Transferring a 3D imaging POC to beef cattle breeding stakeholders to do on-farm high-throughput phenotyping: the PHENO3D example

How to organize on-farm data collection to build machine learning models ?

A. Lebreton, C. Gillé-Perrier, L. Delattre, C. Allain, M. Bruyas

Contact : adrien.lebreton@idele.fr

Funded by : APIS-GENE

Avec la contribution financière du compte d'affectation spéciale développement agricole et rural CASDAR

MINISTÈRE DE L'AGRICULTURE ET DE LA SOUVERAINETÉ ALIMENTAIRE

Context : Need to highlight synergies between Research and Private stakeholders

Today's Focus : How To Organize On-farm Data Collection To Build Machine Learning Models By Engaging Private Stakeholers ?

« How much data is required for Machine Learning models in PLF ?»

« How to improve data quality to train Machine Learning models in PLF ?»

« How to improve data variance collection to train Machine Learning models in PLF ? »

Context : PHENO3D use case

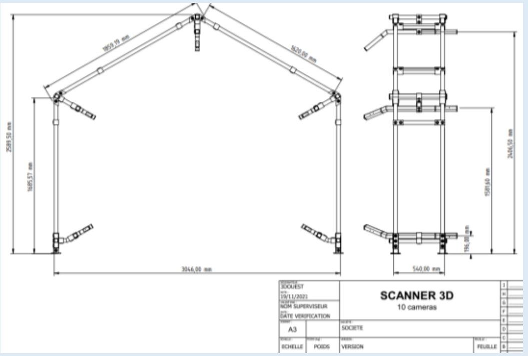
- Performance monitoring = Body Weight + 19 linear scores
  - Used for consultancy and animal selection
- 435 000 calves phenotyped in 2022 (on 10 french breeds)

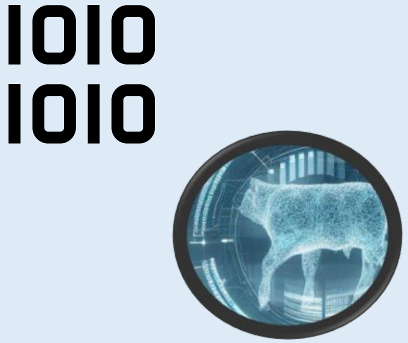
A system that has worked for decades but which have now limits (cost of maintaining the skills in the operators)

The example of the PHENO3D project

Goal : Automate the collection of live weight and the 19 notes of scoring (done visually today) on 10 beef breeds

3 bricks of project deliverables :





A 3D scanner for high-throughput phenotyping in farm

3D OUEST

Prediction AI

Service for breeders

Consortium Animal 3D

Validated by Lebreton et al. (2023)

Presented by Dechaux et al. (2024) in session 28

How to organize on-farm data collection to build machine learning models ?

How much data do you want to train your model ?

Data scientist

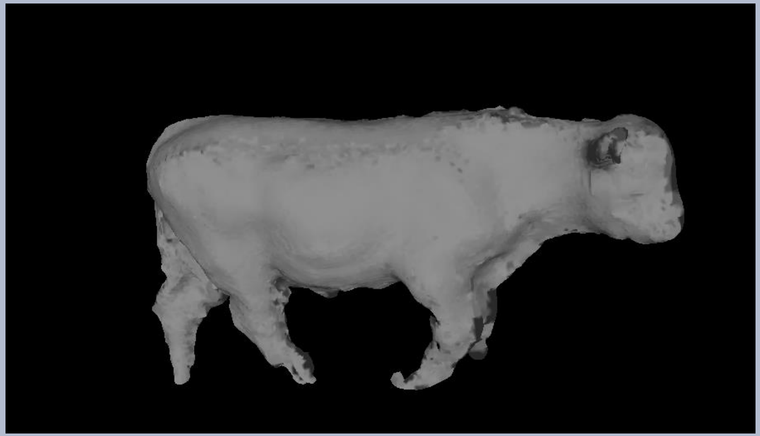
I don't know ! What variance and data quality will you offer me ?

10 000 to 50 000 data

ChatGPT

Objective : Define a framework of on-farm data collection to maximise the variance and the quality of the data collected to train Machine learning models

M&M : Data collected



3D images (2 per animal)

Morphological linear scores:
Skeletal Development,
Muscular Development
(visually made by 3 selected operators)

Body Weight


https://www.marechalle-pesage.fr

Today's Focus : on Charolais Calves around weaning

M&M : the team to ensure quality and variance in the data collection

National federation of performance monitoring organisations (55 regional organisations)

- National and local vision of performance monitoring
- Local advisors and Scoring Operators
  - Knowledge of the farms
  - Linear scoring operators

National federation of breed organisations (17 beef breed societies)

- Governance of the breed selection programs
- National and local vision of performance monitoring
- Local advisors and Scoring operators

NPO Research organization

- Quality management of the phenotypes collected
  - Scoring Operators training evaluation
- 3D acquisition experience
- AI developers

Avec la contribution financière du compte d'affectation spéciale développement agricole et rural CASDAR

MINISTÈRE DE L'AGRICULTURE ET DE LA SOUVERAINETÉ ALIMENTAIRE

Funders

M&M : stakeholders' organization for data collection

Step 1 : Regions and period selection

Volume: Consult with national stakeholders for insights into regional practices.
Validate availability with local stakeholders.

Variance: National stakeholders provide data on past performance and variance.
Local stakeholders validate with their regional knowledge.

Quality: Local stakeholders check global availabilities and the welcoming nature of the region.

Step 2 : Farms selection

Volume: National stakeholders specify complementary data needs across regions.
Local stakeholders select the best farms based on availability.

Variance: National stakeholders provide past performance data.
Local stakeholders investigate current herd variance.

Quality: Local stakeholders assess and educate farmers on the process.

M&M : stakeholders' organization for data collection : PHASE I

Step 3 : Scoring Operator Selection

Quality: *Select 3 operators based on their performance and evaluation.*

Step 4 : Data Collection Management

Variance: ***Prioritize calves with the most variance for phenotyping if all calves cannot be scored.***

Quality: *local advisors facilitate contacts with the farmers*

Ensure daily calibration.

Ensure supervision by an R&D project manager.

M&M : stakeholders' organization for data collection (Real Life)

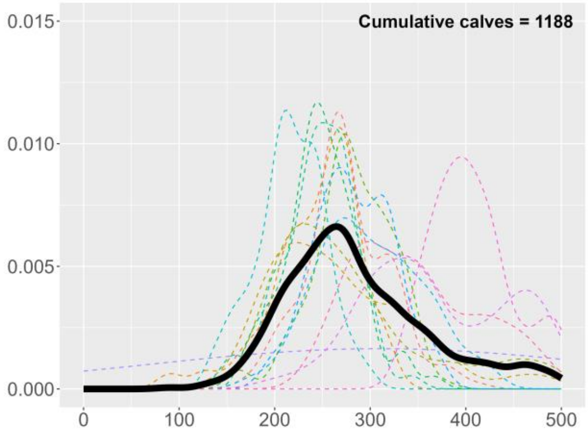
Week 1 : Warm up on very « welcoming » conditions

Week 2 : Target « High » phenotypes

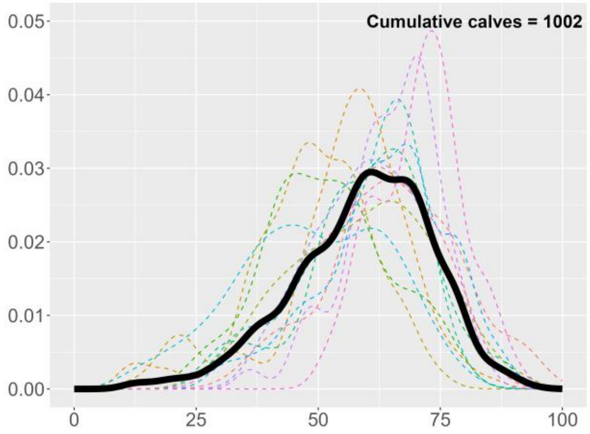
Week 3 : Ultra-targeted data collection

Results : data distribution by days of data collection

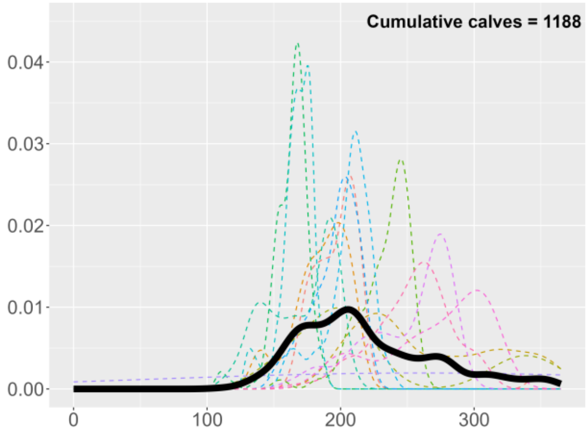
Density



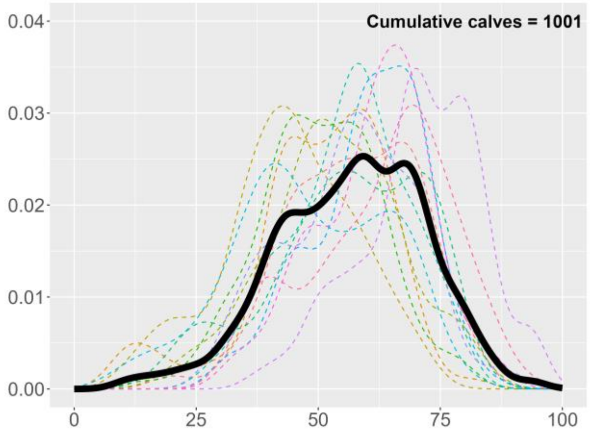
Body Weight (kg)



Skeletal Development score (0:100)



Age (days)



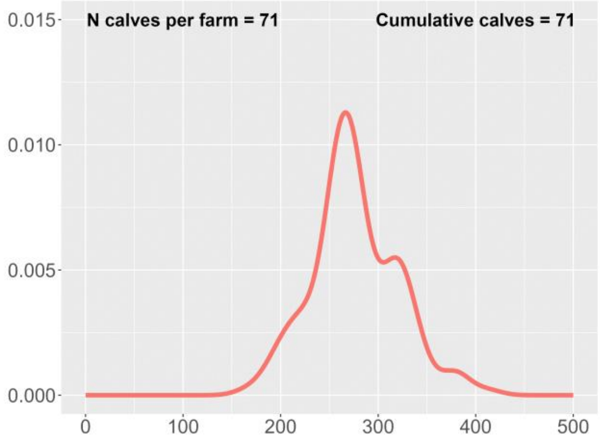
Muscular Development score (0:100)

One colored dispersion curve per day of data collection

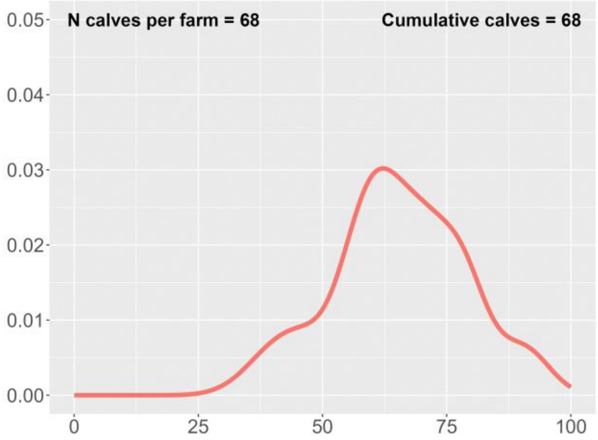
Black dispersion curve : whole days distribution

Results : data distribution by days of data collection

Density

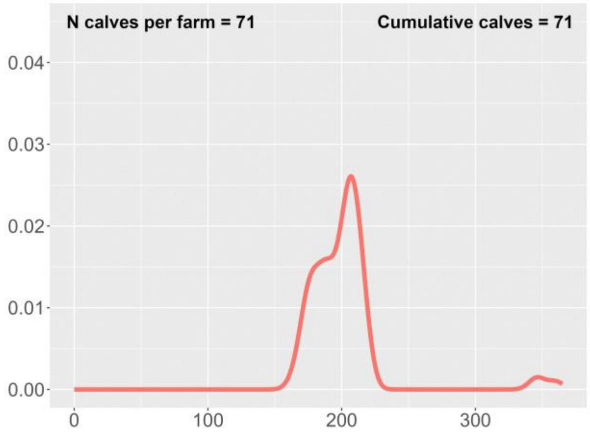


Body Weight (kg)

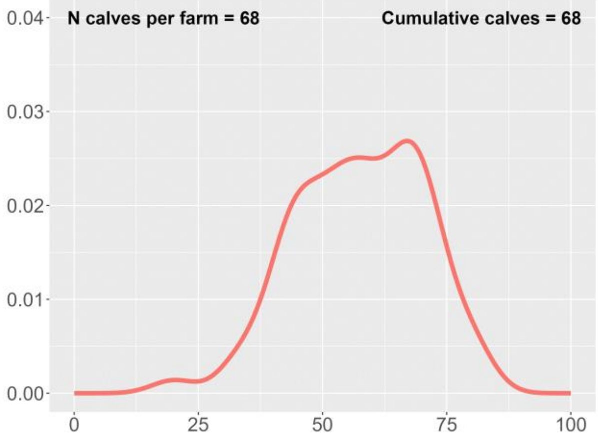


Skeletal Development score (0:100)

One colored dispersion curve per day of data collection



Age (days)

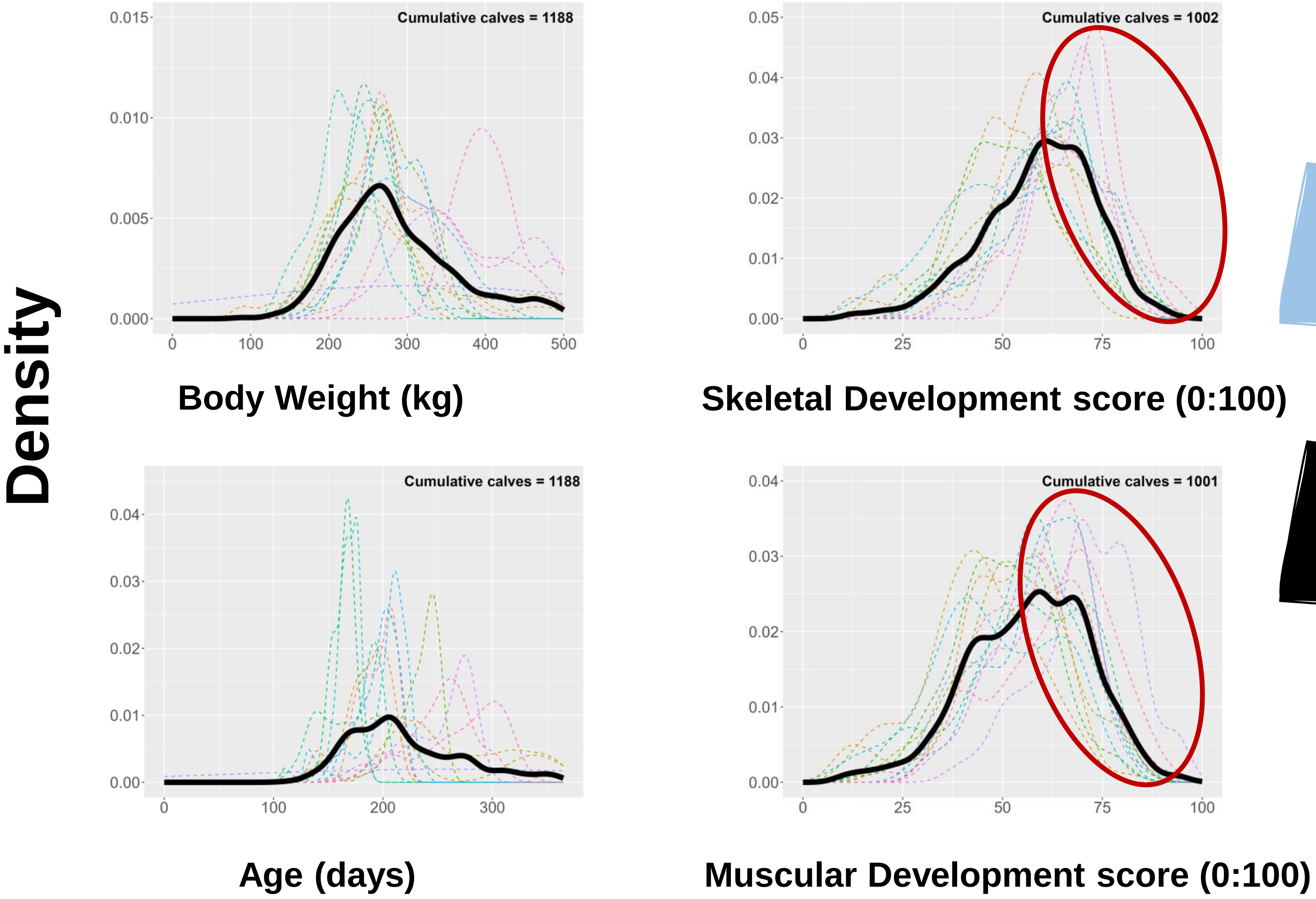


Muscular Development score (0:100)

Black dispersion curve : whole days distribution

From collecting data Volume to data Variance

Results : data distribution by days of data collection

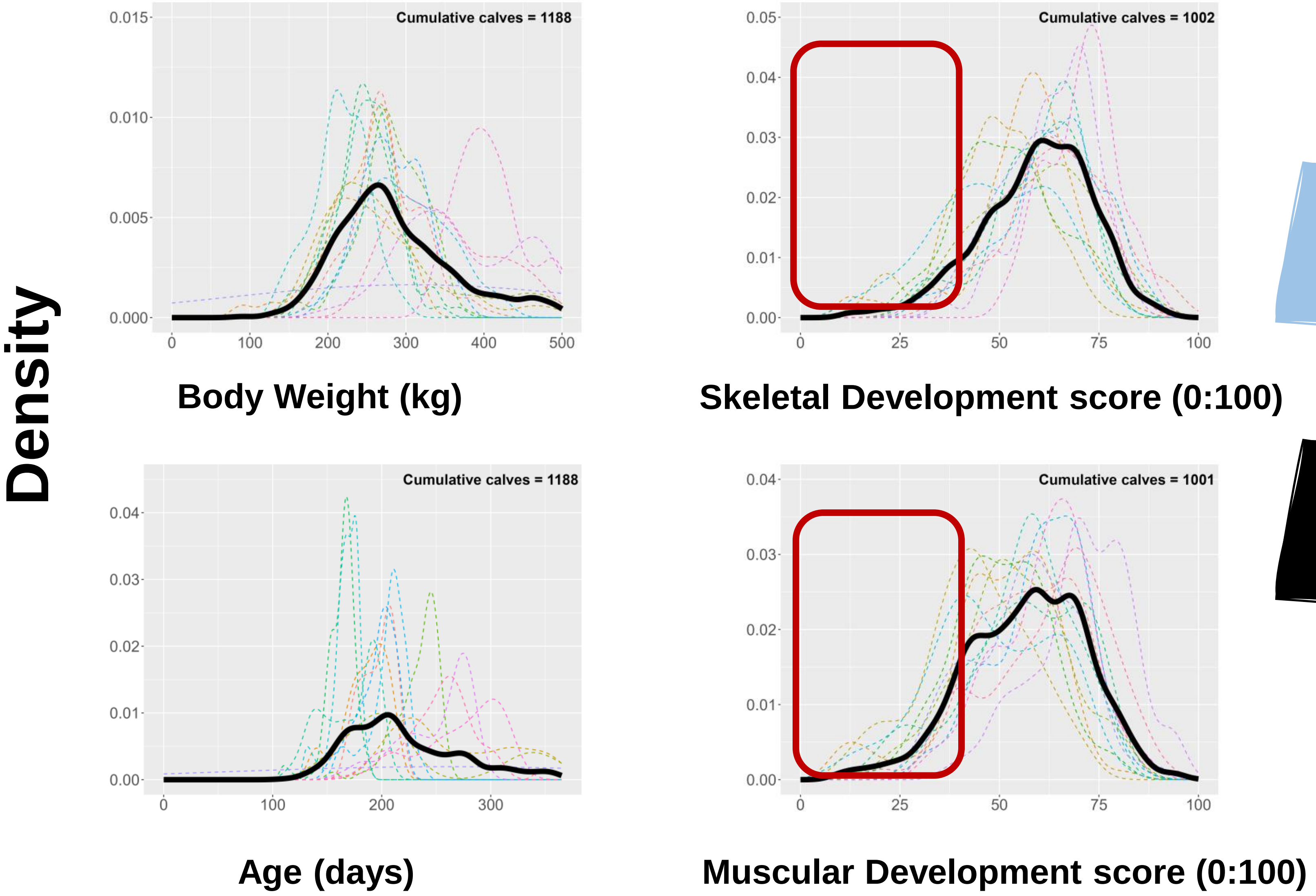


One colored dispersion curve per day of data collection

Black dispersion curve : whole days distribution

Our framework worked to provide « high performance » variance

Results : data distribution by days of data collection



One colored dispersion curve per day of data collection

Black dispersion curve : whole days distribution

« Low performance » animals are harder to find

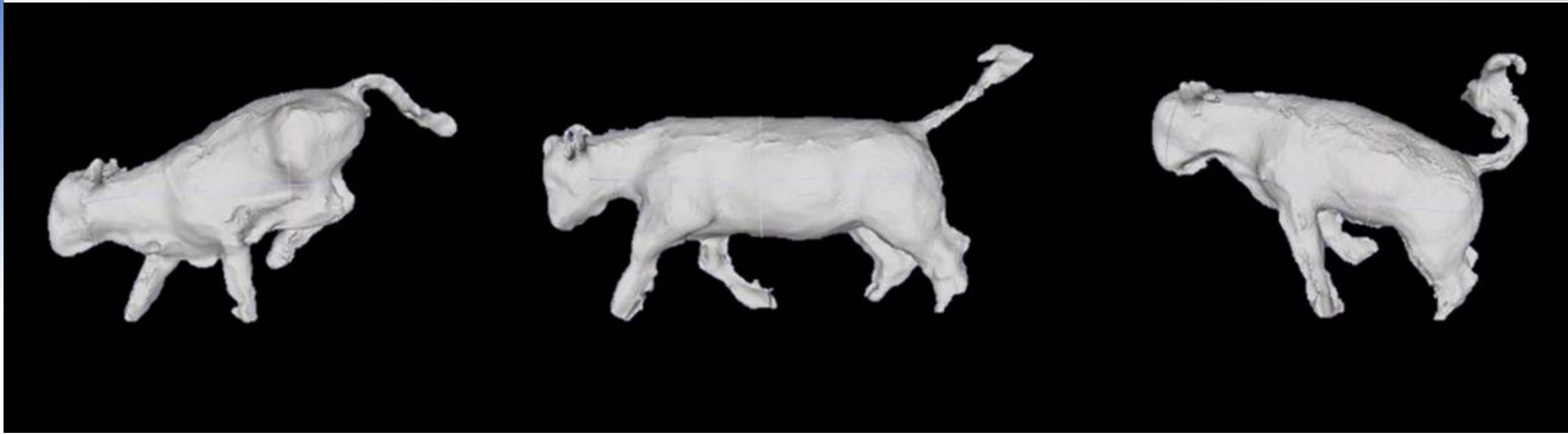
Take-Home messages

- **Prioritize collecting high-quality, varied data over merely amassing large volumes of “average” data.**

- **To do that :**
  - **Step outside the research facility**—gather data from commercial farms to capture population-level variance.
  - **Engage Private stakeholders**, at various levels, to successfully collect data on commercial farms.

*In PHENO3D project, **500 animals scanned by breed** with High quality data and Variance between animals seems enough to predict BW and linear scores accurately.*
Framework repeated in 10 French breeds.

Even with optimal organization, **capturing variance remains a significant challenge.**



Thank you for your attention !

Contact : adrien.lebreton@idele.fr

Funded by :

Adaptation à la contention de l'élevage