Using Data Analytics in Agriculture to Make Better Management Decisions

Douglas Liebe

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Animal and Poultry Science

Robin R. White, Advisor
Zachary Easton
Mike Persia
Leanna House

February 18, 2020
Blacksburg, Virginia

Keywords: data analytics, modeling, decision-making, agriculture

Using Data Analytics in Agriculture to Make Better Management Decisions

Douglas Liebe

Abstract

The goal of this body of work is to explore various aspects of data analytics (**DA**) and its applications in agriculture. In our research, we produce decisions with mathematical models, create models, evaluate existing models, and review how certain models are best applied. The increasing granularity in decisions being made on farm, like individualized feeding, sub-plot level crop management, and plant and animal disease prevention, creates complex systems requiring DA to identify variance and patterns in data collected. Precision agriculture requires DA to make decisions about how to feasibly improve efficiency or performance in the system. Our research demonstrates ways to provide recommendations and make decisions in such systems.

Our first research goal was to clarify research on the topic of endophyte-infected tall fescue by relating different infection-measuring techniques and quantifying the effect of infection-level on grazing cattle growth. Cattle graze endophyte-infected tall fescue in many parts of the U.S and this feedstuff is thought to limit growth performance in those cattle. Our results suggest ergovaline concentration makes up close to 80% of the effect of

measured total ergot alkaloids and cattle average daily gain decreased 33 g/d for each 100ppb increase in ergovaline concentration. By comparing decreased weight gain to the costs of reseeding a pasture, producers can make decisions related to the management of infected pastures.

The next research goal was to evaluate experimental and feed factors that affect measurements associated with ruminant protein digestion. Measurements explored were 0-h washout, potentially degradable, and undegradable protein fractions, protein degradation rate and digestibility of rumen undegradable protein. Our research found that the aforementioned measurements were significantly affected by feedstuff characteristics like neutral detergent fiber content and crude protein content, and also measurement variables like bag pore size, incubation time, bag area, and sample size to bag area ratio. Our findings suggest that current methods to measure and predict protein digestion lack robustness and are therefore not reliable to make feeding decisions or build research models.

The first two research projects involved creating models to help researchers and farmers make better decisions. Next, we aimed to produce a summary of existing DA frameworks and propose future areas for model building in agriculture. Machine learning models were discussed along with potential applications in animal agriculture. Additionally, we discuss the importance of model

evaluation when producing applicable models. We propose that the future of DA in agriculture comes with increasing decision making done without human input and better integration of DA insights into farmer decision-making.

After detailing how mathematical models and machine learning could be used to further research, models were used to predict cases of clinical mastitis (CM) in dairy cows. Machine learning models took daily inputs relating to activity and production to produce probabilities of CM. By considering the economic costs of treatment and non-treatment in CM cases, we provide insight into the lack of applicable models being produced, and why smarter data collection, representative datasets, and validation that reflects how the model will be used are needed.

The overall goal of this body of work was to advance our understanding of agriculture and the complex decisions involved through the use of DA. Each project sheds light on model building, model evaluation, or model applicability. By relating modeling techniques in other fields to agriculture, this research aims to improve translation of these techniques in future research. As data collection in agriculture becomes even more commonplace, the need for good modeling practices will increase.

Using Data Analytics in Agriculture to Make Better Management Decisions

Douglas Liebe

General Audience Abstract

Data analytics (**DA**) has become more popular with the increasing data collection capabilities using technologies like sensors, improvement in data storage techniques, and expanding literature on algorithms that can be used in prediction and summarization. This body of work explores many aspects of agricultural DA and its applications on-farm. The field of precision agriculture has risen from an influx of data and new possibilities for using these data. Even small farms are now able to collect data using technologies like sensor-equipped tractors and drones which are relatively inexpensive. Our research shows how using mathematical models combined with these data can help researchers produce more applicable tools and, in turn, help producers make more targeted decisions. We examine cases where models improve the understanding of a system, specifically, the effect of endophyte infection in tall fescue pastures, the effect of measurement on protein digestibility for ration formulation, and methods to predict sparse diseases using big data. Although DA is widely applied, specific agricultural research on topics such as model types, model performance, and

model utility needs to be done. This research presented herein expands on these topics in detail, using DA and mathematical models to make predictions and understand systems while utilizing applicable DA frameworks for future research.

Acknowledgments

"The life of the creative man is lead, directed and controlled by boredom. Avoiding boredom is one of our most important purposes." -Susan Sontag

List of Tables

List of Figures

Management with Exports - a scenario where animals are kept under current management and dairy products are not consumed in the U.S.; and Retirement - dairy animals are retired to a pasture-based system.

Figure 3-3. Nutrient supply of current production compared to that of three dairy cow removal strategies in terms of human nutrient requirement years (HNRY) met, in millions.

Figure 3-4. Comparison of land use classifications under current conditions (Baseline) and a scenario where grain land previously used for dairy cattle feed is repurposed for fruit and vegetable production (LU-2). Although land use shifts involve only 6.8 million ha of land, the shifts in consumable food produced and in carbon emissions produced are substantial because of the high yields of fruit and vegetable products per unit of land area.

Figure 3-5. Consumable product and tonnes of CO2-equivilent (CO2-e), broken down by food category, of a land reallocation strategy where grain land previously used for dairy cattle feed is repurposed for fruit and vegetable production. Fruits and vegetables account for an increased amount of consumable product, but also disproportionately increase total GHG costs.

Figure 3-6. Comparison to current production system (baseline) of total greenhouse gas emissions under land reallocation scenarios using various crop replacement methods. (A) Replant vacated land with crops in proportion to their current production in the U.S. (B) Replant land with only fruits and vegetables (C) Replant land with only nuts and pulses (D) Replant land in proportion to current U.S. crop production, without planting any additional grain. No Grain Allocation = only reallocate land directly freed from removal of dairy animals or used for silage. Grain Allocation = reallocate land directly freed from removal of dairy animals, used for silage, and used for dairy feed grain.

Figure 3-7. Comparison of nutrient supplies among land use options in a scenario where dairy cattle are depopulated from the agricultural system. Nutrients shown are those that had reduced supply in one or more land use option compared with the baseline scenario.

Figure 4-1. Example of Principal Component Analysis using three gene variables. The three gene factors are combined into two components that maximize variance between groups in 2-dimensional space, as seen on the right (Scholz, 2006).

Table of Contents

Chapter 1: Introduction

Broadly, data can be any set of values (Shannon, 1948). Data become information or knowledge when they can be represented in context or through analysis (Long et al., 2016). Statistician Nate Silver says, "The numbers have no way of speaking for themselves. We speak for them. We imbue them with meaning" (Silver, 2012). In agriculture, we collect data on many decisions and actions. The intent of any data collection is that these data will be useful in making future decisions and learning from past experience. "Big data" is a popular term born from the dramatically increasing ability to collect and store more data because of improvements in technology (Francis, 2012; Reinsel et al., 2017). Big data does not have a single definition, but typically deals with data stored digitally, as opposed to analog, that is too complex to be analyzed by a human alone (Mashey, 1998).

Because big data may be too complex for human analysis, mathematical modeling has become a burgeoning field focused on finding patterns, explaining variance, and predicting outcomes using data without direct human insight. In general, mathematical models, or models, define a system that creates data. Models can be theoretical, based on laws or widely accepted truths, or experimental, meaning the model is derived from observations. Theoretical models can also be applied to observations to determine

the validity of our beliefs or to identify our relative uncertainty with a system. Model building is the first step to using data to derive or inform decisions regarding a given system.

A model, mathematical or otherwise, typically begins with information that a human accepts as true before seeing data, or assumptions (Boghossian 2017). An example of an assumption for building models would be assuming that gravity will govern the movement of an object. If someone throws a ball at you, you know that the ball will slowly drop over time as it approaches you. This assumption is a good starting point, but not enough information to help someone know exactly how to catch a ball or where it will land. In the same way, we may start building mathematical models using equations that we know will govern a system, like the equation for gravity to predict the endpoint of a thrown ball. We will still need to measure or estimate other properties of the ball, e.g. the angle of release and speed at release, but this *a priori* information will improve model performance. In research, once a model is built to understand a system, the model should be compared against reality whenever possible in an attempt to discern from where error in a model comes. Continuing the example with a ball, we may use a set of equations to predict the trajectory and endpoint of a thrown ball, but in reality, the ball lands in a different place. Because reality is not straightforward like equations, there can be

definite sources of error or inherent uncertainty in a system. We may not have accounted for wind in our model, thus producing slightly biased predictions. To determine the accuracy and precision of a model, model evaluation techniques are useful.

Model evaluation is a key part of mathematical modeling, as evaluation makes sure that a model is useful in the context of the system at hand. With empirical data to test models, fit statistics are a common way to evaluate the performance of a model. A statistic like mean squared error (MSE), which measures the average squared deviation of each prediction from the true value, increasingly weights greater deviations from the truth (Berger and Lehmann, 1984). Using MSE as a measure of model performance will select models that tend to predict values closer to the average. Other functions for measuring performance focus on the median error instead, which is less sensitive to outliers. Splitting data into two groups, one to parameterize the model and one to test the parameters using some statistical accuracy measure, is common when data are readily available. Splitting data aims to prevent models from only applying to specific data and not the overarching system. Cross validation, as this splitting technique is called, does not guarantee no overfit models (Wani et al., 2018). If the two halves of split data are dissimilar, the model will not be robust. Accuracy measures should be tailored to the model, as there are many ways for a model to be considered accurate. A model that is

only accurate on average would not be useful in modeling a system where high impact, low frequency events occur (high variability), e.g. earthquakes (Dieterich and Kilgore, 1996; Stewart et al., 2015). In a case like earthquake prediction, there is massive amounts of seismic data on non-earthquake events, but the few earthquake events are of much greater importance and their relative infrequency should be reflected in the model training data. Systems that generate great amounts of data are reliant on data storage methods to capture the depth of said system.

A byproduct of increased storage capacity is a trade-off between data quality and data quantity. Analysis with more "noisy" data becomes a major big data challenge. Where once we were concerned about being able to collect enough information to detect a trend, we now are concerned about collecting so much information we might miss the trend in the noise (Saha and Srivastava, 2014; Liu et al., 2016). This is a defining feature of the new age of "big data." Although some agricultural practices still involve relatively little data, others are beginning to collect great amounts of data on a daily basis (Bronson and Knezevic, 2016; Carolan, 2017; Kamilaris et al., 2017). Consider a large-scale dairy farm, where 2,000 cows being milked 3-times per day have GPS-enabled ear tags and an in-line milking system. GPS-enabled ear tags can currently collect data up to once every second (Wolfger et al., 2017), while accelerometers in dairy cows can

measure up to 10 times per second (Alsaaod et al., 2015). The in-line milking system analyses all milk for components, volume, and somatic cells. The complexity of these data become great when considering each cow individually. However, with the increasing complexity of collected data comes the opportunity to benefit by deriving insights from this data through data analytics (**DA**).

Creating applicable DA tools and research is important because of the number of decisions farmers are making on a daily basis. Research in agricultural DA is focused on the decisions farmers make that are most economically impactful. Popular areas for DA are disease prediction, nutrient digestion, and genetic potential. Being able to predict and prevent costly diseases in plants or animals is vital for the farmer to profit and sustain their operation (Pennypacker, 1980; Berger and Lehmann, 1984; Garrett et al., 2004; Rutten et al., 2013). Models like the Molly Cow model (Baldwin et al., 1987) utilize our understanding of ruminant nutrition to predict digestion of feedstuffs and help to inform feeding decisions. A natural place for DA is genetics, where a full genome contains billions of base pairs (Koonin, 1998) interacting in near-infinite ways to influence an animal's biology. Identifying important genes and determining the genetic merit of plants and animals are fields where analytics helps humans discern patterns on a micro-scale. All of these examples illustrate how DA can be applied in agriculture to help make decisions.

As data collection in agriculture continues to grow, there is an opportunity to bolster our understanding of systems like digestion, genetics, and disease, but also explore new fields like personalized treatment, autonomous decision-making systems on-farm, or higher resolution weather and plant health models. Agricultural practices are steadily increasing the amount of data collected, and in that data are patterns and signals that will help shape the way producers make decisions. By applying DA techniques to agriculture, we aim to improve the breadth and depth of decision-making through example. This work includes examples of modelling systems for prediction and assessment of management practices, model types along with their strengths and shortcomings, and model evaluation in practice and the importance of appropriate context.

Chapter 2: Literature Review

Data Analytics

Steps of Data Analytics

To draw from our original description of data, DA is the tool employed to transform data into information and insights, and to separate the signal from the noise (Shamoo and Resnik, 2009). The process of DA can be generally broken down into four steps: 1) processing raw data; 2) cleaning data; 3) exploratory DA; and 4) modeling and algorithms (O'Neil and Schutt, 2013). This process can be used in the vast majority of DA cases.

First, to process data means to take raw data and organize it into a structure that is better suited for analysis, such as rows and columns (O'Neil and Schutt, 2013). Doing this allows statistical software to handle the data more easily and allows for faster outlier detection as well because data is ordered and unusual values will stand out. Though often taken for granted, the way we organize our data in the initial processing step can greatly influence our subsequent analysis. This organization of raw data is called the "data structure" or "data model." Data models abstractly relate the elements of the data to each other and inherently require knowledge of the system (McCaleb, 1999). As an example, individual dairy cow data might be structured by milk yield, milk components, and weight separately, but also be

separated in a hierarchy by individual cow ID. By relating observations, this data structure creates more flexibility for analysis and allows for easier analysis in the later steps of traditional DA because data can be reshaped in many ways.

After data are structured, collection errors need to be addressed through data cleaning in order to extract the most useful information. Data errors, missing data, and outlier data can occur as a result of data collection in the real world, and the presence of these errors is usually unpredictable. There is no specific way to remove "dirty" data, but some helpful ways to clean data are comparing to previously published data (Koomey, 2006); statistical analysis of the characteristics of data for changes in mean, variance, or another statistic; or strict validation: the removal of values that do not match a set list of values (such as English words or zip codes). A trade-off in data cleaning is making sure not to remove data that could be useful in later analysis, while still allowing the greatest proportion of "true" data to make it through to the analysis steps (Lloyd, 1993). For example, an extremely low milk yield measurement may be an outlier, given a cow's milking history, or it could be an indication of disease. Human intervention in cleansing slows the process, but also, the ingestion of new data requires re-running of all cleansing procedures because all data must be compared against the entire dataset. There are algorithms useful for outlier removal, however,

their appropriateness is best determined on a case-by-case basis. In fact, like outlier detection, many aspects of the DA process require a scenario-specific procedure.

Exploratory data analysis (**EDA**) is step 3 in the traditional DA process. John W. Tukey pioneered EDA in the 60s and 70s, writing a book of the same name in 1977. Formally, EDA is the process of analyzing data sets to summarize their main characteristics (Tukey, 1977). Tukey encouraged other statisticians to use EDA to explore data to formulate hypotheses then collect more data. This may seem backward, even now. Traditionally, data is collected after a hypothesis is formed, and the data is used as a means to reject or fail to reject the null hypothesis. Tukey argued that oftentimes, coming up with an appropriate research question was the most difficult part of the scientific process and that collecting more data to test the hypothesis was just "confirmatory data analysis" (Tukey, 1980). Indeed, the formulation of good research questions is essential to good experiments, and EDA is a means to this end. Albert Einstein is attributed with versions of a quote, most likely actually coming from an anonymous Yale professor, saying, "If I had only one hour to solve a problem, I would spend up to two-thirds of that hour in attempting to define what the problem is" (Markle, 1966). Tukey's ideas help explore and clean data, as we will see using a real-world example.

Consider an example dataset of daily milking data from a large dairy farm (2000 milking head; milking 3x per day). The dataset is from the previous year (i.e., 2,190,000 individual milkings recorded). Along with the milk yield data, there are also components (fat, protein, lactose, and somatic cell count) and individual cow data on age, DIM, and lactation number. When doing EDA the first question to ask is usually: "what research question are we trying to answer?" A good second question comes from Bourke (2019): "what kind of data do you have and how do you treat different types?" If the research problem is predicting milk yield, you should first examine the data to familiarize yourself with it. Data may have missing values, erroneous values, or other oddities that need to be noted before continuing. Notice that even if we knew exactly what models we wanted to employ, issues with data quality can hinder analysis. Knowing what our data look like is essential for having confidence in any later DA.

Especially important to John Tukey's strategy was the visualization of data for understanding. In our above example, observing milk yield curves for each cow plotted on a graph will give a better idea of the variation in data and the impact of possible missing data. Another important tool in EDA can be feature engineering. In our example, creating 5-day moving averages of milk yield may help smooth yield curves and decrease impact of any single erroneous value. While there are useful tools, there are no

set steps for EDA. Consider that Tukey himself said that EDA is not a set of tools, rather an attitude to have while doing DA (Tukey, 1977).

The final step of the DA process is creating models and algorithms to represent the data. Models are one or more algorithms that represent the inherent structure of the data. Algorithms can informally be defined as a set of rules that precisely defines a sequence of operations (Stone, 1971). Stone (1971) goes even further to say that an algorithm can be any set of rules such that a "robot is guaranteed to be able to obey [them]." More formally, consider that a mathematical formula like $y = \beta x + \varepsilon$ is a set of instructions, and therefore an algorithm that describes a linear relationship.

By representing the data, we can discover patterns and relationships in the data that we may not have understood otherwise. In order for models to inform our analysis, the model choice must be appropriate for the given dataset. This is necessarily ambiguous because all datasets will require different models and algorithms for summarization and characterization. For example, a dataset of monthly fluid milk prices and the dates of those recorded prices may require a time-series plot showing the change in milk price over time. An algorithm that took the 6-month moving-average of milk price and a subsequent plot of this moving-

average along with the monthly milk prices would help to visualize variation between the statistic and the data (Figure 2-1).

Although research using DA and mathematical modeling is a rapidly developing field, advances have been slower to take hold in agricultural research. Improved evaluation leads to easier application of DA across specialties within agriculture. Many research efforts employ an analytical framework as described here, but papers are narrowly applicable and researchers seeking to employ the same framework outside of the given field may find it hard to translate the work. Our research goal was to produce an evaluation of DA techniques and provide example situations for their use that would assist future researchers in choosing appropriate analytical frameworks.

In agriculture, as in any other scientific field and the world, to make proper assumptions of our surroundings and then make inferences as to how the world works, we need data. Data informs decision-making, informs assumptions, and, when we know where to look, it can disprove previously held assumptions about our world. But data must be collected and structured in order to be used. The sources of data in agriculture are extensive, but in recent years sensors have become essential to informing the analytics revolution.

Data Collection

Data are any collection of observations, often numeric, and can be collected from an infinite number of sources on a farm, at a market, in a field, or in an individual animal. Given this diversity, it is useful to catalog the ways in which we gather data in practice. Data is usually gathered from sources that have the most potential to return value to the collector, relative to the cost of data collection. For example, weather data can be costly to capture (although it is widely available as a free service of the U.S. government), but is a major driver of decisions on the farm, day-to-day and long-term (Potgieter et al., 2005; Kantanantha et al., 2010). Although the producer may not be the one paying the majority of the costs associated with predicting the weather, weather predictions help many other sections of the population. In order to make money, farmers rely on crop prices and commodity prices to help decide which crops to plant (Eales et al., 1990; Hoffman, 2019). The pattern of paying to collect data until the returns offset the costs is common across agriculture.

Sensors in animal agriculture are becoming far more widespread, as newer technologies allow for smaller, longer-lasting, more versatile tools to be placed on or near animals to collect data. GPS sensors can track individual cow movements within a barn with a deviation of 1.22 meters using just an ear tag (Wolfger et al., 2017). Knowing a cow's location reduces labor

costs and improves productivity by limiting time spent in the barn searching for cows to deliver treatments or checking a cows' health. Other wearable devices include halters, collars, and leg bands that measure things like chewing, grazing behavior, or rumination (Borchers et al., 2016; Zehner et al., 2017; Reiter et al., 2018); ankle monitors that measure steps taken and laying time (Maatje et al., 1997; Alsaaod et al., 2012); and tail sensors that detect heat (Miura et al., 2017). Data collection devices can also be inside the animal, like a ruminal sensor, which can measure things like rumination, rumen pH, and even rumen contents in some cases (Enemark et al., 2003; Ipema et al., 2008; Mottram et al., 2008). Although animals provide a necessary challenge to track and record from a data collection standpoint, crop agriculture also utilizes sensors to improve yields and management practices.

Understanding the fields that crops grow in is obviously of utmost importance to the crop producer. Using soil sensors, variables like pH, nutrient levels, moisture, or airflow can be recorded at any desired granularity in the field (Hamilton et al., 2007). This understanding of soil dynamics helps the producer target specific areas for special treatment. Spectral reflectance is commonly measured to assess crop growth in fields, specifically nitrogen deficiency (Solie et al., 1996; Stone et al., 1996). These recordings can be used to better predict growth and adjust crop management in order to maximize yields. Bushong et al. (2016)

showed that combining spectral readings and soil moisture profiles to improve prediction capabilities of their models. Moisture profiles can even affect the planting of crops, because it has been shown that optimal soil moisture can allow farmers to plant 10 to 20% more grain (Rehm and Schmitt, 1989).

Precision GPS allows for better topography of any field, allowing for better planting and interpretation of yield and other sensor data (Price and Gaultney, 1993). With the advent and popularization of drones, more individualized plant care is possible (Burema and Filin, 2016). Drones, or unmanned aerial vehicles (UAV), have been shown to assist with estimation of crop yield or height, mapping of plant locations, and fertilizing (Luna and Lobo, 2016; Feng et al., 2018; Li et al., 2019). The goals of using UAV technology are to optimize field management and evaluate crop performance (Feng et al., 2018). Because traditional methods to measure yield, height, or the health of plants are costly, typically involving tractors or other large vehicles, advantages of UAV are the low cost, as well as the improved accuracy. A downside of many precision technologies that involve complex data collection is the increased need for on-farm processing power or improved networks to transmit data from the field to a server.

Crop systems are becoming more efficient with the use of sensors and livestock systems are adopting some similar technologies. For instance, a “Virtual Dairy Farm Brain” is being

developed by researchers at the University of Wisconsin that aims to take advantage of technological innovations to improve whole farm decision-making (Lianga et al., 2018). Examples of decisions being made in this framework are optimal cow replacement decisions, reproductive decisions, insemination outcome prediction, feeding efficiency, and natural language processing (Cabrera, 2010, 2012; Kalantari et al., 2016; John et al., 2017). Although the goals of this project are far-reaching, a downside of large, multi-disciplinary efforts is the speed of progress. New technologies will be developed with time and applicable tools produced may become obsolete quickly. Another problem with large-scale efforts is the inability to produce an applicable tool for farmers that is easily understood, but still efficient and effective (Hunt et al., 2011). By integrating many different data sources, this research group hopes to generate applicable predictions that will change the way dairy decision-making is done.

A sensor in a field, on a cow, or in the ground, is rarely equipped with computation power to process raw data. Therefore, the data must be passed from the sensors to a computer for cleaning, exploration, and modeling, the final three steps of DA as mentioned above. The internet of things (IoT) refers to the interconnected nature of electronic devices in the current internet-age (Evans, 2011). Many barns are outfitted with wireless internet to connect animal sensors to a central computer or cloud

computing (Kwong et al., 2012). This allows data collected out in the fields to be collated in one place for analysis. Although the exact definitions and devices that are encompassed in the IoT are ever-changing, the IoT is an important addition to farming and agriculture in general as it provides a digital infrastructure for data to be collected and utilized.

The rise of precision technologies to aid in the management of animals and crops has also exposed general problems, some referenced above. Precision technologies rely on improved network capabilities for sharing and transmitting data around the farm. Another issue associated with more sensors for data collection is the need for greater power to run these technologies (Keshtgary and Deljoo, 2012). The architecture needed to implement precision technologies is necessarily greater than that needed for human-measurement. Besides the physical limitations, precision data technologies frequently rely on models to interpret data collected. That is, the data collected is a proxy for the actual variable of interest. For instance, measurements of spectral reflectance act as a proxy for nitrogen status in crops. Without additional modeling, spectral reflectance is not useful to the farmer. The use of models to translate collected data into insights is a great feat, but also drives reliance on those same models. Despite these concerns, precision agriculture and modeling will be key factors in improving decision-making on-farm.

Modeling in Agricultural Systems

Linear Regression Models

Perhaps the most common statistical models used in agriculture are linear regressions, regression being the estimation of the relationship between variables. By knowing the relationship between the variables, we can attempt to predict a response variable given a set of independent variables. The use of regression to predict a specific variable's value, *Y*, at any fixed values for all independent variables, *X*, is called conditional expectation, defined by:

$$E[X \mid Y](y) = E[X \mid Y = y]$$

which suggests that *Y* is also a fixed value in the solution. The use of a conditional expectation framework allows us to solve for the expected value of any variable (the variable we are interested in) by fixing the values of all other variables. Using linear regression we have

$$Y_i = a + \beta x_i + \varepsilon_i$$

Where the expectation of $\varepsilon_i = 0$, for normally distributed error (which is to be assumed), so:

$$EY_i = a + \beta x_i$$
$$E(Y_i \mid x_i) = a + \beta x_i$$

The Y_i are conditionally dependent on each x_i as mentioned above, giving us the conditional expectation form of the linear regression equation. Being a linear regression means that the conditional expectation of *Y* (the result), given the parameters, must be a linear function of the input parameters only (Casella and Berger, 2002).

Linear regressions rely only on the assumption that the dependent variables be linearly related to the independent variable. This assumption may not always be justified, but is extremely convenient when fitting models or interpreting results. When fitting a line to data, a least squares estimate of the line that minimizes the residual sum of squares is always possible. This is a powerful tool, meaning that a solution does not need to be searched for or optimized for, it can be mathematically solved for using, assuming any line through the data $y = c + dx$:

$$\min \sum_{i=1}^{N} (y_i - (c + dx_i))^2 ,$$

and minimizing c and d. Additionally, when interpreting the coefficients of a linear regression, the value is the expected change in the independent variable when the given dependent variable is changed and all other variables are held fixed. In general, the coefficients relate to the change per one-unit in change of each independent variable. The linear framework also allows the user to theoretically change multiple input values and

predict the change in the response variable. Sometimes is it unrealistic to only change one variable without changing others slightly, because of intrinsic correlation. Correlation tables or using a single variable linear model, called a simple linear regression would work to account for this variation. The first linear models were used as early as 1805 (Yan and Su, 2009), and their advantages of having a specific solution and intuitive interpretation make it clear to see why linear models are still used so often today in agricultural modelling.

As previously mentioned, linear models can have downsides when their assumptions are not met. If a given descriptive variable does not relate to the response variable in a linear manner, a linear model may be misleading. In 1973, Francis Anscombe created four relatively notorious datasets of 11 (x, y) points each, now called "Anscombe's quartet" (Figure 2-2). Each dataset possessed the same x and y means, sample variances, correlation, linear regression line, and coefficient of determination (Anscombe, 1973). This data outlines the weakness of fitting linear regression models on data that does not have a linear relationship with the response variable.

Multicollinearity can also skew results in a linear regression. Multicollinearity occurs when there is a strong correlation between the dependent variables. A consequence of

multicollinearity is that it can be hard to estimate the effect of any single independent variable. This is based on the assumption in regression that controlling all other independent variables and changing the variable of interest will reflect the impact that variable of interest alone. However, when independent variables are correlated, this assumption breaks down. It is also important to note that, as long as the multicollinearity in the data holds in new data, predictions using a model with multicollinearity are not hindered (Gujarati and Porter, 2003).

Homoscedasticity, or constant error along the regression line, is also assumed in regression (Statistics Solutions, 2013). If the dependent variable's variance depends on the value of the independent variable, the data is heteroscedastic. Predictions from regression models trained on heteroscedastic data will be more uncertain than in certain areas than others. The Breusch-Pagan test is used to measure heteroscedasticity. This test uses the least squares of the fitted model's residuals. If the mean of the least squares residuals is far from 0, than the data is likely heteroscedastic (Breusch and Pagan, 1979). Although linear regression models offer an easy and succinct way to analyze data, the pitfalls and assumptions can hinder good analysis.

Growing in popularity is another type of linear model: the mixed effects linear model, or just "mixed model", containing fixed effects and random effects. These models are useful in scientific

research because the user knows, to a degree, which factors were random and which were manipulated. A more comprehensive definition of fixed and random effects comes from Green and Tukey (1960): "When a sample exhausts the population, the corresponding variable is fixed; when the sample is a small (i.e., negligible) part of the population the corresponding variable is random." This statement is true of most agricultural experiments in areas like supplement effect, effect of feed changes, or management strategy effects, where the variable of interest relates to samples across possible values, while effects like possible location or individual animal do not. Mixed models use a best linear unbiased predictor to estimate random effects and their variance (Robinson, 1991). To estimate the effects of independent variables the best linear unbiased estimator is used. The mixed model equation is then

$$y = X\beta + Z\mu + \varepsilon$$

Where y are the observations; β are the fixed effects; μ are the random effects, with mean = 0; ε are the random errors; and X and Z are the observations. A joint estimate for β and μ must be made to solve this equation, which is traditionally done using expectation maximization of the joint likelihood (Lindstrom and Bates, 1988).

Mixed models improve on the traditional simple linear regression model for agricultural research in their ability to

handle repeated measures of the same statistical unit, which is common in research trials. A driving piece in the Journal of Animal Science (St-Pierre, 2001) illustrating the shortcomings of linear regression in repeated measures experiments helped shift practitioner's opinions towards a more robust modelling framework. This work by St. Pierre also illustrated the importance of using mixed models to control for experiment as a random effect in meta-analyses, a common practice in the agricultural field. Additionally, mixed models allow missing data, as long as that data is missing-at-random, which traditional linear regression would remove (Seltman, 2012).

A downside of using a mixed-effects model comes when the number of samples within a random variable is small. It is difficult to estimate differences in slopes of regression lines across random groups, and the subsequent estimation of overall model variance will be great (Mok, 1995; Clarke and Wheaton, 2007; Bell et al., 2008). Another issue with using mixed-effects models is that it is necessary to make informed decisions as to the nature of the variables measured (Seltman, 2012). Determining which variables are random effects and which are fixed will affect the resulting model. Because applied research will face challenges associated with a small sample size, it is important to consider the implications of model choice on the accuracy and precision of the predictions. In practice, mixed- and fixed- effects models are

easily run as part of a statistical computing package, like R (R Team, 2018). Because of the ease of use, linear regression is extremely common in agricultural research. However, there are other methods of generating models seen in the literature, principally among them Bayesian methods.

The two most common schools of statistical inference are Bayesian and frequentist. The difference between the two approaches revolves around their interpretation of probability (Orloff and Bloom, 2014). While a frequentist sees probability as the long-term frequency of an event, a Bayesian refers to probability as the degree of belief generated from past information and the data. A frequentist does not place a probability on a hypothesis, only the probability of the data given a hypothesis. The result of a frequentist analysis will be an estimation of a statistic's value, and a confidence interval around that value. The confidence interval does not give probabilities to the values around the estimated value, except to say the true value will be within the range some proportion of the time when the scenario is repeated infinite times. A 95% confidence interval does not mean that a single sampled statistic will be within the interval 95% of the time. A Bayesian sees hypotheses, statistics, and data as having probabilities. A Bayesian would give a probability curve relating all values under a given hypothesis. Here, the estimated value is not set, but a range of values with varying chances of

being the true value. Although these two views are thought to be opposing, the majority of frequentist concepts have an equivalent in the Bayesian framework and vice versa. For instance, optimal estimates of a value sampled from a normal population would be the same under both frameworks, despite being derived from different starting places.

Bayesian Models

Bayesian models rely on the Bayesian interpretation of probability, that probability represents a degree of belief in an event (Bayes, 1763). Thomas Bayes, the namesake of Bayesian practices, formulated Bayes' Theorem which is generally stated, for any two events *A* and *B*:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

P(A) is called the prior, because it is a distribution reflective of prior knowledge about A, the proposition. P(B|A) is the likelihood, or the likelihood of observing *B* given *A*. P(A|B) is the posterior probability or the likelihood of observing *A* given *B*. The posterior is what we aim to solve, the likelihood of our proposition being true, given the evidence we observe. P(B) is the probability of the evidence observed in any scenario. The probability of the evidence can be rewritten as:

$$P(B) = \int_{i=1}^{N} P(B \mid A_i) P(A_i) dA$$

or $P(B) = \sum_{i=1}^{N} P(B \mid A_i) P(A_i)$ in the discrete case

Oftentimes, calculating P(B) directly can be challenging and unnecessary because the evidence does not change in the analysis, meaning P(B) is always the same for any given evidence. The posterior likelihood can be written like so:

$$P(A \mid B) \propto P(B \mid A) P(A) \text{ or}$$

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

The important thing to realize is that the maximum likelihood of this posterior distribution will be the same, regardless of whether or not we include P(B) (Gelman et al., 2013). Using the proportionality instead of the full equation simplifies the problem greatly, avoiding the need to integrate over all possible propositions. To predict using Bayes' theorem analytically, you must integrate the posterior over all possible evidence (or sum in the discrete case).

In practice, unless the product of the prior and likelihood distributions are in the same family of distributions and produce a closed-form expression (called "conjugate"), it will be hard to evaluate the posterior. For instance, if the likelihood is normally distributed with a known mean and variance, a normally distributed prior will produce a closed-form, normal, posterior distribution.

In cases where the prior and likelihood are not conjugate, numerical integration can be used to find approximate solutions to describe the resulting distribution.

Numerical integration is useful because many priors, like uniform or normal, are not conjugate in many cases. That is, the use of only conjugate priors can be constraining to analysis. Numerical integration is the process for finding numerical approximations to a differential equation, like the differential equation needed to predict from a posterior distribution:

$$P(x|X,a) = \int P(x|\theta)P(\theta)\,d\theta$$

where x = new data to be predicted, X = sample, a = the hyperparameters of the parameters of the new data, and θ= parameters of the new data's distribution θ= P(θ| a). Markov chain Monte Carlo (MCMC) methods refer to a class of algorithms used to do numeric integration by sampling from a distribution with a probability density proportional to the distribution of interest. By repeatedly sampling and using points in proportion to the "true" distribution, the new distribution will converge towards the true distribution. Because this proportional sampling will produce points, we can calculate an expected value or variance using this data. The key of MCMC is having an algorithm that can "walk" in the potential parameter space, moving into areas of high

probability according to the integral. The process of searching the space is called a random walk.

An example of a random walking MCMC algorithm is the Metropolis-Hastings algorithm. Here, in order to move in the space and sample, a move is proposed and either rejected or accepted based on the density of the integral in the new space compared to the old space. Importantly, the move is not always rejected in the case of moving to a space with lower density, rather, the move is rejected proportionally to the density change between the old and new point, called the acceptance ratio. Written formally, the acceptance ratio is:

$$\text{Acceptance ratio} = f(x')/f(x)$$

where x = the previous point and x′ = the proposed point and *f(•)* is the function of interest. If the acceptance ratio > 1, the algorithm will always accept the move. Repeating this random walk many times, usually from multiple starting positions, will produce a distribution that follows the desired distribution. Figure 2-3 shows an example of a random walk over many runs and the produced distribution of a parameter (Tran, 2014). The popularity of Bayesian techniques in research stems from the ability to extend the techniques to hypothesis testing.

In research, Bayesian inference is frequently used in hypothesis testing, with the proposition being a hypothesis and

the evidence being collected or observed data: $P(H_0|X) \propto P(X|H_0) \times P(H_0)$. In recent agricultural research, some examples of Bayesian inference have been to estimate the energy requirements of cattle (Moraes et al., 2014), water requirements of fields (Padalalu et al., 2017), and yield prediction of crops (Savla et al., 2015). Another Bayesian use example is for genome-wide association studies (GWAS). Freua et al. (2017) used a database of phenotypic and genotypic information for 893 cattle. Using three established differential equations relating to DNA accretion and energy of maintenance requirement, the study estimated the distributions of two parameters using the information on the single nucleotide polymorphisms (SNP) present in the data. To generate probability distributions for each parameter, Gibbs sampling, a sub-type of Metropolis-Hasting sampling was used for MCMC. Predictions of DNA accretion and energy requirements were estimated with SNPs shown to have 95% confidence intervals of their parameter estimates greater than 0 (Freua et al., 2017). Phenotypic variation was better explained using a Bayesian approach compared to previous models (Freua et al., 2017). This example in GWAS illustrates the Bayesian framework's ability to estimate complex parameter distributions and improve predictions.

Hypothesis Testing

In research, oftentimes we are testing hypotheses using data. In order to test hypotheses, we start with a null hypothesis, which is a model of our belief, given that nothing is out of the ordinary (Everitt, 2006). The null hypothesis is assumed to be true until disproved. This is the challenge of all scientists: to prove that their theories have validity against previously held truths. The null hypothesis is part of the field of inferential statistics, used to derive properties of underlying distributions and make inferences about a population, given a sample from the population. When a researcher rejects the null hypothesis, sometimes based on a *p*-value, they are rejecting the belief that conditions are as previously believed, in favor of a new belief.

Assumptions are key in inference and therefore important in testing hypotheses in research. The *p*-value, which we use to decide whether to reject the null hypothesis, reflects the probability of seeing data at least as extreme, or more extreme, given that the data was generated from a model reflecting our null assumptions (Wasserstein and Lazar, 2016). Notice that the *p*-value's definition does not include any ties to the alternate hypothesis. Just because data is unlikely to be drawn under the null hypothesis' distribution does not mean the data does come from the proposed alternative distribution. Model assumptions can be a problem when testing hypotheses because poor model assumptions can

guarantee that collected data will disagree with the null hypothesis. *P*-values have been criticized for sometimes not producing coherent results in a two-sided test. The two-sided hypothesis test compares two datasets and determines the probability of the two samples being different (either one dataset being greater or less than the other). One would expect as the average difference between the two datasets diverged, that the *p*-value of this test would decrease, but this is not always the case. The property of test statistics to never increase as the average difference between samples increases is called monotonicity. *P*-values in a two-sided hypothesis test are not monotonic. Statisticians have called for varying solutions to the issue that arise from *p*-value use in research.

Research has shown that a disproportionate number of published studies have *p*-values directly below 0.05, most likely stemming from the acceptance of 0.05 as a threshold for statistical significance (Hartgerink et al., 2016). The most widely accepted origins of the 0.05 threshold come from a 1925 book, Statistical Methods for Research Workers by R.A. Fisher, where he claimed about 0.05: "The value for which P=0.05 … it is convenient to take this point as a limit in judging whether a deviation ought to be considered significant or not" (Fisher, 1925). However, less than 100 pages later in the same book, Fisher calls a *p*-value between 0.02 and 0.05 to be "significant, though barely so..." (Fisher,

1925). The extreme popularity of a 0.05 threshold for significance in research is likely as a result of the need for enough research to be published (Bross, 1971). If a lower threshold, say 0.01, was common, much less research would be publishable, and Bross (1971) supposed that the "evolutionary process would have eliminated it." Although alternatives have been proposed and many other hypothesis testing frameworks exist, *p*-values and a 0.05 is by far the norm.

The hypothesis testing method used in Bayesian inference, Bayes factors, have a different interpretation than classical *p*-values (Morey et al., 2016). The Bayes factor is the likelihood ratio of two models' marginal likelihoods. While classical *p*-values measure the likelihood of the data or more extreme data under the null hypothesis, Bayes factors compare two models (like the null and alternative hypothesis) and represent the support of one model over the other. That is, a greater Bayes factor only suggests that the alternative model is more likely than the null model, given the data and priors. When both models are equally likely *a priori*, the Bayes factor is equal to the ratio of the posterior probabilities of the two models. A Bayesian framework can be easily interpreted because it compares the alternative hypothesis to the null directly. This interpretation differs from traditional *p*-values which compare the null hypothesis to all other possible hypotheses. The Bayesian framework is gaining popularity

in research because its interpretation is more in line with the research process.

Creating mathematical models in research requires hypothesis testing to disprove our previous beliefs. Hypothesis testing is invaluable when used correctly, but there are dangers of hypothesis testing, sometimes stemming from a misunderstanding of the tool or what a *p*-value means to research results (Chow, 1996; Morrison and Henkel, 2006). For instance, the classical *p*-value from a *t*-test does not indicate anything about the validity of the alternative hypothesis, only the null hypothesis' fit under the data. Researchers commonly associate lower *p*-values with more significant results, which is not the case (Ziliak and McCloskey, 2008). If a poor null hypothesis is chosen, like a model that is infeasible, the resulting *p*-value will almost certainly be significant. Despite these concerns, good model selection involves using statistical tests like a *t*-test or Bayesian framework to choose the best models to represent complex systems.

Model Types

In agriculture, modeling natural processes usually involves complex models with multiple sub-models to produce predictions. There are many different types of models that are used in agriculture, differentiated by the model's structure. Some common

classifications of models are static vs. dynamic; deterministic vs. stochastic; and empirical vs. mechanistic. Static and dynamic models refer to the handling of time in model calculations; deterministic and stochastic models deal with randomness in the system; and mechanistic and empirical models differ regarding how the model is created, based on theory or observed data.

Many modeled systems vary over time and our research interests require knowledge of these changes. A dynamic system is defined as a system in which a function describes the time dependence of a point in phase space (Katok and Hasselblatt, 1997). A rule of dynamical modeling is that given all points up to time $t-1$, we can predict the state of a point at time t, called the time evolution law. While systems can vary over time, Henri Poincaré provided the Poincaré recurrence theorem showing that, given enough time, all systems will reach near-steady state (Poincaré, 1890). This additional information is useful because, if we are interested in only the steady state behavior of a system, we can model the behavior without consideration of time. Models that are time-indifferent are called static models. Using static models, we can represent systems that are in steady state or systems at a specific point in time. The use of static versus dynamic models in agriculture depends mostly on the system of interest and whether time-dependence is assumed or necessary.

When modeling a system, the randomness therein is also important when deciding which type of model to use. Deterministic models assume all inputs are fixed and known, meaning the output will always be the same (Meiss, 2007). Deterministic models are useful when most variance in outcomes is described by the model's variables. Stochastic models use probability distributions for variables, meaning that outputs will vary. Because of the probabilistic framework, stochastic models are frequently called statistical models. A common example of a statistical model is one of height in children based on age. A deterministic interpretation of this system for the *i*th child's height would be:

$$height_i = \beta_0 + \beta_1 age_i$$

where β_0 = the intercept; β_1 = the slope; and age_i = the age of the *i*th child. The related stochastic model would include an error-term for the prediction of each child:

$$height_i = \beta_i + \beta_0 age_i + \varepsilon_i$$

where ε_i = the error of the prediction for the *i*th child. The error term would be an assumed probability distribution. Each prediction from the statistical model involves randomly realizing one value from the error distribution, providing a different estimate for height each time. Because stochastic models require draws from probabilistic distributions for each output, the models can be very slow to near infinitely slow, for all intents and purposes, if the distributions are complex. When solutions to stochastic

models are practically feasible, solutions can be more robust than deterministic models because error terms can simulate variance inherent in the system not described by the input variables.

Mechanistic and empirical models are both common in agricultural literature. Mechanistic models directly relate to an assumed underlying mechanism in the modeled system, while empirical models are created based on observed data and make no assumptions about system mechanisms (Thakur, 1991). A model cannot be completely mechanistic because lower-level mechanisms cannot be completely described (Gill et al., 1989). Consider a model of digestion that accounts for the mechanisms of the organs and tissues, which accounts for the mechanisms in the individual cells. In this case, there is no finite end to the number of systems that would need to be accounted for to keep the model completely mechanistic (France et al., 1984). Another limitation of mechanistic models is their ability to describe nonlinear and dynamic systems, which describe most biological processes (Sauvant, 1991).

When mechanistic models cannot describe complex systems, empirical models are a good alternative, able to closely mimic the system's outputs without knowing the underlying actions at play. To evade needing to understand every part of a given process, empirical models can produce a confidence measure in predictions or error estimates. This uncertainty allows a single model to

reflect complex interactions or inherent variation in a process. Empirical models tend to have a more narrow focus, must rely on the accuracy of previous observations, and limit the array of possible statistical methods (Hristov et al., 2018). This property of empirical models makes them better for prediction (Tedeschi et al., 2005). In systems like digestion, there are many pathways that must be modeled, with varying levels of understanding of the mechanisms present. Because of the breadth and depth of modeling systems like digestion, many agricultural models incorporate both mechanistic and empirical models. There are many examples of models used in agriculture that illustrate the different potential model types.

Applied Animal Models

In the field of animal science, there are examples of mathematical models for decision-making, at the individual-, farm-, or national-level. These models are especially prescient, as demand for food and natural resources in 2050 is expected to increase to double 2009-levels (FAO, 2009). All animal models allow the integration of knowledge about feed, digestion, metabolism, energy movement, and other processes (Tedeschi et al., 2005). The integration of knowledge allows us to make predictions about areas like nutrient requirements or management practices to improve the

overall decision-making process. Although most animal models have the same goal of providing insight, varying of models have different advantages.

The Molly cow model represents nutrient digestion and metabolism of the rumen (Baldwin et al., 1987). After the first iteration of Molly in 1987, many changes and updates have been proposed and integrated into the model like reparameterization of the digestion models, prediction of methane emissions, and volatile fatty acid production (Offner and Sauvant, 2004; Gregorini et al., 2013; Hanigan et al., 2013; Ghimire et al., 2014). Molly is dynamic, deterministic, and primarily mechanistic. The Molly cow model was designed primarily for lactating cows and has been used on-farm to make decisions regarding greenhouse gas emissions (GHGe) (Beukes et al., 2011), nitrogen excretion (Johnson and Baldwin, 2008), and ration formulation (Kohn et al., 1998). The ability to constantly improve models via additional empirical data and studies has allowed the Molly model to remain useful 30 years after its initial creation. However, Molly also illustrates a weakness in mechanistic, individual models: that complex models rely on greater equation parameterization to be accurate (Hanigan et al., 2013). The parameterization may not be the same for all scenarios, making the derivation of one specific prediction cumbersome.

The Cornell Net Carbohydrate and Protein System (CNCPS) is another example of a commonly used model in agriculture to help make decisions (Fox et al., 1992; Russell et al., 1992; Sniffen et al., 1992). The CNCPS model has mechanistic and empirical components, used to produce predictions of various dairy and beef cow processes. The CNCPS submodels use knowledge of maintenance energy requirements and growth patterns for predictions in the absence of data on a certain set of cattle (Fox et al., 1992). When producers provide their own data and parameterizations, CNCPS allows integration of this information for improved predictions (Fox et al., 1992, 1995). Because of the flexibility and use of production data to improve the applicability, the CNCPS has been actively used and updated for over 25 years (Lanzas et al., 2007; Van Amburgh et al., 2013, 2015; Higgs et al., 2015). Similarly to the Molly cow model, in order to be used by farmers, the CNCPS needs to be distilled into specific parts like a growth model, energy needs, or feed digestion and inputs must be understood as the system is sensitive to initial conditions (Tedeschi et al., 2005).

Where Molly and CNCPS models predict individual animal performance, other models can be used to simulate the effects of budgetary farm-level decisions, for example. Tall fescue toxicosis is a problem specific to grazing animals, especially beef cattle, in the southern U.S. Endophytes, a fungus found in tall fescue,

can produce ergot alkaloids that are vasoconstricting and cause a variety of ailments like thermoregulation issues and decreased fertility. Non-ergot alkaloid endophyte cultivars are becoming more common, such as MaxQ (Pennington Seeds, Madison, GA), because they still possess endophytes that help maintain the hardy nature of tall fescue while not producing the toxin. With new non-ergot alkaloid endophyte, completely endophyte-free, and traditional cultivars, farmers face a decision regarding which cultivar is economically advantageous. Research on a healthy limit for ergot endophyte concentration brings mixed conclusions. Additionally, what alkaloids to measure and their respective effects have been debated. Our goal is to produce a meta-analysis of endophyte-infected tall fescue to create a rule-of-thumb as to the potential effect of varying concentrations of ergot alkaloids in fields. Our research objectives were two-fold: determine the relationship between varying measurements of ergot alkaloid concentration and cattle ADG; and identify a threshold past which increasing effects of infection are not seen. Both objectives will improve decision-making of farmers when assessing which cultivars to graze their cattle.

Disease modeling is another popular field where many types of models are employed. Hierarchical models of chronic disease incidence (Boshuizen et al., 2017), physical representations of kidney tissue to model kidney disease (Morizane and Bonventre,

2018), and agent-based models of noncommunicable diseases (Nianogo and Arah, 2015) are just a few examples of varying model frameworks used in human disease modeling. Translating this research into agriculture is already being done, with studies like Østergaard and colleagues' study (2005) of mastitis spread and control. In this study, 35 different scenarios of clinical and sub-clinical mastitis infection in a herd were simulated (Østergaard et al., 2005). Various control strategies were implemented and the economic consequences and input variable sensitivity were observed. Predictions from the models were compared to the true consequences of mastitis in a dairy herd (Østergaard et al., 2005). When modeling a system with many potential variables, simulation models are especially helpful to compare treatments and examine the variability of predictions.

Complex model frameworks allow the user to predict responses, but not only in the previously observed space. Put another way, a model can give us an idea of how a process works, allowing us to predict what will happen in scenarios we haven't yet observed. For example, atomic particle behavior can be modeled to predict reactions in collisions or laboratory experiments that have not been performed (Hockney and Eastwood, 1988). Modeling can illuminate the mechanism of a system, or just provide a rule-of-thumb for decisions in the future. However, when data is scarce,

complex models may not provide appropriate predictions of error around predictions may be great.

Model Evaluation

Overfitting is the creation of a model that too closely mirrors the data (Everitt, 2006). The use of excessive parameters to describe data that is not appropriately complete will produce overfit models. An overfit model generally will perform poorly on held-out data or in the prediction of future data. Measures can be taken to prevent choosing overfit models, despite the models performing well on a given dataset. Cross-validation is frequently used in the literature to test model performance. Cross-validation involves holding out part of a dataset to use for testing. If a model does not produce similar performance metrics on held-out data, overfitting may be involved.

When assessing the value of a model, the purpose of the model is important to consider. Models are used to help understand and predict a given system. Models being used to predict can either be interpolating or extrapolating beyond the scope of the data used to train the model. A model that performs well on a subset of data used for training is not necessarily adequate for extrapolation. This inconsistency in a model's effective scope is a challenge when validating a model. Another issue that arises from models

that poorly extrapolate is the assumption of causality. When a model performs well, it is common to assume these variables cause the variation in the dependent variable. However, without knowing the performance of a model on data outside the scope of the training data, causation is not guaranteed. In order to create useful models, the model's purpose must align with its scope.

Model evaluation, often referred to as model validation, offers many ways to test the robustness of a given model to independent data. In research, it is unlikely that a modeler will have access to two complete datasets, one for training, and another completely independent source of data for evaluation. Methods like cross-validation, as mentioned above, help achieve an estimation of performance on independent data without needing two datasets. K-folds cross-validation is very common in agricultural research, where the data is divided into k subsamples and one subsection is held out for evaluation. This process is repeated k times with each subsection being the test set. The advantage of this approach over randomly held-out samples is the guaranteed inclusion of every point exactly one time. When $k = n$, where n is the number of points in the dataset, this is called leave-one-out cross-validation. Leave-one-out validation is more common with smaller datasets, as the training set is $n-1$ points and the resulting test performance is averaged over n trials. There are many methods of evaluation models and limitations to evaluation also need to be considered.

Another popular model evaluation procedure involves partitioning the data randomly into two splits for training and testing (the testing split is usually smaller), then repeatedly testing the performance of the models on the test splits. This process is called repeated random sub-sampling or Monte Carlo cross validation (Xu and Liang, 2001). This process does not guarantee the same performance results each time, due to randomness in the selection of the test datasets but is powerful because it allows the user the freedom to choose the train/test proportion and iterations without being bound by the dataset's size, like in K-folds validation. In any evaluation scheme that includes proportions of the data for training and testing, the user must consider the properties of the performance metrics generated. If F is the true performance of a model, than our cross-validation estimate is F^* and E[F] is the expectation of our model fit. With each iteration of cross-validation, F* will vary, with most of the error being attributable to variance, assuming our proportion of training to test data is reasonable (Christensen, 2015). Consider that the V[F*] is not directly correlated to model fit, rather the evaluation process that generated F*. That is, models with a lesser performance do not guarantee better E[F].

The idea of model validation, or evaluation, itself has been questioned. There are at least two schools of thought on the

definition of statistical model validation, as observed in Lewandowski (1982).

"... (model validation is) substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy [consistent] with the intended application of the model." - (SCS Technical Committee, 1979)

This definition suggests model validation as subjective to the user, only needing to possess "satisfactory" performance.

"... a model is valid if its behaviour corresponds to system behaviour under all conditions of interest. A model is considered invalid if we can devise an experiment in which the model outputs disagree with system measurements within the specified area of interest…" - (Mankin et al., 1977)

Here, the definition is more in-line with general validity; the model must match all conditions and be globally accurate to be valid. Validity is "the quality of being well-grounded, sound, or correct," as defined by the Merriam-Webster Dictionary (Merriam-Webster and Inc. Staff, 2016). The "correct" definition of model validation may change depending on the application on the model (Lewandowski, 1981). Mankin et al. (1977) suggests two alternative

measures of model performance when making predictions: reliability and usefulness. Reliability is how often a model is right. While a valid model must be invariably correct, a reliable model will perform acceptably in prediction without perfect accuracy (Mankin et al., 1977). The difference between terms that describe a similar process illustrate why the term "evaluation" is becoming more common, as it is more ambiguous. Since the 1970s, the objectives of model evaluation have moved towards an understanding of a model's error structure and areas of uncertainty (National Research Council et al., 2012). With the increasing complexity of models in general, because of improved computing power, access to data, among other reasons, the idea that a valid model must be invariably correct can be disregarded. Instead of absolute accuracy, comparisons between models are often used to determine the quality of any individual model's performance (National Research Council et al., 2012). Additionally, determining the causes of uncertainty, like input parameter variation are emphasized for creating better performing models (National Research Council et al., 2012). For practical purposes, the user must be cognizant of the limitations of the model in prediction, understanding a system, or scenario analysis. An evaluation method can only prevent the selection of overfit models, but to avoid creating overfit models is the responsibility of the practitioner considering the dimensionality and complexity of their data.

In agriculture, models that make predictions about the future or about unknown processes are of particular interest. A challenge that comes with building prediction models is comparing the performance of models and determining the applicability of a model. Although cross-validation is a method to assess results, there are many statistical tests available to measure performance or fit. Measures like root-mean-square error (**RMSE**), mean-square error (**MSE**), or median absolute deviation (**MAD**) for continuous predictors and log-loss, specificity (**Sp**), sensitivity (**Se**), or positive predictive value (**PPV**) for discrete cases are common. The RMSE and MSE both measure average deviation between predicted and actual value; MSE squares this deviation before averaging while RMSE does the same, but then takes the square root of this value. The RMSE and MSE both are sensitive to outliers because both increase proportional to the square of the error (Bermejo and Cabestany, 2001). Although MSE estimates the variance of unbiased estimators, RMSE estimates the standard deviation and is reported in the same units as the estimator. The MAD differs from both RMSE and MSE by measuring the median distance between predictions and the actual value, making it robust with respect to outliers (Leys et al., 2013).

For discrete predictors, there are only certain values that can be realized, therefore assessing model fit is different than for models with continuous possible outcomes. Frequently there

will be only two possible outcomes - e.g.,"disease" or "no disease" - in agricultural models, while the predictions take on all values between the two outcomes. Logarithmic loss, or log-loss, is used to assess fit on a continuous scale for discrete predictors by penalizing the model proportional to the log-distance between the prediction and outcome (always 0 or 1). Alternatively, continuous predictions can be converted to discrete predictions using a threshold value. When a prediction is greater than threshold, a positive prediction will be reported and vice versa. When predictions and outcomes are both discrete, a confusion matrix can be used to assess fit by putting predictions in four groups: true- and false-positives (**TP**; **FP**) and true- and false-negatives (**TN**; **TP**). An example confusion matrix is shown in Figure 2-4.

Using a confusion matrix, many helpful measures can be derived. The PPV, which is calculated as Σ(TP) / Σ(all positive predictions), represents the percentage of positive predictions that will result in actual positive cases. This number can be extremely useful to producers, as a model that can predict all positive cases of a disease, but also mis-labels many non-diseased cases as positive is not useful. False-positive predictions can be a waste of time and cause distrust in a model (Bell, 2010). Commonly used in tandem, Se and Sp are common in discrete model evaluation (Hogeveen et al., 2010). Specificity represents *1 - type I error*, or the ability of the model to detect positive cases,

relative to all true positive cases. Sensitivity is the ability to detect true negative cases and is equal to *1 – type II error*. A model's Se and Sp are statistics concerned with the proportion of true cases, while PPV is concerned with the relationship between predicted and true cases. None of the aforementioned statistics can completely describe the applicability of the model, rather they provide pieces of a complete analysis (Dominiak and Kristensen, 2017). Choosing which metrics to use to measure model fit is up to the researcher and the goals of the model.

Data in Decision Making

Creating a model to predict an outcome does not inherently produce better decision making. In research of sensor systems used in production agriculture, there are four levels of information that can be provided by the sensor: (I) raw data from the sensor; (II) interpretations based on sensor data; (III) advice derived from interpretations and outside knowledge or data; and (IV) a decision being made (Rutten et al., 2013). A model prediction trained on data falls into level II of the information used in decision making. For example, a model prediction of average daily gain based on various feed compositions does not explicitly indicate the "best" course of action or carry out a decision. Rutten et al. (2013) showed that all reviewed sensor information

falls into level I or II, failing to generate advice or actual decisions directly (Figure 2-5). From a modeling perspective, the lack of higher level sensor information suggests a lack of applicability of models. That is, models are currently not producing predictions that need further context or more information before an actual decision can be made on-farm.

Decision support systems (**DSS**) are intended to help producers with the great number of decisions they must make (Steeneveld et al., 2010). Most producers have to use intuition, or herdsmanship, to estimate variables relevant to each decision, like economics, possible outcomes, or herd health (Groenendaal et al., 2004). Because profitability is the major concern for more agricultural decisions, models can help provide better estimations of the range of possible outcomes for a decision than the producer could alone (Aramyan et al., 2007). Insights like economic consequences of decisions integrated into existing and future models would create the necessary DSS needed to derive clear advice on-farm.

Models that aim to produce actual decisions, although sparse, are present in the agricultural literature. A DSS describes the situation where collected data produces advice and decisions for the farmer, while a decision support tool (**DST**) can suggest optimal decision paths and act as evidence for the user to make decisions (Rose et al., 2016). Mastitis prediction models (Hogeveen et al., 2010; Mollenhorst et al., 2012), locomotion-based models (Bicalho

et al., 2007; Bruijnis et al., 2010; Pastell et al., 2010), and fertility models (Williams et al., 1981; Galon, 2010) are all examples of DST. The methods used to produce predictions are varied, but show the statistical methods often employed resulting in better decision making.

In the case of locomotion-based models on dairy farms, models have been created to predict the presence of lesions, sores, or other ailments that affect the cow's welfare and production. Pastell et al. (2010) use weight distribution and visual observation to detect lameness. This system identified a relationship between weight distribution and a cow's gait that was statistically significant. However, the reliance on human observation means that this model cannot truly be a DSS, as the model cannot make decisions alone and does not make recommendations on what the farmer should do in cases of lameness or lesions. Another locomotion modeling study used veterinarians to assess lameness, which was then fed to a model to predict lameness (Bicalho et al., 2007). Here, the study found that veterinarians' scores to be a better predictor than an automatic gait scoring software (Bicalho et al., 2007). Again, these studies indicate the inability to rely entirely on a model for lameness predictions and the lack of clear advice that current locomotion models give.

Estrus detection using sensors aims to aid farmers in making decisions about when inseminate their cows. Pedometry is a common

method for detecting estrus and is used around the world (Roelofs et al., 2005). In a review of pedometry for estrus detection, Galon (2010) found that negative pregnancy check rate varied from 9.5 to 53.4%. This variation makes accurate predictions difficult, especially when considering farm-to-farm variability. Other methods of estrus detection can be effective, like tail painting, but require entirely human inputs (Xu et al., 1998). There has also been recent research showing that vaginal temperatures alone may be better than pedometery in predicting estrus (Sakatani et al., 2016). Overall, research in the area of estrus detection shows promise for DA to improve on-farm decision-making while decreasing the human intervention needed.

Decision-making systems in place for clinical mastitis (CM) are traditionally interpretations of raw data into insights, but not actual decisions. For instance, a CM model that takes milk component and production data to derive a prediction of disease likelihood does not indicate whether or not treatment is warranted, making the model a DST. The current model evaluation scheme uses Se and Sp as the gold standards to measure performance and applicability. Although these two measures are important for measuring model performance, they only represent part of the set of classification statistics. Of particular importance to producers is the PPV, or proportion of predicted positive cases that are correct. The goal of our research was to illuminate the

importance of other model performance measures and show how to implement weighting schemes in model-building to more closely mirror producers' costs associated with predicting CM.

Simulation models can be especially useful in on-farm decision making because of the abilities to forecast outputs far into the future, explore many different decision paths, and be tailored directly to a given farm. A simulation is simply a method for implementing a model over time (U.S. Department of Defense, 2018). An example simulation study measured the economic value of implementing information technologies, specifically activity meters, concentrate feeders, and automated parlour systems (van Asseldonk et al., 1999). By using models of animal performance, fertility, and feeding strategy, van Asseldonk et al. (1999) were able to compare a range of inputs by expected monetary return. The specific ranges of each variable were determined using past research and were inputted into the model in all combinations. By accounting for all expected combinations of inputs, simulation models predict variation in a process over time in many scenarios.

One specific animal example of simulation modeling is GHGe. Greenhouse gas emissions are an important topic within the research community and the public. Animal agriculture is often attributed to being a significant contributor of GHGe, and many people are proponents of the removal of animal production systems as a means to decrease global GHGe. Although there is research that focuses

on the impacts of complete removal of animals from a GHGe, land-use, and nutrient requirement standpoint (White and Hall, 2017), less is understood about the implications of removal of select species. Simulation models can be used to estimate GHGe contributions of varying animal populations. White et al. (2010) used a whole-farm model along with nutrient budgeting software to simulate the effects of intensifying the production of beef cows in New Zealand. In another example, O'Brien et al. (2010) used a simulation model to assess the effects of genetic merit on GHGe. Each of these examples explores the results of a specific change in animal production systems aimed at reducing GHGe, but considerations for exactly how we would reach a goal of complete animal production system removal have not been explored in this context. The goal of this research was to elucidate the effects of removal of all dairy cow production systems from the U.S. The effects of removal were measured in terms of GHGe, land-use, and nutrient requirements, and several scenarios for removal of dairy cows were explored. Our compilation of pertinent data and simulations of dairy cow removal scenarios will improve the body of research relating to decision-making for policy changes and clarify dairy cow production's impact on GHGe.

Much of the literature on decision-making in agriculture is specific to the domain. Although there may be literature on a specific DA topic, it may not be applied to agriculture in a way

that makes translation easy. Gaps could be literal gaps in knowledge or simply gaps in translation. Translating analytical work done in other fields is necessary for advancing the field of DA in agriculture and creating better decision-making tools and systems.

References

Alsaaod, M., J.J. Niederhauser, G. Beer, N. Zehner, G. Schuepbach-Regula, and A. Steiner. 2015. Development and validation of a novel pedometer algorithm to quantify extended characteristics of the locomotor behavior of dairy cows. J. Dairy Sci. 98:6236–6242. doi:10.3168/jds.2015-9657.

Alsaaod, M., C. Römer, J. Kleinmanns, K. Hendriksen, S. Rose-Meierhöfer, L. Plümer, and W. Büscher. 2012. Electronic detection of lameness in dairy cows through measuring pedometric activity and lying behavior. Appl. Anim. Behav. Sci. 142:134–141. doi:10.1016/j.applanim.2012.10.001.

Anscombe, F.J. 1973. Graphs in Statistical Analysis. Am. Stat. 27:17–21. doi:10.1080/00031305.1973.10478966.

Aramyan, L.H., A.G.J.M.O. Lansink, and J.A.A.M. Verstegen. 2007. Factors underlying the investment decision in energy-saving systems in Dutch horticulture. Agric. Syst. 94:520–527. doi:10.1016/j.agsy.2007.01.005.

van Asseldonk, M.A.P.M., A.W. Jalvingh, R.B.M. Huirne, and A.A. Dijkhuizen. 1999. Potential economic benefits from changes in management via information technology applications on Dutch dairy farms: a simulation study. Livestock Production Science 60:33–44. doi:10.1016/s0301-6226(99)00039-1.

Baldwin, R.L., J.H. Thornley, and D.E. Beever. 1987. Metabolism of the lactating cow. II. Digestive elements of a mechanistic model. J. Dairy Res. 54:107–131.

Bayes, T. 1763. LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. Philosophical transactions of the Royal Society of London 370–418.

Bell, B.A., J.M. Ferron, and J.D. Kromrey. 2008. Cluster size in multilevel models: The impact of sparse data structures on point and interval estimates in two-level models. JSM proceedings, section on survey research methods 1122–1129.

Bell, L. 2010. Monitor alarm fatigue. Am. J. Crit. Care 19:38. doi:10.4037/ajcc2010641.

Berger, J.O., and E.L. Lehmann. 1984. Theory of Point Estimation. Journal of the American Statistical Association 79:941. doi:10.2307/2288729.

Bermejo, S., and J. Cabestany. 2001. Oriented principal component analysis for large margin classifiers. Neural Netw. 14:1447–1461.

Beukes, P.C., P. Gregorini, and A.J. Romera. 2011. Estimating greenhouse gas emissions from New Zealand dairy systems using a mechanistic whole farm model and inventory methodology. Anim. Feed Sci. Technol. 166-167:708–720. doi:10.1016/j.anifeedsci.2011.04.050.

Bicalho, R.C., S.H. Cheong, G. Cramer, and C.L. Guard. 2007. Association between a visual and an automated locomotion score in lactating Holstein cows. J. Dairy Sci. 90:3294–3300. doi:10.3168/jds.2007-0076.

Boghossian, P.A. 2017. Analyticity. B. Hale, C. Wright, and A. Miller, ed. John Wiley & Sons, Ltd, Chichester, UK.

Borchers, M.R., Y.M. Chang, I.C. Tsai, B.A. Wadsworth, and J.M. Bewley. 2016. A validation of technologies monitoring dairy cow feeding, ruminating, and lying behaviors. J. Dairy Sci. 99:7458–7466. doi:10.3168/jds.2015-10843.

Boshuizen, H.C., M.J.J.C. Poos, M. van den Akker, K. van Boven, J.C. Korevaar, M.W.M. de Waal, M.C.J. Biermans, and N. Hoeymans. 2017. Estimating incidence and prevalence rates of chronic diseases using disease modeling. Popul. Health Metr. 15:13. doi:10.1186/s12963-017-0130-8.

Bourke, D. 2019. A Gentle Introduction to Exploratory Data Analysis. Accessed. https://towardsdatascience.com/a-gentle-introduction-to-exploratory-data-analysis-f11d843b8184.

Breusch, T.S., and A.R. Pagan. 1979. A Simple Test for Heteroscedasticity and Random Coefficient Variation. Econometrica 47:1287–1294. doi:10.2307/1911963.

Bronson, K., and I. Knezevic. 2016. Big Data in food and agriculture. Big Data & Society 3:2053951716648174. doi:10.1177/2053951716648174.

Bross, I. 1971. Critical Levels, Statistical Language and Scientific Inference. Foundations of Statistical Inference.

Bruijnis, M.R.N., H. Hogeveen, and E.N. Stassen. 2010. Assessing economic consequences of foot disorders in dairy cattle using a dynamic stochastic simulation model. J. Dairy Sci. 93:2419–2432. doi:10.3168/jds.2009-2721.

Burema, H., and A. Filin. 2016. Aerial farm robot system for crop dusting, planting, fertilizing and other field jobs. US Pat. No. 9382003.

Bushong, J.T., J.L. Mullock, E.C. Miller, W.R. Raun, A.R. Klatt, and D. Brian Arnall. 2016. Development of an in-season estimate of yield potential utilizing optical crop sensors and soil moisture data for winter wheat. Precis. Agric. 17:451–469. doi:10.1007/s11119-016-9430-4.

Cabrera, V.E. 2010. A large Markovian linear program to optimize replacement policies and dairy herd net income for diets and nitrogen excretion. J. Dairy Sci. 93:394–406. doi:10.3168/jds.2009-2352.

Cabrera, V.E. 2012. A simple formulation and solution to the replacement problem: a practical tool to assess the economic cow value, the value of a new pregnancy, and the

cost of a pregnancy loss. J. Dairy Sci. 95:4683–4698. doi:10.3168/jds.2011-5214.

Carolan, M. 2017. Publicising food: big data, precision agriculture, and co-experimental techniques of addition. Sociol. Ruralis 57:135–154.

Casella, G., and R.L. Berger. 2002. Statistical Inference. Duxbury Pacific Grove, CA.

Chow, S.L. 1996. Statistical Significance: Rationale, Validity and Utility. Sage.

Christensen, R. 2015. Thoughts on prediction and cross-validation.

Clarke, P., and B. Wheaton. 2007. Addressing Data Sparseness in Contextual Population Research: Using Cluster Analysis to Create Synthetic Neighborhoods. Sociol. Methods Res. 35:311–351. doi:10.1177/0049124106292362.

Dieterich, J.H., and B. Kilgore. 1996. Implications of fault constitutive properties for earthquake prediction. Proc. Natl. Acad. Sci. U. S. A. 93:3787–3794. doi:10.1073/pnas.93.9.3787.

Dominiak, K.N., and A.R. Kristensen. 2017. Prioritizing alarms from sensor-based detection models in livestock production - A review on model performance and alarm reducing methods. Comput. Electron. Agric. 133:46–67. doi:10.1016/j.compag.2016.12.008.

Eales, J.S., B.K. Engel, R.J. Hauser, and S.R. Thompson. 1990. Grain price expectations of illinois farmers and grain merchandisers. Am. J. Agric. Econ. 72:701–708. doi:10.2307/1243040.

Enemark, J.M.D., G. Peters, and R.J. Jørgensen. 2003. Continuous monitoring of rumen pH--a case study with cattle. Journal of Veterinary Medicine Series A 50:62–66.

Evans, D. 2011. The Internet of Things: How the Next Evolution of the Internet Is Changing Everything.

Everitt, B.S. 2006. The Cambridge Dictionary of Statistics. Cambridge University Press.

FAO. 2009. Global agriculture towards 2050.

Feng, A., K. Sudduth, E. Vories, M. Zhang, and J. Zhou. 2018. Cotton Yield Estimation based on Plant Height From UAV-based Imagery Data. Page 1 in 2018 ASABE Annual International Meeting. American Society of Agricultural and Biological Engineers.

Fisher, R.A. 1925. Statistical Methods For Research Workers. Genesis Publishing Pvt Ltd.

Fox, D.G., M.C. Barry, R.E. Pitt, D.K. Roseler, and W.C. Stone. 1995. Application of the Cornell Net Carbohydrate and Protein model for cattle consuming forages. J. Anim. Sci. 73:267–277. doi:10.2527/1995.731267x.

Fox, D.G., C.J. Sniffen, J.D. O'Connor, J.B. Russell, and P.J. Van Soest. 1992. A net carbohydrate and protein system for evaluating cattle diets: III. Cattle requirements and diet adequacy. J. Anim. Sci. 70:3578–3596. doi:10.2527/1992.70113578x.

France, J., J.H.M. Thornley, and Others. 1984. Mathematical Models in Agriculture. Butterworths.

Francis, M. 2012. Future telescope array drives development of exabyte processing. Ars Technica.

Freua, M.C., M.H. de Almeida Santana, R.V. Ventura, L.O. Tedeschi, and J.B.S. Ferraz. 2017. Using a system of differential equations that models cattle growth to uncover the genetic basis of complex traits. J. Appl. Genet. 58:393–400.

Galon, N. 2010. The use of pedometry for estrus detection in dairy cows in Israel. J. Reprod. Dev. 56 Suppl:S48–52. doi:10.1262/jrd.1056s48.

Garrett, K.A., L.V. Madden, G. Hughes, and W.F. Pfender. 2004. New applications of statistical tools in plant pathology. Phytopathology 94:999–1003. doi:10.1094/PHYTO.2004.94.9.999.

Gelman, A., J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. 2013. Bayesian Data Analysis. Chapman and Hall/CRC.

Ghimire, S., P. Gregorini, and M.D. Hanigan. 2014. Evaluation of predictions of volatile fatty acid production rates by the Molly cow model. J. Dairy Sci. 97:354–362. doi:10.3168/jds.2012-6199.

Gill, M., D.E. Beever, and J. France. 1989. Biochemical bases needed for the mathematical representation of whole animal metabolism. Nutr. Res. Rev. 2:181–200. doi:10.1079/NRR19890014.

Green, B.F., and J.W. Tukey. 1960. Complex analyses of variance: General problems. Psychometrika 25:127–152. doi:10.1007/BF02288577.

Gregorini, P., P.C. Beukes, M.D. Hanigan, G. Waghorn, S. Muetzel, and J.P. McNamara. 2013. Comparison of updates to the Molly cow model to predict methane production from dairy cows fed pasture. J. Dairy Sci. 96:5046–5052. doi:10.3168/jds.2012-6288.

Groenendaal, H., D.T. Galligan, and H.A. Mulder. 2004. An economic spreadsheet model to determine optimal breeding and replacement decisions for dairy cattle. J. Dairy Sci. 87:2146–2157. doi:10.3168/jds.S0022-0302(04)70034-X.

Gujarati, D., and D. Porter. 2003. Multicollinearity: What happens if the regressors are correlated. Basic econometrics 363.

Hamilton, M.P., E.A. Graham, P.W. Rundel, M.F. Allen, W. Kaiser, M.H. Hansen, and D.L. Estrin. 2007. New Approaches in Embedded Networked Sensing for Terrestrial Ecological Observatories. Environ. Eng. Sci. 24:192–204. doi:10.1089/ees.2006.0045.

Hanigan, M.D., J.A.D.R.N. Appuhamy, and P. Gregorini. 2013. Revised digestive parameter estimates for the Molly cow model. J. Dairy Sci. 96:3867–3885. doi:10.3168/jds.2012-6183.

Hartgerink, C.H.J., R.C.M. van Aert, M.B. Nuijten, J.M. Wicherts, and M.A.L.M. van Assen. 2016. Distributions ofp-values smaller than .05 in psychology: what is going on?. PeerJ 4:e1935. doi:10.7717/peerj.1935.

Higgs, R.J., L.E. Chase, D.A. Ross, and M.E. Van Amburgh. 2015. Updating the Cornell Net Carbohydrate and Protein System feed library and analyzing model sensitivity to feed inputs. J. Dairy Sci. 98:6340–6360. doi:10.3168/jds.2015-9379.

Hockney, R.W., and J.W. Eastwood. 1988. Computer Simulation Using Particles. Taylor & Francis, Inc., Bristol, PA, USA.

Hoffman, L. 2019. Using Futures Prices to Forecast the Season-Average U.S. Corn Price. USDA.

Hogeveen, H., C. Kamphuis, W. Steeneveld, and H. Mollenhorst. 2010. Sensors and Clinical Mastitis—The Quest for the Perfect Alert. Sensors 10:7991–8009. doi:10.3390/s100907991.

Hristov, A.N., E. Kebreab, M. Niu, J. Oh, A. Bannink, A.R. Bayat, T.M. Boland, A.F. Brito, D.P. Casper, L.A. Crompton, J. Dijkstra, M. Eugène, P.C. Garnsworthy, N. Haque, A.L.F.

Hellwing, P. Huhtanen, M. Kreuzer, B. Kuhla, P. Lund, J. Madsen, C. Martin, P.J. Moate, S. Muetzel, C. Muñoz, N. Peiren, J.M. Powell, C.K. Reynolds, A. Schwarm, K.J. Shingfield, T.M. Storlien, M.R. Weisbjerg, D.R. Yáñez-Ruiz, and Z. Yu. 2018. Symposium review: Uncertainties in enteric methane inventories, measurement techniques, and prediction models. J. Dairy Sci. 101:6655–6674. doi:10.3168/jds.2017-13536.

Hunt, G., M. Moloney, and A. Fazio. 2011. Embarking on large-scale qualitative research: reaping the benefits of mixed methods in studying youth, clubs and drugs. Nordisk Alkohol Nark. 28:433–452. doi:10.2478/v10199-011-0040-1.

Ipema, A.H., D. Goense, P.H. Hogewerf, H.W.J. Houwers, and H. van Roest. 2008. Pilot study to monitor body temperature of dairy cows with a rumen bolus. Comput. Electron. Agric. 64:49–52. doi:10.1016/j.compag.2008.05.009.

John, R.J.L., N. Potti, and J.M. Patel. 2017. Ava: From Data to Insights Through Conversations in CIDR.

Johnson, H.A., and R.L. Baldwin. 2008. Evaluating model predictions of partitioning nitrogen excretion using the dairy cow model, Molly. Anim. Feed Sci. Technol. 143:104–126. doi:10.1016/j.anifeedsci.2007.05.007.

Kalantari, A.S., L.E. Armentano, R.D. Shaver, and V.E. Cabrera. 2016. Economic impact of nutritional grouping in dairy herds. J. Dairy Sci. 99:1672–1692. doi:10.3168/jds.2015-9810.

Kamilaris, A., A. Kartakoullis, and F.X. Prenafeta-Boldú. 2017. A review on the practice of big data analysis in agriculture. Comput. Electron. Agric. 143:23–37. doi:10.1016/j.compag.2017.09.037.

Kantanantha, N., N. Serban, and P. Griffin. 2010. Yield and Price Forecasting for Stochastic Crop Decision Planning. J.

Agric. Biol. Environ. Stat. 15:362–380. doi:10.1007/s13253-010-0025-7.

Katok, A., and B. Hasselblatt. 1997. Introduction to the Modern Theory of Dynamical Systems. Cambridge University Press.

Keshtgary, M., and A. Deljoo. 2012. An efficient wireless sensor network for precision agriculture. Canadian Journal on Multimedia and Wireless Networks 3:1–5.

Kohn, R.A., K.F. Kalscheur, and M. Hanigan. 1998. Evaluation of models for balancing the protein requirements of dairy cows. J. Dairy Sci. 81:3402–3414. doi:10.3168/jds.S0022-0302(98)75904-1.

Koomey, J.G. 2006. Best practices for understanding quantitative data.

Koonin, S.E. 1998. An independent perspective on the Human Genome Project. Science 279:36–37. doi:10.1126/science.279.5347.36.

Kwong, K.H., T.-T. Wu, H.G. Goh, K. Sasloglou, B. Stephen, I. Glover, C. Shen, W. Du, C. Michie, and I. Andonovic. 2012. Practical considerations for wireless sensor networks in cattle monitoring applications. Comput. Electron. Agric. 81:33–44. doi:10.1016/j.compag.2011.10.013.

Lanzas, C., C.J. Sniffen, S. Seo, L.O. Tedeschi, and D.G. Fox. 2007. A revised CNCPS feed carbohydrate fractionation scheme for formulating rations for ruminants. Anim. Feed Sci. Technol. 136:167–190. doi:10.1016/j.anifeedsci.2006.08.025.

Lewandowski, A. 1981. Issues in model validation.

Leys, C., C. Ley, O. Klein, P. Bernard, and L. Licata. 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. J. Exp. Soc. Psychol. 49:764–766. doi:10.1016/j.jesp.2013.03.013.

Lianga, D., H. Delgadob, and V. Cabrerac. 2018. A virtual dairy farm brain. Pages 1–19 in 13th European International Farming Systems Association (IFSA) Symposium, Farming systems: facing uncertainties and enhancing opportunities, 1-5 July 2018, Chania, Crete, Greece. International Farming Systems Association (IFSA) Europe.

Li, J., Q. Liu, H. Wang, Z. Zheng, S. Yang, and Y. Zheng. 2019. Research on Relative Height Detection System of Plant Protection UAV Based on Kalman Filter. Page 1 in 2019 ASABE Annual International Meeting. American Society of Agricultural and Biological Engineers.

Lindstrom, M.J., and D.M. Bates. 1988. Newton—Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data. J. Am. Stat. Assoc. 83:1014–1022. doi:10.1080/01621459.1988.10478693.

Liu, J., J. Li, W. Li, and J. Wu. 2016. Rethinking big data: A review on the data quality and usage issues. ISPRS J. Photogramm. Remote Sens. 115:134–142. doi:10.1016/j.isprsjprs.2015.11.006.

Lloyd, T.A. 1993. Outliers in agriculture: An intervention analysis of agricultural land values. Journal of Agricultural Economics 44:443–455. doi:10.1111/j.1477-9552.1993.tb00286.x.

Long, P., S. Lawrey, and V. Ellis. 2016. Cambridge International AS and A Level IT Coursebook with CD-ROM. Cambridge University Press.

Luna, I., and A. Lobo. 2016. Mapping Crop Planting Quality in Sugarcane from UAV Imagery: A Pilot Study in Nicaragua. Remote Sensing 8:500. doi:10.3390/rs8060500.

Maatje, K., R.M. de Mol, and W. Rossing. 1997. Cow status monitoring (health and oestrus) using detection sensors.

Comput. Electron. Agric. 16:245–254. doi:10.1016/S0168-1699(96)00052-X.

Mankin, J.B., R.V. O'Neill, H.H. Shugart, and B.W. Rust. 1977. The importance of validation in ecosystem analysis. New directions in the analysis of ecological systems, part 1:63–71.

Markle, W.H. 1966. The Manufacturing Manager's Skills. The Manufacturing Man and His Job.

Mashey, J.R. 1998. Big Data and the next wave of infraStress in Computer Science Division Seminar, University of California, Berkeley.

McCaleb, M.R. 1999. A Conceptual Data Model of Datum Systems. J. Res. Natl. Inst. Stand. Technol. 104:349.

Meiss, J.D. 2007. Differential Dynamical Systems. SIAM.

Merriam-Webster, and Inc. Staff. 2016. The Merriam-Webster Dictionary, International Edition. Merriam-Webster, Incorporated.

Miura, R., K. Yoshioka, T. Miyamoto, H. Nogami, H. Okada, and T. Itoh. 2017. Estrous detection by monitoring ventral tail base surface temperature using a wearable wireless sensor in cattle. Anim. Reprod. Sci. 180:50–57. doi:10.1016/j.anireprosci.2017.03.002.

Mok, M. 1995. Sample size requirements for 2-level designs in educational research. Multilevel modelling newsletter 7:11–15.

Mollenhorst, H., L.J. Rijkaart, and H. Hogeveen. 2012. Mastitis alert preferences of farmers milking with automatic milking systems. J. Dairy Sci. 95:2523–2530. doi:10.3168/jds.2011-4993.

Moraes, L.E., E. Kebreab, A.B. Strathe, and J. France. 2014. Bayesian analysis of energy balance data from growing

cattle using parametric and non-parametric modelling. Anim. Prod.

Morey, R.D., J.-W. Romeijn, and J.N. Rouder. 2016. The philosophy of Bayes factors and the quantification of statistical evidence. Journal of Mathematical Psychology 72:6–18. doi:10.1016/j.jmp.2015.11.001.

Morizane, R., and J.V. Bonventre. 2018. Chapter 12 - Organoids for modeling kidney disease. J.A. Davies and M.L. Lawrence, ed. Academic Press.

Morrison, D.E., and R.E. Henkel. 2006. The Significance Test Controversy: A Reader. Transaction Publishers.

Mottram, T., J. Lowe, M. McGowan, and N. Phillips. 2008. Technical note: A wireless telemetric method of monitoring clinical acidosis in dairy cows. Comput. Electron. Agric. 64:45–48. doi:10.1016/j.compag.2008.05.015.

National Research Council, Division on Engineering and Physical Sciences, Board on Mathematical Sciences and Their Applications, and Committee on Mathematical Foundations of Verification, Validation, and Uncertainty Quantification. 2012. Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification. National Academies Press.

Nianogo, R.A., and O.A. Arah. 2015. Agent-based modeling of noncommunicable diseases: a systematic review. Am. J. Public Health 105:e20–31. doi:10.2105/AJPH.2014.302426.

O'Brien, D., L. Shalloo, C. Grainger, F. Buckley, B. Horan, and M. Wallace. 2010. The influence of strain of Holstein-Friesian cow and feeding system on greenhouse gas emissions from pastoral dairy farms. Journal of Dairy Science 93:3390–3402. doi:10.3168/jds.2009-2790.

Offner, A., and D. Sauvant. 2004. Comparative evaluation of the Molly, CNCPS, and LES rumen models. Anim. Feed Sci. Technol. 112:107–130. doi:10.1016/j.anifeedsci.2003.10.008.

O'Neil, C., and R. Schutt. 2013. Doing Data Science: Straight Talk from the Frontline. "O'Reilly Media, Inc."

Orloff, J., and J. Bloom. 2014. Comparison of frequentist and Bayesian inference. Massachusetts Institute of Technology. Recuperado de https://ocw. mit. edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading20. pdf.

Østergaard, S., M.G.G. Chagunda, N.C. Friggens, T.W. Bennedsgaard, and I.C. Klaas. 2005. A stochastic model simulating pathogen-specific mastitis control in a dairy herd. J. Dairy Sci. 88:4243–4257. doi:10.3168/jds.S0022-0302(05)73111-8.

Padalalu, P., S. Mahajan, K. Dabir, S. Mitkar, and D. Javale. 2017. Smart water dripping system for agriculture/farming. Pages 659–662 in 2017 2nd International Conference for Convergence in Technology (I2CT). ieeexplore.ieee.org.

Pastell, M., L. Hänninen, A.M. de Passillé, and J. Rushen. 2010. Measures of weight distribution of dairy cows to detect lameness and the presence of hoof lesions. J. Dairy Sci. 93:954–960. doi:10.3168/jds.2009-2385.

Pennypacker, S.P., H.D. Knoble, C.E. Antle, and L.V. Madden. 1980. A flexible model for studying plant disease progression. Phytopathology 70:232–235.

Poincaré, H. 1890. Sur le problème des trois corps et les équations de la dynamique. Acta Math. 13:A3–A270.

Potgieter, A.B., G.L. Hammer, A. Doherty, and P. de Voil. 2005. A simple regional-scale model for forecasting sorghum yield across North-Eastern Australia. Agric. For. Meteorol. 132:143–153. doi:10.1016/j.agrformet.2005.07.009.

Price, R.R., and L.D. Gaultney. 1993. Soil moisture sensor for predicting seed planting depth. Trans. ASAE 36:1703–1711.

Rehm, G., and M. Schmitt. 1989. Setting realistic crop yield goals.

Reinsel, D., J. Gantz, and J. Rydning. 2017. Data Age 2025: The Evolution of Data to Life-Critical (IDC White Paper).

Reiter, S., G. Sattlecker, L. Lidauer, F. Kickinger, M. Öhlschuster, W. Auer, V. Schweinzer, D. Klein-Jöbstl, M. Drillich, and M. Iwersen. 2018. Evaluation of an ear-tag-based accelerometer for monitoring rumination in dairy cows. J. Dairy Sci. 101:3398–3411. doi:10.3168/jds.2017-12686.

Robert, C.P., J.-M. Cornuet, J.-M. Marin, and N.S. Pillai. 2011. Lack of confidence in approximate Bayesian computation model choice. Proc. Natl. Acad. Sci. U. S. A. 108:15112–15117. doi:10.1073/pnas.1102900108.

Robinson, G.K. 1991. That BLUP is a Good Thing: The Estimation of Random Effects. Stat. Sci. 6:15–32. doi:10.1214/ss/1177011926.

Roelofs, J.B., F.J.C.M. van Eerdenburg, N.M. Soede, and B. Kemp. 2005. Pedometer readings for estrous detection and as predictor for time of ovulation in dairy cattle. Theriogenology 64:1690–1703. doi:10.1016/j.theriogenology.2005.04.004.

Rose, D.C., W.J. Sutherland, C. Parker, M. Lobley, M. Winter, C. Morris, S. Twining, C. Ffoulkes, T. Amano, and L.V. Dicks. 2016. Decision support tools for agriculture: Towards effective design and delivery. Agric. Syst. 149:165–174. doi:10.1016/j.agsy.2016.09.009.

R Team. 2018. R Foundation for Statistical Computing.

Russell, J.B., J.D. O'Connor, D.G. Fox, P.J. Van Soest, and C.J. Sniffen. 1992. A net carbohydrate and protein system for

evaluating cattle diets: I. Ruminal fermentation. J. Anim. Sci. 70:3551–3561. doi:10.2527/1992.70113551x.

Rutten, C.J., A.G.J. Velthuis, W. Steeneveld, and H. Hogeveen. 2013. Invited review: sensors to support health management on dairy farms. J. Dairy Sci. 96:1928–1952. doi:10.3168/jds.2012-6107.

Saha, B., and D. Srivastava. 2014. Data quality: The other face of Big Data. Pages 1294–1297 in 2014 IEEE 30th International Conference on Data Engineering. ieeexplore.ieee.org.

Sakatani, M., M. Takahashi, and N. Takenouchi. 2016. The efficiency of vaginal temperature measurement for detection of estrus in Japanese Black cows. J. Reprod. Dev. 62:201–207. doi:10.1262/jrd.2015-095.

Sauvant, D. 1991. The use of modelling to predict animal responses to diet. RALSTON PURINA INTERNATIONAL SCIENTIFIC ADVISORY BOARD, Paris 34.

Savla, A., N. Israni, P. Dhawan, A. Mandholia, H. Bhadada, and S. Bhardwaj. 2015. Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture. Pages 1–7 in 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). ieeexplore.ieee.org.

SCS Technical Committee. 1979. Terminology for model credibility. Simulation 32:103–104.

Seltman, H.J. 2012. Experimental design and analysis. Online at: http://www. stat. cmu. edu/, hseltman/309/Book/Book. pdf.

Shamoo, A.E., and D.B. Resnik. 2009. Responsible Conduct of Research. Oxford University Press.

Shannon, C.E. 1948. A Mathematical Theory of Communication. Bell System Technical Journal 27:379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.

Silver, N. 2012. The Signal and the Noise: Why So Many Predictions Fail--but Some Don't. Penguin.

Sniffen, C.J., J.D. O'Connor, P.J. Van Soest, D.G. Fox, and J.B. Russell. 1992. A net carbohydrate and protein system for evaluating cattle diets: II. Carbohydrate and protein availability. J. Anim. Sci. 70:3562–3577.

Solie, J.B., W.R. Raun, R.W. Whitney, M.L. Stone, and J.D. Ringer. 1996. Optical sensor based field element size and sensing strategy for nitrogen application. Trans. ASAE 39:1983–1992.

Statistics Solutions. 2013. Homoscedasticity. Accessed 2019. https://www.statisticssolutions.com/homoscedasticity/.

Steeneveld, W., L.C. van der Gaag, W. Ouweltjes, H. Mollenhorst, and H. Hogeveen. 2010. Discriminating between true-positive and false-positive clinical mastitis alerts from automatic milking systems. J. Dairy Sci. 93:2559–2568. doi:10.3168/jds.2009-3020.

Stewart, J.P., J. Douglas, M. Javanbarg, Y. Bozorgnia, N.A. Abrahamson, D.M. Boore, K.W. Campbell, E. Delavaud, M. Erdik, and P.J. Stafford. 2015. Selection of Ground Motion Prediction Equations for the Global Earthquake Model. Earthquake Spectra 31:19–45. doi:10.1193/013013EQS017M.

Stone, H.S. 1971. Introduction to Computer Organization and Data Structures. McGraw-Hill, Inc., New York, NY, USA.

Stone, M.L., J.B. Solie, and W.R. Raun. 1996. Use of spectral radiance for correcting in-season fertilizer nitrogen deficiencies in winter wheat. Transactions of the.

St-Pierre, N.R. 2001. Invited review: Integrating quantitative findings from multiple studies using mixed model methodology. J. Dairy Sci. 84:741–755. doi:10.3168/jds.S0022-0302(01)74530-4.

Tedeschi, L.O., D.G. Fox, R.D. Sainz, L.G. Barioni, S.R. de Medeiros, and C. Boin. 2005. Mathematical models in ruminant nutrition. Sci. Agric. 62:76–91. doi:10.1590/S0103-90162005000100015.

Thakur, A.K. 1991. Model: Mechanistic vs Empirical. A. Rescigno and A.K. Thakur, ed. Springer US, Boston, MA.

Tran, K. 2014. Adaptive Metropolis Hastings and Factor Slice Sampling. Accessed November 14, 2019. https://www.mathworks.com/matlabcentral/fileexchange/45976-adaptive-metropolis-hastings-and-factor-slice-sampling.

Tukey, J.W. 1977. Exploratory Data Analysis. Addison-Wesley Publishing Company.

Tukey, J.W. 1980. We Need Both Exploratory and Confirmatory. Am. Stat. 34:23–25. doi:10.1080/00031305.1980.10482706.

U.S. Department of Defense. 2018. Modeling and Simulation (M&S) Verification, Validation, and Accreditation.

Van Amburgh, M.E., E.A. Collao-Saenz, R.J. Higgs, D.A. Ross, E.B. Recktenwald, E. Raffrenato, L.E. Chase, T.R. Overton, J.K. Mills, and A. Foskolos. 2015. The Cornell Net Carbohydrate and Protein System: Updates to the model and evaluation of version 6.5. J. Dairy Sci. 98:6361–6380. doi:10.3168/jds.2015-9378.

Van Amburgh, M.E., A. Foskolos, E. Collao-Saenz, R. Higgs, and D.A. Ross. 2013. Updating the CNCPS feed library with new feed amino acid profiles and efficiencies of use: Evaluation of model predictions, version 6.5 in 75th Cornell Nutrition Conference for Feed Manufacturers. pure.aber.ac.uk.

Wani, F.J., S.E.H. Rizvi, M.K. Sharma, and M. Iqbal Jeelani Bhat. 2018. A study on cross validation for model selection and estimation. INTERNATIONAL JOURNAL OF AGRICULTURAL SCIENCES 14:165–172. doi:10.15740/has/ijas/14.1/165-172.

Wasserstein, R.L., and N.A. Lazar. 2016. The ASA Statement on p-Values: Context, Process, and Purpose. The American Statistician 70:129–133. doi:10.1080/00031305.2016.1154108.

White, R.R., and M.B. Hall. 2017. Nutritional and greenhouse gas impacts of removing animals from US agriculture. Proc. Natl. Acad. Sci. U. S. A. 201707322. doi:10.1073/pnas.1707322114.

White, T.A., V.O. Snow, and W.M. King. 2010. Intensification of New Zealand beef farming systems. Agric. Syst. 103:21–35. doi:10.1016/j.agsy.2009.08.003.

Williams, W.F., D.R. Yver, and T.S. Gross. 1981. Comparison of estrus detection techniques in dairy heifers. J. Dairy Sci. 64:1738–1741. doi:10.3168/jds.S0022-0302(81)82757-9.

Wolfger, B., B.W. Jones, K. Orsel, and J.M. Bewley. 2017. Technical note: Evaluation of an ear-attached real-time location monitoring system. J. Dairy Sci. 100:2219–2224. doi:10.3168/jds.2016-11527.

Xu, Q.-S., and Y.-Z. Liang. 2001. Monte Carlo cross validation. Chemometrics Intellig. Lab. Syst. 56:1–11. doi:10.1016/S0169-7439(00)00122-2.

Xu, Z.Z., D.J. McKnight, R. Vishwanath, C.J. Pitt, and L.J. Burton. 1998. Estrus detection using radiotelemetry or visual observation and tail painting for dairy cows on pasture. J. Dairy Sci. 81:2890–2896. doi:10.3168/jds.S0022-0302(98)75849-7.

Yan, X., and X. Su. 2009. Linear Regression Analysis: Theory and Computing. World Scientific.

Zehner, N., C. Umstätter, J.J. Niederhauser, and M. Schick. 2017. System specification and validation of a noseband pressure sensor for measurement of ruminating and eating behavior in stable-fed cows. Comput. Electron. Agric. 136:31–41. doi:10.1016/j.compag.2017.02.021.

Ziliak, S., and D.N. McCloskey. 2008. The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives. University of Michigan Press.

Figures

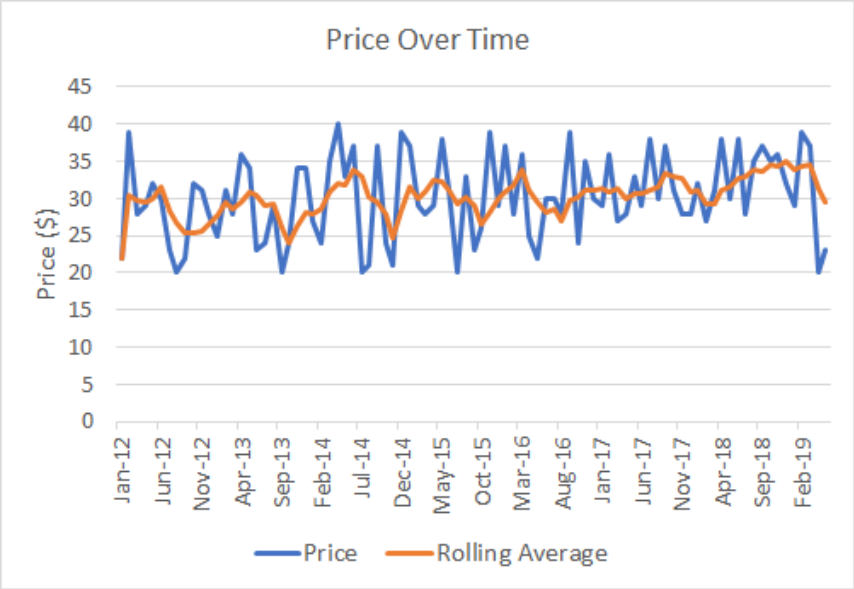

Figure 2-1. Example graph of milk price data over time. Lines for actual price and 6-month moving average are shown.

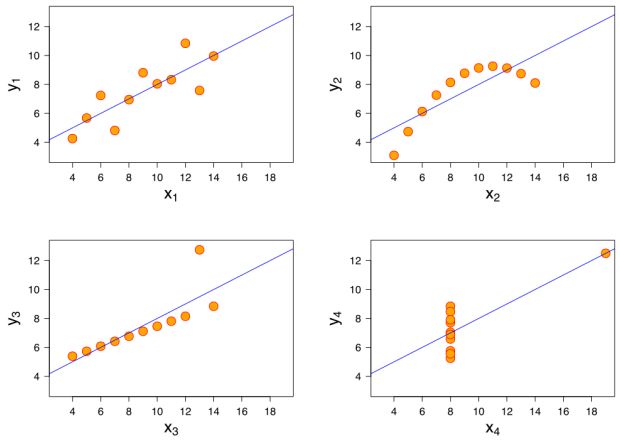

Figure 2-2. Anscombe's quartet. A series of four datasets with equal x and y means, sample variances, correlation, linear regression line, and coefficient of determination (Anscombe, 1973)

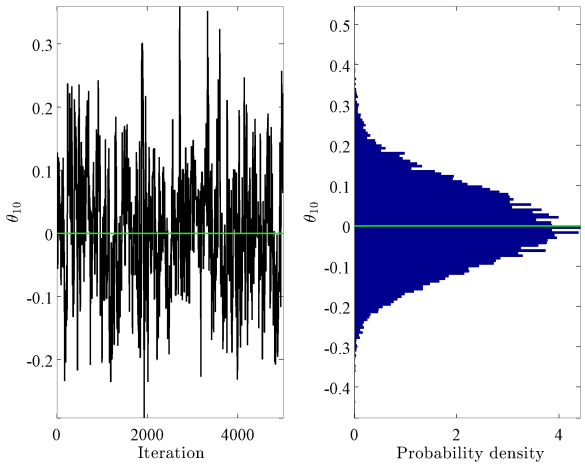

Figure 2-3. Resulting samples from an algorithm used to sample from a parameter distribution with unknown density. The plot on the left is called a "random walk" because each generated point is different from the last and is proposed randomly, although sampling algorithms have different operations to decide whether to accept the proposed move in parameter space.

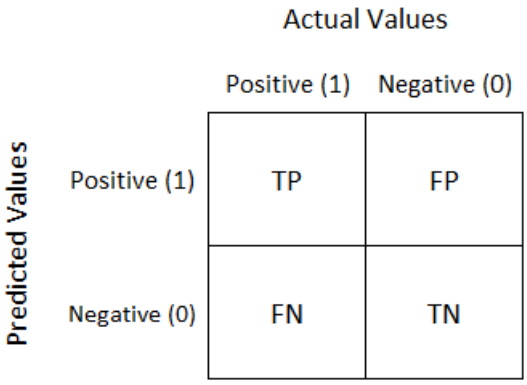

Figure 2-4. Confusion matrix showing the possible outcomes of a binary prediction model. TP = True positive; FP = False positive; TN = True negative; FN = False negative.

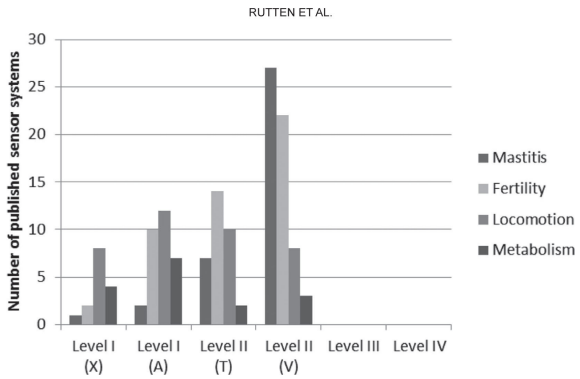

Figure 2-5. Number of studied sensor systems for mastitis, fertility, locomotion, and metabolism per development level (Rutten et al., 2013): Level I = technique, Level II = data interpretation, Level III = integration of information, and Level IV = decision making. Levels I and II are subdivided into solely measuring a parameter (Level I, X), an assumed relation between gold standard and sensor data (Level I, A), a statistically tested relation between gold standard and sensor data (Level II, T), or a validated algorithm for detection (Level II, V).

CHAPTER 3: Contributions of dairy products to environmental impacts and nutritional supplies from United States agriculture

**To be submitted to Journal of Dairy Science

Abstract

Questions regarding the balance between contribution to human nutrition and environmental impact of livestock food products rarely evaluate specific species or how to accomplish the recommended depopulation. The objective of this study was to assess current contributions of the U.S. dairy industry to the supply of nutrients and environmental impact, characterize potential impacts of alternative land use for land previously used for crops for dairy cattle, and evaluate the approaches' impacts on U.S. dairy herd depopulation. Data on U.S. dairy production were obtained from the analysis conducted by White and Hall (2017; https://doi.org/10.1073/pnas.1707322114). We modeled three scenarios to reflect different sets of assumptions for how and why to remove dairy cattle from the U.S. food production system coupled with four land use strategies for the potential newly available land previously cropped for dairy feed. Scenarios also differed in assumptions of how to repurpose land previously used to grow grain for dairy cows. The current system provides sufficient fluid milk to meet the annual energy, protein, and calcium requirements of 71.2, 169, and 254 million people, respectively. Vitamins supplied by dairy products also make up a high proportion of total domestic supplies from foods, with dairy providing 39% of the vitamin A,

54% of the vitamin D, 47% of the riboflavin, 57% of the vitamin B12, and 29% of the choline available for human consumption in the U.S. Retiring, or cows maintained without product collection, dairy cattle under their current management resulted in no change in absolute greenhouse gas emissions (GHGE) relative to the current production system. Both depopulation and retirement to pasture resulted in modest reductions (6.8 to 12.0%) in GHGE relative to the current agricultural system. Most dairy cow removal scenarios reduced availability of essential micronutrients such as alpha-linolenic acid, vitamins A, D, B12 and choline, as well as Ca. Those removal scenarios that did not reduce micronutrient availability also did not improve GHGE relative to the current production system. These results suggest that reducing GHGE without reducing the supply of the most limiting nutrients to the population is difficult.

Keywords: dairy, calcium, protein, greenhouse gases

Introduction

The UN Food and Agriculture Organization (FAO) recommends that food production nearly double from 2009 levels by 2050 to ensure sufficient human nutrition worldwide (FAO, 2009). Increased food production is a major challenge because of existing limitations in land and water availability, food distribution and storage solutions to eliminate food waste, and yield efficiencies, among other factors (Gupta and Deshpande, 2004; Bruinsma, 2009; Sauer et al., 2010). A common recommendation when considering this impending food crisis is to eliminate or reduce animal production in favor of plant sources (Aleksandrowicz et al. 2016). Many consider livestock and poultry production resource-intensive in terms of land-use, GHGE, and water use per kilocalorie of food produced, having significantly greater environmental impacts when compared to plant-source foods (Clark and Tillman, 2017). Despite the simplification of this issue, the public frequently views animal production as resource-intensive without considering variability or other factors like nutrient profile or viability of alternatives. Accordingly, some research has called to reduce consumption of animal-derived foods (Pimentel and Pimentel, 2003; Weber and Matthews, 2008). Some suggest the near elimination of animal agriculture based on environmental, human health, and ethical bases (Willett et al., 2019). However, an assessment of U.S. agriculture revealed increased micronutrient deficiencies, in

terms of human nutrient requirements per year (HNRY), despite greater food availability in a simulated system without farmed animals as food resources (White and Hall, 2017). A major difference between the White and Hall (2017) assessment and other studies is the assumption regarding land use in a system without animals: White and Hall (2017) assumed similar land use to the current agricultural system while other assessments assume land use will adapt to meet food demand (Emery, 2018). Neither of these strategies is ideal (Springmann et al., 2018) given that agricultural land use is dynamic and governed by physical, chemical, climatological, biological, economic, and social factors. As such, there is a need to consider the mechanics of how land use within the agricultural system might adapt under a scenario where society moves toward reduced consumption of animal products.

An additional important consideration in these assessments involves the assumptions about what happens to the supporting animal populations when we reduce consumption of animal-source foods. Although may be easy to recommend a world without livestock, it is less comfortable to discuss how we might get to such a world. Assessing elimination of dairy cattle production is a logical starting place for this type of assessment for several reasons. First, dairy cattle predominantly consume total-mixed rations in confinement systems, making cows accessible, which allows for

easier implementation of strategies aimed at reducing production (entire elimination and movement to pasture-based production, among others). It is important to consider these strategies for scaling back production because the environmental and human food benefits from entire depopulation of cattle will undoubtedly differ from the benefits if cattle persist as a feral or semi-managed population. Dairy is also an interesting case because dairy products have unique nutrient composition (USDA, 2018b) and production of milk from dairy cattle has a lower environmental impact than meat production (Nijdam et al., 2012; Luo et al., 2015) and some plant source products such as lettuce (Marvinney, 2016). As such, understanding what role dairy products, specifically, play in the U.S. agricultural system and the nutritional and environmental impacts associated with removing dairy production would be of use in assessing dairy production's utility in the U.S. food production system.

The objectives of this study were: 1) to ascertain the current contributions of dairy products to the nutrient supply in the U.S.; 2) to evaluate impacts of approaches to depopulation of the U.S. dairy herd by estimating GHGE, land use, and HNRY; and 3) to evaluate the potential impact of alternatives for land use for land previously used for crops fed to dairy cattle.

Materials and Methods

Data on U.S. dairy production were obtained from the analysis conducted by White and Hall (2017), which utilized data from the U.S. Department of Agriculture (USDA, 2018a) and Economic Research Service (USDA/ERS, 2018a; USDA/ERS, 2018b) and Food Composition databases (USDA, 2018b); the U.S. Environmental Protection Agency (US Environmental Protection Agency, 2010); the United Nations Food and Agriculture Organization (FAO, 2013); and other peer-reviewed, published sources to estimate nutritional and greenhouse gas contributions of livestock to U.S. agriculture. In the current work, we disaggregated the reported animal metrics to specifically assess the contributions of the U.S. dairy industry to nutrient supply and GHG emission (GHGE) within the agricultural system. Unless otherwise specified, metrics are estimated as described in White and Hall (2017). All modeling was based on estimated population sizes from the above data. All assessments terminated at fluid milk production and did not consider further processing and does not include foods derived from dairy animal carcasses.

Scenarios for Removing Dairy Products for U.S. Use

We assessed three scenarios which differed in their assumptions about the removal of dairy cattle from the agricultural system. The method of animal removal was an attribute in the scenarios studied because it addresses societal concerns about the fates of

animals and affects alterations in GHGE and land use. Assuming the U.S. stopped consuming dairy products, we considered three possible scenarios:

- Depopulation (DEP)
- Current management; export dairy (CME)
- Retirement (RET)

In DEP, dairy animals would be depopulated in response to consumers ceasing consumption of dairy products. In CME, dairy cattle would be kept under current management and milk produced would go to products other than human food or would be exported from the U.S. for human consumption. In RET, dairy cattle would be retired to a pasture-based management system. In this third scenario, the number of lactating cows in the national herd was reduced to that which could be supported by the available pastureland. Land use was a focus in all animal removal scenarios because of the concerns raised in response to previous work (White and Hall, 2017; Emery, 2018; Springmann et al., 2018; Van Meerbeek and Svenning, 2018) and in the surprising findings related to increasing fruit and vegetable production reported in White and Hall (2018).

Depopulation

In simulating DEP, we only compared diets for the U.S. human population and outputs pre- and post-cattle depopulation and

considered the transition period to be instantaneous. That is, no food product resulted from the slaughter of the dairy cattle population, given the short duration and non-renewable nature of the event. If dairy cattle are no longer present in U.S. agriculture, we must consider downstream effects like handling of pasture and grain land previously used for producing dairy feed, disposition of byproduct feeds, and sourcing fertilizer.

We modeled several cropland allocation options to reflect different sets of assumptions for repurposing land for crop production that previously grew feed for dairy cattle. White and Hall (2017) assumed that all cropland used to grow grain crops for animal feed would continue to be used for growing grain, though others contended that it may be more appropriate to reallocate this land along with the land used for silage production for the cultivation of non-grain crops (Emery, 2018; Springmann et al., 2018; Van Meerbeek and Svenning, 2018). Here, 2 options for dairy land reallocation to other crops were tested: 1) reallocate silage land only or 2) reallocate silage and grain land. All DEP scenarios did not repurpose pastureland because we assumed it could be repurposed for beef cattle production. To test how land use change might influence scenario outcomes, we tested each of four land use options (LU-1 through LU-4) with crop reallocation of newly available land previously used to produce silage, or grain and silage, for dairy cattle: LU-1) *Current Proportions*, all newly

available land was replaced with crops according to the current proportions of crops grown in the U.S.; LU-2) *Fruits and Vegetables*, all newly available land was planted to fruits and vegetables only, according to their current proportion of crops in the U.S.; LU-3) *Nuts and Legumes*, all newly available land was planted to nuts and legumes only, according to their current proportion of crops in the U.S.; and LU-4) *Non-grain/oilseed/sugar*, all newly available land was planted to any crop except those used to produce grains, oilseeds and sugar, according to their current proportion of crops in the U.S. Figure 3-1 shows land reallocation and land use options within DEP.

We assumed land used for silage crops (3.1 million ha) to be dairy-specific and repurposed for production of other crops. To test the effect of re-allocating the land for grain consumed by dairy cattle, it is essential to calculate the land area used for producing grain for dairy cattle. Eshel et al., (2014) estimated the proportions of grain consumed by the dairy industry (kg consumed by dairy cattle / kg produced). Yield data from USDA NASS (2018) was used to estimate grain land-area (proportion of grain consumed by dairy multiplied by land area for grain production). That liberated grain land-area (3.7 million ha) was then re-allocated based on the previously described land use options (Table 3-1).

Other important assumptions in each dairy scenario included handling of dairy cattle' byproduct feeds, like bakery products or wheat middlings, and fertilizer. In DEP, other livestock utilized all dairy production byproducts, resulting in no net GHGE from the disposal of byproducts. Synthetic fertilizers replaced fertilizer produced using dairy manure and these synthetic fertilizers accounted for additional agricultural GHG.

Current Management; no human products/exports

Under CME, we assumed that dairy products would be exported with none entering the U.S. food system, but the industry would continue to house and manage cows and bulls in a similar manner to current practice. In this scenario, we assume no land liberation because cows would continue to eat silage, grain, and pasture as they do today. As such, none of the land use options applied to CME. Similarly, dairy cows would continue to consume byproduct feeds and all manure would still be available for use as fertilizer. Essentially, CME retained all aspects of the system the same as our current system, except nutrient availability from dairy products and meat from culled animals to the U.S. population.

Retirement

RET reflects the idealistic perspective of ending milk production while allowing remaining dairy cattle to continue in a pasture-based setting. This scenario reflects what might happen if we allowed a reduced population of cattle to roam and breed freely on pastureland. RET addresses the magnitude of possible animal numbers given the carrying capacity of the land and impacts of the retained herd. We assumed available pastureland was equal to combined, current pasture and silage land areas. We did not assess costs of fencing, infrastructure, or other peripherals. The conversion of silage land and use of pastureland means that none of this land would be available for additional crop production, as a result, this scenario did not consider any of the land use options described in DEP. The carrying capacity of cows on this land was calculated based on maintenance intake (National Research Council, 2001) of pasture (12.8 kg DM/d) and an estimated annual yield of 6,200 kg DM per ha of pasture. The carrying capacity was estimated at 4.176 million individual animals, approximately 44% of the current population (USDA, 2018a). To achieve this population size either animals would have to be released onto the fenced lands, leading to issues of survival, and oscillations in population size with changes in pasture availability, or humans would have to intervene to cull the animals in excess of the carrying capacity of the available pastureland.

Handling of byproduct feeds and fertilizer in RET was assumed to be a hybrid of the DEP and CME. Byproduct feeds previously consumed by dairy cattle were assumed to be repurposed for consumption by other livestock industries, meaning that no environmental penalty was considered for byproduct feed disposal. Nutrients in dairy manure were assumed to be deposited directly onto the pasture and not recovered for use as fertilizer. As such, we assumed that additional synthetic fertilizer would be used to replace the manure fertilizer that would previously have been produced by the dairy industry and used on existing cropland.

In previous work by White and Hall (2017), the production of food products estimates the carbon footprint of the agricultural system. The kilograms of milk produced by all dairy cows estimated the GHGE associated with the dairy industry, approximately 1.23 kg of carbon dioxide equivalents (CO_2-e) per kg of fat and protein corrected milk (Thoma et al., 2013); however, the carbon footprint estimate for milk is only valid for the current U.S. dairy industry and would not be an appropriate reflection of the emissions from the dairy cattle RET system. The diets of dairy cows raised in confinement systems is quite different than that of dairy cows on pasture. As such, we only used enteric and manure methane and nitrous oxide emissions to estimate emissions from RET. Enteric methane emissions were estimated based on the equations listed in (Ellis et al., 2007) and pasture composition data from the DairyOne

feed library (dairyone.com). Manure methane and nitrous oxide emissions were calculated using IPCC tier II methodology (Intergovernmental Panel on Climate Change and Intergovernmental Panel on Climate Change). Methane and nitrous oxide were converted to carbon dioxide equivalents (CO_2-e) assuming 25 kg CO_2-e per kg CH_4 and 289 kg CO_2-e per kg N_2O. These CO_2-e were used for consistency with other GHGE estimates, though it should be noted that other research has called into question the validity of CO_2-e estimates on enteric methane estimates (Allen et al., 2018).

Comparisons Among Scenarios

The proposed dairy cow removal scenarios, land reallocation and land use options, described above were intended for specific comparisons. The way in which cattle are removed from the food production system is examined by comparing DEP LU-1 (current proportions) with silage and grain land reallocation, CME and RET. This comparison is important because the way in which we eliminate, or export production has potential relevance on the environmental impacts and nutritional profile, in terms of HNRY, of the food produced by the agricultural system. A second set of comparisons relies on evaluating the different land use options within the DEP scenario. If we were to remove dairy cattle entirely from U.S. agriculture, it is important to consider what agricultural products might take their place.

Results and Discussion

Current Contributions of the Dairy Industry to Nutrient Supplies and GHG Emissions

Dairy products contribute substantially to the supply of human-edible nutrients in the current U.S. agricultural system. The current system provides sufficient fluid milk to meet the annual energy, protein, and calcium requirements of 71, 169, and 254 million people, respectively. Calcium content and availability in dairy products makes it feasible to meet calcium nutrient requirements from foods, whereas achieving that on a strictly plant-based diet is largely impractical without fortification or supplements (Weaver et al., 1999). Dairy products are a significant component of the protein supply in the U.S., providing 20% of the protein and 20 to 30% of many essential AA. According to previous assessments, whole milk protein's digestible indispensable AA score (DIAAS), a reflection of the nutritional value of proteins to humans, is anywhere between 15.5% (Ertl et al., 2016) and 30% (Rutherfurd et al., 2015), with greater than values reported for legume protein sources. The new DIAAS system measures the true ileal digestibility of proteins. The DIAAS system gives greater credit to the AA quality of animal protein sources for meeting human needs than did the previous protein digestibility-corrected AA score system (PDCAAS) which truncates the values' sum of animal proteins at 100% and generally overestimated nitrogen

digestibility of plant-based proteins (Rutherfurd et al., 2015). A report by the U.N. Food and Agriculture Organization (2013) recommended the adoption of DIAAS as a replacement for PDCAAS as a more accurate descriptor of protein nutritional value.

Vitamins supplied by dairy products also make up a high proportion of total domestic supplies, with dairy providing 39% of the vitamin A, 54% of the vitamin D, 47% of the riboflavin, 57% of the vitamin B12, and 29% of the choline available for human consumption in the U.S. These vitamins are often in low supply in the U.S. food production system (White and Hall, 2017), but are essential for eye (vitamin A), bone (vitamin D), brain (vitamin B12), and organ (choline) health and energy metabolism (riboflavin). A study of the contribution of dairy products to essential micronutrient intakes in France identified vitamins (B_{12}, choline, D, and A) as important contributors of dairy to human diets (Coudray, 2011).

In agreement with the present study, numerous other reports have identified the nutritional importance of dairy products in developed (Hess et al., 2015) and developing countries (Hoddinott et al., 2013; Murphy et al., 2016). In particular, studies find that dairy products are an important source of Ca (Murphy et al., 2016), a macromineral essential for bone and tooth health, muscle and enzyme function, and blood clotting, among other functions. Dairy products provide a greater amount of absorbable Ca per

serving than the majority of vegetable sources (Weaver et al. 1999). Although this study only followed fluid milk production, it is also important to note that microorganisms in fermented dairy products can also contribute to human health (Fernández et al., 2015) both directly (probiotics) and indirectly through the production of metabolically active compounds like vitamins, linoleic acid, and others. Dairy products are not free from speculation about negative effects on human health (Thorning et al., 2016; Zhu and Kannan, 2018); however, their role in providing a substantial supply of essential, bioavailable nutrients for human consumption is clear.

The U. S. dairy industry accounts for 16% of GHGE from all of U.S. agriculture (White and Hall, 2017). Based on most recent estimates, the U.S. Environmental Protection Agency found that 8.4% of total U.S. GHGE were the result of agricultural activities (U.S. Environmental Protection Agency, 2019). Using the assumption that dairy production accounts for 16% of agricultural GHGE and agricultural emissions makeup 8.4% of total U.S. GHGE, our numbers suggest the U.S. dairy industry responsible for about 1.28% of total U.S. GHGE. All subsequent results regarding GHG values and reductions reported herein will be in terms of their proportion of current agricultural GHGE.

Impacts of Cattle Removal Strategy

If the U.S. were to discontinue dairy production, the question of what should happen to the current dairy herd has animal welfare and public perception concerns Figure 3-2 shows the estimated total agricultural GHGE with each of the 3 dairy cow removal scenarios, DEP, CME, and RET, compared with current production. Figure 3-3 includes the nutrient supplies estimated from these scenarios. By design, CME showed no difference (±0.0%) from current production or baseline in terms of GHGE (Figure 3-2), as dairy production continued, and products were simply moved elsewhere. Additionally, using dairy cows for non-consumable products, like leather, animal feed or manure, and exports yielded a decrease in many domestic, human-edible nutrient supplies in terms of HNRY compared to current diets and when compared to the DEP scenario (Figure 3-3). A RET scenario showed a 11.97% decline in total agricultural GHG (Figure 3-2) compared to current emissions. This GHG decline with RET is likely because of the reduced population of cows sustainable on available pastureland. Along with this reduction in agricultural GHGE under RET, domestically available supplies of all nutrients decreased. The CME and RET scenarios use the same amount of land and therefore both averaged an 18% reduction in HNRY supply compared with current production across all nutrients measured (Figure 3-3). All 39 nutrients either declined or remained the same in CME and RET. Total energy HNRY harvested from the

agricultural system in RET decreased by 11% compared with current production. Although CME and RET could be considered more publicly-favored because they retain dairy cows in the U.S., the DEP scenario allows more freedom in terms of land reallocation. Under DEP assumptions, GHGE declined 7.2% compared to current levels (Figure 3-2). Nutrient supplies under DEP rose 42% on average, with 30 of the 39 nutrients measured increasing compared to levels in our current system. Comparing potential dairy removal scenarios demonstrates the likely trade-offs inherent in affecting change in agricultural systems: namely, it is difficult to find scenarios that simultaneously increase supplies of critically limiting nutrients and decrease GHGE. Table 3-2 compares GHGE changes on the basis of total agricultural GHGE, dairy GHGE only, and total U.S. GHGE.

Although RET has limited economic justification, it was important to assess from the social dimension. The approximately 44% of the national dairy herd retired to pastureland would produce 11.6% less of national agricultural GHGE, declining from the current 16% because of two major factors: reduced numbers of cows, and a change in the management of those animals. Because existing pastureland would need to sustain all retired cows, RET would sustain only an estimated 44% of the current dairy cattle population. Additionally, whereas current dairy production relies heavily on high intakes (25 to 30 kg of DM/cow/d) of total-mixed

rations comprised of silage, grains, and byproduct feeds to support milk production, the current scenario assumed low intake (12.8 kg DM/cow/d) of pasture only. Despite methane emissions on forage-based diets being higher than diets with greater inclusion of cereal grains and byproduct feeds, total feed (energy) intake, which decreases, is the major driver of emissions (Johnson and Johnson, 1995). A lesser factor contributing to the decline in GHG under RET is the accounting of emissions associated with fertilizer synthesis. In RET, we assumed fertilizer production would increase because of the challenges associated with harvesting manure from pasture-based housing systems. This synthesis of fertilizer, through processes like the Haber Bosch process (Haber, 1905), accounted for a 1.0% increase in agricultural GHGE in RET, or about 9% of the GHG that would have been lost by removing excess dairy cows.

Although GHGE considerations are important, dairy cows must ultimately produce foods for human consumption. In addition to achieving minimal or no reductions in GHGE, production decreased in many nutrients when considering future scenarios which retain dairy cows (CME or RET) when compared to the current HNRY supply. Calcium, alpha-linolenic acid, vitamin A, vitamin D, vitamin B_{12}, choline, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine all decreased 20% or more in HNRY in CME and RET scenarios compared to current

production system (Figure 3-3). Protein is not in the figure, because its decline relative to the baseline was only 19%. Relative to the current contributions of dairy to the U.S. agricultural system, domestically produced supplies of Ca (-72.6%), vitamin B_{12} (-56.7%) and vitamin D (-53.9%) were the nutrients most markedly affected by retiring dairy cattle. These reductions translate to 254 million less people meeting their Ca requirements for the year, 500 million less for B_{12}, and 16 million less for vitamin D. The nutritional importance of Ca to humans was discussed above. Vitamin B_{12} is essential for normal function of the central nervous system, in the formation of red blood cells, and in cellular metabolism (Wokes et al., 1955), and is particularly important for the correct development of the brain (Müller-Wielsch et al., 2010). Vitamin D is essential for skeletal homeostasis and prevention of bone disorders. Additionally, suboptimal vitamin D status is implicated in chronic autoimmune and cardiovascular disease, hypertension, and common cancers (Hewison, 2012). When considering the impact of removing any animal production system, we should consider the nutrients produced and the proportion of the population whose nutrient requirements will be met.

Impacts of Land Allocation Strategy

Of the approximately 134 million ha of land considered in this analysis, the 3.1 million ha of arable land previously

allocated to silage was available for reallocation in DEP. Under LU-1 with silage and grain land reallocation, the grain supply decreased by 2%, while under LU-2, LU-3, and LU-4, grain production decreased by 5% compared to current production. Grains provide an energy- and nutrient-dense food source (Macdiarmid et al., 2012). Optimization of food production for human diets for either cost, environmental impact, or both, tend to have high amounts of grains because grains can be produced efficiently and can also be fortified with missing, but required, nutrients (Clydesdale, 1994; Cook et al., 1997; Macdiarmid et al., 2012). The relatively small change in grain land when accounting for land previously used to feed dairy cattle suggests that dairy cattle consume very minimal quantities of human-edible grains, thus are minimally competitive for human food.

Assessing the different land-use options makes it clear that shifts in use of relatively small land areas (e.g., the 3.1 million ha of land in the U.S. allocated to silage currently compared to the approximately 134 million ha of U.S. cropland in the baseline scenario) can have substantial effects on the production of nutrients from the agricultural system. Figure 3-4 compares current or baseline GHGE to DEP with silage and grain reallocation and land use for fruit and vegetable production (LU-2). Figure 3-5 illustrates food production (in kg) and CO_2-e, broken down by food product, under LU-2. Under LU-2 with silage and grain land

reallocation, GHGE increased by 9.9% when compared to current emissions. However, LU-2 without grain land reallocation resulted in a net neutral effect (+0.04%) on GHGE. Following the assumptions of DEP with no grain land reallocation, using LU-1, LU-3, and LU-4 resulted in decreased GHGE of 6.88%, 8.18%, and 7.59%, when compared to current production. Fruits and vegetables tend to be more carbon-intensive, GHG emitting crops compared to grains, making them less likely to appear in optimized diets (Macdiarmid et al., 2012; Wilson et al., 2013; Gephart et al., 2016). Although fruits and vegetables can provide some of the same nutrients as dairy products, like vitamin A and vitamin C, the increased GHG cost makes this option less desirable. As described above, the non-fruit and vegetable scenarios also resulted in reduced availability of critical micronutrients supplied in high concentration in dairy products. As a result, our analysis suggests that all proposed land reallocation options appear to be suboptimal for limiting GHGE while still meeting the nutrient requirements of the population. Figure 3-6 compares the GHGE of all land reallocation scenarios and land-use options.

The removal of dairy cows from the U.S. agricultural system under DEP, with each land use option, increases land available and crops yields (approximately 17%) in terms of total energy of nutrients supplied. However, under any dairy removal scenarios, the land allocation options further reduce the supply of vitamin

D, choline, calcium, vitamin A, and alpha-linolenic acid, which are all within the 11 least abundant nutrients in our analysis. In the current food production system, calcium is in sufficient supply to meet the requirements of approximately 350 million humans. Under DEP with grain land reallocation and LU-1, LU-2, LU-3, and LU-4, calcium supply would change by -89.1, -57.6, -87.4, -85.3 million HNRY respectively. Figure 3-7 shows the changes in calcium and other least-abundant nutrients within the DEP scenario under all land use options with grain land reallocation. The data suggest reduced availability of these micronutrients regardless of land use and grain land reallocation. The declines in supply of the most limiting nutrients of the U.S. food production system illustrates the nutritional impact of removing dairy cows from agriculture: dairy cows provide a relatively efficient, nutrient-dense source of valuable micronutrients that cannot currently be mirrored in common plant-source foods.

Practical Feasibility of Land Use Options

Another important consideration for the different land use options evaluated is the actual suitability of land for various agricultural practices. There is a diverse literature evaluating indices to characterize the suitability of agricultural land (Littleboy et al., 1996; Reshmidevi et al., 2009; Singha and Swain, 2016; Senagi et al., 2017). Although some horticultural crops,

tree nuts, and fruits can be competitive with grain crops in terms of land use (Wolz and DeLucia, 2019), many are particularly sensitive to climate and there is concern over how climate change will influence the productivity of these crops (Luedeling et al., 2009; Parker and Abatzoglou, 2017). Indeed, a geospatial analysis of the United States based on suitability of land for growing selected fruits and vegetables suggests only 144,000 ha of agricultural land area is suitable for repurposing for fruit and vegetable production (Conrad et al., 2017). For reference, the current assessment reflects a 3.1 to 6.8 million ha change in land use (21 to 47 times the expected suitable land area). To effect such a change in land use, there would need to be substantial technological improvements to support growing fruits and vegetables on land currently unsuited to the purpose. Overall, analyses of the soil characteristics, climatological parameters, and other factors for agricultural areas across the U.S. suggest limited opportunity to expand fruit and vegetable production, in particular, and we must take these practical challenges into account when planning alternative land uses in agricultural systems.

Conclusion

Our investigations into the impacts and alternatives when removing dairy cows from U.S. production agriculture suggest that

GHGE changes would be minor, equivalent of 0.7% of total U.S. GHGE. Emissions may increase if we optimize cropland through the production of more carbon-intensive crops to improve the current nutrient supply to the US population. At the same time, supplies of some limiting essential nutrients for the human population would decline. Lastly, any reductions in GHGE or increases in available cropland come at the cost of culling more or all dairy cattle. Scenarios involving such culling incurs ethical costs not assessed in the current work.

Acknowledgements

Research was supported by Dairy Management Inc.

References

Aleksandrowicz, L., R. Green, E. J. M. Joy, P. Smith, and A. Haines. 2016. The Impacts of Dietary Change on Greenhouse Gas Emissions, Land Use, Water Use, and Health: A Systematic Review. PLoS One. 11:e0165797. doi:10.1371/journal.pone.0165797. Available from: http://dx.doi.org/10.1371/journal.pone.0165797

Allen, M.R., Shine, K.P., Fuglestvedt, J.S., Millar, R.J., Cain, M., Frame, D.J. and Macey, A.H., 2018. A solution to the misrepresentations of CO 2-equivalent emissions of short-lived climate pollutants under ambitious mitigation. *Npj Climate and Atmospheric Science*, *1*(1), p.16.

Bruinsma, J. 2009. The resource outlook to 2050: by how much do land, water and crop yields need to increase by 2050? Page in How to feed the World in 2050. Food and Agriculture Organization of the United Nations (FAO).

Clark, M., and D. Tilman. 2017. Comparative analysis of environmental impacts of agricultural production systems, agricultural input efficiency, and food choice. Environ. Res. Lett. 12:064016. doi:10.1088/1748-9326/aa6cd5.

Clydesdale, F. M. 1994. Optimizing the diet with whole grains. Crit. Rev. Food Sci. Nutr. 34:453–471. doi:10.1080/10408399409527675. Available from: http://dx.doi.org/10.1080/10408399409527675

Conrad, Z., C. Peters, K. Chui, L. Jahns, and T. Griffin. 2017. Agricultural Capacity to Increase the Production of Select Fruits and Vegetables in the US: A Geospatial Modeling Analysis. Int. J. Environ. Res. Public Health. 14:1106. doi:10.3390/ijerph14101106. Available from: http://dx.doi.org/10.3390/ijerph14101106

Cook, J. D., M. B. Reddy, J. Burri, M. A. Juillerat, and R. F. Hurrell. 1997. The influence of different cereal grains on

iron absorption from infant cereal foods. Am. J. Clin. Nutr. 65:964–969. doi:10.1093/ajcn/65.4.964. Available from: http://dx.doi.org/10.1093/ajcn/65.4.964

Coudray, B. 2011. The contribution of dairy products to micronutrient intakes in France. J. Am. Coll. Nutr. 30:410S–4S. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22081686

Ellis, J. L., E. Kebreab, N. E. Odongo, B. W. McBride, E. K. Okine, and J. France. 2007. Prediction of methane production from dairy and beef cattle. J. Dairy Sci. 90:3456–3466. doi:10.3168/jds.2006-675. Available from: http://dx.doi.org/10.3168/jds.2006-675

Emery, I. 2018. Without animals, US farmers would reduce feed crop production. Proc. Natl. Acad. Sci. U. S. A. 115:E1703. doi:10.1073/pnas.1720760115. Available from: http://dx.doi.org/10.1073/pnas.1720760115

Ertl, P., W. Knaus, and W. Zollitsch. 2016. An approach to including protein quality when assessing the net contribution of livestock to human food supply. Animal 10: 1883-1889. Available from doi:10.1017/S1751731116000902

Eshel, G., A. Shepon, T. Makov, and R. Milo. 2014. Land, irrigation water, greenhouse gas, and reactive nitrogen burdens of meat, eggs, and dairy production in the United States. Proc. Natl. Acad. Sci. U. S. A. 111:11996–12001. doi:10.1073/pnas.1402183111. Available from: http://dx.doi.org/10.1073/pnas.1402183111

FAO. 2009. Global agriculture towards 2050.

FAO. 2013. FAOSTAT. Available from: www.fao.org/faostat/en/#data/QA

Fernández, M., J. A. Hudson, R. Korpela, and C. G. de los Reyes-Gavilán. 2015. Impact on human health of microorganisms present in fermented dairy products: an overview. Biomed

Res. Int. 2015:412714. doi:10.1155/2015/412714. Available from: http://dx.doi.org/10.1155/2015/412714

Gephart, J. A., K. F. Davis, K. A. Emery, A. M. Leach, J. N. Galloway, and M. L. Pace. 2016. The environmental cost of subsistence: Optimizing diets to minimize footprints. Sci. Total Environ. 553:120–127. doi:10.1016/j.scitotenv.2016.02.050. Available from: http://dx.doi.org/10.1016/j.scitotenv.2016.02.050

Gupta, S. K., and R. D. Deshpande. 2004. Water for India in 2050: first-order assessment of available options. Curr. Sci. 86:1216–1224. Available from: http://www.jstor.org/stable/24109929

Haber, F., and G. van Oordt. 1905. Über die Bildung von Ammoniak den Elementen. Z. Anorg. Chem. 44:341–378. doi:10.1002/zaac.19050440122.

Hess, J. M., S. S. Jonnalagadda, and J. L. Slavin. 2015. Dairy Foods: Current Evidence of their Effects on Bone, Cardiometabolic, Cognitive, and Digestive Health. Compr. Rev. Food Sci. Food Saf. 15:251–268. doi:10.1111/1541-4337.12183. Available from: http://dx.doi.org/10.1111/1541-4337.12183

Hewison, M. 2012. An update on vitamin D and human immunity. Clin. Endocrinol. . 76:315–325. doi:10.1111/j.1365-2265.2011.04261.x. Available from: http://dx.doi.org/10.1111/j.1365-2265.2011.04261.x

Hoddinott, J., D. Headey, and M. Dereje. 2013. Cows, Missing Milk Markets and Nutrition in Rural Ethiopia. SSRN Electronic Journal. doi:10.2139/ssrn.2326144. Available from: http://dx.doi.org/10.2139/ssrn.2326144

Intergovernmental Panel on Climate Change, and Intergovernmental Panel on Climate Change. Agriculture. In: Climate Change

2007. p. 497–540. Available from: http://dx.doi.org/10.1017/cbo9780511546013.012

Johnson, K. A., and D. E. Johnson. 1995. Methane emissions from cattle. J. Anim. Sci. 73:2483–2492. Available from: https://www.ncbi.nlm.nih.gov/pubmed/8567486

Littleboy, M., D. M. Smith, and M. J. Bryant. 1996. Simulation modelling to determine suitability of agricultural land. Ecol. Modell. 86:219–225. doi:10.1016/0304-3800(95)00055-0. Available from: http://dx.doi.org/10.1016/0304-3800(95)00055-0

Luedeling, E., M. Zhang, and E. H. Girvetz. 2009. Climatic changes lead to declining winter chill for fruit and nut trees in California during 1950-2099. PLoS One. 4:e6166. doi:10.1371/journal.pone.0006166. Available from: http://dx.doi.org/10.1371/journal.pone.0006166

Luo, T., Q. Yue, M. Yan, K. Cheng, and G. Pan. 2015. Carbon footprint of China's livestock system--a case study of farm survey in Sichuan province, China. J. Clean. Prod. 102:136–143. Available from: https://www.sciencedirect.com/science/article/pii/S0959652615004527

Macdiarmid, J. I., J. Kyle, G. W. Horgan, J. Loe, C. Fyfe, A. Johnstone, and G. McNeill. 2012. Sustainable diets for the future: Can we contribute to reducing greenhouse gas emissions by eating a healthy diet? Am. J. Clin. Nutr. 96:632–639. doi:10.3945/ajcn.112.038729. Available from: http://dx.doi.org/10.3945/ajcn.112.038729

María, G. A. 2006. Public perception of farm animal welfare in Spain. Livest. Sci. 103:250–256. doi:10.1016/j.livsci.2006.05.011. Available from: http://www.sciencedirect.com/science/article/pii/S1871141306001909

Marvinney, E. M. 2016. A Life Cycle Based Assessment of Energy Use, Greenhouse Gas, and Pollutant Emissions in California Nut Production. Available from: https://books.google.com/books/about/A_Life_Cycle_Based_Assessment_of_Energy.html?hl=&id=8NvSswEACAAJ

Müller-Wielsch, K., P. Huppke, and J. Gärtner. 2010. Developmental retardation due to dietary vitamin B12 deficiency during infancy. Neuropediatrics. 41. doi:10.1055/s-0030-1265572. Available from: http://dx.doi.org/10.1055/s-0030-1265572

Murphy, M. M., L. M. Barraj, L. D. Toth, L. S. Harkness, and D. R. Bolster. 2016. Daily intake of dairy products in Brazil and contributions to nutrient intakes: a cross-sectional study. Public Health Nutr. 19:393–400. doi:10.1017/S1368980015001780. Available from: http://dx.doi.org/10.1017/S1368980015001780

National Research Council. 2001. Nutrient Requirements of Dairy Cattle. National Academies Press, Washington, DC.

Nijdam, D., T. Rood, and H. Westhoek. 2012. The price of protein: Review of land use and carbon footprints from life cycle assessments of animal food products and their substitutes. Food Policy. 37:760–770. doi:10.1016/j.foodpol.2012.08.002. Available from: http://www.sciencedirect.com/science/article/pii/S0306919212000942

Parker, L. E., and J. T. Abatzoglou. 2017. Shifts in the thermal niche of almond under climate change. Clim. Change. 147:211–224. doi:10.1007/s10584-017-2118-6. Available from: http://dx.doi.org/10.1007/s10584-017-2118-6

Pimentel, D., and M. Pimentel. 2003. Sustainability of meat-based and plant-based diets and the environment. Am. J.

Clin. Nutr. 78:660S–663S. doi:10.1093/ajcn/78.3.660S. Available from: http://dx.doi.org/10.1093/ajcn/78.3.660S

Reshmidevi, T. V., T. I. Eldho, and R. Jana. 2009. A GIS-integrated fuzzy rule-based inference system for land suitability evaluation in agricultural watersheds. Agric. Syst. 101:101–109. doi:10.1016/j.agsy.2009.04.001. Available from: http://dx.doi.org/10.1016/j.agsy.2009.04.001

Rutherfurd, S. M., A. C. Fanning, B. J. Miller, and P. J. Moughan. 2015. Protein digestibility-corrected amino acid scores and digestible indispensable amino acid scores differentially describe protein quality in growing male rats. J. Nutr. 145:372-379. Available from: doi:10.3945/jn.114.195438.

Sauer, T., P. Havlík, U. A. Schneider, E. Schmid, G. Kindermann, and M. Obersteiner. 2010. Agriculture and resource availability in a changing world: The role of irrigation. Water Resour. Res. 46. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1029/2009WR007729

Senagi, K., N. Jouandeau, and P. Kamoni. 2017. Using parallel random forest classifier in predicting land suitability for crop production. Journal of Agricultural Informatics. 8. doi:10.17700/jai.2017.8.3.390. Available from: http://dx.doi.org/10.17700/jai.2017.8.3.390

Singha, C., and K. C. Swain. 2016. Land suitability evaluation criteria for agricultural crop selection: A review. Agricultural Reviews. 37. doi:10.18805/ar.v37i2.10737. Available from: http://dx.doi.org/10.18805/ar.v37i2.10737

Springmann, M., M. Clark, and W. Willett. 2018. Feedlot diet for Americans that results from a misspecified optimization algorithm. Proc. Natl. Acad. Sci. U. S. A. 115:E1704–E1705.

doi:10.1073/pnas.1721335115. Available from: http://dx.doi.org/10.1073/pnas.1721335115

Thoma, G., J. Popp, D. Shonnard, D. Nutter, M. Matlock, R. Ulrich, W. Kellogg, D.S. Kim, Z. Neiderman, N. Kemper, F. Adom, and C. East. 2013. Regional analysis of greenhouse gas emissions from USA dairy farms: A cradle to farm-gate assessment of the American dairy industry circa 2008. Int. Dairy J. 31:S29–S40. doi:10.1016/j.idairyj.2012.09.010.

Thorning, T. K., A. Raben, T. Tholstrup, S. S. Soedamah-Muthu, I. Givens, and A. Astrup. 2016. Milk and dairy products: good or bad for human health? An assessment of the totality of scientific evidence. Food Nutr. Res. 60:32527. doi:10.3402/fnr.v60.32527. Available from: http://dx.doi.org/10.3402/fnr.v60.32527

United Nations. 2017. World Population Prospects: The 2017 Revision, Key Findings and Advance Tables.

USDA. 2018a. USDA NASS QuickStats Database. Available from: https://quickstats.nass.usda.gov/

USDA. 2018b. USDA Food Composition Database. Available from: https://ndb.nal.usda.gov/ndb/search/list

USDA/ERS. 2018a. Oil Crops Yearbook. Available from: www.ers.usda.gov/data-products/oil-crops-yearbook.aspx

USDA/ERS. 2018b. Feed grains:Year book tables. Available from: ers.usda.gov/data-products/feed-grains-database/feedgrains-yearbook-tables.aspx.

US Environmental Protection Agency. 2010. Inventory of US Greenhouse Gas Emissions and Sinks. US EPA, Washington, DC.

Van Meerbeek, K., and J.-C. Svenning. 2018. Causing confusion in the debate about the transition toward a more plant-based diet. Proceedings of the National Academy of Sciences. 115:E1701–E1702. doi:10.1073/pnas.1720738115. Available from: http://dx.doi.org/10.1073/pnas.1720738115

Weaver, C. M., W. R. Proulx, and R. Heaney. 1999. Choices for achieving adequate dietary calcium with a vegetarian diet. Am. J. Clin. Nutr. 70 (suppl): 543S-548S.

Weber, C. L., and H. S. Matthews. 2008. Food-miles and the relative climate impacts of food choices in the United States. Environ. Sci. Technol. 42:3508–3513. Available from: https://www.ncbi.nlm.nih.gov/pubmed/18546681

White, R. R., and M. B. Hall. 2017. Nutritional and greenhouse gas impacts of removing animals from US agriculture. Proc. Natl. Acad. Sci. U. S. A. 201707322. doi:10.1073/pnas.1707322114. Available from: http://www.pnas.org/content/early/2017/11/07/1707322114

White, R. R., and M. B. Hall. 2018. Reply to Van Meerbeek and Svenning, Emery, and Springmann et al.: Clarifying assumptions and objectives in evaluating effects of food system shifts on human diets. Proceedings of the National Academy of Sciences. 115:E1706–E1708. doi:10.1073/pnas.1720895115. Available from: http://dx.doi.org/10.1073/pnas.1720895115

Willett, W., J. Rockström, B. Loken, M. Springmann, T. Lang, S. Vermeulen, T. Garnett, D. Tilman, F. DeClerck, A. Wood, M. Jonell, M. Clark, L. J. Gordon, J. Fanzo, C. Hawkes, R. Zurayk, J. A. Rivera, W. De Vries, L. Majele Sibanda, A. Afshin, A. Chaudhary, M. Herrero, R. Agustina, F. Branca, A. Lartey, S. Fan, B. Crona, E. Fox, V. Bignet, M. Troell, T. Lindahl, S. Singh, S. E. Cornell, K. Srinath Reddy, S. Narain, S. Nishtar, and C. J. L. Murray. 2019. Food in the Anthropocene: the EAT-Lancet Commission on healthy diets from sustainable food systems. Lancet. 393:447–492. doi:10.1016/S0140-6736(18)31788-4. Available from: http://dx.doi.org/10.1016/S0140-6736(18)31788-4

Wilson, N., N. Nghiem, C. Ni Mhurchu, H. Eyles, M. G. Baker, and T. Blakely. 2013. Foods and dietary patterns that are healthy, low-cost, and environmentally sustainable: a case study of optimization modeling for New Zealand. PLoS One. 8:e59648. doi:10.1371/journal.pone.0059648. Available from: http://dx.doi.org/10.1371/journal.pone.0059648

Wokes, F., J. Badenoch, and H. M. Sinclair. 1955. Human Dietary Deficiency of Vitamin B12. Am. J. Clin. Nutr. 3:375–382. doi:10.1093/ajcn/3.5.375. Available from: http://dx.doi.org/10.1093/ajcn/3.5.375

Wolz, K. J., and E. H. DeLucia. 2019. Black walnut alley cropping is economically competitive with row crops in the Midwest USA. Ecol. Appl. 29:e01829. doi:10.1002/eap.1829. Available from: http://dx.doi.org/10.1002/eap.1829

Zhu, H., and K. Kannan. 2018. Continuing Occurrence of Melamine and Its Derivatives in Infant Formula and Dairy Products from the United States: Implications for Environmental Sources. Environmental Science & Technology Letters. 5:641–648. doi:10.1021/acs.estlett.8b00515. Available from: http://dx.doi.org/10.1021/acs.estlett.8b00515

Tables

Table 3-1. Comparisons of all scenarios and land use allocation options used in study.

		Change from Baseline Land Use, %[1]							
Category	Baseline Land Use (Ha)	Current Land Use, No Re-Allocation	Current Land Use, Re-Allocation	Fruits and Veg Land Use, No Re-Allocation	Fruits and Veg Land Use, Re-Allocation	Nuts and Pulses Land Use, No Re-Allocation	Nuts and Pulses Land Use, Re-Allocation	Non-grain land use, No Re-Allocation	Non-grain land use, Re-Allocation
Fruit	1181058	2.8%	6.3%	119.7%	261.0%	0.0%	0.0%	8.6%	18.7%
Grain	69749095	2.8%	0.7%	0.0%	-5.3%	0.0%	-5.3%	0.0%	-5.3%
Legume	32416697	2.8%	6.3%	0.0%	0.0%	9.2%	20.1%	8.6%	18.7%
Nut	1205757	2.8%	6.3%	0.0%	0.0%	9.2%	20.1%	8.6%	18.7%
Oil	4849841	2.8%	6.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sugar	832723	2.8%	6.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Vegetable	1410691	2.8%	6.3%	119.7%	261.0%	0.0%	0.0%	8.6%	18.7%
Hay	19578913	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Silage	3101221	-100.0%	-100.0%	-100.0%	-100.0%	-100.0%	-100.0%	-100.0%	-100.0%
Croppable pasture	57278	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

[1]Change is the percentage increase or decrease of hectares used in the alternative land use allocation compared to the baseline.

Table 3-2. Percentage change in greenhouse gas (GHG) emissions of four select scenarios on the basis of U.S. dairy GHG, total U.S. agricultural GHG, and total U.S. GHG.

	Depopulated, Current Land Use, No Grain Land Re-Allocation	Depopulated, Current Land Use, Grain Land Re-Allocation	Retirement - Current Management, No Land Reallocation, No Grain Land Reallocation	Retirement - To Pasture, No Land Use Change, No Grain Land Re-Allocation
U.S. Dairy GHG, %	-42.98%	-31.93%	0.00%	-74.84%
U.S. Agricultural GHG, %	-6.88%	-5.11%	0.00%	-11.97%
U.S. Total GHG, %	-0.58%	-0.43%	0.00%	-1.01%

Figures

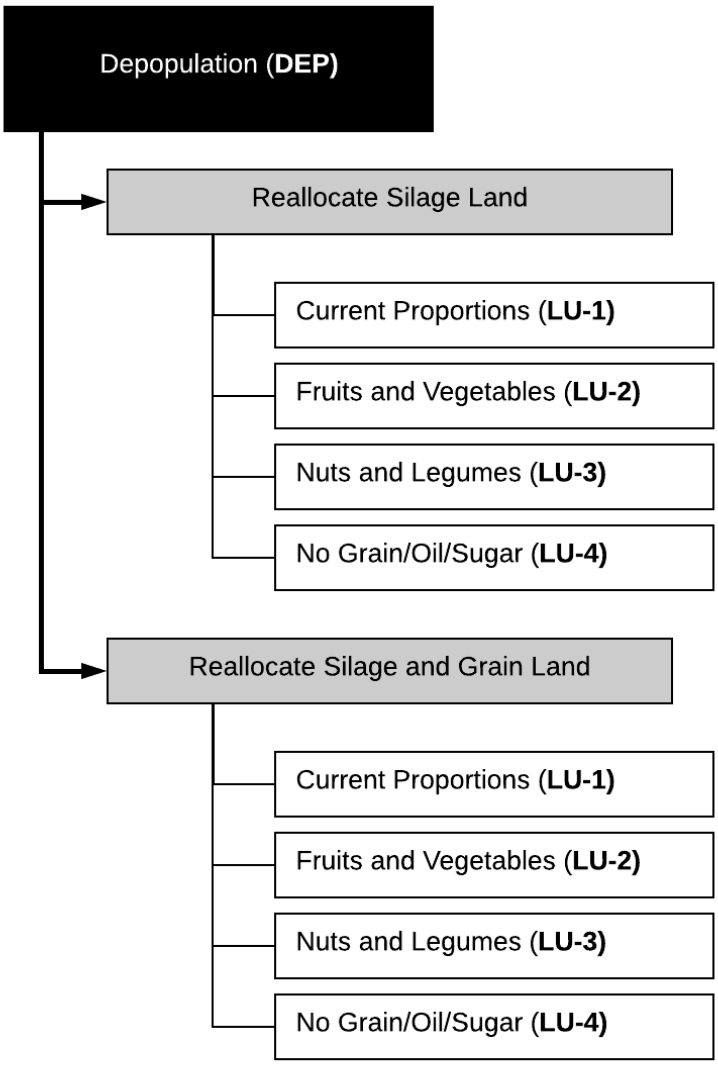

Figure 3-1. Description of the depopulation scenario describing how dairy cattle would be removed from U.S. agriculture. Two different land reallocation options were used: re-allocate only silage land previously used for dairy or additionally take grain land previously used for dairy out of production. The four uses for reallocated land are shown as LU-1 through LU4: allocate new land based on current land use, based on increasing production of fruits and vegetables, based on increasing production of nuts and legumes, or based on increasing production of all products except grains, oils and sugar.

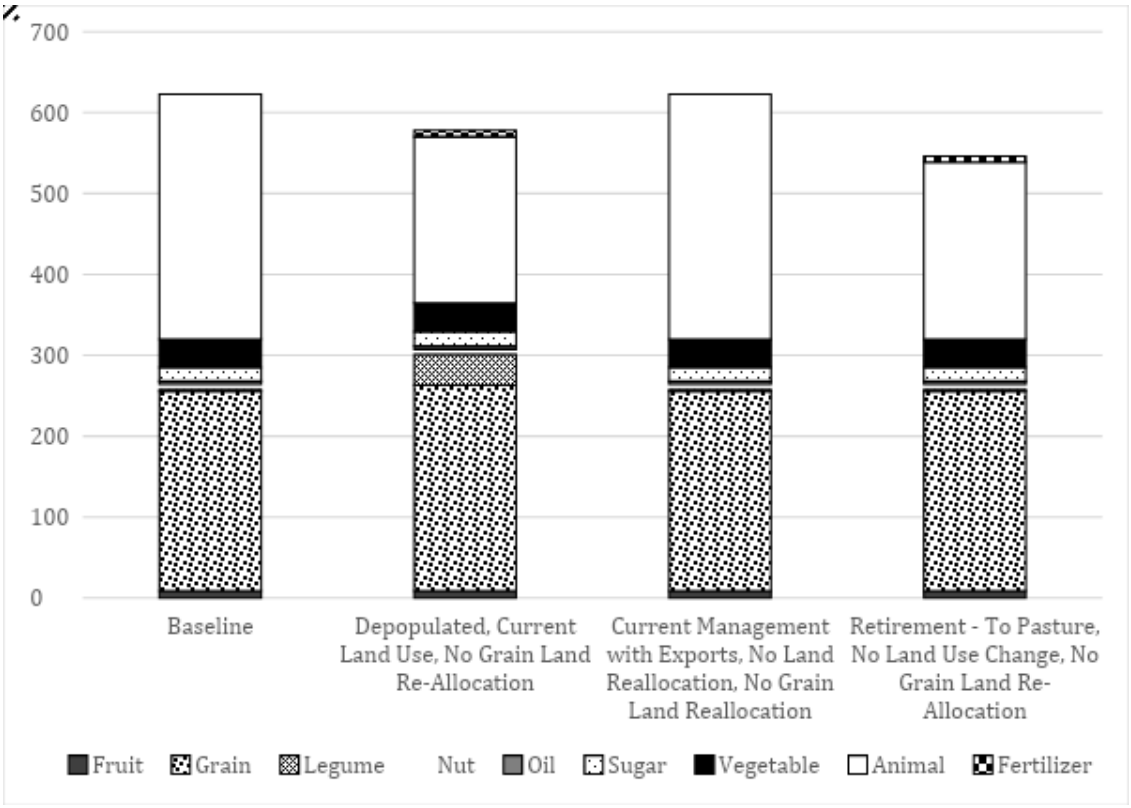

Figure 3-2. Comparison of GHG emissions from agriculture in kg CO_2 equivalents between Baseline - the current production system; Depopulation - all dairy animals are removed; Current Management with Exports - a scenario where animals are kept under current management and dairy products are not consumed in the U.S.; and Retirement - dairy animals are retired to a pasture-based system.

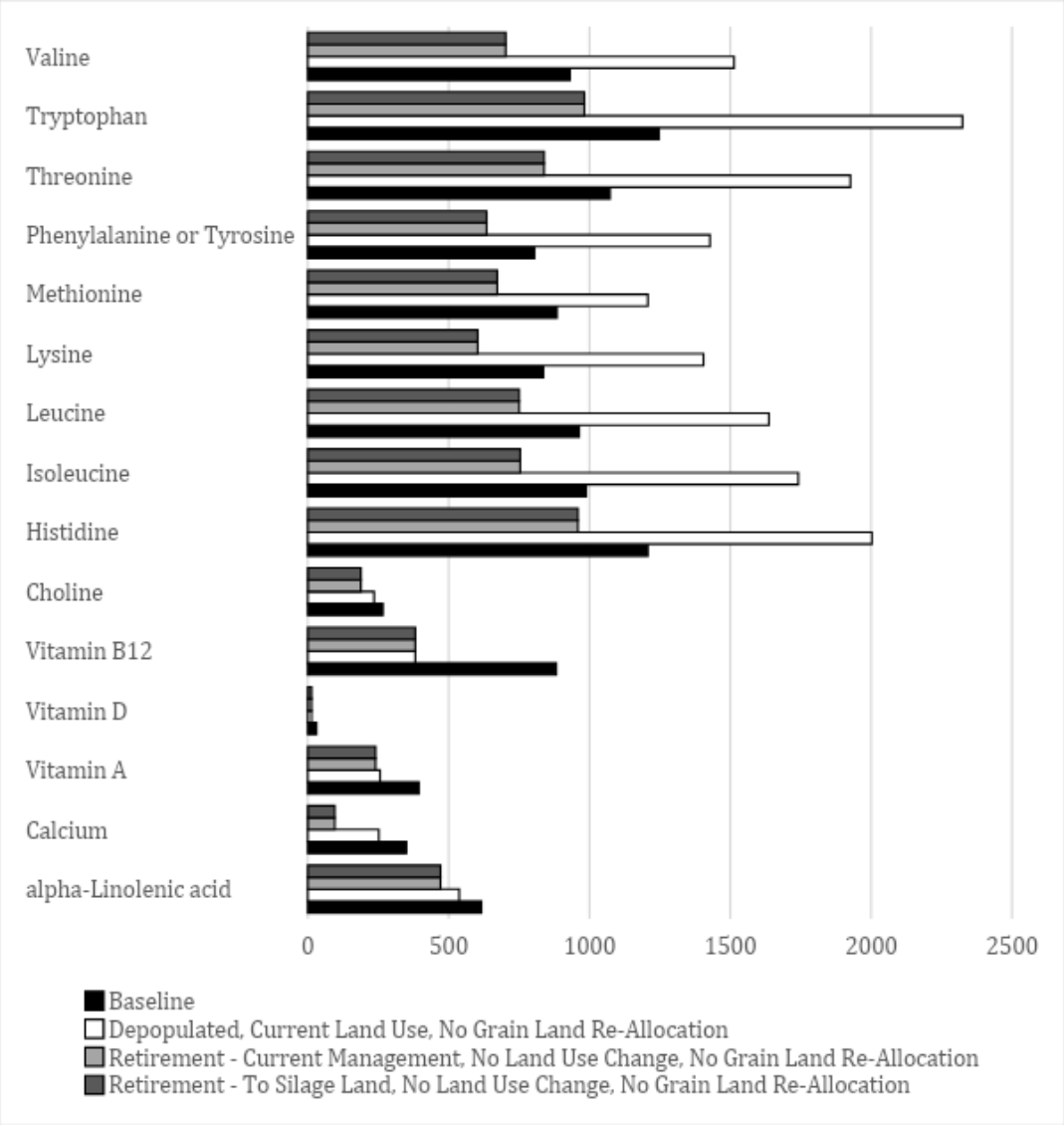

Figure 3-3. Nutrient supply of current production compared to that of three dairy cow removal strategies in terms of human nutrient requirement years (HNRY) met, in millions.

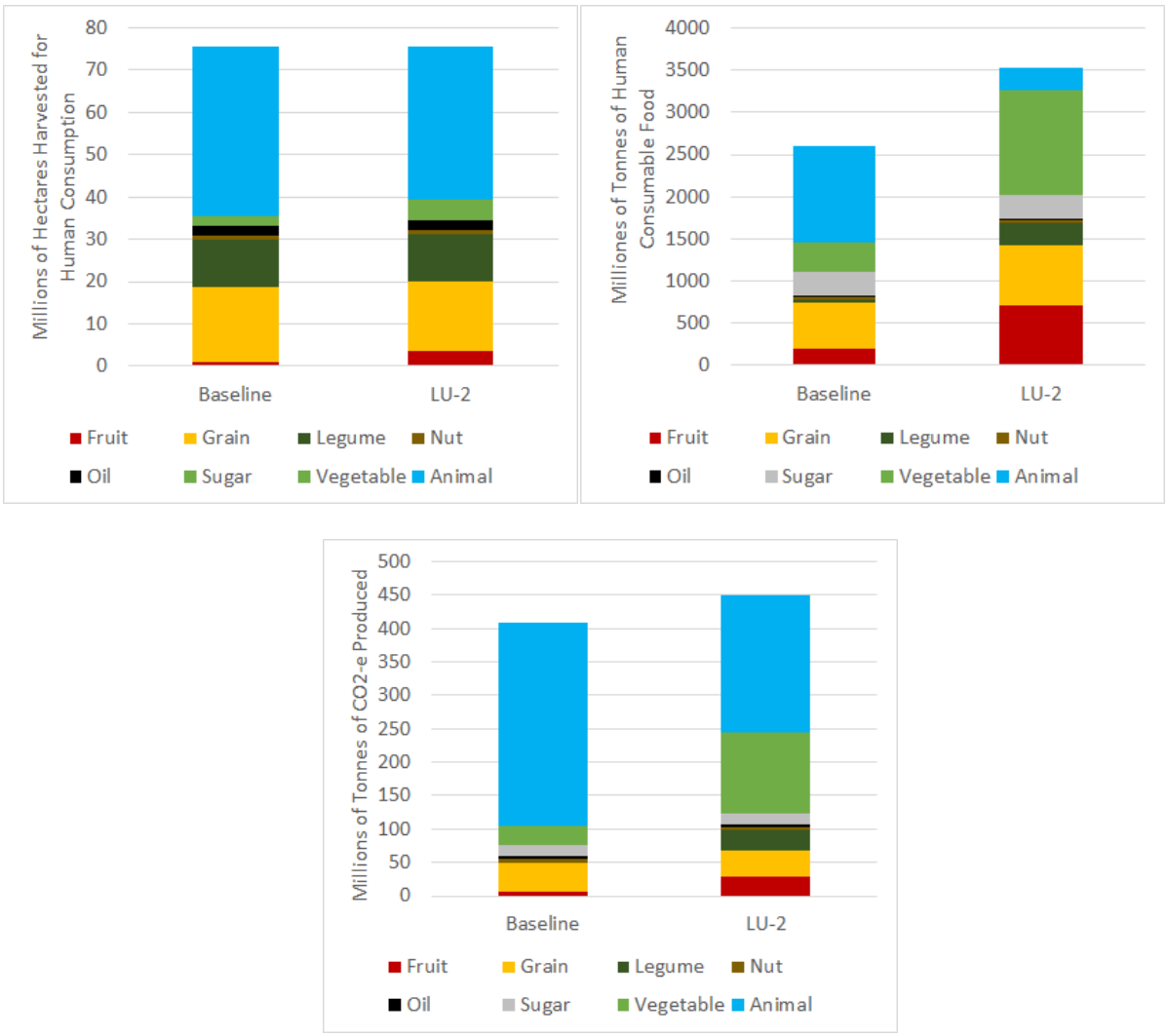

Figure 3-4. Comparison of land use classifications under current conditions (Baseline) and a scenario where grain land previously used for dairy cattle feed is repurposed for fruit and vegetable production (LU-2). Although land use shifts involve only 6.8 million ha of land, the shifts in consumable food produced and in carbon emissions produced are substantial because of the high yields of fruit and vegetable products per unit of land area.

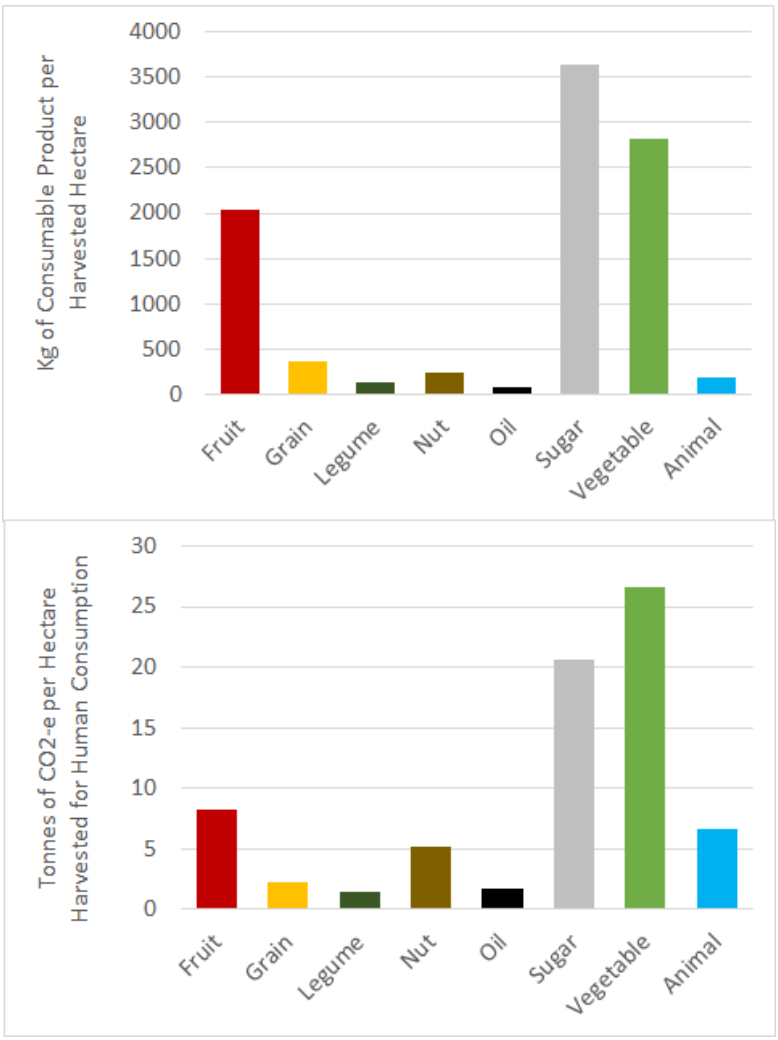

Figure 3-5. Consumable product and tonnes of CO_2-equivilent (CO_2-e), broken down by food category, of a land reallocation strategy where grain land previously used for dairy cattle feed is repurposed for fruit and vegetable production. Fruits and vegetables account for an increased amount of consumable product, but also disproportionately increase total GHG costs.

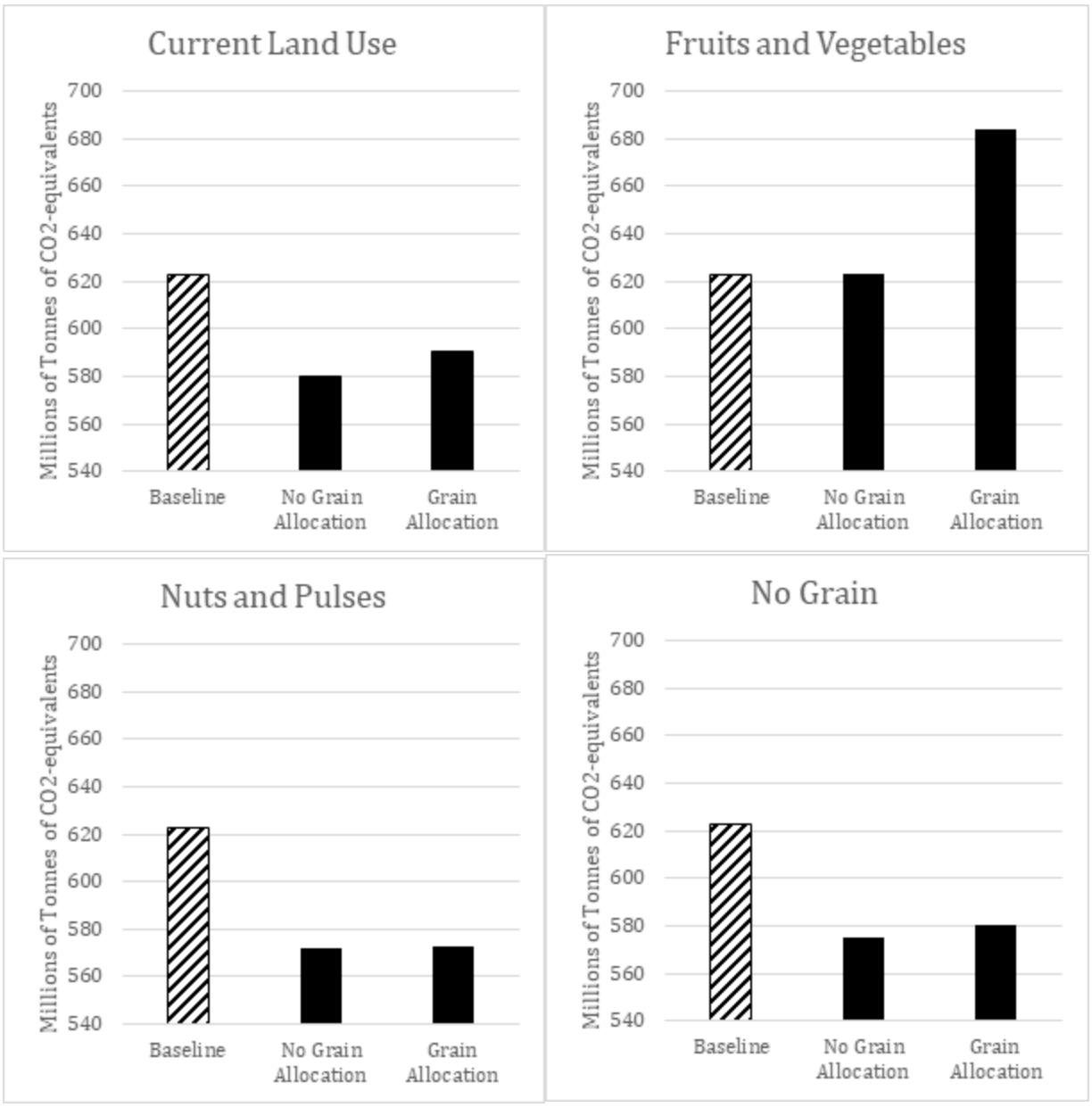

Figure 3-6. Comparison to current production system (baseline) of total greenhouse gas emissions under land reallocation scenarios using various crop replacement methods. (A) Replant vacated land with crops in proportion to their current production in the U.S. (B) Replant land with only fruits and vegetables (C) Replant land with only nuts and pulses (D) Replant land in proportion to current U.S. crop production, without planting any additional grain. No Grain Allocation = only reallocate land directly freed from removal of dairy animals or used for silage. Grain Allocation = reallocate land directly freed from removal of dairy animals, used for silage, and used for dairy feed grain.

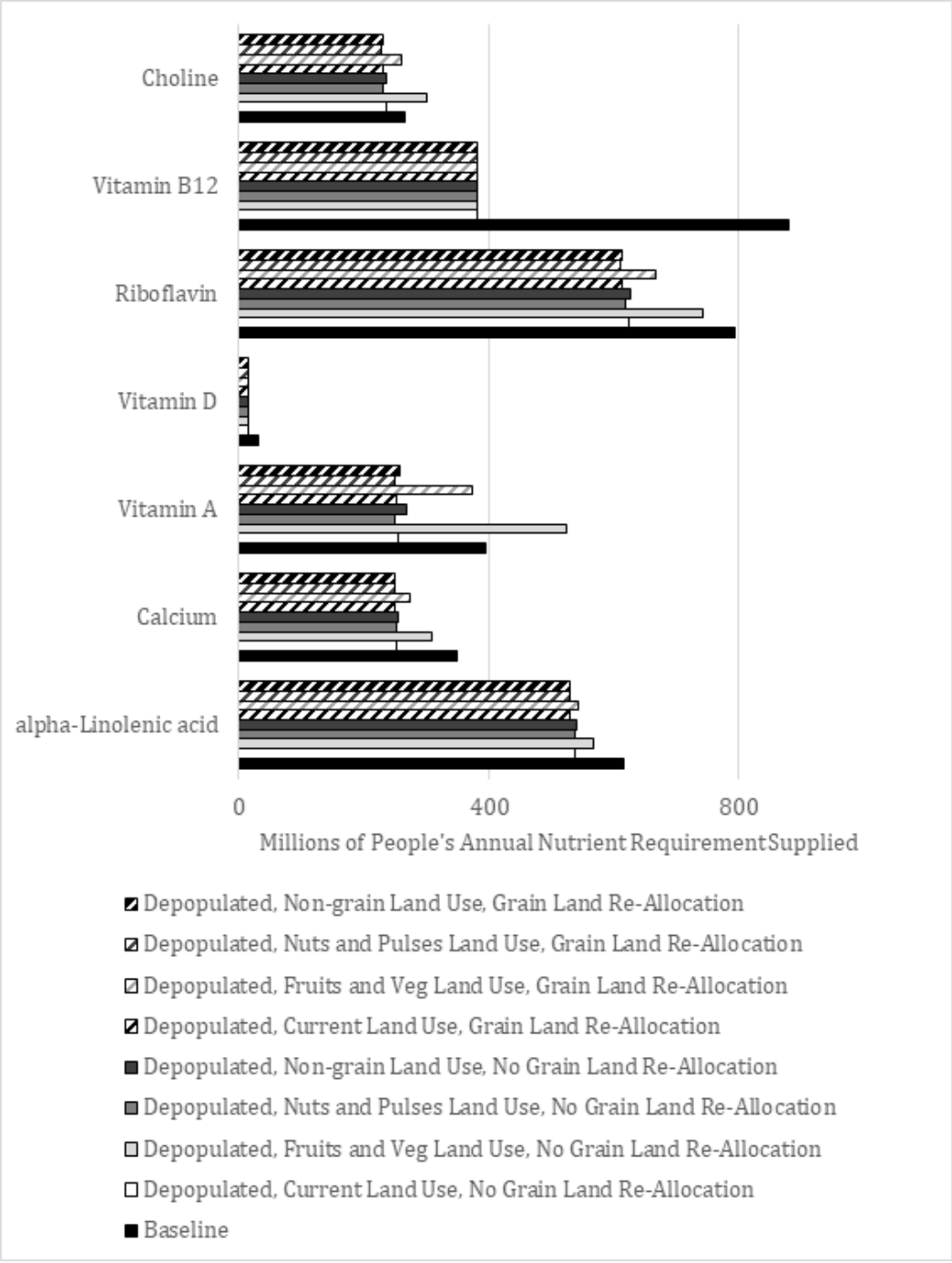

Figure 3-7. Comparison of nutrient supplies among land use options in a scenario where dairy cattle are depopulated from the agricultural system. Nutrients shown are those that had reduced supply in one or more land use option compared with the baseline scenario.

Chapter 4: Analytics in Sustainable Precision Animal Nutrition

Published as: Liebe, D.M., and R.R. White. 2019. Analytics in sustainable precision animal nutrition. Anim Fron 9:16–24. doi:10.1093/af/vfz003.

Implications

The global population, resource, and climate dynamics suggest we must improve sustainability of food production systems; precision feeding of livestock may be one way to accomplish this goal.

Analytics for precision management can be classified according to four levels: I) technique, II) data interpretation, III) integration of information, and IV) decision making. Most current animal agricultural analytics fall under categories I and II. Moving toward analytics that address integration of information and decision making is of critical importance.

Data analytical techniques such as linear modeling and machine learning provide unique and important tools for interpreting data obtained from on-farm sensors. These techniques each apply to the different levels of precision management classification.

Assessing adequacy and performance of analytics tools must, by default, depend on the objective of those tools and the type of response considered. As more advanced level III and IV systems are developed, integration of expert opinion into analytics may be essential to optimize performance and relevance on-farm.

Keywords Machine Learning, Data Mining, Computer Vision, Internet of Things

Main Text

How will precision nutrition influence sustainability?

The global population, resource, and climate dynamics suggest we must improve sustainability of food production systems (Ohlsson, 2014; Kleinman et al., 2018). Improving livestock production sustainability is particularly important because a significant portion of the projected increases in global food demand is anticipated to come from livestock (Thornton, 2010). Improving sustainability of livestock production systems can be achieved through optimized reproductive, genetic, nutritional, and health management (White et al., 2014; White et al., 2015). Management decisions within livestock production can be thought of as two interleaved feedback loops. The first feedback loop is between the animal and the environment: the animal is influenced by its environment and, in turn, influences its environment. The second feedback loop is between the animal and the manager: the manager takes information about the animals behavior and attempts to influence the environment to optimize the animals performance. Managers make management decisions on different timescales ranging from immediate to relaxed. An example of an immediate management decision would be a farmer identifying an animal as sick, isolating the animal, and treating the animal for the illness. We term this immediate because the farmer must identify the sick animal as soon

as possible and must react to the diagnosis as soon as possible. An example of a relaxed management decision would be the farmer electing to change the feed provided to his animals in response to something observed about their production (i.e., the cows are producing poorly, so change the ration to provide higher nutrient density to correct a nutrient shortfall). This decision is more relaxed because its formulation and response are subjected to natural biological delays (i.e., it may take days to weeks to see a production response to a new diet). Improving the precision of these decision-making processes and reducing the burden of decision making on farmers are two critical steps toward improving sustainability of livestock production. Precision agricultural technologies have been identified as one possible solution (Berckmans, 2014; Tullo et al., 2019).

Precision field crop agriculture has dramatically expanded and industrialized over the last several decades, demonstrating substantial opportunity for using precision technologies in agriculture (Thorp and Tian, 2004; Nash et al., 2009; Zhang and Kovacs, 2012). Such technologies include global positioning system (GPS) guided equipment, unmanned aerial vehicles, robotic harvesting and monitoring equipment, automated application of agrochemicals, among many others. Precision animal agriculture, on the other hand, has had limited expansion. Although technologies such as temperature monitors, rumen sensors, robotic milkers, and

others exist, the uptake and industrialization of precision animal agriculture has not paralleled crop agriculture. There are several differences between crop and livestock management that may contribute to this difference in technology uptake. For example, the management time scales for crop agriculture interventions, while highly profitable, are often measured in days or weeks. In animal agriculture, timescales for certain management can range from hours to days. For issues of nutrition, health, productivity, and efficiency, animal agriculture must treat both the individuals and the collective, whereas crop agriculture focuses primarily on the field-scale. Animal losses are also perceived differently than crop losses due to moral concerns, possibly imposing higher standards on animal-based decision technology. Collectively, these challenges mean that animal agriculture will likely require different types of technological interventions than have been pioneered in crop systems. Exploring opportunities for where precision technologies may be relevant in the livestock nutrition space exemplifies this.

Management applications for precision animal nutrition

Optimizing Rumen Fermentation.

The idea that fermentation can be optimized if degradable carbohydrate sources and degradable protein sources are properly

matched has been contemplated for decades (Sinclair, 1995; Sinclair et al., 1995). The theory behind optimizing nutrient synchrony suggests that fermentations will be optimized if they are never limited by energy or N (i.e., supplies are balanced). Despite this theory being sound, achieving nutrient synchrony within rumen fermentations is extremely difficult to accomplish with currently available technologies (Hall and Huntington, 2008). One potential reason for this challenge is the limited real-time data available on the fermentation environment. Several models attempt to account for nutrient degradation kinetics (Hanigan et al., 2013; Higgs et al., 2015; Van Amburgh et al., 2015; Li et al., 2018), however, obtaining data to construct and evaluate models of degradation kinetics *in vivo* often requires expensive experiments. The advent of technologies such as indwelling rumen sensors have enabled more precise understanding of how pH changes over the course of a day. Expanding these sensors to include recording other important metabolites could enable development of feeding recommendations that take fermentation profile into account more precisely.

Detection of Metabolic Diseases.

It is possible to use analytics to identify risk of metabolic diseases. Existing efforts to identify other disease states (e.g.,

mastitis) have shown moderate promise. Although mastitis is extremely costly to the dairy industry, it can be difficult to predict due to the imbalance of positive results (disease cases) relative to the population. The incidence rate of clinical mastitis ranges depending on many factors like housing or location, but is near 15 cases per 100 cow lactations, or 1 case per 2033 cow days, assuming a 305 day lactation (McDougall et al., 2007). Put another way, a priori, a randomly selected lactating cow from a random herd is only approximately 0.05% likely to exhibit clinical mastitis. Sparse datasets, the analytical term for the issue of having a disproportionate amount of positive test cases in a dataset, is a common problem in present-day analytics (Han et al., 2015; Greenland et al., 2016). However, due to the widespread nature of the issue, new analytical techniques like modified tree-based algorithms can learn patterns while maintaining the underlying proportion of cases in the training data (Ushikubo et al., 2017). Alternatively, the collation of larger datasets is also advantageous for producing better metabolic disease predictions. There is a tendency to collect new data to train new models, but in cases with sparse data, the combination of past data and new data will lead to richer training sets. Consider that each additional positive training case will greatly improve accuracy compared to each new negative case. In fact, removing additional negative cases to artificially improve the proportion

of positive cases can help to train models. The caveat to training on stratified datasets is that they must be properly validated on datasets with the appropriate proportion of positive cases to determine real-world use. By utilizing strategies designed for the problem of sparse data in machine learning, predicting metabolic disease will become easier, and most importantly, more accurate, providing decreased false-positives.

Response-based Nutrient Requirement Recommendations.

A major limitation of existing nutrient requirement systems like the National Research Council Requirements for dairy cattle (National Research Council et al., 2001) is the requirement-based nature of the recommendations. Maximizing production mass is often not the same as optimizing production efficiency. Multicriteria optimization has previously been used to formulate rations to simultaneously achieve multiple environmental goals (White et al., 2014; White et al., 2015). Optimizing productivity or economic parameters could also be accomplished with this technique if the underlying equations linked dietary inputs with productive outputs in a responsive way. A challenge with response-based nutrient requirements systems is that most of our current data that could be used to develop such a system relies on pen-fed cattle. Responses of individuals are likely unique and such a response-

based model would be more useful if feeding systems and nutrition models did a better job of representing the individual, rather than the collective.

Precision Nutrition Research.

In a wide variety of ruminant nutrition research, access to the rumen is obtained through rumen cannulae; however, sampling through this orifice is physically difficult and often results in mixing of naturally stratified (vertical and horizontal) rumen contents. The physical difficulty in sampling the rumen can impede precision monitoring of difficult-to-reach areas. Additionally, disrupting the rumen environment through sampling physically or chemically alters the unique microclimates that are thought to exist within the rumen, and thus precluding accurate and representative sampling. Collectively, these challenges make accessing unique microclimates within the rumen a challenge. The availability of a platform that can monitor rumen sensors would be valuable to the study of these unique rumen microclimates.

What limitations exist for current technologies?

Rutten et al. (2013) summarized 126 publications describing 139 dairy sensor systems from the period 2002 to 2012. The systems were then compared based on the four levels of: I) technique, II)

data interpretation, III) integration of information, and IV) decision making. Systems that accomplish all 4 of these levels are often referred to as cyberphysical systems (CPS). These CPS are often an automated network of sensors, networking technologies, analytics, and actuation technologies that work in combination with or independent of the farmer to affect management changes based on real-time sensed information on-farm. None of the 139 sensor systems evaluated by Rutten et al. (2013) included integration of sensed metrics with other information available on the farm to produce management advice or automated decision making (Rutten et al., 2013). Most sensor systems that were used in the farmer's decision process only provided the raw data measured by the sensor, or a probability (such as the probability of disease given the sensor data). In both cases, the farmer is left to their intuition to integrate and actually make a management decision.

Although basic linear models or logit models produce predictions that are correct on average over a group, these models cannot account for increased variation in individuals. The models being used to interpret data, as referenced in level II of Rutten et al. (2013) can be slow or fail to converge on a suggested action under the complexity of decision-making. For example, although there may be a manageable number of factors that affect the prediction of ketosis, the number of factors affecting the costs and benefits of the treatment of said ketosis is surely greater.

Put another way, knowing that a cow is 35% ± 2 likely to be ketotic tomorrow does not say anything about whether the farmer should check the cow, treat the cow, cull her, or do something else. To properly assess the promise of analytics in creating CPS capable of filling all four levels of the Rutten et al. (2013) summary of agricultural systems, we will present a common precision nutrition aim: automated individualized feeding of dairy cows. Using this example objective, we highlight several possible alternative analytical approaches and discuss their strengths and potential pitfalls relevant to this objective.

A Nutrition Analytics Example: Automated Individual Feeding

Automated Individualized Feeding.

Given the variation among individual animals, it is reasonable to assume that by using data specific to each animal, we can make better decisions on what, and how much, to feed. As we have previously noted, model-based feeding can optimize productivity for the whole farm because individuals likely have differing and unique requirements. Individual feeding requires the ability to collect data specific to each animal coupled with the analytics capable of estimating individual requirements from that data. Feeding individuals eliminates the need to over-feed some animals to avoid under-feeding others, leading to more targeted feeding practices. One does not necessarily need to feed each

animal individually; this same reduction in over/under feeding can be accomplished simply by reducing the variation in the feeding group, either by feeding more like-animals together or by feeding animals in small groups. An example of variance reduction through smaller groupings of animals would be the use of different feeding groups by lactation number in dairy cows. It is clear that nutrient requirements are vastly different for first and fourth lactation cows, so they are separated to reduce the feed requirement variance. Another more targeted example of individualized feeding is concentrate supplement feeding. A larger group of animals can receive the same basal diet and the supplement is provided separately to smaller groups (Dela Rue and Eastwood, 2017). However, this type of individualized feeding, as noted by Dela Rue and Eastwood (2017), has not been shown to provide marginal benefits to farmers. Multiple recent studies which suggested individualized supplement feeding saw no improvement in milk production, BCS, or BW (Lawrence et al., 2015; Dale et al., 2016; Little et al., 2016).

Although it seems intuitive that more individualized feeding regimens would lead to better performance, this is not always what occurs in practice. These limitations may be because of the aforementioned issues with requirement models, which are based on data from groups of animals, not individuals. Another limitation might be the complexity of analytics used for feeding

recommendations. Of the three citations above that showed no increase in performance on individualized concentrate feeding, all studies used only one variable (milk yield) to inform concentrate requirement. In one study, only two levels of concentrate based on milk yield were fed, and a linear multiplier of milk yield was used in the other two studies to determine concentrate. Such low-dimensionality models, using only one variable to predict a response, limits the robustness of the predictions and results. We will examine potentials of higher-level modeling approaches by examining the current infrastructure to support CPS in the four levels described by Rutten and colleagues.

Current CPS Infrastructure.

Level I, the techniques for data collection, is comprised of technologies like radio frequency identification (RFID) tags, accelerometers, and other output measurement software like inline milking parlor sensors. We can use this data that is collected daily, or even in real-time, to broadly evaluate the performance of animals. One of the issues with the techniques of collecting raw data is the interpretation. With only raw data, it is hard to determine the cause-effect relationship between feeding and performance. For example, the fact that the daily step count of an animal has increased on a new diet does not inform the farmer

whether or not to continue feeding this diet or what needs to be changed. Rather, raw data must be interpreted before it can be used effectively to make diet decisions. Level II, or the interpretation of sensor data, seeks to add context to sensor data with emphasis on explaining such relationships. Many models attempt to predict intake requirements of dairy cows using raw data as predictors (Jensen et al., 2015). Jensen et al. (2015) evaluated models that were used on a national scale in different countries. All models were fit to held-out intake data to determine the residual error in each prediction model. The root mean square prediction error for each model ranged between 1.2 kg dry matter (DM) per day to 3.2 kg DM per day (Jensen et al., 2015). The held-out data included 94 treatment means derived from 917 lactating dairy cows. A given model's average prediction was near 2.0 kg of DM greater or less than a cow's average intake. If these results were applied to individual cow days, the variance would necessarily be greater than the variance in predictions for a cow's average intake. Models predicting DMI can be simple, lending themselves to being correct on average, which is not as useful in individualized feeding because response variance increases.

In a review of linear models predicting DMI (Jensen et al., 2015), models referred to as "advanced" were those that incorporated interaction terms into the linear model, specifically

the models "TDMI" and "NorFor" (Huhtanen et al., 2011; Volden et al., 2011). Many recent publications involve predicting intake using less than 10 total predictor variables and rely on basic linear regression (McParland et al., 2014; de Haas et al., 2015; Shetty et al., 2017; White et al., 2017). Most models attempt to find the few variables that will reduce the variance better than previous models. At some point, we will not be able to find a selection of 10 or fewer variables that continue to reduce variance in a meaningful way. One advance in data analytics is hierarchical modeling, which works well in the case where there are many models using varying parameters to predict the same response. Making a "model of models" can improve accuracy beyond that of any one model in the group (Gelman, 2006). This is possible due to uncorrelated error structures in different sub-models. To create an example hierarchical model for predicting DMI in dairy cows, we could combine the outputs of models built on herd level data into models built on models using different individual cow measurements to make a more accurate prediction of individual DMI than using a single model alone. Although hierarchical modeling is just a framework, there are many useful ways to combine existing models that can improve model accuracy. Models can be weighted based on accuracy in a test dataset, the variance of predictions, or even on prior knowledge.

With over 9 million dairy cows in the United States, it intuitively seems easy to collect sufficient data to predict intake; however, this is not necessarily the case (McParland et al., 2014). First, many data sources must be collated to create better-trained models. There are incentives now for farmers to continue to collect individual intake data and genetic data relating to intake to help inform farmers in the future (Berry et al., 2014). An estimated 89% of genetic variation in DMI could be explained with only four common animal characteristics, according to one meta-analysis of genetic studies (Berry and Crowley, 2013). Although we have great amounts of data, there are near-infinite permutations of cow characteristics that would need to be predicted in order to improve DMI prediction. Luckily, data analytics offers a way to reduce the dimensionality of problems and also group similar animals together to make the prediction space more manageable. Principal components analysis (PCA) attempts to reduce dimensionality while maintaining maximal variance in the remaining dimensions using an orthogonal transformation (Pearson, 1901). Consider a 3-dimensional set of data, shown in Figure 4-1. If we know the groupings ahead of time, we can find two linear descriptors using all three factors that maximizes variance in the data set. By using all three factors (in this example, genes), but condensing the desciprtors into two values for each point, we have reduced the dimensionality at minimal varaince cost between

groups. This is evident in the image on the right in Figure 4-1. Using PCA can also help discern groups, as PCA is sensitive to scale changes and can be used to determine the distance between two multi-dimensional points in space. Traditionally, a machine learning technique like k-nearest neighbors (Altman, 1992) or k-means (Lloyd, 1982) is used to determine the similarity between points. In our example with a herd of cows that we need to predict and feed individually, a linear model trained on the entire herd will only be right on average. If we do not have sufficient data to make low-variance predictions for individual cows, we could employ PCA on the individual cow data to determine cows that are most similar, combine their data and train models on these smaller combined datasets of similar cows to achieve more accurate results. By using a fixed modeling procedure and measure of accuracy, we could iteratively test models using data from smaller groups until we no longer saw an improvement in accuracy. Consider the scenario outlined in Figure 4-2 which explains the framework for using PCA to find the optimal groupings for a given model.

It is important to note that although 2-dimensional PCA is easiest to visualize, PCA results should be retained in the number of dimensions that explains a specified amount of variance. Figure 4-3 shows a plot of the variance explained as the number of dimensions included in PCA is increased. With fewer dimensions there is less variance explained by the components and the

proportion of variance explained by each additional component is high. As we increase dimensions, the cumulative variance explained increases but the proportion of variance explained by each additional component decreases. Humans tend to interpret best in two dimensions, but we can see that if we wanted our PCA to explain at least 80% of the variance in our dataset, 2 dimensions would not be sufficient. Also keep in mind that not all datasets will produce such steady reductions in variance with each component. There is no rule of thumb for how many components to condense. With PCA, and many algorithms in data analytics, we must trade-off interpretability for accuracy.

Opportunities to leverage machine learning in precision livestock nutrition

In level III, integration of information, the predictions made by models are used to created recommendations for the farmer. Level IV is the culmination of the prediction, leading to action, either by the system itself or the farmer. A lack of level III and IV CPS was noted in Rutten et al. (2013). We would expect that, by utilizing the most appropriate modeling techniques to generate predictions at levels I and II, appropriate decision-making models would be possible. However, this is obviously not the case, as we see minimal examples of decision making algorithms present in the current animal nutrition literature. One factor that traditional modeling frameworks do not allow for is the ability to update based

on feedback. If a level II model predicts DMI at 50 kg, but the farmer continuously adjusts this to 45 kg, based on his/her knowledge of something outside the model scope, a traditional model does not "learn." Here neural networks, and other recurrent machine learning algorithms provide a promising approach to decision-making frameworks by allowing for revising predictions in practice. In a traditional individualized feeding modeling framework, a model is built for each cow and the model itself does not change, only the predictions. In a machine learning framework, the predicted DMI for a cow each day could be predicted and, using all data available along with the actual response of the animal, the algorithm may change the weights of certain factors in the model. This dynamic feedback loop allows the model to "learn" on-farm and produce more accurate predictions.

Neural networks, or artificial neural networks, are actually a combination of many algorithms in a network, where layers of nodes, representing algorithms, feed outputs from the previous layer of nodes as inputs to the next layer, until the final layer's output is used as the prediction (McCulloch and Pitts, 1943). Figure 4-4 shows a typical framework for neural network, with raw information being fed into the left and predictions coming from the right. Nodes each represent a nondescript function, typically those that make small changes to inputs, allowing for better control at each node over the final prediction. The real power for

a problem with the complexity of individualized feeding is the idea of backpropagation, where the accuracy of prediction is back-propagated through the nodes of a network to re-weight the importance of each node, thereby ensuring better accuracy on the same example datum if presented again (Werbos, 1974). Put simply, backpropagation allows us to distribute error through the existing network. Neural networks have been shown to detect patterns in highly nonlinear data, which is nearly-impossible for linear models (Fukushima, 1980).

Reinforcement learning is another key concept in the field of machine learning and is crucial for problems where cost functions are not explicit, like in predicting feed intake. That is, we do not know the exact cost of overfeeding or underfeeding. Suppose we are training a model to tell a farmer how to feed each cow, but the farmer is well-informed and keeps adjusting the predictions. If we were trying to minimize the need for farmer intervention, our feedback loop would weight errors based on the farmer's adjustment to each prediction. That is to say the recurrent neural network is estimating the model that limits error under the unknown cost function. The framework starts with substantial uncertainty about the cost function and the network performs poorly; then, the network is trained and the model parameterized to decrease the cumulative costs. This is done in an updating manner called a Markov decision process (Howard, 1960). In the real world, our

farmers are likely not omniscient, but the ability to estimate models under cost uncertainty can still be utilized to choose better models for actual decision making, because the cost of feeding decisions is not fixed or known, but predictions must be made every day for every cow. In fact, reinforcement models are seen in many places where decisions must be made, despite uncertainty about their costs, like game-playing algorithms and resource allocation problems (Damas et al., 2000).

Making prediction under uncertainty can make modeling more difficult and is surely a reason why reliable level III and IV CPS are not seen in animal agriculture. For example, a model built to predict crop demand would have a large amount of training data, because these data have been collected for many years. But how will a model predict the appropriate desire for crops during a global pandemic? Many predictions made in animal agriculture are based on rare scenarios relative to the amount of data collected. Assume a scenario where predictions for a cow's intake have been very accurate, then she gets her foot caught in the parlor and is in a great deal of pain, the injury is not caught immediately and will not be fed into the model as an explicit variable. Is it correct to punish the model for incorrectly predicting intake on this day? Likely not, because a known, but unanticipated, event can explain the variation. This example points to a major challenge with deploying these modeling techniques on-farm. If allowed to

iterate and update in an unrestricted manner, the model will try to assign weights to other factors to explain why the cow reduced intake the day she injured herself. For example, if activity data were included in the model, the weight on activity responses might be updated because we would anticipate activity to also change with the injured hoof. However, the model may take some time to recover from this prediction to correct the weight on activity under a non-injured scenario, resulting in a period of time where predictions were poor. A solution to this type of challenge would be to include an injury variable in the model to account for these types of cases; however, the point of the example is that there is always opportunity for factors exogenous to the model to influence the behavior of the response variable. When building and deploying these analytics, we must consider that reality. Another solution to the challenge is to omit data from the day in question. However, that opportunity introduces the issue of human perception with respect to identifying exogenous causes and correctly differentiating them from endogenous causes. It is important to keep in mind that we cannot leave out predictions that are not correct without reason, because every cow needs to get a prediction every day. A different solution might be found in the training of the model. Instead of focusing on minimizing the average cost of a prediction, it is possible to train the model on minimizing the maximum cost of prediction. The measure of costs relates to a

secondary problem plaguing models of all varieties today: how to choose the cost functions, or, how to know which model is best.

Challenges with Model Selection and Evaluation

There are a number of model evaluation statistics used commonly to assess the precision and accuracy of predictions; however, when models are applied as analytics in conjunction with sensors and in the context of CPS, the system as a whole is often evaluated on the basis of sensitivity (Se) and specificity (Sp). Indeed, in an example outside nutrition, there are actually International Standards Organization standards for Se and Sp for CPS formulated initially for automated detection of mastitis (Rutten, N., A. G. J. Velthuis, W. Steeneveld, H. Hogeveen, 2013-27, 2013). Sensitivity is a model's ability to detect positive cases, that is, the percentage of all true positives that are detected. Specificity is the same metric applied to negative cases, namely the percentage of total negative cases that the model detects correctly. High Sp and low Se leads to models that rarely detect (predict) a positive case, while the opposite would be true of high Se, low Sp models. In the case of precision feeding, a positive case may be overfeeding and a negative case may be underfeeding. Alternatively, if detecting metabolic disease is an important attribute of the precision feeding system, a positive case might be an animal with metabolic disease whereas a negative case might be an animal free from disease. Although both of these

calculations are extremely important for a useful CPS model in animal agriculture, false alarms can become an issue, especially in cases where the proportion of positive to negative cases is skewed in the overall population. In the case of models that detect animal conditions to alert farmers, the positive predictive value (**PPV**) is a third measure of model accuracy that should be considered. The PPV can be thought of as the probability that an alert (predicted positive case) actually is positive. Models with low PPV will have more false alarms. Although PPV would not be useful in the proportion of positive to negative cases in the population was equal, in many disease detection, less than 1% of cow-days on a typical farm will be positive.

When we consider the example of predicting intake, or designing an ideal supplementation strategy for a cow, the use of Se and Sp for model evaluation becomes more nebulous. Undoubtedly, it is more important to know by how much you over- or under-predicted a response like intake or milk yield than it is to know the binary directionality of the residual. A number of statistics (root mean squared error, mean absolute error, etc.) are available to quantify fit in this manner. However, as discussed above, when making recommendations on-farm, incorporating the cost of these decisions is perhaps most important. Working more explicitly to tie performance predictions to economic data on-farm will be an important step in advancing analytics of precision feeding.

When are the Analytics Good Enough?

As John von Neumann said, "truth … is much too complicated to allow anything but approximations" (Szász, 2011). Approximations are a necessary evil, particularly in the business of feeding animals. Livestock nutrition is a complex science, verging on an art form, and successful nutritionists combine analytics and exogenous information to optimize productivity of their farms. A CPS, almost by design, limits the opportunity for exogenous data or, at a minimum, change the way that exogenous data will influence the system. To assess gold standards for when a CPS is good enough for deployment to farms, it may be useful to evaluate the standards professional nutritionists use for making feeding recommendations. Many nutritionists have a dollar value or a milk response cutoff that they believe a product, or feeding recommendation, must be expected to achieve before it should be recommended to a farmer. Gaining consensus on those cutoffs may be one way to evaluate the relevance of precision nutrition analytics from an industry context. Although it is possible to set more objective cutoffs, creating such an objective cutoff implies that a given model's knowledge completely covers that of the experts, which is very unlikely. Although models can help weigh options in complex environments, they are only as complex as the data they are trained on, and thus by default are less informed than an expert who has the opportunity to see exogenous and endogenous variables. Further

work is needed to identify the best strategies to combine and incorporating expert opinion/knowledge into CPS focused on animal feeding.

References

Altman, N. S. 1992. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. Am. Stat. 46:175–185. doi:10.1080/00031305.1992.10475879. Available from: https://amstat.tandfonline.com/doi/abs/10.1080/00031305.1992.10475879

Berckmans, D. 2014. Precision livestock farming technologies for welfare management in intensive livestock systems. Revue Scientifique et Technique de l'OIE. 33:189–196. doi:10.20506/rst.33.1.2273. Available from: http://dx.doi.org/10.20506/rst.33.1.2273

Berry, D. P., M. P. Coffey, J. E. Pryce, Y. de Haas, P. Løvendahl, N. Krattenmacher, J. J. Crowley, Z. Wang, D. Spurlock, K. Weigel, K. Macdonald, and R. F. Veerkamp. 2014. International genetic evaluations for feed intake in dairy cattle through the collation of data from multiple sources. J. Dairy Sci. 97:3894–3905. doi:10.3168/jds.2013-7548. Available from: http://dx.doi.org/10.3168/jds.2013-7548

Berry, D. P., and J. J. Crowley. 2013. Cell Biology Symposium: genetics of feed efficiency in dairy and beef cattle. J. Anim. Sci. 91:1594–1613. doi:10.2527/jas.2012-5862. Available from: http://dx.doi.org/10.2527/jas.2012-5862

Dale, A. J., S. McGettrick, A. W. Gordon, and C. P. Ferris. 2016. The effect of two contrasting concentrate allocation strategies on the performance of grazing dairy cows. Grass Forage Sci. 71:379–388. doi:10.1111/gfs.12185. Available

from: http://doi.wiley.com/10.1111/gfs.12185

Damas, M., M. Salmeron, A. Diaz, J. Ortega, A. Prieto, and G. Olivares. 2000. Genetic algorithms and neuro-dynamic programming: application to water supply networks. In: Proceedings of the 2000 Congress on Evolutionary Computation. CEC00 (Cat. No.00TH8512). Vol. 1. p. 7–14 vol.1. Available from: http://dx.doi.org/10.1109/CEC.2000.870269

Dela Rue, B. T., and C. R. Eastwood. 2017. Individualised feeding of concentrate supplement in pasture-based dairy systems: practices and perceptions of New Zealand dairy farmers and their advisors. Anim. Produc. Sci. 57:1543–1549. doi:10.1071/AN16471. Available from: http://www.publish.csiro.au/an/AN16471

Fukushima, K. 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern. 36:193–202. doi:10.1007/BF00344251. Available from: https://doi.org/10.1007/BF00344251

Gelman, A. 2006. Multilevel (Hierarchical) Modeling: What It Can and Cannot Do. Technometrics. 48:432–435. doi:10.1198/004017005000000661. Available from: https://doi.org/10.1198/004017005000000661

Greenland, S., M. A. Mansournia, and D. G. Altman. 2016. Sparse data bias: a problem hiding in plain sight. BMJ. 352:i1981. doi:10.1136/bmj.i1981. Available from: http://dx.doi.org/10.1136/bmj.i1981

de Haas, Y., J. E. Pryce, M. P. L. Calus, E. Wall, D. P. Berry, P. Løvendahl, N. Krattenmacher, F. Miglior, K. Weigel, D. Spurlock, K. A. Macdonald, B. Hulsegge, and R. F. Veerkamp. 2015. Genomic prediction of dry matter intake in dairy cattle from an international data set consisting of

research herds in Europe, North America, and Australasia. J. Dairy Sci. 98:6522–6534. doi:10.3168/jds.2014-9257. Available from: http://www.sciencedirect.com/science/article/pii/S0022030215004919

Hall, M. B., and G. B. Huntington. 2008. Nutrient synchrony: sound in theory, elusive in practice. J. Anim. Sci. 86:E287–92. doi:10.2527/jas.2007-0516. Available from: http://dx.doi.org/10.2527/jas.2007-0516

Hanigan, M. D., J. A. D. R. N. Appuhamy, and P. Gregorini. 2013. Revised digestive parameter estimates for the Molly cow model. J. Dairy Sci. 96:3867–3885. doi:10.3168/jds.2012-6183. Available from: http://dx.doi.org/10.3168/jds.2012-6183

Han, X., Z. Shen, W.-X. Wang, and Z. Di. 2015. Robust reconstruction of complex networks from sparse data. Phys. Rev. Lett. 114:028701. doi:10.1103/PhysRevLett.114.028701. Available from: http://dx.doi.org/10.1103/PhysRevLett.114.028701

Higgs, R. J., L. E. Chase, D. A. Ross, and M. E. Van Amburgh. 2015. Updating the Cornell Net Carbohydrate and Protein System feed library and analyzing model sensitivity to feed inputs. J. Dairy Sci. 98:6340–6360. doi:10.3168/jds.2015-9379. Available from: http://dx.doi.org/10.3168/jds.2015-9379

Howard, R. A. 1960. Dynamic Programming and Markov Processes. Published jointly by the Technology Press of the Massachusetts Institute of Technology and. Available from: https://market.android.com/details?id=book-fXJEAAAAIAAJ

Huhtanen, P., M. Rinne, P. Mäntysaari, and J. Nousiainen. 2011. Integration of the effects of animal and dietary factors on total dry matter intake of dairy cows fed silage-based

diets. Animal. 5:691–702. doi:10.1017/S1751731110002363. Available from: http://dx.doi.org/10.1017/S1751731110002363

Ivezić, Ž., A. J. Connolly, J. T. VanderPlas, and A. Gray. 2014. Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data. Princeton University Press. Available from: https://market.android.com/details?id=book-2fM8AQAAQBAJ

Jensen, L. M., N. I. Nielsen, E. Nadeau, B. Markussen, and P. Nørgaard. 2015. Evaluation of five models predicting feed intake by dairy cows fed total mixed rations. Livest. Sci. 176:91–103. doi:10.1016/j.livsci.2015.03.026. Available from: http://www.sciencedirect.com/science/article/pii/S1871141315001547

Kleinman, P. J. A., S. Spiegal, J. R. Rigby, S. C. Goslee, J. M. Baker, B. T. Bestelmeyer, R. K. Boughton, R. B. Bryant, M. A. Cavigelli, J. D. Derner, E. W. Duncan, D. C. Goodrich, D. R. Huggins, K. W. King, M. A. Liebig, M. A. Locke, S. B. Mirsky, G. E. Moglen, T. B. Moorman, F. B. Pierson, G. P. Robertson, E. J. Sadler, J. S. Shortle, J. L. Steiner, T. C. Strickland, H. M. Swain, T. Tsegaye, M. R. Williams, and C. L. Walthall. 2018. Advancing the Sustainability of US Agriculture through Long-Term Research. J. Environ. Qual. 47:1412–1425. doi:10.2134/jeq2018.05.0171. Available from: http://dx.doi.org/10.2134/jeq2018.05.0171

Lawrence, D. C., M. O'Donovan, T. M. Boland, E. Lewis, and E. Kennedy. 2015. The effect of concentrate feeding amount and feeding strategy on milk production, dry matter intake, and energy partitioning of autumn-calving Holstein-Friesian cows. J. Dairy Sci. 98:338–348. doi:10.3168/jds.2014-7905. Available from: http://dx.doi.org/10.3168/jds.2014-7905

Li, M. M., R. R. White, and M. D. Hanigan. 2018. An evaluation

of Molly cow model predictions of ruminal metabolism and nutrient digestion for dairy and beef diets. J. Dairy Sci. 101:9747–9767. doi:10.3168/jds.2017-14182. Available from: http://dx.doi.org/10.3168/jds.2017-14182

Little, M. W., N. E. O'Connell, and C. P. Ferris. 2016. A comparison of individual cow versus group concentrate allocation strategies on dry matter intake, milk production, tissue changes, and fertility of Holstein-Friesian cows offered a grass silage diet. J. Dairy Sci. 99:4360–4373. doi:10.3168/jds.2015-10441. Available from: http://dx.doi.org/10.3168/jds.2015-10441

Lloyd, S. 1982. Least squares quantization in PCM. IEEE Trans. Inf. Theory. 28:129–137. Available from: https://evasion.imag.fr/Membres/Franck.Hetroy/Teaching/ProjetsImage/2007/Bib/lloyd-1982.pdf

McCulloch, W. S., and W. Pitts. 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5:115–133. doi:10.1007/BF02478259. Available from: https://link.springer.com/article/10.1007/BF02478259

McDougall, S., K. E. Agnew, R. Cursons, X. X. Hou, and C. R. W. Compton. 2007. Parenteral treatment of clinical mastitis with tylosin base or penethamate hydriodide in dairy cattle. J. Dairy Sci. 90:779–789. doi:10.3168/jds.S0022-0302(07)71562-X. Available from: http://dx.doi.org/10.3168/jds.S0022-0302(07)71562-X

McParland, S., E. Lewis, E. Kennedy, S. G. Moore, B. McCarthy, M. O'Donovan, S. T. Butler, J. E. Pryce, and D. P. Berry. 2014. Mid-infrared spectrometry of milk as a predictor of energy intake and efficiency in lactating dairy cows. J. Dairy Sci. 97:5863–5871. doi:10.3168/jds.2014-8214. Available from: http://dx.doi.org/10.3168/jds.2014-8214

Nash, E., P. Korduan, and R. Bill. 2009. Applications of open

geospatial web services in precision agriculture: a review. Precis. Agric. 10:546–560. doi:10.1007/s11119-009-9134-0. Available from: http://dx.doi.org/10.1007/s11119-009-9134-0

National Research Council, Board on Agriculture and Natural Resources, Committee on Animal Nutrition, and Subcommittee on Dairy Cattle Nutrition. 2001. Nutrient Requirements of Dairy Cattle: Seventh Revised Edition, 2001. National Academies Press. Available from: https://market.android.com/details?id=book-EgWw2Kd9mb4C

Ohlsson, T. 2014. Sustainability and Food Production. In: Food Safety Management. p. 1085–1097. Available from: http://dx.doi.org/10.1016/b978-0-12-381504-0.00043-3

Pearson, K. 1901. On Lines and Planes of Closest Fit to Systems of Points in Space. University College. Available from: https://market.android.com/details?id=book-uGt_YgEACAAJ

Rutten, C. J., A. G. J. Velthuis, W. Steeneveld, and H. Hogeveen. 2013. Invited review: sensors to support health management on dairy farms. J. Dairy Sci. 96:1928–1952. doi:10.3168/jds.2012-6107. Available from: http://dx.doi.org/10.3168/jds.2012-6107

Rutten, N., A. G. J. Velthuis, W. Steeneveld, H. Hogeveen. 2013-27, 2013. Sensor systems for dairy cow health management: a review. In: Proceedings of the Precision Dairy Conference and Expo. Mayo Civic Center. p. 89–90.

Scholz, M. 2006. Approaches to analyse and interpret biological profile data. Available from: https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/index/index/docId/696

Shah, I. A., I. Khan, S. A. Mir, M. S. Pukhta, and A. A. Lone. 2018. Principal Component Analysis Utilizing R and SAS Software’s. Int.J.Curr.Microbiol.App.Sci. 7:3794–3801. doi:10.20546/ijcmas.2018.705.441. Available from: http://dx.doi.org/10.20546/ijcmas.2018.705.441

Shetty, N., P. Løvendahl, M. S. Lund, and A. J. Buitenhuis. 2017. Prediction and validation of residual feed intake and dry matter intake in Danish lactating dairy cows using mid-infrared spectroscopy of milk. J. Dairy Sci. 100:253–264. doi:10.3168/jds.2016-11609. Available from: http://dx.doi.org/10.3168/jds.2016-11609

Sinclair, L. A. 1995. Effects of Synchronizing the Rate of Dietary Energy and Nitrogen Release in Diets with a Similar Carbohydrate Composition on Rumen Fermentation and Microbial Protein Synthesis in Sheep. Available from: http://books.google.com/books/about/Effects_of_Synchronizing_the_Rate_of_Die.html?hl=&id=NtufDAEACAAJ

Sinclair, L. A., P. C. Garnsworthy, J. R. Newbold, and P. J. Buttery. 1995. Effects of synchronizing the rate of dietary energy and nitrogen release in diets with a similar carbohydrate composition on rumen fermentation and microbial protein synthesis in sheep. J. Agric. Sci. 124:463–472. doi:10.1017/S0021859600073421. Available from: https://www.cambridge.org/core/journals/journal-of-agricultural-science/article/effects-of-synchronizing-the-rate-of-dietary-energy-and-nitrogen-release-in-diets-with-a-similar-carbohydrate-composition-on-rumen-fermentation-and-microbial-protein-synthesis-in-sheep/03A795D6B463E34556AC25B14FFF80C5

Szász, D. 2011. John von Neumann, the Mathematician. Math. Intelligencer. 33:42–51. doi:10.1007/s00283-011-9223-6. Available from: https://doi.org/10.1007/s00283-011-9223-6

Thornton, P. K. 2010. Livestock production: recent trends, future prospects. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365:2853–2867. doi:10.1098/rstb.2010.0134. Available from: http://dx.doi.org/10.1098/rstb.2010.0134

Thorp, K. R., and L. F. Tian. 2004. A Review on Remote Sensing

of Weeds in Agriculture. Precis. Agric. 5:477–508. doi:10.1007/s11119-004-5321-1. Available from: http://dx.doi.org/10.1007/s11119-004-5321-1

Tullo, E., A. Finzi, and M. Guarino. 2019. Review: Environmental impact of livestock farming and Precision Livestock Farming as a mitigation strategy. Sci. Total Environ. 650:2751–2760. doi:10.1016/j.scitotenv.2018.10.018. Available from: http://dx.doi.org/10.1016/j.scitotenv.2018.10.018

Ushikubo, S., C. Kubota, and H. Ohwada. 2017. The Early Detection of Subclinical Ketosis in Dairy Cows Using Machine Learning Methods. In: Proceedings of the 9th International Conference on Machine Learning and Computing. ACM, New York, NY, USA. p. 38–42. Available from: http://doi.acm.org/10.1145/3055635.3056625

Van Amburgh, M. E., E. A. Collao-Saenz, R. J. Higgs, D. A. Ross, E. B. Recktenwald, E. Raffrenato, L. E. Chase, T. R. Overton, J. K. Mills, and A. Foskolos. 2015. The Cornell Net Carbohydrate and Protein System: Updates to the model and evaluation of version 6.5. J. Dairy Sci. 98:6361–6380. doi:10.3168/jds.2015-9378. Available from: http://dx.doi.org/10.3168/jds.2015-9378

Volden, H., N. I. Nielsen, M. Åkerlind, M. Larsen, Ø. Havrevoll, and A. J. Rygh. 2011. Prediction of voluntary feed intake. In: H. Volden, editor. NorFor - The Nordic feed evaluation system. Wageningen Academic Publishers, Wageningen. p. 113–126. Available from: https://doi.org/10.3920/978-90-8686-718-9_10

Werbos, P. 1974. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. Harvard. Available from: https://books.google.com/books/about/Beyond_Regression.html?hl=&id=r4GToAEACAAJ

White, R. R., M. Brady, J. L. Capper, and K. A. Johnson. 2014. Optimizing diet and pasture management to improve sustainability of U.S. beef production. Agric. Syst. 130:1–12. doi:10.1016/j.agsy.2014.06.004. Available from: http://dx.doi.org/10.1016/j.agsy.2014.06.004

White, R. R., M. Brady, J. L. Capper, J. P. McNamara, and K. A. Johnson. 2015. Cow-calf reproductive, genetic, and nutritional management to improve the sustainability of whole beef production systems. J. Anim. Sci. 93:3197–3211. doi:10.2527/jas.2014-8800. Available from: http://dx.doi.org/10.2527/jas.2014-8800

White, R. R., M. B. Hall, J. L. Firkins, and P. J. Kononoff. 2017. Physically adjusted neutral detergent fiber system for lactating dairy cow rations. I: Deriving equations that identify factors that influence effectiveness of fiber. J. Dairy Sci. 100:9551–9568. doi:10.3168/jds.2017-12765. Available from: http://dx.doi.org/10.3168/jds.2017-12765

Zhang, C., and J. M. Kovacs. 2012. The application of small unmanned aerial systems for precision agriculture: a review. Precis. Agric. 13:693–712. doi:10.1007/s11119-012-9274-5. Available from: http://dx.doi.org/10.1007/s11119-012-9274-5

Figures

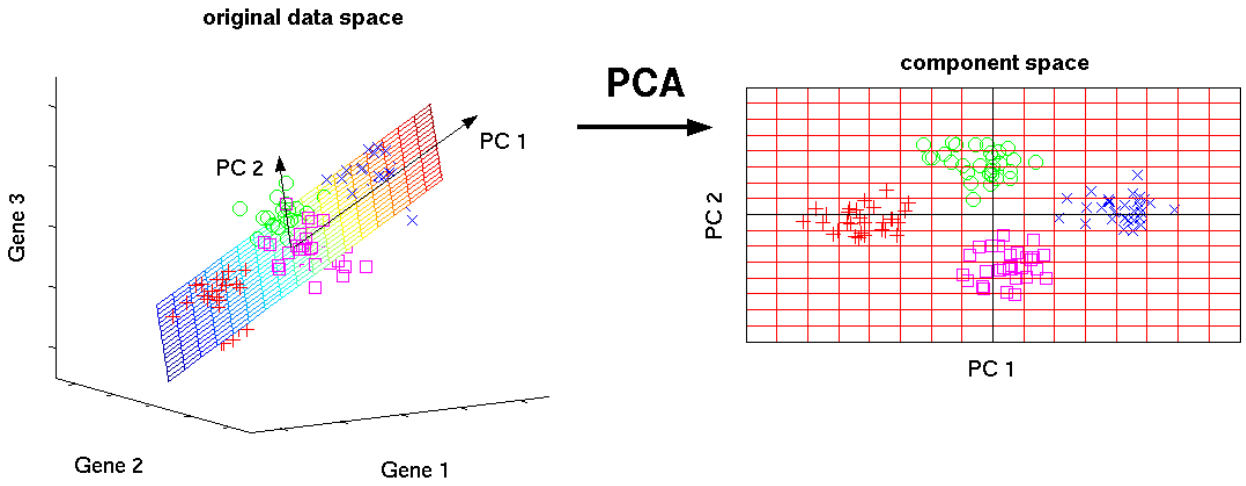

Figure 4-1. Example of Principal Component Analysis using three gene variables. The three gene factors are combined into two components that maximize variance between groups in 2-dimensional space, as seen on the right (Scholz, 2006).

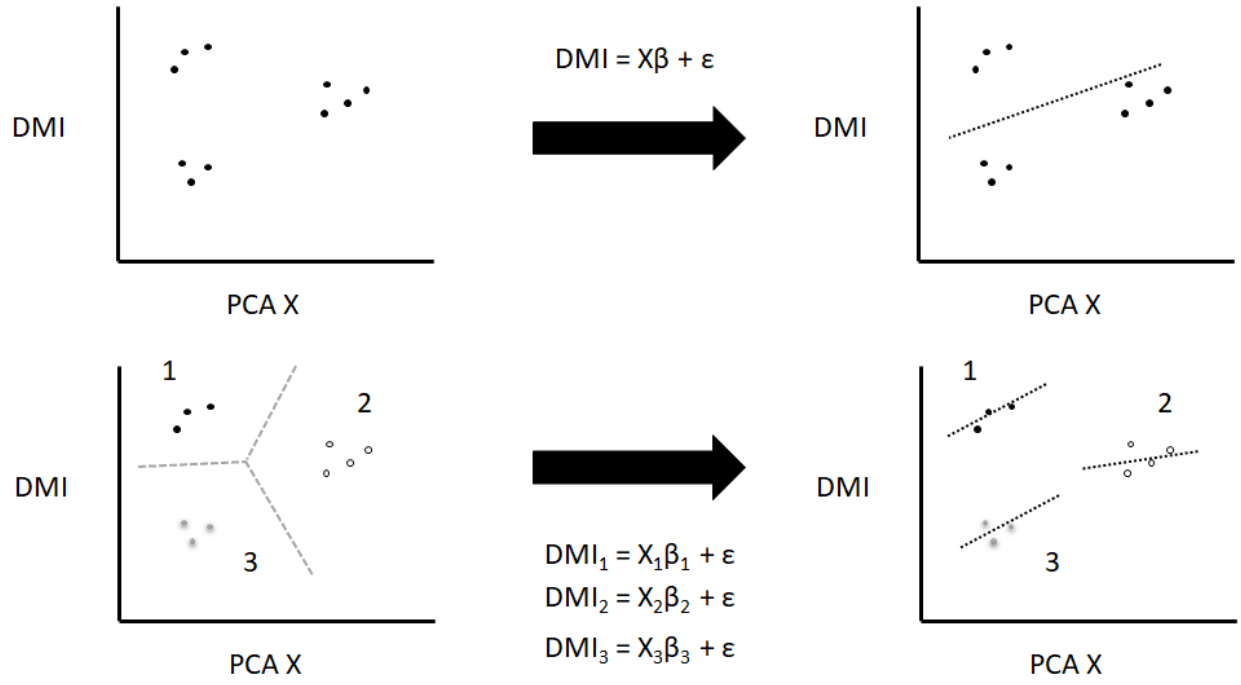

Figure 4-2. Comparison of fitting models after grouping PCA results. Top) Fitting single linear model to data after PCA. Bottom) Fitting group of linear models based on clustering algorithm to data after PCA. Grouping data based on a clustering algorithm allows the same model increased flexibility when making predictions. Notice that the linear model used does not change, only the data used to train the model is varied.

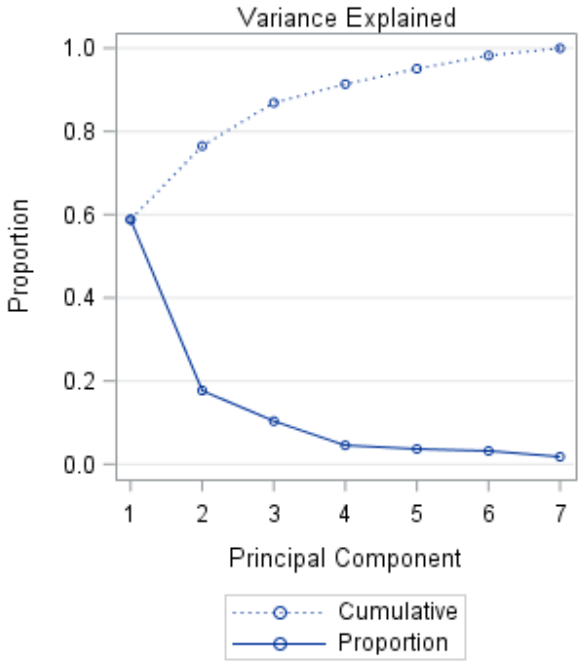

Figure 4-3. A potential example plot of the proportion of variance explained by each additional component in principal component analysis. Variance explained by each additional component can vary considerably based on the data you are working with. (Shah et al., 2018)

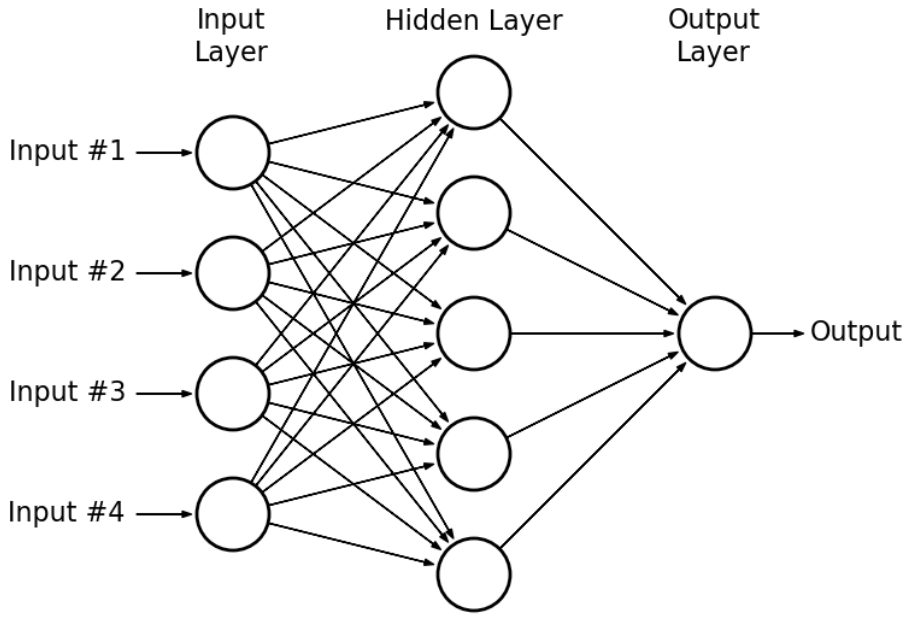

Figure 4-4. An example of a neural network framework. Circles represent individual equations which are fed data from all connected nodes. The lack of a 1-1 ratio of nodes in each layer of the network forces the model to condense information and leads to the most important information being determined iteratively through backpropagation of error. (Ivezić et al., 2014)

Chapter 5: Methodological and feed factors affecting measurement of protein A, B, and C fractions, degradation rate, and intestinal digestibility of rumen-undegraded protein

Published as: Liebe, D.M., J.L. Firkins, H. Tran, P.J. Kononoff, and R.R. White. 2018. Technical note: Methodological and feed factors affecting measurement of protein A, B, and C fractions, degradation rate, and intestinal digestibility of rumen-undegraded protein. J. Dairy Sci. 101:8046–8053. doi:10.3168/jds.2018-14837.

Abstract

When formulating dairy cow rations, characterization of protein in feed requires estimation of protein degradation in both the rumen and intestine. The objective of this work was to evaluate experimental and feed related factors that affect characterization, using in-situ, in-vitro, or mobile bag techniques, of 0-h washout (A), potentially degradable (B), and undegradable (C) protein fractions, protein degradation rate (Kd), and digestibility of rumen undegradable protein (dRUP). Datasets of 136 studies on A, B, C, and Kd and 113 studies on dRUP were amassed from the literature. Mixed-effect linear models were used to relate these variables to methodological and feed factors while accounting for random differences among studies. Predictions of A, B, and C protein fractions were significantly ($P < 0.05$) influenced by CP and NDF interactions with sample grind size, bag pore size, incubation time, bag area, and sample-to-bag area ratio. For example, a 20% decrease in CP of a theoretical legume silage sample

would increase A fraction prediction by 20.1%, but 34.7% with bag incubation time -1 SD below the mean. A shift in measurement method nearly doubles the predicted A fraction of the same feed. Similarly, reported Kd values were significantly ($P < 0.05$) influenced by CP interactions with sample grind size, bag area, and sample-to-bag area ratio. Feed variables and measurement variables influencing protein digestibility measures suggest that these analytical factors are likely associated with variance among differing methodologies and within unique samples of the same feed. When predicting dRUP, pepsin-acid incubation time and use of mobile bag method produced significantly different ($P < 0.05$) estimates compared to the traditional in-vitro 3-step method. The use of mobile bag resulted in a 12% (±3.1%) higher estimate of dRUP compared to the in-situ technique. In 618 and 977 samples, sample variation to sample mean ratio for ADF and pepsin-acid incubation time was 63% and 58%, respectively. Variation in feedstuff content and lack of standardization of methods used to measure protein disappearance lead to a lack of robustness in the measurements commonly employed.

Keywords: ruminal degradability of protein, in-situ, in-vitro

TECHNICAL NOTE

Data Collection and Preparation

Data were collected from published, peer-reviewed papers that reported ruminal disappearance of protein from nylon bags and intestinal disappearance of protein using either in vitro or mobile bag procedures. Keywords used in searches for relevant articles included: protein, digestibility, nylon bag, in situ, mobile bag, protein disappearance, and rumen undegradable protein. Searches were conducted in the Fall of 2015 using Google (http://www.scholar.google.com/) and PubMed (https://www.ncbi.nlm.nih.gov/pubmed), as well as using the searches located on the websites of the following journals: Journal of Dairy Science (http://www.journalofdairyscience.org/) , Canadian Journal of Animal Science (http://www.nrcresearchpress.com/journal/cjas), Animal (https://www.cambridge.org/core/journals/animal), and Animal Feed Science and Technology (http://www.sciencedirect.com/science/journal/03778401). Every article recovered was also screened for references with relevant titles for subsequent searches. If studies used animals other than cattle, the article was not considered.

Reported estimates of feed A, B, C protein fractions, protein degradation rate (Kd) of the B fraction (Ørskov and McDonald, 1979), and intestinal digestibility of RUP (dRUP), using either in vitro or mobile bag methods (Calsamiglia and Stern, 1995; Hvelplund

at al. 1992; Gargallo et al., 2006), were gathered from published studies. Additional data gathered from each study included feed sample name, proportion of forage fed, bag characteristics (dimensions and pore size), and incubation times, and all criteria outlined in the NRC (2001) for standardization of in-situ procedures (NRC, 2001). Studies that failed to report microbial nitrogen corrections were included to compile a robust dataset. A total of 187 studies reporting A, B, C, and Kd and 143 studies reporting dRUP were identified. Of these, studies from which no A or C fraction could be derived due to log-linear transformation, short incubation time, or otherwise were removed. Studies that incubated bags for less than 48 hours were also removed. If the sum of the fractions (A+B+C) diverged from 100% by more than 10%, the study was excluded. Studies reporting feed crude protein above 100% of dry matter or RUP digestibility less than 0% of RUP were also excluded from the analysis. After applying these exclusion criteria, 136 studies on A, B, C and Kd and 113 studies on dRUP were used for analysis. The summary statistics of the degradability dataset are included in Table 5-1 and summary statistics of the digestibility dataset are included in Table 5-2.

Some studies failed to report all protein fractions or reported three protein fractions that did not sum to 100. For studies that used log-linear approaches without a computer model, the B pool was adjusted until A + B + C equaled 100%. For studies

fitting A and B using a computer model, C was estimated as 100 – A – B. If a study fitting A and B using a computer model had A and B summing to greater than 100, the A fraction was decreased until A and B summed to 100. Finally, if B and C were fit by a computer model, A was assumed to be 100 – A - C.

Mathematical Models

Multiple regression models were used to estimate all dependent variables of interest. All models included a random effect of study and fixed effects for each independent variable. Statistical analysis was conducted as described by Roman-Garcia et al. (2016) using the lmer package (Kuznetsova et al., 2013) in R version 3.1.0. (R Core Team, 2014).

Model Derivation Approach

Key groupings of explanatory variables (Table 5-1) were included in an initial multiple regression model. Variables were iteratively eliminated based on removing the highest *P*-value for each iteration until all variable P-values suggested at least a tendency ($P < 0.10$) for significance. For A, B, C, and Kd, the initial model included: feed NDF (% of DM), feed CP (% of DM), forage type, feed category, feed name, grind size, bag area, bag pore size, ratio of sample size to bag area, and incubation length. Interaction terms for feed category, forage type, or NDF and CP with pore size, grind size, bag area, bag pore size, ratio of sample size to bag area, and incubation length of time were also

tested. For the dRUP dataset, the initial model included feed NDF, ADF, and CP, feed name, total time incubated in the rumen, time exposed to acid, and method (in vitro or mobile bag). Two-way interaction terms for CP, NDF, ADF, pore size, incubation time, and acid incubation time with method were also evaluated.

A critical reason for the interest in understanding the influence of methodological factors in estimating A, B, C, Kd, and dRUP are the results of White et al. (2017a,b). This study concluded that existing A, B, and C fraction estimates, in combination with Kd values, did not yield adequate predictions of ruminally degradable protein. One reason for the failure of these feed-specific factors to function as effective predictors of dietary rumen degradable protein would be methodological inconsistencies in estimating A, B, and C fractions or Kd among feeds. The presence of significant interactions between feed chemical components and methodological variables would suggest a possible reason for why the A, B, and C fraction estimates and Kd values are not representative predictors of ruminal N outflows.

Chemical and Methodological Factors Influencing Measurement of Protein Fractions

For the final A, B, C, and Kd models (Table 5-3), a combination of surface area of the bag, grind size, pore size, incubation time, sample size to bag area ratio, NDF, and CP were

identified as significant variables. The model for B fraction did not include area or grind size. The C fraction model did not use grind size. All protein fractions (A, B, C) had significant interaction terms for CP and NDF interactions with methodological variables. Significant interaction terms with CP were present for incubation time, pore size, bag area, grind size, and sample size to bag area ratio for the A, B, and C protein fraction models. Significant interaction terms with sample NDF were also present for pore size, bag area, and sample to bag area ratio in the A, B, and C fraction models. The model for Kd only included interaction terms with CP, for bag area, grind size, and sample size to bag area ratio.

These results suggest challenges with current practices measuring A, B, and C fractions of protein. A major challenge is the fact that methodological variables significantly influenced reported A, B, and C fractions within a feed. Measuring the same feed but grinding to different sample particle sizes would result in two different predictions for A, B, and C fraction protein. For example, estimating fraction protein disappearance in legume silage (NDF = 44.5%, CP = 19.4%) using incubation time ±1 SD of those reported in the dataset (all other variables represented by the means in Table 5-1) would result in a 36.5% increase, or a 12.5 percentage point increase, in estimated B fraction protein.

An additional challenge is illustrated by the models: the observation that feed chemical composition interacted with methodological factors to influence reported A, B, and C fractions. These interactions suggest that fraction measurement is not consistent across feed types or perhaps even within feeds of differing chemical composition (low CP forage vs high CP forage). For example, a 20% decrease in the CP content of legume hay (from 20.5% to 16.4%, all other variables represented by the means in Table 5-1) would result in an increase the predicted A fraction by 20.1% and decrease the predicted B fraction by 7.9%. However, if an incubation time-1 SD from the mean was used for measurement, the same 20% decrease in CP content of legume hay would increase A fraction by 34.7% and decrease B fraction by 8.7%. The use of a shorter incubation time increases the effect of decreasing CP content on fraction predictions. To further demonstrate the challenges revealed by the significant variables and their interactions, Table 5-4 lists the effects of changing the variables in the model for B fraction protein by ±1 standard deviation for a variety of feeds included most often in the dataset.

An additional challenge with measuring protein fractions is artifact from bacterial CP contamination. Research has shown that rumen-exposed feedstuffs contained microbial nitrogen between 8 and 26%, as a percentage of total nitrogen (Erasmus et al., 1994). Bacterial contamination can decrease RDP by up to 5% (Alexandrov,

1998) and also potentially underestimate Kd through the mismeasurement of the B fraction (Wanderley et al., 1993). To overcome the challenge of contamination, researchers have attempted to quantify microbial nitrogen in samples by using various microbial markers in the residue (Michalet-Doreau and Ould-Bah, 1992; Machado et al., 2013) most recently employing real-time PCR to measure microbial DNA (Sylvester et al., 2004; Paz et al., 2014). Increasing NDF and decreasing CP increased the potential for bacterial contamination to effect fractions and Kd estimates, along with adding starch to grasses and increasing incubation times (De Visser et al., 1998; Yang et al., 1999). Because Kd is derived using estimates of the material undigested in the rumen, Kd is highly correlated with the protein fraction estimations (Woods et al., 2003).

Chemical and Methodological Factors Influencing Measurement of Kd

For the final Kd model, bag area, grind size, sample size to bag area ratio, NDF, and CP were identified as significant. Additionally, CP had significant interactions with bag area, grind size, and sample size to bag area ratio. The measurement of Kd appears to be affected by similar challenges described above for the A, B, and C protein fractions within a feed. Namely, method significantly influenced outcomes such that interactions were

detected between chemical composition and methodological factors that influenced measured Kd. For example, a 10% decrease in the CP content of legume hay using the dataset's mean grind size would result in a 5.5% increase in predicted Kd. The same 2 CP concentrations measured with a grind size -1 SD from the mean would result in a 9.0% increase in predicted Kd. The varying relationship between Kd and CP using 2 different grind sizes for the chemical analysis exhibits almost double the predicted Kd. In practice, this means that samples measured with a finer grind will result in elevated Kd estimates, possibly leading to an overestimation in RDP. Because of the methodological implications on measuring Kd and the resulting effect on calculating RDP, these inconsistencies might lead to a failure to accurately supply MP to the animal.

The inaccuracy of standard protein digestibility practices has been noted in studies going back to the 1980's. Weakley et al. (1983) reported variation in DM and CP disappearance measurements based on the grind and pore size of the in-situ bags used in the analysis. A later meta-analysis pointed to myriad factors that appeared to cause variation in in-situ degradability measurements (Vanzant et al. 1998). Most relevant analyses of protein digestibility measurement have suggested a standardization in procedure in some way; however, standardizing all aspects of these procedures is challenging. In particular, an approach to correct for bacterial contamination has not been standardized, although

quantitative PCR shows promise. Paz et al. (2014) showed how isolating bacterial contamination via DNA or purine markers on bacteria showed potential for better feedstuff digestibility estimation. Although no consensus has been reached on standardizing measurements for bacterial contamination, Michalet-Doreau and Ould-Bah (1992) provided several suggestions on how to reduce variation in measurement including standardizing pore size, pre-treatment bag preparation, grind size, and incubation time. Despite these suggestions, time-series analysis of the data collected herein suggests standardization does not appear to be occurring in practice.

In the current dataset, with studies from 1985 to 2015, the average protein fractionation digestibility measurement parameters did not change significantly over the 30-year period for grind size, bag area, or pore size (P = 0.81, 0.98, 0.19). Year had a positive relationship with incubation time (P = 0.007), with incubation time increasing by an average of 0.9 hours each year. The variance of each parameter did not change over time for grind size, bag size, pore size, or incubation time (P = 0.80, 0.49, 0.20, 0.53, respectively). Lack of change in incubation time variation indicates that, despite the apparent increase in mean incubation time up to 2015, there is still a similar range of incubation times being used. These observations suggest that recommendations for proper disappearance measurement have not been

heeded or a consensus has not been reached in the scientific community. Future meta-analyses focused on defining robust and consistent A, B, C, and Kd estimates for use in feed libraries should consider removing studies that contain sample size to bag area ratios or pore sizes that vary significantly from current recommendations (Vanzant et al. 1998) to ensure consistent and quality data. This elimination of variance associated with measurement methods related to protein fraction measurement would improve the accuracy of future feed libraries and hopefully address the issues with the A, B, C system raised in White et al. (2017a, b).

Chemical and Methodological Factors Influencing Measurement of dRUP

The three-step procedure is currently used in feed testing laboratories for estimating intestinal digestibility of RUP in ruminants. This procedure involves grinding and bagging feed samples to be placed in the rumen of a cow for incubation, then exposing the remaining sample through pepsin and pancreatin enzymes which is also designed to remove any undigested residues (Calsamiglia and Stern, 1995). This procedure has since been modified (Gargallo et al., 2006), but remains similar to the original procedure proposed in 1995. After the three steps, the sample is precipitated out of solution and tested for remaining

protein. This protein represents that which would be undegraded in the rumen but digested in the intestine of a ruminant. In contrast, the mobile bag procedure estimated protein digestion but utilizes the small intestine of the ruminant in addition to the rumen to incubate samples to test for intestinal protein digestibility (Hvelplund et al., 1985). In the mobile bag procedure, samples are incubated in the rumen, then in pepsin, and finally placed in the small intestine of the animal via a duodenal cannula and collected in the feces and analyzed for protein content. Although it is possible to test intestinal digestibility without first incubating samples in the rumen, White et al. (2017a) showed that lack of exposure to ruminal microbes can lead to underestimation in digestibility. Although the two procedures measure the disappearance of protein in the intestine, previous research has also identified a significant difference in the measurements made on the same feedstuffs (Kopečný et al., 1998; Mesgaran et al., 2008; Jahani-Azizabadi et al., 2009).

In the models predicting dRUP of feeds, ADF, acid hydrolysis time, and the method of bagging the sample all had a significant effect on the measurement (Table 5-5). For example, using the mobile bag method increased predicted dRUP by 12 percentage points compared with the 3-step procedure. The proportion of SE to mean observation (Table 5-2) reflects the variability in reported values in the literature. For two key variables identified to be

significant (ADF and acid hydrolysis time), this ratio indicated substantial variation within the literature (63% and 58% with 618 and 977 observations, respectively). Previous work done by Harstad and Prestlokken (2000; 2001) discussed the significance of bag pore size in mobile bag testing on estimates of dRUP. The lack of significance for pore size found in this study could be a result of the relatively large mean pore size in the dataset. The authors cited a pore size of 15 µm as potentially having a significant effect on dRUP, but the mean pore size in the current dataset was 45.9 µm for in-vitro samples and 42.9 µm for mobile bag samples. Despite the caution listed in these papers, pore size is still not uniform. Perhaps one way to improve consistency of dRUP is to focus on more stringent standardization of procedures.

Although measurement of dRUP was initially intended only to evaluate high protein concentrates (Ørskov and McDonald, 1979; Stern and Satter, 1984), the technique has since been used to measure many forages and fibers, as well as rumen protection in certain amino acid supplements (Overton et al., 1996; Berthiaume et al., 2000; Koenig and Rode, 2001; Wu et al., 2012). Measuring a variety of feedstuffs will lead to the results being limited in three ways. First, feeds are not exposed to chewing, nor changes in environment, feed pH, or physical mixing (Wu and Papas, 1997; Ji et al., 2016). Second, these methods measure resistance to degradability in the rumen environment, not the absorption by the

animal after the feed leaves the bag (Wu and Papas, 1997). Lastly, testing on fine particles may allow the feed to be solubilized and leave the bag without necessarily being digested (Erasmus et al., 2014; Whitehouse et al., 2017). As a consequence of these factors, when evaluating information on rumen protection of rumen protected amino acid supplements it is possible that the analytical procedure may interact with method of protection.

A myriad of measurement variables play a significant role in determining the measured fraction protein percentages. Accurate estimates of the protein digestibility and makeup of a diet is invaluable in the dairy industry for making sure that cows are feed adequate protein for growth, maintenance, and lactation, but also to control nitrogen excretion into the environment and many studies acknowledge the growing concern over environmental costs associated with nitrogen excretions (Cooperband and Good 2002; Davidson 2009; Dong et al. 2014). The lack of consistent measurement protocols across methods suggests that measuring fraction proteins lacks accuracy and standardization of measurement methods is essential before compositing literature estimates into feed library values. Similarly, any analyses of AA should include ruminal degradation and intestinal digestibility of total AA for better integration with the CP protocols described herein (White et al., 2017c).

Conclusion

The use of A, B, and C fractions of protein and Kd of the B fraction for estimating RDP is a well-established technique for describing the protein quality in feedstuffs. Measurement methods for protein fractionation are similar, but not identical, because factors like pore size of the bags used, the grind of the feed samples tested, and the size of the bags used can vary. The assumption when measuring these values is that measurements will not vary significantly within feedstuff or by measurement method. Because specific methodological and feed factors change predictions, no unified approach to measurement will completely alleviate the errors. However, standardization of the entire measurement process would lessen the variance among similar feedstuffs. Without standardization, diet formulation will continue to suffer from poor estimations of ruminal protein fraction dynamics and digestibility of rumen-undegradable protein.

Acknowledgements

Research was supported by state and federal funds appropriated to the Virginia Tech Agricultural Research Station. The authors would like to thank C. G. Schwab (Schwab Consulting LLC, Boscobel, WI) and N. L. Whitehouse (University of New

Hampshire, Durham, NH) for providing portions of data used in this study.

References

Adams, Richard S. 1995. "Dairy Reference Manual." Northeast Regional Agricultural Engineering Service, Cooperative Extension.

Alexandrov, A.N. 1998. Effect of ruminal exposure and subsequent microbial contamination on dry matter and protein degradability of various feedstuffs. Anim. Feed Sci. Technol. 71:99–107.

de Boer, G., J. J. Murphy, and J. J. Kennelly. 1987. Mobile nylon bag for estimating intestinal availability of rumen undegradable protein. J. Dairy Sci. 70:977–982.

Boila, R. J. and J. R. Ingalls. 1994. The post-ruminal digestion of dry matter, nitrogen and amino acids in wheat-based distillers' dried grains and canola meal. Anim. Feed Sci. Technol. 49:173-188.

Boucher, S. E., S. Calsamiglia, C. M. Parsons, H. H. Stein, M. D. Stern, P. S. Erickson, P. L. Utterback, and C. G. Schwab. 2009a. Intestinal digestibility of amino acids in rumen undegradable protein estimated using a precision-fed cecectomized rooster bioassay: I. Soybean meal and soyplus. J. Dairy Sci. 92:4489-4498.

Boucher, S. E., S. Calsamiglia, C. M. Parsons, M. D. Stern, M. Ruiz Moreno, M. Vázquez-Añón, and C. G. Schwab. 2009b. In vitro digestibility of individual amino acids in rumen-undegraded protein: The modified three-step procedure and the immobilized digestive enzyme assay. J. Dairy Sci. 92:3939-3950.

Calsamiglia, S. and M. D. Stern. 1995. A three-step in vitro procedure for estimating intestinal digestibility of protein in ruminants. J. Anim. Sci. 73:1459-1465.

Cooperband, Leslie R., and Laura Ward Good. 2002. "Biogenic Phosphate Minerals in Manure: Implications for Phosphorus

Loss to Surface Waters." Environmental Science & Technology 36 (23):5075–82.

Davidson, Eric A. 2009. "The Contribution of Manure and Fertilizer Nitrogen to Atmospheric Nitrous Oxide since 1860." Nature Geoscience 2 (August). Nature Publishing Group:659.

De Visser, H., A. Klop, C.J. Van der Koelen, and A.M. Van Vuuren. 1998. Starch supplementation of grass harvested at two stages of maturity prior to ensiling: intake, digestion, and degradability by dairy cows. J. Dairy Sci. 81:2221–2227.

Dong, R. L., G. Y. Zhao, L. L. Chai, and K. A. Beauchemin. 2014. "Prediction of Urinary and Fecal Nitrogen Excretion by Beef Cattle." Journal of Animal Science 92 (10):4669–81.

Erasmus, L., Z. Bester, and R. Coertze. 2014. Milk composition as technique to evaluate the relative bioavailability of a liquid rumen protected methionine source. South Afr. J. Anim. Sci. 43:93. doi:10.4314/sajas.v43i5.16.

Erasmus, L.J., P.M. Botha, C.W. Cruywagen, and H.H. Meissner. 1994. Amino acid profile and intestinal digestibility in dairy cows of rumen-undegradable protein from various feedstuffs. J. Dairy Sci. 77:541–551.

Fathi Nasri, M. H., J. France, M. Danesh Mesgaran, and E. Kebreab. 2007. Effect of heat processing on ruminal degradability and intestinal disappearance of nitrogen and amino acids in iranian whole soybean. Livest. Sci. 113:43-51.

Gargallo, S., S. Calsamiglia, and A. Ferret. 2006. Technical note: A modified three-step in vitro procedure to determine intestinal digestion of proteins. J. Anim. Sci. 84:2163-2167.

Harstad, O. M. and E. Prestløkken. 2000. Effective rumen degradability and intestinal indigestibility of individual amino acids in solvent-extracted soybean meal (sbm) and xylose-treated sbm (soypass®) determined in situ. Anim. Feed Sci. Technol. 83:31-47.

Harstad, O. M. and E. Prestløkken. 2001. Rumen degradability and intestinal indigestibility of individual amino acids in corn gluten meal, canola meal and fish meal determined in situ. Anim. Feed Sci. Technol. 94:127-135.

Hvelplund, T. 1985. Digestibility of rumen microbial protein and undegraded dietary protein estimated in the small intestine of sheep and by in sacco procedure. Acta Agriculturae Scandinavica, Section A-Animal Science 25:132-144.

Hvelplund, T., M.R. Weisbjerg, and L.S. Andersen. 1992. Estimation of the true digestibility of rumen undegraded dietary protein in the small intestine of ruminants by the mobile bag technique. Acta Agriculturae Scandinavica A-Animal Sciences 42:34–39.

Jahani-Azizabadi, H., M. Danesh Mesgaran, R. Valizadeh, and H. Nassiri Moghaddam. 2009. Comparison of in vivo with in situ mobile bag and three step enzymatic procedures to evaluate protein disappearance of alfalfa hay and barley grain. Iranian Journal of Veterinary Research. 10:260–266.

Ji, P., H.A. Tucker, R.E. Clark, M. Miura, and C.S. Ballard. 2016. Short communication: Effect of on-farm feeding practices on rumen protected lysine products. J. Dairy Sci. 99:1242–1246. doi:10.3168/jds.2014-9197.

Kopečný, J., O. Tománková, and P. Homolka. 1998. Comparison of protein digestibility of rumen undegraded protein estimated by an enzymatic and mobile bag method: feeds for ruminants and anaerobic fungus. Anim. Feed Sci. Technol. 71:109–116.

Lee, C., A. N. Hristov, T. W. Cassidy, K. S. Heyler, H. Lapierre, G. A. Varga, M. J. de Veth, R. A. Patton, and C. Parys. 2012. Rumen-protected lysine, methionine, and histidine increase milk protein yield in dairy cows fed a metabolizable protein-deficient diet. J. Dairy Sci. 95:6042-6056.

Ljøkjel, K., O. M. Harstad, and A. Skrede. 2000. Effect of heat treatment of soybean meal and fish meal on amino acid digestibility in mink and dairy cows. Anim. Feed Sci. Technol. 84:83-95.

Machado, P.A.S., S.C. Valadares Filho, E. Detmann, S.A. Santos, R.F.D. Valadares, C. Ducatti, P.P. Rotta, and L.F. Costa e Silva. 2013. Development of equations to estimate microbial contamination in ruminal incubation residues of forage produced under tropical conditions using N as a label. J. Anim. Sci. 91:3836–3846.

Mesgaran, M. D., H. Jahani-Azizabadi, and A. R. H. Moussavi. 2008. Comparison of techniques to determine the ruminal and post-ruminal protein disappearance of various oilseed meals. Res. J. Biol. Sci. 3:1028–1033.

Michalet-Doreau, B., and M. Y. Ould-Bah. 1992. In vitro and in sacco methods for the estimation of dietary nitrogen degradability in the rumen: a review. Anim. Feed Sci. Technol. 40:57–86.

Ørskov, E.R., and I. McDonald. 1979. The estimation of protein degradability in the rumen from incubation measurements weighted according to rate of passage. The Journal of Agricultural Science 92:499–503. doi:10.1017/S0021859600063048.

Patton, R. A., A. N. Hristov, and H. Lapierre. 2014. Protein feeding and balancing for amino acids in lactating dairy cattle. Vet. Clin. Food Anim. 30:599-621.

NRC. 2001. Nutrient Requirements of Dairy Cattle. 7th rev. ed. Natl. Acad. Press, Washington, DC

Paz, H. A., T. J. Klopfenstein, D. Hostetler, S. C. Fernando, E. Castillo-Lopez, and P. J. Kononoff. 2014. Ruminal degradation and intestinal digestibility of protein and amino acids in

high-protein feedstuffs commonly used in dairy diets. J. Dairy Sci. 97:6485-6498.

Stern, M. D., A. Bach, and S. Calsamiglia. 1997. Alternative techniques for measuring nutrient digestion in ruminants. J. Anim. Sci. 75:2256-2276.

Stern, M.D., and L.D. Satter. 1984. Evaluation of nitrogen solubility and the dacron bag technique as methods for estimating protein degradation in the rumen. J. Anim. Sci. 58:714–724.

Sylvester, J.T., S.K.R. Karnati, Z. Yu, M. Morrison, and J.L. Firkins. 2004. Development of an assay to quantify rumen ciliate protozoal biomass in cows using real-time PCR. J. Nutr. 134:3378–3384.

Taghizadeh, A., M. Danesh Mesgaran, R. Valizadeh, F. E. Shahroodi, and K. Stanford. 2005. Digestion of feed amino acids in the rumen and intestine of steers measured using a mobile nylon bag technique. J. Dairy Sci. 88:1807-1814.

Vanzant, E. S., R. C. Cochran, and E. C. Titgemeyer. 1998. Standardization of in situ techniques for ruminant feedstuff evaluation. J. Anim. Sci. 76:2717–2729.

Wanderley, R.C., J.T. Huber, Z. Wu, M. Pessarakli, and C. Fontes Jr. 1993. Influence of microbial colonization of feed particles on determination of nitrogen degradability by in situ incubation. J. Anim. Sci. 71:3073–3077.

Weakley, D. C., M. D. Stern, and L. D. Satter. 1983. Factors affecting disappearance of feedstuffs from bags suspended in the rumen. J. Anim. Sci. 56:493–507.

White, R.R., Y. Roman-Garcia, J.L. Firkins, P. Kononoff, M.J. VandeHaar, H. Tran, T. McGill, R. Garnett, and M.D. Hanigan. 2017a. Evaluation of the National Research Council (2001)

dairy model and derivation of new prediction equations. 2. Rumen degradable and undegradable protein. J. Dairy Sci. 100:3611–3627.

White, R.R., Y. Roman-Garcia, J.L. Firkins, M.J. VandeHaar, L.E. Armentano, W.P. Weiss, T. McGill, R. Garnett, and M.D. Hanigan. 2017b. Evaluation of the National Research Council (2001) dairy model and derivation of new prediction equations. 1. Digestibility of fiber, fat, protein, and nonfiber carbohydrate. J. Dairy Sci. 100:3591–3610.

White, R.R., P.J. Kononoff, and J.L. Firkins. 2017c. Technical note: Methodological and feed factors affecting prediction of ruminal degradability and intestinal digestibility of essential amino acids. J. Dairy Sci. 100:1946–1950.

Whitehouse, N.L., C.G. Schwab, and A.F. Brito. 2017. The plasma free amino acid dose-response technique: A proposed methodology for determining lysine relative bioavailability of rumen-protected lysine supplements. J. Dairy Sci. 100:9585–9601. doi:10.3168/jds.2017-12695.

Wu, S.H., and A. Papas. 1997. Rumen-stable delivery systems. Adv. Drug Deliv. Rev. 28:323–334.

Yang, J.H., G.A. Broderick, and R.G. Koegel. 1993. Effect of Heat Treating Alfalfa Hay on Chemical Composition and Ruminal In Vitro Protein Degradation1. J. Dairy Sci. 76:154–164.

Tables

Table 5-1. Summary Statistics of Dataset Used for Evaluating A, B, and C Protein Fractions

Variable	N. Obs	Mean	SD	Min	Max
NDF, % DM	325	46.39	15.28	8.0	87.8
CP, % DM	744	25.60	16.90	3.9	97.0
A Fraction, % CP	941	33.67	18.61	0.0	93.2
B Fraction, % CP	941	55.89	21.73	4.9	100.0
C Fraction, % CP	737	10.44	12.42	0.0	86.5
Degradation Rate, %/h	941	10.10	8.19	0.2	64.0
Grind, mm	774	10.14	6.12	1.0	25.0
Pore Size, µm	883	20.11	9.82	1.0	39.0
Height, mm	182	15.22	5.67	3.0	42.0
Width, mm	182	8.53	4.48	2.5	35.0
Area, mm^2	862	34.76	24.17	1.0	74.0
Density, g/mm^2	850	21.79	20.95	1.0	69.0
Calculated C Fraction, % CP	941	10.43	12.79	0.0	91.0
Lag Time, min	47	0.77	0.87	0.0	4.5
Sample:Bag Area Ratio, mg/mm^2	846	0.90	1.35	0.0	9.0
Incubation Time, h	822	58.90	34.00	18.0	336.0

Table 5-2. Summary Statistics for Dataset Reporting RUP Digestibility Values

Variable	N. Obs	Mean	SD	Min	Max
RUP Digestibility, % RUP	1,344	75.4	20.6	1.24	100
CP, % DM	1,344	29.9	18.2	2.4	99.6
ADF, % DM	618	17.6	11.1	0.0	86.0
NDF, % DM	654	34.2	17.8	0.66	82.7
RUP, % DM	994	41.9	23.4	1.2	100
ADIN, % N	329	11.1	12.9	0.0	105.6
NDIN, % N	209	24.2	15.3	2.4	89.6
Incubation Time, h	1,255	15.9	12.6	0.0	96.0
Acid Hydrolysis Time, h	977	1.62	0.934	0.0	6.0
Pore Size, µm	1,211	42.7	10.1	9.0	55.0

Table 5-3. Model of A, B, and C Fraction Protein and Kd As Influenced by Feed and Methodological Factors

	A			B			C			Kd		
Variable	**Estimate**	**SE**	**P-value**	**Estimate**	**SE**	**P-value**	**Estimate**	**SE**	**P-value**	**Estimate**	**SE**	**P-Value**
Intercept	61.7	17.6	<0.001	76.6	19.6	<0.001	-35.2	10.9	0.002	23.2	5.7	<0.001
Area, mm^2	-0.391	0.19	0.037				0.394	0.14	0.005	0.143	0.085	0.098
Grind, mm	0.435	0.38	0.26							-0.794	0.20	<0.001
Pore size, µm	1.52	0.57	0.008	-1.84	0.68	0.007	0.197	0.18	0.268			
Incubation time, h	-0.177	0.14	0.22	-0.249	0.11	0.027	0.333	0.10	0.001			
Sample:bag area ratio, mg/mm^2	2.01	1.21	0.10	-6.27	3.62	0.085	4.98	1.71	0.004	3.80	0.80	<0.001
NDF, % DM	-0.463	0.17	0.008	-0.177	0.23	0.45	0.685	0.11	<0.001	-0.235	0.055	<0.001
CP, % DM	-1.05	0.49	0.035	0.485	0.35	0.17	0.379	0.29	0.19	-0.0661	0.14	0.636
CP Interactions:												
Incubation time, h	0.0189	0.0063	0.003				-0.012	0.0040	0.0032			
Pore size, µm	-0.0208	0.011	0.068	0.0350	0.014	0.016	-0.011	0.0055	0.041			
Area, mm^2										0.039	0.0088	<0.001
Grind, mm	-0.0312	0.017	0.064							-0.0077	0.032	0.014
Sample:bag area ratio, mg/mm^2	-0.131	0.047	0.006	0.121	0.067	0.074				-0.206	0.032	<0.001
NDF Interactions:												
Pore size, µm	-0.0190	0.0079	0.017	0.0225	0.0088	0.012						
Area, mm^2	0.00680	0.0035	0.054				-0.0088	0.0025	<0.001			
Sample:bag area ratio, mg/mm^2				0.0849	0.051	0.098	-0.10	0.028	<0.001			
Fit Statistics:												
N	277			277			277			277		
σ_s^2 [1]	10.2			12.0			8.59			8.26		
σ_e^2 [2]	7.07			8.32			4.19			3.35		
Ratio of σ_s^2 to σ_e^2	1.45			1.44			2.05			2.46		

[1]Square root of the estimated variance associated with study

[2]Square root of the estimated variance associate with residual error.

Table 5-4. Sensitivity Analysis for Predicting A Fraction Protein in Various Feedstuffs.

	Changed Variables[1]		
Feedstuff	Pore Size, µm	Incubation Time, h	Ratio[2]
Legume Hay	5.04	23.2	0.22
Grass Hay	1.13	23.4	-1.81
Barley Grain Rolled	29.0	17.8	10.3
Fresh Grass	6.78	24.6	0.69
Soybean Meal	-4.28	15.4	-2.21
Corn Grain Dry Ground	29.4	12.9	10.8
Grass Silage	4.50	38.4	-1.38
Legume Silage	9.17	36.5	0.92
Canola Meal	-6.61	14.8	-3.84

[1]Values are percentage change between predictions using +1 SD and -1 SD for each feed variable, with all other variables in the model as the mean from the dataset.

[2]Sample size to bag area ratio (mg/mm^2).

Table 5-5. Model of RUP Digestibility as Influenced by Feed and Methodological Factors

Variable	Estimate	SE	P-value
Intercept	60.0	12.6	<0.001
ADF	-0.235	0.122	0.056
Acid Incubation Time, h	-4.01	2.01	0.052
Method = Mobile Bag	12.0	3.1	<0.001
Fit Statistics			
N	280		
σ_s^2 [1]	8.98		
σ_e^2 [2]	6.42		
Ratio of σ_s^2 to σ_e^2	1.40		

[1]Square root of the estimated variance associated with study

[2]Square root of the estimated variance associate with residual error.

Chapter 6: Meta-Analysis of Endophyte-Infected Tall Fescue Effects on Cattle Growth Rates

Published as: Liebe, D.M., and R.R. White. 2018. Meta-analysis of endophyte-infected tall fescue effects on cattle growth rates. J. Anim. Sci. 96:1350–1361. doi:10.1093/jas/sky055.

Abstract

The objective of this study was to quantitatively summarize literature reporting endophyte-infected (*Neotyphodium coenophialum*) tall fescue (*Festuca arundinacea*) effects on cattle average daily gain (ADG). This meta-analysis evaluated endophyte infection level, weather, and forage yield using a literature dataset of 138 treatments from 20 articles. Three infection level measurements were tested: endophyte infection as a percentage of infected tillers (E%); ergovaline concentration in ppb ([E]); and total ergot alkaloid concentration ([TEA]). Three types of weather variables were used: base values (temperature, humidity, and relative humidity), weather indices (heat index and temperature-heat index (THI)), and novel weather variables based on time of year which accounted for the duration of weather effects. Mixed effect models, weighted by the inverse of the standard error (SEM), including a random effect of study were built for each factorial combination of measurement method and weather variable group. Because many studies were missing SEM, two datasets were used: one containing only data with SEM reported and one that also included missing-SEM data. For the complete-SEM dataset (CSD), models were

weighted by 1/SEM. In the missing-SEM dataset (MSD) the mean reported 1/SEM was assigned as the weight for all missing SEM treatments. Although 18 initial models were created (2x3x3 factorial approach), the backward stepwise derivation resulted in models that included only endophyte infection level, suggesting a negative relationship between infection level and ADG. The CSD models predicted ADG to decrease 39 and 33 g/d with each increase of 100 ppb of [TEA] and [E], and by 39 g/d for each increase of 10% E%. In the MSD dataset, predicted ADG decreased by 39 and 33 g/d with each increase of 100 ppb of [TEA] and [E], and by 47 g/d for each increase of 10% E%. All relationships reported had $P<0.05$. After visual inspection of the data, piecewise regression was used to identify an infection threshold (IT) of 60 ppb [E] and 11 E%, where the effect of infection level was constant on either side of the IT. The ADG was 40% and 49% greater for infection levels below the IT for [E] and E%, respectively. Across THI values in the analysis, ADG decreases ranged from 11.2 to 45.0% for cattle grazing endophyte-infected tall fescue compared to non-ergot alkaloid endophyte infected tall fescue. Pasture E%, [E], and [TEA] have a negative relationship with ADG in growing cattle, and increasing temperature decreases ADG when infection level is greater than the IT.

Keywords: cattle, endophyte, ergovaline, fescue, growth, meta-analysis

Introduction

Non-ergot alkaloid endophyte-infected (**NEA**) tall fescue is an alternative to toxic wild-type endophyte infected tall fescue because it has greater stand persistence than endophyte-free tall fescue (**EFF**; Clay, 1988). Accordingly, cattle grazing NEA fescue have greater ADG than those grazing EFF (Parish et al., 2003; Nihsen et al., 2004). A challenge with adopting NEA tall fescue cultivar is the initial investment required to convert to novel cultivars; time required for transition is estimated at 3 to 7 years (Bouton et al., 2002; Gunter and Beck, 2004; Beck et al., 2008). Without planning tools to evaluate this transition, evaluating economic viability is difficult. Quantitative cattle performance expectations on different cultivars is necessary for such planning.

Because of the large number of cultivars and the limited number of studies on each, it is more feasible to evaluate cattle responses to endophyte infection level. Three infection level metrics are reported: endophyte percentage (**E%**) measured as the proportion of infected tillers; concentrations of ergovaline (**[E]**); and total ergot alkaloid concentration (**[TEA]**). The literature on endophytes currently contains studies using E%, [E], and [TEA], and production expectations must also account for these variable reporting methods.

The objective of this study was to develop equations to describe performance of cattle grazing infected tall fescue pastures using meta-analysis of literature reporting on endophyte infection level and cattle ADG. Weather variables were investigated for effects on growth, and novel weather variables were created to better model the cyclic nature of climate's effect on endophyte growth. A secondary objective was to investigate the possibility of an infection threshold (**IT**), where infection effect differs in relation to IT. It was hypothesized that ADG would be decreased with increasing infection level and that other factors, such as weather, would exacerbate these effects.

Materials and Methods

To properly assess the effects of grazing endophyte-infected tall fescue on stocker cattle ADG, a dataset was compiled, and the papers were screened based on the inclusion and exclusion criteria detailed below. Weather variables not reported in papers were obtained from public weather station data repositories based on the dates and durations of the appropriate studies. Weather indices were calculated to more succinctly describe weather variables. To account for changes in environment over time, novel parameters for the effect of study start date, temperature, and duration were created using sinusoidal curves to mimic the changes in endophyte growth over the year.

All models were built using the lmer package (Kuznetsova et al., 2013) in R version 3.1.0. (R Core Team, 2014) and data was weighted based on study standard error (SEM). For data reported without SEM, a second dataset was tested with unreported SEM data being given the average SEM of all other studies. This resulted in two sets of data for model-building. Initial models were designed using a 2x3x3 factorial approach with factors being weighting strategy, weather variable, and measurement type. The 18 resulting models were derived using backward, stepwise elimination multiple regression and tested for significance. Models were compared using corrected Akaike information criterion (**AICc**) and estimated variance of study and error. Only models that resulted in significant parameters are reported as results of the meta-analysis.

Data Collection

Data were collected from peer-reviewed published journal articles through a comprehensive literature search. Key words used to search for relevant articles were: fescue, endophyte, infected, ergot alkaloid, toxicosis, ergovaline, gain, and cattle. Subsequent titles were searched from the references of recovered articles, allowing an increased search space. To be included in the dataset, papers needed to be published before 2017 in English and include data for ADG using live BW. Articles reporting the

effects of tall fescue grazing on animals other than cattle were also excluded. Test animals that were fed an endophyte-infected seed concentrate were not considered for inclusion in the dataset because the focus of the study was on grazed tall fescue.

The complete dataset included 138 treatment means from 20 articles. Summary statistics for key variables reported within the dataset are provided in Table 6-1 and a listing of article citations is included in Supplementary Table 6-1. Of the 20 articles, 6 included growth data for heifers and cows instead of, or in addition to, data on steers. The growth curves of these animals would differ, but the addition of a random effect for trial should have accounted for this variation. All animals included were *Bos taurus* species with a combination of predominantly Angus and Angus crossbreeds.

Weighting Strategies

A challenge with meta-analysis is incomplete reporting of SEM. Often, papers that fail to report SEM are removed entirely from the dataset. However, not weighting these studies at all is likely inappropriate because they do have some value, it is just difficult to determine the exact value because the precision of the reported means is unknown. As an alternative to weighting the studies with a value of 0, studies were weighted equally to the average study in the dataset. Because this weighting approach is

not conventional, two approaches were employed: the complete-SEM dataset (**CSD**), including only studies with reported SEM and the missing-SEM dataset (**MSD**), including all data with the mean SEM given to all incomplete data.

Data were weighted for 1/SEM to limit the weight of studies with very small trials and error. Optimal weighting using this method has been previously documented (Roman-Garcia et al., 2016; White et al., 2016) and works well with mixed models (St-Pierre, 2001). All papers included in the dataset were checked to determine whether the statistical analysis used a fixed or mixed effect model. The weighting factor was calculated as each study's SEM divided by the mean SEM of the data within analysis type and taking the reciprocal so the lesser the SEM of the study, the greater weight in the model. Dividing by the mean reciprocal SEM normalized the weighting factors to 1 irrespective of analysis type, effectively standardizing the weights across fixed effect or mixed models (White et al., 2015; Roman-Garcia et al., 2016). In studies that reported SEM as less than one-fourth the mean SEM, the SEM was set to one-fourth of the mean SEM across all studies to prevent over weighting (Firkins et al., 2001; Roman-Garcia et al., 2016; White et al., 2016). The curtailing of SEM resulted in 5.2% and 3.6% of errors being trimmed for CSD and MSD, respectively. The SEM trimming was conducted separately for mixed and fixed effect models, because mixed-models had greater SEM. The result of this

cleaning was weighting factors equivalent to the reciprocal SEM without bias for statistical method, without overweighting extremely precise studies.

Variables to Represent Weather

Raw weather variables of interest within this study included mean, maximum, and standard deviation of temperature by month. Mean humidity was also considered. Weather data were gathered for each study from the National Centers for Environmental Information's local climatological database (National Climatic Data Center (NCDC)). Data were downloaded from the database and mean temperatures, dew points, and maximum temperatures for the duration of the studies were recorded. The Relative Humidity (**RH**), which the sourced weather data did not supply, was calculated using the formula:

[1] $$RH,\% = 100 \times e^{\frac{17.625 \times TD}{243.04 \times TD}} / e^{\frac{17.625 \times T}{243.04 \times T}},$$

where *T* is Temperature in degrees Celsius and *TD* is the Dew Point (Alduchov and Eskridge, 1996).

As a means to potentially reduce complexity in the models, and better describe weather effects, various indices were employed to represent the effect of weather on ADG. Because previous work has suggested cattle consuming endophyte infected tall fescue are hypersensitive to heat, specific focus was placed on variables

that would reflect heat stress, including: heat index (HI) and temperature-humidity index (THI).

The HI was calculated as:

$$HI = 0.5 \times \left\{ T + 61.0 + \left[\left(T - 68.0 \right) \times 1.2 \right] + \left(RH \times 0.094 \right) \right\}, \quad [2]$$

where *T* is Temperature in degrees Celsius and *RH* is Relative Humidity as a percentage (Rothfusz and Headquarters, 1990).

A metric combining temperature and humidity is often employed to more accurately represent the heat load on the animal. The most common technique, used in The Livestock Weather Safety Index (LCI, 1970), is the temperature-humidity index (THI). The THI was calculated using the equation:

$$THI = 0.8 \times T + [\frac{RH}{100} + (T - 14)] + 46.4, \quad [3]$$

where *T* is Temperature in degrees Celsius and *RH* is Relative Humidity as a percentage. The THI has been shown to effectively indicate heat stress in cattle, and continues to be refined with more recent adjustments for wind speed and solar radiation (Mader et al., 2006). Although wind speed and radiation have been shown to influence heat stress, these data were not available in the current study. However, Mader et al. (2006) suggest that THI is an adequate representation of thermal load with or without adjustment for wind speed and radiation.

It is established knowledge that increased endophytic infection leads to greater vasoconstriction, making it more difficult for animals to dissipate heat. Research has also shown that E% and [E] change throughout the year relative to maximum infection potential, regardless of location (Ju et al., 2006). Given the concurrent change in temperature and infection percentage, it was hypothesized that a variable representing weather impacts on tall fescue ADG responses should also represent the temporal behavior of stand infection. Table 6-2 includes data from Ju et al. (2006), which was used to derive a curve to describe the effect of time of year on endophyte levels. A curve for E% and [E] effects as a percentage of max infection level were derived using nonlinear least-squares regression and the following formula:

$$Effect\% = A \times \sin(2\pi ft + \varphi) + \beta \quad [4]$$

where *A* is Amplitude, *f* is Frequency, *t* is Time in months, φ is the Phase and β is the y-intercept ($P < 0.001$ for all parameters). By fitting sinusoidal curves to E% and [E] the effects of infection level over time were quantified. Trial duration was accounted for by integrating over the curve in 1/30th-month intervals from the start date of the trial to determine the mean effect of infection level experienced by the animals on any specific trial. The

equation for average endophyte level over a given duration (**AEL**) was:

$$AEL = \frac{1}{duration} \sum_{i=start}^{start+duration} Effect\%_i \quad [5]$$

where *start* is the start month of the trial and *duration* is the length of the trial as a fraction of a month. This resulted in a number that, in theory, more accurately described the combined thermal and endophyte consumption effects experienced within each treatment group.

Infection Threshold

The IT, defined as the threshold to produce clinical tall fescue toxicosis, has been measured as 300 to 750 pbb [E] (Hovermale and Craig, 2001; Tor-Agbidye et al., 2001; Craig et al., 2014). Along with the three common measurement methods previously described: E%, [E], and [TEA], an IT was derived by piecewise regression to analyze the effect of splitting the response surface. This approach fits one slope to infection level greater than a threshold and a different slope below the threshold. The threshold value was identified by iteratively testing models over the sample space of each infection level measurement to determine the optimal threshold. The IT value was chosen based on the infection level identified to generate the smallest AICc

(Hurvich and Tsai 1989). Once a threshold was identified, forward stepwise regression was used to test for significance of weather variables.

Model Derivation Procedure

All models were derived using the lmer (Bates et al., 2017) function in R version 3.1.0. (R Core Team, 2017). Growing cattle ADG was used as the response variable for all models. All explanatory variables are summarized in Table 6-1. For the set of 18 factorial models, models were refined through backward stepwise elimination multiple regression as described in Roman-Garcia et al. (2016) and White et al. (2016). The variable with the greatest non-significant P-value ($P > 0.05$) was iteratively eliminated from the model unless the term was a linear term with a significant quadratic effect ($P \leq 0.05$). The piecewise models were derived as previously described. After the threshold infection percentage was identified, forward stepwise regression was used to test for parameter significance.

Final models were also checked to ensure all variance inflation factors (VIF) factors were acceptable. The VIF measures the severity of the multicollinearity in a regression and the resulting severity of the inflation of variance due to this collinearity in the parameter estimations. The square root of the VIF indicates the inflation of the variance of the parameter

estimate compared with an ideal scenario of no collinearity. This study used a VIF cutoff of VIF < 10 for linear factors not involved in interactions or quadratic terms, meaning the variance due to collinearity was one factor greater than that of a regression with no collinearity in parameter estimates. The cutoff for quadratic terms, interaction terms and linear terms involved in either quadratic or interaction terms was VIF < 100. These cutoffs are in line with current research practices, although no clear rules have been established (Roman-Garcia et al., 2016).

Evaluating Model Performance

Models were evaluated based on AICc and root estimated variance due to error ($\hat{\sigma}_{\varepsilon}$, i.e. the estimated variance for error) and study ($\hat{\sigma}_{s}$). Both $\hat{\sigma}_{\varepsilon}$ and $\hat{\sigma}_{s}$ are expressed as a percentage of the dependent variable mean. The AICc was the predominant indicator of model performance because all models were derived from the same size datasets using the same response variable. The reporting of RMSE was avoided because of the inclusion of a random effect of study. When a random study effect is included and models are chosen based on RMSE, the models perform poorly on new data with different studies because they underestimate error (Boerman et al., 2015). All variables were assessed for simple correlation to evaluate collinearity. Residual plots used data adjusted for the random effect of study, and the linear regressions were weighted for the

SEM to check for patterns in the data. Slope and mean bias as a percentage of the dependent variable mean were recorded and any model displaying a significant bias was adjusted or removed. When models were comparable, the one with more observations was deemed more desirable.

Results and Discussion

Equation Descriptions

A set of linear models was derived using a 2x3x3 factorial approach with factors for weighting strategy, weather variable, and measurement type. The final equations for each of the 18 factor combinations can be viewed in Table 6-3. When different factorial combinations lead to the same equation, the resulting equation was only listed once.

Average Daily Gain Responses

After backward elimination, infection level was the only significant variable in all tested models. The 3 equations using E%, [E], and [TEA] were considered the basis of comparison for all other models and can be found in Table 6-3. Each of the three measurement methods had significant ($P < 0.05$), negative relationships with ADG. Using the derived equations from the CSD: for a 10% increase in E%, a 100 ppb increase in [E] and a 100 ppb

increase in [TEA] was associated with a decrease in ADG of 38 g (5.3%), 33 g (4.4%), and 39 g (6.6%) per day, respectively. Barker et al. (2009) noted that a rule of thumb for E% is a loss of 45 g/d for each 10% increase in E%, which is marginally greater but generally in line with the CSD estimates.

Measurement Methods

Equations utilized one of the three base methods for estimating endophyte levels in tall fescue: [E] in parts per billion, E% and [TEA] in parts per billion. From 1983 to 1993, the popular method for measuring endophyte levels involved taking samples from tillers in the field and inspecting the samples for the presence of endophytes. The results of this microscopy work were reported as infected tillers as a percentage of total samples taken (E%). This method was used until Rottinghaus et al. (2001) developed a method of using High-Performance Liquid Chromatography (HPLC) to measure [E], the most abundant toxic alkaloid produced by the endophytes (Rottinghaus et al., 1991). This HPLC measurement remains the most common measurement method in research on endophyte-infected tall fescue due in part to the recent creation of non-ergot alkaloid endophyte strains which do not produce ergovaline.

Non-ergot alkaloid endophytes present in tall fescue still appear using microscopy, but do not create the harmful ergot

alkaloids that are toxic to cattle. As a result, novel cultivars cannot be reliably distinguished from endophyte infected tall fescue when using the E% method. Because of the growth in popularity and interest in endophyte-infected tall fescue that does not produce ergot alkaloids, HPLC is distinctly more effective at properly quantifying the infection level key to decreased ADG in cattle. Using E% as the measurement method (Eq. 6) yielded the greatest slope and mean error as a percentage of the mean SE in the CSD. In terms of variation, the E% equation parameter estimate CV was greatest among all equations, with CV of 130, 93, and 77% for Eq. 6, 7, and 8, respectively. When coupled with inability to differentiate toxic from novel, nontoxic cultivars, these observations suggest E% is a poor indicator of ADG.

The [E] has been cited as making up between 85 to 97% of [TEA] in tall fescue (Lyons et al., 1986). Using the models derived herein, the proportion of ADG loss associated with [E] and [TEA] can be calculated. This calculation was done using the following simple arithmetic and Eq. 6 and 7.

[6] $$ADG, kg/d = ([E] \times -3.254 \times 10^{-4}) + 0.7443$$

[7] $$ADG, kg/d = ([TEA] \times -3.934 \times 10^{-4}) + 0.5969$$

[8] $$[E] = [TEA] \times 1.209 + 453.0$$

Using this equation, the effect of any given [TEA] in terms of an [E] can be derived. Taking into consideration the almost identical slopes of Eq. 6 and 7, the difference in the effect of an increase in [TEA] versus [E] is approximately 20%. This suggests that ergovaline accounts for at most 80% of the effect on ADG of the [TEA] produced by a given tall fescue cultivar. 80% reflects a maximum because there could be effects common between [E] and [TEA] such as shade, water availability, mineral supplementation, or some other factor. Note that this percentage is different than the aforementioned 85 to 97% for the proportion of ergovaline in the [TEA]. Although [E] makes up at most 80% of the effect on ADG, the remaining 3 to 15% of ergot alkaloids, along with factors like genetics or pasture management, are likely responsible for the other 20% of the ADG response.

Because there is a linearly correlated relationship between the effect of [E] and [TEA] ($r \approx 1$, $P < 0.001$), measuring one should provide a good proxy for the other at any concentration. Although the difference in slope between the two equations was negligible, the proportion of effect attributed to [E] may decrease at high concentrations; [E] makes up about 65% of the total effect at [TEA] of 800 ppb.

Weather Variables

Although variables were systematically added back into the final models to identify the best combination, no weather variables were identified as significant for the three measurement methods using the CDS ($P>0.05$). There are several reasons why the weather variables did not appear to have a significant effect on ADG. By utilizing a mixed effects model, publication was a random variable to remove study-to-study variation. Because of the small number of trials included for each measurement type, 3, 5 and 6 publications for [E], E% and [TEA] in the CSD and 4, 8, and 7 publications for the MSD, this random effect may explain most of the weather variation. In the 14 different publications included in the CSD, states in which trials were performed were Arkansas, Oregon, Kentucky, Georgia, Tennessee, Alabama, Oklahoma, Louisiana and North Carolina. The state with the most studies in the dataset was Georgia, with 5 different treatment groups. By isolating points from only Georgia, the resulting box plot (Figure 6-1) shows a visual relationship between THI and ADG. When only these data were analyzed, a 75% decrease in ADG was observed as THI moved up approximately 3.5 units ($P<0.01$). This relationship may be an overestimate due to limited sample size; however, this illustrates the possibility of weather variables having significance that was not apparent in Eq. 6, 7, and 8.

A duration by starting month interaction variable was included at the beginning of backwards regression in each model using the raw weather data factor, with the effect being removed due to non-significance each time. The *P*-value of duration by starting month when combined with each measurement method alone was only significant ($P = 0.014$) for [TEA]; however, the AICc value was greater in this model and was considered inferior when compared to predictions using [TEA] alone. For models using [E] and E%, adding a duration by starting month interaction variable either showed a tendency towards significance or was non-significant ($P = 0.069$ and 0.8114, respectively) in the CSD. This result could be due to the small sample size of starting months.

A reason for failure to identify significant weather effects in most models might be that THI and HI were not within ranges typically considered to be severe. Because THI and HI were calculated using the average temperatures provided, the indices may not have properly captured times of heat stress that may have occurred intermittently during the trials. Eigenburg et al. (2005) cites a maximum threshold THI of 74 for "normal" conditions; there were no trials with a calculated THI greater than 73. Although THI was assumed using average temperatures and RH would be a good proxy for severity of heat stress experienced, this may not have been the case in the current dataset.

The research goal was to utilize previous data to extrapolate possible weather and infection curves, then use those curves to better account for the average effect experienced by the animal using integration to account for duration of study. The CSD showed a significant (P<0.001) negative correlation between duration of the studies included and the starting month (Table 6-4), meaning that longer studies tended to start earlier in the year and shorter studies at the end of the year. This makes sense from a logical standpoint because the prime grazing season is predominantly in the spring and summer. However, the correlation between duration and starting date made it hard to test AEL.

Piecewise Regression Model Performance

Studies have identified an [E] threshold value needed to produce clinical signs of toxicity. The range of suggested IT was large, but these studies suggest a differing effect of infection greater and less than a threshold (Hovermale and Craig, 2001; Tor-Agbidye et al., 2001; Craig et al., 2015). To evaluate whether this relationship existed in the current data, piecewise regression was used to test models with two linear response surfaces based on IT. Previous studies have cited thresholds for clinical tall fescue toxicosis (Hovermale and Craig, 2001; Tor-Agbidye et al., 2001; Craig et al., 2015); however, these thresholds have not been derived based on quantitative summary of

the available literature. In this analysis, IT models were only derived for [E] and E% due to the small sample size of the [TEA] data in the CSD and MSD relative to any threshold. Using the IT, infection levels less than the IT resulted in a 40 and 49% increase in ADG for [E] and E%, respectively. The AICc graphs and visuals of the gaps in data are shown in Figure 6-2 and suggest that infection levels do not appear uniformly in the literature, but rather in discrete ranges.

Results of the IT models are shown in Table 6-5. For the [E] IT, a negative relationship was significant for the infected by average temperature interaction variable. The ADG decreased as temperature increased when the tall fescue infection level was greater than the IT. This is logical due to temperature exacerbating the effects of heat stress caused by cattle grazing infected tall fescue. The fact that temperature alone did not show a negative relationship with ADG also makes sense considering that increased temperature would indicate better growing conditions for forages, leading to greater ADG in the absence of infected tall fescue (lower than the IT). Despite the increase in predictor variables, the AICc value for the infection by average temperature model was lower than for Eq. 7 using the same data. Using THI as the weather variable in place of average temperature also yielded significant results and a negative relationship between the infected by THI interaction variable (Table 6-5).

The use of a AEL variable to account for duration of study improved the results of the IT model, with the intercept, infected, AEL, and AEL by infected interaction variable all showing significance ($P<0.05$). As with the raw weather indices, the interaction of AEL by infected had a negative relationship with ADG. As cattle graze infected tall fescue in times of greater infection levels, ADG decreases. In times where AEL would be high but the pastures are not infected, ADG increases.

As noted previously, research has shown a change in infection rates over the year, with greater infection levels occurring in the warmer months (Barker et al., 2009). The fact that both infection levels and temperature rise in parallel could be another reason why a significant interaction between infection level and temperature was not detected. In the models, some of the effects of increasing infection rates may be attributed to rises in weather variables that compound on an animal's ability to dissipate heat. As such, within the data it is likely that infection level represents both the actual infection level of the forage but also the season in which the measurements were collected.

Weighting Strategies

In the analysis of the MSD, a larger dataset was utilized by including data that did not report SEM of ADG responses. The value

and significance of parameter estimates derived from the CSD and MSD datasets differed (Table 6-3). At this point, it is unclear whether these inconsistencies are caused by the weighting procedure (filling in missing SEM as equal to the mean SEM) or by the added data revealing new relationships. Future work on synthetic data evaluating different ways to handle missing SEM data are needed to better understand the best way to deal with this data challenge.

Missing-SEM Data. Results for models derived using the MSD are listed in Table 6-3. The estimation for ADG with no infection (intercept) changed +8.8, -1.9, and +31.2% from the estimations using the CSD for E%, [E], and [TEA], respectively. Changes in slopes for E%, [E], and [TEA] were +19.6, +0.2, and -0.6% compared to the CSD equations. Both the intercept and parameter estimations increased from Eq. 8 to Eq. 11, likely because the mean ADG within the MSD dataset was greater than the CSD dataset (0.6771 and 0.5650 kg/d).

Implications

Total endophyte concentration has been shown to increase starting in the spring, with lower concentrations in the winter (Ju et al., 2006). A curve for [E] was created using year-long measurements of endophyte concentration because it was assumed

that the fraction of ergot-alkaloid producing endophytes is proportional to total endophyte concentration throughout the year, which is reasonable considering E% fluctuates over a year in a cyclic manner as well. Studies have reported cyclic changes in [E] coinciding with head emergence and seed development in perennial ryegrass, and also suggest the highest [E] occurs during peak host growth in the summer (Reed et al. 2011). The values in Table 6-2 were used to calculate a curve for percent [E] of maximum exhibited in the year for each day of the year. For example, on March 1st, [E] in a field will be 34% of the year's potential maximum [E]. Using three levels of field infection, 0, 100, and 300 ppb [E], differences in cumulative growth were calculated for cattle spring and fall calving seasons. Cattle BW gain was 33 and 19% greater in the 4-month period from October through January for cattle on 0 ppb [E] fields compared to 300 and 100 ppb, respectively. From June through September, BW gain was 128% greater for cattle on 0 ppb [E] fields compared to both 300 and 100 ppb. A 0 ppb [E] field compared to a 300 ppb [E] results in an extra 20 and 65 kg BW gain in the fall and spring calving seasons, respectively. The cumulative BW gain over 1 year, October through January, and June through September are graphed in Figure 6-3. Value per kg of BW gain was calculated using Missouri purchase prices from 2012 to 2017 of 249-kg BW, medium-framed, number 1 steers in September for fall grazing and February for spring grazing. The values calculated

were $1.27/kg and $1.21/kg in US dollars for fall and spring purchase prices, the growth difference among these systems would give the 0 ppb pasture a benefit of $25 and $79 per head per season. These values are considerable, but very much aligned with the data compiled. In two studies that reported ADG for cattle grazing tall fescue on both sides of the IT, the average depression in ADG for each was 356 and 354 g/d (Parish et al. 2013; Parish et al. 2003), whereas the model predicted average ADG depressions of 262 and 431 g/d. The effect of endophyte-infection level on ADG should be an important consideration for farmers when grazing cattle.

When deciding whether to plan for spring or fall calving (cow/calf) or grazing (stocker) seasons, the effects of endophyte infection level should be considered. One study specifically comparing the profitability of fall-calving versus spring-calving herds and concluded that, when grazing endophyte-infected fescue, it is more cost effective to calve in the fall (Caldwell et al., 2013). The same study went on to note that higher daily gain could be seen in spring-calving herds if the pasture was nontoxic endophyte-infected instead (Caldwell et al., 2013). Multiple studies have shown the increased profitability of calving in the fall, due in part to higher daily gain for calves (Bagley et al. 1987; McCarter et al., 1991; Gaetner et al., 1992; Henry et al., 2016). The present study also supports the idea that farmers

maintaining pastures with endophyte-infected tall fescue should consider switching to a fall calving system as a means to minimize BW gain losses.

Limitations

The biggest limitations of this meta-analysis were sample size and data structure. Two weighting strategies were employed to combat this limitation by including more incomplete studies with incomplete data reporting. Some of the recorded variables did not have consistent distributions. For example, there were no data for studies beginning in 5 months of the year, which limited ability to derive effects by change in weather. Another result was an inconsistent distribution of trial durations for each starting month, making it hard to establish a duration effect. The inconsistency of measurement methods required splitting the data into three groups, severely limiting the size of each of the training sets. Ideally, as more studies are conducted on the effects of novel endophyte-infected tall fescue cultivars, a more consistent dataset utilizing one measurement method will become available.

References

Alduchov, O. A., and R. E. Eskridge. 1996. Improved Magnus Form Approximation of Saturation Vapor Pressure. J. Appl. Meteorol. 35:601–609. doi:2.0.CO;2">10.1175/1520-0450(1996)035<0601:IMFAOS>2.0.CO;2. Available from: 2.0.CO;2">https://doi.org/10.1175/1520-0450(1996)035<0601:IMFAOS>2.0.CO;2

Bacon, C. W. 1995. Toxic endophyte-infected tall fescue and range grasses: historic perspectives. J. Anim. Sci. 73:861–870. Available from: https://www.ncbi.nlm.nih.gov/pubmed/7608021

Bagley, C. P., J. C. Carpenter, J. I. Feazel, F. G. Hembry, D. C. Huffman, and K. L. Koonce. 1987. Influence of Calving Season and Stocking Rate on Beef Cow-Calf Productivity. J. Anim. Sci. 64:687–694. doi:10.2527/jas1987.643687x. Available from: http://dx.doi.org/10.2527/jas1987.643687x

Barker, D. J., N. S. Hill, J. G. Andrae, H. A. Fribourg, D. B. Hannaway, and C. P. West. 2009. Measuring the Endophyte—Plants, Fields, and Farms. In: Agronomy Monographs. Available from: http://dx.doi.org/10.2134/agronmonogr53.c15

Boerman, J. P., J. L. Firkins, N. R. St-Pierre, and A. L. Lock. 2015. Intestinal digestibility of long-chain fatty acids in lactating dairy cows: A meta-analysis and meta-regression. J. Dairy Sci. 98:8889–8903. doi:10.3168/jds.2015-9592. Available from: http://dx.doi.org/10.3168/jds.2015-9592

Bouton, J. H., G. Latch, N. S. Hill, C. S. Hoveland, M. A. McCann, R. H. Watson, J. A. Parish, L. L. Hawkins, and F. N. Thompson. 2002. Reinfection of tall fescue cultivars with non-ergot alkaloid--producing endophytes. Agron. J. 94:567–574. Available from: https://dl.sciencesocieties.org/publications/aj/abstracts/94/3/567

Clay, K. 1988. Fungal Endophytes of Grasses: A Defensive Mutualism between Plants and Fungi. Ecology. 69:10–16. doi:10.2307/1943155. Available from: http://dx.doi.org/10.2307/1943155

Clay, K. 1990. Fungal Endophytes of Grasses. Annu. Rev. Ecol. Syst. 21:275–297.

Craig, A. M., L. L. Blythe, and J. M. Duringer. 2014. The role of the Oregon State University Endophyte Service Laboratory in diagnosing clinical cases of endophyte toxicoses. J. Agric. Food Chem. 62:7376–7381. doi:10.1021/jf5027229. Available from: http://dx.doi.org/10.1021/jf5027229

Craig, A. M., J. L. Klotz, and J. M. Duringer. 2015. Cases of ergotism in livestock and associated ergot alkaloid concentrations in feed. Front Chem. 3:8. doi:10.3389/fchem.2015.00008. Available from: http://dx.doi.org/10.3389/fchem.2015.00008

Firkins, J. L., M. L. Eastridge, N. R. St-Pierre, and S. M. Noftsger. 2001. Effects of grain variability and processing on starch utilization by lactating dairy cattle. J. Anim. Sci. 79:E218–E238. Available from: https://dl.sciencesocieties.org/publications/jas/abstracts/79/E-Suppl/JAN00790ESE218

Gaertner, S. J., F. M. Rouquette Jr, C. R. Long, and J. W. Turner. 1992. Influence of calving season and stocking rate on birth weight and weaning weight of Simmental-sired calves from Brahman-Hereford F1 dams. J. Anim. Sci. 70:2296–2303. Available from: https://www.ncbi.nlm.nih.gov/pubmed/1506293

Gunter, S. A., and P. A. Beck. 2004. Novel endophyte-infected tall fescue for growing beef cattle. J. Anim. Sci. 82 E-Suppl:E75–82. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15471817

Henry, G. W., C. N. Boyer, A. P. Griffith, J. Larson, A. Smith, and K. Lewis. 2016. Risk and returns of spring and fall calving for beef cattle in Tennessee. J. Appl. Agric. Econ. 48:257–278. doi:10.1017/aae.2016.11.

Hoveland, C. S. 1993. Importance and economic significance of the Acremonium endophytes to performance of animals and grass plant. Agric. Ecosyst. Environ. 44:3–12. doi:10.1016/0167-8809(93)90036-O. Available from: http://www.sciencedirect.com/science/article/pii/016788099390036O

Hovermale, J. T., and A. M. Craig. 2001. Correlation of ergovaline and lolitrem B levels in endophyte-infected perennial ryegrass (Lolium perenne). J. Vet. Diagn. Invest. 13:323–327. doi:10.1177/104063870101300407. Available from: http://dx.doi.org/10.1177/104063870101300407

Hurvich, C. M., and C.-L. Tsai. 1989. Regression and time series model selection in small samples. Biometrika. 76:297–307. doi:10.1093/biomet/76.2.297. Available from: https://academic.oup.com/biomet/article-abstract/76/2/297/265326

Ju, H.-J., N. S. Hill, T. Abbott, and K. T. Ingram. 2006. Temperature Influences on Endophyte Growth in Tall Fescue. Crop Sci. 46:404. doi:10.2135/cropsci2005.0282. Available from: http://dx.doi.org/10.2135/cropsci2005.0282

Lyons, P. C., R. D. Plattner, and C. W. Bacon. 1986. “Occurrence of Peptide and Clavine Ergot Alkaloids in Tall Fescue Grass.” Science 232 (4749): 487–489.

Mader, T. L., M. S. Davis, and T. Brown-Brandl. 2006. Environmental factors influencing heat stress in feedlot cattle. J. Anim. Sci. 84:712–719. Available from: https://www.ncbi.nlm.nih.gov/pubmed/16478964

Mathews, K., and M. Haley. 2014. Livestock, Dairy, and Poultry Outlook. Economic Research Service, US Department of Agriculture. Available from: http://usda.mannlib.cornell.edu/usda/ers/LDP-M/2010s/2014/LDP-M-12-16-2014.pdf

McCarter, M. N., D. S. Buchanan, and R. R. Frahm. 1991. Comparison of crossbred cows containing various proportions of Brahman in spring- or fall-calving systems: IV. Effects of genotype x environment interaction on lifetime productivity of young cows. J. Anim. Sci. 69:3977–3982. Available from: https://www.ncbi.nlm.nih.gov/pubmed/1778810

National Climatic Data Center (NCDC). Local Climatological Data Publication | IPS | National Climatic Data Center (NCDC). Available from: https://www.ncdc.noaa.gov/IPS/lcd/lcd.html

Nihsen, M. E., E. L. Piper, C. P. West, R. J. Crawford Jr, T. M. Denard, Z. B. Johnson, C. A. Roberts, D. A. Spiers, and C. F. Rosenkrans Jr. 2004. Growth rate and physiology of steers grazing tall fescue inoculated with novel endophytes. J. Anim. Sci. 82:878–883. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15032446

Parish, J. A., M. A. McCann, R. H. Watson, N. N. Paiva, C. S. Hoveland, A. H. Parks, B. L. Upchurch, N. S. Hill, and J. H. Bouton. 2003. Use of nonergot alkaloid-producing endophytes for alleviating tall fescue toxicosis in stocker cattle. J. Anim. Sci. 81:2856–2868. Available from: https://www.ncbi.nlm.nih.gov/pubmed/14601890

Parish, J. A., J. R. Parish, T. F. Best, H. T. Boland, and C. A. Young. 2013. Effects of selected endophyte and tall fescue cultivar combinations on steer grazing performance, indicators of fescue toxicosis, feedlot performance, and carcass traits. J. Anim. Sci. 91:342–355.

doi:10.2527/jas.2011-4725. Available from: http://dx.doi.org/10.2527/jas.2011-4725

Reed, K. F. M., Z. N. Nie, L. V. Walker, and G. Kearney. 2011. Fluctuations in the concentration of ergovaline and lolitrem B produced by the wild-type endophyte (Neotyphodium lolii) in perennial ryegrass (Lolium perenne) pasture. Anim. Produc. Sci. 51:1098–1108. doi:10.1071/AN11143. Available from: http://www.publish.csiro.au/an/AN11143

Roman-Garcia, Y., R. R. White, and J. L. Firkins. 2016. Meta-analysis of postruminal microbial nitrogen flows in dairy cattle. I. Derivation of equations. J. Dairy Sci. 99:7918–7931. doi:10.3168/jds.2015-10661. Available from: http://dx.doi.org/10.3168/jds.2015-10661

Rothfusz, L. P., and N. S. R. Headquarters. 1990. The heat index equation (or, more than you ever wanted to know about heat index). Fort Worth, Texas: National Oceanic and Atmospheric Administration, National Weather Service, Office of Meteorology. 9023. Available from: http://204.227.127.200/images/oun/wxsafety/summerwx/heatindex.pdf

Rottinghaus, G. E., G. B. Garner, C. N. Cornell, and J. L. Ellis. 1991. HPLC method for quantitating ergovaline in endophyte-infested tall fescue: seasonal variation of ergovaline levels in stems with leaf sheaths, leaf blades, and seed heads. J. Agric. Food Chem. 39:112–115. doi:10.1021/jf00001a022. Available from: http://dx.doi.org/10.1021/jf00001a022

St-Pierre, N. R. 2001. Invited review: Integrating quantitative findings from multiple studies using mixed model methodology. J. Dairy Sci. 84:741–755.

doi:10.3168/jds.S0022-0302(01)74530-4. Available from: http://dx.doi.org/10.3168/jds.S0022-0302(01)74530-4

Strickland, J. R., J. W. Oliver, and D. L. Cross. 1993. Fescue toxicosis and its impact on animal agriculture. Vet. Hum. Toxicol. 35:454–464. Available from: https://www.ncbi.nlm.nih.gov/pubmed/8249272

Tor-Agbidye, J., L. L. Blythe, and A. M. Craig. 2001. Correlation of endophyte toxins (ergovaline and lolitrem B) with clinical disease: fescue foot and perennial ryegrass staggers. Vet. Hum. Toxicol. 43:140–146. Available from: https://www.ncbi.nlm.nih.gov/pubmed/11383653

White, R. R., P. S. Miller, and M. D. Hanigan. 2015. Evaluating equations estimating change in swine feed intake during heat and cold stress. J. Anim. Sci. 93:5395–5410. doi:10.2527/jas.2015-9220. Available from: http://dx.doi.org/10.2527/jas.2015-9220

White, R. R., Y. Roman-Garcia, and J. L. Firkins. 2016. Meta-analysis of postruminal microbial nitrogen flows in dairy cattle. II. Approaches to and implications of more mechanistic prediction. J. Dairy Sci. 99:7932–7944. doi:10.3168/jds.2015-10662. Available from: http://dx.doi.org/10.3168/jds.2015-10662

Young, C. A., N. D. Charlton, J. E. Takach, G. A. Swoboda, M. A. Trammell, D. V. Huhman, and A. A. Hopkins. 2014. Characterization of Epichloë coenophiala within the US: are all tall fescue endophytes created equal? Front Chem. 2(95):1-11.

Tables

Table 6-1. Summary statistics of the key variables in the study dataset, including missing-SEM data.

Item	N[1]	Mean	Median	Minimum	Maximum
Studies	20				
ADG, kg/d	138	0.683	0.618	-0.340	2.32
Initial Weight, kg	70	256	247	220	437
Final Weight, kg	70	354	346	176	687
Forage Yield, kg/ha	33	3020	2850	1350	8980
Weather					
Max Temp, ^{0}C	131	28.0	26.8	12.7	43.3
Max Temp S.D, ^{0}C	80	6.95	6.8	3.01	9.62
Average Temp, ^{0}C	135	18.4	20.7	2.83	36.0
Average Temp S.D, ^{0}C	80	6.48	6.36	2.84	9.06
Relative Humidity, %	127	63.7	60.2	52.1	77.0
Relative Humidity S.D, %	80	16.3	16.1	8.42	21.1
THI	127	63.8	66.7	48.1	73.4
Heat Index	127	64.1	68.7	41.3	77.6
Tall Fescue					
Endophyte, %	55	42.7	35.0	0.00	98.3
Ergovaline, ppb	44	196	31.5	0.00	1210
Total Ergot Alkaloids, ppb	24	145	20.0	0.00	820

[1]Number of data points in missing-SEM dataset

Table 6-2. Data for ergovaline concentrations* and average temperature by month from the missing-SEM dataset.

Month	Endo[1]	EndoMax, %[2]	Temp[3]	TempMax, %[4]
January	0.57	0.28	3.77	0.14
February	0.77	0.37	5.20	0.20
March	0.66	0.32	10.2	0.38
April	0.93	0.45	15.1	0.57
May	1.25	0.60	20.2	0.76
June	1.52	0.73	24.5	0.92
July	1.93	0.93	26.2	0.98
August	1.80	0.87	26.6	1.00
September	1.87	0.90	22.0	0.83
October	2.07	1.00	16.3	0.61
November	1.61	0.78	10.5	0.39
December	1.16	0.56	12.0	0.45

*Ergovaline data adapted from Ju et al., 2006
[1]Total Endophyte Concentration, mg/g
[2]Ergovaline as a decimal percentage of the maximum concentration recorded
[3]Temperature, °C
[4]Temperature percentage as a decimal of maximum temperature recorded

Table 6-3. Parameter estimates in models of ADG using either the complete-SEM dataset or including the missing-SEM data. Models were further divided by method used to measure infection level.

	Complete-SEM Dataset			Missing-SEM Dataset		
Item[1]	E%	[E]	[TEA]	E%	[E]	[TEA]
Eq. no.	6	7	8	9	10	11
Intercept	0.7342	0.7443	0.5969	0.7985	0.7302	0.7834
E%	-3.885×10^{-3}			-4.647×10^{-3}		
[E]		-3.254×10^{-4}			-3.259×10^{-4}	
[TEA]			-3.934×10^{-4}			-3.912×10^{-4}
Fit Statistics						
n	32	41	22	55	44	24
AICc	5.09	-33.4	-4.70	68.3	-34.1	1.52
$\hat{\sigma}_s^2$	0.037	0.120	0.274	0.440	0.103	0.555
$\hat{\sigma}_\varepsilon^3$	0.115	0.106	0.076	0.242	0.110	0.074

[1]E% = Endophyte percentage; [E] = Ergovaline, ppb; [TEA] = Total Ergot Alkaloids, ppb; AICc = corrected Akaike's Information Criterion.
[2]Square root of the estimated variance associated with study
[3]Square root of the estimated variance associate with residual error.

Table 6-4. Correlation table for variables reported in studies

	Dur	Mon	RH	RH.sd	Max.T	Max.Tsd	Avg.T	Avg.Tsd	ADG	Forage	HI	THI	[TEA]	[E]	E%	AEL
Dur	1.00															
Mon	-0.22[a]	1.00														
RH	0.52	0.02	1.00													
RH.sd	-0.30[a]	-0.59[a]	-0.32[a]	1.00												
Max.T	0.62	-0.16[b]	0.83	0.88	1.00											
Max.Tsd	0.05	0.05	0.11	0.52[a]	0.59[a]	1.00										
Avg.T	-0.23[a]	0.25[a]	0.49[a]	0.88	0.70	-0.60[a]	1.00									
Avg.Tsd	0.08	-0.06	-0.21[b]	0.39[a]	0.43[a]	0.98	0.44[a]	1.00								
ADG	-0.14[b]	0.07	0.03	-0.44[a]	0.01	-0.16	0.26[a]	-0.11	1.00							
Forage	0.52[a]	-0.16	-0.10	0.72[a]	0.06	0.09	0.00	0.02	0.02	1.00						
HI	-0.39[a]	-0.30[a]	0.52	-0.88	0.72	-0.59[a]	1.00	-0.44[a]	0.26[a]	0.00	1.00					
THI	-0.41[a]	-0.27[a]	0.54	-0.87	0.73	-0.60[a]	1.00	-0.46[a]	0.27[a]	-0.01	1.00	1.00				
[TEA]	0.00	-0.02	-0.05	-0.20	-0.09	0.00	-0.14	0.01	-0.46[a]	1.00	0.20	0.19	1.00			
[E]	-0.07	0.01	0.12	-0.04	0.07	-0.12	0.07	-0.07	-0.57[a]	0.05	0.07	0.07	NA	1.00		
E%	0.16	0.13	0.22	NA	0.23	NA	0.21	NA	-0.32[a]	NA	0.23	0.23	NA	NA	1.00	
AEL	-0.20[a]	0.57	-0.17[b]	0.82	0.45[a]	-0.55[a]	0.54	-0.45[a]	0.14	0.48[a]	-0.47[a]	-0.45[a]	0.03	0.01	0.03	1.00

[1]Dur = Duration; Mon = Starting Month; RH = Relative Humidity; RH sd = standard deviation of the RH; Max.T = Average Maximum Temperature, °C; Max.T sd = standard deviation of Max.T; Avg.T = Average Temperature, °C; Avg.T sd = standard deviation of Avg.T; Forage = Forage Yield; HI = Heat Index; THI = Temperature-Humidity Index; [TEA] = Total Ergot Alkaloid Concentration, ppb; [E] = Ergovaline concentration, ppb; E% = Endophyte percentage; AEL = Average Infection Level over duration of study.

[a] $P < 0.05$

[b] $0.05 < P < 0.10$

Table 6-5. Parameter estimates for models of ADG using the complete-SEM dataset with thresholds derived from E% and [E] separately.

	Infection-Only Equations		Weather-Included Equations	
Item[1]	E%	[E]	E%	[E]
Eq. no.	15	16	17	18
Intercept	0.745	0.754	1.017	-0.172
E% Infected	-0.269		-0.226	
[E] Infected		-0.249		0.882
THI				0.016
AEL			-0.004	
[E] Infected x THI				-0.020
Fit Statistics				
n	32	41	32	41
AICc	-3.5	-39.4	2.56	-50.3
$\hat{\sigma}_s^2$	<0.001	0.098	<0.001	0.041
$\hat{\sigma}_\varepsilon^3$	0.117	0.114	0.104	0.078

[1]E% = Endophyte percentage; [E] = Ergovaline, ppb; [TEA] = Total Ergot Alkaloids, ppb; AICc = corrected Akaike's Information Criterion.
[2]Square root of the estimated variance associated with study
[3]Square root of the estimated variance associate with residual error

Figures

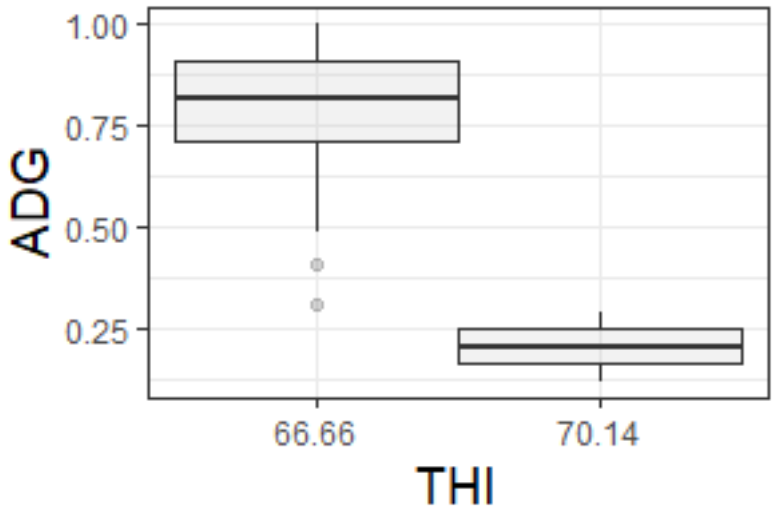

Figure 6-1. Boxplot of relationship between Temperature-Humidity Index (THI) and ADG in all Georgia data. Dataset was comprised of all studies done in Georgia within the dataset to look for a visual relationship between THI and ADG not seen in the models derived for ADG.

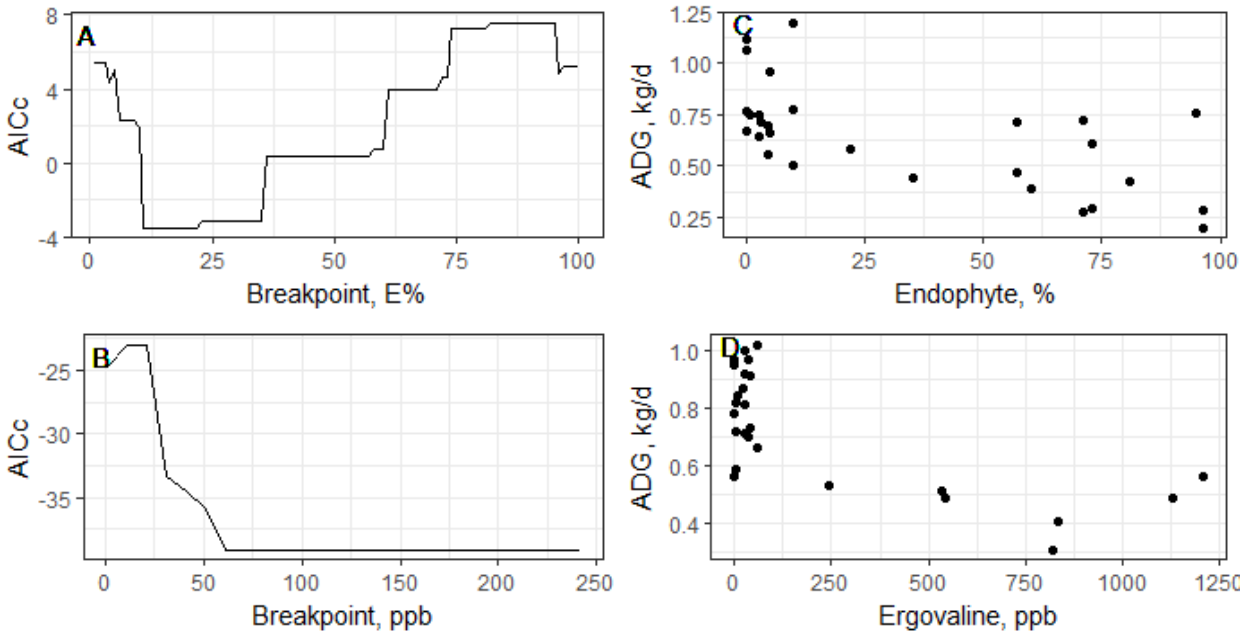

Figure 6-2. (A,B) Corrected Akaike's Information Criterion (AICc) values associated with pairwise regression models built at an infection threshold (IT). The optimal IT was selected as the first point at which the AICc value reached its minimum. (C,D) Plots of infection measurement method, illustrating the sparsity of intermediate level infection data.

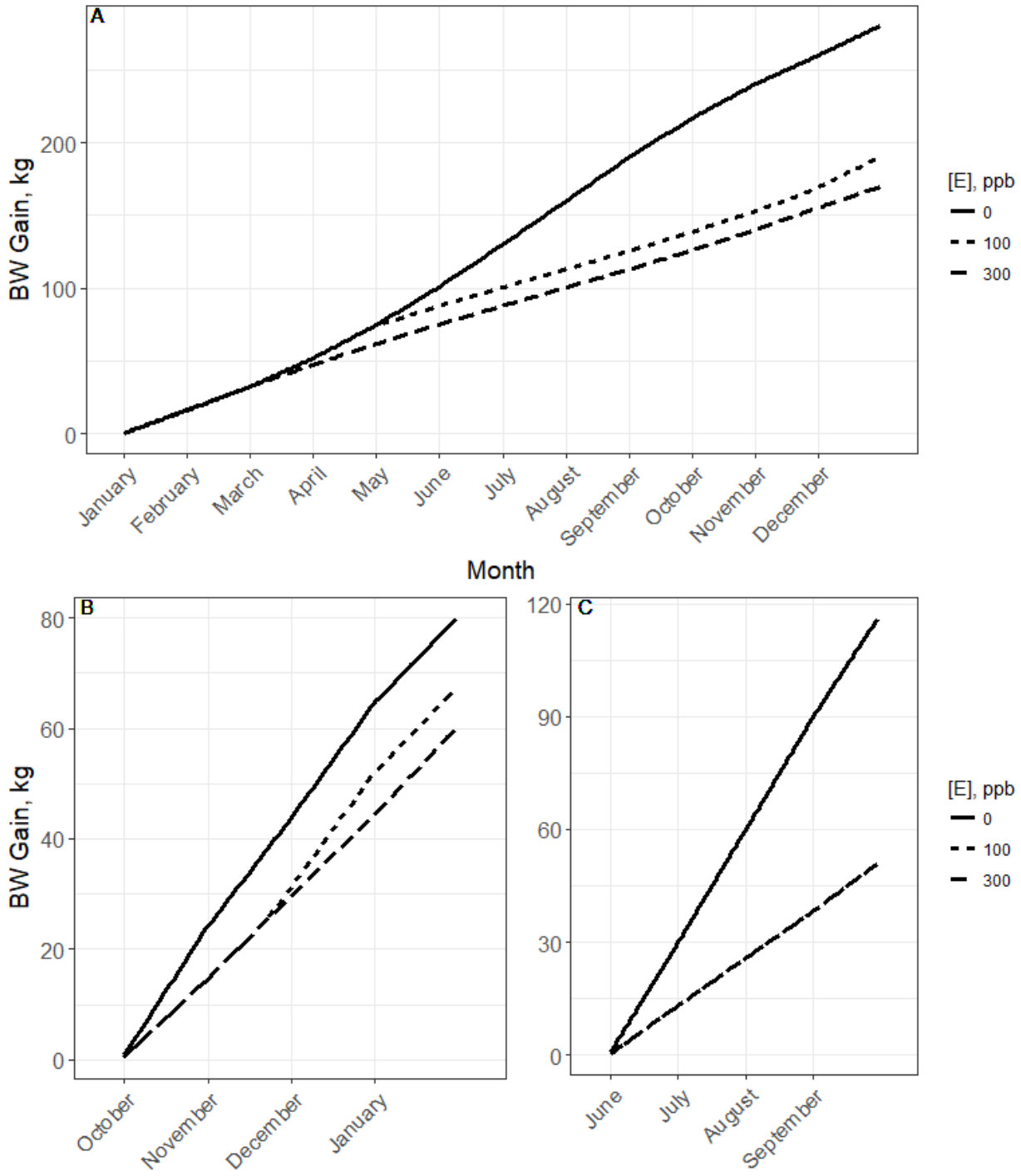

Figure 6-3. (A) Cumulative BW gain over 1 year by maximum infection level of a given pasture. Infection level was corrected using Ju et. al (2006) and temperature-humidity index (THI) data for each month was derived from the weather data collect for each study. ADG was calculated using Eq. 12 with inputs for THI and infection level based on 60 ppb ergovaline concentration ([E]) threshold. (B) Cumulative BW gain over a

theoretical fall growing season. (C) Cumulative BW gain over a theoretical spring growing season.

Chapter 7: Practical challenges and potential approaches to predicting low-incidence diseases on-farm using individual cow data: a clinical mastitis example

**To be submitted to Journal of Animal Science

Abstract

Mastitis is an extremely costly disease in the US dairy industry. Despite this, clinical mastitis (CM) can be challenging to model because of its low prevalence, affecting less than 1% of cows per day. This low prevalence means that a model can nearly always be correct at predicting the negative instance (no clinical infection). But the sparsity of clinical mastitis cases makes training, evaluating, and applying CM prediction models a challenge. The objective of this study was to build models for predicting CM incidence using time-series sensor data and choose models that maximize net return based on a cost matrix. This work used data from two university dairy farms, the University of Florida and Virginia Polytechnic Institute and State University to gather representative CM incidence data including 110,156 milkings and 333 total CM cases. Variables used in the models were milk yield, protein, lactose, fat, activity, conductivity, DIM (7 bins), and lactation number (1st, 2nd, or 3+). Models that predicted either likelihoods of Gram-positive (GP) or Gram-negative (GN) CM on each day were derived using extreme gradient

boosting with weighting favoring true positive cases, logistic responses, and log-loss errors. Model accuracies were determined using data randomly held-out from the training set on each run. All variables considered were in terms of change (slope) over previous days including the day of visual mastitis. The GN and GP models were each run 100 times on random 75% samples of the data and the GN models had a median sensitivity (Se) of 52.6% and specificity (Sp) of 99.8%, while the GP models had a median Se of 37.5% and Sp of 99.9% when tested on the held-out data. The International Standard ISO/FDIS 20966 describes a minimum Se of 80% with Sp greater than 99%; however, comparison of results in the present work suggests that CM models might benefit from greater importance placed on Sp. Results also highlight the importance of positive predictive value along with Sp and Se. The calculated partial net return of our GN and GP models were -$0.15 and -$0.10 per cow per lactation, respectively; while ISO-standard models with Se of 80% and Sp of 99% would return -$1.32 per cow per lactation. Models chosen that minimized the cost to the farmer were in stark contrast to models that met ISO guidelines, showing the asymmetry in targets between Sp and Se when the incidence rate of the disease is low.

Introduction

Mastitis is considered one of the costliest diseases facing dairy producers. Costs associated with mastitis are estimated to be $2 billion in the US annually. Despite the high cost, the prevalence of mastitis is between 25 to 41 cases per 100 cows per lactation, equivalent to roughly one clinical mastitis (CM) case per 1,460 to 890 cow days (0.07% to 0.1%) (Pol and Ruegg, 2007; USDA-APHIS-VS-CEAH-NAHMS, 2014). This low prevalence poses a challenge to modelers attempting to predict CM. A predictive model that is either unable to identify a majority of true positive cases or true negative cases would not be useful. Historically, the "usefulness" or appropriateness of models or technologies designed to predict mastitis have been evaluated based on two outcomes, Sensitivity (Se) and Specificity (Sp) (Sargeant et al., 2001; Pyörälä, 2003; Koskinen et al., 2009; Ganda et al., 2016; Khatun et al., 2017). Specificity refers to the true negative rate, that is, the proportion of negative cases correctly identified (true negatives) of the total negative cases in the dataset. Sensitivity is the true positive rate or proportion of positive cases correctly identified from the total positive cases in the dataset. Although Se and Sp are commonly used to evaluate mastitis prediction approaches, the Se and Sp measurements on low-prevalence datasets will tend to skew interpretation of the models. For example, a predictive model with Se = 80% and Sp = 80% on a dataset with 50%

CM cases compared to a dataset with only 5% CM cases would yield the confusion matrices in Table 7-1.

When the prevalence of CM is 50%, there are 4 true positive predictions of CM for every incorrect signal (false positive). When the prevalence is decreased to only 5%, the result is 4.75 false positive cases for every true positive identification. In terms of positive predictive value (PPV), or likelihood of a positive prediction truly being a positive result, the 10-fold decline in CM prevalence results in a 19-fold change in PPV. This is an unbalanced change in PPV and highlights the challenge of using only Se and Sp to evaluate models describing low prevalence datasets. The International Standard ISO/FDIS 20966 (Automatic milking installations-requirements and testing) of the International Standard Organization (ISO) includes an annex describing a minimum Se of 80% with a Sp greater than 99%. These two values are commonly reported in the mastitis prediction literature as a gold standard for minimum model performance (Hogeveen et al., 2010). Because of CM's low prevalence in the data, a model with ISO standard performance in terms of Sp and Se will produce more false alarms compared to a model predicting a greater prevalence disease. The incidence of false alarm rate is a critical consideration because both false positive and false negative detection can be an economic and practical burden. For example, one study found that an automated milking system reported

11,156 alerts for CM, of which 159 were true cases (Steeneveld et al., 2010). In this study, farmers used their judgment and non-milking system data to help filter these alerts in an attempt to improve efficiency (Steeneveld et al., 2010). In order for a sensor-based mastitis detection system to be successful in the industry, the benefit of using the tool must outweigh the costs. As such, evaluating the net return of model predictions may be a more logical way of evaluating model success.

One strategy to account for the net return of a model is to weight model outcomes proportional to the cost of each outcome. Without a weighting scheme, a model's error structure will be unbiased between false positives and false negatives. This may appear advantageous, if the model-builder has no preconceived notions about false positives or false negatives, but from an applicability standpoint this does not reflect the costs the producer is facing. It is much more costly for the producer to incorrectly treat an animal than to wait an additional day to correctly identify CM in that animal. This asymmetry is accentuated in a disease with a low prevalence because CM is less likely to randomly occur *a priori* *(Dominiak and Kristensen, 2017)*. Weighting schemes can mean setting limits for Se or Sp before running models or actually changing the error function to penalize false negatives or false positives to be weighted greater. Previously unweighted mastitis detection algorithms, meaning no additional penalty for

false positives or false negatives, have been successfully published using ranging datasets and methods (de Mol et al., 1997; Maatje et al., 1997; Cavero et al., 2006; Kamphuis et al., 2008). Although models have been built using the principles of weighting schemes and imbalanced Se and Sp (Kamphuis et al., 2010; Miekley et al., 2012; Huybrechts et al., 2014), the idea of a cost matrix to assess applicability is novel to our knowledge and the approach can be widely applied to models in this and other similar fields.

As one review reports, "no sensor-based detection model has fulfilled the performance demands needed to generate a satisfyingly low level of false positive alarms." (Dominiak and Kristensen, 2017). In most in-line milk monitoring systems, data such as conductivity, somatic cell count, and milk components are logged and recorded in data management software almost instantaneously. One major challenge preventing detection of mastitis is the derivation of precise and accurate algorithms to convert this quantity of raw, individual animal, time-series data into management insights. Techniques such as neural networks (Nielen et al., 1995; Cavero et al., 2008; Sun et al., 2009; Samarasinghe et al., 2017), fuzzy logic (Cavero et al., 2006; Kramer et al., 2009), moving averages (Maatje, 1992), and others appear in the research as means to estimate likelihood of CM given sensor and non-sensor data being collected on farm. Irrespective of the approach, this real-time sensor data is prone to two

challenges : 1) missing data (missed measurements or poor farmer record keeping) and 2) outlier data (sensor or recording errors). In the four unweighted model papers referenced above, none reference any data cleaning technique relating to missing data (de Mol et al., 1997; Maatje et al., 1997; Cavero et al., 2006; Kamphuis et al., 2008). In Kamphuis et al. (2010) and Miekley et al. (2012) removal of data outside of certain preset bounds was used to determine outlier points and data missing important variables was removed prior to analysis. In Huybrechts et al. (2014), maximum likelihood estimation was used to estimate values for milk yield if a value was missed to allow their model to work as intended. Addressing these data challenges with various data cleaning techniques is essential to deriving industry-relevant models.

Another shortfall of previously published mastitis detection algorithms, as they pertain to on-farm application, is the lack of pathogen specificity (Dominiak and Kristensen, 2017). Mastitis infections are predominantly caused by two major categories of pathogen: Gram-positive (GP) or Gram-negative (GN). Because the protocol for treating mastitis infections caused by these two pathogen types differs, an ideal mastitis detection algorithm would report the probable pathogen. In general, GP cases are treated with antibiotics which require milk be discarded until the antibiotic clears the cow's system; GN cases are not treated with

antibiotics. Recent research has shown differences in sensor variables leading up to CM detection in GP and GN cases, suggesting it might be possible to expand the use of sensor-based mastitic detection algorithms to also suggest which pathogen is implicated in the infection (Vasquez et al., 2018). Future efforts to advance the sensor-based prediction of CM should be complementary with efforts to evaluate models on the basis of net returns, and efforts to reduce challenges associated with missing and outlier data.

The objective of this study was to build separate GN and GP models for predicting CM incidence using time-series sensor data and compare models that maximize net return based on a partial costs matrix to those created to maximize Sp and Se. It was hypothesized that creating models using a partial costs matrix as the accuracy measure would produce models that would produce greater net returns on farm than maximizing model Sp and Se.

Materials and Methods

Data

Data were collected from lactating cows on the Virginia Tech (VT) and University of Florida (UF) dairy farms between August 2015 and April 2017. Milk yield and conductivity were measured via milk meter (AfiMilk MPC) and composition data (protein, fat, lactose, and somatic cell count) were collected using an in-line milk analyzer (AfiLab©, Afimilk Ltd., Kibbutz Afikim, Israel).

Daily activity measures, daily rest bout, rest duration, and total activity, were collected using Afi PedometerPlus© at VT and Afi Pedometer© at UF. At both farms, any variables measured twice daily were reported as the average of the two values in model building. Cows were milked twice daily on both farms and milk yield was summed rather than averaged when used in the models. Cows were labeled as clinically mastitic on the day of detection and the previous 2 days or 5 days for GN or GP, respectively. This practice of labelling mastitis differently based on Gram-type is uncommon with respect to similar research on modeling CM (Steeneveld et al., 2009; Kamphuis et al., 2011). Actual cases of CM were identified by farm staff and recorded in the Afifarm Software. The use of farm staff to determine CM cases is consistent with the literature (de Mol et al., 1997; Kamphuis et al., 2010; Miekley et al., 2012). Another common way to label CM cases is to set variable thresholds, like conductivity or SCC, and all cows that exceed these thresholds are considered mastitic (Maatje, 1992; Kamphuis et al., 2008). This dataset provided a representative proportion of CM cases to healthy cows (milkings = 136,127, CM% = 0.47%). Table 7-2 shows the raw data's summary statistics prior to any data cleansing.

Data Preparation

On both farms, daily milk yield and composition were comprised of morning (AM) and evening (PM) measurements, reflecting twice-daily milkings. Because the parlor system used to collect milk data is automated, there are occasional failures in the system that lead to missing data. Such failures include: milking events where the in-parlor radio-frequency identification tag reader mis-reads the animal's tag; milk composition or yield readings linked to an animal are not reported, often because they are outside the feasible range; and events that require manual recording are missed or forgotten. Computer error and truly missed-milkings can be hard to differentiate simply by looking back at the data. These missing data create a computational challenge during analysis because the maximum number of days used in the slope calculation was seven, meaning each missing day had high potential to impact the resulting slope. Approximately 0.78% of milking yield and composition records were missing either the AM or PM measurements. To handle these missing data, days with single data points were repeated in the empty records. We used this method because cows were milked every 12 hours, lending to similar AM and PM measurements. After adjusting for these missing values, daily average milk yield and composition were calculated by averaging over paired samples each day, by cow. The use of this data-filling technique helped with calculating the change in each variable over time.

Data were then standardized to mean = 0 and standard deviation = 1 and values greater than 5 standard deviations from the mean were considered outliers and removed. This approach was chosen because it eliminated approximately 5% of values in each variable measured, which is consistent with other research (Kamphuis et al., 2010). Conductivity, daily rest bout, rest duration, and activity measurements had a higher proportion of outlier data than milk yield and composition variables. Outliers were also identified based on the average of previous values to deal with variables with significant variance. In cases like milk yield, an outlier at one point in lactation may be normal later in lactation. In these cases, the previous 4-day average was used as a comparator and the absolute difference between the target value and the 4-day average was calculated. By curtailing the top 5% of these values to the average of the previous 4 days, we attempted to dampen noise in the sensed variables and better characterize the changes in variables over time, as illustrated in Figure 7-1.

When cleaning and standardizing the data, data from each farm was standardized separately. When combined standardization was used, data were found to be too dissimilar between farms and made slopes less indicative of changes. Standardization allows for comparison across variables in terms of change, is common in multivariate model-building, and has been used to standardize conductivity scores in past CM sensor models (de Mol and Woldt,

2001). A summary of the combined dataset with standardization and outlier removed can be found in Table 7-3 (n = 110,156, CM% = 0.30%). Although the CM incidence in this dataset was greater than the estimated 0.1% of the average U.S. farm (Pol and Ruegg, 2007; USDA–APHIS–VS–CEAH–NAHMS, 2014), it is important to consider that CM incidence is not distributed normally about the mean. Farms with greater CM incidence will skew the distribution to the right. This being the case, the total CM incidence in the dataset is similar to other research in the field of sensor-based detection modeling, with de Mol et al. (1997) reporting a CM incidence of 0.14% per cow day in 75,000 milkings or Miekley et al. (2012) reporting 0.5% CM incidence per day in 46,000 cow days. These data cleaning techniques are useful not only in the preparation of training data sets, but also in actual farm monitoring systems, because missing data, outliers, and skewed values are commonplace.

Boosting

Gradient boosting trees algorithms (sometimes referred to as just "boosting") have become a popular machine learning technique recently because of their accuracy and speed. Boosting is a method of learning a dataset by deriving iteratively more accurate "weak learners". Weak learners are decision trees that are simple (few nodes) and have predictable biases that can be corrected with

additional decision trees. By iteratively correcting these trees by adding another tree, improved accuracy can be attained. This method is useful for learning the importance of features of the dataset as well, because of the ability to analyze the Gini indices of the trees. Gini indices measure the ability of a tree to produce classifications proportional to the labels true proportion in the data and are commonly used to identify feature importance of decision tree-based algorithms (Ushikubo et al., 2017).

Extreme Gradient Boosting (XGBoost) is currently one of the most popular gradient boosting tree algorithms (Chen and Guestrin, 2016). Chen and Guestrin (2016) cite problems like store sales prediction; high energy physics event classification; and massive online course dropout rate prediction as just some of the problems in which XGBoost has won prediction competitions. The XGBoost algorithm was also used to discover the Higgs Boson at the Large Hadron Collider because of its applicability to complex datasets (Chen and He, 2014). Although these examples may seem far from applicable to CM detection, the underlying data all share similar characteristics. Because of its previous success in other sparse data problems, it is logical to test XGBoost as a means of CM prediction. Already, techniques advancing beyond linear regression, like the aforementioned boosting, bagging, or markov chains, have been used to tackle the issue of predicting disease from sensor data (Kamphuis et al., 2010; Ostersen et al., 2010).

An added advantage of the XGBoost algorithm is that it uses efficient computing techniques such as parallelization and cache optimization, to make classifications on sparse datasets more accurate than traditional tree algorithms. Parallelization allows models to be built and run on all cores simultaneously, increasing speed and improving search space. Cache optimization reduces the computational investment in accessing information stored in a computer's memory.

Model Building

XGBoost version 0.82.1 (Chen and He, 2014) was used in R (R Team, 2018) for all model building. Each iteration of each model was trained on 75% of the dataset, with 25% being held out randomly for testing. Each model was iterated 100 times with new, random samplings for the train-test data split. All models were run using the xgboost() command in R, with a scaled learning rate of 0.5, maximum tree depth of 100, 4 parallel threads, 40 rounds of boosting, model objective set to "binary:logistic", and binary classification error.

The response variable used in all algorithms was the presence of CM, either GN CM or GP CM, depending on the model being built. To better account for prevalence, GN cases were predicted using all data, using healthy cows and GP cases as negative examples. Conversely, GP cases were predicted including healthy cows and GN

cases as negative examples. Using cases with the opposite pathogen type as a negative example in the training data allowed for a larger dataset and should theoretically improve the algorithm's ability to differentiate between CM types. Potential explanatory variables offered into the algorithms included slopes for daily rest bout, rest duration, activity, conductivity, fat, lactose, protein, and milk yield. Additional variables were included for DIM, binned into seven discrete groups, and lactation number, either first, second, or third or greater. Seven bins were chosen to describe the lactation curve because this best balanced the need to separate differently sloping areas of the curve, while maintaining great enough sample sizes within each bin.

The model itself was a binary logistic model and predictions of CM (0 or 1) were determined by using the model outputs (probability between 0.0 and 1.0). A threshold for positive predictions was set at 0.5 and a log loss function was utilized to propagate error in the model. Because of the significant imbalance in the costs associated with misprediction in a CM positive case compared to a CM negative case, the error associated with positive predictions was scaled 1,000,000 times that of a negative in the XGBoost framework using the "scale_pos_weight = 1000000" command. This strategy was implemented to mimic the cost matrix outlined in Table 7-4, where the importance of positive predictions are infinitely greater than that of negative predictions. The value of

1,000,000 was chosen because accuracy was found to not increase any more if the value was any greater.

Model Evaluation

Evaluation of models was completed using held-out data, with 75% of the data used to train models and 25% to test. Random splits of this proportion were repeated 100 times and Se and Sp were recorded. Change in potential net return (Δ_{cost}) based on model decision was used as a metric for cost associated with each model. Change in potential net return was calculated as the return from a healthy cow minus the cost associated with milk loss, minus the cost associated with increased mortality, minus the cost associated with treatment. Using the values from Bar et al. (2008), we considered a healthy cow to return $426, and considered losses as $115 due to milk loss, $14 due to increased mortality, and $50 due to treatment-associated costs. Using these values, the return of a cow treated that was not actually infected would be $426 - $115 - $50 = $261 and the Δ_{cost} = -$165. The return of a cow left untreated for CM would be the milk lost over that period, $115. Therefore, the return for treating a cow would be $426 - $115 - $14 - $50 = $247 and the Δ_{cost} = +$132. In both the true and false negative cases, the Δ_{cost} = $0 because these cases reflect the action given that no models were used at all. Table 7-4 illustrates the costs associated with each outcome of the confusion matrix.

The threshold for prediction in the training data was always 0.5 or 50%. Using receiver operator characteristic curves and the cost matrix of missed CM cases and treated CM cases, we determined the optimal threshold for each resulting model. The optimal threshold minimized the value of [TP x 132] - [FP x 165] at a given Sp and Se. To prevent the threshold from being near the extremes, that is 0% or 100%, thresholds were chosen as that which maximized the net return and was between 10% and 90%. Excluding thresholds near the extremes prevented models that never predicted CM cases, but still produced good net returns per prediction. The resulting optimal thresholds for all 100 model iterations are shown in Figure 7-2. The lack of a consistent threshold that improves net returns, combined with the scale of improvement in returns, lead us to choose 0.5 as the threshold for evaluating all proceeding models.

Results and Discussion

The GN models had a median Se of 52.6% and Sp of 99.8% when tested on held-out data. The GP models had a median Se of 37.5% and Sp of 99.9% when tested on the test data. Table 7-5 shows prediction results for all GN and GP models. These values represent improvement in mean Sp and Se compared to the values derived on the same data in Steele et al., (2019, *in press*). Steele et al. (2019, *in press*) describes 3-day linear models which utilized the same data. These models generated greater Se but lesser Sp; GN

models resulted in Se of 73.3% and Sp of 74.1%, while for GP models, Se = 47.4% and Sp = 89.4%. These differences are, in part, because of the different algorithms used and because of the approach for selection of prediction thresholds. When considering the practical implications of greater Sp, the lessening of false positives makes our models more appropriate for on-farm use.

Model performance, in terms of Se and Sp, was optimized for net cost of each prediction. Assuming a cost of $165 for each FP and a net gain of $132 for each TP case, the GN prediction model with Se of 52.63% and Sp of 99.78% yielded a net return of -$0.15 per cow per lactation, assuming 0.30% CM incidence rate, as in the full dataset. Using the unweighted linear model under the same cost matrix produced a net return per cow per lactation of -$42.32 (Steele et al., 2019). The optimization parameters of the model being built clearly have importance in the CM prediction problem. For the GP CM prediction model, Se = 37.5% and Sp = 99.9% also resulted in net returns per prediction, yielding an estimate of -$0.10 per cow per lactation. Although the net returns per prediction are still negative, they are close to zero in comparison to the previous unweighted linear model, which produced a net loss per cow per lactation of $17.25 (Steele et al. 2019; *in press*). Both these comparisons assume no producer input into the decisions, while, in reality, producers provide intuition and insight into treatment decisions (Steeneveld et al., 2010).

Plots of all variables and the distribution of their respective gains across models are shown in Figure 7-3 for GN and Figure 7-4 for GP. Milk yield slope over the previous 3 days was consistently the most important variable in the GN models, while lactose and protein slopes had the greatest median gains in accuracy in the GP models. Lactation number was consistently the least-influential variable in the models. In the GN models, the median gain in accuracy attributed to milk yield was more than twice that of the next most important variable, activity, measured as steps per day. Gram-negative CM is associated with greater losses in milk yield during first-case occurrences when compared to GP cases (Gröhn et al., 2004). Gram-negative CM is also associated with a faster onset of infection and subsequently faster decline in milk yield compared to GP CM (Smith et al., 1985; Pyörälä et al., 1994; Bannerman et al., 2004); these easily measurable variations attributable to GN CM suggest that classification methods are viable. The model results show differing variables as most important in the days leading up to CM incidence, which supports the supposition that GP and GN cases present in different manners. Pathogen-specific treatment of CM has been shown to be cost effective in treating GN CM and helps reduce the use of antibiotics (Schukken et al., 2011; Fuenzalida and Ruegg, 2019). The variation in CM presentation coupled with the fact that targeted treatment has shown promise suggests that

retroactive classification models built on a cost matrix framework could produce tools for recommending CM treatment protocol using on-farm data, but deriving such retroactive models was outside the scope of this study.

The cost matrix can be used in the evaluation of existing models and as a loss function for training models. Using a loss function that values true positives and penalizes false negatives leads to models with greater Sp than models with no weighted loss function. For example, a model with 95% Se and 95% Sp implemented on a herd with 1% incidence of CM/cow-day would have a net cost of $6.91 per cow per lactation. Although this model appears to be exceptional based on Se and Sp, its implementation would cost a farm with 250 lactating cows over $2,067 per year. This assessment of cost is not perfect because farmers are likely to integrate their own judgment into any treatment decision (Steeneveld et al., 2010), meaning any model would most likely be employed in addition to human observation, which would likely reduce costs. However, it allows for a more realistic and objective comparison among models and highlights major challenges with the existing ISO standards for precision dairy technologies (International Standard Organization, 2007).

In agriculture, as in other sectors, profitability a primary importance to producers. By analyzing the results of the model's predictions and using appropriate costs, the savings from a

correctly identified case of mastitis is approximately equal to the loss of treating an uninfected cow (Bar et al., 2008). With this cost assumption, the hypothetical model trained on 5% positive cases would have a negative net return per cow treated. In order to reach net positive returns on predictions, we needed to reach a ratio of:

$$\frac{\text{Benefit of TP} \times \text{TP}}{\text{Cost of FP} \times \text{FP}} > 1.0$$

Because the costs of FP and benefits of TP are similar, this suggests that measuring the number of true positives for each predicted positive instance, called positive predictive value (PPV), adds insights that Se and Sp lack on their own. The formula for PPV is:

$$PPV = \frac{TP}{TP + FP}$$

The implications of models with high Sp and Se but low PPV would be high false alarm rates or high unnecessary treatment. In the context of mastitis prediction, farmers would be checking a majority of cows and not finding clinical signs. In this scenario, farmers may become insensitive to the alerts of a system like this. "Alarm fatigue" is a well-documented problem in critical care medicine (Graham and Cvach, 2010; Borowski et al., 2011). Results of high rates of false positive alerts lead to increased workload, decreased Se, and increased missed critical events (Graham and

Cvach, 2010). In 2014, alarms were the Emergency Care Research Institute's #1 hazard to care (Emergency Care Research Institute, 2013). In order to prevent cases similar to this in mastitis prevention, an appropriate threshold PPV value (or similar statistic) should be established.

By using a cost matrix to evaluate the models produced in this work, predicted net returns per cow per lactation were -$0.15 for GN and -$0.10 for GP models. Although these net returns are still negative, consider the value of using a cost matrix-based evaluation to choose models over maximizing Sp and Se simultaneously. Consider that a dataset with 0.3% CM incidence, models with Se/Sp = 80/95; 95/98; 99/99 would have net returns per lactation of -$7.91, -$2.91, -$1.25, respectively, under our partial costs matrix. The best of these models, Se/Sp = 99/99, would still be more than 8-times more costly per lactation than the GN model described above (-$1.25 versus -$0.15 per cow per lactation), producing 40% more false-positives per true-positive (0.3 versus 0.72 PPV).

Conclusion

Building models that predict CM using sensor-based detection data requires thoughtful collection, analysis, and implementation. When collecting data for predicting CM, where incidence is very low, consider that building a model on this data will inherently

lead to severely limiting results, such as higher false alarm rates. When analyzing the success of the models built, mastitis prediction guidelines specify Sp and Se thresholds, but other measures like PPV, can be useful in building a practical model. Consider that, on-farm, the cost comes in falsely predicting CM. Building models that weight all contingencies equally does not properly assess the situation the producer faces.

Acknowledgements

We wish to acknowledge Afimilk Ltd. for assistance with data extraction and the farm staff at the Virginia Tech Dairy Center and the University of Florida Dairy for collecting milk samples.

References

Bannerman, D.D., M.J. Paape, J.-W. Lee, X. Zhao, J.C. Hope, and P. Rainard. 2004. Escherichia coli and Staphylococcus aureus elicit differential innate immune responses following intramammary infection. Clin. Diagn. Lab. Immunol. 11:463–472. doi:10.1128/CDLI.11.3.463-472.2004.

Bar, D., L.W. Tauer, G. Bennett, R.N. González, J.A. Hertl, Y.H. Schukken, H.F. Schulte, F.L. Welcome, and Y.T. Gröhn. 2008. The cost of generic clinical mastitis in dairy cows as estimated by using dynamic programming. J. Dairy Sci. 91:2205–2214. doi:10.3168/jds.2007-0573.

Borowski, M., M. Görges, R. Fried, O. Such, C. Wrede, and M. Imhoff. 2011. Medical device alarms. Biomed. Tech. 56:73–83. doi:10.1515/BMT.2011.005.

Cavero, D., K.-H. Tölle, C. Buxadé, and J. Krieter. 2006. Mastitis detection in dairy cows by application of fuzzy logic. Livest. Sci. 105:207–213. doi:10.1016/j.livsci.2006.06.006.

Cavero, D., K.-H. Tölle, C. Henze, C. Buxadé, and J. Krieter. 2008. Mastitis detection in dairy cows by application of neural networks. Livest. Sci. 114:280–286. doi:10.1016/j.livsci.2007.05.012.

Chen, T., and C. Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. CoRR abs/1603.02754. doi:10.1145/2939672.2939785.

Chen, T., and T. He. 2014. Higgs Boson Discovery with Boosted Trees. Pages 69–80 in Proceedings of the 2014 International Conference on High-Energy Physics and Machine Learning - Volume 42. JMLR.org.

Dominiak, K.N., and A.R. Kristensen. 2017. Prioritizing alarms from sensor-based detection models in livestock production

- A review on model performance and alarm reducing methods. Comput. Electron. Agric. 133:46–67. doi:10.1016/j.compag.2016.12.008.

Emergency Care Research Institute. 2013. Top 10 Health Technology Hazards Report For 2014. Accessed May 21, 2018. https://www.ecri.org/components/HDJournal/Pages/hd201310guid_Top10.aspx.

Fuenzalida, M.J., and P.L. Ruegg. 2019. Negatively controlled, randomized clinical trial to evaluate intramammary treatment of nonsevere, gram-negative clinical mastitis. Journal of Dairy Science 102:5438–5457. doi:10.3168/jds.2018-16156.

Ganda, E.K., R.S. Bisinotto, D.H. Decter, and R.C. Bicalho. 2016. Evaluation of an On-Farm Culture System (Accumast) for Fast Identification of Milk Pathogens Associated with Clinical Mastitis in Dairy Cows. PLoS One 11:e0155314. doi:10.1371/journal.pone.0155314.

Graham, K.C., and M. Cvach. 2010. Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms. Am. J. Crit. Care 19:28–34; quiz 35. doi:10.4037/ajcc2010651.

Gröhn, Y.T., D.J. Wilson, R.N. González, J.A. Hertl, H. Schulte, G. Bennett, and Y.H. Schukken. 2004. Effect of pathogen-specific clinical mastitis on milk yield in dairy cows. J. Dairy Sci. 87:3358–3374. doi:10.3168/jds.S0022-0302(04)73472-4.

Hogeveen, H., C. Kamphuis, W. Steeneveld, and H. Mollenhorst. 2010. Sensors and Clinical Mastitis—The Quest for the Perfect Alert. Sensors 10:7991–8009. doi:10.3390/s100907991.

Huybrechts, T., K. Mertens, J. De Baerdemaeker, B. De Ketelaere, and W. Saeys. 2014. Early warnings from automatic milk

yield monitoring with online synergistic control. J. Dairy Sci. 97:3371–3381. doi:10.3168/jds.2013-6913.

International Standard Organization. 2007. Automatic milking installations—requirements and testing. International Standard Organization, Geneva, Switzerland.

Kamphuis, C., H. Mollenhorst, A. Feelders, D. Pietersma, and H. Hogeveen. 2010. Decision-tree induction to detect clinical mastitis with automatic milking. Comput. Electron. Agric. 70:60–68. doi:10.1016/j.compag.2009.08.012.

Kamphuis, C., H. Mollenhorst, and H. Hogeveen. 2011. Sensor measurements revealed: Predicting the Gram-status of clinical mastitis causal pathogens. Comput. Electron. Agric. 77:86–94. doi:10.1016/j.compag.2011.03.012.

Kamphuis, C., R. Sherlock, J. Jago, G. Mein, and H. Hogeveen. 2008. Automatic detection of clinical mastitis is improved by in-line monitoring of somatic cell count. J. Dairy Sci. 91:4560–4570. doi:10.3168/jds.2008-1160.

Khatun, M., C.E.F. Clark, N.A. Lyons, P.C. Thomson, K.L. Kerrisk, and S.C. García. 2017. Early detection of clinical mastitis from electrical conductivity data in an automatic milking system. Anim. Produc. Sci. 57:1226–1232. doi:10.1071/AN16707.

Koskinen, M.T., J. Holopainen, S. Pyörälä, P. Bredbacka, A. Pitkälä, H.W. Barkema, R. Bexiga, J. Roberson, L. Sølverød, R. Piccinini, D. Kelton, H. Lehmusto, S. Niskala, and L. Salmikivi. 2009. Analytical specificity and sensitivity of a real-time polymerase chain reaction assay for identification of bovine mastitis pathogens. J. Dairy Sci. 92:952–959. doi:10.3168/jds.2008-1549.

Kramer, E., D. Cavero, E. Stamer, and J. Krieter. 2009. Mastitis and lameness detection in dairy cows by application of

fuzzy logic. Livest. Sci. 125:92–96. doi:10.1016/j.livsci.2009.02.020.

Maatje, K. 1992. The efficacy of in-line measurement of quarter milk electrical conductivity, milk yield and milk temperature for the detection of clinical and subclinical mastitis. Livestock Production Science 30:239–249. doi:10.1016/0301-6226(92)90072-c.

Maatje, K., R.M. de Mol, and W. Rossing. 1997. Cow status monitoring (health and oestrus) using detection sensors. Comput. Electron. Agric. 16:245–254. doi:10.1016/S0168-1699(96)00052-X.

Miekley, B., I. Traulsen, and J. Krieter. 2012. Detection of mastitis and lameness in dairy cows using wavelet analysis. Livest. Sci. 148:227–236. doi:10.1016/j.livsci.2012.06.010.

de Mol, R.M. de, and W.E. Woldt. 2001. Application of Fuzzy Logic in Automated Cow Status Monitoring. Journal of Dairy Science 84:400–410. doi:10.3168/jds.s0022-0302(01)74490-6.

de Mol, R.M., G.H. Kroeze, J.M.F.H. Achten, K. Maatje, and W. Rossing. 1997. Results of a multivariate approach to automated oestrus and mastitis detection. Livestock Production Science 48:219–227. doi:10.1016/S0301-6226(97)00028-6.

Nielen, M., M.H. Spigt, Y.H. Schukken, H.A. Deluyker, K. Maatje, and A. Brand. 1995. Application of a neural network to analyse on-line milking parlour data for the detection of clinical mastitis in dairy cows. Prev. Vet. Med. 22:15–28. doi:10.1016/0167-5877(94)00405-8.

Ostersen, T., C. Cornou, and A.R. Kristensen. 2010. Detecting oestrus by monitoring sows’ visits to a boar. Computers and Electronics in Agriculture 74:51–58. doi:10.1016/j.compag.2010.06.003.

Pol, M., and P.L. Ruegg. 2007. Treatment practices and quantification of antimicrobial drug usage in conventional and organic dairy farms in Wisconsin. J. Dairy Sci. 90:249–261. doi:10.3168/jds.S0022-0302(07)72626-7.

Pyörälä, S. 2003. Indicators of inflammation in the diagnosis of mastitis. Vet. Res. 34:565–578. doi:10.1051/vetres:2003026.

Pyörälä, S., L. Kaartinen, H. Käck, and V. Rainio. 1994. Efficacy of Two Therapy Regimens for Treatment of Experimentally Induced Escherichia coli Mastitis in Cows. Journal of Dairy Science 77:453–461. doi:10.3168/jds.s0022-0302(94)76973-3.

R Team. 2018. R Foundation for Statistical Computing.

Samarasinghe, S., M. Kohli, and D. Kulasiri. 2017. Neural Networks for Robotic Detection of Mastitis in Dairy Cows: Netherlands and New Zealand Perspectives.

Sargeant, J.M., K.E. Leslie, J.E. Shirley, B.J. Pulkrabek, and G.H. Lim. 2001. Sensitivity and specificity of somatic cell count and California Mastitis Test for identifying intramammary infection in early lactation. J. Dairy Sci. 84:2018–2024. doi:10.3168/jds.S0022-0302(01)74645-0.

Schukken, Y.H., G.J. Bennett, M.J. Zurakowski, H.L. Sharkey, B.J. Rauch, M.J. Thomas, B. Ceglowski, R.L. Saltman, N. Belomestnykh, and R.N. Zadoks. 2011. Randomized clinical trial to evaluate the efficacy of a 5-day ceftiofur hydrochloride intramammary treatment on nonsevere gram-negative clinical mastitis. J. Dairy Sci. 94:6203–6215. doi:10.3168/jds.2011-4290.

Smith, K.L., D.A. Todhunter, and P.S. Schoenberger. 1985. Environmental pathogens and intramammary infection during the dry period. J. Dairy Sci. 68:402–417.

Steele, N.M., A. Dicke, A. De Vries, S.J. Lacy-Hulbert, D.M. Liebe, R.R. White, and C.S. Petersson-Wolfe. 2019.

Identifying Gram-negative and Gram-positive clinical mastitis using daily milk component and behavioral sensor data.

Steeneveld, W., L.C. van der Gaag, H.W. Barkema, and H. Hogeveen. 2009. Providing probability distributions for the causal pathogen of clinical mastitis using naive Bayesian networks. J. Dairy Sci. 92:2598–2609. doi:10.3168/jds.2008-1694.

Steeneveld, W., L.C. van der Gaag, W. Ouweltjes, H. Mollenhorst, and H. Hogeveen. 2010. Discriminating between true-positive and false-positive clinical mastitis alerts from automatic milking systems. J. Dairy Sci. 93:2559–2568. doi:10.3168/jds.2009-3020.

Sun, Z., S. Samarasinghe, and J. Jago. 2009. Detection of mastitis and its stage of progression by automatic milking systems using artificial neural networks. J. Dairy Res. 77:168. doi:10.1017/s0022029909990550.

USDA–APHIS–VS–CEAH–NAHMS. 2014. Health and Management Practices on U.S. Dairy Operations. USDA.

Ushikubo, S., C. Kubota, and H. Ohwada. 2017. The Early Detection of Subclinical Ketosis in Dairy Cows Using Machine Learning Methods. Pages 38–42 in Proceedings of the 9th International Conference on Machine Learning and Computing. ACM, New York, NY, USA.

Vasquez, A.K., D.V. Nydam, C. Foditsch, M. Wieland, R. Lynch, S. Eicker, and P.D. Virkler. 2018. Use of a culture-independent on-farm algorithm to guide the use of selective dry-cow antibiotic therapy. J. Dairy Sci. 101:5345–5361. doi:10.3168/jds.2017-13807.

Tables

Table 7-1. Confusion matrices for models with Se = 80% and Sp = 80% for datasets with varying incidence rates of clinical mastitis. Percentages reported are proportion of dataset that would predicted with the given label by the model.

		True Negatives	True Positives
CM Rate = 50%	Predicted Negatives	40%	10%
	Predicted Positives	10%	40%
CM Rate = 5%	Predicted Negatives	76%	1%
	Predicted Positives	19%	4%

Table 7-2. Summary of variables used in models before cleaning and standardization.

Variable	n[1]	mean	sd	min	max	% NA[2]
DIM	135596	173.6	110.5	0	687	0.4%
Lactation number	135596	1.994	1.091	1	6	0.4%
Milk yield	131920	33.69	11.15	0.209	108	3.1%
Conductivity	127741	9.413	0.906	5.7	22.7	6.2%
Fat	131610	3.994	0.677	0.9	9.12	3.3%
Protein	131099	3.017	0.519	0.21	6.72	3.7%
Lactose	131169	3.860	0.745	0.1	11.2	3.6%
Milking time (s)	131117	353.7	116.8	10	1290	3.7%
SCC	126431	102144	7260742	1	1.46E+09	7.1%
Live weight (lb)	122487	1364	244.9	719	2198	10.0%
Activity	135005	124	55.04	1	1136	0.8%
TotalRestTime (s)	122006	725	150	3	1466	10.4%
DailyRestBout (n)	122980	11.62	11.8	0	508	9.7%
Rest duration (s)	121818	71.94	28.72	0.123	733	10.5%
CM, Y/N	136127	0.0047	0.068	0	1	0.0%
Repeat Case (n)	218	2.243	1.375	-5	4	-

Days between repeat CM	202	73.97	53.07	11	223	-
Day relative to CM	3551	-0.112	8.221	-14	14	-

[1]n = number of instances
[2]% NA = Percentage of given data that was missing for a given variable. Values not given are for variables that were not expected every milking.

Table 7-3. Summary of standardized values after data cleaning; AS = Activity Score; CS = Conductivity Score; FS = Fat Score; LS = Lactose Score; MY = Milk Yield; PS = Protein Score; RBS = Rest Bout Score; RDS = Rest Duration Score. Number suffixes refer to slopes from day 7 to n (e.g. AS1 is slope of activity from day 7 to 1).

Variable	n	mean	sd	min	max
Weight	103,007	-0.0005	0.999	-2.586	3.968
GramNeg	110,156	0.0008	0.029	0.000	1.000
GramPos	110,156	0.0022	0.047	0.000	1.000
AS1	110,145	-0.0007	0.107	-2.349	5.684
AS2	110,145	-0.0005	0.134	-2.679	5.684
AS3	110,121	-0.0003	0.168	-3.158	5.684
CS1	110,145	0.0021	0.093	-2.402	1.526
CS2	110,145	0.0013	0.116	-2.712	1.892
CS3	110,121	0.0005	0.150	-4.803	2.543
FS1	109,983	0.0036	0.141	-2.387	2.280
FS2	109,808	0.0029	0.171	-3.367	2.602
FS3	109,601	0.0020	0.223	-4.652	5.127
LS1	110,134	0.0001	0.138	-3.268	2.644
LS2	110,134	0.0009	0.169	-3.268	3.305
LS3	110,110	0.0009	0.214	-5.061	3.305
MYS1	110,145	0.0012	0.074	-2.022	1.830
MYS2	110,145	0.0020	0.090	-2.022	2.344
MYS3	110,121	0.0029	0.115	-2.022	3.123
PS1	110,145	0.0021	0.196	-3.812	4.892
PS2	110,145	0.0067	0.258	-3.812	5.030
PS3	110,121	0.0106	0.332	-3.812	5.030
RBS1	110,145	-0.0018	0.132	-3.644	2.603
RBS2	110,145	-0.0021	0.166	-3.644	2.863
RBS3	110,121	-0.0015	0.216	-3.644	2.993
RDS1	110,145	0.0037	0.116	-1.895	2.581
RDS2	110,145	0.0033	0.143	-2.627	2.581
RDS3	110,121	0.0032	0.184	-3.366	2.581
RTS1	110,145	0.0037	0.130	-2.469	2.577

RTS2	110,145	0.0035	0.169	-2.906	2.577
RTS3	110,121	0.0037	0.228	-4.297	3.300

Table 7-4. Partial costs matrix or change in the net return of an animal classified as either CM positive or negative, given that the animal was either CM positive or negative.

	True Negatives	True Positives
Predicted Negatives	$0	$0
Predicted Positives	-$165	+$132

Table 7-5. Median sensitivity and specificity values (interquartile range) for 100 models build on 75% of data and tested on 25% of held-out data for Gram-positive and Gram-negative prediction models.

	Sensitivity	Specificity
Gram-Negative	0.5263 (0.4615-0.6087)	0.9978 (0.9976-0.9979)
Gram-Positive	0.3750 (0.3367-0.4152)	0.9985 (0.9983-0.9987)

Figures

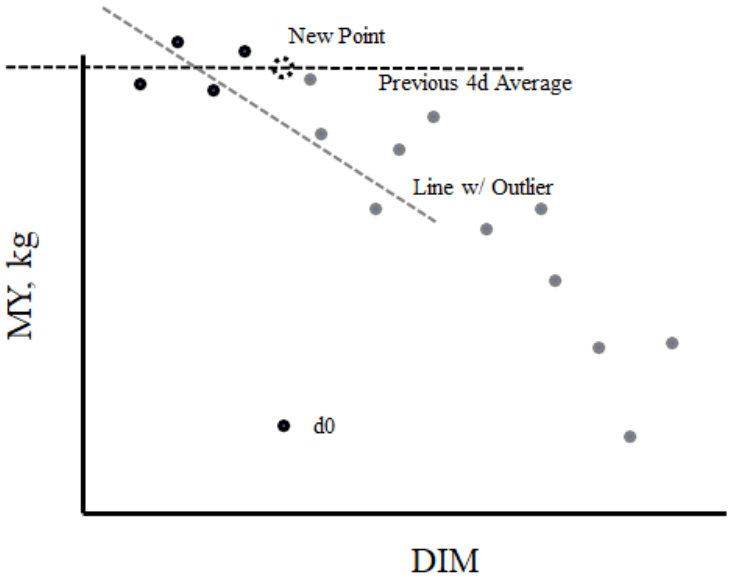

Figure 7-1. Example of removing points by absolute distance from average measurements of four previous days in milk (DIM). Here, the d0 measurement is well within the range of measurements for the entire dataset, but clearly abnormal within the trend of milk yield (MY) over time. This method detected outliers that were erroneous relative to the cow's DIM.

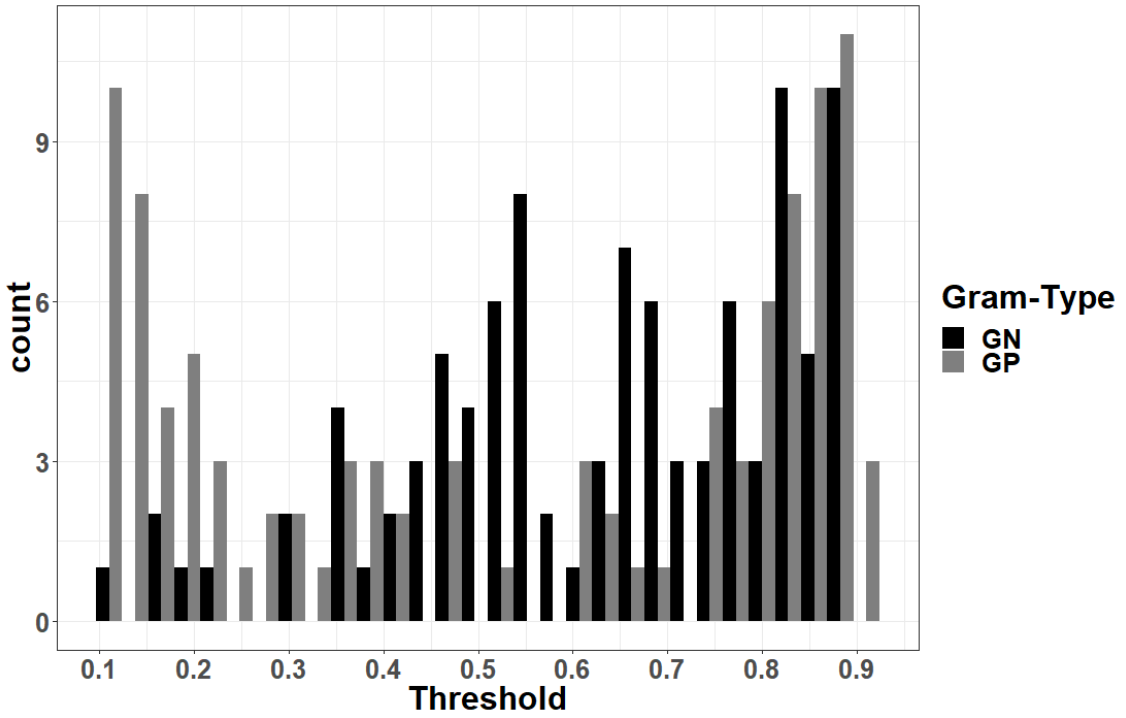

Figure 7-2. Optimal thresholds for differentiating disease and non-disease cases in Gram-negative (GN) and Gram-positive (GP) models. Each model was derived from a sample of the total data and thresholds were derived using the remaining data along with a cost matrix.

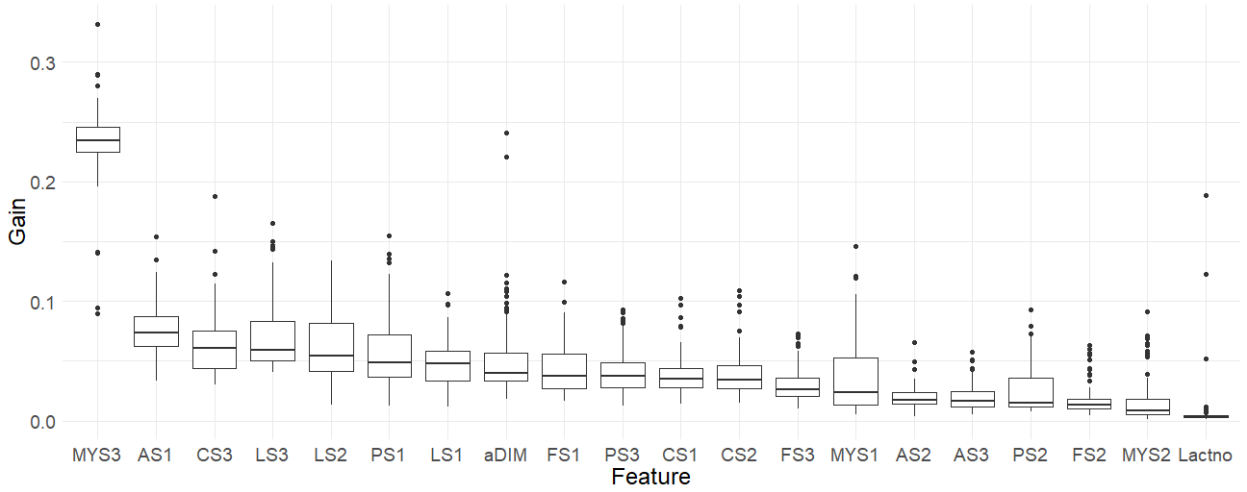

Figure 7-3. Model gain for all variables in the Gram-negative CM models. Arranged by median gain over the 100 model iterations tested on held-out data. All variables are the change (slope) of the variable over time. MYS = Milk Yield; AS = Activity (steps); CS = Conductivity; LS = Lactose; PS = Protein; FS = Fat; aDIM = adjusted DIM (7 bins); Lactno = Lactation number (1st, 2nd, 3+). The number following the abbreviation indicates the number of days prior to the day of prediction that the slope is estimated.

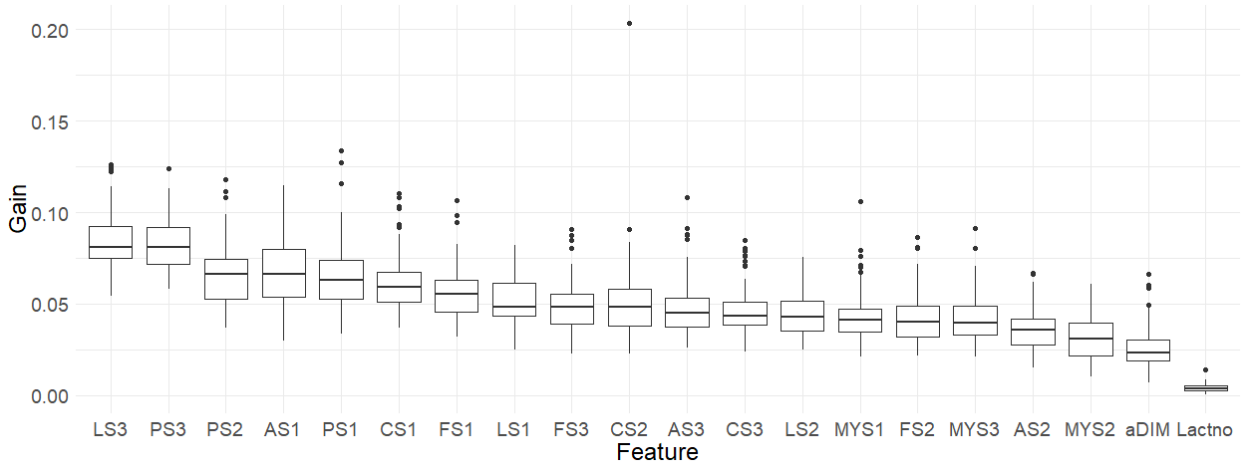

Figure 7-4. Model gain for all variables in the Gram-positive CM models. Arranged by median gain over the 100 model iterations tested on held-out data. All variables are the change (slope) of the variable over time. MYS = Milk Yield; AS = Activity (steps); CS = Conductivity; LS = Lactose; PS = Protein; FS = Fat; aDIM = adjusted DIM (7 bins); Lactno = Lactation number (1st, 2nd, 3+). The number following the abbreviation indicates the number of days prior to the day of prediction that the slope is estimated.

Conclusion

Ultimately, this body of work aimed to demonstrate some of the many ways that data analytics can be applied to make better management decisions on livestock operations. In this vein, many different facets of data analytics were explored, including model building, model evaluation, data collection, hypothesis testing, and data parsing. In the field of agricultural research, we are only scratching the surface of the analytical techniques available. Agriculturalists can learn from other fields, like data science, computer science, and statistics to combat challenges associated with greater amounts of data, different sources and types of data, and using data to forecast.

Machine learning is one specific field of data analytics that shows promise in solving complicated problems and being integrated into decision making systems. Existing research seeks to leverage computer vision to identify locomotive disease in cattle on-farm and crop density in fields, just to illustrate some examples.

In the first chapter of this work, simulation modeling was done to approximate the effects of removing dairy cows and dairy products from the U.S. The simulations are an example of how management decisions can be analyzed in a mathematical framework to inform policy decisions and make predictions about drastic

changes to the U.S. agricultural landscape. Next, specific analytical techniques for common management practices like feeding and disease detection were explored. Machine learning techniques were discussed that are uncommon currently in agricultural research, but more common in computer science and data analytics, like neural networks. The concept of reinforcement learning in machine learning was also discussed, illustrating a path between understanding and implementing machine learning algorithms into on-farm decision-making processes.

We next used data analytics to coalesce information on a management decision, grazing beef cattle, and showed how to use this information to improve decision-making. By understanding the effect of unreported study standard-error on our predictions' precision, we are ultimately making smarter, more economically-sound decisions. We also examined how to use data analytics to assess current best-practices, in the form of feed digestible rumen-undegradable protein measurements. These findings have implications on future feed testing and the accuracy of those results, with feed composition being an important factor in formulation rations for animals. Lastly, we applied XGBoost, a popular boosting algorithm, to the prediction of sparse clinical mastitis cases. Because producers are economically incentivized to only treat mastitic cows that are

sure to need treatment, incorporating case-specific knowledge was imperative to creating useful models. This example showed the importance of synergizing algorithms and domain knowledge to create more useful predictions.

As we move into a research era involving more observations and readily available information, we must be even more mindful of the techniques we employ to analyze data and make predictions. Analytical techniques like machine learning are complicated but have potential to expand our capabilities to analyze data in agriculture. With this potential, there are also some risks, specifically misuse or misunderstanding of the techniques being used. Data analytics is a fledgling field, with advancements being made every day across many disciplines. In agriculture, there is a high ceiling for improving data analytical practices, but how researchers approach learning new techniques will determine the realized benefit.