examples of extreme morphologies for calves around 6 months age.

The details concerning the animals' characteristics used for the trial are presented in table 1 below.

Machine learning models and data analysis

In our database, a total of 2210 3D images were initially available for analysis. However, after careful consideration, a subset of these images was excluded from the study due to factors impacting image quality and data reliability. These factors encompassed issues such as insufficient image clarity, inappropriate animal positioning during imaging (e.g., jumping or kicking), absence of duplicate images necessary for calculating repeatability, and errors in animal identification. The exclusion of these images was crucial to uphold the integrity of our study's findings. Ultimately, we utilized 1818 images for predicting body weight (BW) and 1575 images for predicting muscle (MUS) and skeletal (SKE) development. For both predictions, the models were trained using 80% of the images and tested on the remaining 20%, ensuring no overlap between train and test sets.

For BW prediction, 173 features were initially extracted from the 3D images. To enhance predictive performance and reduce dataset dimensionality, we employed

Table 1. Animals characteristics.

n=1194	Age	Weight	MUS	SKE
Average	221	287	56,2	59,1
SD	61,2	80,1	15,1	14,1
Min	44	130	10	12
Max	559	568	95,7	94

the Recursive Feature Elimination (RFE) method. RFE iteratively eliminates the least important features from the dataset, resulting in 61 selected features after training on the Random Forest estimator.

Subsequently, four machine learning models (Extreme Gradient Boosting, Random Forest, SVM Linear, and Lasso Regression) were trained on 1462 images from the learning dataset. To mitigate overfitting, models underwent training using a 4-fold cross-validation method with 5 repetitions. Following training, each model was evaluated on a test set comprising 356 images.

Similar methodologies were applied for predicting MUS and SKE synthetic scores. The RFE algorithm was used to select the most important features, resulting in 16 features for SKE score prediction and 51 features for MUS score prediction. The same four algorithms were trained on 1208 images using a 4-fold cross-validation method, repeated 5 times, and evaluated on a test set of 367 images.

Model evaluation metrics included Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Spearman correlation (rs). Additionally, the repeatability of predictions was assessed by calculating the Spearman correlation between estimates from duplicate images when available.

Results and discussion

Weight prediction

Table 2 below illustrates the performance of the ML models. Among the models evaluated, Extreme Gradient Boosting emerged as the top performer in terms of predictive accuracy following feature selection. On the training set, BW was predicted with an RMSE of 15kg (5.2%) and a MAE of 11.3kg (4.0%). This resulted in a high correlation between BW and the model's predictions, with a rs of 0.97 and an average R2 of 0.98.

The model's performance remained consistent across both datasets, with an RMSE of 15.6kg (5.4%) and a MAE of 12.1kg (4.2%). This consistency suggests that our model has effectively captured the underlying data patterns without overfitting to the training set, enhancing its reliability for real-world applications.

While Random Forest (RF) also demonstrated competitive performance, it slightly trailed behind Extreme Gradient Boosting. These ensemble methods excel in capturing intricate data interactions. Interestingly, SVM Linear and Lasso regression, despite their reputation for excellence in prediction tasks, exhibited relatively lower performance in terms of MAE and RMSE.

Figure 4 illustrates the relationship between estimated weight and ground truth values for both the train and test sets. With an overall R2 of 0.964, predictions and actual weights are distributed around the line of perfect prediction. With such high-performance levels across both train and test sets, the model demonstrates exceptional accuracy and reliability in predicting weight, rendering it suitable for practical applications.

The model's repeatability was assessed to confirm the consistency of BW predictions when different images of the same individual were provided. This was determined by calculating the Spearman correlation between two images of the same cattle when two 3D images were available. Our findings demonstrate a high level of repeatability of the model across two images of the same animal, with a rs of 0.98 for 738 cattle.

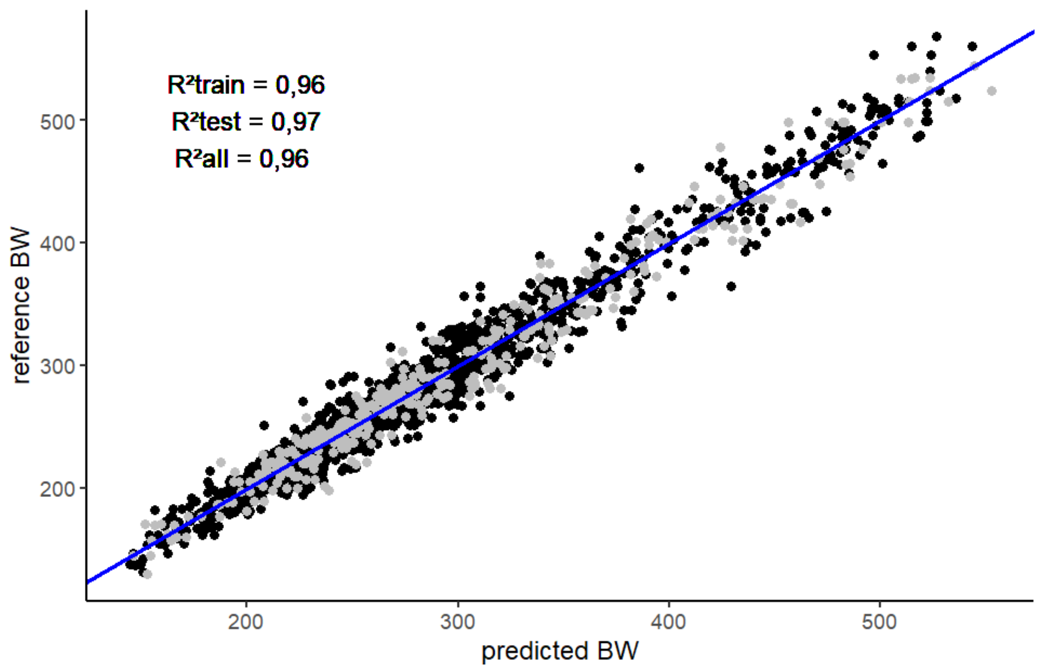

Figure 4. Predicted BW vs real BW for training set (black) and test set (grey). The blue line indicates the optimal fitting.

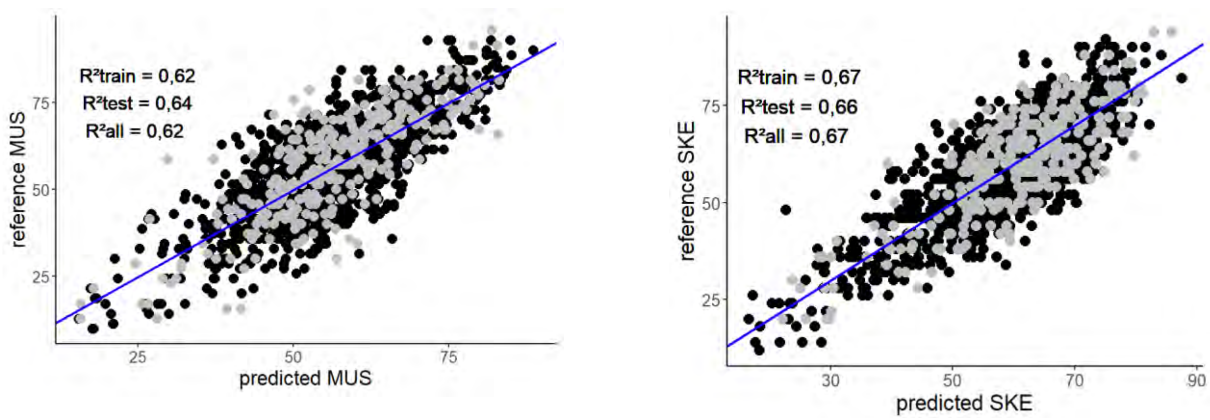

Figure 5. Predictions vs reference MUS and SKE scores for training set (black) and test set (grey). The blue line indicates the optimal fitting

Table 2. Models' performances on BW and synthetic scores prediction.

Predicted traits	Data sets	ML models / Statistics	Random Forest	Extreme Gradient Boosting	Lasso Regression	SVM Linear
BW	Train data set n=1462	MAE (kg)	11.7	11.3	12.9	12.8
		MAPE (%)	4.2%	4.0%	4.6%	4.6%
		RMSE (kg)	15.4	15.0	16.9	16.7
		r_s	0.97	0.97	0.96	0.96
		R^2	0.96	0.96	0.95	0.95
	Test data set n=356	MAE (kg)	12.4	12.1	13.6	13.3
		MAPE (%)	4.3%	4.2%	4.7%	4.5%
		RMSE (kg)	16.2	15.6	17.7	17.7
		r_s	0.97	0.98	0.97	0.97
		R^2	0.96	0.97	0.95	0.95
MUS	Train data set n=1267	MAE	7.6	7.1	7.6	7.6
		MAPE (%)	15.6%	14.5%	15.8%	15.5%
		RMSE	9.4	9.0	9.6	9.7
		r_s	0.73	0.75	0.72	0.73
		R^2	0.58	0.62	0.56	0.56
	Test data set n=308	MAE	7.4	7.1	8.0	7.9
		MAPE (%)	15.5%	14.5%	16.9%	16.7%
		RMSE	9.4	9.1	10.0	10.0
		r_s	0.77	0.78	0.74	0.74
		R^2	0.64	0.65	0.58	0.58
SKE	Train data set n=1267	MAE	6.3	6.1	7.2	7.2
		MAPE (%)	11.9%	11.3%	13.9%	14.0%
		RMSE	7.8	7.7	9.0	9.0
		r_s	0.78	0.79	0.74	0.73
		R^2	0.67	0.68	0.56	0.55
	Test data set n=308	MAE	6.3	6.5	6.9	7.0
		MAPE (%)	11.3%	11.9%	12.9%	13.1%
		RMSE	7.7	8.0	8.5	8.5
		r_s	0.78	0.75	0.76	0.75
		R^2	0.69	0.67	0.62	0.62

Synthetic scores prediction

To predict the synthetic scores, we utilized a dataset comprising 1575 images, with predicted weight included as a predictive variable. Employing the same methodology, we trained models using a 4-fold cross-validation and evaluated their performance on the test set. The results across the entire dataset are presented in Table 2.

Both Random Forest and Extreme Gradient Boosting emerged as the top-performing algorithms. For SKE score predictions, Extreme Gradient Boosting exhibited greater precision, with an MAE of 6.2 (11.4%) and an RMSE of 7.7 (15.2%). Figure 5 illustrates the relationship between reference SKE and the predicted SKE of this model, with an R2 of 0.67 suggesting a correct linear relation between predictions and reference. Similar performances were observed in both the train and test sets, indicating good generalization of the model. To enhance this performance further, it may be advantageous to include a certain proportion of extreme SKE values, particularly those below 40, where the number of animals in our study is limited. Moreover, the model demonstrated good repeatability, with a rs of 0.87, significantly surpassing the repeatability target of 0.78.

For MUS score predictions, Extreme Gradient Boosting also emerged as the top model, achieving an MAE of 7.1 (14.5%) and an RMSE of 9.0 (21.6%). Figure 5 illustrates a strong correlation between predictions and references, with an R2 of 0.62 for the train set, 0.64 for the test set, and 0.62 across the entire dataset. Additionally, the model exhibited good repeatability, with a rs of 0.84, notably exceeding the repeatability target of 0.75.

Conclusion

In conclusion, this study underscores the feasibility of employing three-dimensional imaging in conjunction with artificial intelligence methods to accurately estimate body weight (BW) and linear scores in calves. By leveraging machine learning models, we achieved robust predictions for BW, muscle development (MUS), and skeletal development (SKE), surpassing the accuracy of experienced human scorers. The high repeatability of these predictions underscores the reliability of our approach, promising improved phenotypic assessment in livestock breeding programs..

Acknowledgments

The authors extend their gratitude to all scoring technicians and participating farmers for their contributions to the trial. Furthermore, we express our appreciation to the French Ministry of Agriculture and APIS-GENE for their generous funding support of the PHENO3D project.

List of references

Bruyas, M., Lecomte, C., Allain, C., Lebreton, A., Delattre, L., Delouard, J.M., Gautier, J.M., Leudet, O., Brochard, M., Salvetti, P., Griffon, L. (2022). PHENO3D: phenotyping calves at weaning through automatic 3D image collection and instantaneous processing. ICAR conference Montréal 2022.

Bruyas, M., Lebreton, A., Griffon, L., Delattre, L., Allain, C., 2023. PHENO3D: Validation of a high-throughput movable 3D device for the acquisition of the whole cattle body. ICAR conference Toledo 2023.

Lajudie, P., Berrechet, P., Robergeot, F., Deconchat, P., Tournadre, O., Roy, M., Gout, P., Roussie, C., Hanriot, J. C., Laloe, D., 2014. Guide pratique du pointage des bovins de race à viande du pointage à l'âge adulte. https://idele.fr/fileadmin/medias/Documents/Manuel_pointage_vdef3.pdf

Le Cozler, Y., Allain, C., Caillot, A., Delouard, J. M., Delattre, L., Luginbuhl, T., and Faverdin, P. (2019a). High-precision scanning system for complete 3D cow body shape imaging and analysis of morphological traits. Computers and electronics in Agriculture, 157, 447-453

Le Cozler, Y., Allain, C., Xavier, C., Depuille, L., Caillot, A., Delouard, J.M., and Faverdin, P. (2019b). Volume and surface area of Holstein dairy cows calculated from complete 3D shapes acquired using a high-precision scanning system: Interest for body weight estimation. Computers and Electronics in Agriculture, 165, 104977.

Griffon, L., Boulesteix, P., Delpeuch, A., Gion, A., Guerrier, J., Leudet, O., Miller, S., Saintilan, R., Venot, E., and Tribout, T. (2017). La sélection génétique des races bovines allaitantes en France : un dispositif et des outils innovants au service des filières viande. INRA Productions Animales, 30(2), 107-124.