From Empirical to Mechanistic to Al Models, Making Livestock Production More Efficient (a case study from my lab group)

Dr J. L. Ellis, Associate Professor, Animal Systems Modelling, Centre for Nutrition Modelling, The University of Guelph, ON, Canada; jellis@uoguelph.ca











Agriculture 4.0 – The 4th Agricultural Revolution

 We are witnessing the fusion of animal production systems with emerging digital technologies and automation

Barriers of the past:

Data capture abilities

Efficient data capture/Lack of automation

Appropriate analytical tools for data processing and analysis

Opportunities of the future:

System Optimization and Decision Support

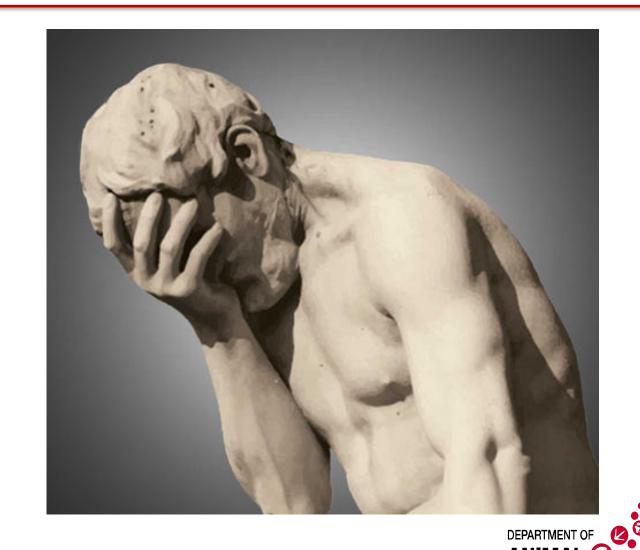
Full automation of some aspects of the animal production chain





A Case Study: Feed Manufacturing

The Story of:
A botched
experiment
(oops)





Feed Pelleting

"Agglomerated feed formed by extruding individual ingredients or mixtures of ingredients by compacting and forcing them through die openings by any mechanical process" (Okewole and Igbeka, 2016)

Benefits

- 1. Improve animal performance
 - Increase digestibility
 - Reduced feed sorting/wastage
 - Improved FCR
- 2. Improve physical characteristics
 - Increase flowability
 - Increase density





Pellet Quality

Pellet Durability Index (PDI)

 The ability of pelleted feed to withstand fragmentation and abrasion during handling and maintain structural integrity



Good quality

Poor quality

- The widely used measure for pellet quality
- The proportion of the remaining pellets to the whole pellets after the pellet durability test
- Two pellet durability test methods: Tumbling and Holmen



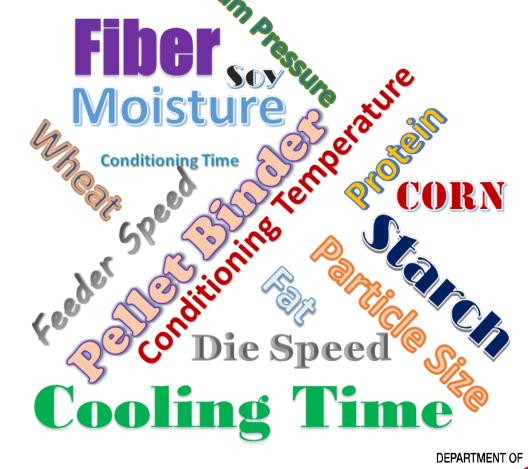
Holmen tester





Pellet Quality Is Impacted by:

- Conditioning temperature (Massuquetto et al., 2018)
- Fibre content (Zimonja et al., 2008)
- Moisture (Moritz et al. 2003)
- Particle size (Stevens, 1987)
- Fat content (Hossein et al, 2019)
- ...
- Anecdotally Mill Operator





Predicting PDI - The Knowledge Gap

 Most studies consider only a few factors at a time, and under controlled experimental conditions

Example:

Gilpin, A. S., Herrman, T. J., Behnke, K. C., & Fairchild, F. J. (2002). Feed moisture, retention time, and steam as quality and energy utilization determinants in the pelleting process. *Applied Engineering in Agriculture*, 18(3), 331.

Moisture (Two levels: 12 and 14%)

Conditioning time (Two levels: short and long) **Steam quality** (Four levels: 70, 80, 90, and 100%)

Only one study had built empirical equations to describe pellet quality under commercial settings

Example

Schroeder, B., Andretta, I., Kipper, M., Franceschi, C. H., & Remus, A. (2020). Empirical modelling the quality of pelleted feed for broilers and pigs. *Animal Feed Science and Technology*, 265, 114522.





Research Question

Can we predict PDI in the commercial feed mill setting?

Application – can we use the model to optimize pellet quality (and other aspects of manufacturing)?





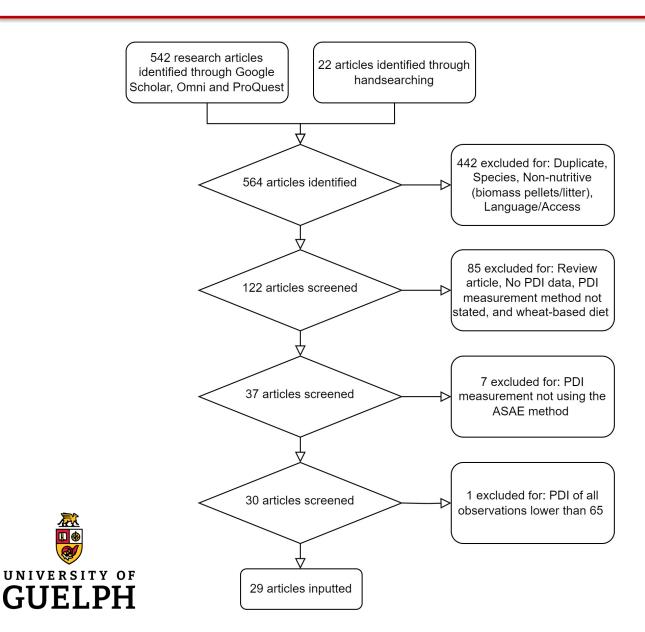
Study 1 - Research Question

Can we develop PDI prediction models via a meta-analysis of the published literature?





Study 1 - Data



- Systematic review conducted in 2024
- 280 treatment means from 29 studies
- PDI measured via ASAE method
- 16 variables:
 - Feed Ingredients (5)
 - Manufacturing parameters (6)
 - Nutrient composition (5)



Study 1 – Model Building Pipeline

Variable Selection:

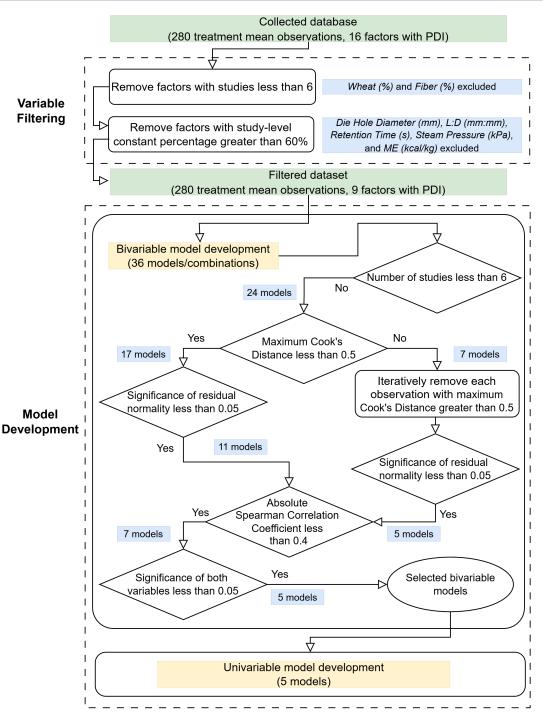
- Removed features reported in less than
 6 studies
- Removed features with limited withinstudy variation

Model Development:

- Mixed model analysis, treating study as a random effect
- Evaluated for:

UNIVERSITY OF

- Outliers via CooksD test
- Normality of residuals
- Multicollinearity between features
- Significance (P < 0.05)



Study 1 – Model Evaluation

3-fold cross-validation

studies moved into folds

Evaluation measures

Mean square prediction error (MSPE) (Bibby & Toutenburg, 1977)

Concordance Correlation Coefficient (CCC) (Lin, 1989)

Visual evaluation of plots





Study 1 – Results

Table 1 Estimated parameters of the selected bivariable models.

Model	Parameter	Estimate	Standard Error	P-Value
_	Intercept	2.975	0.3266	< 0.0001
Model 1	Corn	-0.011	0.0048	0.0269
	Lipid	-0.144	0.0280	< 0.0001
	Intercept	1.094	0.7425	0.1434
Model 2	Corn	-0.013	0.0047	0.0090
	Steam Temperature	0.020	0.0082	0.0135
	Intercept	1.948	0.2736	< 0.0001
Model 3	Lipid	-0.137	0.0283	< 0.0001
	Soybean Meal	0.013	0.0049	0.0130
	Intercept	0.003	0.6918	0.9966
Model 4	Soybean Meal	0.016	0.0052	0.0026
	Steam Temperature	0.020	0.0081	0.0179
	Intercept	1.501	0.3636	0.0661
Model 5 ¹	Lipid	-0.132	0.0280	< 0.0001
	Protein	0.041	0.0130	0.0019

¹ One observation was removed for Model 3 due to the Maximum Cook's Distance greater than 0.5.

Table 1 Model evaluation for the selected bivariable models using 3-fold cross-validation.

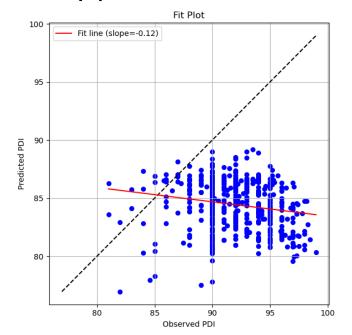
Model ¹	rMSPE (%)	MSPE	ECT (%)	ER (%)	ED (%)	CCC
Model 1	7.79 ± 0.835	42.65 ± 8.541	15.2 ± 23.75	0.9 ± 0.51	83.9 ± 23.75	0.25 ± 0
Model 2	8.01 ± 0.206	48.51 ± 2.062	6.4 ± 9.96	12.1 ± 12.88	81.6 ± 8.29	$-0.02 \pm ($
Model 3	7.62 ± 1.702	41.76 ± 16.864	2.3 ± 2.60	2.0 ± 1.75	95.7 ± 1.04	0.17 ± 0
Model 4	7.81 ± 0.285	46.16 ± 2.930	3.8 ± 6.38	6.0 ± 8.99	90.1 ± 8.50	0.02 ± 0
Model 5	10.33 ± 1.167	77.90 ± 20.929	17.0 ± 17.28	29.8 ± 29.51	53.2 ± 24.37	0.29 ± 0





Study 1 – Conclusions/Application

- Simple models were developed based on published literature easy to implement in practice (ie. simple formula to add into Brill/BESTMIX etc. software to predict PDI)
- How would they perform in a commercial mill...?



Root MSPE (%): 9.2604

MSPE: 72.48748

ECT (%): 77.448389

ER (%): 9.810557

ED (%): 12.741053

CCC: -0.033753

Pearson Correlation: -0.181764

Bias Correction Factor (Cb): 0.185695



Can we do better ...?







Predicting Pellet Quality at the Mill Level Using Machine Learning

Jihao You (PhD Candidate)

Supervisor: Dr Jennifer Ellis

Co-supervisor: Dr Dan Tulpan

Why Might ML be a Good Tool for the Feed Mill?

Feed mills are naturally 'big data' generators:

Volume – large enough that it prevents visual inspection & processing on a conventional computer

Variety – both the type and nature of the data. digital images, on-line & off-line video recordings, environmental sensor output (e.g. temp, movement, activity), animal biosensor output (e.g. indwelling pH probe, temperature, redox potential), sound recordings, omics data

Velocity – typically produced and (mostly) analyzed in real-time

Veracity – data quality - it is often not 'clean', containing missing observations, confounding variables and outliers





Study 2 - Data

Based on data hand-collected between March 31, 2020 – Nov 27, 2020, at Trouw Nutrition - St. Marys Plant 2

- Data included 1,434 observations and 17 variables
- Data were split into two subsets:
 - 80%: For model construction with a 5-fold cross-validation technique
 - 20%: For independent model evaluation
- 12 Machine learning (ML) algorithms were applied and compared





Study 2 - Data

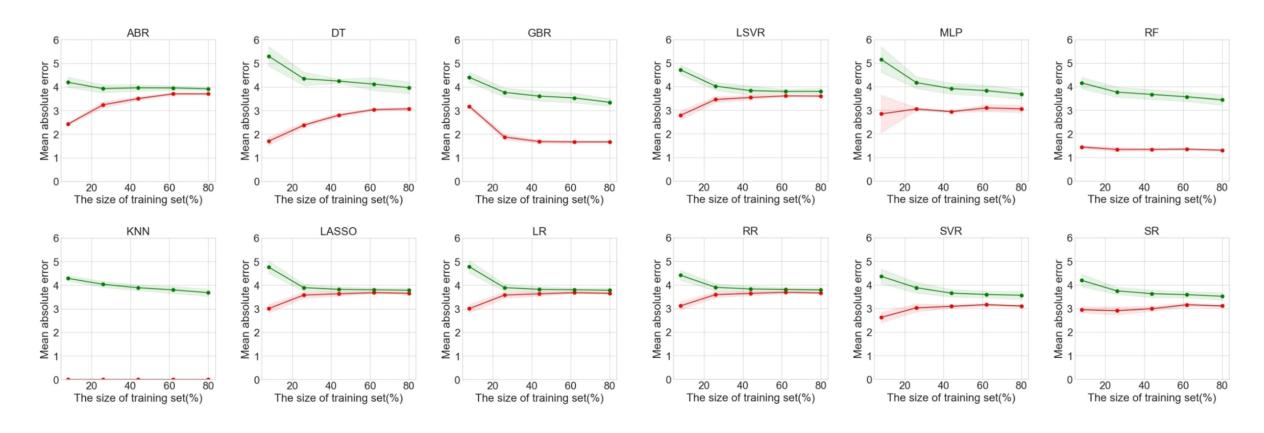
Index	Feature	Sources	Types	
1	Aver. Daily Temp. (°C)			
2	Aver. Daily Humidity (%)	Environment		
3	Aver. Daily Pressure			
4	Wheat (%)			
5	Bakery (%)			
6	Canola (%)	Feed	Continuous	
7	Soy (%)	formulation	features	
8	DDGS (%)		leatures	
9	Corn (%)			
10	Fat (%)			
11	Order Size (Tonne)			
12	Feeder Speed (%)			
13	Cond. Temp. (°C)	 Manufacturing		
14	Product Line		Catagorical	
15	Spec. (Types)	parameters	Categorical features	
16	Operator			
17	PDI		Output	

Environmental Data: https://www.wunderground.com/history





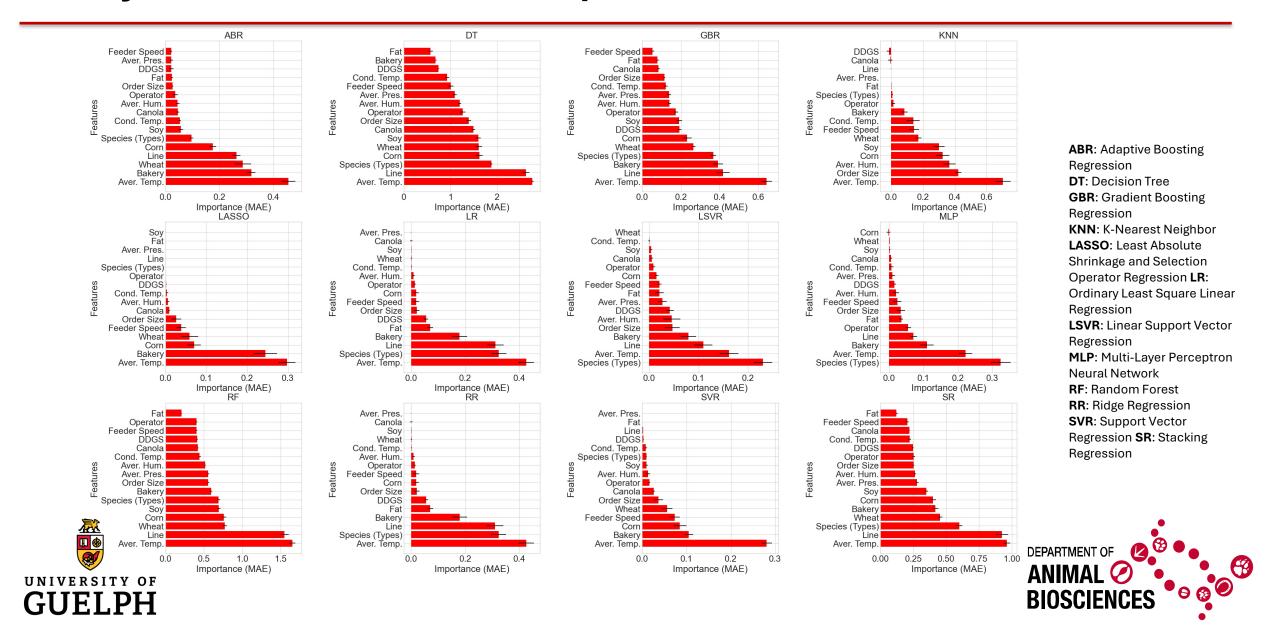
Study 2 – Results – Learning Curves





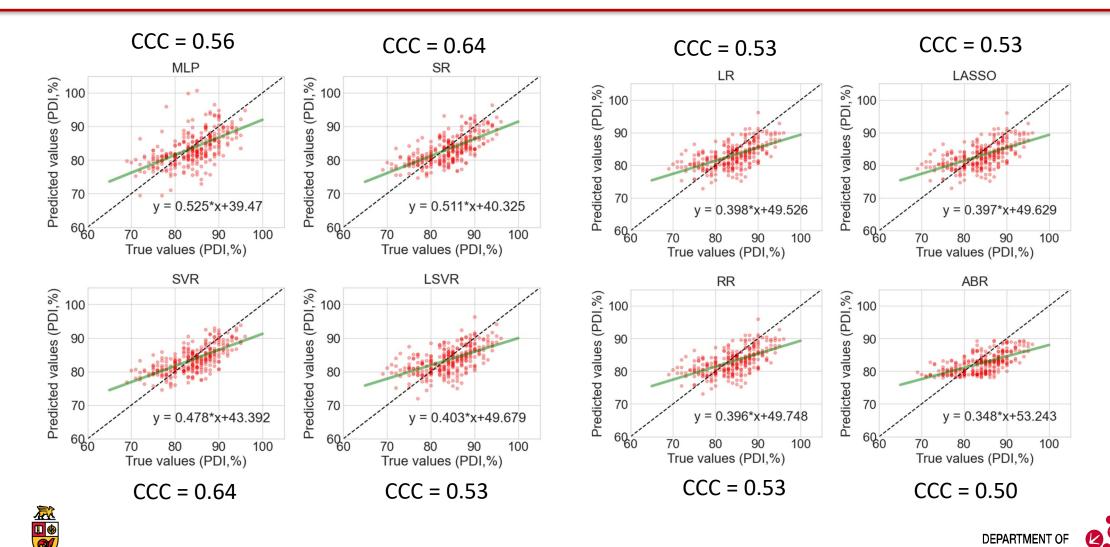


Study 2 – Results – Feature Importance



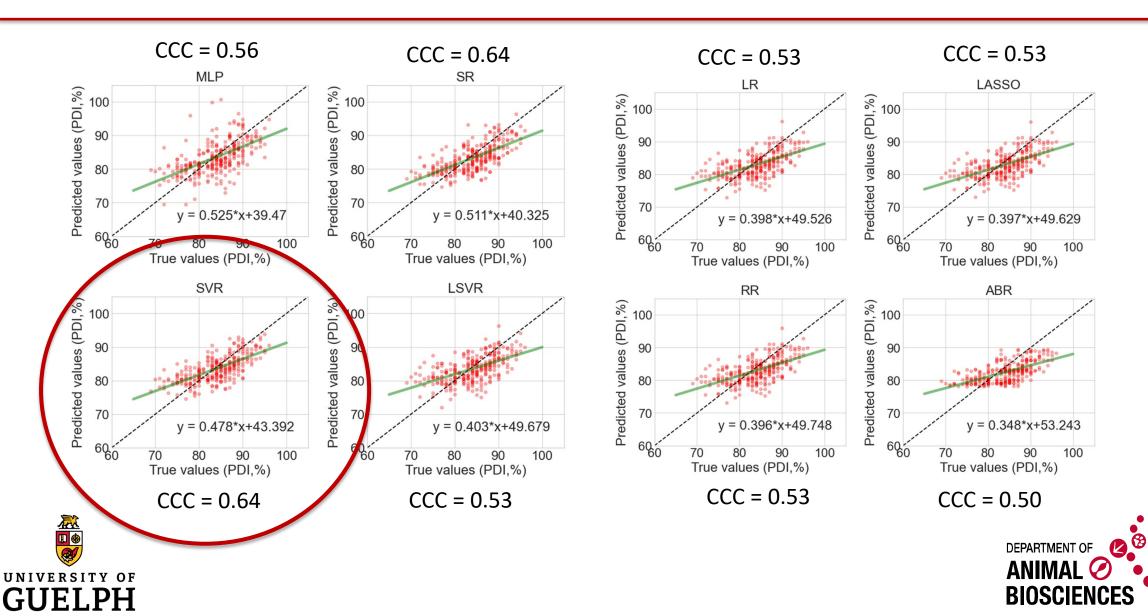
Study 2 – Results – Model Evaluation

UNIVERSITY OF GUELPH



ANIMAL

Study 2 – Results – Model Evaluation



Study 2 - Publication

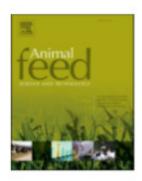
Animal Feed Science and Technology 293 (2022) 115443



Contents lists available at ScienceDirect

Animal Feed Science and Technology

journal homepage: www.elsevier.com/locate/anifeedsci



Using machine learning regression models to predict the pellet quality of pelleted feeds



Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada N1G 2W1

b Trouw Nutrition Canada, Guelph, ON, Canada



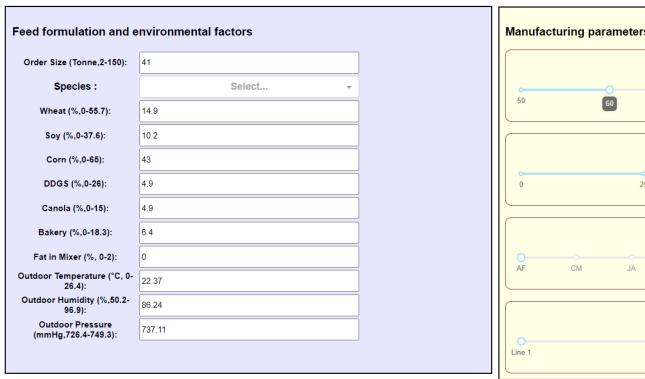


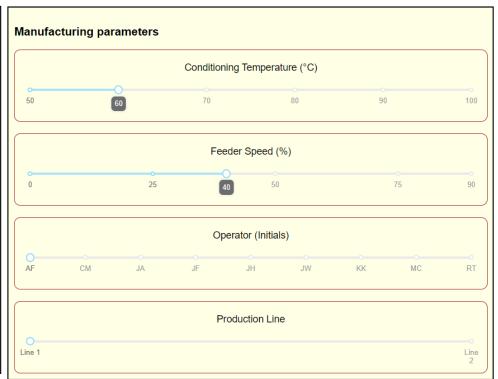


Study 2 - Application Development

PDI Simulator (Version 1.0)

Based on the research article: "You, J., Tulpan, D., Malpass, M. C., & Ellis, J. L. (2022). Using machine learning regression models to predict the pellet quality of pelleted feeds. Animal Feed Science and Technology, 293, 115443."





PDI (Pellet Durability Index, %) =





Study 2 – Lessons Learned

Hand collection and recording of data is not feasible long-term!





Study 3 - Data

Based on automated data capture (ENGIE) between 2021-2022, at Trouw Nutrition - St. Marys Plant 2

- Data included 2,691 observations and 76 variables (expanded manufacturing, nutrition and environmental data compared to Study 1)
- Data were split into two subsets:
 - 80%: For model construction (10-fold cross-validation),
 - 20%: For model evaluation
- Comparative research:

Study 3a: Statistical models (3 models with linear effects)

Study 3b: Machine learning models (12 ML models)



Study 3 - Data

Factors Considered:

- General information on the feed run (4 variables)
- Manufacturing parameters (11 variables)
- Usage of ingredients (41 variables)
- Environmental factors (12 variables)
- Nutrient composition of diets (8 variables)

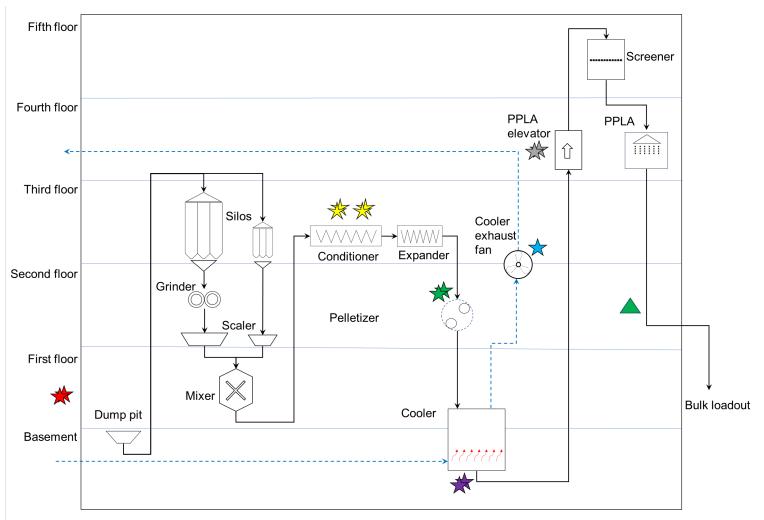
Collected by Engie

Recorded by data loggers

Provided by Trouw



Study 3 – Data – Temperature/Humidity Probes



Data loggers were located in the feed plant for ambient and indoor environmental factors



Black arrows with solid lines: The pipeline for ingredients or formulated feed

Blue arrows with dash lines: The pipeline for air

Stars: The locations of data logger installations (Red: Outdoor; Yellow: Conditioner; Green: Pelletizer; Purple: Cooler; Blue: Cooler exhaust

fan; Grey: PPLA elevator)

Triangle: Sample taken for PDI measurements



Study 3a – Results – Statistical Model

Variables selected for modelling in the best statistical model:

Predictor

ADF content (%)

Ambient humidity (%) ■

Amino acids (%)

Cumulative production (tonnes)

Dehydrated bakery meal (%)

Expanding temperature (°C)

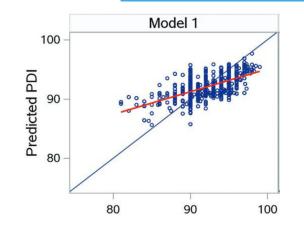
Fat content (%)

Indoor humidity (pelletizer) (%) 📥

Processing aid water (%)

Independent Model Evaluation (20%):

Measure	Model 1 ¹
MAE	1.93 ± 0.063
RMSPE	2.45 ± 0.079
MSPE	5.99 ± 0.389
ECT (%)	0.6 ± 0.83
ER (%)	0.1 ± 0.11
ED (%)	99.3 + 0.83
CCC	0.549 ± 0.0273
$C_{\rm b}$	0.891 ± 0.0181
R	0.616 ± 0.0222







Study 3 – Publication

Journal of Animal Science, 2025, 103, skaf021 https://doi.org/10.1093/jas/skaf021 Advance access publication 4 February 2025 Feeds





Prediction of Pellet Durability Index in a commercial feed mill using multiple linear regression with variable selection and dimensionality reduction

Jihao You,† Dan Tulpan,† Oheryl Krziyzek,‡ and Jennifer L. Ellis†,1,0

Abstract

Pellet quality, measured as the Pellet Durability Index (**PDI**), is an important key performance indicator for commercial feed manufacturing, as it can impact both mill efficiency and downstream performance of animals fed the manufactured diets. However, it is an ongoing challenge for the feed industry to control pellet quality, due to the complexity of feed manufacturing and the large number of variables influencing the process. Previous studies have explored the prediction of pellet quality using either simple empirical models with a few variables or machine learning models with many variables. The objective of the current study was to develop statistical regression models to predict PDI and to describe the





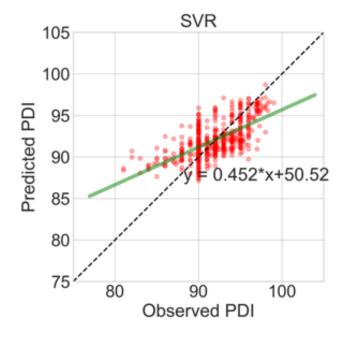
[†]Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada

[‡]Trouw Nutrition Canada, Guelph, ON, Canada

¹Corresponding author: jellis@uoguelph.ca

Study 3b – Results – ML Models

Independent Model Evaluation (20%):



SVR is the best model

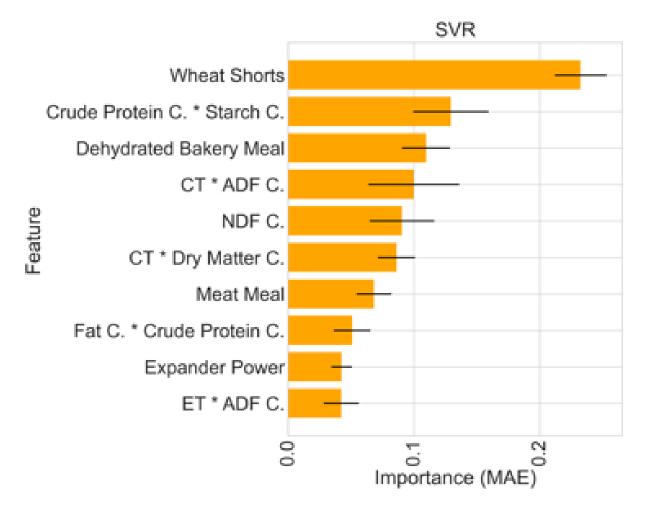
Measure	SVR	
MAE	1.869	
RMSE	2.358	
MSPE	5.562	
ECT	0.167	
ER	0.356	
ED	99.477	
CCC	0.608	
R	0.649	
C_b	0.937	





Study 3b – Results – ML Models

Variable importance ranking of SVR (the best model):



CT: Conditioning Temperature

ET: Expanding Temperature



Study 1, 2, 3a, 3b – Take Home Messages

Important variables for PDI prediction - across models:

- Fat, Protein, Starch Content in diet (%)
- Expanding Temperature (°C)
- Soybean Meal (inclusion %)
- Cumulative Production (since the change of pellet mill die)
- Indoor Humidity (around Pelletizer, %)
- Species
- Feeder Speed (%)
- Wheat shorts(%)
- Bakery (%)

Combination of nutrition, manufacturing and environmental factors

Is anything missing in data collection?

- Perhaps particle size
 - How 'reproducible' is the PDI measure (sampling error?)



...So what next?





Optimization of Pellet Quality using Machine Learning - 2.0

OMAFA Alliance Grant

UG-T1-2023-102187

(Jan 2024 - Dec 2025)



PDI 2.0 – Project Objectives

- To expand the current ML PDI prediction project to include multiple mills across several feed manufacturing companies,
- Development of a 'generalized model' for application across mills (and comparison to models developed on each mill separately)
 More variable data -> more robust model
- Development of an optimization algorithm (given a desired PDI, what inputs should change?)



Study 4 – MFS - Data

Objective: Develop an algorithm to predict both PDI and Duration (hrs):

Data: 5721 observations (May 5, 2022 – April 30, 2024)

- Manufacturing parameters (4 variables)
- Usage of ingredients (22 variables)
- Environmental factors (8 variables)
 Recorded by data loggers
- Nutrient composition of diets (7 variables)
 Provided by MFS





Study 4 – MFS – Model Development

Machine learning model construction using all features

80%: 5-fold crossvalidation,

20%: evaluation



Feature selection



Machine learning model construction using the selected features

80%: 5-fold cross-

validation,

20%: evaluation



Feature importance ranking



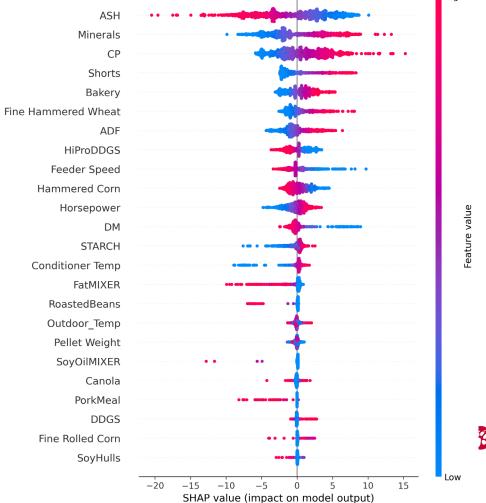


Study 4 – MFS – PDI Model Results

SVR is the best model for PDI prediction (No. of features = 41)

1				
	Measure	SVR		
		Cross-validation	Evaluation	
	RMSPE·(%)	5.006·±·0.162	4.794	
	MSPE	18.364·±·1.151	16.955	
	ECT·(%)	1.376·±·0.941	0.660	
	ER·(%)	0.315·±·0.392	0.027	
	EDP	98.31·±·0.784	99.314	
	CCC	0.731·±·0.015	0.758	
	PCC	0.757·±·0.016	0.785	
	C_b	0.965·±·0.009	0.967	

Beeswarm plot of SHAP values for PDI (No. of features = 24)



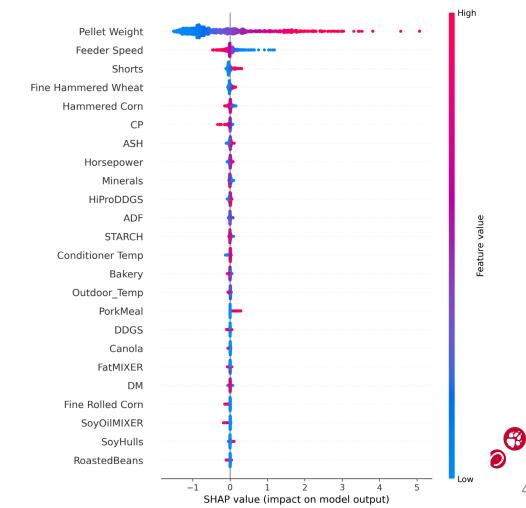


Study 4 – MFS – Duration Model Results

SVR is the best model for Duration prediction (No. of features = 41)

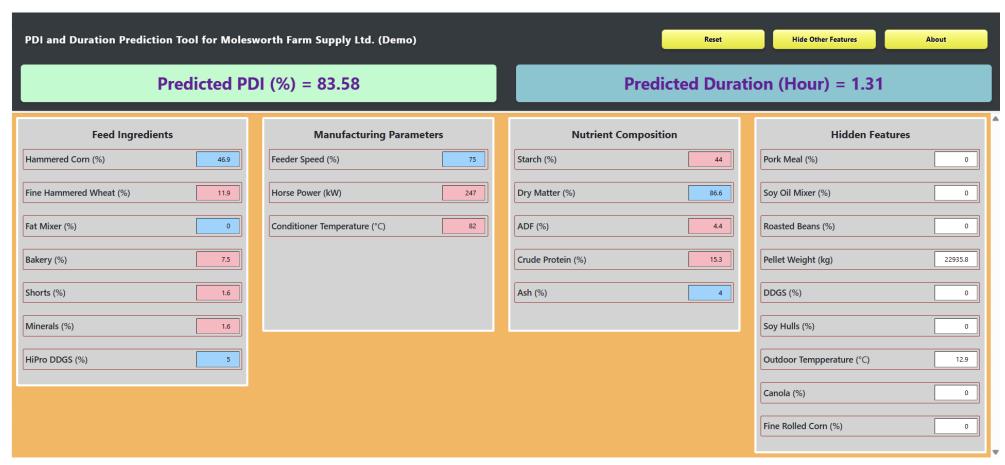
1							
Model	SVR						
	Cross-validation	Evaluation					
RMSPE·(%)	15.872·±2.607	11.470					
MSPE	0.074·±0.025	0.036					
ECT·(%)	0.176·±·0.159	0.335					
ER·(%)	1.954·±·1.658	0.282					
ED·(%)	97.87·±·1.768	99.383					
CCC	0.968·±·0.01	0.984					
PCC	0.97·±·0.009	0.984					
C_b	0.998·±·0.001	0.999					
	·						

Beeswarm plot of SHAP values for Duration (No. of features = 24)





Study 4 – MFS – Model Interface







Results Summary for PDI prediction (Thus Far)

Model	Model Type	No. Observations	Number of features	Top Influential Features	CCC ⁷
Meta- Analysis ¹	Statistical	280 Obs from 29 Studies	3	Protein(%), Lipid(%)	0.29
Model 1 ²	ML	1,434	16	Outdoor Temperature (DegC), Bakery(%), Wheat(%), Corn(%), Feeder Speed	0.64
Model 2a ³	Statistical	2,691	9	Expanding Temperature (DegC), Lipid(%), ADF(%), Indoor Humidity (at Pelletizer, %)	0.55
Model 2b ⁴	ML	2,691	74+	Bakery(%), Ambient Temp(DegC), CP(%), Starch(%), Conditioning Temp (DegC), ADF(%)	0.61
Model 3 ⁵	ML	5,721	41	Hammered Corn(%), Mixer Fat (%), Feeder Speed (%), Horsepower (Kw), Fine Hammered Wheat %)	0.76
Model 3b ⁶	ML	5,721	24	Ash (%), Minerals (%), CP (%), Shorts (%), Mixer Fat (%), Fine Hammered Wheat %), Bakery(%)	0.74

¹ Models not shown today



² TN ML model developed on hand-collected data

³ TN statistical model developed on automated data collection

⁴ TN ML model developed on automated data collection

⁵ MFS ML model – all features

⁶ MFS ML model – feature selection, refit

⁷ Independent evaluation

PDI Prediction

Can we do any better?





Mechanistic Modelling?

Prediction of particle size reduction using a hammer mill

(Rijpert et al., 2025)

Prediction of:

outflowing particle size distribution

Driving variables:

inlet mass feed rate

Product density

Hammer rotational speed

Mill residence time

Sieve aperture size

Feedstuffs:

corn, barley, faba beans, SBM, rapeseed meal, sunflower meal, wheat middling's, corn DDGS

A population balance model to predict size reduction using a hammer mill of eight commonly used feedstuffs in animal feed formulation: practical application and sensitivity analyses

Rijpert J.H.M.^{1,2*}, Bosch G. ¹, Illera-Rodriguez M. ³, Bastiaansen T.M.M.^{1,2}, Gerrits W.J.J. ¹

¹Animal Nutrition Group, Wageningen University & Research, The Netherlands

²Wageningen Livestock Research, Wageningen University & Research, The Netherlands

³Wageningen Food & Biobased Research, Wageningen University & Research, The Netherlands

Modelled components:

breakage kernel, breakage rate, screen selection functions (8 modelled parameters)

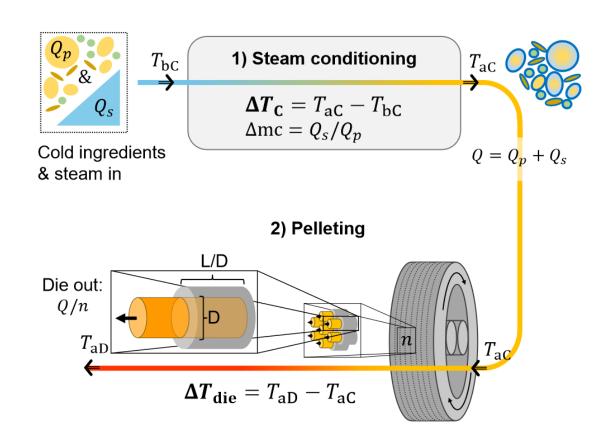
Outcome:

Sensitivity analysis revealed that mill residence time explained the majority of output variance.



Mechanistic Modelling?

"Under which processing conditions do loose pellet ingredients bond into rigid, durable pellets?" (Benders et al., 2025)



Conducted systematic pilot-scale extrusion trials, manipulating:

- Steam conditioning temperature
- Production (throughput) rate
- Die geometry and residence time

Introduction of "Stickiness Temperature" (T*)

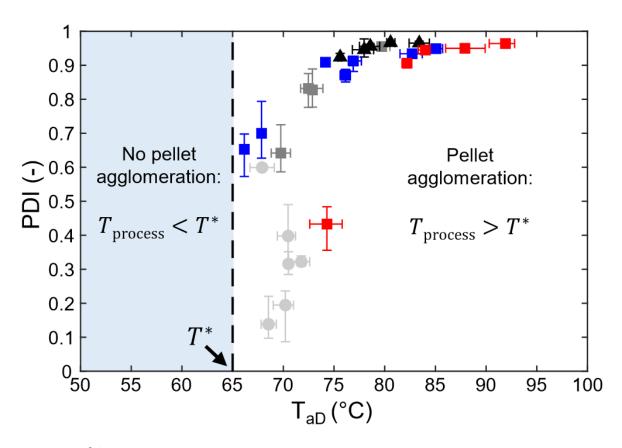
The critical onset temperature at which enough enthalpic activity occurs – typically from moisture, heat and friction – to enable bond formation within the pellet





Mechanistic Modelling?

"Under which processing conditions do loose pellet ingredients bond into rigid, durable pellets?" (Benders et al., 2025)



Pellet agglomeration does not occur below T*, the boundary condition for stickiness

T* - not a fixed universal value, but rather specific to each mash formulation

- Defined as: the lowest in-barrel temperature where:
 - Pellet durability and strength increased sharply
 - A visible structural transformation in the pellet occurred
 - Energy efficiency plateaued





Opportunity for Hybridization of Approaches?

- Introduction of new features to the ML models:
 - Delta Tc
 - Delta Td
- ML models could be used to define T* per mash?
- Mechanistic prediction of particle size -> input to ML model?





What Next?

- The 'generalized' model
- Model approach hybridization/combination
- LCA trade offs between the mill and farm (energy expenditure vs animal performance)
- Established data pipelines -> endless opportunities for **further data** analysis (in-house or in collaboration)



Challenges

- Data pipelines need to be established to allow analysis (but the data is there) - Hand collection of big data is not sustainable
- Some important information is not routinely collected (particle size?)
- Not all data is useful but we won't know unless we look?



Acknowledgements

Dr. Jihao You (UofG, PDF)

Dr. Dan Tulpan (UofG, Co-PI)

Dr. Cheryl Krziyzek (Trouw Nutrition)

Dr. Meeka Capozzalo (Molesworth Farm Supply)

Dr. Mark Malpass (MasterFeeds)

Mike Martin (Jones Feed Mill)













