Beyond Accuracy in Precision Livestock Farming

Distribution-Aware Evaluation for Large-Scale Poultry Weight Monitoring

G. Toth[2,3], M. Alexy[1,2,3]

[1]Obudai University, John von Neumann Faculty of Informatics, Budapest, Hungary
[2]Eötvös Loránd University, Faculty of Informatics, Budapest, Hungary
[3]University Research and Innovation Centre, Budapest, Hungary

1. Original research setup

CAMERA, SCALE AND PRODUCTION ENVIRONMENT

Productive camera setup

visual monitoring in a real poultry environment

Reference weight source

paired data for model development and validation

The original task was practical: connect camera observations to reference weights.

2. From images to a rotation-level average

HOW THE FIRST OUTPUT WAS CREATED

Mean error per rotation

The first evaluation was simple: compare the estimated average with the slaughterhouse average.

3. Ross308 average results by rotation

THE MEAN-BASED VIEW

Initial evaluation: average error of the daily estimates

$$E_r = \left| \frac{\widehat{\mu}_r - \mu_r}{\mu_r} \right| \cdot 100$$

5–10%

absolute relative error in earlier Ross308 rotations

At this point, the problem looked like model error.

4. The turning point

SAME MODEL DIRECTION, DIFFERENT OUTCOME

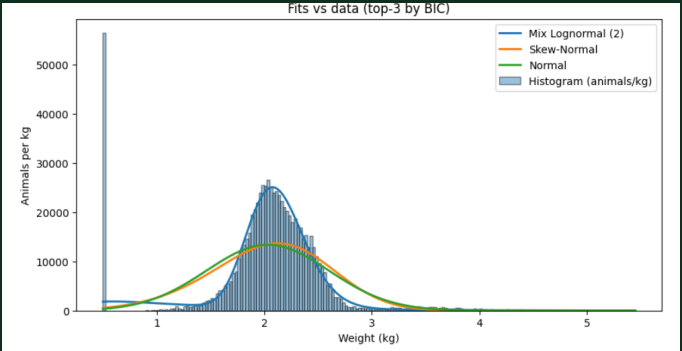
2025.08.16 slaughterhouse validation

slaughterhouse reference	model estimate	mean difference
2032.6 g	2109.2 g	+3.8%
observed grill-weight mean	evaluated model output	relative error

The result changed, so the interpretation had to change too.

5. The reference distribution we looked at

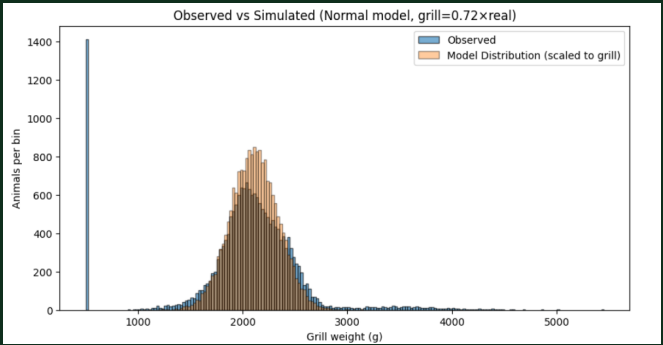
THE MOMENT THE QUESTION CHANGED



The average was acceptable, but the reference distribution was not a simple normal shape.

6. What the August report showed

DISTRIBUTIONS, NOT ONLY AVERAGES



The validation question moved from mean error to distributional shape.

7. Two possible explanations

WHAT CAUSED THE DIFFERENCE?

Model error

bias / features / training

Sampling error

unique animals / visibility / distribution

In practice, both can contribute. Here we follow the sampling line.

A bad final result is not automatically a bad model.

8. First sampling question

IMAGE COUNT IS NOT SAMPLE SIZE

$$N_{images} \neq n_{animals}$$

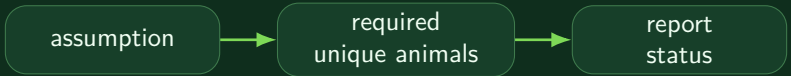
The key question became: do we have enough independent animal-level measurements?

9. The answer depends on assumptions

NORMAL, MODEL-BASED OR DISTRIBUTION-FREE?

Required number of **unique animals** to estimate the flock mean
$N = 20{,}000$ animals, $\sigma \approx 203$ g, 95% reliability; FPC = finite-population correction

Assumption	± 10 g	± 50 g	± 100 g
Distribution-free, Chebyshev + FPC	$\sim$5837	$\sim$325	$\sim$83
Normal approximation + FPC	$\sim$1468	$\sim$64	$\sim$16



10. Backward-compatible output

AVERAGE PLUS SAMPLE SUPPORT

$$\widehat{\mu} = X$$

remains the core output.

$$\widehat{\mu} = X \quad + \quad \textbf{supported?}$$

The average stays, but the system also reports whether the sample supports the claim.

11. Distribution-based reporting model

FROM OBSERVATIONS TO A POPULATION CLAIM

The report is a claim about the real animal distribution.

$$F \longrightarrow O_d \longrightarrow \widehat{F}_d \longrightarrow \widehat{\mu}_d$$

F	=	real flock distribution
O_d	=	observed animals under design d
$\widehat{F}_d$	=	estimated distribution
$\widehat{\mu}_d$	=	reported average

Before trusting the average, the system checks whether enough unique animals were observed.

12. Cost-aware design problem

WHAT SHOULD WE IMPROVE?

More information has a cost. Wrong reports also have a cost.

$$d^* = \arg\min_{d \in \mathcal{D}} \left[C(d) + \lambda L(\widehat{F}_d, F) \right]$$

$C(d)$	=	sensing and data-collection cost
$L(\widehat{F}_d, F)$	=	loss from the wrong distribution
λ	=	importance of report reliability

The optimal system is the best cost–reliability trade-off, not simply the most accurate model.

13. Beyond average error

TRAINING AND EVALUATION SHOULD ALSO BE DISTRIBUTIONAL

If the target is a flock distribution, the metric should see the distribution.

mean error $\longrightarrow$ distributional loss

MAE / MAPE	=	average-level prediction error
Wasserstein distance	=	mass transport between weight distributions
MMD	=	kernel-based distribution difference
KL / JS divergence	=	shape mismatch under density assumptions

The model should not only predict individual weights well; it should preserve the flock-level distribution.

15. Take-home message

CONCLUSION

Accurate predictions are not enough.

The flock distribution must also be identifiable.

The system should know when it does not have enough independent data.

Model accuracy $\neq$ population-report reliability.

Thank you for your attention.

AI4AS – Artificial Intelligence 4 Animal Science | Ghent, Belgium / 29–30 June 2026

Beyond Accuracy in Precision Livestock Farming