

**UC Davis**

**UC Davis Electronic Theses and Dissertations**

**Title**

Precision Feeding Dairy Cow Pens by Predicting and Formulating for the Distribution of Dry Matter Intake

**Permalink**

<https://escholarship.org/uc/item/8w49j3p9>

**Author**

Lucey, Padraig

**Publication Date**

2024

Peer reviewed|Thesis/dissertation

Precision Feeding Dairy Cow Pens by Predicting and Formulating for the Distribution of Dry  
Matter Intake

By

PÁDRAIG LUCEY  
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Animal Biology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

---

Heidi Rossow, Chair

---

Sharif Aly

---

Timothy Hackmann

Committee in Charge

2024

Copyright © 2024 by Pádraig Lucey

## ACKNOWLEDGEMENTS

Firstly, to my advisor, Dr. Heidi Rossow. She planted the idea, encouraged, mentored me, and trusted in my process. This work is the result of her drive to challenge and improve animal nutrition modeling. Combined with long conversations on the history and beginnings of computer programming, all the way to state-of-the-art processing and advanced statistical learning, she grew my vision for this project from a small idea to its realization.

To the staff at the Veterinary Medicine Teaching and Research Center of UC Davis: Staci Slaght, who introduced and recommended me to Dr. Rossow on my first day, Dr. Terry Lehenbauer, for tirelessly writing and signing letters, and providing encouragement from residency all the way to PhD, and all others who make this center feel like a home. My first mentors at UC Davis: Drs. Deniece Williams, John Champagne, and Sharif Aly. They introduced me to large herd management, research and epidemiology. You all taught me to take no information for granted, the importance of getting the fundamentals right, and welcomed me as a friend.

I would also thank my parents, Pat and Eleanor Lucey, and all my siblings, who always believed in my ability, and supported my journey. Even when the pathway was unclear, you trusted my vision. My fellow Rossow lab students: Ashley Niesen, Morgan Wukadinovich, Leslie Jacobsen, Logan Widmer, Emily Meissner, Tyler Batchelder, Sammie Poldervaart, Randy Edwards, Oscar Martinez, Christine Sousa, Matthew Brazil and Julie Cardin, you all at one point were subjected to commenting on the readability of my most technical language. Your input, sacrifice, and friendship are highly appreciated.

Finally, to my wife, Allison Lucey. Your capacity for excitement and enthusiasm at each step of my PhD, research, and dissertation writing was inexhaustible. Thank you for encouraging me, believing in me, and waiting patiently for the time and work I needed to realize this success.

## ABSTRACT

Dairy cows are managed and fed in pens. These pens are constructed for uniform handling and management, and cluster cows on stage of lactation, reproduction status, or milk production. The variation in nutrient requirement of the cows is not accounted for, and different distributions of dry matter intake (**DMI**) occur depending on cow selection. These pens are fed by a ration supplied collectively. This ration is solved to meet a nutrient requirement and scaled to the number of cows in the given pen. But the solution cannot account for within pen variation because DMI is not individually recorded. The common measurement is the average nutrient or DMI per cow, dividing the quantity of ration consumed by the number of cows present. Another approach is predicting a DMI value for a hypothetical mean cow to represent the pen. Both systems result in one DMI value, rather than a distribution. The only scenario where this would be considered precise or accurate nutrition is if the cows within the pen were completely uniform in their requirements. Failing this, if the distribution of the pen were normal, and the tails symmetrical, the lower requirement cows could have their feed consumed by the higher requirement cows, striking a crude balance.

The first chapter of this work examined this assumption. Its hypothesis was the DMI of cow pens are not normally distributed, and the total pen DMI can be better calculated by describing it with an appropriate distribution shape. Cow DMI was first described by week of lactation for skew and kurtosis, then a large dataset of cow pens representing Fresh, High, and Low milk yield lactation stages were assembled from individual observations. The distribution of DMI for each pen was fit to the best distribution shape, then the total pen DMI was calculated from the area under its curve. This was compared to a second model that calculated the total

DMI by applying an empirical equation to the pen's mean values. The Beta distribution shape was the most common best fit of all pens, and the distribution shape predicted pen DMI with less than 1 % error, and more accurately than the comparison model, with an error of 11 – 22 %.

The second chapter predicted the distribution shape of DMI. In chapter I the distribution was demonstrated as a good calculation of pen DMI, but the shape is only known when described from individual DMI values. To build a prediction model in the case of unknown individual DMI, we trained machine learning algorithms to a known dataset. A large dataset of pens was assembled, and these were described with their best fit distribution, either the Beta or Generalized Normal distribution shape. Machine learning algorithms were trained to classify these pens to their best shape and predict DMI values that fit the distribution; the Distribution Shape Model (**DSM**). A second model was contrasted to this that predicted pen DMI by applying the mean descriptive statistics of the pen to an empirical equation, the NASEM equation model (**NSM**). Both models were validated for their DMI prediction performance on a naïve dataset and compared by model diagnostics. The DSM was very precise and accurate, with low error, or bias, and outperformed the NSM in consistently predicting pen DMI.

The last chapter of this work performed precision nutrition with the distribution shape of DMI for ration formulation. Three pens were considered: Fresh, High and Low. Each pen had its DMI predicted by the DSM, and by applying an empirical equation to each cow's characteristics, individual NASEM equation model (**iNSM**). These DMI values were used to formulate individual rations for every cow, and compared by cost, DMI, and metabolizable energy. As individual rations cannot be practically applied on dairies, precision nutrition pen rations were

also solved. Each pen had a nutrient requirement calculated by summing its constituent cow's requirements, and rations were solved with these totals. The DSM and the iNSM were compared to an ideal ration solved to the observed DMI values (**TRU**) to evaluate these two precision nutrition approaches. Two imprecise approaches were also considered. One where an empirical equation was applied to the mean of each pen, the average NASEM equation model (**aNSM**), and one where it was applied to the 75<sup>th</sup> of the pen, the adjustment factor NASEM equation (**fNSM**). In the individual, and pen ration solutions, the DSM was the only model that was not significantly different to the TRU. It was also cheaper and less nutrient wasting than all other ration models, as its solution was almost identical to the TRU for individual cows, and pen rations.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	ii
ABSTRACT.....	iii
LIST OF TABLES .....	ix
LIST OF FIGURES .....	xi
LIST OF EQUATIONS .....	xii
CHAPTER I. DESCRIBING THE DISTRIBUTION TYPE OF DRY MATTER INTAKE FOR DAIRY COW PENS BASED ON PEN CHARACTERISTICS .....	1
Abstract.....	1
Introduction.....	2
Materials and methods .....	4
Database description .....	4
Model description .....	6
Model results.....	8
Results and discussion .....	10
Database construction .....	11
Distribution of DMI by week of lactation.....	12
Distribution fitting by pen type.....	14
Total DMI prediction .....	16
Model assessment .....	18
Conclusion .....	19
References.....	20
Tables .....	25



Figures.....	31
<b>CHAPTER II. PREDICTING THE DISTRIBUTION SHAPE AND DRY MATTER INTAKE</b>	
<b>OF COW PENS WITH MACHINE LEARNING.....</b>	<b>33</b>
Abstract.....	33
Introduction.....	34
Materials and methods .....	36
Individual cow observations .....	36
Creation of pen training dataset based on individual cow observations.....	37
Distribution shape model - DSM .....	37
Predicting the distribution shape of DMI using DSM and NSM.....	39
Statistical Analysis.....	41
Results and discussion .....	43
Individual cow observations and pen assortment .....	43
Selection of distribution shape.....	44
DSM machine learning algorithm selection .....	45
DSM pen DMI prediction validation .....	50
NSM pen DMI validation and diagnostics.....	51
Comparison of DSM and NSM.....	52
Conclusion .....	54
References.....	55
Tables.....	59
Figures.....	64
Appendix I .....	68

CHAPTER III. PRECISION RATION FORMULATION BY PREDICTING THE DISTRIBUTION OF DRY MATTER INTAKE OF LACTATING COW PENS .....	73
Abstract .....	73
Introduction.....	74
Materials and methods .....	76
Pen selection and description.....	76
Ration formulation program development.....	77
Ration solutions .....	80
Statistical analysis.....	80
Results and discussion .....	82
Pen DMI description and diagnostics of the distribution predictions.....	82
Individual cow ration solutions.....	83
Pen ration solutions.....	85
Conclusion .....	87
References.....	88
Tables .....	90
Figures.....	97
CHAPTER IV. CONCLUSION .....	104

## LIST OF TABLES

<b>Table I-1.</b> Description of the trial datasets used in the database of this dairy cow study .....	25
<b>Table I-2.</b> Skew, kurtosis and Shapiro-Wilk test of goodness of fit to normal distribution for dry matter intake values of all dairy cows by parity .....	26
<b>Table I-3.</b> The proportion of best distribution fit for 100 replicates of each dairy cow pen class .... .....	28
<b>Table I-4.</b> DMI predictions by best distribution type and by NASEM equation of all pen replicates of each dairy cow pen class and model fit statistics .....	29
<b>Table II-1.</b> Description of the cow records from the separate datasets used to build the database of this study.....	59
<b>Table II-2.</b> Performance of the XGB algorithm with different feature sets by 5-fold cross validation.....	60
<b>Table II-3.</b> Model performance of the hyperparameter tuned XGB algorithm with both testing and validation data. ....	61
<b>Table II-4.</b> Feature importance of the three XGB1 trained and hyperparameter tuned models. ..	62
<b>Table II-5.</b> Model diagnostics of the Dry Matter Intake prediction of both the distribution shape model and the NASEM equation model when compared to observed. ....	63
<b>Table III-1.</b> Description of pens .....	90
<b>Table III-2.</b> Ingredient nutrient composition, inclusion, and cost .....	91
<b>Table III-3.</b> Model diagnostics of the Dry Matter Intake prediction of both the distribution shape model (DSM) and the NASEM equation model (iNSM) when compared to TRU.....	92
<b>Table III-4.</b> Values of DMI, ME and ration cost for the Welch’s ANOVA and pairwise comparison between the TRU, DSM, and iNSM .....	93
<b>Table III-5.</b> Fresh pen ration solutions .....	94

<b>Table III-6.</b> High pen ration solutions .....	95
<b>Table III-7.</b> Low pen ration solutions .....	96

## LIST OF FIGURES

<b>Figure I-1.</b> Line plots of the mean and median of (Figure A) Milk Yield, (Figure B) DMI and (Figure C) feed efficiency of dairy cows by week of lactation and dataset separately. ....	31
<b>Figure I-2.</b> Scatter plots of observed versus predicted model diagnostics for each of the seven dairy cow pen types for both the Distribution and NASEM model.....	32
<b>Figure II-1.</b> A confusion matrix of the actual and predicted classification classes of the Extreme Gradient Boosted (XGBoost) trained and hyperparameter tuned model.....	64
<b>Figure II-2.</b> A confusion matrix of the actual and predicted classification classes of the Extreme Gradient Boosted (XGBoost) trained and hyperparameter tuned model for the validation data..... .....	65
<b>Figure II-3.</b> Observed versus Predicted plots of Dry Matter Intake by pen plotted for the Distribution Shape Model (DSM), and the NASEM Equation Model (NSM), for a) pens classified as Beta distribution, and b) pens classified as Generalized Normal Distribution.....	66
<b>Figure III-1.</b> Kernel density plots of the observed (TRU), Distribution Shape Model (DSM), and NASEM Equation Model (iNSM) dry matter intake values for the a) Fresh pen, b) High pen, and c) Low pen .....	97
<b>Figure III-2.</b> Regression plots of the mean squared error of prediction of observed versus predicted dry matter intake values for the Distribution Shape Model (DSM) and the NASEM Equation Model (iNSM) for a) the Fresh pen, b) the High pen, and c) the Low pen .....	100
<b>Figure III-3.</b> Violin plots of the ration solution DMI, ME and cost for the observed and predicted DMI constraint inputs of a) the Fresh pen, b) the High pen, and c) the Low pen .....	103

## LIST OF EQUATIONS

<b>Equation I-1.</b> Calculation of 3.5% Fat Corrected Milk from Milk Yield and Milk Fat .....	4
<b>Equation I-2.</b> Calculation of Feed Efficiency from Fat Corrected Milk and Dry Matter Intake....	5
<b>Equation I-3.</b> Analysis of variance to compare the outcomes of Milk Yield, Dry Matter Intake, and Feed Efficiency on the levels of dataset, and treatment group within dataset .....	6
<b>Equation I-4.</b> Prediction of Dry Matter Intake based on cow Parity, Milk Energy, Bodyweight, Body Condition Score, and Days in Milk.....	9
<b>Equations I-5.</b> Calculation of percentage error between predicted and observed values.....	9
<b>Equation I-6.</b> Calculation of mean squared prediction error from the mean error and variance..	10
<b>Equation II-1.</b> Prediction of Dry Matter Intake based on cow Parity, Milk Energy, Bodyweight, Body Condition Score, and Days in Milk.....	41
<b>Equation II-2.</b> Calculation of milk energy based on fat and protein percentage.....	41
<b>Equation II-3.</b> Calculation of classification accuracy .....	42
<b>Equation II-4.</b> Calculation of classification precision.....	42
<b>Equation II-5.</b> Calculation of classification recall.....	42
<b>Equation II-6.</b> Calculation of classification F1-score .....	42
<b>Equation III-1.</b> Objective function of the ration formulation linear programming.....	77
<b>Equation III-2.</b> Metabolizable energy calculated from Digestible energy and ether extract of feed stuffs.....	78
<b>Equation III-3.</b> Bounded constraint equation of Dry Matter Intake in the ration formulation linear programming solution.....	78
<b>Equation III-4.</b> Minimum and maximum nutrient proportion constraint equation in the ration formulation linear programming solution .....	79

**Equation III-5.** Minimum and maximum constraint equation for the proportion of feeds as forage in the ration formulation linear programming equation .....79

## CHAPTER I.

### DESCRIBING THE DISTRIBUTION TYPE OF DRY MATTER INTAKE FOR DAIRY COW PENS BASED ON PEN CHARACTERISTICS

*P.M. Lucey and H. A. Rossow, animal, Volume 17, 2023*

#### ABSTRACT

In practice cows are fed by pen, but a diet is formulated to the nutrient requirements of a single cow. If the dry matter intake (**DMI**) of a pen were equal for all cows this approach would have no error, but cows are grouped into pens on pregnancy and other management factors creating a distribution of DMI. The goal of precision feeding is to meet the requirements of individual animals to increase efficiency and reduce environmental impact but is not achieved when a group is fed as if the individuals have uniform requirements, and the DMI distribution is not normal. The hypothesis of this work was that the DMI of cow pens are not normally distributed and the total DMI from the best fit distribution shape for a cow pen would have lower percentage error to the observed DMI than a prediction of a single DMI that is fed at a uniform level and assumes a normal distribution. Our objective was to describe the distribution shape of DMI by week of lactation, and for different pen types. Pens were generated by randomly assorting cows by week of lactation from a database into different categories of pen for size and lactation period. These pens were fitted to the best distribution type, and its parameters were used to randomly generate distribution plots that predict the total DMI for each pen. A second predictive model estimated the DMI of each pen using an empirical equation of DMI that was multiplied by the number of cows in the pen to represent feeding of a uniform DMI quantity. The percentage error for the distribution shape model was significantly lower than the empirical model with pen errors being less than 1 %. The Beta distribution type was the most common distribution to best represent the data of pen DMI. Describing the distribution and using it to predict a total pen DMI provided accurate estimates of feed quantity for a group. Reducing error



by using the distribution of DMI for feed formulation, instead of the nutrient requirements of an individual animal can provide a precision nutrition approach to group feeding.

## **INTRODUCTION**

In practice, cows are fed by pen, and a diet is formulated to a single estimate of DMI that represents the pen's mean cow. This is scaled to the number of cows in the pen and a uniform quantity of feed per cow is provided. If the DMI of a pen were constant, there would be zero feed quantity error in this approach. However, if DMI is not constant, the mean is not an appropriate parameter. It is accepted in the dairy industry that milk production and DMI are not constant within cow pens. Feeding cows uniformly, consistently underestimates feed quantity. Then, mathematical corrections are employed to ensure sufficient feed to high production cows, such as feeding to the 83rd percentile nutrient requirements of the pen (Stallings and McGilliard, 1984), or over-formulation where the estimated DMI quantity is increased by a fixed proportion (Weiss, 2019). Precision feeding of dairy cows requires accurate estimates of the pen feed quantity, as the goal of precision feeding is to minimize nutrient waste through minimizing leftover feed. But pen level management prohibits individual cow feeding (Schulze, Spilke, and Lehner, 2007), as diets formulated for an individual do not meet the requirements of a group.

Precision feeding addresses between-animal variation to deliver the correct amount of feed to each animal and can reduce nutrient excretion, feed cost, environmental impact and increase nutrient efficiency (Bewley, 2010; Pomar et al., 2011; Capper and Cady, 2020). To employ precision feeding in a pen, the distribution of DMI by pen cannot be represented by a constant as cows are not uniform across a pen. For example, cows are sorted into pens based on

pregnancy status, milk production level (low and high) and nutrient efficiency (St-Pierre and Thraen, 1999). Variation in DMI is affected by milk yield (45 %), feed management (22 %), bodyweight (**BW**) (17 %), climate (10 %) and body condition score (**BCS**) (6 %) (Roseler et al., 1997). Grouping cows by these same factors will create a distribution of DMI. The current assumption of uniformity when formulating pen diets does not address between-cow variation. There is concern in the applicability of individual cow studies to commercial operations where cows are grouped in pens (St-Pierre, 2007). Empirically derived predictions of DMI and nutrient requirements for an individual cow, such as Nutrient Requirements of Dairy Cattle (NASEM, 2021) or the Cornell Net Carbohydrate and Protein System (Van Amburgh et al., 2015), assume all cows in a pen eat the same amount, and do not consider how distribution of DMI differs for groups of cows at different production levels.

The hypothesis of this chapter is that feeding a uniform DMI does not represent the distribution shape of DMI for cow pens, and the total DMI quantity for a pen calculated as the density of the best fitting distribution shape will be closer to the true DMI when compared to the empirically calculated DMI for the mean cow of the pen. Using a database of individual cow DMI by week we first described the skew and kurtosis of DMI of cows when grouped by week of lactation, then generated sample pens of cows within a range of weeks of lactation, described the distribution type and parameters, and predicted a total pen DMI. This was compared to the observed total DMI. The total pen DMI was also predicted by applying the NASEM 2021 DMI equation to the mean of the pens and compared to the observed DMI. Both prediction models were contrasted by their performance to the observed.

## MATERIALS AND METHODS

No approval from the University of California, Davis, Animal Care and Use Committee was needed for this study as it was conducted using only previously collected data.

### *Database description*

A database was constructed from datasets of nine research trials for this mathematical modelling study. All trials were nutrition or management-based interventions on lactating cows and seven of the nine have been published in peer reviewed journals. Each trial included data for different ranges of the lactation period, with the recording period starting at 1 wk (first 7 d immediately after parturition) and extended up to 44 wk (Table I-1). All cows had individual milk yield and components recorded at each milking daily, and DMI recorded daily using Calan gates (American Calan, NH, USA). Data were averaged by week per cow for 426 cows in total.

### *Inclusion criteria*

Every dataset needed to list individual records by unique cow identification number, identify treatment assignment, parity, diet composition, days in milk (**DIM**), weekly milk yield, weekly milk fat and weekly DMI consecutively for at least 4 wk. Some of the datasets also included milk protein, bodyweight, and BCS, but these were not required for inclusion.

### *Data cleaning*

Cows with missing values for some of their weekly observations were not removed. Datasets were imported to a SQL database using an open database connector pipeline. Within the database, a variable for fat corrected milk was created using  $3.5\% FCM = (0.4324 * kg) +$

( $16.216 * \text{milkfat}$ ) (Tyrrell and Reid, 1965; Eq. I-1). We used fat corrected milk instead of energy corrected milk because one dataset did not record milk protein values. A variable for feed efficiency (**FE**) was created using the formula  $FE = FCM/DMI$  (Korver, 1988; Eq. I-2). We used FE to compare the metabolic efficiency of cows across trials despite differences in diet and environment to determine if data were reasonable and if these cows could hypothetically exist on a single dairy and be grouped together into pens.

#### *Database variance assessment*

All datasets were compared for similarity of variance in DMI, milk yield and FE. The variance of FE by dataset and treatment group within dataset was used as the main measure of comparability, as this represents metabolic efficiency and allows the comparison of cows with different levels of feed intake and milk yield. Data were retained in the overall database if 50% of the FE data overlapped with all other datasets on boxplot analysis, and less than 20% of the weeks of lactation had a FE mean significantly different from other datasets on analysis of variance (**ANOVA**) analysis. The database was imported to R 4.2.1 (R Core Team, Vienna, Austria, 2018) and milk yield, DMI, and FE were examined visually using scatter plots to identify and remove outliers and assess the range of the three variables (not shown). The mean and median of milk yield, DMI and FE was plotted by week for each dataset (Figure I-1). Boxplots were examined at every week by dataset, and treatment within dataset, for visual comparison of the mean and variance of milk yield, DMI and FE (not shown). We used ANOVA models with the R 4.2.1 base package to compare the means of the variables of milk yield, DMI and feed efficiency with cow as the unit of interest. Data were first evaluated using treatment group within dataset as the subgroup for the outcomes of milk yield, DMI and FE. Next the

database was evaluated by ANOVA using dataset within database as the subgroup for the outcomes of milk yield, DMI and FE. Using the equation  $Y_{i,j} = \mu + \beta_i + \epsilon_{i,j}$  (Eq. I-3), where  $Y_{i,j}$  is the  $j$ -th observation of the  $i$ -th group ( $i=1, 2, \dots, 9$  datasets or  $i = 1, 2, \dots, 35$  treatment within dataset),  $\beta_i$  is the  $i$ -th subgroup effect and  $\epsilon_{i,j}$  is the random error present in the  $j$ -th observation on the  $i$ -th treatment of these fixed effect models.

### ***Model description***

#### *Skew and Kurtosis by week of lactation*

The skew, kurtosis and Shapiro-Wilk test for normality (Shapiro and Wilk, 1965) was calculated based on DMI data for each week of lactation separately starting at 1 wk to 44 wk using the Moments package of R 4.2.1 (Komsta, L., Novomestky, F., 2022). This was performed on all cows in the database, then with primiparous cows only, and then multiparous cows (Table I-2). The null hypothesis of the Shapiro-Wilk test is that the data comes from a normal distribution. The null hypothesis was rejected at a significance level of  $P < 0.05$ .

#### *Creation of pens to simulate DMI distributions*

The database was used to generate multiple cow pens for a set number of cows within a range of weeks of lactation. Three pen types were created: fresh, high and low. Fresh pens were lactation range 1 – 3 wk, high range 4 – 18 wk and low range 19 – 44 wk. We also generated a large and small dairy pen size for each type. We chose pen size and lactation range as representative of management in the California dairy industry (M. Wukadinovich, personal communication, November 2022). Fresh pens were set at 20 (**FRESH20**) and 60 cows (**FRESH60**), high at 50 (**HIGH50**) and 200 (**HIGH200**), and low at 50 (**LOW50**) and 150

(**LOW150**). One additional high pen was created with 50 primiparous cows only (**HIGHP50**), resulting in seven pen types (Table I-3). To create replicates of each pen type, data was prepared in longform with one row per cow per week of data observation. The data were subset into the week of lactation ranges for each pen type, and 100 unique pens for each type were generated by selecting one weekly observation of DMI per unique cow within that week range. This pen generating code was nested in a “FOR” loop with a changing random seed number in R 4.2.1 for 100 replications. Observations were replaced between each pen generation. This resulted seven sets of 100 unique pen replicates via bootstrapping. Each bootstrapped pen contained unique weekly DMI observations by cow within the lactation range for that pen, no cow occurred at multiple time points as each pen represents a temporal snapshot of potential grouping on a dairy. Each replicate selected a random weekly DMI observation per cow from the data pool available for the pen constraints, and uniqueness comes from the diversity of arrangements of different cow week observations, and different cows grouped within pens together. Bootstrap selection restores the data pool between each replicate, and random selection of DMI observations provides the potential for many unique pens. All weekly observations included DMI and milk information, cows with missing information of BW and BCS were included in pen generation, and their missing information was excluded when calculating summary statistics of these variables.

#### *Best distribution type fitting of DMI*

All 100 replicates of each seven pen types were fit to the best probability distribution type using the “ExtDist” package in R 4.2.1 (Wu, Godfrey, and Pirikahu, 2020). The distributions Normal, 4-parameter Beta, Weibull and Logistic were selected by preliminary

visual assessment of the distribution shape of DMI of replicated pens, then tested with maximum log likelihood estimation. The value closest to zero represented the best fit to the data of the four tested distributions. For each of the 100 replicates of the seven pen types, the best distribution type to the DMI data of that pen replicate was calculated by this method. The number of occurrences of best fit for each tested distribution type were summed for each pen type to calculate the proportion of each's 100 replicates that were best fit by each distribution candidate. The skew and kurtosis values of every replicate were calculated, and the mean of both values for each pen type was reported (Table I-3). This method described the distribution of DMI for multiple randomly generated pens to report the most common distribution shape that best fits these pen types. The observed total DMI for each of the 700 replicated pens was known and calculated as the sum of observed DMI values per pen.

### ***Model results***

#### *Predicting total DMI for a distribution shape*

A total DMI for every replicate of the seven pen types was predicted using its best fitting distribution. This distribution type had appropriate parameters for the data it described calculated using the function for that distribution in the R 4.2.1 package "ExtDist" by maximum likelihood estimation (Wu et al., 2020). For each distribution and its parameters, we randomly generated DMI values that fit the shape of the curve ("rBeta", "rNormal", "rLogistic", "rWeibull" functions, base R 4.2.1). The sum of these values was considered the prediction of DMI for each pen for that fitted distribution shape. This random generation and summing of DMI values was performed 1000 times for each pen replicate, then the mean value of total DMI for the 1000 iterations was taken as the final pen DMI estimate. This was performed because of variance

between the randomly predicted values for a given distribution's parameters, and to standardize the prediction of total DMI across all replicates. (Table I-4).

#### *Predicting total DMI by empirical equation*

The descriptive statistics of each of the 700 generated pen replicates of the seven pen types were used to calculate the predicted DMI by the NASEM 2021 equation  $DMI(kg/d) = [(3.7 + Parity \times 5.7) + 0.305 \times MilkE(Mcal/d) + 0.022 \times BW(kg) + (-0.689 + Parity \times -1.87) \times BCS] \times [1 - (0.212 + Parity \times 0.136) \times e(-0.053 \times DIM)]$  (de Souza et al., 2019; NASEM, 2021; Eq. I-4) in R 4.2.1. The mean values of parity, milk energy (NRC, 2001), BW, BCS, and DIM for each pen replicate were used as the input values. The predicted DMI was multiplied by the n of cows in that pen replicate to produce a total DMI of the pen at an assumption of uniform DMI.

#### *Comparing predicted DMI to observed and model diagnostics*

Every pen had predicted DMI compared to the observed DMI for percentage error with the formula  $PercentageError = (Value_{observed} - Value_{predicted}) / Value_{observed} \times 100$  (Eq. I-5). Observed total DMI was considered the sum of the observed DMI values of every cow in each pen replicate. The predicted total DMI for a distribution shape, and the predicted total DMI by empirical equation of every pen were both compared to the observed total DMI for their percentage error of total pen DMI prediction. The mean error for each of the seven pen types by both methods of prediction was compared for a difference in means by student's t-test at a significance level of  $P < 0.05$  (Table I-4). Both prediction models, distribution fitting vs. NASEM, were compared to the observed data for their model fit to the data using mean squared



prediction error (**MSPE**) ( $MSPE = ME^2 + VAR$ , **ME** = mean error, **VAR** = variance; Eq. I-6) using Microsoft Excel (Microsoft, WA, USA) and partitioned into error due to mean, slope and random bias as described by Benchaar et al., 1998 and Theil, 1966).

## **RESULTS AND DISCUSSION**

The goals of this modeling study were to create a pool of cows that could be assembled into virtual pens that represent possible grouping of cows as may occur on a commercial dairy, determine the best fit DMI distribution, and then estimate DMI of the pen using the distribution. With bootstrap sampling, we generated virtual pens by selecting a different weekly observation per unique cow and assembling them into pens of cows for a range of weeks of lactation. To observe the distribution in DMI of a pen and represent how this truly may occur on a dairy, it was important to examine many possible cow groupings for the consistency of distribution shape and estimate the group level variables that defined the distribution. Because dairies are composed of many different cows, variance between cows in the dataset was desired. But to avoid generating pens of cows that would not realistically be grouped together, FE was used to compare metabolism between cows. Cows are most commonly grouped by reproductive status, lactation stage and milking performance (Contreras-Govea et al., 2015), so we assumed FE should be comparable across the database as a measure of the cow's ability to produce milk from feed, and to avoid skew by introducing an artifact from combining metabolically dissimilar cows.

## ***Database construction***

### *Inclusion Criteria*

Three of the nine datasets were different ( $P < 0.05$ ) in milk yield or DMI to all others due to the intervention in their publication. All datasets had overlap of at least 50 % of the data on boxplot analysis by treatment group. Results of the ANOVA analysis showed differences between treatment groups for multiple datasets at occasional weeks of lactation. But there were no differences observed for a period of three or more consecutive weeks, and differences did not occur for more than 20% of all weeks of lactation of the trial. Therefore, all treatment levels of each dataset were retained. Datasets were compared to each other visually using boxplots and statistically by ANOVA with dataset within database as the level for the means of DMI, milk yield and FE. Dataset seven had consistently lower DMI and milk yield than other datasets, and the range of these did not overlap by at least 50 % of the data of other datasets at all weeks of lactation. This dataset also had repeated weeks of milk yield, DMI and FE that were different from other datasets ( $P < 0.05$ ). All observations from dataset seven were removed from the database, resulting in a total of 375 cows with 8,982 weekly DMI observations retained and used in this study.

### *Database variance assessment*

Bootstrapping of pen groups allowed the generation of many virtual cow pens, but as the pen size approached the total number of unique cows available, or number of replicates increases, the variation between replicates would diminish. To generate reasonable pens from this database, a high tolerance for variation in milk yield and DMI within and between datasets was accepted, including reported significant differences by the published studies. Dairies

intentionally group cows of similar milk production levels, and low milk yield accounts for 21 % of all cow culling removal (USDA, 2014). To satisfy that the cows in the database could be grouped together on a dairy, metabolic efficiency as FE, and the overlap in the variance of milk yield and DMI was used to validate that generated virtual pens are reasonable, as dairies will group cows to reduce variance in these variables.

### ***Distribution of DMI by week of lactation***

Week of lactation was the grouping factor for this study and is an important determinant of DMI as energy demand of milk production changes with DIM. To understand how this contributes to the distribution of DMI for pens, we first described DMI distribution at each week of lactation. Skewness is an asymmetry in a distribution, and a normal distribution is considered to have no skew. Positive skewness indicates the mass of the distribution is shifted to the lower value of the variable with a right tail (right skew), and the median is lower than the mean. Negative skewness indicates the mass of the distribution is shifted to the higher value of the variable with a left tail (left skew), and the median is higher than the mean. Kurtosis is a measure of the data in a distribution's tails. The normal distribution has a kurtosis of three, with higher values indicating more data in the tails.

Cows were first examined together, then the primiparous and multiparous separately (Table I-2) for skew, kurtosis, and Shapiro-Wilk test of normality for DMI. The best distribution shape fit was not presented for these data as cows would not be grouped in pens or fed by individual week of lactation on a dairy. For all cows together: 2 – 10 wk of lactation were non-normal except for 6 wk. Normality was observed from 11 – 35 wk for these cow groups at

almost every week, then non-normality was again observed occasionally between 37 – 44 wk. Almost all non-normal weeks were left skewed. For primiparous, non-normality was observed at 20 and 28 wk with left skew. For multiparous, non-normality was observed for the fresh and early high period (2 – 6 wk), and 16 – 23, 36, 40 and 43 wk. For all three animal groups (mixed parity, primiparous, and multiparous) for most weeks of lactation the DMI distribution shape was negatively skewed (left skew).

Under the central limit theorem, we expect the random variable of DMI at each week of lactation would assume a near-normal distribution for a large sample size. This was only observed when looking at primiparous cows. For all cows during the early lactation period the effects of parturition on energy balance, metabolism and immunity (Pascottini, Leroy, and Opsomer, 2020) may be creating a non-normal distribution of DMI. This distribution in the early lactation period is not considered in empirical prediction equations of DMI. The NASEM 2021 and Cornell Net Carbohydrate and Protein System are both calculated using observed data of cows after their peak lactation milk yield. Normality was only observed for early lactation primiparous cows. Primiparous cows may be more uniform in their milking potential and DMI as these animals still have growth energy demands and lower incidence of production diseases or intramammary infections than older animals. Further non-normality was observed in the multiparous cows after the milk yield peak (6 – 9 wk, Figure I-1), and this distribution may be introduced by factors such as the age diversity, pregnancy status, disease history or management differences.

The mean and median of all cows were presented by week for the variables milk yield, DMI and FE (Figure I-1). There was a rapid increase in DMI and milk yield across the early lactation period. This difference in DMI level early in lactation will impact the DMI distribution of fresh pens. Later in lactation we see a gradual decline in DMI. Pens at this period of the lactation usually include cows across a wider range of weeks, and the number of cows at each week will influence the DMI distribution for a pen. The median and mean lines of Figure I-1 deviate when skew occurs, as the mean is drawn away from the median by outliers. These factors, the construction of pens across a range of the lactation, and skew in DMI within cows at the same week, both highlight the concern in assuming uniformity of DMI for a pen.

### ***Distribution fitting by pen type***

One hundred replicates of the seven pen types were randomly generated and each one was fit to the best distribution. The 4-parameter Beta was the best distribution for all the pen types. For pens FRESH20 and FRESH60 the Beta was the best fit for 85 and 87 % of replicates respectively. For HIGH50, HIGHP50 and HIGH200 the Beta was the best fit for 77, 80 and 60 % of replicates respectively, and for LOW50 and LOW150 it was the best fit for 62 and 36 % respectively (Table I-3). The shape and scale parameters of the Beta distribution allow it to accommodate right and left skewness in its shape and made it the best distribution to represent these data. For FRESH20 and FRESH60, the mean skewness of all the replicates were close to 0, indicating an even number of positive and negative skew values in the replicates, and the mean kurtosis was less than three, describing a platykurtic distribution with flat tails and small outliers. The HIGH50 and HIGH200 pens both had negative mean skewness, indicating a left sided tail, and mean kurtosis close to three, describing tails similar in weight to a normal distribution. The

HIGHP50 pen had a negative mean skew value and normal mean kurtosis. Both LOW50 and LOW150 had a mean negative skewness and mean kurtosis values higher than three. This positive excess kurtosis indicates a leptokurtic distribution with heavier weighted tails, describing large outliers. This may be due to the largest week range of the LOW50 and LOW150 pens at 19 – 44 wk. The skew values of all pen types led to the beta distribution as the best overall fit. The pens FRESH20 and FRESH60 had a mean skewness close to zero, but when examining each replicate separately pens were equally dispersed as either heavily negative or heavily positive in skewness (not shown), resulting in a mean skewness close to zero.

This demonstrates that in the early lactation fresh cow pens may have non-normal distributions, with outliers to either the right or left side, depending on the cows in the group. Pens later in the lactation, HIGH50, HIGH200, LOW50 and LOW150 were consistently left skewed and best described by the Beta distribution as it accommodates this spread of data. The proportion of replicates best described by the Normal was higher for HIGH200, LOW50 and LOW150 than other pens. The HIGH200 pen may move closer to the Normal as the n of this pen was the largest, but Beta was still a better distribution fit for most of these replicates. Pens LOW50 and LOW150 had the Normal as the best fit for 10 and 29% respectively, and the Logistic distribution as best fit for 24 and 34 % of replicates respectively. These pens were spread across the longest lactation range, and at the lower milk production period, when most cows that remain in the herd would be pregnant and at a lower intake level so the decline in DMI is less steep for these weeks. The HIGHP50 pen was still described best by the Beta distribution but had a skewness and kurtosis close to the Normal. This pen only consisted of primiparous cows, and as only three of the eight final datasets in the database included primiparous, this pen

had less variation by DMI. But the Beta distribution was still most appropriate, indicating that skewness was an important component of these pen groups at that lactation range.

The description of these pens as a Beta distribution indicates that uniform feeding of cows within a pen to a fixed DMI value is not appropriate. When a DMI value is estimated for a hypothetical cow of a pen and that number is multiplied by the number of cows, not all cows are eating at the same level. There is an assumption that the symmetric tails of a normal distribution will correct for this feeding rate. By feeding as if every cow will eat the same, it is expected that cows who eat less will have their excess feed consumed by cows who require more DMI. But if distributions are not normal, the pen will be over or underfed depending on how the distribution is skewed. Correction factors are routinely applied to increase the pen DMI as this method under-supplies the pen in practice. This is proven by the description of the pen distributions in this study. All high and low yielding pens, except for HIGHP50, had consistent left skew and non-normal distributions. With left sided outliers the main mass of the data is in the upper range of values for these pens. The mean is only an appropriate parameter of the Normal distribution, and the mean is pulled towards extreme outliers. For these pens, the mean value will be below the center of mass of this distribution leading to an underestimate of total DMI when the pen is fed at a uniform rate.

### ***Total DMI prediction***

#### *Predicting total DMI for a distribution shape*

All 100 replicates of the seven pen types had a total DMI by pen predicted using the best fit distribution shape for those data (Table I-4). The observed DMI for every pen replicate of the

seven pen types was calculated by summing the observed DMI of the cows in each replicate, and the pen DMI predicted by its best distribution shape was compared to the observed for percentage error. A percentage error was calculated for every replicate and the mean percentage error by pen type is presented. All seven pens had a percentage error less than 1 % for the best distribution prediction of DMI.

#### *Predicting total DMI by empirical equation*

Pen DMI was also predicted for all replicates by calculating a DMI value with the NASEM 2021 equation with the inputs as the mean values of the pen for each and multiplying that value by the n of cows in the pen. These predicted values were compared to the observed DMI by pen type for mean percentage error and ranged from 11 to 22 %.

The MSPE of each pen type was calculated for both prediction models to assess the fit of the predicted values to the observed (Table I-4, Figure I-2). Every pen type had a lower MSPE for the distribution shape prediction model as compared to the empirical equation model showing a better fit to data of the distribution shape model. This analysis demonstrates the model fit of a DMI predicted by an estimated distribution shape is more appropriate than assuming uniform feeding. The model error was partitioned into error due to bias, random variance and slope not equaling one. For the distribution prediction model 77 – 98 % of error was due to random variation in the data indicating that the model is well fitted to the data. For HIGHP50 the mean bias error was 21 % suggesting systematic error in prediction of the DMI mean by the model for this group. For the NASEM prediction model random error ranged from 6 – 68 % with the HIGHP50 pen type at 6 % error due to random variation and 93 % due to mean bias. This



HIGHP50 group was made up of primiparous only animals and from a smaller subset of possible animals and both models demonstrated systematic error in their prediction of mean producing this proportion of error in the model fit.

The NASEM 2021 empirical equation was derived using post milking peak cow data so estimates for early lactation pens may be inaccurate. High and low lactation range pens were all underestimated by this equation. High pen underestimation by the empirical model of this study agrees with the observed practice of necessary over-formulating diet on commercial dairies, as underfeeding cows can limit milk production. Describing a distribution type for a pen DMI predicts a DMI closer to the observed DMI of a pen, allowing diets to be calculated to a more accurate amount at formulation rather than utilizing an imprecise correction factor. The practice of formulating diets by increasing DMI by a fixed level is not precision feeding, and ingredient composition may change if DMI is a binding constraint in the formulation.

### ***Model assessment***

This modeling study was conducted by generating virtual pens with data of known DMI values. But the application of this work relies on the future ability of predicting the distribution shape of a cow pen with unknown DMI. For group housed dairy cows, DMI estimation is crudely approximated by measuring the feed remaining after a given time-period from the supply of a known feed quantity. Individual DMI measurements, such as Calan gates, are not feasible for large groups as one gate would be needed per cow and would probably change inter-cow feeding behavior (Seymour et al., 2019). Sensor technology attempts to measure individual cow

DMI with accelerometers and cameras at feed bunks, and if the data are validated could contribute to pen calculation of total DMI.

Pens with known individual cow DMI values can describe the distribution shape and parameters. This model generated virtual pens to build a labelled database of known distribution shapes. For pens on commercial dairies where individual DMI is not known, supervised learning could predict the distribution characteristics of a pen by a model trained with a labeled database. The model presented in this study demonstrates that predicting pen DMI with a distribution can provide accurate estimates. Current group feeding approaches supply a uniform level of DMI and are not precision feeding. Utilizing the distribution of DMI for each pen of cows is a necessary factor in achieving precision feeding aims for group feeding methods.

## **CONCLUSION**

Precision feeding for cow pens is not achieved by formulating a DMI value for an individual cow and feeding it uniformly. Describing the distribution shape of DMI for cow pens predicts a pen DMI with less error and can provide an approach to precision feeding in this industry.

## REFERENCES

- Bewley J., 2010. Precision Dairy Farming: Advanced Analysis Solutions for Future Profitability. Proceedings of The First North American Conference on Precision Dairy Management, 2 – 5 March 2010, Toronto, Canada, p. 16.
- Benchaar C, Rivest J, Pomar C, Chiquette J. 1998. Prediction of methane production from dairy cows using existing mechanistic models and regression equations<sup>1</sup>. *Journal of Animal Science*. 76:617–27
- Capper, JL., Cady, RA., 2020. The effects of improved performance in the U.S. dairy cattle industry on environmental impacts between 2007 and 2017. *Journal of Animal Science*. 98:skz291
- Chalupa, W., Vecchiarelli, B., Galligan, DT., Ferguson, JD., Baird, LS., et al. 1996. Responses of Dairy Cows Supplemented with Somatotropin During Weeks 5 Through 43 of Lactation. *Journal of Dairy Science*. 79:800–812
- Contreras-Govea, F. E., V. E. Cabrera, L. E. Armentano, R. D. Shaver, P. M. Crump, D. K. Beede, and M. J. VandeHaar. “Constraints for Nutritional Grouping in Wisconsin and Michigan Dairy Farms.” *Journal of Dairy Science*. 98:1336–44
- Dann, HM., Varga, GA., Putnam, DE., 1999. Improving Energy Supply to Late Gestation and Early Postpartum Dairy Cows. *Journal of Dairy Science*. 82:1765–78
- DePeters, EJ., Smith, NE., Acedo-Rico, J., 1985. Three or Two Times Daily Milking of Older Cows, and First Lactation Cows for Entire Lactations. *Journal of Dairy Science*. 68:123–32
- de Souza, RA., Tempelman, RJ., Allen, MS., VandeHaar, MJ., 2019. Updating predictions of dry matter intake of lactating dairy cows. *Journal of Dairy Science*. 102:7948–60

- Dhiman, TR., Satter, LD., 1997. Yield Response of Dairy Cows Fed Different Proportions of Alfalfa Silage and Corn Silage. *Journal of Dairy Science*. 80:2069–82
- Greenfield, RB., Cecava, MJ., Johnson, TR., Donkin, SS., 2000. Impact of Dietary Protein Amount and Rumen Undegradability on Intake, Peripartum Liver Triglyceride, Plasma Metabolites, and Milk Production in Transition Dairy Cattle. *Journal of Dairy Science*. 83:703–10
- Hamner, B., Frasco, M., 2018. Metrics: Evaluation Metrics for Machine Learning. R package version 0.1.4, Retrieved on 10 January from <https://CRAN.R-project.org/package=Metrics>.
- Huyler, MT., Kincaid, RL., Dostal, DF., 1999. Metabolic and Yield Responses of Multiparous Holstein Cows to Prepartum Rumen-Undegradable Protein. *Journal of Dairy Science*. 82:527–36
- Komsta, L., Novomestky, F., 2022. moments: Moments, Cumulants, Skewness, Kurtosis and Related Tests. R package version 0.14.1, Retrieved on 10 January 2023 from <https://CRAN.R-project.org/package=moments>.
- Korver, S., 1988. Genetic aspects of feed intake and feed efficiency in dairy cattle: A review. *Livestock Production Science*. 20:1–13
- Livesey, CT., Harrington, T., Johnston, AM., May, SA., Metcalf, JA., 1998. The effect of diet and housing on the development of sole haemorrhages, white line haemorrhages and heel erosions in Holstein heifers. *Animal Science*. 67:9–16
- Lucey, P.M., and H.A. Rossow. 2023. Describing the distribution type of DM intake for dairy cow pens based on pen characteristics. *animal* 17:100888.  
doi:10.10168/j.animal.2023.100888.

- National Academies of Sciences, Engineering, and Medicine (NASEM)., 2021. Nutrient Requirements of Dairy Cattle: Eighth Revised Edition. Washington, DC: The National Academies Press. <https://doi.org/10.17226/25806>.
- National Research Council (NRC)., 2001. Nutrient Requirements of Dairy Cattle: Seventh Revised Edition, 2001. Washington, DC: The National Academies Press. <https://doi.org/10.17226/9825>.
- Pascottini, OB., Leroy, JLMR., Opsomer, G., 2020. Metabolic Stress in the Transition Period of Dairy Cows: Focusing on the Prepartum Period. *Animals*. 10:1419
- Pomar, C., Hauschild, L., Zhang, GH., Pomar, J., Lovatto, PA., 2011. Precision feeding can significantly reduce feeding cost and nutrient excretion in growing animals. In *Modelling nutrient digestion and utilisation in farm animals*, (ed. Sauvant, D., Van Milgen, J., Faverdin, P., Friggens, N.) Wageningen Academic Publishers, Wageningen, Netherlands, pp. 327–34.
- Roseler, DK., Fox, DG., Chase, LE., Pell, AN., Stone WC. 1997. Development and Evaluation of Equations for Prediction of Feed Intake for Lactating Holstein Dairy Cows. *Journal of Dairy Science*. 80:878–93
- Schulze, C., Spilke, J., Lehner, W., 2007. Data modeling for Precision Dairy Farming within the competitive field of operational and analytical tasks. *Computers and Electronics in Agriculture*. 59:39–55
- Seymour, DJ., Cánovas, A., Baes, CF., Chud, TCS., Osborne, VR., et al. 2019. Invited review: Determination of large-scale individual dry matter intake phenotypes in dairy cattle. *Journal of Dairy Science*. 102:7655–63

- Shapiro, S. S., and M. B. Wilk. “An Analysis of Variance Test for Normality (Complete Samples).” *Biometrika* 52. 591–611.
- Stallings, CC., McGilliard, ML., 1984. Lead Factors for Total Mixed Ration Formulation. *Journal of Dairy Science*. 67:902–7
- St-Pierre, NR., Thraen, CS., 1999. Animal grouping strategies, sources of variation, and economic factors affecting nutrient balance on dairy farms. *Journal of Animal Science*. 77:72–83
- St-Pierre NR., 2007. Design and Analysis of Pen Studies in the Animal Sciences. *Journal of Dairy Science*. 90:87–99
- Theil H., 1966. Applied economic forecasting. Amsterdam: North-Holland Publishing Company
- Tyrrell, HF., Reid, JT., 1965. Prediction of the Energy Value of Cow’s Milk. *Journal of Dairy Science*. 48:1215–23
- USDA., Dairy, 2014, “Trends in Dairy Cattle Health and Management Practices in the United States, 1991-2014” USDA–APHIS–VS–CEAH-NAHMS. Fort Collins, CO. #711.0821
- Van Amburgh, ME., Collao-Saenz, EA., Higgs, RJ., Ross, DA., Recktenwald, EB., et al., 2015. The Cornell Net Carbohydrate and Protein System: Updates to the model and evaluation of version 6.5. *Journal of Dairy Science*. 98:6361–80
- Weiss, B., Setting Nutrient Specifications for Formulating Diets for Groups of Lactating Dairy Cows. Retrieved on 10 January 2023 from <https://dairy-cattle.extension.org/setting-nutrient-specifications-for-formulating-diets-for-groups-of-lactating-dairy-cows>

Wu, H., Godfrey, AJR., Govindaraju, K., Pirikahu, S., 2020. ExtDist: Extending the range of functions for probability distributions. R package version 0.6-4, Retrieved on 10 January 2023 from <https://CRAN.R-project.org/package=ExtDist>

Wu, H., Godfrey, AJR., Pirikahu, S., 2020. Parameter estimation and distribution selection by ExtDist. Retrieved on 10 January 2023 from <https://cran.microsoft.com/snapshot/2022-02-19/web/packages/ExtDist/vignettes/ParaEst-and-DistSel-by-ExtDist.pdf>

## TABLES

**Table I-1.** Description of the trial datasets used in the database of this dairy cow study.

Trial	Index <sup>2</sup>	n <sup>1</sup>			Breed	Lactation range, wk <sup>5</sup>	Intervention <sup>6</sup>	Difference ( $P < 0.05$ ) <sup>7</sup>
		Total	Primi <sup>3</sup>	Multi <sup>4</sup>				
DePeters et al. 1985 <sup>8</sup>	1	55	15	40	Holste in	1 – 44	Management	No
Dhiman and Satter 1997 <sup>9</sup>	2	74	29	45	Holste in	1 – 36	Diet	Yes
Greenfield et al. 2000 <sup>10</sup>	3	37	0	37	Holste in & Jersey	1 – 8	Therapy	Yes
Unpublished <sup>10</sup>	4	45	0	45	Not stated	3 – 18	Not stated	Not stated
Unpublished <sup>10</sup>	5	40	0	40	Not stated	1 – 21	Diet	Not stated
Huyler et al. 1999 <sup>10</sup>	6	31	0	31	Holste in	1 – 10	Diet	No
Livesey et al. 1998 <sup>11</sup>	7	51	0	51	Holste in	1 – 14	Diet & management	No
Chalupa et al. 1996 <sup>8</sup>	8	36	12	24	Holste in	3 – 43	Diet	Yes
Dann et al. 1999 <sup>10</sup>	9	57	57	0	Holste in	1 – 9	Diet	No
Total		426	113	313				

Primi = Primiparous

Multi = Multiparous

<sup>1</sup> Cow population per trial as the total, and primiparous and multiparous separately.

<sup>2</sup> Identifying index of each dataset.

<sup>3</sup> Primiparous dairy cows in their first lactation.

<sup>4</sup> Multiparous cows.

<sup>5</sup> The weeks of lactation for which each dataset recorded weekly individual cow milk yield and dry matter intake.

<sup>6</sup> The type of treatment intervention studied by each trial.

<sup>7</sup> Publication reported significant difference in the milk yield or dry matter intake of a single treatment group in the dataset.

<sup>8</sup> Dataset recorded weekly average dry matter intake, milk, milkfat, milk protein and bodyweight.

<sup>9</sup> Dataset recorded weekly average dry matter intake, milk, milkfat, and milk protein.

<sup>10</sup> Dataset recorded weekly average dry matter intake, milk, milkfat, milk protein, bodyweight, and body condition score.

<sup>11</sup> Dataset recorded weekly average dry matter intake and milk, and monthly milkfat, bodyweight and body condition score.



**Table I-2.** Skew, kurtosis and Shapiro-Wilk test of goodness of fit to normal distribution for dry matter intake values of all dairy cows by parity,

Week <sup>1</sup>	All cows				Primiparous				Multiparous			
	n <sup>2</sup>	Skew <sup>3</sup>	Kurtosis <sup>4</sup>	Shapiro-Wilk P-value <sup>5</sup>	n	Skew	Kurtosis	Shapiro-Wilk P-value	n	Skew	Kurtosis	Shapiro-Wilk P-value
1	294	-0.10	2.70	0.5	101	-0.49	3.24	0.08	193	-0.04	2.31	0.1
2	294	-0.32	2.93	0.01	101	-0.012	2.61	0.2	193	-0.47	3.43	0.03
3	342	-0.34	3.02	0.05	102	0.10	2.28	0.2	240	-0.56	3.87	0.004
4	360	-0.45	3.22	0.001	109	0.10	2.38	0.5	251	-0.77	4.40	< 0.001
5	364	-0.30	3.24	0.06	110	-0.031	2.52	0.2	254	-0.47	4.04	0.001
6	374	-0.24	3.21	0.5	113	0.24	2.78	0.4	261	-0.42	4.20	0.008
7	375	-0.37	3.15	0.006	113	0.072	3.00	0.4	262	-0.29	3.19	0.1
8	373	-0.30	2.69	0.003	113	0.014	2.33	0.4	260	-0.23	2.79	0.005
9	337	-0.36	2.79	0.003	113	-0.022	2.37	0.3	224	-0.30	3.07	0.1
10	280	-0.11	2.33	0.02	56	0.063	2.69	0.6	224	-0.18	2.68	0.2
11	250	-0.21	2.61	0.09	56	-0.61	3.46	0.09	194	-0.16	2.79	0.3
12	250	-0.21	2.47	0.05	56	-0.13	2.46	0.5	194	-0.37	3.29	0.1
13	250	-0.15	2.91	0.8	56	-0.47	3.70	0.4	194	-0.11	3.37	0.5
14	250	0.052	2.65	0.9	56	-0.012	3.11	0.8	194	-0.071	3.21	0.7
15	250	-0.091	2.72	0.8	56	-0.26	2.58	0.7	194	-0.33	3.52	0.1
16	250	-0.27	3.05	0.3	56	0.35	3.06	0.4	194	-0.70	4.67	< 0.001
17	250	-0.22	2.78	0.4	56	-0.32	3.18	0.4	194	-0.55	3.68	0.01
18	250	-0.20	3.40	0.1	56	0.25	2.83	0.8	194	-0.65	4.71	< 0.001
19	205	-0.23	3.69	0.2	56	-0.12	3.06	0.7	149	-0.67	4.14	0.003
20	205	-0.42	3.96	0.002	56	-1.8	9.34	< 0.001	149	-0.66	3.93	0.003
21	205	-0.18	3.67	0.1	56	-0.40	3.16	0.3	149	-0.54	4.05	0.01
22	164	-0.16	3.65	0.04	56	0.21	3.03	0.9	108	-0.68	4.33	0.008
23	162	-0.16	4.01	0.2	56	-0.062	2.73	0.6	106	-0.59	4.59	0.02
24	162	0.34	2.93	0.2	56	-0.50	2.78	0.04	106	0.051	2.59	0.3
25	162	0.32	2.49	0.03	56	-0.052	2.45	0.5	106	-0.023	2.21	0.1
26	162	0.21	3.03	0.5	56	-0.53	3.32	0.4	106	-0.082	2.73	0.9
27	162	0.52	3.01	0.01	56	0.012	2.79	0.6	106	0.20	2.48	0.2
28	162	-0.082	3.46	0.6	56	-1.2	5.94	0.001	106	-0.081	2.63	0.7
29	162	0.16	2.58	0.2	56	0.19	2.29	0.3	106	-0.22	2.49	0.2

30	162	0.11	2.88	0.8	56	-0.17	3.63	0.4	106	-0.062	2.59	0.9
31	162	0.033	2.60	0.9	56	0.48	3.33	0.4	106	-0.23	2.43	0.3
32	162	0.11	2.74	0.2	56	0.68	3.55	0.05	106	-0.19	2.44	0.2
33	161	-0.18	2.66	0.4	56	-0.062	2.91	0.9	105	-0.39	2.45	0.03
34	161	-0.11	2.74	0.09	56	-0.13	3.55	0.5	105	-0.31	2.37	0.02
35	161	-0.21	2.84	0.1	56	0.17	3.47	0.2	105	-0.42	2.57	0.03
36	160	-1.2	8.10	< 0.001	56	-0.082	2.70	0.8	104	-1.4	7.31	< 0.001
37	86	-0.29	3.44	0.2	27	0.053	2.18	0.7	59	-0.29	3.07	0.3
38	85	0.094	4.16	0.08	26	-0.19	2.14	0.6	59	0.034	3.64	0.3
39	83	-0.27	4.15	0.07	26	-0.23	2.64	0.5	57	-0.22	4.10	0.1
40	82	-0.36	4.27	0.2	26	-0.20	2.07	0.4	56	-0.45	4.45	0.09
41	65	-0.46	3.72	0.3	21	-0.64	2.62	0.1	44	-0.39	4.14	0.5
42	62	-0.80	3.62	0.02	21	-0.86	2.48	0.01	41	-0.72	4.28	0.1
43	58	-0.87	3.92	0.01	21	-0.64	2.63	0.3	37	-0.97	5.19	0.03
44	26	-0.26	2.39	0.5	8	-0.091	1.68	0.5	18	-0.15	2.69	0.9

<sup>1</sup> The week of lactation of data observation

<sup>2</sup> The population of cows observed at the given week of lactation

<sup>3</sup> Measure of data skew. Negative values indicate left skew, positive values indicate right skew.

<sup>4</sup> Measure of data in tails of distribution. Kurtosis of 3 is considered normal.

<sup>5</sup> A  $P < 0.05$  rejects the null hypothesis that the data are normally distributed.

**Table I-3.** The proportion of best distribution fit for 100 replicates of each dairy cow pen class.

Pen type index <sup>3</sup>	Cows, n	Lactation range <sup>4</sup> , wk	Pen replicates <sup>5</sup> , n	Mean DMI, cow <sup>6</sup> (kg)	DMI Distribution description <sup>1</sup>		Proportions of best distribution of DMI for all replicates <sup>2</sup> (%)			
					Skew	Kurtosis	Normal	Beta	Weibull	Logistic
FRESH20	20	1 – 3	100	17.06	0.07	2.76	8	85	1	6
FRESH60	60	1 – 3	100	16.97	-0.03	2.49	8	87	0	5
HIGH50	50	4 – 18	100	22.71	-0.2	2.99	8	77	0	15
HIGHP50	50	4 – 18	100	20.51	0.05	2.87	7	80	8	5
HIGH200	200	4 – 18	100	22.97	-0.2	3.12	28	60	0	12
LOW50	50	19 - 44	100	20.90	-0.2	3.22	10	62	4	24
LOW150	150	19 – 44	100	20.86	-0.2	3.43	29	36	1	34

FRESH20 = Pens of 20 random cows within the lactation range 1 – 3 week.

FRESH60 = Pens of 60 random cows within the lactation range 1 – 3 week.

HIGH50 = Pens of 50 random cows within the lactation range 4 – 18 week.

HIGHP50 = Pens of 50 random primiparous cows within the lactation range 4 – 18 week.

HIGH200 = Pens of 200 random cows within the lactation range 4 – 18 week.

LOW50 = Pens of 50 random cows within the lactation range 19 – 44 week.

LOW150 = Pens of 150 random cows within the lactation range 19 – 44 week.

<sup>1</sup> The mean distribution statistics of skew and kurtosis for all replicates of each pen type.

<sup>2</sup> The proportion of each distribution type occurring as the best fit for DMI of a pen across all replicates of pen types.

<sup>3</sup> Index of generated virtual pen type for given lactation range and population size.

<sup>4</sup> The range of weeks of lactation the pen type was constructed across.

<sup>5</sup> The number of replicates generated for each pen type with a random and unique selection of cows in each replicate.

<sup>6</sup> The mean cow DMI of all replicates of each pen type.

**Table I-4.** DMI predictions by best distribution type and by NASEM equation of all pen replicates of each dairy cow pen class and model fit statistics.

Pen type index <sup>3</sup>	Observed pen DMI (kg) <sup>4</sup>	Predicted pen DMI by distribution <sup>1</sup>						Predicted pen DMI by NASEM <sup>2</sup>					
		DMI (kg)	% Error of distribution <sup>5</sup>	MSPE <sup>6</sup>	% Bias <sup>7</sup>	% Slope not 1 <sup>8</sup>	% Random variation <sup>9</sup>	DMI (kg)	% Error of NASEM	MSPE	% Bias	% Slope not 1	% Random variation
FRESH20	341.09	342.43	+ 0.4 <sup>a</sup>	55.64	0	5	95	291.66	- 14.37 <sup>b</sup>	211.4	0	32	68
FRESH60	1,018.17	1,019.8	+ 0.09 <sup>a</sup>	60.42	0	2	98	876.96	- 13.82 <sup>b</sup>	641.6	0	42	58
HIGH50	1,135.61	1,136.6	+ 0.09 <sup>a</sup>	43.63	0	2	98	937.45	- 17.42 <sup>b</sup>	545.9	0	35	65
HIGHP50	1,025.69	1,026.2	+ 0.05 <sup>a</sup>	55.28	21	2	77	921.81	- 10.11 <sup>b</sup>	7 612	93	1	6
HIGH200	4,593.08	4,595.1	+ 0.04 <sup>a</sup>	182.8	2	2	96	3761.0	- 18.11 <sup>b</sup>	1 573	0	33	67
LOW50	1,044.79	1,045.6	+ 0.08 <sup>a</sup>	47.44	0	2	98	810.67	- 22.38 <sup>b</sup>	587.1	0	64	36
LOW150	3,127.82	3,131.1	+ 0.1 <sup>a</sup>	81.99	0	3	97	2 427.9	- 22.37 <sup>b</sup>	803.2	0	82	18

DMI = Dry matter intake

MSPE = Mean squared prediction error

FRESH20 = Pens of 20 random cows within the lactation range 1 – 3 week.

FRESH60 = Pens of 60 random cows within the lactation range 1 – 3 week.

HIGH50 = Pens of 50 random cows within the lactation range 4 – 18 week.

HIGHP50 = Pens of 50 random primiparous cows within the lactation range 4 – 18 week.

HIGH200 = Pens of 200 random cows within the lactation range 4 – 18 week.

LOW50 = Pens of 50 random cows within the lactation range 19 – 44 week.

LOW150 = Pens of 150 random cows within the lactation range 19 – 44 week.

<sup>1</sup> Distribution type prediction model. The mean total pen DMI of 100 replicates of the given pen type from the predictive model of the best fitted distribution shape for each replicate.

<sup>2</sup> Empirical predication model (NASEM, 2021). The mean total pen DMI of 100 replicates of the given pen type from the predictive model of a single empirical DMI estimate that is fed uniformly to the population of cows.

<sup>3</sup> Index of generated virtual pen type for given lactation range and population size.

<sup>4</sup> The observed mean total pen DMI for 100 replicates of each pen type.

<sup>5</sup> Mean percentage error of each predicted pen DMI estimate from the observed value for each replicate. Means of percentage error for each pen replicate were compared across both prediction models using Students t-test.

<sup>6</sup> Model fit evaluated as the predicted total pen DMI of both prediction models to the observed values of each replicate. Mean square prediction error to evaluate the error in model fit.

<sup>7</sup> Proportion of model error as MSPE due to bias (Benchaar et al., 1998).

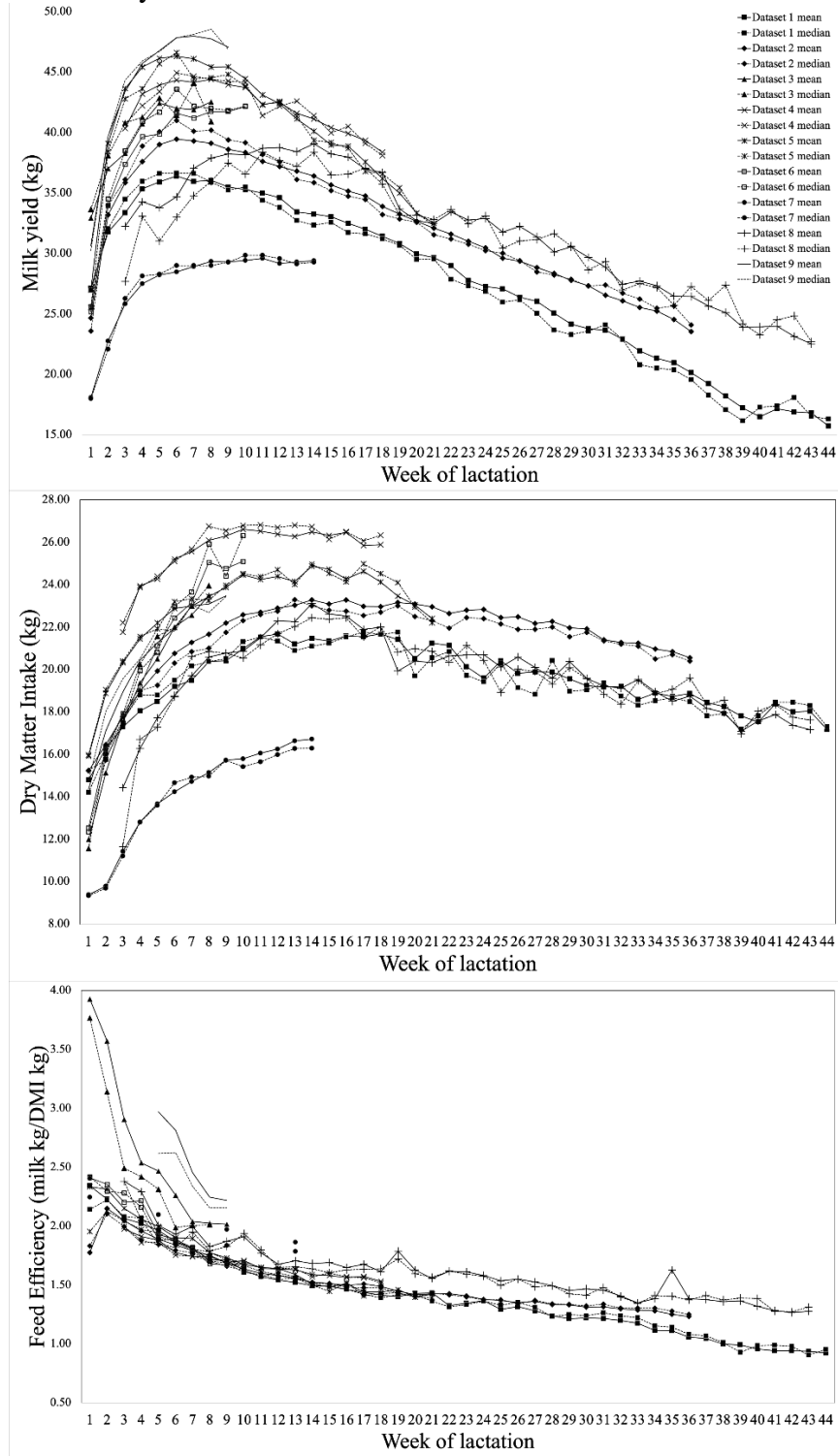
<sup>8</sup> Proportion of model error as MSPE due to the slope of data not equal to 1 (Benchaar et al., 1998).

<sup>9</sup> Proportion of model error as MSPE due to random variation in the data (Benchaar et al., 1998).

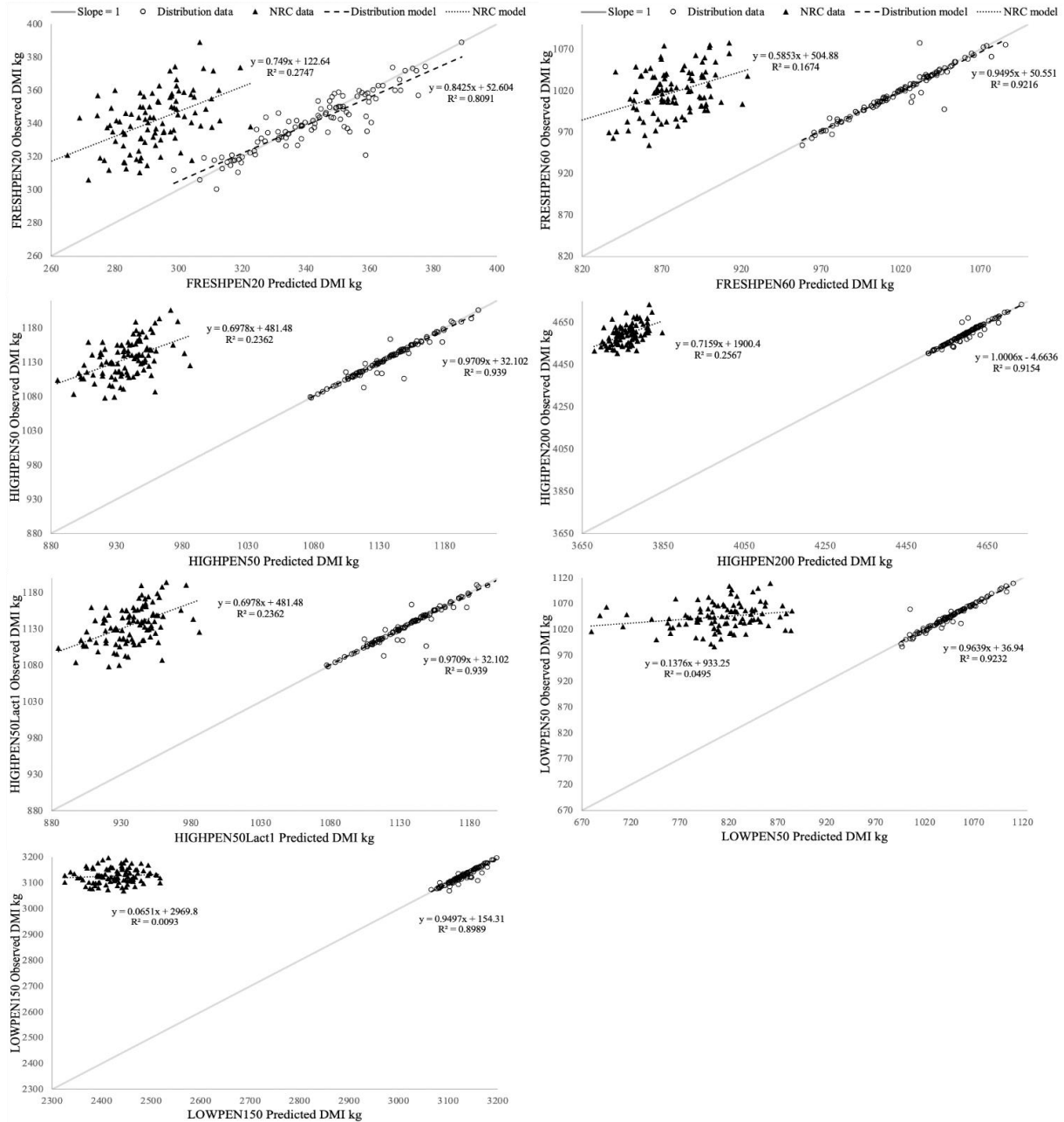
<sup>a-b</sup> Values within a row with different superscripts differ significantly at  $P < 0.05$  for t-test comparison of mean percentage error of both prediction models.

## FIGURES

**Figure I-1.** Line plots of the mean and median of (Figure A) Milk Yield, (Figure B) DMI and (Figure C) feed efficiency of dairy cows by week of lactation and dataset separately. Abbreviations: DMI = Dry matter intake.



**Figure I-2.** Scatter plots of observed versus predicted model diagnostics for each of the seven dairy cow pen types for both the Distribution and NASEM model. Abbreviations: DMI = DM intake. FRESH20 = Pens of 20 random cows within the lactation range 1–3 weeks. FRESH60 = Pens of 60 random cows within the lactation range 1–3 weeks. HIGH50 = Pens of 50 random cows within the lactation range 4–18 weeks. HIGH50Lact1 = Pens of 50 random primiparous cows within the lactation range 4–18 weeks. HIGH200 = Pens of 200 random cows within the lactation range 4–18 weeks. LOW50 = Pens of 50 random cows within the lactation range 19–44 weeks. LOW150 = Pens of 150 random cows within the lactation range 19–44 weeks.



## CHAPTER II.

# PREDICTING THE DISTRIBUTION SHAPE AND DRY MATTER INTAKE OF COW PENS WITH MACHINE LEARNING

### ABSTRACT

The individual cow is not representative of the dry matter intake (**DMI**) of a pen, but we formulate rations based on an average cow with a correction factor. This is not precision feeding, and we feed by pen. A normal distribution of DMI is assumed, but it usually has skew or kurtosis. If the distribution of DMI is described, rations can be formulated to the unit of pen more precisely. Two models of estimating pen DMI were compared. The first used a trained Extreme Gradient Boosting (**XGB**) algorithm to predict the distribution shape of pen DMI: a Beta or Generalized Normal distribution. Another XGB algorithm predicted the input features that describe that distribution shape. The best selected input features for the XGB algorithms were parity, week in milk, and descriptive statistics of milk yield. A set of DMI values that fit the predicted shape were randomly generated to estimate total pen DMI for validation of the model's prediction. Each set was summed to estimate total pen DMI. The second model was the NASEM 2021 DMI equation of cow variables, which used a representative cow to predict DMI for each pen. The pen means of each input value were applied to the NASEM equation: milk energy, days in milk, parity, bodyweight (**BW**), and BCS. To calculate total pen DMI, this value was multiplied by the number of cows, simulating the scenario where DMI is assumed to be uniform. Results from both models were compared to the observed pen DMI for assessment of mean squared error of prediction (**MSEP**) using a dataset of 229,909 pens. Pens were subset by their distribution shape prediction. For Beta pens: the distribution shape model, and NASEM equation



model had a MSEP of 5,249 and 50,320, with 3% and 85% due to mean bias, 5% and 10% due to regression slope not equal to one, and 92% and 5% due to random bias respectively. For Generalized Normal pens: the distribution shape, and NASEM DMI models had a MSEP of 2,085 and 45,779 respectively, with 1% and 73% due to mean bias, 7% and 7% due to regression slope not equal to one, and 92% and 20% due to random bias. The NASEM equation model consistently underestimated pen DMI, contributing to the high error. Applying the equation to each cow, instead of the pen mean, may improve the estimation but is not how it is commonly applied. The machine learning algorithm described the distribution for a highly precise and accurate DMI prediction, only requiring the inputs of pen size, week in milk, milk yield and parity.

## **INTRODUCTION**

On commercial dairies, the unit for ration formulation should be pen, not individual cow. The individual cow is not representative of the DMI of a pen, but we formulate rations based on an average cow or the 75<sup>th</sup> percentile cow (Linn et al., 2021). This is not precision feeding. Most pens contain cows with a broad range of DMI, milk yield, and nutrient requirements within the pen. All ration formulation programs include a DMI constraint. When the DMI of a single cow is known, a ration can be formulated precisely to her requirements (NASEM, 2021), but if that ration is fed to a pen, it assumes that the cows have a uniform DMI value. Precision and accuracy are lost, and the distribution of feed intake is unknown. This raises a concern on the applicability of individual cow studies to operations where cows are housed in pens (St-Pierre, 2007). Rations should be formulated for the unit they are feeding, the pen.

Practically, individual DMI values are difficult to measure. Pen weigh backs (feed refusals subtracted from feed delivered divided by the number of cows in the pen) estimates DMI for an average cow in the pen. But the number of cows in the pen is not always known and depends on the day relative to when cows are moved in and out of pens on the dairy. Individual feed bins are impractical and cost-prohibitive; spatial imaging attempts to circumvent this are yet to be validated or implemented (Lassen et al., 2023). Common practice is to use the herd mean as input values in diet formulation (Seymour et al., 2019). Lucey and Rossow (2023) reported the Beta distribution as the most common shape of DMI for dairy pens of multiple types. This proves the mean is an inappropriate input value, as it is only a property of the Normal distribution. For example, a distribution shape predicted DMI values for pens with an error of less than 1% from the observed, in comparison to a 22% error when the mean was assumed as the DMI of all cows. A robust method of estimating this distribution shape without knowledge of individual cow DMI will improve diet precision and accuracy.

Cow DMI has been calculated by formulas using biological and environmental information: milk yield, milk energy, BCS and BW (Fox et al., 2004; NASEM, 2021). This is constrained by the relevant input information being available from commercial dairies (Higgs et al., 2012). Data that describe the level of the pen are not observations of these listed variables, but instead their descriptive statistics: minimum, maximum, mean, and median of milk yield, week in milk (**WIM**) or parity, etc. Lactating cows are usually grouped in pens by WIM, parity, reproductive status, or milk production (Contreras-Govea et al., 2015) and this affects the distribution of cow's nutrient requirements. There are biologically plausible relationships between cow factors, such as milk yield and DMI, facilitating the empirical deduction of DMI

equations. But the minimum or maximum WIM, milk yield, or parity, etc., of a pen does not have a biological relationship to the distribution that fits the DMI. Since empirical modeling assumes biological plausibility, construction of a pen level model empirically is limited. Instead, techniques such as machine learning algorithms could connect these arbitrary relationships (Halevy et al., 2009).

The objectives of this paper were to construct and compare two models of pen DMI prediction: A) Create a large training dataset of pens based on real cow weekly observations, B) Develop a distribution shape model (**DSM**) by training machine learning algorithms to predict the distribution shape of DMI. Validate it with an external dataset by comparing predicted DMI to observed, C) Use a NASEM equation model (**NSM**; NASEM, 2021) of DMI to predict the DMI of each pen and validate it with an external dataset by comparing predicted DMI to observed, D) Compare the model diagnostics of the DSM and NSM. Therefore, this modeling study will test if the pen DMI predicted by the DSM will be closer to the true value than the NSM.

## **MATERIALS AND METHODS**

No approval from the University of California, Davis, Animal Care and Use Committee was needed for this study as it was conducted using only previously collected data.

### ***Individual cow observations***

Weekly individual cow observations of DMI were joined into a single dataset from nine separate studies. To be included, each observation required an identifier, parity, WIM, average

weekly DMI (kg) and average weekly milk yield (kg). Some of the studies recorded milk protein (kg), milk fat (kg), BCS or bodyweight (kg) (Table II-1). Outliers were removed from the variables milk fat and milk protein if they were greater than the 99<sup>th</sup> percentile.

### ***Creation of pen training dataset based on individual cow observations***

Similar to Lucey and Rossow (2023), individual cow observations were randomly assorted into pens of different sizes, ranges of WIM, and parity with a function in Python 3.10 (Python Software Foundation, <https://www.python.org/>; Appendix 1). All values of parity level, primiparous, multiparous, or both, pen size ranging from 25 to 250 cows, and range of WIM from the individual cow records in the data were combined to create the pens. Only observations from unique cows were selected to meet the pen size; if the number of qualifying cows was less than the pen size then it was not assorted at those criteria. This was performed 20 times with a different random seed for every pen, providing 20 pens of different cows for every pen possibility. This methodology delivered snapshots of pens at different ranges of WIM, with no cow present more than once per pen.

### ***Distribution shape model - DSM***

Descriptive statistics of mean, median, minimum, maximum, 5<sup>th</sup>, 25<sup>th</sup>, 75<sup>th</sup> and 95<sup>th</sup> percentile, standard deviation, skew and kurtosis of milk yield, milk fat, milk protein, BW, BCS and WIM were calculated for every pen (Appendix 2) to act as features in the model. Missing values of these variables were excluded from the calculations, except for milk yield, parity, and WIM as these had no missing data. The unit of interest in this modeling study was pen. The percentile values from 5<sup>th</sup> to 95<sup>th</sup> represented the distribution shape of the variables, the mean and

median inferred information about data skew, kurtosis and skew statistics described the spread of data.

The distribution shape of DMI for each pen was fit to either the Beta or Generalized Normal shape based on the highest value of the Kolmogorov-Smirnov goodness of fit test (Smirnov, 1939) using the “SciPy” package (Virtanen et al., 2020) of Python 3.10. Preliminary analysis of these data demonstrated the Beta and Generalized Normal were the most frequent shapes to describe pen DMI from a candidate pool of 60 shapes (Appendix 3), so fitting was restricted to these distributions. Beta distributions can represent skewed or symmetrical data. The Generalized Normal is the Normal distribution with an additional shape parameter and can describe data with a concentration around the mean, and different tails than the Normal distribution. Two distributions were chosen so a binary decision algorithm could be constructed for the outcome of the best fitting distribution shape. After selecting the distribution, the pens were split by distribution shape, and the respective parameter values of each were calculated using maximum likelihood estimation. The four parameters of the Beta distribution were shape,  $\alpha$  and  $\beta$  (which together explain symmetry, peak location, spread and skewness), location and scale (which explains dispersion of data). The three parameters of the Generalized Normal were shape (tails), location, and scale. These parameters replicated the curved line of the distribution of DMI, and the area under the curve was the total DMI for each pen. These data were the training dataset.

## ***Predicting the distribution shape of DMI using DSM and NSM***

### *Feature preparation*

The descriptive statistics of each pen of the training dataset were included as features for training the machine learning algorithms. All continuous variables were visually assessed for normality to fulfil their assumption. The parity variable was encoded to columns using the “OneHotEncoder” function of the Scikit-learn package (Pedregosa et al., 2011) in Python 3.10.

### *Machine learning training, and algorithm selection*

For the DSM, three separate machine learning systems were used: one classification, and two non-linear regressions. The classification algorithm was trained to the outcome label of DMI distribution shape; Beta or Generalized Normal. The pens were split by distribution shape and two multi-output regression algorithms, one for each distribution, were trained to the outcome labels of the respective shape and scale parameters. All were trained on 75% of the training dataset, which was split randomly by the “train\_test\_split” function of Scikit-learn. The other 25% of data were not included in algorithm training and were used in testing. These training and testing data together were the DSM training data and were not used in validation.

### *Preparation of validation data*

Data from a single site study of 88 multiparous Holstein cows and 2,086 weekly observations (Lean et al., 1992) of DMI, milk yield, milk fat, milk protein and BW were assorted into pen data that was prepared and cleaned as described in the creation of pen training dataset section of the methods of this paper. The pen size for these data was a maximum of 88 cows, as only one observation per cow was included per pen. The WIM range of these observations was 1

to 15, and pens were only assorted within this window. Pen DMI was calculated as the sum of the observed cow DMI values in each pen, and the true DMI distribution shape was selected from the Beta or Generalized Normal shapes by the highest value of the Kolmogorov-Smirnov goodness of fit test.

#### *Pen DMI DSM prediction of validation data*

We validated the DSM prediction of pen DMI calculation with these data using the pipeline of the trained algorithms that predicted the DMI distribution shape of these pen data, the distribution shape parameters, and plotted a set of DMI values for that distribution (Lucey and Rossow, 2023). The classification algorithm predicted the pen DMI data as Beta or Generalized Normal. These data were split on the prediction and separated by the distribution shape selected, and the non-linear regression algorithms predicted the parameter values for each shape separately. A set of individual cows' DMI values were plotted to fit the predicted distribution, and they were summed for a calculation of total pen DMI of these validation data. This step was repeated 100 times for each pen so that 100 random sets of DMI values for the predicted distribution were generated, with the mean taken as the final pen DMI.

#### *Pen DMI NSM prediction of validation data*

To compare the DSM calculation of pen DMI to a conventional method, the DMI of each pen in the validation data were predicted with a model of the NASEM empirical equation, the NSM using Eq. II-1 (NASEM, 2021).

$$DMI = [(3.7 + parity \times 5.7) + 0.305 \times MilkE + 0.022 \times BW(kg) + (-0.689 + parity \times -1.87) \times BCS \times [1 - (0.212 + parity \times 0.136) \times \exp^{(-0.053 \times DIM)}]] \quad \text{Eq. II-1}$$

$$MilkE = ((0.0929 \times milk\ fat\ (\%)) + (0.0547 \times (milk\ protein\ (\%) \div 0.93)) + 0.192) \quad \text{Eq. II-2}$$

The input parameters of this equation included milk energy (Eq. II-2); NRC, 2001), days in milk (**DIM**), parity, BW and BCS. Information on BCS was not available for any pens in these data, so this value was fixed at three. All other inputs were set at the mean of each pen, and a DMI value was calculated. Assuming a uniform DMI (as is the case when one DMI represents a pen) for all cows, this calculated value was multiplied by the pen size for the pen DMI estimate. These pen DMI predictions were compared to the observed, and model diagnostic evaluation was performed.

### Statistical Analyses

For the DSM, the best performing algorithm from three possibilities was selected: Decision Tree, Random Forest, and Extreme Gradient Boosting (**XGB**). The default hyperparameters of these algorithms (Pedregosa et al., 2011), and the features of pen size, parity, WIM, and milk yield descriptive statistics were used for evaluation with the training data. For classification, the best was chosen by the highest combined skill scores calculated with a 5-fold cross validation using Eq. II-3, 4, 5:

$$\text{Accuracy} = \left( \frac{\text{True Positives} + \text{True Negatives}}{\text{True Positives} + \text{True Negatives} + \text{False Negatives} + \text{False Positives}} \right) \quad \text{Eq. II-3}$$



$$\text{Precision} = \left( \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \right) \quad \text{Eq. II-4}$$

$$\text{Recall} = \left( \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \right) \quad \text{Eq. II-5}$$

$$\text{F1-score} = \left( \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \right) \quad \text{Eq. II-6}$$

Each skill score was calculated on one of five data folds of the training data, and the mean of the folds was the result. For the non-linear regression algorithms, the root mean squared error (RMSE) was evaluated by a 3-fold cross validation, and the algorithm with the lowest score was selected.

#### *DSM Feature inclusion*

The best algorithm of classification and non-linear regression were evaluated with additional features. The base feature set included: pen size, parity, WIM, and milk yield descriptive statistics. The algorithms were trained again with the addition of milk protein and milk fat descriptive statistics, followed by the further addition of BW and BCS descriptive statistics. For both additional feature sets, the algorithms were evaluated by skill scores and RMSE, and compared to the scores of the base for selection of the best feature set.

#### *DSM Hyperparameter tuning*

Hyperparameters were the properties of each algorithm that affected its decision process, including the depth of decision trees, number of decision splits, or the learning rate, loss function, and number of boosting rounds. The machine learning algorithms of the DSM were trained with different combinations of their hyperparameter values by a Random Search cross validation using the “RandomizedSearchCV” function of Scikit-learn. This was done in 3-folds

for 50 iterations of possible hyperparameters. Using three folds ensured training was robust and not dependent on any subset of the data. The best combination of hyperparameters were selected at the lowest RMSE.

### *DSM and NSM model diagnostics*

Predicted pen DMI was compared to the observed of every pen for model diagnostics by mean square error of prediction (**MSEP**) partitioned into error due to mean bias, error due to the slope not equal to 1 and error due to random variation. The coefficient of determination (**R<sup>2</sup>**), root mean square error of prediction (**RMSEP**), ratio of RMSEP/standard deviation (**RSR**), concordance correlation coefficient (**CCC**), correlation coefficient (**r**) and bias correction factor (**Cb**) were calculated to evaluate model performance (Bibby and Toutenberg, 1977).

## **RESULTS AND DISCUSSION**

### *Individual cow observations and pen assortment*

The dataset from all nine studies included 9,029 weekly observations of 609 cows, after the exclusion of outliers or observations with missing values of DMI or milk yield (Table II-1). Observations with missing values of milk protein (2,434), milk fat (197), bodyweight (2,948) and BCS (7,397) were not removed. To achieve objective A, pen assortment was designed to represent pen types on commercial dairies and was based on four pen types: Fresh pens (1 to 3 – 5 WIM), Early (1 to 6 – 41 WIM), Peak (3 – 16 to 8 – 41 WIM), and Late (10 – 20 to 18 – 45) (Appendix 1). The total number of resulting pens after outlier removal was 5,203,685, and of these 871,956 contained no BCS information, and their BCS descriptive statistics were assumed to be the mean of all other pens. Cow observations had a maximum WIM of 45 wk.

### *Selection of distribution shape*

Before data cleaning, 3,533,115 pens were best fit by Beta, and 3,657,925 pens were best fit by Generalized Normal. After removal of outliers from these pen data, there were 1,547,611 Beta and 3,656,074 Generalized Normal pens. This is a loss of 56% of Beta, and 0.05% of Generalized Normal data. Based on visual assessment of the distributions, pens with values for shape, location, and scale outside of the range  $-40$  to  $+80$  were removed; these plotted distributions did not describe curves of the expected shape and were considered an overfitting error. This error was introduced when a pen was fit to one of the two possible distribution shapes, when neither may have been appropriate for those data.

For example, neither of the two candidate distribution shapes can describe multimodal data, but both candidates can represent almost identical shapes to each other at certain parameter values. Because of these properties, it may be difficult to predict the best shape when pens have DMI distributions well described by both Beta and Generalized Normal. If pen DMI was multimodal, or an atypical shape, neither Beta nor Generalized Normal were a good fit, and insensible distribution parameters were calculated when one of these shapes were forced to the data. Since the Generalized Normal shape does not describe skew, it would be expected that the Beta was selected for any non-normal shapes, including multimodal. This may explain the large attrition of pens classified as Beta. The addition of other candidate distribution shapes to the best fitting algorithm could reduce this but would increase the complexity of the classifier. The best fit distribution shape algorithm could also be constrained so that a distribution shape with extreme outliers was not chosen instead of a distribution with appropriate parameter values, even if its maximum likelihood score was less optimal than the distribution with extreme parameters.

### ***DSM machine learning algorithm selection***

Objective B was to develop the DSM by training and selecting the best machine learning algorithms to predict the distribution shape and its parameter values for DMI. For the classification algorithm, the XGB had the highest performance, achieving an average accuracy of 71%, 72% precision, 96% recall, and an F1-score of 82% (Appendix 4). For the non-linear regression algorithms, the Beta and Generalized Normal distributed pens were trained separately. The XGB algorithm with a feature set of pen size, milk yield descriptive statistics, parity, and WIM had the best performance for both distribution shape subsets with the lowest RMSE (Appendix 4). Extending the feature set to include milk fat and milk protein descriptive statistics, and further with the addition of bodyweight and BCS yielded no improvement in performance of the classification or non-linear regression algorithms as the skill scores and RMSE were not different between feature sets (Table II-2).

The decision tree had the worst performance and was a simple algorithm that predicts the outcome by splitting the data at different feature values. It was prone to overfitting, sensitive to outliers, and can have high variance. Random Forest was an ensemble of decision trees generated on random splits of the data. The entire set of these decision trees was the forest, and their results were combined for the final output. This results in lower variance and makes it less sensitive to outliers than a decision tree. The XGB algorithm was a gradient boosting technique that uses tree-based learning, but instead of combining trees like Random Forest (bagging), it builds the best model sequentially and passed the error (loss function) of each learning tree onto the next (boosting). The residuals of each tree were used to fit the next, and regularization was included to reduce overfitting. The XGB algorithm is a compiled C++ library that efficiently employs

CPU, memory, and parallelization. For these large training data of this project, XGB was more time efficient than Random Forest (20 minutes for XGB, 25 minutes for random forest with an AMD Ryzen 9 7950X 16-Core Processor 4.50 GHz, 32 GB RAM, NVIDIA GeForce RTX 4070). The XGB was faster than Random Forest in the order of tens of hours in hyperparameter fitting because of its parallelization use and produced higher skill scores for both classification and regression. However, boosting algorithms such as XGB are more sensitive to outliers, which is why features were assessed for normality in the preparation step.

More features provide more information for the algorithms to learn from, but do not guarantee better performance. Part of objective B was to assess the features necessary for prediction of the DMI distribution, as it impacts the application in commercial settings. Empirical equations of DMI utilize physiologic parameters: age, milk production, milk composition, bodyweight and BCS at the level of cow. To apply these to the unit of pen, it is necessary to predict a DMI value for each cow and combine the results. But with more variables, the quality of data and the labor time for data collection impacts application. In these machine learning systems, parity and WIM information were included in all training, and these data were easily available for dairy pens and provided information on the stage of lactation and cow makeup. Milk production features were added to the base feature set, as these could be recorded routinely when the equipment and technology are present. There is an established relationship between milk yield and DMI, but milk components and BW, and BCS also impacted energy requirements (NASEM 2021), so these additional features were added to test for improvements in performance. The lack of difference in skill scores (Table II-2) demonstrates that information on milk production, pen size, WIM and parity was sufficient to predict the distribution of DMI

for a pen. No information on BCS was included in the final models, removing the limitation of missing BCS information in these data.

#### *Algorithm testing after hyperparameter tuning*

The final XGB algorithms were evaluated with testing data to assess overfitting, generalization, and model performance on previously unseen data. The classification algorithm had an overall accuracy of 71% on testing. For Beta classification it had a precision of 58%, 10% recall, and an F1-score of 17%. For Generalized Normal classification it had a precision of 72%, 97% recall and an F1-score of 82% (Table II-3). Accuracy is the total number of correctly classified instances over the total data and is a simple measure of correctness but can be misleading as the dataset was imbalanced with 1,547,611 Beta and 3,656,074 Generalized Normal pens. Precision (positive predictive value) was the ratio of correctly classified positive predictions from all positive classifications and decreases with more false positives. Recall (sensitivity) was the ratio of correctly predicted positive observations from all observations of that class and decreases with greater false negatives. Good classifiers have values close to one for both precision and recall. The F1-score was the harmonic mean of precision and recall and can be a better measure of classifier performance than accuracy in class imbalance. The RMSE measures error by the square root of the squared differences between predicted and true values. For these multi-output regressions, it is the average error of the multiple outcomes. The non-linear regressions had a testing performance of 10.9 RMSE for Beta, and 0.46 for Generalized Normal (Table II-3).

The classifier only had good performance for the Generalized Normal class. The performance for Beta pens was impacted by the low number of Beta pens in the training data, with a high number of false negatives (347,626), a low number of false positives (28,246), a low number of true positives (39,559), and high true negatives (885,491) (Figure II-1). Generalized normal had 885,491 true positives, 39,559 true negatives, 347,626 false positives, and 28,246 false negatives (Figure II-1) and had high skill scores for classification with an F1 of 82%. The imbalanced data exhibited a bias towards the majority class, Generalized Normal, in its performance. Many Beta pens were removed as outliers, and their shape cannot handle atypical or multimodal DMI distributions. The addition of other distributions to the classification may balance the dataset. Or the Generalized Normal class could be under sampled prior to fitting. Given the high proportion of pens fit to Generalized Normal, a system where all pens are assumed as Generalized Normal distribution should also be evaluated.

The issue in classification was due to the nature of the candidates. The Beta can represent skewed data but will also describe data similar to a normal distribution. The Generalized Normal represents normal data with flexibility of kurtosis, and well describes data concentrated around the mean. This made it appropriate for livestock management as our practical goal should be to group animals by similarity in nutrient requirements. As the Beta distribution can represent shapes very similar to the Generalized Normal, there may be overlap in pens described, and the training features may not well distinguish them. Neither describe multimodal data, which could fail to describe pens with both primiparous and multiparous animals because of the difference in energy requirements.

For both non-linear regressions, the prediction error was low relative to the parameter values. The RMSE of the Beta model on testing was 10.9. The minimum and maximum of each parameter for Beta ranged from 0.32 – 80, 0.38 – 50, -60 – 22 and 6.8 – 161. Given the large ranges of some of these parameters (maximum of 154), an RSME of 10.9 is up to 6% from the true value. The RMSE of the Generalized Normal model on testing was 0.46, this error is averaged across the three Generalized Normal parameters, and their minimum and maximum ranges were 0.32 – 33, 12 – 28 and 0.033 – 12. The ranges of these parameters are smaller, and the error in prediction is also lower (up to 1%). The low error demonstrates that these parameters can be estimated to provide a way to describe the distribution curves of DMI. Given the similarity of Beta and Generalized Normal shapes, we hypothesized that in classification error (Generalized Normal predicted instead of Beta, or vice versa), DMI could be calculated by the incorrect shape and still produce a final DMI prediction with low error and this was tested in the DSM validation.

### *Feature importance*

Feature importance was calculated by their weights. The number of times a feature was used to split the data increased its weight value and this statistic was used to interpret the algorithm predictions (Table II-4). For the classification algorithm the parity category of a pen was the top three features. This was expected, as primiparous, and multiparous cows have systematically different DMI due to growth requirements. Only the Beta distribution can represent skewed data, making the parity category the most important feature. For the non-linear regression algorithms, parity category was also in the top four most important features. The maximum and minimum milk yield were also top features for all models, as they influence the



spread of DMI data. Pen size was one of the top three features for the Beta regression algorithm, but the fourth lowest feature of importance for the Generalized Normal regression algorithm. This may indicate that pen size had a large importance in fitting non-normal or skewed distributions such as the Beta (Table II-4).

### ***DSM pen DMI prediction validation***

To complete objective B a naïve dataset of pens was used for validation. After outlier removal these data were 229,909 pens, 148,311 fit to Generalized Normal and 81,598 fit to Beta. With the DSM, the classification algorithm predicted the distribution shape of DMI with a performance of 63% accuracy. The classification performance for Beta was a precision of 47%, recall 25% and F1-score 33%. The classification performance for Generalized Normal was precision of 67%, recall of 84% and F1-score of 75% (Table II-3). The classification of Beta pens on these validation data was higher performing than to the testing data because the validation data was less imbalanced by classification level (Figure II-2).

The non-linear regression algorithms of the DSM were applied to the validation data for the respective Beta and Generalized Normal pens as they were classified by the DSM. Pens correctly and incorrectly classified as Beta were applied to the Beta regression algorithm for parameter prediction, and similarly for Generalized Normal. Performance metrics were not calculated on these predictions, as incorrectly classified pens would falsely elevate the RMSE. A pen DMI was calculated for each predicted distribution curve, and observed versus predicted was plotted for total pen DMI (Figure II-3, A & B).

### *DSM pen DMI diagnostics*

For the DMI prediction of pens classified as Beta: the  $R^2$  was 0.95, MSEP was 5,249, with 3% of error due to mean bias, 5% due to the regression slope not equal to one, and 92% due to random bias. The RSR, CCC, r and Cb of the model were 0.22, 0.98, 0.98 and 0.99 (Table II-5). For the DMI prediction of pens classified as Generalized Normal: the  $R^2$  was 0.99, MSEP was 2,085, with 1% of error due to mean bias, 7% due to the regression slope not equal to one, and 92% due to random bias. The statistics of RSR, CCC, precision and accuracy were 0.12, 0.99, 0.99 and 0.99 respectively (Table II-5). The Generalized Normal model had lower total MSEP for prediction of pen DMI, and an RSR closer to 0, but both models indicated good performance based on these statistics. The CCC, precision and accuracy close to one for both models indicate a good ability to correctly predict this outcome.

### *NSM pen DMI validation and diagnostics*

Objective C was to contrast the DSM to a conventional method of calculating pen DMI, the NSM, with a deterministic equation to the average cow values of the pen, but with no correction factor. The DSM and NSM were compared to the observed DMI, and performance of both models was contrasted. The calculated total pen DMI of the NSM were plotted as observed versus predicted (Figure II-3). For the DMI prediction by the NSM of pens classified as Beta: the  $R^2$  was 0.55, the MSEP was 50,320, with 85% of the error due to mean bias, 10% due to the regression slope not equal to one, and 5% due to random bias. The RSR, CCC, r and Cb were 0.67, 0.84, 0.99 and 0.85 respectively (Table II-5). For the DMI prediction by the NSM of pens classified as Generalized Normal: The  $R^2$  was 0.69, the MSEP was 45,779, with error due to

mean bias of 73%, error due to regression slope not equal to one at 7%, and error due to random bias of 20%. The RSR, CCC, r and Cb were 0.55, 0.87, 0.97 and 0.90 respectively (Table II-5).

### ***Comparison of DSM and NSM***

Objective D compared the model diagnostics of the DSM and the NSM. The error for both Beta and Generalized Normal using the NSM was much higher than the DSM, with a 10-fold and 22-fold increase in MSEP respectively. This error was calculated on kg of DMI per pen, and the magnitude is large. The NSM had numerically large error as the DMI of every cow was assumed to be uniform, so every cow observed above or below this (in the distribution shape of that pen) added error in its prediction of pen DMI. The dimensionless statistics of RSR, CCC, r and Cb were also superior for the DSM, indicating better prediction ability. The error due to mean bias for the NSM in both Beta and Generalized Normal was high at 85% and 17% of the total error. This indicates systematic error, due to a consistent underprediction, seen in the observed versus predicted plots (Figure II-3). The RSR of 0.22 and 0.12 for Beta and Generalized Normal in the DSM signifies high predictive precision and reliable performance as the error is small relative to the variability of the observed values, compared to the higher values of RSR for the NSM of 0.67 and 0.55 for Beta and Generalized Normal. The NSM had values for CCC, r and Cb all above 0.84, showing consistent prediction, but the distribution shape DMI models had higher values for each of these metrics, at over 0.98.

The NSM is simplistic, as a DMI value was calculated using the pen means for each of its input parameters, but this is akin to the real-world application where pen averages are commonly used. A more robust approach would be to calculate the DMI for each cow in every pen by this

equation and sum those values. For this comparison that approach was not possible as the individual cow information of each of the 229,909 validation pens was not captured, but if individually assessing pens this may provide a better performing pen DMI prediction. The pens of these validation data only included a WIM up to 15, and the NASEM 2021 equation was empirically derived from mid-lactation cows (75 – 175 DIM), meaning it may not be appropriate for cows less than 10 WIM.

The consistent underprediction of the NSM illustrates that the mean is not an appropriate parameter for assessing dairy cow pens. Since the mean is a parameter of the Normal distribution, it should not be used if a different distribution shape is expected. In practice nutritionists may choose a percentile, such as the 75<sup>th</sup> of milk yield for DMI prediction, rather than the mean, to overcome the underprediction of high producing pens (Linn et al., 2021). Using either the mean or 75<sup>th</sup> percentile still provides no information on how the DMI is distributed within that pen. Understanding the distribution of DMI leads to calculation of nutrition quantities required by that animal group, rather than the uniform supply of each. This can enhance nutrient utilization and excretion, precision feeding, and ration cost if understood and applied to commercial dairy pens; formulating a ration to the pen's requirements will provide a more precise solution than using the mean, 75<sup>th</sup> percentile, or applying a correction factor to the ration. These approaches are a crude scaling of quantity, rather than the optimized solution appropriate to the correct requirements.

Both the Beta and Generalized Normal distribution shapes from the machine learning system of this paper appropriately predicted pen DMI, and both shapes can tolerate data

clustered at its center. The goal of dairy management is to group cows in pens that are similar in reproductive status, production, and nutrient need. In a normal distribution, 95% of the data falls within two standard deviations of the mean, but if the data is skewed left, or has heavy tails, then using the mean for a DMI equation like the NASEM 2021 will not well predict the true value. Starting first with a prediction of the expected shape of DMI distribution based on pen descriptive statistics, then calculating a pen DMI from this shape, we demonstrated high precision and accuracy. Explaining the distribution of DMI by pen will improve the precision of diet formulation, as the input constraint of DMI will be closer to the true value. This work also demonstrated that the distribution could be predicted from only descriptive information on the milk yield, WIM, and parity of the pen, removing the need to collect further data on individual cows.

## **CONCLUSION**

The distribution shape that described pen DMI was accurately predicted by a machine learning system of classification and non-linear regression with only information on pen level milk yield, parity, WIM, and pen size as features. This system was constructed into a model of DMI prediction, by calculating a set of DMI values that fit this distribution shape. This model precisely, accurately and consistently predicted total pen DMI in validation with naïve data. A similar model of DMI prediction was constructed by applying the NASEM 2021 DMI equation to each pen and assuming a uniform DMI of cows in each pen. This model had large mean bias and model error when compared to the observed DMI of these pens. Predicting the distribution of DMI with this model provides nutrient constraints for a pen that can be used to optimize a precise ration formulation.

## REFERENCES

- Bibby, J., Toutenburg, H. 1977. Prediction and Improved Estimation in Linear Models. Wiley & Sons.
- Chalupa, W., B. Vecchiarelli, D.T. Galligan, J.D. Ferguson, L.S. Baird, R.W. Hemken, R.J. Harmon, C.G. Soderholm, D.E. Otterby, R.J. Annexstad, J.G. Linn, W.P. Hansen, F.R. Ehle, D.L. Palmquist, and R.G. Eggert. 1996. Responses of Dairy Cows Supplemented with Somatotropin During Weeks 5 Through 43 of Lactation. *Journal of Dairy Science* 79:800–812. doi:[10.3168/jds.S0022-0302\(96\)76428-7](https://doi.org/10.3168/jds.S0022-0302(96)76428-7).
- Contreras-Govea, F.E., V.E. Cabrera, L.E. Armentano, R.D. Shaver, P.M. Crump, D.K. Beede, and M.J. VandeHaar. 2015. Constraints for nutritional grouping in Wisconsin and Michigan dairy farms. *Journal of Dairy Science* 98:1336–1344. doi:[10.3168/jds.2014-8368](https://doi.org/10.3168/jds.2014-8368).
- Council, N.R. 2001. Nutrient Requirements of Dairy Cattle: Seventh Revised Edition, 2001. The National Academies Press, Washington, DC.
- Dann, H.M., G.A. Varga, and D.E. Putnam. 1999. Improving Energy Supply to Late Gestation and Early Postpartum Dairy Cows. *Journal of Dairy Science* 82:1765–1778. doi:[10.3168/jds.S0022-0302\(99\)75407-X](https://doi.org/10.3168/jds.S0022-0302(99)75407-X).
- DePeters, E.J., N.E. Smith, and J. Acedo-Rico. 1985. Three or Two Times Daily Milking of Older Cows, and First Lactation Cows for Entire Lactations. *Journal of Dairy Science* 68:123–132. doi:[10.3168/jds.S0022-0302\(85\)80805-5](https://doi.org/10.3168/jds.S0022-0302(85)80805-5).
- Dhiman, T.R., and L.D. Satter. 1997. Yield Response of Dairy Cows Fed Different Proportions of Alfalfa Silage and Corn Silage. *Journal of Dairy Science* 80:2069–2082. doi:[10.3168/jds.S0022-0302\(97\)76152-6](https://doi.org/10.3168/jds.S0022-0302(97)76152-6).

- Fox, D.G., L.O. Tedeschi, T.P. Tylutki, J.B. Russell, M.E. Van Amburgh, L.E. Chase, A.N. Pell, and T.R. Overton. 2004. The Cornell Net Carbohydrate and Protein System model for evaluating herd nutrition and nutrient excretion. *Animal Feed Science and Technology* 112:29–78. doi:10.1016/j.anifeedsci.2003.10.006.
- Greenfield, R.B., M.J. Cecava, T.R. Johnson, and S.S. Donkin. 2000. Impact of Dietary Protein Amount and Rumen Undegradability on Intake, Peripartum Liver Triglyceride, Plasma Metabolites, and Milk Production in Transition Dairy Cattle. *Journal of Dairy Science* 83:703–710. doi:10.3168/jds.S0022-0302(00)74932-0.
- Halevy, A., P. Norvig, and F. Pereira. 2009. The Unreasonable Effectiveness of Data. *IEEE Intell. Syst.* 24:8–12. doi:10.1109/MIS.2009.36.
- Higgs, R.J., L.E. Chase, and M.E. Van Amburgh. 2012. Development and evaluation of equations in the Cornell Net Carbohydrate and Protein System to predict nitrogen excretion in lactating dairy cows. *Journal of Dairy Science* 95:2004–2014. doi:10.3168/jds.2011-4810.
- Huyler, M.T., R.L. Kincaid, and D.F. Dostal. 1999. Metabolic and Yield Responses of Multiparous Holstein Cows to Prepartum Rumen-Undegradable Protein. *Journal of Dairy Science* 82:527–536. doi:10.3168/jds.S0022-0302(99)75264-1.
- Lassen, J., J.R. Thomasen, and S. Borchersen. 2023. Repeatabilities of individual measures of feed intake and body weight on in-house commercial dairy cattle using a 3-dimensional camera system. *Journal of Dairy Science* 106:9105–9114. doi:10.3168/jds.2022-23177.
- Lean, I.J., R.L. Baldwin, H.F. Troutt, M.L. Bruss, J.C. Galland, T.B. Farver, J. Rostami, L.D. Weaver, and C.A. Holmberg. 1992. Impact of bovine somatotropin administration beginning at day 70 of lactation on serum metabolites, milk constituents, and production

- in cows previously exposed to exogenous somatotropin. *American Journal of Veterinary Research* 53:731–741. doi:10.2460/ajvr.1992.53.05.731.
- Linn, J., Hutjens, M., Shaver, R., Otterby, D., Howard, W. T., & Kilmer, L. 2021. Formulating dairy cow rations. University of Minnesota Extension. Accessed June 10, 2024. <https://extension.umn.edu/dairy-milking-cows/formulating-dairy-cow-rations>.
- Lucey, P.M., and H.A. Rossow. 2023. Describing the distribution type of dry matter intake for dairy cow pens based on pen characteristics. *animal* 100888. doi:10.1016/j.animal.2023.100888.
- National Academies of Sciences, Engineering, and Medicine. 2021. *Nutrient Requirements of Dairy Cattle: Eighth Revised Edition*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/25806>.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12:2825–2830.
- Seymour, D.J., A. Cánovas, C.F. Baes, T.C.S. Chud, V.R. Osborne, J.P. Cant, L.F. Brito, B. Gredler-Grandl, R. Finocchiaro, R.F. Veerkamp, Y. de Haas, and F. Miglior. 2019. Invited review: Determination of large-scale individual dry matter intake phenotypes in dairy cattle. *Journal of Dairy Science* 102:7655–7663. doi:10.3168/jds.2019-16454.
- Smirnov, N.V. 1939. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bull. Math. Univ. Moscou* 2:3–14.
- St-Pierre, N.R. 2007. Design and Analysis of Pen Studies in the Animal Sciences. *Journal of Dairy Science* 90:E87–E99. doi:10.3168/jds.2006-612.



Virtanen, P., R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, Í. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 17:261–272. doi:10.1038/s41592-019-0686-2.

## TABLES

**Table II-1.** Description of the cow records from the separate datasets used to build the database of this study.

Trial	n <sup>1</sup>			Breed	Weeks in Milk, wk <sup>5</sup>	DMI recording
	Total	Primi <sup>2</sup>	Multi <sup>3</sup>			
DePeters et al., 1985 <sup>6</sup>	55	15	40	Holstein	1 – 44	Calan gate <sup>10</sup>
Dhiman and Satter, 1997 <sup>7</sup>	74	29	45	Holstein	1 – 36	Calan gate
Greenfield et al., 2000 <sup>8</sup>	37	0	37	Holstein & Jersey	1 – 8	Calan gate
Unpublished <sup>8</sup>	45	0	45	Not stated	3 – 18	Calan gate
Unpublished <sup>8</sup>	40	0	40	Not stated	1 – 21	Calan gate
Huyler et al., 1999 <sup>10</sup>	31	0	31	Holstein	1 – 10	Calan gate
Chalupa et al., 1996 <sup>6</sup>	36	12	24	Holstein	3 – 43	Calan gate
Dann et al., 1999 <sup>8</sup>	57	57	0	Holstein	1 – 9	Calan gate
A. N. Hristov <sup>6</sup>	308	119	189	Holstein	1 – 48	Calan gate
Lean, 1992 <sup>9</sup>	88	0	88	Holstein	1 - 15	Calan gate
<b>Total</b>	<b>771</b>	<b>232</b>	<b>539</b>			

<sup>1</sup> Cow population per trial.

<sup>2</sup> Primiparous cows in their first lactation.

<sup>4</sup> Multiparous cows.

<sup>5</sup> The weeks of lactation for which each dataset recorded weekly individual cow milk yield and dry matter intake.

<sup>6</sup> Dataset recorded weekly average dry matter intake, milk, milkfat, milk protein and bodyweight.

<sup>7</sup> Dataset recorded weekly average dry matter intake, milk, milkfat, and milk protein.

<sup>8</sup> Dataset recorded weekly average dry matter intake, milk, milkfat, milk protein, bodyweight, and BCS.

<sup>9</sup> Dataset reserved for validation and was not included in model development.

<sup>10</sup> American Calan, NH, USA.

**Table II-2.** Performance of the XGB algorithm with different feature sets by 5-fold cross validation.

XGB algorithm: <sup>1</sup>	Milk only <sup>2</sup>	Milk and solids <sup>3</sup>	Full feature set <sup>4</sup>
<u>Classification model<sup>5</sup></u>			
Accuracy	0.71	0.71	0.71
Precision	0.72	0.72	0.72
Recall	0.96	0.96	0.96
F1-score	0.82	0.82	0.82
<u>Regression models</u>			
Beta			
RMSE <sup>6</sup>	10.9	10.9	10.9
Generalized Normal			
RMSE	0.47	0.47	0.46

<sup>1</sup> Extreme Gradient Boosting.

<sup>2</sup> Training features included descriptive statistics of milk yield, parity, and week in milk.

<sup>3</sup> Training features included descriptive statistics of milk yield, parity, week in milk, milk fat, and milk protein.

<sup>4</sup> Training features included descriptive statistics of milk yield, parity, week in milk, milk fat, milk protein, bodyweight, and BCS.

<sup>5</sup> Scores of Precision, Recall, and F1-score were averaged across both classification levels.

<sup>6</sup> Root mean squared error.

**Table II-3.** Model performance of the hyperparameter tuned XGB algorithm with both testing and validation data.

	Classification				Regression
	Accuracy <sup>2</sup>	Precision	Recall	F-1 score	RMSE <sup>3</sup>
<hr/>					
Performance to testing data					
Beta		0.58	0.10	0.17	10.9
Generalized Normal		0.72	0.97	0.82	0.46
	0.71				
<hr/>					
Performance to validation data					
Beta		0.47	0.25	0.33	
Generalized Normal		0.67	0.84	0.75	
	0.63				

<sup>1</sup> Extreme Gradient boosting.

<sup>2</sup> Accuracy is calculated on both classification levels.

<sup>3</sup> Root mean squared error.

**Table II-4.** Feature importance of the three XGB1 trained and hyperparameter tuned models.

Rank	Classification model	Beta Regression model	Generalized Normal Regression model
1	Parity: Heifers & Cows	Parity: Heifers only	Parity: Heifers only
2	Parity: Heifers only	Parity: Cows only	Parity: Cows only
3	Parity: Cows only	Pen size	Parity: Heifers & Cows
4	Maximum milk yield	Parity: Heifers & Cows	75 <sup>th</sup> percentile milk yield
5	Minimum WIM <sup>2</sup>	Maximum milk yield	Maximum milk yield
6	25 <sup>th</sup> percentile WIM	Milk yield SD	95 <sup>th</sup> percentile milk yield
7	5 <sup>th</sup> percentile WIM	Minimum milk yield	Mean milk yield
8	Maximum WIM	95 <sup>th</sup> percentile milk yield	Median milk yield
9	Pen size	Minimum WIM	Median WIM
10	95 <sup>th</sup> percentile milk yield	25 <sup>th</sup> percentile WIM	25 <sup>th</sup> percentile WIM
11	75 <sup>th</sup> percentile milk yield	5 <sup>th</sup> percentile WIM	Milk yield SD
12	Median WIM	75 <sup>th</sup> percentile milk yield	Minimum WIM
13	75 <sup>th</sup> percentile WIM	Skew of milk yield	5 <sup>th</sup> percentile WIM
14	Minimum milk yield	Median WIM	75 <sup>th</sup> percentile WIM
15	Milk yield SD	5 <sup>th</sup> percentile milk yield	25 <sup>th</sup> percentile milk yield
16	WIM SD	Mean WIM	95 <sup>th</sup> percentile WIM
17	Mean milk yield	95 <sup>th</sup> percentile WIM	Mean WIM
18	25 <sup>th</sup> percentile milk yield	Maximum WIM	Maximum WIM
19	Median milk yield	Median milk yield	5 <sup>th</sup> percentile milk yield
20	Kurtosis of milk yield	Mean milk yield	WIM SD
21	Mean WIM	75 <sup>th</sup> percentile WIM	Pen size
22	95 <sup>th</sup> percentile WIM	25 <sup>th</sup> percentile milk yield	Kurtosis of milk yield
23	5 <sup>th</sup> percentile milk yield	WIM SD	Skew of milk yield
24	Skew of milk yield	Kurtosis of milk yield	Minimum milk yield

SD = Standard Deviation

<sup>1</sup> Extreme Gradient Boosting.

<sup>2</sup> Week in milk.

**Table II-5.** Model diagnostics of the Dry Matter Intake prediction of both the distribution shape model and the NASEM equation model when compared to observed.

Model:	MSEP <sup>1</sup>	% bias	% Slope not 1	% Random Variation	R- squared	RSR <sup>2</sup>	CCC <sup>3</sup>	r <sup>4</sup>	Cb <sup>5</sup>
<b>DMI distribution</b>									
Beta	5,249	3	5	92	0.95	0.22	0.98	0.98	0.99
Generalized Normal	2,085	1	7	92	0.99	0.12	0.99	0.99	0.99
<b>NASEM equation</b>									
Beta	50,320	85	10	5	0.55	0.67	0.84	0.99	0.85
Generalized Normal	45,779	73	7	20	0.69	0.55	0.87	0.97	0.90

<sup>1</sup> MSEP = mean square error of prediction

<sup>2</sup> RSR = ratio of RMSEP/Standard deviation

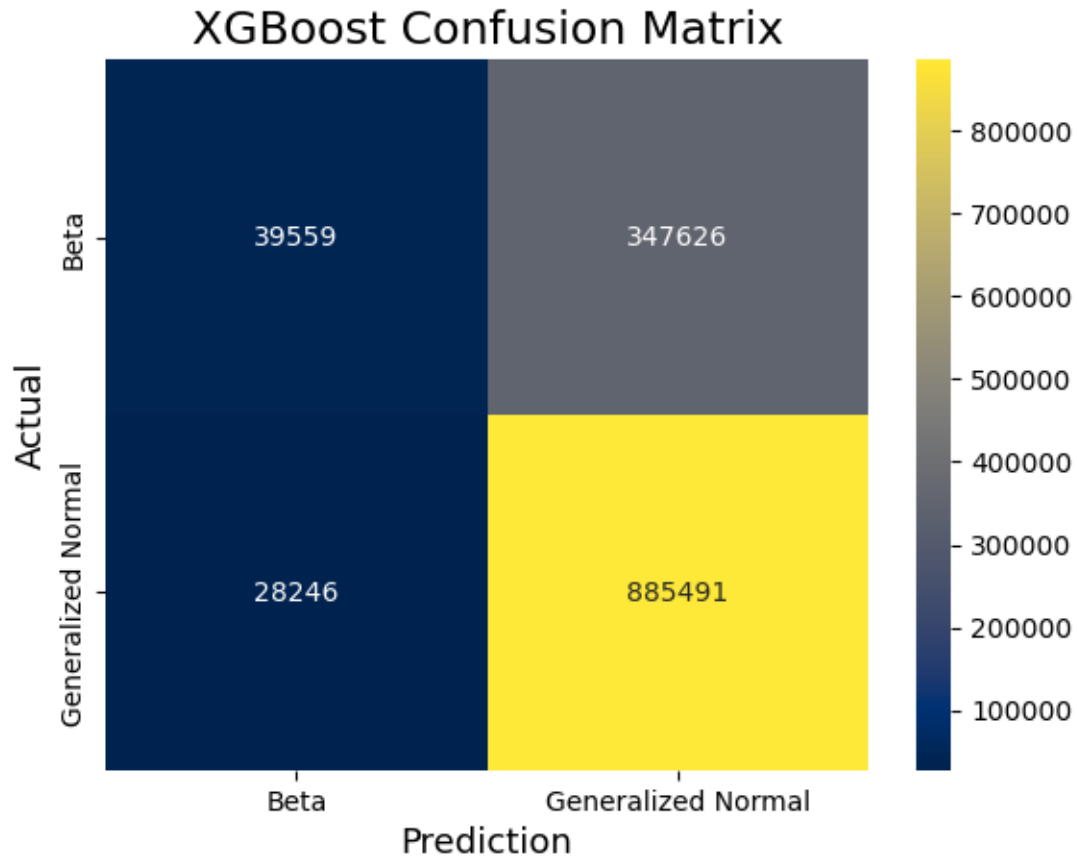
<sup>3</sup> CCC = Concordance correlation coefficient. CCC = r\*Cb

<sup>4</sup> r = Correlation coefficient.

<sup>5</sup> Cb = Bias correction factor.

## FIGURES

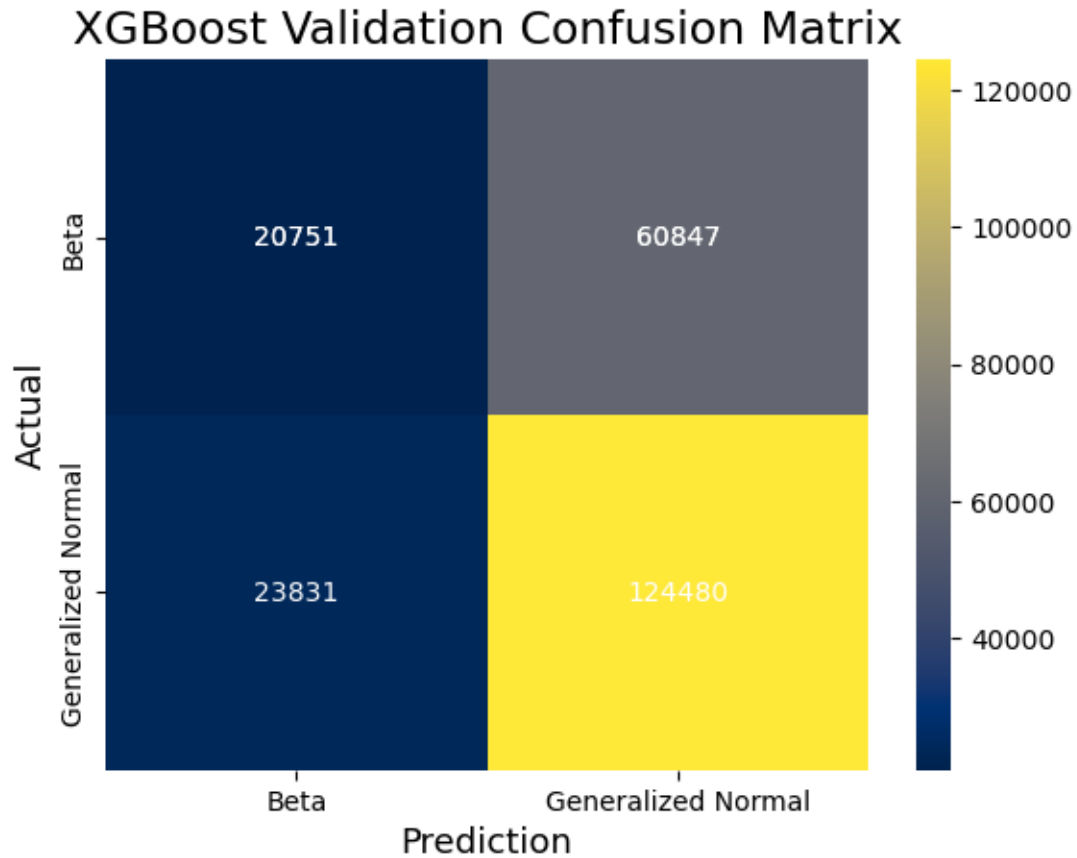
**Figure II-1.** A confusion matrix of the actual and predicted classification classes of the Extreme Gradient Boosted (XGBoost) trained and hyperparameter tuned model.<sup>1,2</sup>



<sup>1</sup> True positives and negatives are the quadrants with a matching class for both actual and prediction.

<sup>2</sup> False positives and negatives are the quadrants where the class of actual and prediction are mismatched.

**Figure II-2.** Figure 2. A confusion matrix of the actual and predicted classification classes of the Extreme Gradient Boosted (XGBoost) trained and hyperparameter tuned model for the validation data.<sup>1,2</sup>



<sup>1</sup> True positives and negatives are the quadrants with a matching class for both actual and prediction.

<sup>2</sup> False positives and negatives are the quadrants where the class of actual and prediction are mismatched.



**Figure II-3.** Observed versus Predicted plots of Dry Matter Intake by pen plotted for the Distribution Shape Model (DSM), and the NASEM Equation Model (NSM), for a) pens classified as Beta distribution, and b) pens classified as Generalized Normal Distribution.

Figure II-3a.

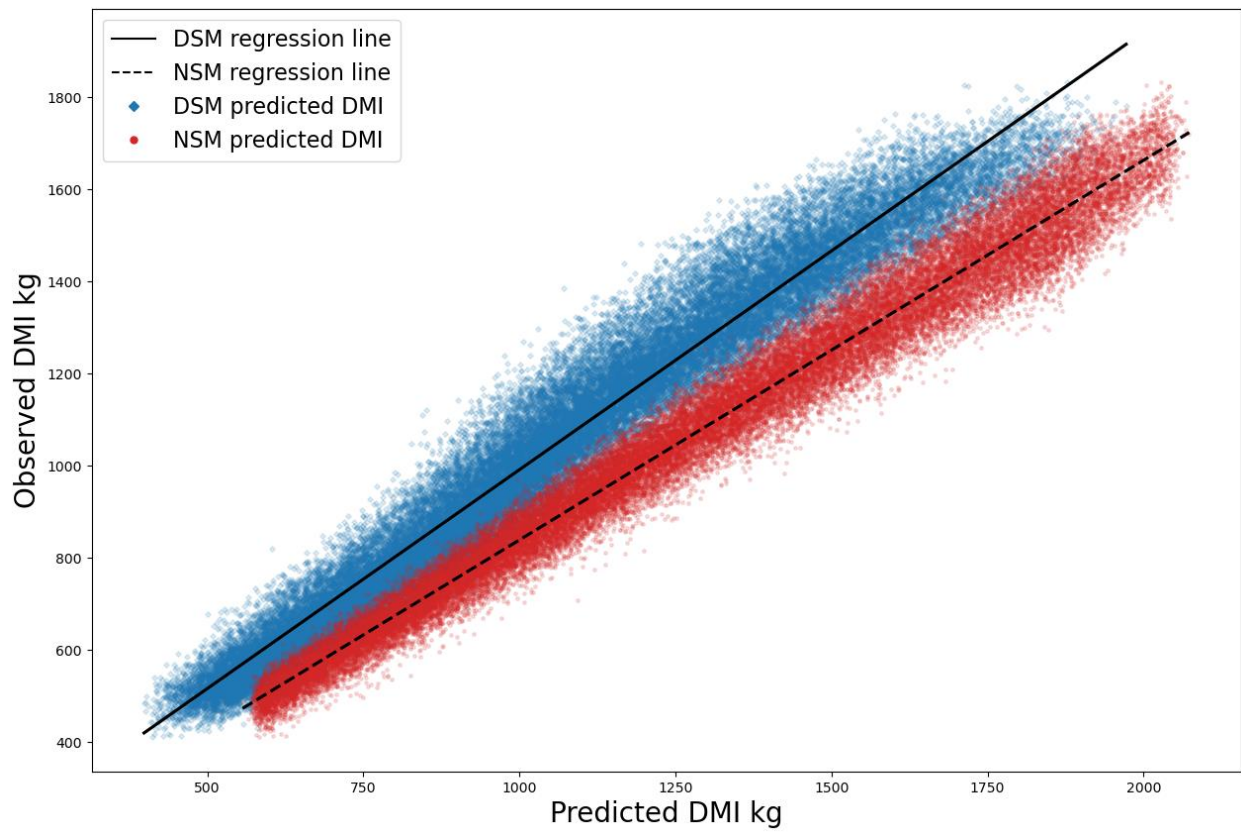
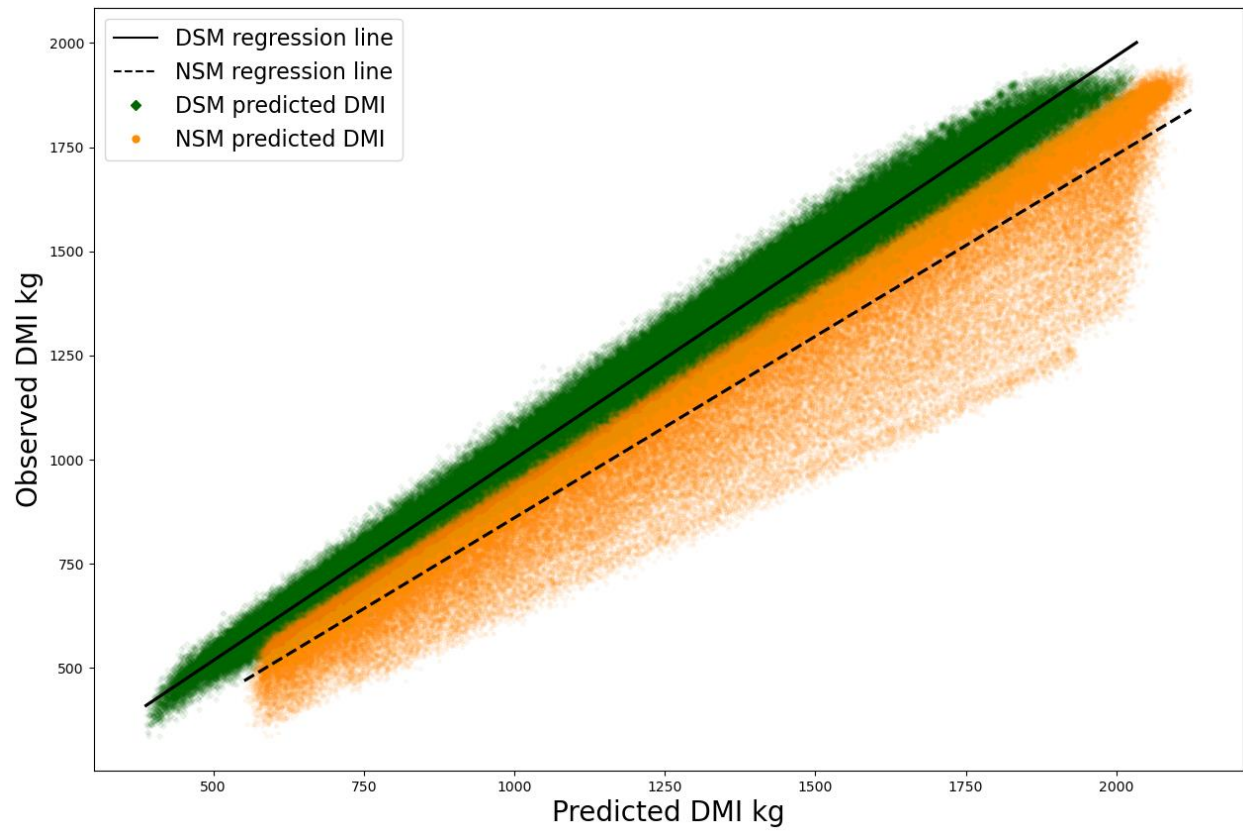


Figure II-3b.



## APPENDIX

### 1. Pen level bounds for assortment from individual cow weekly observations.

Pen Type	Value range			n of pens
	Earliest Week	Latest Week	Pen Size <sup>1,2</sup>	
Fresh	1	3, 4, 5	25 – 100	10,360
Early	1	6 – 41	25 – 250	390,300
Peak	3 - 16	8 – 41	50 – 250	4,092,140
Late	10 - 20	18 – 45	50 – 250	2,698,240
Total				7,191,040

<sup>1</sup> Increments of 1.

<sup>2</sup> All pens were assorted with primiparous cows only, multiparous cows only, or both.

## 2. The range of values of descriptive statistics of all pens in the database after outlier removal.

	Mean	Median	Min	Max	5 <sup>th</sup> quantile	25 <sup>th</sup> quantile	75 <sup>th</sup> quantile	95 <sup>th</sup> quantile	Variance	Skew	Kurtosis
DMI (kg)	12.2-27.3	12-27.8	5.3-22.2	15.4-36.4	7.5-22.8	10.6-25.5	13.7-30.7	15.1-35.6	2.4-49	-2.3-2.4	-1.5-10.4
Milk (kg)	22.9-47.3	23.5-48.4	22.9-37.3	29.1-68.7	7.7-29.3	18.1-43.5	26.2-54.9	28.5-63.4	8.8-193.9	-1.9-1.7	-1.5-4.9
Milk fat (kg)	0.9-3.7	0.9-4.2	0.07-1.3	1.3-5.1	0.2-1.4	0.6-3.7	1.1-4.5	1.2-4.9	1.9-2.9	-2.2-3.9	-1.9-15.9
Milk protein (kg)	0.7-3.1	0.7-3.3	0.2-1.2	0.9-3.9	4.1-2.8	0.6-3.1	0.8-3.4	0.8-3.7	0.004-1.5	-3.9-5.3	-1.9-3.5
Bodyweight (kg)	488-667	485-681	363-606	542-839	367-608	443-640	511-728	538-795	352-11285	-2.5-2.3	-1.8-8
BCS	1.25-4.5	1.25-4.5	1-4.5	1.25-4.5	1-4.5	1.25-4.5	1.25-4.5	1.25-4.5	0-2.25	-3-3	-2-1.1
Week	2-30	1-32	1-21	3-44	1-21	1-26	2-37	3-43	0-159	-	-

### 3. List of all distributions tested to the cow pens for selection of best fitting shape.

All distributions and their probability density functions were obtained from the SciPy package of

Python 3.10 (Python Software Foundation, <https://www.python.org/>) (Virtanen, 2020).

norm: Normal (Gaussian) distribution  
beta: Beta distribution  
expon: Exponential distribution  
gamma: Gamma distribution  
genextreme: Generalized Extreme Value distribution  
logistic: Logistic distribution  
lognorm: Log-normal distribution  
triang: Triangular distribution  
uniform: Uniform distribution  
fatiguelife: Fatigue life (Birnbaum-Saunders) distribution  
gengamma: Generalized Gamma distribution  
gennorm: Generalized normal distribution  
dweibull: Double Weibull distribution  
dgamma: Double Gamma distribution  
gumbel\_r: Right-skewed Gumbel distribution (Extreme Value Type I)  
powernorm: Power normal distribution  
rayleigh: Rayleigh distribution  
weibull\_max: Maximum Weibull (inverse Weibull) distribution  
weibull\_min: Minimum Weibull distribution  
laplace: Laplace (double exponential) distribution  
alpha: Alpha distribution  
genexpon: Generalized Exponential distribution  
bradford: Bradford distribution  
betaprime: Beta prime distribution  
burr: Burr (Type XII) distribution  
fisk: Fisk (Log-logistic) distribution  
genpareto: Generalized Pareto distribution  
hypsecant: Hyperbolic Secant distribution  
halfnorm: Half-normal distribution  
halflogistic: Half-logistic distribution  
invgauss: Inverse Gaussian distribution  
invgamma: Inverse Gamma distribution  
levy: Levy distribution  
loglaplace: Log-Laplace distribution  
loggamma: Log-Gamma distribution  
maxwell: Maxwell distribution  
mielke: Mielke's Beta-Kappa distribution  
ncx2: Non-central Chi-squared distribution  
ncf: Non-central F distribution  
nct: Non-central Student's t distribution

nakagami: Nakagami distribution  
pareto: Pareto distribution  
lomax: Lomax distribution (Pareto Type II)  
powerlognorm: Power log-normal distribution  
powerlaw: Power-law distribution  
rice: Rice distribution  
semicircular: Semicircular distribution  
trapezoid: Trapezoidal distribution  
invweibull: Inverse Weibull distribution  
foldnorm: Folded normal distribution  
foldcauchy: Folded Cauchy distribution  
cosine: Cosine distribution  
exponpow: Exponential power distribution  
exponweib: Exponential Weibull distribution  
wald: Wald (Inverse Gaussian) distribution  
wrapcauchy: Wrapped Cauchy distribution  
truncexpon: Truncated Exponential distribution  
truncnorm: Truncated Normal distribution  
t: Student's t distribution  
rdist: R-distribution

**4. Algorithm performance comparison using the default hyperparameters with the features of pen size, milk yield, parity, and week in milk by 5-fold cross validation.**

Model algorithms:	Decision tree	Random forest	XGB <sup>1</sup>
<u>Classification model</u>			
Accuracy	0.62	0.70	0.71
Precision	0.73	0.72	0.72
Recall	0.72	0.95	0.96
F1-score	0.73	0.82	0.82
<u>Regression models</u>			
Beta			
RMSE <sup>2</sup>	16.1	11.2	10.9
Generalized Normal			
RMSE	0.68	0.51	0.47

<sup>1</sup> Extreme Gradient Boosting.

<sup>2</sup> Root mean squared error

## CHAPTER III.

### PRECISION RATION FORMULATION BY PREDICTING THE DISTRIBUTION OF DRY MATTER INTAKE OF LACTATING COW PENS

#### ABSTRACT

Precision nutrition cannot be achieved for individual dairy cows, as they are managed and fed by pen. When individual dry matter intake (**DMI**) is not recorded, the distribution within the pen can be accurately predicted with machine learning. Rations are conventionally formulated imprecisely for an approximation of a pen's DMI and nutrient requirement. This work used least cost linear programming to assess precision nutrition formulation approaches for cow pens with different estimates of DMI. The observed DMI (**TRU**) of each cow was known, individual DMIs were predicted by the distribution shape generated from a machine learning algorithm (**DSM**), and individual DMIs were predicted with an empirical equation (**iNSM**). Cow observations were sorted into three pens, Fresh, High lactating, and Low lactating, based on their weeks in milk. The different DMI values were used to formulate an individual ration for each cow, and the resulting DMI, metabolizable energy, and ration cost were compared by pen. Next, a total pen nutrient requirement was calculated as the sum of the constituent cows. A whole pen ration solution was formulated for three (Fresh, High lactating, Low lactating) pens using this nutrient requirement, and the DMI of the TRU, DSM, iNSM, and two additional models: a DMI calculation using the pen average in the empirical equation (**aNSM**), and a DMI calculation using the pen 75<sup>th</sup> percentile in the empirical equation (**fNSM**). The aNSM and the fNSM represented imprecise dairy nutrition approaches, and their solutions were compared to the TRU diet, which was the ideal. The DSM model was the closest to the TRU in individual cow, and



pen ration formulation solutions. In the individual cow rations, there was no difference between DSM and TRU in DMI, cost and metabolizable energy, and for the pen rations its solutions were almost identical to the TRU for Fresh, High and Low pens. All models using an empirical DMI equation, including the precision approach of the iNSM, were inaccurate in their prediction of individual and pen rations. This demonstrates that precision nutrition can and should be applied on a population level when cows are managed in this manner to avoid systematic bias.

## INTRODUCTION

Precision nutrition is defined as minimizing the error between differences in nutrient requirements and nutrients provided. Knowing and supplying an exact ration for each dairy cow would maximize sustainability and milk production efficiency, less nutrients would be wasted, and cow performance would not be limited by dietary factors (Bewley, 2010; Meijer & Peeters, 2010; Maltz, 2020). This is not practiced; cows are fed by pen at an assumed constant DMI. Individually feeding cows for their energy needs does increase milk yield, solids, and energy-corrected milk per kg of ration dry matter (Maltz, 2013), but is not practical compared to the time and fuel-efficient large group feeding of a TMR (Schingoethe, 2017). Ration formulation should be adapted to precisely feed pens by setting the pen as the unit of feeding because dairy cows are managed by pen on commercial dairy farms.

Pens are assembled to manage reproduction, milk yield, stage of lactation (Weiss, 2017), or grouped by metabolizable energy and protein requirements (Cabrera, 2012). This results in a distribution of DMI that is usually not normal (Lucey and Rossow, 2023). In all these scenarios, rations are formulated based on the ‘average cow’ of the pen, which does not exist. The mean milk yield, composition, bodyweight, and BCS are used to estimate DMI and energy for the

ration solution. Therefore, half the cows will be underfed, and the other half overfed (Maltz, 2020), necessitating over-formulation to sustain high milk production (Linn et al., 2021). The distribution of DMI is ignored, and over-formulation imprecisely supplies nutrients.

Measuring the individual DMI of cows within a pen is not practical in commercial settings (Seymour, 2019), so the true distribution is unknown. But machine learning technology can predict this shape with accuracy and precision from information based on milk yield (Chapter II). The total DMI of the pen is known from the distribution shape and provides ration solutions that account for inter-cow variation within the pen.

The hypothesis of this work is that formulating rations by pen and using estimates of pen DMI distribution shape will lead to improved precision nutrition. The objective of this research is to use five DMI prediction models to formulate least cost rations for fresh, high and low milking pens to compare the impact of precision of pen DMI prediction on ration solutions. The first model represents the ‘true’ ration DMI (**TRU**), which is the actual DMI of the pens. The second is a model that represents the distribution shape of DMI for the pens (**DSM**; Chapter II) The third is the (NASEM 2021; Eq. II-1 & 2) DMI empirical model that can predict individual cow DMI that is summed to represent pen DMI (**iNSM**), the fourth model also uses the DMI equation from NASEM 2021 to predict pen DMI but uses data from the average cow in the pen (**aNSM**), and the 5<sup>th</sup> model uses the DMI equation from NASEM 2021 to predict DMI but uses data from the 75<sup>th</sup> percentile of the pen which in effect adds an adjustment factor to the DMI (**fNSM**) (Linn et al., 2021). Ration ingredient profiles, nutrients and cost will be compared between these models to determine which is closest to the TRU ration formulation.

## MATERIALS AND METHODS

No approval from the University of California, Davis, Animal Care and Use Committee was needed for this study as it was conducted using only previously collected data.

### *Pen selection and description*

Three pens were assembled from a database of 8,239 weekly observations of 443 individual lactating dairy cows, after removal of outliers above the 95<sup>th</sup> percentile for milk protein % and milk fat %. They were assembled by randomly selecting cows from this observation pool for the size of each pen, as per Lucey and Rossow (2023). The three pens were Fresh: fifty cows of WIM from one to three, High: one hundred and fifty cows of WIM from four to twenty-seven, and Low: one hundred cows of WIM from twenty-eight to forty-four (Table III-1). These ranges were chosen to reflect current industry practices of pen groups (Campos et al., 2023). All three pens included random mixtures of primiparous and multiparous cows. Only one observation per cow was included, and the WIM, and weekly average milk yield (kg) and DMI (kg) were required for inclusion into the dataset. Some observations also included bodyweight (kg), BCS, milk fat (%), and milk protein (%), but were not excluded if these data were missing. If these data were missing, the mean observation value of the pen was imputed for the missing values. For example, BCS was missing in all the Low pen data, so BCS was set to three for all cows.

### *Distribution shape prediction of pen DMI*

As previously described, the distribution shape model (DSM) with a machine learning pipeline was used to predict the distribution of DMI for the three pens. This required the inputs

pen size, parity category (primiparous, multiparous, or mixed), and descriptive statistics (mean, median, minimum, maximum, 5<sup>th</sup>, 25<sup>th</sup>, 75<sup>th</sup>, 95<sup>th</sup> percentile, standard deviation, skew and kurtosis) of milk yield and WIM. The model predicted the distribution shape as Beta or Generalized Normal, and the parameters of the selected shape. The DMI was calculated as the area under the distribution curve.

#### *The three NSM empirical predictions of pen DMI*

The three NSM models for DMI used Eq. II-1 & 2 to calculate pen DMI but differed by what data was used in the calculations. The iNSM used each cow's observed milk yield (kg), DMI (kg), milk fat (%), milk protein (%), parity, bodyweight (kg), BCS and DIM to calculate a DMI, and the sum was the pen DMI. The aNSM used the mean data from the pen for each input, and fNSM used the 75<sup>th</sup> percentile of the pen for the input variables, this calculated one DMI value for each pen, which was multiplied by the number of cows for each pen DMI estimate.

#### *Ration formulation program development*

A least cost linear program was developed using the PuLP package (Roy, J., et al, 2005) in Python 3.12 (Python Software Foundation, <https://www.python.org/>) to solve rations to the nutrient requirements of dairy cows (Moraes et al., 2012). The ration objective function was to minimize cost (Eq. III-1) where  $x$  is the amount (kg of DM) of feed  $i$ , and  $c$  is the cost (\$/kg of DM) of feed  $i$ , and  $n$  is the number of feeds available to the solution.

$$\text{Minimize } COST = \sum_{i=1}^n c_i x_i \quad \text{Eq. III-1}$$

Every cow of each pen was assigned nutrient requirements for metabolizable energy (**ME**) Mcal/kg, Rumen-degradable protein (**RDP**) %, Rumen-undegradable protein (**RUP**) %, Crude protein (**CP**) %, and Neutral Detergent Fiber (**NDF**) % based on their days in milk, and parity (Table 21-1, NASEM 2021). A feed bank was available for the Fresh, High and Low pen based on common ingredient usage (Table III-2; Rossow and Aly, 2013), and the nutrient composition of each was the listed value of table 19-1 NASEM 2021. The listed digestible energy (DE) of the feeds was converted to metabolizable energy (ME) using equation 2-10 of the NRC 2001 (Eq. III-2) where EE is ether extract of the feed as a percent of DM.

$$ME(Mcal/kg) = [1.01 \times DE - 0.45] + 0.0046 \times (EE - 3) \quad \text{Eq. III-2}$$

Feed costs were obtained in \$/kg of DM from the Penn State feed prices for December 2023 (Penn State Feed Price List, 2023) and personal communication with a nutrition professional (Table III-2).

#### *Ration program constraints*

Maximum inclusion limits on a kg/kg of DM were included for Corn Silage, Dried Distillers Grains, Molasses and Urea (Table III-2). The DMI was constrained to 2.5% above and below the predicted value for every solution where  $y$  is the cow's observed or predicted DMI and  $x$  is the DM in kg of feed  $i$  (Eq. III-3).

$$y \times 0.975 \leq \sum_{i=1}^n x_i \leq y \times 1.025 \quad \text{Eq. III-3}$$

The ration value of ME and proportions of CP, RDP, RUP, NDF and starch were constrained by Eq. III-4. A minimum ME value was constrained to the nutrient requirement of Mcal/kg of DM. The CP was limited to a maximum of 20% kg DM, and RUP and RDP were constrained to minimums at the nutrient requirement (10% and 6.2 – 7.5% kg DM). The NDF was constrained between 25 and 35% kg DM, and starch between 18 and 33% kg DM.

$$\mathbf{u}_t \leq \frac{\sum_{i=1}^n x_i \omega_{ti}}{\sum_{i=1}^n x_i} \leq \mathbf{q}_t \quad \text{Eq. III-4}$$

Where  $x_i$  is the amount of feed  $i$ ,  $\omega_{ti}$  is the nutrient  $t$  proportion of feed  $i$  in kg/kg DMI,  $u_t$  is the minimum proportion of nutrient  $t$ , and  $q_t$  is the maximum proportion of nutrient  $t$ .

The ration NDF as forage proportion was constrained between 19 and 25% of kg/kg DM of the solutions by Eq. III-5. The feeds Legume hay-mature, Corn silage-typical, and Wheat straw were considered the forage feeds.

$$\mathbf{df} \leq \frac{\sum_{i'=1}^3 x_{i'} f_{i'}}{\sum_{i=1}^n x_i} \leq \mathbf{ef} \quad \text{Eq. III-5}$$

Where  $df$  is the minimum and  $ef$  is the maximum ration proportion of NDF from forage as kg/kg DM,  $x_{i'}$  is the amount of feed for the  $i'$  forage,  $f_{i'}$  is the NDF proportion of the  $i'$ th forage, and  $x_i$  is the amount of feed  $i$ . The summation of variable  $x_{i'}$  is the feeds that are considered forages and is a subset of  $x_i$ .

## ***Ration solutions***

### *Individual cow ration solutions*

A least cost ration was solved using the described formulator for every cow within a pen. The feeds available to each solution were restricted to the feed bank of the respective pen of the cow. A solution was generated three times for each cow: for TRU, DSM, and the iNSM models. The aNSM and fNSM were not used for individual cow ration solutions, as one DMI value was assigned to every cow in each pen. The distribution of the solutions for DMI, ME, and ration cost were plotted by pen for the three different models.

### *Pen ration solutions*

Combining individual cow ration solutions would incorporate different ingredients by cow, introducing low inclusions of certain ingredients. Single ration solutions for each pen were formulated to prevent this. For all models, the nutrient constraints other than DMI were a sum of the nutrient requirements for all cows within the pen. The % of CP, RDP, RUP, NDF and starch required by each cow were multiplied by their observed DMI to calculate the quantity required, then summed for the kg DM required by the pen. This total quantity of CP, RDP, RUP and NDF were divided by the pen DMI of each model to calculate their % requirements. These requirements were the constraints to solve a single ration for each pen by all five models.

## ***Statistical Analysis***

### *Model diagnostics of pen DMI predictions*

The distribution of DMI was plotted by kernel density plots for the TRU, DSM, and iNSM models for each pen (Figure III-1). The DSM and iNSM were individually compared to

the observed values from TRU for every cow in each pen by the mean square error of prediction (MSEP) partitioned into error due to mean bias, error due to slope not equal one, and error due to random variation. They were also evaluated by the coefficient of determination ( $R^2$ ), relationship between observed and predicted, root mean square error of prediction (RMSEP), ratio of RMSEP/standard deviation (RSR), concordance correlation coefficient (CCC), correlation coefficient ( $r$ ) and bias correction factor (Cb) (Bibby and Toutenberg, 1977). Observed (TRU) versus predicted regression plots for the DSM and iNSM models DMI predictions were evaluated (Figure III-2).

#### *Individual cow ration solutions*

The final ration solution DMI, ME, and cost values were compared by pen for homogeneity of variance (Levene, 1960), then with Welch's ANOVA (Welch, 1951) and Games-Howell post-hoc pairwise comparisons (Games & Howell, 1976) to test for a difference between groups of unequal variances for the DMI, ME and cost of the observed, DSM, and iNEM, aNSM, and fNSM ration solutions.

#### *Pen ration solutions*

All pen ration solutions for the models: DSM, iNSM, aNSM, and fNSM, were compared for their total values of DMI, ME to the solution of TRUE for a calculation of RMSE.



## RESULTS AND DISCUSSION

### *Pen DMI description and diagnostics of the distribution predictions*

The three pens were assembled from a database of individual cow observations then their predictions of DMI distribution by the two models: DSM and iNSM were compared by Kernel density plots (Figure III-1) and MSEP (Table III-3) to the DMI distribution of the TRU model. For the Fresh pen the DMI distribution of the DSM was similar to TRU in peak and spread, the iNSM data peak was at a higher value than TRU, and its data spread did not include the lower range of observed data, overpredicting the density. For the High pen the DSM was similar to the TRU in peak and spread; the iNSM data had a similar peak location, but the peak was much higher than the TRU, with narrower tails. The same was observed for the Low pen, the DSM data matched the TRU, but the iNSM described a higher and narrower peak.

Model diagnostics were calculated comparing the DSM and iNSM separately to TRU (Table III-3; Figure III-2). The DSM had lower MSEP, with a maximum of 1.4 for the Low pen, compared to a range of 11 – 35 for the iNSM, indicating higher model accuracy for the DSM. The proportion of error due to % mean bias was low for the Fresh and Low pen with the DSM at 1%, showing little bias in prediction, but was greater for the High pen at 57% suggesting some bias. The proportion of error due to regression slope not equal to one was moderate for the DSM, showing that the model has a different prediction scale to the TRU. Error due to random variation for the DSM was moderately low, indicating a good representation of the data variance by this model. The iNSM had a high proportion of error due to % mean bias for the Fresh and High pen at 71 and 24%, indicating consistent bias in prediction. The proportion of error due to regression slope not equal to one was less than 10% for all pens in the iNSM, showing that it

responded to changes appropriately, even though it is inaccurate overall. The proportion of error due to random variation for the iNSM was high, with 91% for the Low pen. This shows the iNSM is unreliable as it does not represent the variation of the observed data.

These model diagnostics demonstrated that predicting the distribution shape (DSM) was a more accurate and robust method to predict DMI when examining pens of cows. The iNSM employed an empirical equation that calculated a DMI for each individual cow based on their characteristics to describe the distribution by pen. The proportion of error due to regression slope not equal one was the lowest for that iNSM approach as the empirical equation could appropriately scale its response with changes in the individual, but when examined across a population (the cow pen) a large error was introduced as shown by the higher magnitude of MSEF when compared to the DSM. This is an example of the atomistic fallacy and its risk; when studying a group, such as the nutrient requirements of a cow pen, assumptions drawn from the individual level do not apply to the group as a whole. The empirical equation iNSM predicted DMI with a narrow data range and large peaks, in a normal distribution (Figure III-1) for each pen, but when compared to the TRU the pen distribution was not represented.

### ***Individual cow ration solutions***

A ration solution was formulated for every cow's individual nutrient requirement using the TRU, DSM and iNSM predicted DMI as constraints, and compared between the three pens. Violin plots visualized the distribution of the solved ration's DMI, ME and cost (Figure III-3). For the three pens, the distribution of DMI, ME and cost of each cow's ration solution was similar between the TRU and the DSM. The distribution of solved DMI for each pen was higher

and with a narrower data range for the iNSM than the two other ration solutions. For the pen ME and cost, the distribution by pen was similar for TRU, DSM and iNSM, but with a consistently higher mean and right skew for the iNSM.

The variance between the groups for DMI, ME and cost were unequal in all pens when assessed with Levene's test of homogeneity of variance ( $P < 0.05$ ), so Welch's ANOVA was used to compare the means of the observed, DSM, and iNSM ration solutions. On pairwise comparison the solution values of DMI, ME and cost for the iNSM were significantly different from the observed, and the DSM. The largest differences were in the Fresh pen; the iNSM ration cost was \$1.30 greater than the observed at \$5.60, and \$1.10 greater than the DSM. The equation used in the iNSM was empirically derived from observed cow data that did not include Fresh cows, and these ration solutions demonstrate it is inappropriate when predicting the DMI at that stage of lactation. The DMI and ME for the Fresh ration solutions was greater for the iNSM than the TRU or DSM. Employing this equation introduced error when applied to this pen type due to excess feed quantity and energy. The High pen ration solutions were also different with iNSM when compared to TRU and DSM. This ration cost was \$0.50 - \$0.60 more expensive and also supplied excess energy and feed. Given that this pen was the biggest (150 cows), this difference generates an excess cost of \$75/d, waste of 195 kg/d DM, and 585 Mcal/d when compared to the DSM ration solution. The Low ration solution had the smallest differences between the iNSM to TRU and DSM but was still higher for all values. On large dairies these ration inaccuracies compound to a substantial increase in cost, waste of nutrients and feed.

### *Pen ration solutions*

Pen rations were formulated to their total nutrient requirement with each model (Table III-5, 6 & 7). The TRU model used the observed DMI of each cow, so its solution was considered the ideal and most cost-effective. Formulating rations for individual cows is not applicable to commercial dairy management as they are fed by pens. Combining individual rations would affect the composition of a TMR; different ingredients may be present in different solutions and are diluted across the feed bunk at low levels. A single ration per pen is the common application of a TMR (Schingoethe, 2017), with some operations providing separate rations for cow types (e.g. Fresh, High, or Low), or feed the same ration to all pens, but at different quantities for different energy needs. The goal of this work was to provide a ration that can be applied similarly but formulated more precisely. The DSM predicted DMI with a distribution shape; the iNSM applied a DMI prediction equation to each cow. The aNSM and fNSM provided a single DMI for each pen which was multiplied by the number of cows; this assumes a constant DMI and represents industry usage of ration formulation when individual DMIs are not measured.

The DSM predicted a distribution very close to the TRU, and its pen ration was almost identical. The Fresh ration was formulated for 50 cows, and the DSM model under supplied the energy of this group by 6 kg DMI and 14 Mcal ME. In contrast the iNSM, aNSM, and fNSM over predicted this ration in cost by \$68, \$45, and \$78 more than TRU, and DMI and ME by up to 319 kg DM and 852 Mcal, while the DSM was \$2 more expensive. Maintaining energy balance in early lactation cows is important. However, all NSM models overpredicted the required ration, incurring monetary, and nutrient losses. If pens are overfed, the remainder can be

fed to others, but the composition of the remainder is unknown, wasteful, and may promote feeding spoiled rations. The results of this work show that different constraint inputs lead to different combinations of ingredients, even for the same pen. Rations should always be formulated with the best-known requirements, especially for high-value ingredients and nutrients. It is important to note that the Fresh pen in this work consisted of cows between WIM 1 and 3, while the NASEM equation used for all three NSM models was derived from cow data of WIM 10 and greater, so greater inaccuracy in this pen's DMI is expected.

The High and Low pen ration solutions were again almost identical for the TRU and DSM models. The RMSE of cost, DMI and ME were less than 1 for the High pen between these two and ranged from 2.6 – 26.6 for the Low pen. For the iNSM model the cost of High ration was overpredicted by \$33/pen/d, which is a higher cost of \$0.22 per cow/d as 150 cows were in this pen. However, the aNSM and fNSM models overpredicted cost by \$66/pen/d, and \$169/pen/d. The difference in cost for the Low pen was \$3/pen for the DSM, \$21/pen, \$23/pen, and \$58/pen for the iNSM, aNSM, and fNSM. The difference in cost between iNSM and aNSM was lower than for the High pen. Less variation in DMI may exist in later lactation pens as cows are beyond their milking peak, showing that approximating a pen's requirements with a pen descriptor, such as the mean or 75<sup>th</sup> percentile, is more appropriate in these pens, but still inaccurate when compared to the DSM.

These extra costs add unnecessary expense, especially as the number of cows per ration and the cost of the ration increases. The aNSM and fNSM models assumed a uniform DMI for the pens, ignoring distribution, and while the iNSM was closer to the TRU, is not as easily

applied. The NASEM equation requires multiple cow inputs for its use: individual cow milk solids, BW, and BCS, which may not be routinely collected on dairies without the necessary infrastructure. But milk yield, parity and WIM were the main inputs of the DSM, which can be obtained when dairies record the milk yield of each cow. Using the NASEM based models incorrectly balanced the feed ingredients for each pen and did not offer precision feed utilization. Even if excess was fed as waste to other animal groups, nutrient wasting occurs as these would not be balanced correctly for the recipient groups.

## **CONCLUSION**

It is understood that DMI is not a normal distribution for dairy cow pens. When the DSM model was applied, the nutrient requirements of a Fresh, High, and Low pen were accurately predicted. This was compared to models using the NASEM DMI equation for ration solutions. The model that applied the equation to each cow per pen (iNSM), to just the mean of the pen (aNSM), and the 75<sup>th</sup> percentile of the pen (fNSM) all overestimated the energy needs. Ration solutions using these models demonstrated that the DSM could provide a cost and ingredient composition close to the ideal ration, improving precision feeding and nutrient utilization.

## REFERENCES

- Bewley, J. 2010. Precision Dairy Farming: Advanced Analysis Solutions for Future Profitability. The First North American Conference on Precision Dairy Management. Toronto, Canada. <https://www.precisiondairyfarming.com/>
- Bibby, J., Toutenburg, H. 1977. Prediction and Improved Estimation in Linear Models. Wiley & Sons.
- Campos, L.M., H. Ringer, M. Chung, and M.D. Hanigan. 2023. Application of a mathematical framework for the optimization of precision-fed dairy cattle diets. *animal* 17:101001. doi:[10.1016/j.animal.2023.101001](https://doi.org/10.1016/j.animal.2023.101001).
- Games, P.A., and J.F. Howell. 1976. Pairwise Multiple Comparison Procedures with Unequal N's and/or Variances: A Monte Carlo Study. *Journal of Educational Statistics* 1:113–125. doi:[10.3102/10769986001002113](https://doi.org/10.3102/10769986001002113).
- Levene, H. 1960. Robust tests for equality of variances. *Contributions to probability and statistics* 278–292.
- Linn, J., Hutjens, M., Shaver, R., Otterby, D., Howard, Terry., Kilmer, L., University of Minnesota Extension, 2021, Formulating dairy cow rations. Accessed Aug. 16, 2024. <https://extension.umn.edu/dairy-milking-cows/formulating-dairy-cow-rations>
- Lucey, P.M., and H.A. Rossow. 2023. Describing the distribution type of DM intake for dairy cow pens based on pen characteristics. *animal* 17:100888. doi:[10.10168/j.animal.2023.100888](https://doi.org/10.10168/j.animal.2023.100888).
- Maltz, E. 2020. Individual dairy cow management: achievements, obstacles and prospects. *Journal of Dairy Research* 87:145–157. doi:[10.1017/S0022029920000382](https://doi.org/10.1017/S0022029920000382).
- Meijer, R., and K. Peeters. 2010. The Use of Precision Dairy Farming in Feeding and Nutrition.

- The First North American Conference on Precision Dairy Management. Toronto, Canada.  
<https://www.precisiondairyfarming.com/>
- Moraes, L.E., J.E. Wilen, P.H. Robinson, and J.G. Fadel. 2012. A linear programming model to optimize diets in environmental policy scenarios. *J Dairy Sci* 95:1267–1282.  
doi:[10.3168/jds.2011-4651](https://doi.org/10.3168/jds.2011-4651).
- National Academies of Sciences, Engineering, and Medicine (NASEM)., 2021. Nutrient Requirements of Dairy Cattle: Eighth Revised Edition. Washington, DC: The National Academies Press. <https://doi.org/10.17226/25806>.
- Penn State Feed Price List. 2023. Pennsylvania State University Cooperative Extension.  
Accessed June 9th, 2024. <https://extension.psu.edu/programs/files/feed-price-lists>
- Roy, J.S., Mitchell, S.A., Peschiera, F., PuLP: A Linear Programming Toolkit for Python (2.9.0),  
<https://pypi.org/project/PuLP/>
- Schingoethe, D.J. 2017. A 100-Year Review: Total mixed ration feeding of dairy cows. *Journal of Dairy Science* 100:10143–10150. doi:[10.3168/jds.2017-12967](https://doi.org/10.3168/jds.2017-12967).
- Seymour, D.J., A. Cánovas, C.F. Baes, T.C.S. Chud, V.R. Osborne, J.P. Cant, L.F. Brito, B. Gredler-Grandl, R. Finocchiaro, R.F. Veerkamp, Y. de Haas, and F. Miglior. 2019. Invited review: Determination of large-scale individual dry matter intake phenotypes in dairy cattle. *Journal of Dairy Science* 102:7655–7663. doi:[10.3168/jds.2019-16454](https://doi.org/10.3168/jds.2019-16454).
- Weiss, B. 2017. Grouping similar cows has its benefits. *Hoard's Dairyman*, 162, p.192. Accessed Aug. 14, 2024. <https://hoards.com/article-20655-grouping-similar-cows-has-its-benefits.html>



## TABLES

**Table III-1.** Description of pens.

	<b>Fresh pen</b>	<b>High pen</b>	<b>Low pen</b>
Animals (n)	50	150	100
DMI (kg/d)	11.6 – 25.4	10.4 – 32.4	9.4 – 29
Milk yield (kg/d)	14 – 49.6	16.1 – 57	8.7 – 57
Milk fat (%)	0.5 – 5.1	0.7 – 8.2	0.4 – 3.6
Milk protein (%)	0.5 - 5	0.6 – 3.8	0.24 – 2.7
WIM	1 - 3	4 – 27	28 – 43
BW (kg)	483 - 727	383 – 756	432 – 748
BCS	2.75 - 4	2 – 4.5	--- <sup>1</sup>

WIM = Weeks in milk.

BW = Bodyweight.

<sup>1</sup> Not recorded on any cows in this pen.

**Table III-2.** Ingredient nutrient composition, inclusion, and cost.

<b>Ingredient</b>	<b>NASEM Entry No.<sup>1</sup></b>	<b>Fresh<sup>2</sup></b>	<b>High<sup>2</sup></b>	<b>Low<sup>2</sup></b>	<b>Limit<sup>3</sup></b>	<b>Cost</b>
Legume hay, mature	90	✓	✓	✓	No limit	0.22 <sup>4</sup>
Corn silage, typical	48	✓	✓	✓	0.6	0.17 <sup>4</sup>
Oat hay mid-maturity	101	✓	✓	✓	No limit	0.098 <sup>5</sup>
Wheat silage, HEADED	171	✓	✓	✓	No limit	0.0087 <sup>5</sup>
Wheat straw	173	✓	✓	✓	No limit	0.15 <sup>4</sup>
Corn grain dry, coarse grind	255	✓	✓	✓	No limit	0.44 <sup>4</sup>
Corn gluten feed, dry	40	✓	✓	✓	No limit	0.27 <sup>4</sup>
Canola meal	28	✓	✓	✓	No limit	0.49 <sup>4</sup>
DDGS, high protein	60	✓	✓	✓	0.10	0.22 <sup>5</sup>
Molasses	99	✓	✓	✓	0.03	0.42 <sup>4</sup>
Urea	165	✓			0.0005	0.24 <sup>5</sup>
Soybean meal, expellers	142	✓	✓	✓	No limit	0.59 <sup>4</sup>
Wheat middlings	170	✓	✓	✓	No limit	0.27 <sup>4</sup>
Whey, dry	174	✓	✓		No limit	0.78 <sup>4</sup>
Cottonseed, whole	56	✓	✓	✓	No limit	0.44 <sup>4</sup>

<sup>1</sup> Entries from Table 19-1 of NASEM 2021.

<sup>2</sup> Available as an ingredient to the ration solution.

<sup>3</sup> In kg/kg of DM.

<sup>4</sup> In \$/kg of DM. Feed prices from the Penn State University feed price list December 2023.

<sup>5</sup> In \$/kg of DM. Feed prices from personal communication.

**Table III-3.** Model diagnostics of the Dry Matter Intake prediction of both the distribution shape model (DSM) and the NASEM equation model (iNSM) when compared to TRU.

<b>Model:</b>									
	<b>MSEP</b>	<b>% bias</b>	<b>% Slope not 1</b>	<b>% Random Variation</b>	<b>R-squared</b>	<b>RSR</b>	<b>CCC</b>	<b>r</b>	<b>Cb</b>
<b>DSM</b>									
Fresh pen	0.86	1	69	30	0.98	0.28	0.99	0.98	0.98
High pen	0.83	57	19	24	0.99	0.24	0.98	0.99	0.98
Low pen	1.4	1	62	37	0.96	0.34	0.93	0.98	0.95
<b>iNSM</b>									
Fresh pen	35	71	2	27	0.15	1.8	0.14	0.38	0.36
High pen	17	24	8	68	0.25	1.1	0.42	0.49	0.84
Low pen	11	6	3	91	0.22	0.94	0.42	0.47	0.88

DSM = Distribution shape DMI model

iNSM = Individual NASEM equation model

MSEP = mean square error of prediction

RSR = ratio of RMSEP/Standard deviation

CCC = Concordance correlation coefficient.  $CCC = r * Cb$

r = Correlation coefficient.

Cb = Bias correction factor.

**Table III-4.** Values of DMI, ME and ration cost for the Welch’s ANOVA and pairwise comparison between the TRU, DSM, and iNSM.

	<b>Comparison of ration model solutions</b>		
	TRU	DSM	iNSM
<b>Pen:</b>			
<b>Fresh</b>			
DMI (kg)	17.8 <sup>a</sup>	18.2 <sup>a</sup>	22.7 <sup>b</sup>
ME (Mcal)	47.3 <sup>a</sup>	48.4 <sup>a</sup>	60.6 <sup>b</sup>
Cost (\$)	4.3 <sup>a</sup>	4.5 <sup>a</sup>	5.6 <sup>b</sup>
<b>High</b>			
DMI (kg)	23.3 <sup>a</sup>	23.4 <sup>a</sup>	24.7 <sup>b</sup>
ME (Mcal)	62.2 <sup>a</sup>	62.5 <sup>a</sup>	66.4 <sup>b</sup>
Cost (\$)	5.7 <sup>a</sup>	5.8 <sup>a</sup>	6.3 <sup>b</sup>
<b>Low</b>			
DMI (kg)	19.9 <sup>a</sup>	19.9 <sup>a</sup>	20.9 <sup>b</sup>
ME (Mcal)	54.4 <sup>a</sup>	54.7 <sup>a</sup>	57.3 <sup>b</sup>
Cost (\$)	5.3 <sup>a</sup>	5.4 <sup>a</sup>	5.7 <sup>b</sup>

<sup>a-b</sup> Different superscript letters were significantly different by Games-Howell pairwise comparison.

DMI = Dry Matter Intake.

ME = Metabolizable Energy.

TRU = Observed DMI model

DSM = Distribution shape DMI model

iNSM = Individual NASEM equation model

**Table III-5.** Fresh pen ration solutions.

<b>n = 50</b>	<b>TRU</b>	<b>DSM<sup>1</sup></b>	<b>iNSM<sup>2</sup></b>	<b>aNSM<sup>3</sup></b>	<b>fNSM<sup>4</sup></b>
Cost, \$ <sup>5</sup>	217	215	276	262	295
DMI, kg <sup>6</sup>	890	884	1135	1077	1209
ME, Mcal <sup>7</sup>	2366	2352	3019	2863	3218
CP, %	17.4	17.4	17.3	17.3	17.5
RUP, %	7.36	7.37	7.33	7.31	7.5
RDP, %	10	10	10	10	10
NDF, %	35	35	35	35	35
NDF % forage	19	19	19	19	19
Starch, %	25	25	25	25	25
Corn silage, typical, kg	413	411	527	500	561
Wheat silage, kg	77	77	97	92	109
Corn grain dry, kg	70	69	90	86	94
Corn gluten feed, kg	129	128	167	159	165
DDGS, kg	89	88	113	108	121
Molasses, kg	26	26	34	32	36
Urea, kg	0.45	0.44	0.57	0.54	0.60
Soybean meal, kg	84	84	106	99	122
Ration total, kg	890	884	1136	1077	1209

DMI = Dry Matter Intake

TRU = Observed DMI model

DSM = Distribution shape DMI model

iNSM = Individual NASEM equation model

aNSM = Average cow NASEM equation model

fNSM = 75<sup>th</sup> percentile cow NASEM equation model

ME = Metabolizable Energy

CP = Crude Protein

RUP = Rumen Undegradable Protein

RDP = Rumen Degradable Protein

NDF = Neutral Detergent Fiber

DDGS = Dried Distillers Grains with Solubles

<sup>1</sup> Ration constraint inputs from the Distribution Shape Model (DSM).

<sup>2</sup> Ration constraint inputs from the iNSM equation model.

<sup>3</sup> Ration constraint inputs from the aNSM equation model that used the pen means.

<sup>4</sup> Ration constraint inputs from the fNSM equation model that used the pen 75<sup>th</sup> percentile.

<sup>5</sup> RMSE of 1.27, 59, 45, and 78 for the DSM, iNSM, aNSM and fNSM when compared to TRU Cost.

<sup>6</sup> RMSE of 5.39, 245, 187, and 319 for the DSM, iNSM, aNSM and fNSM when compared to TRU DMI.

<sup>7</sup> RMSE of 14, 652, 497, and 851 for the DSM, iNSM, aNSM and fNSM when compared to TRU ME.

**Table III-6.** High pen ration solutions.

<b>n = 150</b>	<b>TRU</b>	<b>DSM<sup>1</sup></b>	<b>iNSM<sup>2</sup></b>	<b>aNSM<sup>3</sup></b>	<b>fNSM<sup>4</sup></b>
Cost, \$ <sup>5</sup>	858	858	891	924	1027
DMI, kg <sup>6</sup>	3498	3498	3632	3771	4186
ME, Mcal <sup>7</sup>	9324	9324	9683	10052	11161
CP, %	17.4	17.4	17.4	17.3	17.5
RUP, %	7.36	7.37	7.33	7.31	7.5
RDP, %	10	10	10	10	10
NDF, %	35	35	35	35	35
NDF % forage	19	19	19	19	19
Starch, %	25	25	25	25	25
Corn silage, typical, kg	1,624	1,623	1,686	1,751	1,943
Wheat silage, kg	264	264	274	283	319
Corn grain dry, kg	262	262	273	283	312
Corn gluten feed, kg	573	573	595	619	677
DDGS, kg	350	350	363	377	419
Molasses, kg	105	105	109	113	126
Soybean meal, kg	321	321	333	345	391
Ration total, kg	3,499	3,498	3,633	3,771	4,177

DMI = Dry Matter Intake

TRU = Observed DMI model

DSM = Distribution shape DMI model

iNSM = Individual NASEM equation model

aNSM = Average cow NASEM equation model

fNSM = 75<sup>th</sup> percentile cow NASEM equation model

ME = Metabolizable Energy

CP = Crude Protein

RUP = Rumen Undegradable Protein

RDP = Rumen Degradable Protein

NDF = Neutral Detergent Fiber

DDGS = Dried Distillers Grains with Solubles

<sup>1</sup> Ration constraint inputs from the Distribution Shape Model (DSM).

<sup>2</sup> Ration constraint inputs from the iNSM equation model.

<sup>3</sup> Ration constraint inputs from the aNSM equation model that used the pen means.

<sup>4</sup> Ration constraint inputs from the fNSM equation model that used the pen 75<sup>th</sup> percentile.

<sup>5</sup> RMSE of 0.01, 33, 67, and 170 for the DSM, iNSM, aNSM and fNSM when compared to TRU Cost.

<sup>6</sup> RMSE of 0.02, 135, 273, and 688 for the DSM, iNSM, aNSM and fNSM when compared to TRU DMI.

<sup>7</sup> RMSE of 0, 360, 728, and 1837 for the DSM, iNSM, aNSM and fNSM when compared to TRU ME.

**Table III-7.** Low pen ration solutions.

<b>n = 100</b>	<b>TRU</b>	<b>DSM<sup>1</sup></b>	<b>iNSM<sup>2</sup></b>	<b>aNSM<sup>3</sup></b>	<b>fNSM<sup>4</sup></b>
Cost, \$ <sup>5</sup>	535	538	556	558	593
DMI, kg <sup>6</sup>	1988	1997	2067	2071	2203
ME, Mcal <sup>7</sup>	5545	5472	5663	5674	6036
CP, %	18.7	18.7	18.7	18.7	18.7
RUP, %	7.7	7.7	7.7	7.7	7.7
RDP, %	11	11	11	11	11
NDF, %	35	35	35	35	35
NDF % forage	19	19	19	19	19
Starch, %	21	21	21	21	21
Corn silage, typical, kg	923	927	959	962	1,023
Wheat silage, kg	0	0	0	0	0
Corn grain dry, kg	0	0	0	0	0
Corn gluten feed, kg	750	754	780	781	832
DDGS, kg	0	0	0	0	0
Molasses, kg	60	60	62	62	66
Soybean meal, kg	255	256	265	265	283
Ration total, kg	1,988	1,997	2,066	2,070	2,204

DMI = Dry Matter Intake

TRU = Observed DMI model

DSM = Distribution shape DMI model

iNSM = Individual NASEM equation model

aNSM = Average cow NASEM equation model

fNSM = 75<sup>th</sup> percentile cow NASEM equation model

ME = Metabolizable Energy

CP = Crude Protein

RUP = Rumen Undegradable Protein

RDP = Rumen Degradable Protein

NDF = Neutral Detergent Fiber

DDGS = Dried Distillers Grains with Solubles

<sup>1</sup> Ration constraint inputs from the Distribution Shape Model (DSM).

<sup>2</sup> Ration constraint inputs from the iNSM equation model.

<sup>3</sup> Ration constraint inputs from the aNSM equation model that used the pen means.

<sup>4</sup> Ration constraint inputs from the fNSM equation model that used the pen 75<sup>th</sup> percentile.

<sup>5</sup> RMSE of 2.61, 21.4, 22.5, and 58.1 for the DSM, iNSM, aNSM and fNSM when compared to TRU Cost.

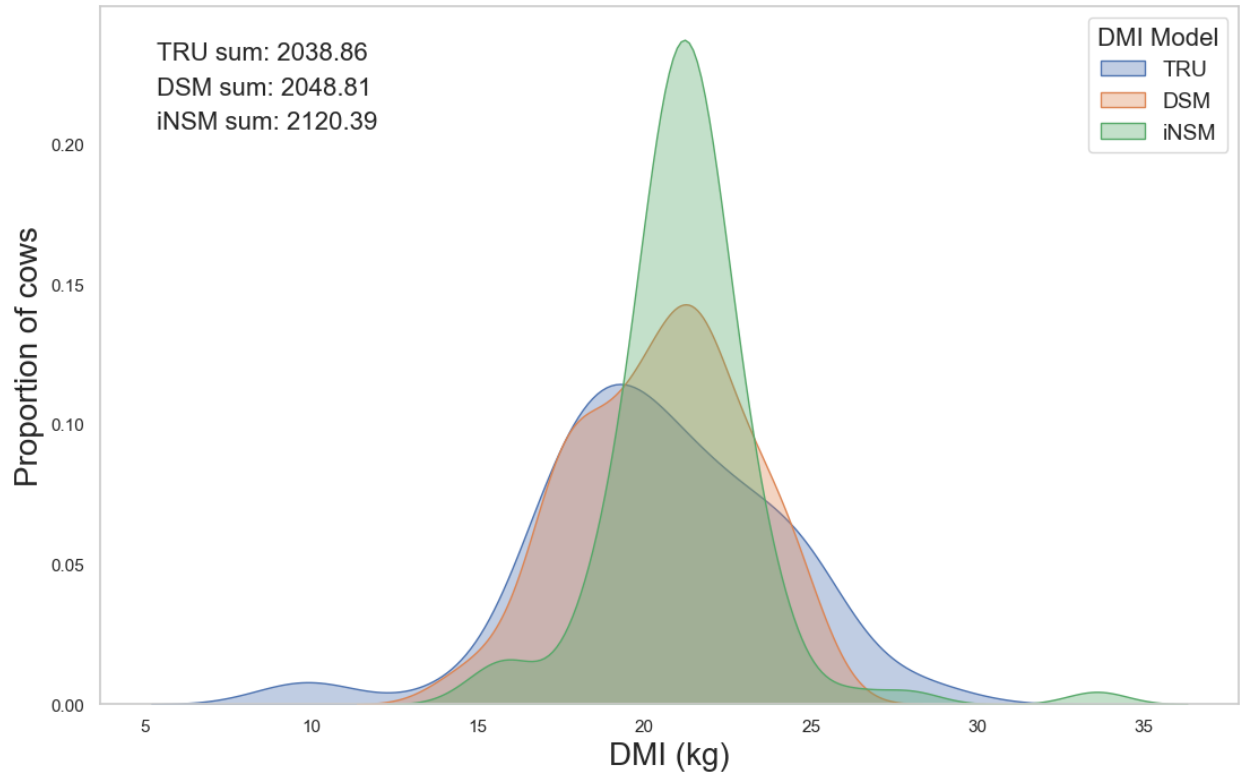
<sup>6</sup> RMSE of 9.71, 79.5, 83.4, and 215 for the DSM, iNSM, aNSM and fNSM when compared to TRU DMI.

<sup>7</sup> RMSE of 26.6, 217, 228, and 590 for the DSM, iNSM, aNSM and fNSM when compared to TRU ME.

## FIGURES

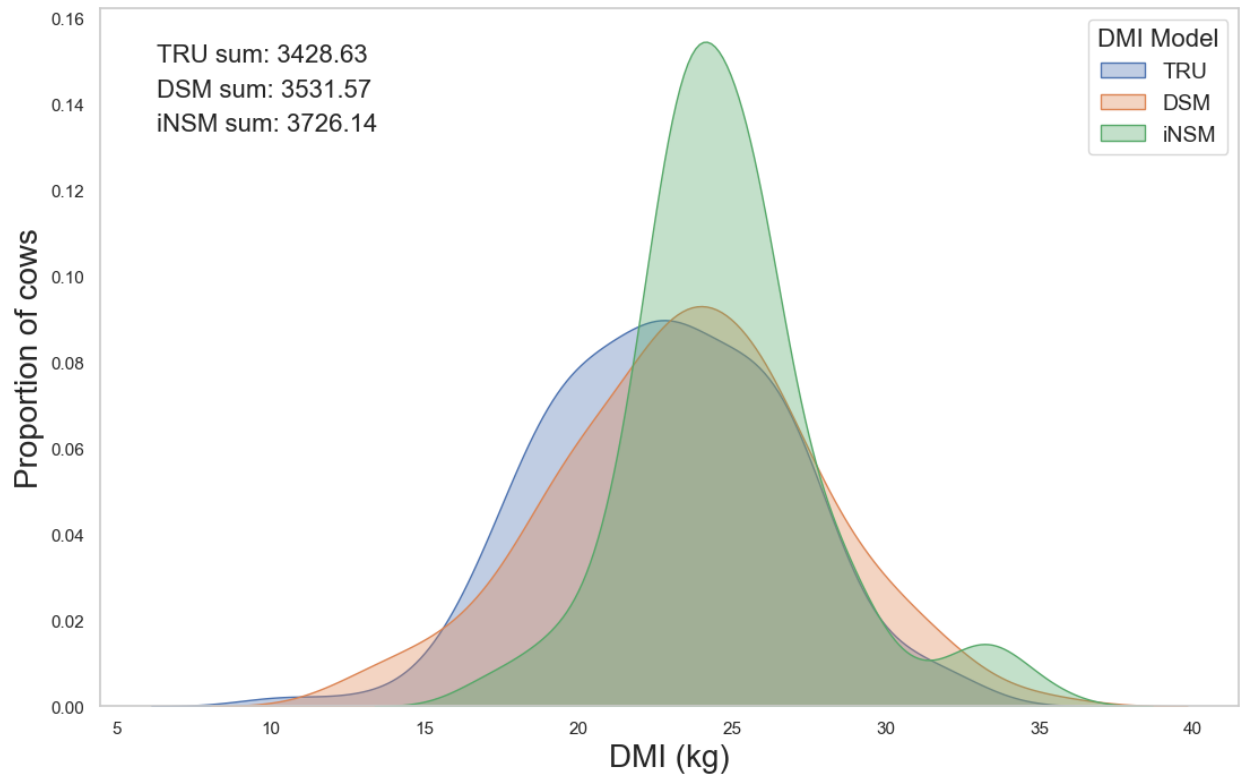
**Figure III-1.** Kernel density plots of the observed (TRU), Distribution Shape Model (DSM), and NASEM Equation Model (iNSM) dry matter intake values for the a) Fresh pen, b) High pen, and c) Low pen.

**Figure III-1a.**

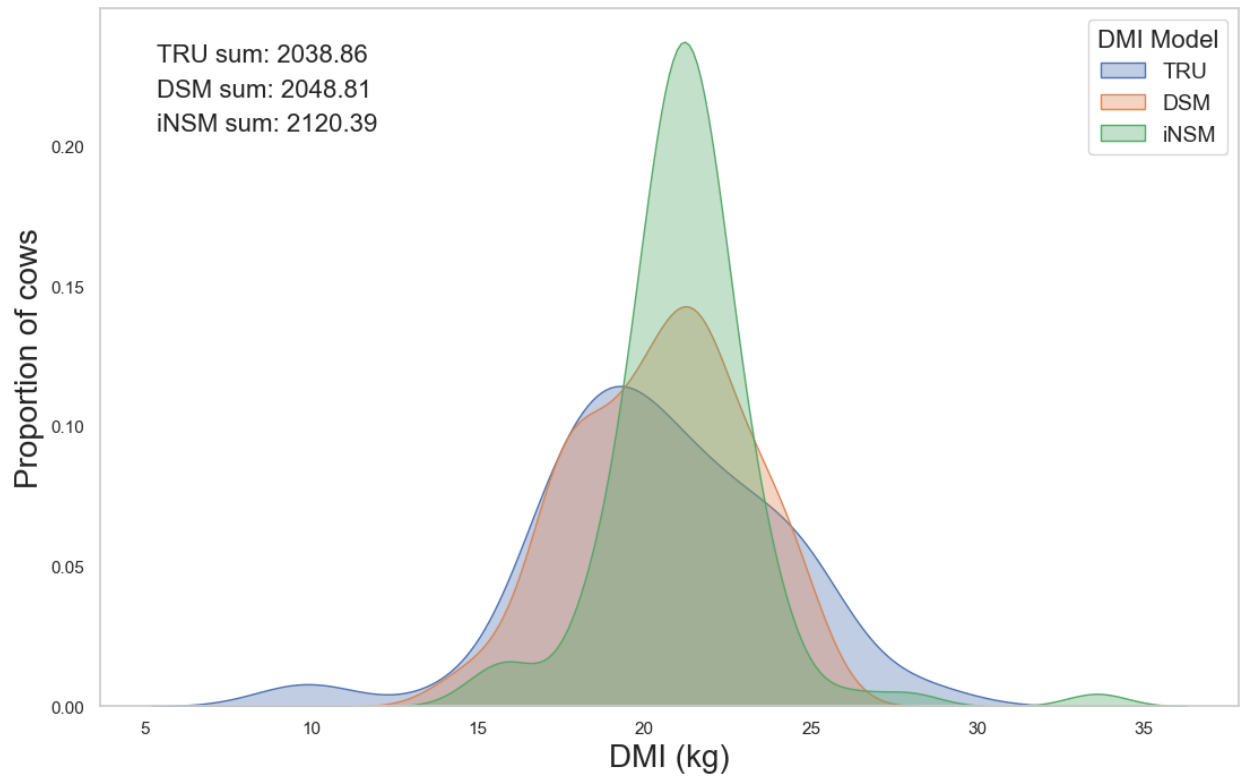




**Figure III-1b.**



**Figure III-1c.**



**Figure III-2.** Regression plots of the mean squared error of prediction of observed versus predicted dry matter intake values for the Distribution Shape Model (DSM) and the NASEM Equation Model (iNSM) for a) the Fresh pen, b) the High pen, and c) the Low pen.

**Figure III-2a.**

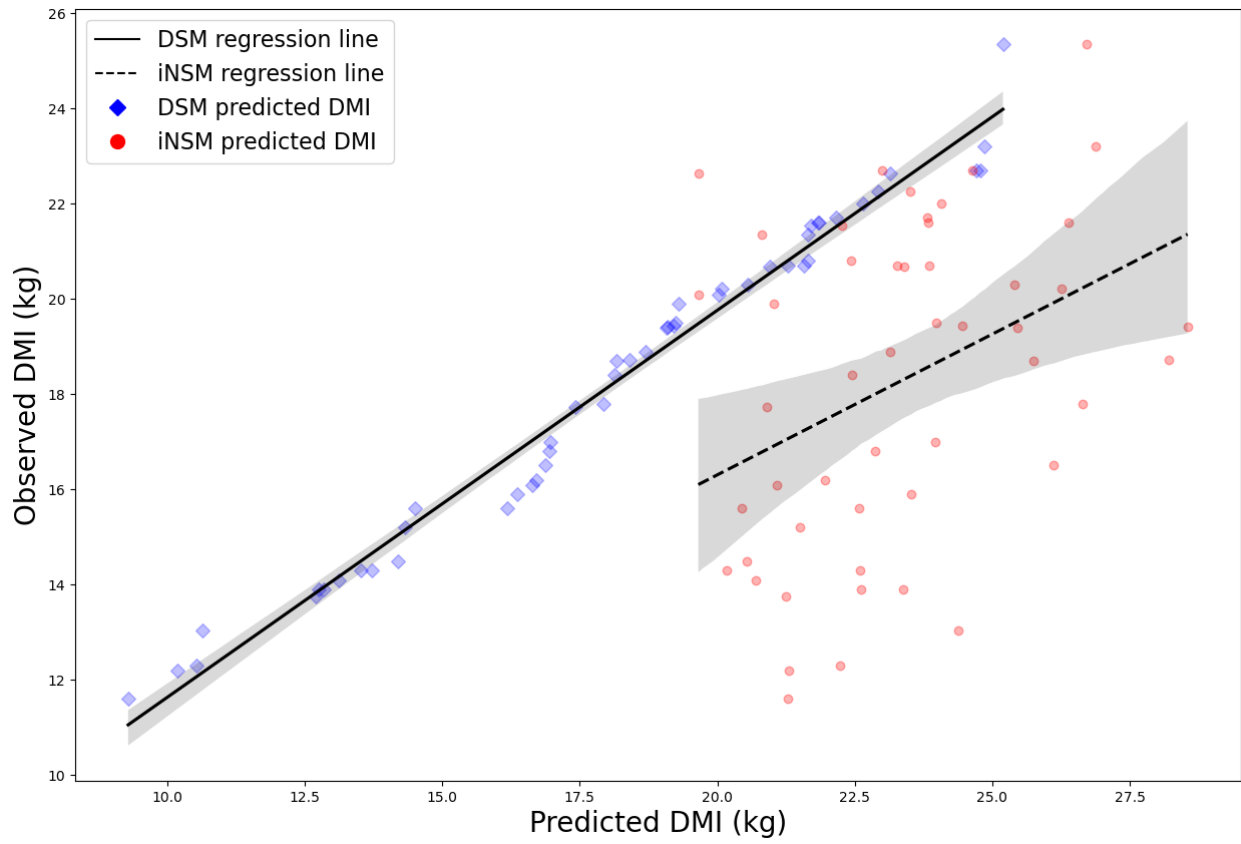


Figure III-2b.

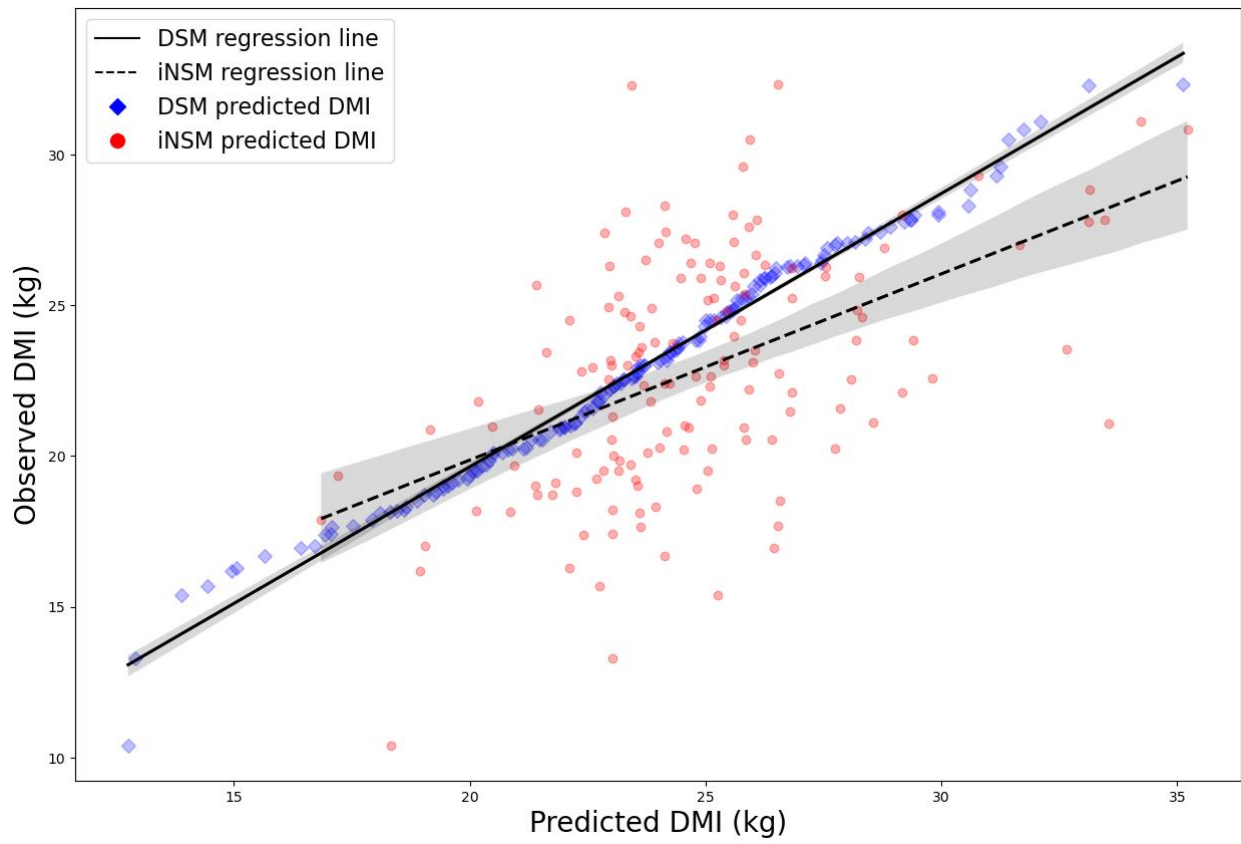
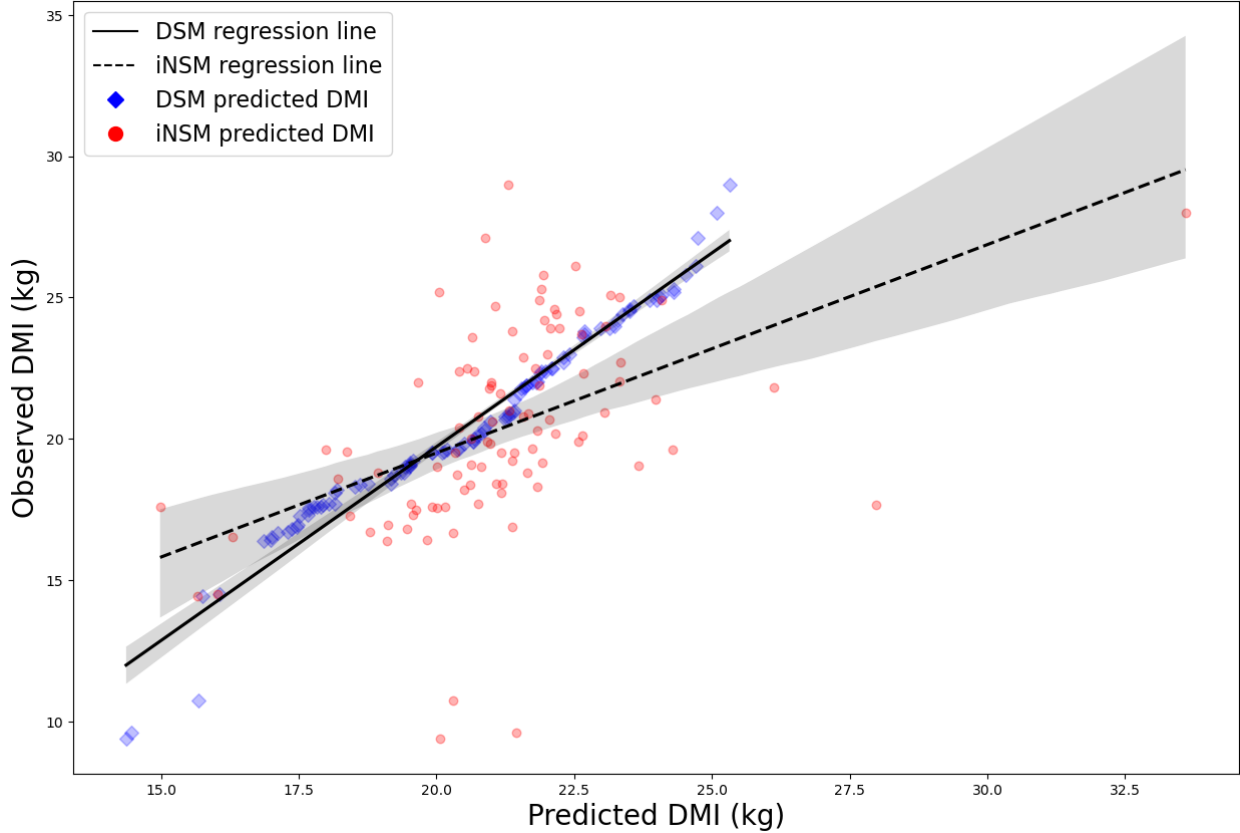
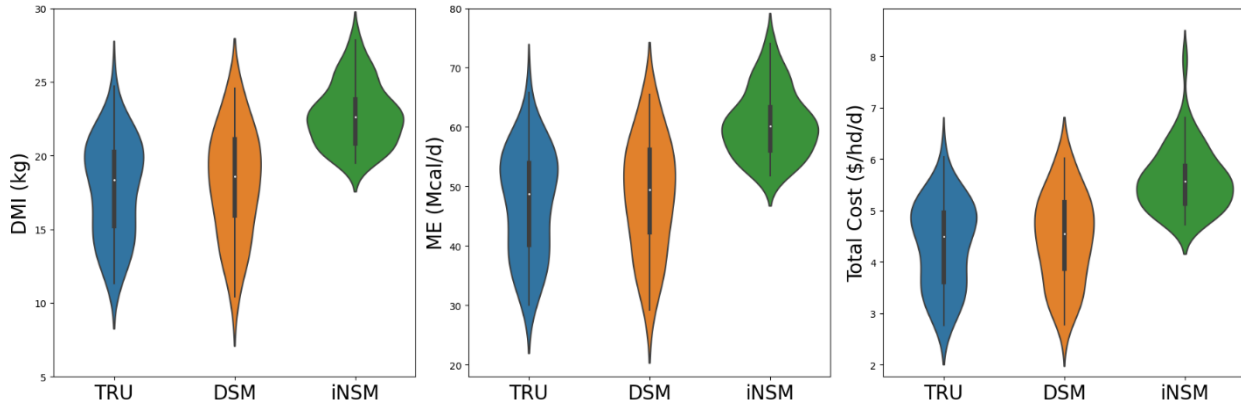


Figure III-2c.

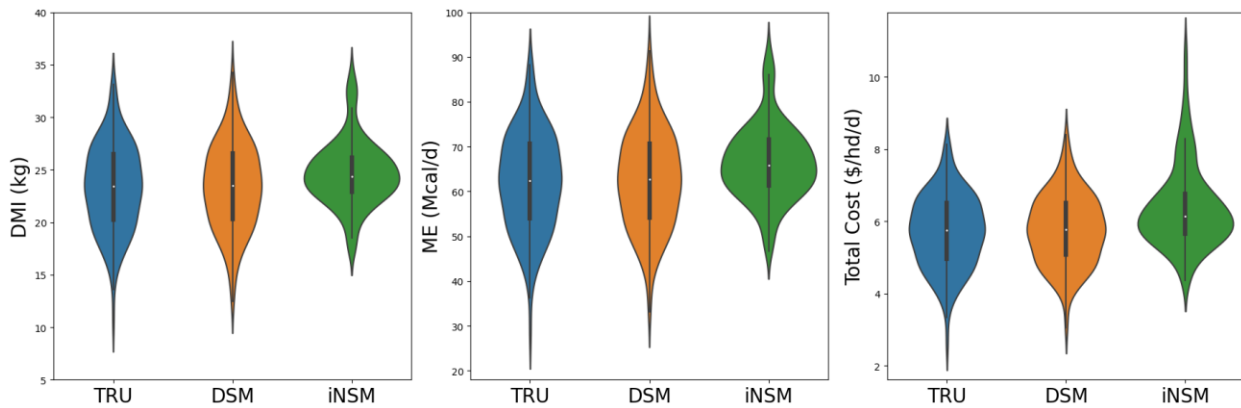


**Figure III-3.** Violin plots of the ration solution DMI, ME and cost for the observed and predicted DMI constraint inputs of a) the Fresh pen, b) the High pen, and c) the Low pen.

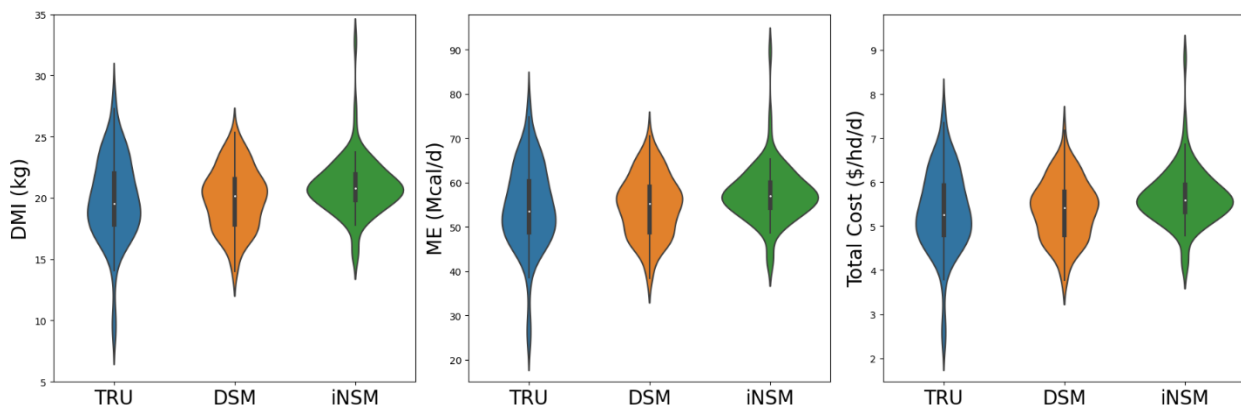
**Figure III-3a.**



**Figure III-3b.**



**Figure III-3c.**



## **CHAPTER IV.**

### **CONCLUSIONS**

Precision nutrition cannot be approached on dairies when the unit of feeding management is a pen, but the unit of ration formulation is a cow. This work approached the knowledge gap between these two levels. We described the distribution of pen DMI and the factors that influence its shape by examining DMI by week, and by pen type. Machine learning algorithms provided a prediction solution for this population descriptor. Previous empirical equations have used individual cow variables to plot a regression line that predicts DMI, but population level variables, such as the parity grouping, or size of a cow pen, do not have biologic relationships with the distribution shape that best fits a pen's DMI. This limits empirical deduction on the level of pen. A labelled dataset for the outcome of DMI distribution shape was able to train an appropriate machine learning model for the prediction of DMI by pen. When this was used for a precision nutrition ration formulation, the distribution information solved a ration that was not different in composition and cost to the ideal ration. Other evaluated approaches were inaccurate and imprecise and did not tolerate within pen variation of DMI. This demonstrated that a pen level approach is the appropriate form of precision nutrition for pen managed dairy cows.