

UC Davis

UC Davis Electronic Theses and Dissertations

Title

Model-Free Approaches To Recovering Complex Behavioral Patterns From Precision Livestock Farming Data Streams Using Unsupervised Machine Learning and Information Theory

Permalink

<https://escholarship.org/uc/item/15m6r6sz>

Author

McVey, Catherine G

Publication Date

2022

Supplemental Material

<https://escholarship.org/uc/item/15m6r6sz#supplemental>

Peer reviewed|Thesis/dissertation

**Model-Free Approaches to Recovering Complex Behavioral Patterns from Precision
Livestock Farming Data Streams Using Unsupervised Machine Learning and
Information Theory**

By

CATHERINE G. MCVEY
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Animal Biology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Kristina Horback, Chair

Fushing Hsieh

Carly Moody

Committee in Charge

2022

“That which is forced can never be beautiful” – Xenophon

“Abandon all models, ye who enter here” – Fushing Hsieh

Contents

Abstract	v
Acknowledgments	vi
INTRODUCTION	1
<i>Comparing Model-Dependent and Model-Free Approaches to Information Compression</i>	5
<i>Comparing Model-Dependent and Model-Free Approaches to Identifying Bivariate Associations</i>	16
References	23
OBJECTIVES	29
CHAPTER 1	31
<i>Mind the Queue: A case study in visualizing heterogeneous behavioral patterns in livestock sensor data using unsupervised machine learning techniques</i>	
Introduction	31
Methods	34
Results & Discussion	49
Conclusions	72
References	74
CHAPTER 2	81
<i>Livestock Informatics Toolkit: A Case Study in Visually Characterizing Complex Behavioral Patterns across Multiple Sensor Platforms, Using Novel Unsupervised Machine Learning and Information Theoretic Approaches</i>	
Introduction	81
Materials & Methods	84
Results & Discussion	105
Conclusion	126
References	128
CHAPTER 3	135
<i>Tensor Mechanics: A case study in recovering systematic temporal patterns from multivariate data streams by employing unsupervised machine learning approaches in the analysis of daily time budget records</i>	
Introduction	135
Description of Data	137
Creating an Ensemble of Data Mimicries using Simulated Noise	138
Encoding Daily Time Budgets	139
Bivariate Associations	147
Conclusions	156
References	157

CHAPTER 4	161
<i>Model-free approaches to quantifying plasticity and resiliency in PLF data streams using unsupervised machine learning and information theoretic approaches</i>	
<i>Introduction</i>	161
<i>Description of Data</i>	164
<i>Fitting Lactation Curves</i>	166
<i>Creating an Ensemble of Data Mimicries using Simulated Noise</i>	169
<i>Information Compression</i>	172
<i>Cross-Sensor Inferences</i>	190
<i>Conclusions</i>	212
<i>References</i>	214

Abstract

Precision Livestock Farming (PLF) technologies can now allow ethologists to continuously monitor the behaviors of large numbers of animals over extended periods of observation. This creates new opportunities to study the behavioral dynamics of livestock in working farm environments, but also raises new methodological challenges. Large and densely sampled sensor datasets can contain a range of complex deterministic and stochastic structures that may difficult to fully accommodate in conventional model-based approaches. This thesis will explore the utility of model-free approaches to recovering complex behavioral patterns from within and across PLF data streams using unsupervised machine learning and information theoretic approaches. Chapter one will explore how data mechanics, an iterative hierarchical clustering algorithm, may be used to leverage the intrinsic behavioral codependences of socially housed animals to recover complex temporal patterns in the order in which cows enter the parlor to be milked, even when the environmental or managerial changes eliciting changes in queueing patterns are not recorded. Chapter two will introduce ensemble mimicry approaches, wherein minimally parametric simulation strategies may be employed to mimic the complex and often codependent sources of measurement error often found in PLF data streams, and demonstrate how this information may be infused into standard hierarchical clustering approaches to recover the most reliable patterns in overall time budgets. A model-free approach to recovering complex nonlinear bivariate associations between data streams is also developed using information theoretic approaches. Chapter three extends the data mechanics framework to recover temporal nonstationarity in multivariate records using tensor data structures to explore systematic patterns in daily time budget records. Finally, chapter four will explore model free approaches to quantifying the relative plasticity in biological and behavioral responses using novel information theoretic estimators, and further demonstrates how an ensemble mimicry may be employed to disentangle this information from sources of measurement error and bias in noisy PLF data streams.

Acknowledgements

First and foremost, I should thank my family, and especially my parents, for loving and supporting me through all my adventures, no matter how hare-brained they have seemed.

I should also thank several of my high school teachers – Mrs. Jones, Mrs. Shaw, Mr. Bob Gotwals – for seeing and nurturing in me the first sparks of a passion for research, again, no matter how hair-brained they may have seemed.

I want to thank a number of professors and mentors from my years as undergraduate, who continued to nurture and shape those research interests: my academic advisors Dr. Vivek Fellner and Dr. Spencer Muse, who helped ensure that I had the academic foundations necessary to make the most of my graduate training; my long-time research mentor Prof. Daniel Egger, who first taught me to approach complex problems in data analysis like an information theorist; and of course the many friends and mentors from my undergraduate internships at Scotland's Rural College and Riverlea Australia, who taught me how to collect quality data from large and often cranky animals without getting myself killed in the process; and Mrs. Judy Day and Dr. Richard Blanton, who helped make those internships possible.

I should thank my advisor and lab-mates from the Colorado State Dairy Systems Group (Pinedo Lab), for breaking me into life in academia during my masters, and for their continued support of my research as collaborators. My ability to move forward with this dissertation with minimal disruptions through the COVID lockdown is attributable in no small part to their incredible generosity in sharing their expertise and high quality data.

And I of course have many people at UC Davis to thank for helping to put the polish on me as a researcher during my PhD: the students and professors in the Center for Animal Welfare, for being a continuous source of personal and academic support and encouragement; and the professors that served on my various advising committees, especially the guidance and advice of Dr. James Fadel during the early part of my PhD. I will always be grateful to Dr. Fushing Hsieh, who patiently labored to help me unlearn much of my statistical education so that I might embrace these model-free approaches to data science. And finally, I have to thank my PI, Dr. Tina Horback, for believing in and always supporting my lofty and not always very coherent aspirations for this dissertation. You've continuously encouraged me to challenge convention and carve out my own niche, and for that I'll always be grateful.

Finally, I should thank my husband – often a source of good information theory references, and always a source of sanity. We did it!

INTRODUCTION

Model-free approaches to recovering complex behavioral patterns from Precision Livestock Farming data streams using unsupervised machine learning and information theoretic tools

Submitted as an Invited Review to the Journal of Applied Animal Science

What is the difference between data and information? Although these two terms are often used interchangeably in biology, entire subfields of data science are dedicated to solving this riddle (Kirby 2001; MacKay 2003). While mathematicians have developed formal definitions for these terms (Shannon 1948; MacKay 2003), we will attempt to illustrate the conceptual differences with a simple example.

Suppose a graduate student wants to analyze the gait dynamics of a dairy cow using a 120 Hz leg-mounted accelerometer. In developing their data collection protocol, they surmise that if one accelerometer would be good, then two accelerometers must be even better. For their pilot study, they strap two sensors onto the same leg of their favorite cow and record her walking normally down the alley way in a straight line. If both sensors were calibrated and synchronized properly, then these sensors might produce data that looks something like the plot in Figure 1A. From this graph, can we learn anything about this cow from the blue sensor that we could not have inferred from the red sensors (or visa versa)? Since these sensors are quite precise, the two signals are virtually identical. Thus, even though this graduate student has doubled the number of datapoints recorded by adding a second sensor, they have not really collected any additional information about the gait dynamics of this cow.

As this example has hopefully highlighted, data can be considered a physical resource – something that can be measured in bytes and memory space. Information, on the other hand, should be regarded as a more nebulous unit of measure that represents how much can be learned from a dataset. In experimental settings, collecting measurements is typically costly, so sampling strategies are often developed to minimize redundancy between datapoints, such that data and information are often functionally equivalent terms (Pinheiro and Bates 2000). With sensor technologies, however, sampling density is often dictated by the hardware, which can result in a considerable amount of redundancy between datapoints (Kirby 2001). With such datasets, information compression strategies are often implemented in an effort to reduce the size of a dataset without losing any potentially useful information about the system. Returning to our previous example, suppose the graduate realizes they have not requisitioned enough hard drives to store data from both accelerometers. They decide to average the output of the two sensors at each timepoint, as illustrated in Figure 1B. This simple information compression step reduces the size of their dataset by half, but no details about the signal have been sacrificed.

Now let's suppose that the graduate returns to the dairy to confirm that this protocol will also work on heifers. In the struggle that ensues to attach the accelerometers, one sensor is accidentally strapped on upside down. As a result, the two signals become inverted, as illustrated in Figure 1C. Subsequently, if the previously validated rolling average filter were applied to this new dataset, the signals would cancel out, leaving only the measurement noise as shown in Figure 1D. Thus, this example serves to illustrate just how easily biologically relevant patterns can be lost when the assumptions employed in an information compression strategy do not match the data to which it is applied.

As the range of Precision Livestock Farming (PLF) technologies that are available to farmers to herds continues to expand (Ellen *et al.* 2019; Brito *et al.* 2020; Stygar *et al.* 2021), so

does the need for robust algorithms that are capable distilling the large quantities of data that these systems produce down to useable knowledge (Wathes 2008; McVey 2021). In data science, there is “no free lunch” – no single algorithm that can be expected to perform optimally in all scenarios (Wolpert and MacReady 1997). In this paper, however, we will demonstrate how model-free unsupervised machine learning approaches can be used to recover and encode complex behavioral patterns from datasets generated outside of controlled experimental contexts, where an appropriate model may not be immediately obvious. We will then explore how model-free information theoretic approaches may be used to recover complex nonlinear associations between farm records and multi-sensor systems in order to provide more comprehensive ethological insights.

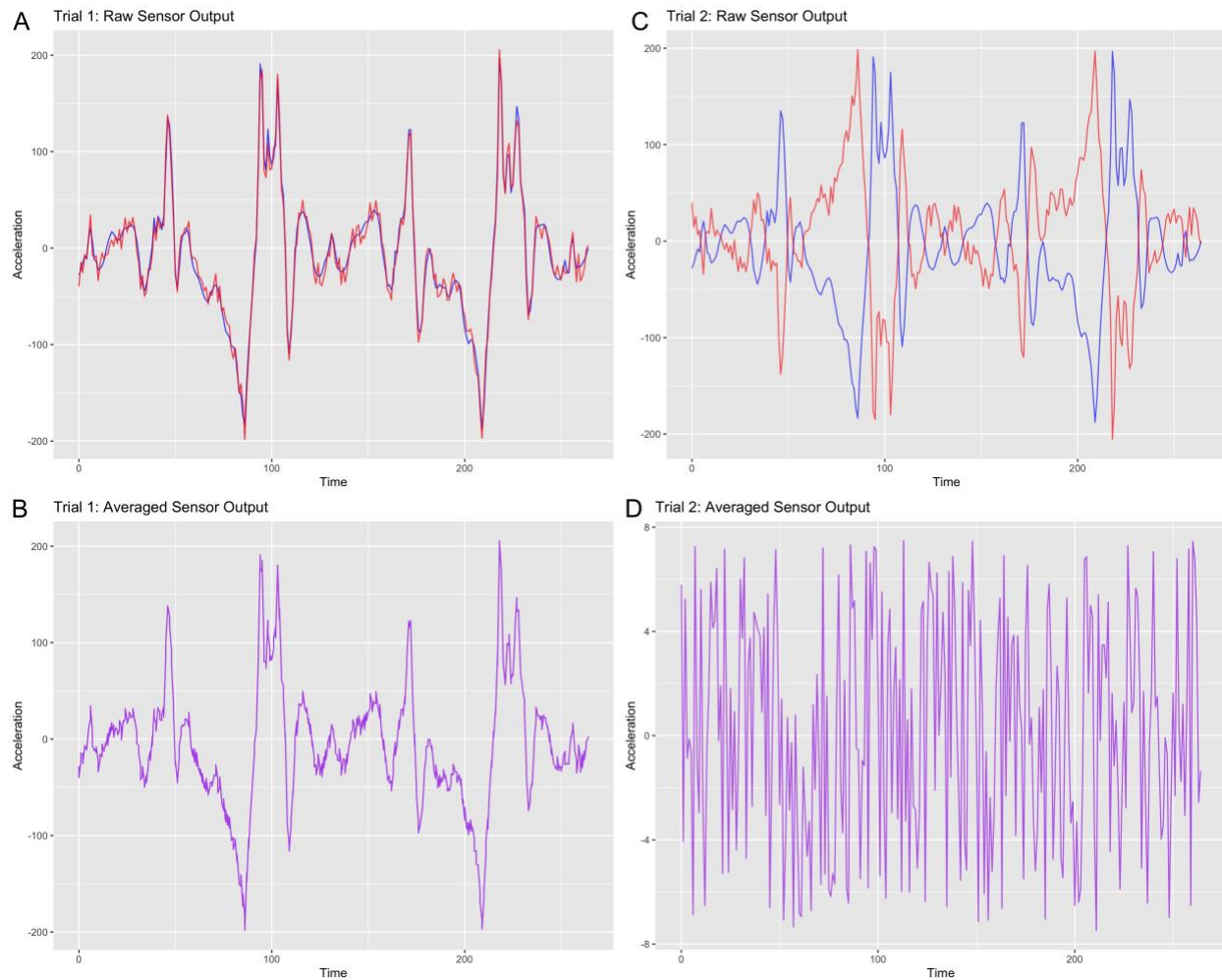


Figure 1: Human gait data collected from the forward axis of a 120Hz triaxial accelerometer has been modified to help visualize the difference between data and information. Upper Left (A): The data collection protocol is implemented flawlessly, such that the red and blue sensors produce nearly identical signals. Lower Left (B): Averaging their results at each time point reduces the size of the dataset by half with no appreciable loss of detail in the signal. Upper Right (C): One sensor is strapped on upside-down so that the signals become inverted. Bottom Right (D): Assumptions employed by this simple information compression algorithm are now violated. In reducing the dataset by half, all information about gait dynamics is lost, the signals having canceled out to leave only measurement noise.

Comparing Model-Dependent and Model-Free Approaches to Information Compression

Animal scientists are seldom forced to work with raw accelerometer data, as in the previous example. Therefore, to further illustrate the fundamental differences between model-dependent and model-free approaches to information compression and knowledge discovery, let's consider a more practical example. Suppose a farmer has a group of 100 cows that are currently overstocked at 200% stocking density in a freestall barn. Concerned they might fail the outcome measures on their upcoming welfare audit, the farmer contracts a consultant to analyze data from the leg-mounted accelerometer system that the farm uses in its estrus detection program. This system provides daily estimates of the proportion of time that each animal spends lying down. The farmer would like to know if there is any clear evidence of animals with compromised welfare in 60 days of archival records? Through structured random sampling from Gaussian distributions, we've simulated a dataset that might be realized from such a scenario. Full details of this simulation are provided in Supplemental Materials, but to avoid spoiling the surprise, consider first how to approach the analysis of this dataset when ignorant of its underlying structure.

Let's suppose the consultant hired is an animal scientist who, while not necessarily a specialist in welfare and behavior, still has plenty of quality training in scientific method, and wants to answer this farmer's question with full statistical rigor using a mixed-effect linear model. They begin by implementing standard exploratory data analysis (EDA) techniques, and produce the plot in Figure 2, wherein each dot represents the proportion of time a given cow is recorded lying down on a given observation day, with a LOESS curve fitted to reflect the mean lying time for the entire herd at each timepoint. From this visualization they glean that there is a considerable amount of variation between cows, and that there certainly are animals on any given day that are not spending a sufficient amount of time lying down (Tucker *et al.* 2021). There is not, however,

any clear evidence in this graph that would indicate that lying time patterns are changing at the herd level over time - quite the opposite actually, the herd mean is incredibly stable over the observation interval.

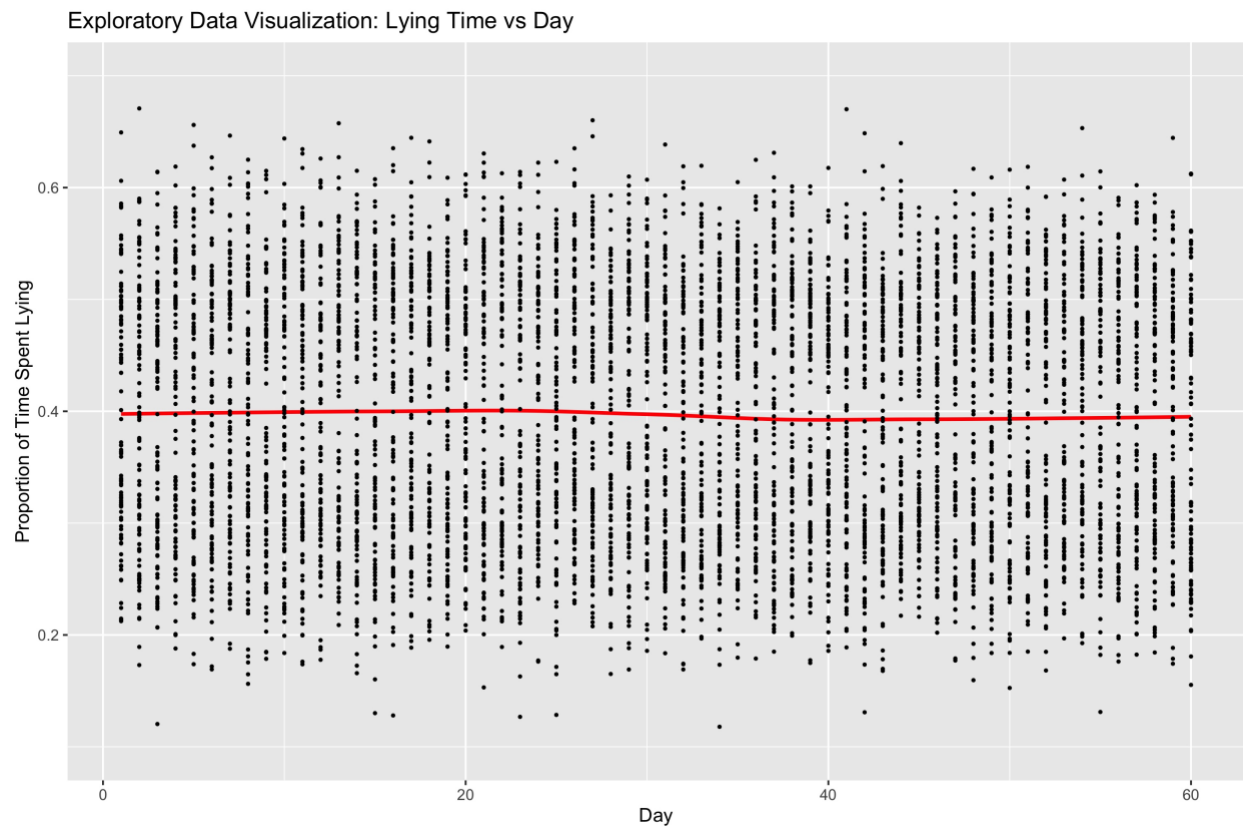


Figure 2: Exploratory Data Analysis visualization of simulated dataset. Each dot represents the proportion of time an individual cow is recorded as lying down on a specific day. The red LOESS fit line represents the average amount of lying time across the herd for a given day.

Based on this visualization, the consultant constructs a mixed effect model with “cow” as the random effect, to account for repeated measurements and avoid pseudo-replication. Since the data appears stationary, they add “day of observation” as a simple continuous fixed effect. Finally, based on their preexisting knowledge of factors that may affect lying time, they add a categorical fixed effect to distinguish between “heifers” and “cows” within the herd. The results of this model

are provided in full in Supplemental Materials. As anticipated from the EDA visualizations, the fixed effect for day is neither biologically nor statistically significant ($\hat{B} = -0.0001$, $SE = 8.53e-5$, $T = -1.41$). The consultant is, however, a bit surprised to find that the difference in average lying time between heifers and cows is also neither statistically or biologically significant ($\hat{B} = -0.0026$, $SE = 3.13e-3$, $T = 0.83$). Thus, the consultant determines that the expected proportion of time spent lying for any animal within this herd for any given day within the observation window is 40% ($SE = 0.34\%$). In appraising the variance estimates for the random effects in this model, the consultant also sees that between-cow variance ($\hat{\sigma} = 2.67e-5$) is considerably smaller than the within-cow error term ($\hat{\sigma} = 1.31e-2$), from which they conclude that there is also no evidence of consistent individual differences in lying time within this herd, and so these results also do not indicate that any individuals are consistently above or below the expected lying time for this herd. Subsequently, they might confirm that, while cows might not get to lie down an adequate amount of time every day, and that the average lying time could certainly be improved, there is also no evidence of cows in this herd that are consistently, and critically, under-rested.

Let's now suppose that the milk buyer is dubious of this result, and so they send the data to a new consultant who specializes in welfare audits. They ask the farmer some additional questions about this herd before constructing their own model and learn that the start of grazing season fell somewhere in the middle of this observation window, at which point cows had free access to pasture from their freestall barn. This consultant, suspecting that heifers and cows might react differently to this pasture access, develops a similar model to the first consultant, except this time an interaction term is included to allow the temporal dynamics to differ between these subgroups. With this small modification to the fixed effects matrix, this model now tells an entirely different story. The discrete effect for parity, the linear term for day, and the interaction effect all

now have a biologically and statistically significant impact on expected lying time. In visualizing this dynamic in Figure 3, the consultant can now clearly see that during the first half of the observation interval, the mature cows are hogging the free stall spaces, forcing the heifers to stand; however, once the grazing season starts, the cows leave their free stalls to go graze, while the now foot weary heifers remain in the free stalls.

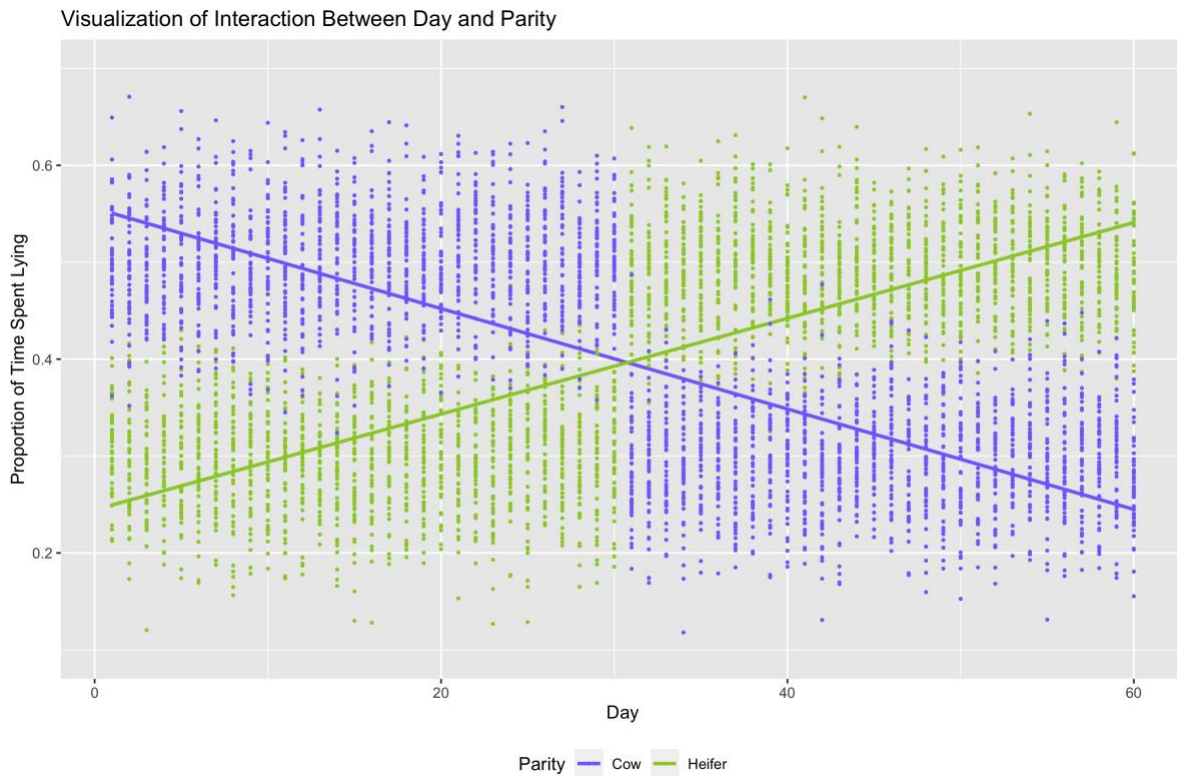


Figure 3: Visualization of the temporal complexity hidden within this dataset. Due to the overstocking in this pen, a strong negative correlation is found between subgroups as cows must compensate for the limited number of freestall spaces. When this physical limitation is removed with the reintroduction of pasture access, this dynamic becomes inverted. If a linear model is not structured to anticipate this inversion, the symmetric nature of this dataset will cause this pattern to be wholly overlooked.

In the preceding analyses, it's important to distinguish between linear models and statistical inferences as two complementary but distinct steps. A linear model is ultimately just a way to produce estimates of mean and variance that are conditional on additional variables. Further probabilistic assumptions can be made in order to draw statistical inferences about these estimates, but they typically are not employed in the estimation step itself. In other words, linear models are just a means of generating summary statistics. And all summary statistics are a form of information compression (Agresti 2013). Thus, we can once again see in this example that, when the biological system that generates a dataset is well understood, this prior knowledge can be infused into an appropriate model to efficiently compress the information contained in such records. If, however, we have an incomplete understanding of the system that gives rise to the behavioral responses, a linear model can become an inefficient means of data compression that hemorrhages information, potentially causing important patterns to go overlooked.

In the preceding example, the simulated behavioral response was intended to be a bit tricky to capture, but it was still fully describable by explanatory variables – date of observation and parity – available to the consultants. But what if this same mechanism driving inversions in lying patterns had been caused by more complex factors? What if, instead of a single persistent shift in lying patterns after pasture access was granted, the desirability of the free stalls had been influenced by more transient environmental factors such as weather or bedding cleanliness? Alternatively, what if an animal's ability to monopolize these free stall spaces had not been determined by a simple trait like parity? Resource holding potential is known to be influenced by a range of factors – size, seniority, energy balance, health status, innate aggressiveness etc – many of which might not be directly measured in standard production systems (Schein and Fohrman 1955; Hurd 2006; Sarova 2013; Hubbard 2021). Clever model parameterization may still succeed

in capturing any one of these patterns, but as such complexities begin to compound within a system, it can become overwhelming to account for all such contingencies within a single model. Subsequently, the farther we move from controlled experimental contexts towards the chaos and complexity of commercial farm environments, the more fundamentally challenging it becomes to use model-based approaches to information compression to extract ethological insights from PLF data streams.

Looking beyond linear models, machine learning (ML) approaches may provide a means to overcome such gaps in background knowledge by more fully leveraging the power of modern computing (Valletta *et al.* 2017). Such algorithms are divisible into two general classes (Kirby 2001; James *et al.* 2013; Valletta *et al.* 2017). In supervised machine learning, the exact structure of the model is, to varying degrees, gleaned algorithmically from the data by itself; however, the user is still required to distinguish between explanatory and response variables. Subsequently, this framework lends itself to optimizing the predictive power of a model, but may still overlook important patterns in a dataset if the factors driving heterogeneity (systematic non-randomness) in the response variables are not measured and included amongst the candidate predictor variables. Unsupervised ML algorithms, on the other hand, employ a more open-ended approach to information compression. Such algorithms do not distinguish between explanatory and response variables, but seek only to progressively extract and visualize the most striking nonrandom features of a dataset until only noise remains.

Several algorithmic approaches exist to explore the latent high dimensional geometry of large datasets via UML. Neural networks are arguably the most powerful of such techniques (Goodfellow *et al.* 2016). While unsupervised neural networks are not as data-hungry as their supervised cousins, their ground-up approach to learning the key features of a dataset still requires

many hundreds to many thousands of observational units. Future advancements in transfer learning techniques and in our understanding of neural architectures may eventually lower the sample sizes required to train such networks (Arac 2019; Mathis and Mathis 2020; Andersen *et al.* 2021), but for many behavioral applications where data may only be available from a single farm or even a single group of animals, less data-hungry algorithms are typically required.

Spectral embedding approaches offer a method of information compression that is suitable for many “wide” problems, where the number of variables in a dataset may be larger than the number of observational units (Kirby 2001). Principal Component Analysis (PCA) is the simplest example of this class of algorithms, but more modern approaches are more suitable for the complex nonlinear geometric features that can often be found in sensor data. Alternatively, several algorithmic strategies may be employed to compress the complex geometric features of high dimensional datasets into discrete clusters (Valletta et al. 2017; Adamczyk et al. 2017; Rufener et al. 2019; Xu et al. 2020; Sibanda et al. 2020; Hou et al. 2020; Clouard et al. 2022; Franceschini et al. 2022). K-means clustering is arguably the simplest to implement, and subsequently has seen some adoption in analyses of livestock data. While no overt model is stated in this approach, the users still must specify *a priori* the appropriate number of clusters needed to represent the latent structures in the data – a metaparameter choice that may not be immediately obvious in all applications. For such datasets, Hierarchical Clustering (HC) algorithms may provide a more open-ended approach to developing an optimal discrete encoding of such high dimensional patterns.

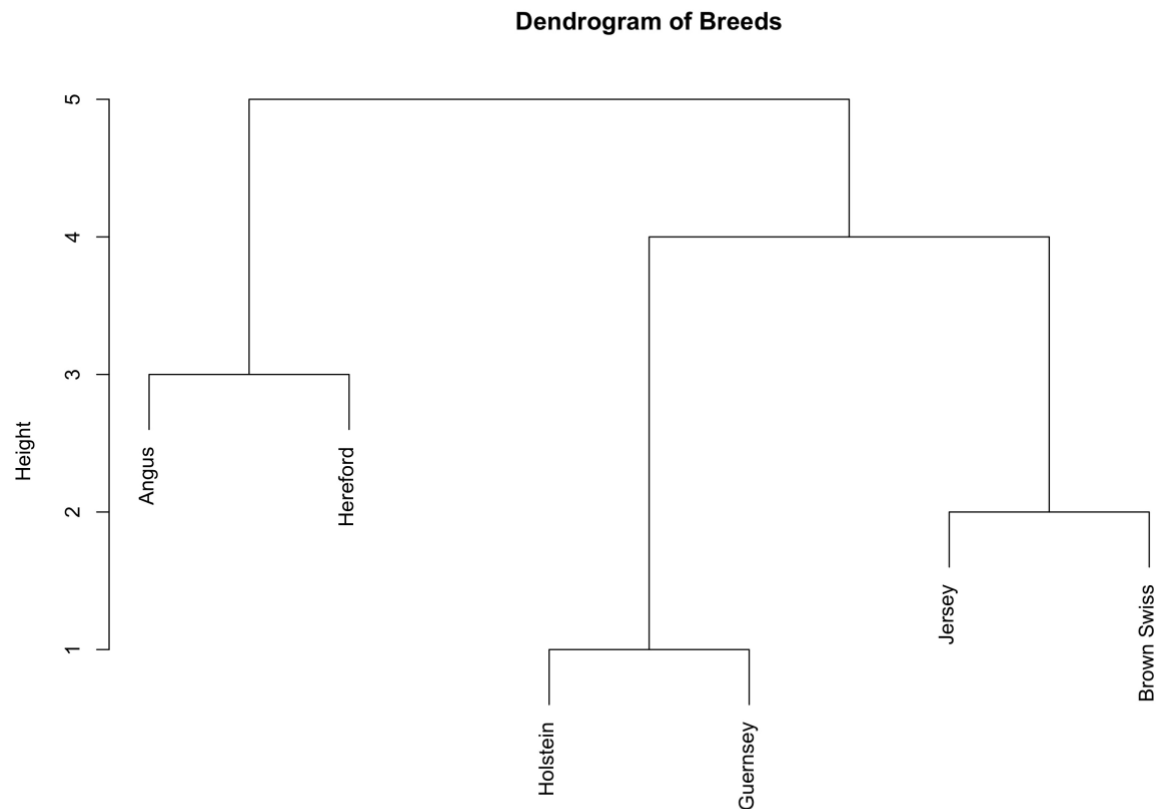


Figure 4: Through sequential agglomerative steps, breeds are progressively paired together based on their relative similarity. The resulting dendrogram provides a 2D visualization of how these breeds are distributed within a much higher dimensional space of phenotypic traits.

As its name implies, HClustering is an intuitive analytical framework that progressively groups data through a series of sequential agglomerative steps (James et al. 2013). To illustrate how this algorithm works, suppose you go to the feed store and get a pack of little plastic toy cows. You line them up in front of a toddler and ask what cows look most alike. They point to the Guernsey and the Holstein, so you pull them forward and stand them together. You ask this question again, they point to the Jersey and the Brown Swiss, so you pull these forward and stand

them together. You ask again, and the Angus and Hereford cows are paired together. You ask again, and the two groups of dairy cows are pushed together. And finally, if you still command said toddler's attention, you end up through one final pairing with all your cows in one pile. If you kept track of all these pairings, you might produce a visual summary of this decision-making process presented in Figure 4. This schematic, properly called a dendrogram, provides a succinct 2D representation of how a fairly large number of phenotypic features are distributed within your plastic herd.

Hierarchical clustering algorithms seek to mimic this agglomerative process, which we can implement fairly intuitively, using objective mathematical constructs. To demonstrate how this might work, let's return to our previous example of the farmer with overstocked dairy cows. As before, our larger mature cows will monopolize the free stalls when they are not on pasture. In this example, however, we will increase the complexity of this analytical problem by now supposing that the animals always had access to pasture, and that rain events were the environmental factor driving inversions in lying times – inversions that will now be randomly scattered throughout the observation window. The first step in hierarchically clustering this dataset is to compute a dissimilarity matrix, which is a square symmetric matrix containing quantitative estimates of the dissimilarity between each pair of observational units in the corresponding row and column indices. In order to cluster cows together with similar lying patterns, we will first calculate a dissimilarity matrix using the Euclidean distance or L2 norm, which here is just the squared differences in observed lying times between a given pair of cows summed over all observation days. Similarly, in order to cluster days together wherein the herd demonstrated similar lying patterns in response to environmental factors, we'll calculate the Euclidean distance between each pair of observation days as the sum of squared differences in lying time over all animals in the

herd. Using these pairwise dissimilarity values, a ground-up agglomeration algorithm can then be applied. Here we will use Ward's (D2) linkage method, wherein clusters are merged in order to produce the largest increase in between-group variance.

In order to visualize the results of both clustering routines simultaneously, a heatmap can be used to directly visualize this simulated data matrix. Here each row will correspond to a cow, each column will correspond to an observation day, and each cell will be colored to represent the proportion of time that a given cow spends lying down on a given day. With rain events scattered randomly throughout the observation window, and cow records provided in no particular order, we see in visualizing the raw data matrix in Figure 5A that the inversions in lying patterns are completely obscured. The dendrograms produced by clustering cows over days and also days over all cows can be used to reorder the rows and columns indices of the raw data matrix so the cows with similar temporal patterns in their lying times are grouped together on the row axis. Visualizing these results again using a heatmap in Figure 5B, we can now see that this clustering algorithm has captured two distinct groups of cows and two distinct groups of observation days, and so the inversion in lying patterns within the data are now quite visually striking. Thus, even though this hierarchical clustering algorithm was never provided information on the factors driving this behavioral pattern – cow parity and weather records – it has still succeeded in recovering the social dynamics hiding within this lying time data. Subsequently, even if no other farm records were made available to us, we would still be able to identify that overstocking is compromising the welfare of this herd using this model-free approach to information compression and knowledge discovery.

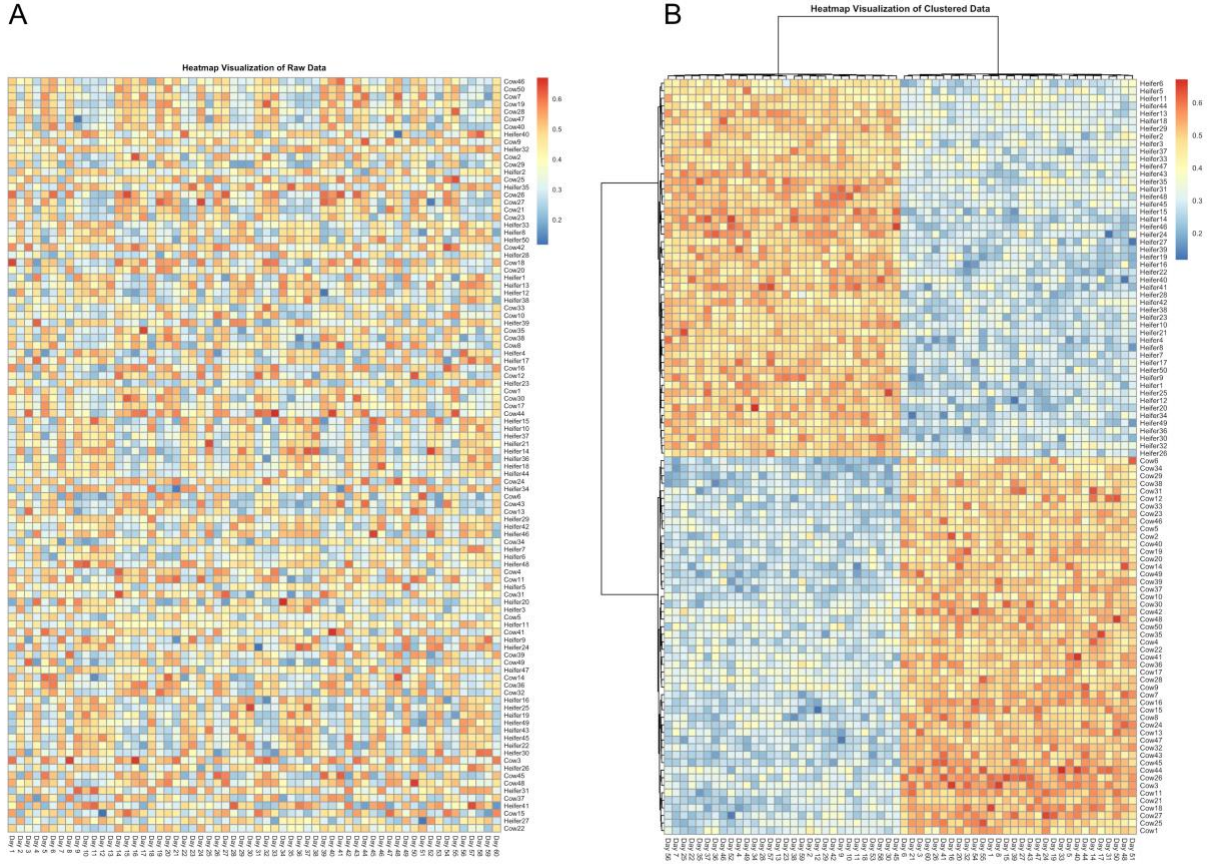


Figure 5: Heatmap visualization of the lying time data matrix. Cows are arranged on the row axis, observation days are arranged on the column axis, and the color of each cell represents the proportion of time that a given cow is recorded as lying down on a given day. In 5A (left) no clear patterns in lying time are visible, as the weather events driving inversions in lying time are distributed randomly over the observation window and cow records are provided in no systematic order. In 5B (right) results of hierarchically clustering cows over days and days over cows are used to rearrange the rows and columns of the data matrix, respectively, making it easy to visualize the clear inversions in lying dynamics within this herd even without having any data on the factors driving this dynamic.

Comparing Model-Dependent and Model-Free Approaches to Identifying Bivariate Associations

The preceding example has hopefully illustrated that, by leveraging intrinsic co-dependencies in the behavioral responses of socially housed animals, model-free machine learning approaches can recover complex behavioral patterns from sensor data absent any assumptions of the causative factors. In practice, however, once such a pattern is detected, we typically want to try to pin down the variables eliciting such reactions. In the preceding simulation, the mechanism linking lying time, age, and weather were pretty simple, and so armed with the insights from the UML analyses, a number of fairly straightforward linear methods might be employed to probe for the causative variables amongst farm records. But what if the link between environmental variables and the behavioral responses had been more complex? We've already seen in the previous example that model-based methods can overlook significant bivariate associations when the structure of the assumed model does not match the dynamics of the system, so let's then consider how model-free approaches might be extended to characterizing complex behavioral patterns not only within but also across datasets.

A number of supervised machine learning approaches are specifically designed to probe for significant associations amongst large sets of candidate predictor variables (James *et al.* 2013; Valletta *et al.* 2017). With LASSO regression, any nonlinear dynamics between candidate predictors and the response must be explicitly coded, whereas neural networks and regression trees can infer a range of nonlinear dynamics when properly parameterized. For all these methods, however, two potential drawbacks should be considered for behavioral applications. First, because the bias-variance tradeoffs of such models must be tuned through cross validation to avoid overfitting and spurious associations, the sample sizes required by these techniques to screen large

number of candidate predictors are often quite large as well (James *et al.* 2013). Even when sufficient sample sizes are available, such techniques inherently prioritize prediction over intuition. While variable importance estimates may provide some insights into the key drivers of a behavioral response, it can be difficult with these approaches to characterize the dynamics between predictor and response variables.

For applications where intuition is more important than prediction, information-theoretic approaches may provide a model-free approach to both identifying and characterizing bivariate associations between datasets that naturally compliments UML clustering techniques. Information entropy, as calculated using Equation 1, is an estimator designed to quantify the relative uncertainty in a categorical response variable (Shannon 1948; MacKay 2003). The discrete values that may be recorded for such a variable is often called the “vocabulary” or “dictionary”, and are here indexed as $k = 1 \dots K$, wherein K is the total number of discrete levels. For applications in animal behavior, however, this dictionary is typically just the ethogram used. While logarithmic terms are not typically known for their interpretability, here they confer entropy estimates with a number of convenient algebraic properties that makes this estimator intuitive as a relative measure.

$$H(x) = \sum_{k=1}^K P(X = k) \log\left(\frac{1}{P(X = k)}\right)$$

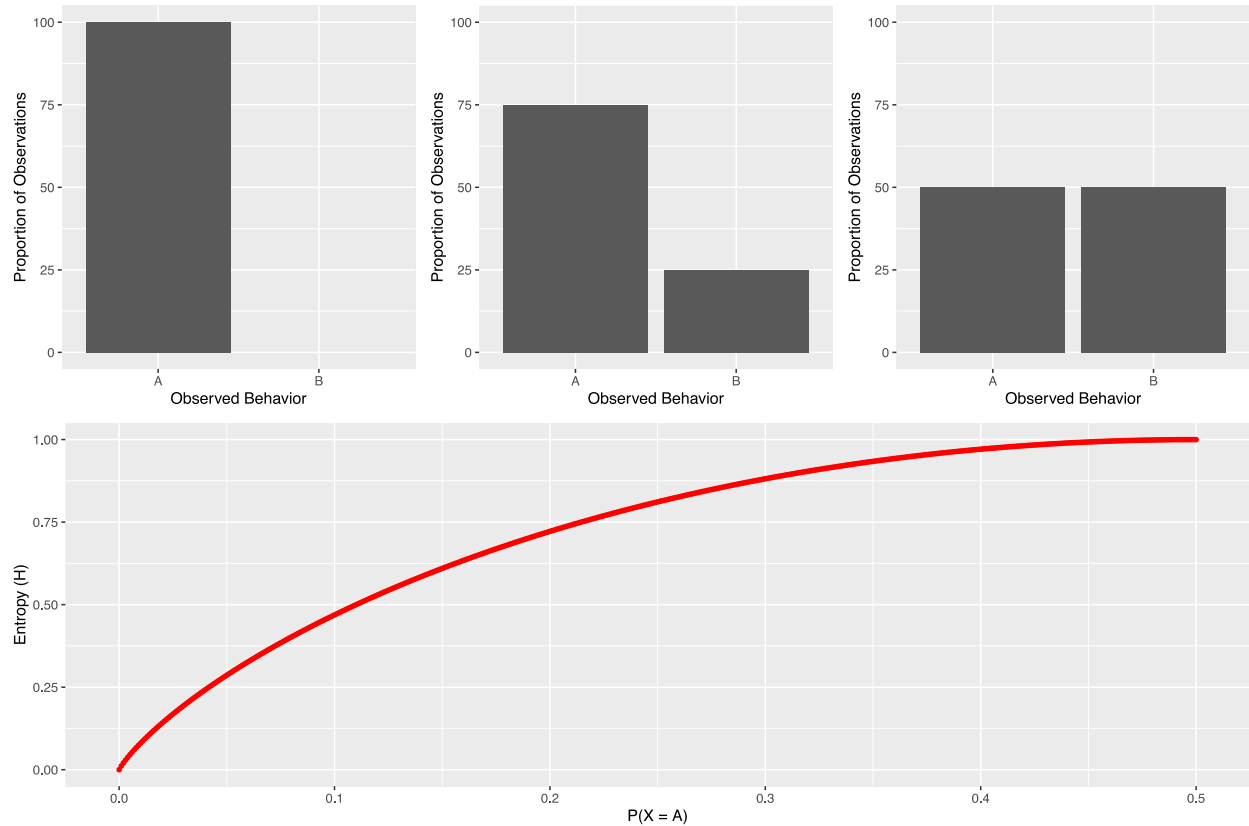


Figure 6: A visualization of how entropy estimates vary with changes in the relative distribution of discrete observations for a simple $K = 2$ encoding (ethogram). When behaviors A and B are recorded at equal frequencies $P(X = A) = P(X = B) = 0.5$, the entropy or uncertainty in this behavioral system is maximized at $\log_2(2) = 1$. As behavior A is observed more frequently, the entropy of the system progressively decreases, dropping to 0 when only behavior A is observed.

Suppose a dataset were collected wherein each observational unit is coded with one of two mutually exclusive behaviors – for example lying or not lying. If both behaviors are observed at equivalent frequencies, and we are provided no additional information than the distribution of this discrete variable, could we make any type of informed guess about what behavior we might see if we selected an observation at random from this data? Since no behavior is no more likely to be observed than another, we would be completely uncertain, and subsequently would compute a

maximum theoretical entropy of $\log_2(2) = 1$. Alternatively, if some behavioral mechanism caused this probability to shift towards or away from one of these two behaviors, then we might be able to venture a guess at what behavior we might expect to see if we picked an observation at random – our uncertainty would decrease and so too would the corresponding entropy estimate. Taken to the extreme, if only one behavior were ever observed, then we could guess the encoded value of any observation in our dataset with complete certainty, and so the corresponding entropy estimate would drop to zero. Figure 6 demonstrates how the entropy estimate of a bivariate variable changes with the symmetry of the distribution of the observed encoding. In calculating such entropy values, the resulting estimates are contingent only upon the assumptions used to develop the discretization scheme. No conditional mean or any other model need be assumed, as with a standard variance estimator utilized in conventional statistical inference frameworks for linear models.

This model-free framework can subsequently be extended to a multivariate estimator with two or more discrete variables. In the bivariate case, the distribution of one variable is compared across each encoded level of the other in order to decompose the total entropy in the joint encoding into three terms: the conditional entropy unique to the first variable, the conditional entropy unique to the second variable, and the mutual information that is redundant between the two encodings (MacKay 2003). This mutual information estimate in turn reflects how much information we learn about one encoded variable if we know the value of the other, and subsequently can be used to reflect the strength of a bivariate association between two sets of encoded data regardless of the underlying dynamic – linear, quadratic, exponential, etc.

To demonstrate the efficacy of an information-theoretic framework for identifying complex bivariate associations over more conventional model-based approaches, let's return to the previous example. Suppose our farmer with the overstocked cows, now fully aware of the welfare

issues this management choice has created, reduces their stocking rate to a 1:1 ratio and continues to monitor the lying time of the animals to provide proof to their milk buyer that the issue has been resolved. After several months at this lower stocking rate, they review the data and are dismayed to find that there are still days where animals have inadequate lying times. To solve this new problem, they hire yet another consultant to help them track down the source of the problem.

Suppose that this new dataset were collected in the summer, and so naturally, this consultant includes Temperature-Humidity Index (THI) as one of many candidate variables to consider as a potential source of the continued welfare concerns for this herd (Tucker *et al.* 2021). Using stochastic sampling techniques, the full details of which are provided in Supplemental Materials, we have simulated a fairly straight forward but nonlinear dynamic between these two variables. On days where the observed THI values are low, animals are not heat stressed and so spend the majority of their day out on pasture grazing. As the THI rises, cows become heat stressed for progressively larger proportions of the day, resulting in a gradual increase in the proportion of each day that cows spend lying down in the shade of their freestall barn. Above a certain high THI threshold, however, cows struggle to thermoregulate while lying down, causing them to stand for extended periods of time. When lying time is plotted against THI, as in Figure 7A, we can see that there is a clear nonrandom pattern in this data that is perhaps best characterized by a threshold model – a dynamic that is commonly found when a single behavioral response is subject to the influence of competing underlying behavioral response mechanisms. If a simple linear effect were utilized to probe for a significant bivariate association between these two variables, however, a near-zero slope would be returned, as shown by the red line overlaid in Figure 7A. In this case, not only would a Pearson correlation test ($R = -0.03$, $p = 0.25$) fail to identify this nonrandom but also nonlinear pattern, but because this pattern is also not monotonically increasing, even a

nonparametric Spearman Rank correlation test ($Rho = 0.03$, $p = 0.12$) would fail to identify THI as a significant influence on lying times within this herd.

Let's consider instead a nonparametric testing framework using information entropy as the test statistic (McVey et al. 2021). Suppose both the THI and lying time measurements are discretized using simple equal-sized binning rules. If we compare the mutual information estimate from the two observed encodings against estimates generated from a simple permutation, the resulting p-value for this test of bivariate association would be highly significant ($p \leq 0.001$). To further characterize this dynamic, a simple contingency table, wherein each cell represents the total number of observations for each joint encoding, can be easily visualized as in Figure 7B. The mutual information estimate for the overall bivariate association can subsequently be decomposed into pointwise mutual information estimates to reflect how much each cell in the observed table differs from the counts that would be expected by multiplying the marginal probabilities, which would be the distribution of joint observations anticipated if no association existed between the two encodings (see McVey *et al.* 2021 for further details). Here blue cells indicate that there are fewer observations with the corresponding joint encoding than would be expected if no association between these variables were present, whereas orange cells are over-represented relative to the null. From this visualization we can clearly see that the probability of observing a given lying pattern is being shifted in different directions based on the level of the THI encoding. Thus, absent any prior intuition or assumptions about the relationship between these two variables, an information theoretic approach has successfully identified a significant bivariate association and provided insights into the underlying dynamic to inform further interpretation of the underlying behavioral mechanisms at play and subsequently the correct management interventions needed to remediate this welfare concern.

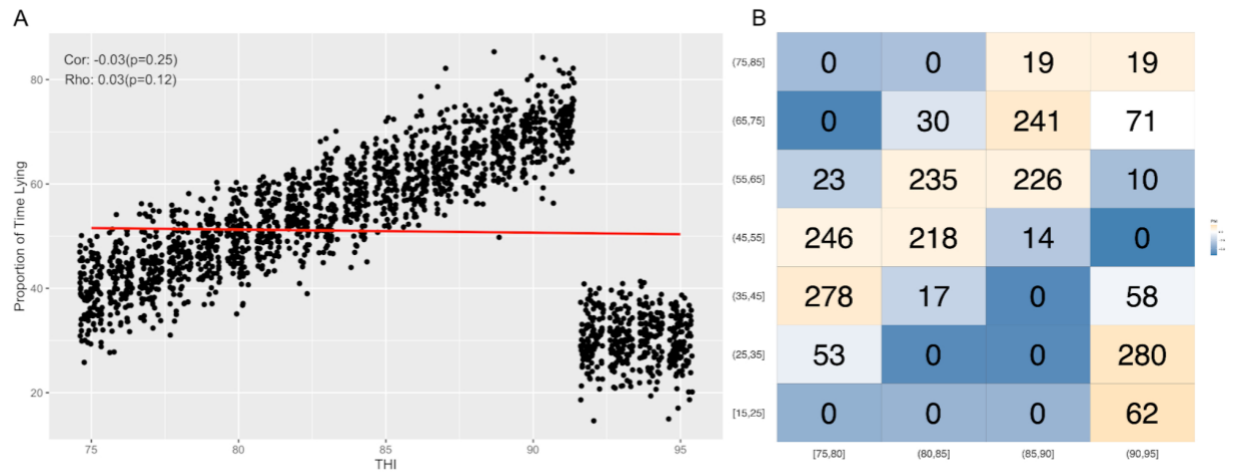


Figure 7: Two visualizations of the simulated relationship between THI and lying time. (A) A simple dot plot reveals a clear bivariate association that is likely best described by a threshold model, with cows becoming heat stressed and standing for extended periods of time in order to move heat more effectively at THI values above 92 degrees. This nonlinear dynamic, however, is not captured by a simple linear association assumed in the Pearson correlations test, which is visualized with the red line imposed over the data, or the monotonically increasing association assumed by the Spearman rank order test. (B) Both variables have been discretized and their joint distribution visualized using a simple contingency table. Cells are here visualized by their pointwise mutual information estimates, which reflects cells that are significantly over- or under-populated when compared to the product of the marginal probabilities, which is the value that would be expected if no bivariate relationship exists. Here we can clearly see that behavioral mechanisms are shifting the lying patterns of this herd in different directions at different levels of heat stress.

References

- Agresti, A., 2013. Categorical Data Analysis, 3rd ed, Wiley Series in Probability and Statistics. John Wiley & Sons, Inc, Hoboken NJ.
- Andersen, P.H., Broomé, S., Rashid, M., Lundblad, J., Ask, K., Li, Z., Hernlund, E., Rhodin, M., Kjellström, H., 2021. Towards Machine Recognition of Facial Expressions of Pain in Horses. *Animals* 11, 1643. <https://doi.org/10.3390/ani11061643>
- Arac, A., Zhao, P., Dobkin, B.H., Carmichael, S.T., Golshani, P., 2019. DeepBehavior: A Deep Learning Toolbox for Automated Analysis of Animal and Human Behavior Imaging Data. *Front. Syst. Neurosci.* 13, 20. <https://doi.org/10.3389/fnsys.2019.00020>
- Beggs, D.S., Jongman, E.C., Hemsworth, P.H., Fisher, A.D., 2018. Short communication: Milking order consistency of dairy cows in large Australian herds. *Journal of Dairy Science* 101, 603–608. <https://doi.org/10.3168/jds.2017-12748>
- Berry, D.P., McCarthy, J., 2012. Genetic and non-genetic factors associated with milking order in lactating dairy cows. *Applied Animal Behaviour Science* 136, 15–19. <https://doi.org/10.1016/j.applanim.2011.11.012>
- Berry, D.P., Veerkamp, R.F., Dillon, P., 2006. Phenotypic profiles for body weight, body condition score, energy intake, and energy balance across different parities and concentrate feeding levels. *Livestock Science* 104, 1–12. <https://doi.org/10.1016/j.livsci.2006.02.012>
- Brito, L.F., Oliveira, H.R., McConn, B.R., Schinckel, A.P., Arrazola, A., Marchant-Forde, J.N., Johnson, J.S., 2020. Large-Scale Phenotyping of Livestock Welfare in Commercial Production Systems: A New Frontier in Animal Breeding. *Front. Genet.* 11, 793. <https://doi.org/10.3389/fgene.2020.00793>

Ellen, E., van der Sluis, M., Siegford, J., Guzhva, O., Toscano, M., Bennewitz, J., van der Zande, L., van der Eijk, J., de Haas, E., Norton, T., Piette, D., Tetens, J., de Klerk, B., Visser, B., Rodenburg, T., 2019. Review of Sensor Technologies in Animal Breeding: Phenotyping Behaviors of Laying Hens to Select Against Feather Pecking. *Animals* 9, 1–21.

<https://doi.org/10.3390/ani9030108>

Enriquez, S., Hsieh, F., 2020. Categorical exploratory data analysis on goodness-of-fit issues.

Franceschini, S., Grelet, C., Leblois, J., Gengler, N., Soyeurt, H., 2022. Can unsupervised learning methods applied to milk recording big data provide new insights into dairy cow health? *Journal of Dairy Science* S0022030222003551. [https://doi.org/10.3168/jds.2022-](https://doi.org/10.3168/jds.2022-21975)

[21975](https://doi.org/10.3168/jds.2022-21975)

Friggens, N.C., Berg, P., Theilgaard, P., Korsgaard, I.R., Ingvarsten, K.L., Løvendahl, P., Jensen, J., 2007. Breed and Parity Effects on Energy Balance Profiles Through Lactation: Evidence of Genetically Driven Body Energy Change. *Journal of Dairy Science* 90, 5291–5305.

<https://doi.org/10.3168/jds.2007-0173>

Gadbury, J.C., 1975. Some preliminary field observations on the order of entry of cows into herringbone parlours. *Applied Animal Ethology* 1, 275–281. [https://doi.org/10.1016/0304-](https://doi.org/10.1016/0304-3762(75)90020-6)

[3762\(75\)90020-6](https://doi.org/10.1016/0304-3762(75)90020-6)

Gareth, J., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning, Springer Texts in Statistics. Springer New York, New York, NY. [https://doi.org/10.1007/978-](https://doi.org/10.1007/978-1-4614-7138-7)

[1-4614-7138-7](https://doi.org/10.1007/978-1-4614-7138-7)

- Guan, J., Hsieh, F., 2018. Coupling Geometry on Binary Bipartite Networks: Hypotheses Testing on Pattern Geometry and Nestedness. *Front. Appl. Math. Stat.* 4, 38.
<https://doi.org/10.3389/fams.2018.00038>
- Hsieh, F., Liu, S.-Y., Hsieh, Y.-C., McCowan, B., 2018. From patterned response dependency to structured covariate dependency: Entropy based categorical-pattern-matching. *PLoS ONE* 13, 1–28. <https://doi.org/10.1371/journal.pone.0198253>
- Hubbard, A.J., Foster, M.J., Daigle, C.L., 2021. Social dominance in beef cattle — A scoping review. *Applied Animal Behaviour Science* 241, 105390.
<https://doi.org/10.1016/j.applanim.2021.105390>
- Hurd, P.L., 2006. Resource holding potential, subjective resource value, and game theoretical models of aggressiveness signalling. *Journal of Theoretical Biology* 241, 639–648.
<https://doi.org/10.1016/j.jtbi.2006.01.001>
- Kilgour, R., Scott, T.H., 1959. Leadership In A Herd Of Dairy Cows. *Proceedings of the New Zealand Society of Animal Production* 19, 36–43.
- Kirby, M., 2001. *Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns*. John Wiley & Sons, Inc, New York, USA.
- Lamb, R.C., 1976. Relationship Between Cow Behavior Patterns and Management Systems to Reduce Stress. *Journal of Dairy Science* 59, 1630–1636. [https://doi.org/10.3168/jds.S0022-0302\(76\)84416-5](https://doi.org/10.3168/jds.S0022-0302(76)84416-5)
- MacKay, D.J.C., 2003. *Information Theory, Inference, and Learning Algorithms*, 4th ed. Cambridge University Press, Cambridge UK.

- Manriquez, D., Chen, L., Albornoz, G., Velez, J., Pinedo, P., 2018. Case Study: Assessment of human-conditioned sorting behavior in dairy cows in farm research trials. *The Professional Animal Scientist* 34, 664–670. <https://doi.org/10.15232/pas.2018-01749>
- Manriquez, D., Chen, L., Melendez, P., Pinedo, P., 2019. The effect of an organic rumen-protected fat supplement on performance, metabolic status, and health of dairy cows. *BMC Vet Res* 15, 1–14. <https://doi.org/10.1186/s12917-019-2199-8>
- Mathis, M.W., Mathis, A., 2020. Deep learning tools for the measurement of animal behavior in neuroscience. *Current Opinion in Neurobiology* 60, 1–11. <https://doi.org/10.1016/j.conb.2019.10.008>
- McVey, C., Hsieh, F., Manriquez, D., Pinedo, P., Horback, K., 2021. Livestock Informatics Toolkit: A Case Study in Visually Characterizing Complex Behavioral Patterns across Multiple Sensor Platforms, Using Novel Unsupervised Machine Learning and Information Theoretic Approaches. *Sensors* 22, 1. <https://doi.org/10.3390/s22010001>
- McVey, C., Hsieh, F., Manriquez, D., Pinedo, P., Horback, K., 2020. Mind the Queue: A Case Study in Visualizing Heterogeneous Behavioral Patterns in Livestock Sensor Data Using Unsupervised Machine Learning Techniques. *Front. Vet. Sci.* 7, 523. <https://doi.org/10.3389/fvets.2020.00523>
- O'Connor, A.H., Bokkers, E.A.M., de Boer, I.J.M., Hogeveen, H., Sayers, R., Byrne, N., Ruelle, E., Shalloo, L., 2019. Associating cow characteristics with mobility scores in pasture-based dairy cows. *Journal of Dairy Science* 102, 8332–8342. <https://doi.org/10.3168/jds.2018-15719>
- Pinheiro, J.C., Bates, D.M., 2000. *Mixed-Effects Models in S and S-PLUS*, Statistics and Computing. Springer-Verlag, New York USA.

- Polikarpus, A., Kaart, T., Mootse, H., De Rosa, G., Arney, D., 2015. Influences of various factors on cows' entrance order into the milking parlour. *Applied Animal Behaviour Science* 166, 20–24. <https://doi.org/10.1016/j.applanim.2015.02.016>
- Rathore, A.K., 1982. Order of cow entry at milking and its relationships with milk yield and consistency of the order. *Applied Animal Ethology* 8, 45–52. [https://doi.org/10.1016/0304-3762\(82\)90131-6](https://doi.org/10.1016/0304-3762(82)90131-6)
- Reinhardt, V., 1982. Movement Orders and Leadership in a Semi-Wild Cattle Herd. *Behav* 83, 251–264. <https://doi.org/10.1163/156853983X00183>
- Rufener, C., Berezowski, J., Maximiano Sousa, F., Abreu, Y., Asher, L., Toscano, M.J., 2018. Finding hens in a haystack: Consistency of movement patterns within and across individual laying hens maintained in large groups. *Sci Rep* 8, 12303. <https://doi.org/10.1038/s41598-018-29962-x>
- Šárová, R., Špinka, M., Stěhulová, I., Ceacero, F., Šimečková, M., Kotrba, R., 2013. Pay respect to the elders: age, more than body mass, determines dominance in female beef cattle. *Animal Behaviour* 86, 1315–1323. <https://doi.org/10.1016/j.anbehav.2013.10.002>
- Schein, M.W., Fohrman, M.H., 1955. Social dominance relationships in a herd of dairy cattle. *The British Journal of Animal Behaviour* 3, 45–55. [https://doi.org/10.1016/S0950-5601\(55\)80012-3](https://doi.org/10.1016/S0950-5601(55)80012-3)
- Shannon, C.E., 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal* 27, 379–423.

Sibanda, T.Z., Welch, M., Schneider, D., Kolakshyapati, M., Ruhnke, I., 2020. Characterising Free-Range Layer Flocks Using Unsupervised Cluster Analysis. *Animals* 10, 855.

<https://doi.org/10.3390/ani10050855>

Soffié, M., Thinès, G., De Marneffe, G., 1976. Relation between milking order and dominance value in a group of dairy cows. *Applied Animal Ethology* 2, 271–276.

[https://doi.org/10.1016/0304-3762\(76\)90060-2](https://doi.org/10.1016/0304-3762(76)90060-2)

Stygar, A.H., Gómez, Y., Berteselli, G.V., Dalla Costa, E., Canali, E., Niemi, J.K., Llonch, P., Pastell, M., 2021. A Systematic Review on Commercially Available and Validated Sensor Technologies for Welfare Assessment of Dairy Cattle. *Front. Vet. Sci.* 8, 634338.

<https://doi.org/10.3389/fvets.2021.634338>

Tucker, C.B., Jensen, M.B., de Passillé, A.M., Hänninen, L., Rushen, J., 2021. Invited review: Lying time and the welfare of dairy cows. *Journal of Dairy Science* 104, 20–46.

<https://doi.org/10.3168/jds.2019-18074>

Valletta, J.J., Torney, C., Kings, M., Thornton, A., Madden, J., 2017. Applications of machine learning in animal behaviour studies. *Animal Behaviour* 124, 203–220.

<https://doi.org/10.1016/j.anbehav.2016.12.005>

Wathes, C.M., Kristensen, H.H., Aerts, J.-M., Berckmans, D., 2008. Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall? *Computers and Electronics in Agriculture* 64, 2–10.

<https://doi.org/10.1016/j.compag.2008.05.005>

Wolpert, D.H., Macready, W.G., 1997. No free lunch theorems for optimization. *IEEE Trans. Evol. Computat.* 1, 67–82.

<https://doi.org/10.1109/4235.585893>

OBJECTIVES

There is no “free lunch” in data science, and so no single analytical framework can be expected to best serve the broad spectrum of datasets encountered in animal science (Wolpert and MacReady 1997). For many PLF data streams, the amount of *a priori* knowledge required to conduct model-based analyses is often fundamentally misaligned with the chaotic and poorly characterized farm environments from which these data sets are compiled. Where PLF technologies are employed in more controlled experimental settings, using a model to infuse into statistical analyses information about key experimental assumptions and design decisions can improve the power of subsequent inferences. For data collected in less structured environments, however, this rigid analytical framework may serve only to create opportunities for researcher biases to creep into such analyses and for important but unanticipated behavioral patterns to go overlooked. For such data streams, a fully model-free data analytical pipeline might therefore provide a more open-ended approach to knowledge discovery capable of creating a more complete picture of the behavioral dynamics at play in a system.

The objectives of this dissertation were to develop the statistical and algorithmic advances necessary to bridge the gap between model free unsupervised machine learning and information theoretic tools in order to create a complete analytical pipeline that could be readily bootstrapped to accommodate the range of data structures found in PLF applications.

Chapter 1 provides preliminary explorations of the basic components of this pipeline. In this chapter, the efficacy of two opposing UML approaches to information compression are compared in analyses of milking order data extracted from parlor metadata records. Using the

patterns recovered from these analyses, the efficacy of information theoretic and standard frequentist statistical inferences are compared in recovering significant bivariate associations between milking order, cow records, and ear tag accelerometer data.

In Chapter 2, a number of novel algorithmic techniques are developed to build upon and further streamline this basic analytical pipeline. This chapter introduces a flexible ensemble simulation approach to more formally distinguish between signal and noise in PLF data streams with more complex sources of measurement error. Information theoretic approaches are also proposed to more fully leverage the geometric information contained in tree structures produced by upstream clustering analyses, and also to more easily characterize any nonlinear patterns recovered through contingency table-based visualizations. The efficacy of these algorithms are subsequently demonstrated by recovering complex patterns in overall time budgets, and relating them back to patterns in milking order.

Chapters 3 and 4 expand the algorithmic framework developed in Chapter 2, in an effort to demonstrate how easily these analytical concepts can be expanded and bootstrapped to a wider range of data structures. Chapter 3 delves more deeply into the temporal complexity hiding within gross patterns in overall time budget data, exploring temporal patterns in daily and hourly time budgets. Finally, Chapter 4 explores how this algorithmic framework can be extended not only to recovering patterns, but patterns in patterns, by developing a fundamentally discrete approach to characterizing individual differences in behavioral and biological plasticity that can be applied to noncontinuous or even multivariate data streams.

CHAPTER 1

“Mind the Queue: A case study in visualizing heterogeneous behavioral patterns in livestock sensor data using unsupervised machine learning techniques”

Published in Frontiers in Veterinary Science (2020, 7:523)

Introduction

For much of its history, ethological research in livestock has relied on human observers to encode behaviors of interest (Dawkins 2007). While developing a detailed ethogram and observer training protocols constitute no simple task, there are several inherent advantages to this approach for subsequent statistical analyses. Continuous involvement of a human in the incoming data stream allows many erroneous data points to be identified and excluded from downstream analyses that they might otherwise destabilize. Extensive involvement of research personnel in the data collection phase also nurtures a deeper familiarity with the system under study. This not only aids in the specification of an appropriate statistical model and interpretation of results, but is often critical in identifying unexpected behavioral patterns that can inspire novel hypotheses.

Unfortunately, the inherent quality of such data imposes practical limitations on the quantity that can be produced. This can restrict both the number of animals utilized in a study and the period of time over which they are observed. The later limitation can overlook important dynamic features of the behavioral patterns under consideration. Restrictions on the number of animals that can be studied, on the other hand, can fundamentally alter the behavioral mechanisms at play in a herd. For example, the linearity of dominance hierarchies are known to change with group size (Pagel & Dawkins 1996).

As commercial herds and flocks become ever larger, this only serves to broaden the gap between experimental findings and the welfare challenges they are meant to inform. Subsampling of animals or observations windows may be employed to reduce the number of observations collected without restricting the size of the study system. If the pre-existing base of scientific literature does not provide clear guidance on the selection of target animals or focal periods, however, such strategies may risk overlooking finer-grain behavioral patterns and skewing inferences about the collective behavior of the group (Biro & Dingemanse 2008; Rosser *et al.* 2013).

In recent years, livestock sensor technologies have become a popular alternative to visual observation (Wathes *et al.* 2008; Banhazi *et al.* 2012; Ellen *et al.* 2019; Halachmi & Guarino 2016). While the behaviors recorded are neither as complex or as detailed as those quantified via an observational ethogram, such devices have the capacity to continuously monitor hundreds or even thousands of animals for extended periods of time. Such a substantial expansion in the bandwidth capacity of ethological studies creates many new opportunities to better understand the behavior of livestock, particularly in large-scale commercial settings, but also raises new methodological challenges. Replacing nuanced human intuition with basic computer logic may increase the risk of erroneous data points, an issue that is only further compounded by the scale of data produced by such technologies, which renders many conventional visualizations techniques ineffective in identifying outliers. Observations recorded over extended time periods with high sampling frequency from large heterogeneous social groups may also contain a range of complex stochastic features – autocorrelation, temporal nonstationary, heterogeneous variance structures, non-independence between experimental units, etc. – that can lead to spurious inferences when not appropriately specified in a conventional linear model. The hands-off and somewhat black-boxed nature of many sensor platforms, however, do not nurture the intuition needed to identify many of these model structures *a priori*. Such insights must

instead be drawn directly from the data itself, but here again, standard visualization tools may not scale to such large datasets.

Unsupervised machine learning (UML) tools offer a distinct empirical approach to knowledge discovery that are purpose built for large and complex datasets (Valletta *et al.* 2017). Whereas conventional linear models excel at providing answers to targeted experimental hypotheses, UML algorithms strive to identify and characterize the nonrandom patterns hiding beneath the stochastic surface of a dataset using model-free iterative techniques that impose few structural assumptions. This open-ended and highly flexible approach to data exploration may offer an empirical means by which to recover much of the familiarity with a study system that is lost with the shift from direct observation to sensor platforms. The purpose of this research was contrast the behavioral insights gleaned from UML algorithms with those recovered using conventional exploratory data analysis (EDA) techniques, and to then explore how such information could be best integrated into standard linear analysis pipelines.

Milking order, or the sequence in which cows enter the parlor to be milked, is recorded in all RFID (Radio Frequency Identification) equipped milking systems, making such records one of the most universal automated data streams to be found on modern dairies. Despite their ubiquity, such records are seldom used to inform individual or herd-level management strategies. This lack of utility, however, has not been for lack of study. Milking order has been the subject of scientific study since 1950's (Kilgour & Scott 1959), with early investigators speculating that such records might contain pertinent information about individual cow productivity (Rathore 1982; Soffie *et al.* 1976), health (Gadbury 1975), and social status (Soffie *et al.* 1976; Reinhardt 1973; Dickson *et al.* 1967). The modest base of scientific literature that has since been compiled on this topic, however, has struggled to recover repeatable evidence of such associations (Polikarpus 2015; Grasso 2007; Beggs *et al.* 2018; Berry & McCarthy 2012). While such inconsistency may simply reflect non-uniformity in the behavioral

strategies driving queueing patterns across different herds and farm environments, misspecification of the linear models used to describe this system could also contribute to volatility in these statistical inferences. The objective of this methodological case study will be visualize the various stochastic aspects of such records using UML tools in an effort to identify erroneous data points and heterogeneous variance structures that may not be recovered using conventional EDA techniques.

Methods

Study Animal Management

Data for this case study was repurposed from a feed trial assessing the effect of an organic fat supplement on cow health and productivity through the first 150 days of lactation (Manriquez *et al.* 2019). All animal handling and experimental protocols were approved by the Colorado State University Institution of Animal Care and Use Committee (Protocol ID: 16-6704AA). The study ran from January to July of 2017 on a certified organic dairy in Northern Colorado. A total of 200 mixed-parity Holstein cows were enrolled over a 1.5 month period as study-eligible animals calved. Cows were maintained in a closed herd for the duration of the study, with sick animals temporarily removed to a hospital pen when necessary. The study pen was an open-sided free stall barn, stocked at just above half capacity with respect to bunk space and beds, with free access to an adjacent outdoor dry lot. At roughly the midpoint of the trial, cows were moved overnight to a grass pasture that conformed with organic grazing requirements (for more details on pen setup see Manriquez *et al.* 2018). Cows had access to total mixed ration (TMR) ration following each milking. Animals were temporarily split into two subsections of the pen following the morning milking to facilitate administration of control and treatment diets. Cows remained locked for roughly 45 minutes following this division so that farm and research staff could collect health and fertility data. Additionally, all animals were fitted with

CowManager® ear tag accelerometers (Agis Automatisering BV, Harmelen, Netherlands). This commercial sensor platform, while designed and optimized for disease and heat detection, also provided hourly time budget estimates for total time (mins) engaged in a range of behaviors - eating, rumination, non-activity, activity, and high activity - as well as average skin temperature.

Data Wrangling

Raw milk logs were exported from the rotary parlor following each morning milking (ALPRO, DeLaval, Tumba, Sweden), and were processed using data wrangling tools available in R version 3.5.1 (R Core Team 2018). To account for missing records due to illnesses and RFID reader errors, ordinal entry positions were normalized by the total number of cows recorded in a given milking (Beggs *et al.* 2018). Transforming the data to an entry quantile served to make the domain restriction uniform across days. Additionally, “dividing out” daily variations in herd size served to prevent this uncontrolled experimental factor from artificially increasing individual variability in entry position. For example, if a cow were always the last animal to enter the parlor, her ordinal entry position might vary widely with herd size, but her entry quantile would always be 1.

The first 55 days of records were excluded from analyses to allow all animals to enter the herd over the rolling enrollment period and become established in their parlor entry position (Polikarpus *et al.* 2015). To avoid irregularities in cow movements, several observation days surrounding management changes were also dropped, including: the two days preceding transition to pasture, the four days following pasture access, and the final seven days on trial. Any days where less than 75% of the herd was successfully recorded in the parlor were also dropped. This left a total of 80 days of milk order observations – 26 recorded while cows remained overnight in their pen, and 54 after the transition to overnight pasture. Finally, cows that were not present in at least 50% of the remaining milkings were

excluded from further analysis. Of the 177 cows with sufficient records, 114 had no recorded health events.

Quantifying Degree of Randomness

The first step in understanding this system was to determine if there was any evidence of non-random patterns in queue formation. Had this data been collected observationally, researchers might have simply noted if the same cows consistently entered the parlor in a given section of the queue. Standard summary statistics do not readily lend themselves to making an equivalent empirical determination, however, as cow identity is a discrete variable with no inherent nominal value. This issue is encountered fairly regularly in ethological studies, where many qualitative behaviors have no natural ordering, such as: locations occupied in a pen at discrete time points, conspecifics an individual interacts with, feedstuffs or enrichment items engaged with, etc. Here we use entropy to quantify the stochasticity of cow-membership within each fixed quantile range (H_q). To compute these values, the queue was divided into 20 equally-sized segments ($q_{0-0.05}, \dots, q_{0.95-1}$). For each queue segment, counts were generated to determine how frequently each individual cow had been observed in that range of entry quantiles. Shannon's entropy was then calculated conditional of queue segment (q) using the formula in Equation 1 (Shannon 1948).

$$H_q = - \sum_{c = \text{cow}} \frac{n_{c|q}}{N_q} * \log_2 \left(\frac{n_{c|q}}{N_q} \right) \quad (1)$$

$n_{c|q}$ = total times a given cow (c) is observed in quantile position (q)

N_q = total animals observed in queue segment (q) across observed milkings

With this metric, the more consistently a smaller set of cows are observed in a given segment of the queue, the smaller the entropy values becomes to reflect less stochasticity in the system. In standard statistical models, the nominal value of estimators such as loglikelihood and AIC scale with the size of the data set, and must be interpreted relative the value of equivalent terms assessed against a null model. Analogously, the nominal value of the entropy estimates scales with the number of discrete categories used. The maximum theoretical value occurs when no underlying deterministic structures are present and all categories are equally likely to occur, which algebraically simplifies to the log of the number of discrete categories used (Shannon 1948). Here the maximum theoretical entropy value would be $\log_2(114) = 6.83$. To visually contrast differences in stochasticity across the queue, the observed entropy values were plotted against the median entry quantile of the corresponding queue segment using the *ggplot2* package, with maximum theoretical entropy added as a horizontal reference line (Wickham 2016).

Nonrandom patterns in queue formation could also be explored by tracking the entry position of individual cows over time. As entry quantile has a numerical value, we can now also use variance to quantify and contrast stochasticity between animals. As with all analytical approaches reviewed in this paper, there are both strengths and shortcomings to either approach (Table 1). In this system there are two potential drawbacks to this conventional summary statistic. The first is that variance estimates are quite sensitive to outliers, making it difficult to empirically distinguish between cows that occupy a wider range of queue positions and animals who typically occupy a narrower range but might have gotten jostled far from their normal position on one or several occasions. The second drawback is that, because variance quantifies dispersion about a central value, it cannot distinguish between cows that demonstrate little consistency in entry position and multimodal queuing patterns. For example, if a cow always entered the parlor either first or last, we would intuitively determine that this pattern is non-random, but the corresponding variance estimate would be the largest in the herd. These issues are

circumvented, however, by discretizing entry quantile values and again using entropy to quantify stochasticity. To evaluate an individual cow's variability in quantile range-memberships (H_c), count data was used to recalculate Shannon's entropy conditional on cow ID (c) using Equation 2 (Shannon 1948).

$$H_c = - \sum_{q = \text{Queue Segment}} \frac{n_{q|c}}{N_c} * \log \left(\frac{n_{q|c}}{N_c} \right) \quad (2)$$

$n_{q|c}$ = total times a given cow (c) is observed in quantile position (q)
 N_c = total number of days a given cow (c) was observed in the queue

Here the maximum possible entropy value, signifying a cow is equally likely to occupy any queue segment, would be $\log_2(20) = 4.32$. Observed entropy and variance values were visually compared using the *ggplot2* package (Wickham 2016). To test if an individual cow demonstrated less stochasticity in entry positions than would be expected with a purely random queueing process, entry quantile values were again permuted within each observation day, and both variance and entropy recalculated. This process was repeated over 5000 iterations to generate empirical cumulative density functions (CDFs) for both stochasticity estimators under the null, which were then used to estimate p-values for the corresponding observed values.

Analytical Goal	Approach	Strengths	Shortcomings
Quantifying randomness	Variance	• Uses continuous measures	• Cannot use categorical data • Assumes unimodality • Sensitive to outliers
	Entropy	• Uses categorical data • Permits Multimodality • Robust to Outliers	• Continuous measures must be discretized, which can result in loss of information
Visualizing inter-animal relationships	PCA	• No metaparameters • Assess embedding via loadings	• Assumes latent structures are linear (additive)
	Diffusion map	• No linearity assumption	• Employs metaparameters • Embedding qualitatively assessed using visualizations
Visualizing temporal non-stationarity	Residual plots of stational repeated measures model	• Considers all data points simultaneously	• Non-homogeneous temporal trends may be overlooked
	Time $\times$ Response scatter plots for each cow	• Easy to create and visually assess for non-stationarity	• Difficult to contextualize trends across entire herd
	Data mechanics plot	• Simultaneously visualize social and temporal structure • Non-homogeneous temporal trends visually enhanced	• Row and column cluster granularity must be determined visually
Association between queue position and cow attributes	Linear mixed effect model	• Targeted hypotheses • Simultaneous estimation of multiple covariates	• Non-independence between animals inflates rate of Type I Errors
	Mutual conditional entropy permutation test	• Robust to between-animal non-independence • Detects non-linear patterns	• Cannot adjust for influence of other variables (confounders)
Association between queue position and accelerometer logs of home pen behavior	Linear mixed effect model	• Targeted hypotheses	• Convergence issues with two large sets of repeated measures
	Mutual conditional entropy permutation test	• Generalized pattern detection • Easily extended to large data	• Cannot adjust for confounding variables

Table 1: Summary of analytical approaches compared in this manuscript, and a comparison of their relative strengths and shortcomings.

Visualization of Inter-Animal Relationships

Having recovered evidence of nonrandom patterns, the next step was to begin characterizing the behavioral mechanisms driving this heterogeneity. The most fundamental question that need be answered to inform further analysis was the degree to which queueing patterns were driven by individual or collective behaviors. Because cows jockey for position with one another in the crowd pen, where they are pushed up to enter the parlor, we know intuitively that entry quantile records cannot be considered truly independent observations. If cows move through this melee as independent agents, such that their position within the queue is determined by individual attributes – preferences, dominance, etc – then a linear model may still provide a reasonable approximation of the underlying

system. Early observational work on milking order, however, has suggested that cows may form consistent associations when entering the milking parlor, particularly when heifers are reared together (Gadbury 1975; Schein & Fohrman 1955). If cows move into the parlor in cohesive units, such that queue position is more determined by clique-level than individual attributes, then network analyses may be a more appropriate.

Principal Component Analysis (PCA) is commonly employed to visualize relationships between observational units in high dimensional datasets. In this approach, redundancy between variables, here each milking record, is captured using either covariance or correlation assessed across all data points, here all animals. An eigenvector decomposition is then used to linearly compress the information contained in the data via rotation of the orthogonal axes. New axes (loadings) are added iteratively such that each new dimension is pointed in the direction of greatest remaining variability until only noise remains (James *et al.* 2013). Each data point is then projected into the resulting low-dimension linear space (Barter & Gross 2018). PCA was here performed only on animals with no recorded health in order to prevent any anomalous queuing behaviors recorded from acutely or chronically ill animals from obscuring the queuing patterns of the broader herd. The correlation matrix was constructed using all pairwise complete observations, and a scree plot was used to determine the dimensionality of the resulting space (see Supplemental Materials). The *plotly* package (Sievert 2018) was then used to visualize the final embedding.

While PCA provides a computationally expedient means of visualizing high dimensional data, the underlying assumption of linearity is not always appropriate (James *et al.* 2013; Barter & Gross 2018). In some data sets complex geometric constraints, such as those commonly found with images or raw accelerometer data, and other latent deterministic features may project data points onto high dimensional geometric surfaces collectively called manifolds (Saul & Roweis 2003). When these topologies are nonlinear (cones, spheres, donuts, etc.), the spatial relationships between data points

cannot always be reliably maintained when projected directly into a linear (Euclidean) space, which can lead to incorrect inferences (Kirby 2001; Barter & Gross 2018). Imagine, for example, you had a round globe of the world and wanted instead a flat map. Applying PCA to this task would be analogous to smooshing the globe flat on a table. Some of the original geographic relationships would be discernable, but some locations would appear erroneously close, and some landscapes would be entirely obscured. Modern manifold learning algorithms strive to more reliably project the complex geometric relationships between observational units into a standard Euclidean space by approximating the surface of a nonlinear manifold with a series of interconnected flat surfaces that can then be ‘unwrapped’ onto a linear space (Izenman 2013). Returning to the previous pedagogical metaphor, this would be analogous to taking pictures of the globe centered around a number of key geographic locations, and then attempting to arrange the overlapping images onto a flat table. Some geographic features will still be lost, particularly over sparsely sampled regions like the oceans, but the spatial relationships between landmarks would collectively prove more representative of the original topography.

To further explore the underlying structure of this data absent assumptions of linearity, and thereby potentially accommodate any complex geometric constraints imposed on milk order records by latent social structures within the herd, a diffusion map algorithm was implemented using functions provided in base R (R Core Team 2018). This was done here by first calculating the Euclidean distance between temporally aligned vectors of parlor entry quantiles for each pairwise combination of cows, scaled to adjust for missing records, and then inverting these values to create a similarity matrix. From this similarity matrix a weighted network was created by progressively adding links for the $k = 10$ nearest neighbors surrounding each data point. A spectral value decomposition was then performed on the corresponding graph Laplacian matrix (Coifman *et al.* 2005; Barter & Gross 2018). The resulting eigenvalues were used to select the appropriate number of dimensions, and the corresponding

eigenvectors visualized using the 3D scatter tools from the *plotly* package (Sievert 2018). Finally, as a means of comparing geometric structures identified in the observed dataset with those of a completely randomized queuing process, the permuted dataset generated in the previous section was also embedded and visualized using *plotly* graphics (Sievert 2018).

Characterization of Temporal Dynamics

Having determined from the previous visualizations that a linear model might be a reasonable representation of the underlying deterministic structures of this system, the next step was to explore the temporal dynamics of this dataset. In a standard repeated measures model, multiple observations from the same animal are assumed to be identically and independently sampled, implying that sampling order should not affect the observed value. If the observation period is sufficiently long to allow the underlying process to shift or evolve over time, however, stationarity cannot be assumed. Failure to statistically accommodate a temporal trend can not only lead to spurious inferences due to incorrect estimation of error variance, but also risks overlooking dynamic features of the behaviors under consideration (Pinheiro & Bates 2000). In practice temporal trends are often assessed by first fitting a stationary model and analyzing the resulting residuals. This may suffice when the temporal trend is uniform across animals, but risks overlooking more complex nonhomogeneous temporal affects. This could occur if only a subset of the larger group displays a nonstationary pattern, a risk that is likely heightened in large socially heterogeneous groups. In this physically constrained system, where we know that every cow moving forwards in the queue must force other cows backwards, compensatory trends could also be easily overlooked in collective assessment of residuals.

We first assessed temporal trend using two conventional EDA techniques. First, the *ggplot2* package (Wickham 2016) was used to generate scatter plots of entry quantile values against the corresponding observation date for each individual cow, with pasture access annotated with a vertical

line. Plots were visually inspected for nonstationary, and are provided in supplementary materials. Next, to further explore the impact of the shift from pen to overnight pasture access on morning queueing patterns, median queue positions from the two subperiods were plotted against using the *ggplot2* package (Wickham 2016), and Pearson correlation (R) and Kendall Tau (τ) were computed using the *stats* package (R Core Team 2018). While these preliminary visualizations were easy to both generate and interpret, both treat cows as independent and somewhat isolated units. With such a large number of animals to consider, the capacity for human pattern detection is quickly overwhelmed, making it difficult to contextualize trends within the broader herd. Further, this approach fails to leverage non-independence between animals entering the parlor, and thus risks overlooking subtler collective responses.

Data mechanics visualizations were implemented to simultaneously explore systematic heterogeneity in milk entry quantiles both between animals and across the temporal axis. This was done by first using entry quantile values to compute two Euclidean distance matrices: one quantifying the similarity between pairwise combinations of cows, the second quantifying similarity between pairwise combinations of daily milking sequences. These distance matrices were then used to generate two independent hierarchical clustering trees using the Ward D2 method (R Core Team 2018; James *et al.* 2013). By cutting both trees at a fixed number of clusters, observation days and cows were both partitioned into empirically defined categories, and a contingency table was then formed with cow clusters as the row variable and day clusters as the column variable. The original distance matrices were then updated, using the clustering structure between cows to create a weighted distance matrix between days and vice versa, thereby allowing mutual information to be shared between the temporal and social axes of the dataset (see Supplemental Materials for details). After several iterations of this algorithm, clusterings converged towards a contingency table with minimal entropy, wherein the entry quantile values within each cell were as homogenous as possible. When the entry quantile values were

subsequently visualized using a heat map, this highly generalizable entropy minimization technique served to visually enhance heterogeneity within the data driven by nonrandom patterns along either axis. Further, by facilitating the transfer of information between axes, interaction effects between the social and temporal dimensions of this system were magnified, which here provided a means to explore nonhomogeneous temporal nonstationary between subgroups within the herd (Hsieh *et al.* 2018; Guan & Hsieh 2018; McCowan *et al.* 2016).

The data mechanics pipeline was used to analyze the temporal dynamics present in both the complete milking order dataset and the subset of animals with no recorded health events. Heat map visualizations were generated using the *pheatmap* package in R (Kolde 2019), with observation days arranged on the column axis and Cow ID's arranged on the row axis. Fixed values for the number of clusters used to divide the row and column axes could not be determined *a priori*. Instead this algorithm was applied on a grid from 1 to 10 clusters for either axis. The resulting 100 heat maps scanned visually to determine the clustering granularity required to bring into resolution any interactions between social and temporal mechanisms. While this process may be computationally cumbersome, it is empirically analogous to systematically varying the focus of a light microscope to bring into resolution microbes of unknown size – a tedious but effective means of identifying all relevant structures within a sample (Hsieh *et al.* 2018). Finally, the *RColorBrewer* package (Neuwirth 2014) was used to add color annotations to the column margin, to clarify temporal patterns, and to the row margins, which served to visualize potential relationships between queue position, a selection of individual cow attribute variables, and the onset of recorded health complications.

Linear Analysis of Cow Attributes

Having thoroughly characterized the stochastic structures present in this dataset, the insights gleaned from the preceding visualizations were incorporated into a linear model to evaluate the

relationship between queue position and several cow attributes. The four days identified as outliers by the data mechanics visualizations were first removed and the dataset converted to long format to be analyzed as a repeated measures model using the *nlme* package (Pinheiro *et al.* 2018). Cow was fit as a random intercept via maximum likelihood method. Guided by the results of entropy and data mechanics visualizations, VarIdent was used to estimate separate error variance terms for each cow, and the necessity of this data-hungry heterogeneous variance model confirmed via likelihood ratio test against the null model with homogenous variance (Pinheiro & Bates 2000). After centering and scaling cow attribute variables, linear fixed effects were added for cow age (days old at start of trial), calving date (approximately the date of entry into the herd), and peak milk yield (estimated via the 95th quantile of each cow's 150-day parlor lactation record). Interaction effects were created for each combination of these linear terms, and a categorical effect added for the control and treatment groups of the fat supplementation trial. Models were generated for both the complete dataset and the subset of animals with no recorded health events, which consisted of 160 and 104 cows respectively after removing animals with incomplete attribute records. The predictive value of each fixed effect term was evaluated via a Wald's test. Where a significant association was identified at the standard $\alpha = 0.05$ (Type I Error) confidence level, this pattern was visualized by plotting the cow attribute variable against the predicted queue position for each cow (fixed effect + BLUP).

While UML insights served to improve the specification of model variance structures within-animal, the validity of statistical insights made at the between-animal level is still contingent upon the correct estimation of model degrees of freedom. A fundamental assumption of frequentist tests is that observations must be independently sampled. When observations are not independent, the effective degrees of freedom present in the model may be lower than the nominal value. This causes the model to be overconfident in its estimation of error terms, increasing the risk of a false positive result. Non-independence due to repeated sampling (pseudoreplication) has here been accounted for by fitting a

random effect for each cow, but non-independence between animals has not been accommodated. The results of the diffusion map and data mechanics visualizations did not recover overwhelming evidence of coordinated movements between animals through the queue, which would have signified non-independence due to social cohesion (positive interclass correlation between animals); however, we both visualized via data mechanics and know intuitively that in this physically constrained system any cow moving forward in the queue must be countered with other cows being forced backwards and vice versa. If this effect extends beyond isolated fluctuations in daily formation of the queue, then the presence of some animals in the herd might systematically dampen or even completely prevent other animals from demonstrating behavioral patterns that they would otherwise display independently or in another herd with a different social composition (negative interclass correlation between animals). This would not only serve to confound the behavioral mechanisms at play, but such cows whose behaviors are suppressed by their herd mates cannot be said to be contributing fully to the model, potentially reducing the effective sample size. This could allow sampling fluctuations to produce misleading statistical inferences, even in this large sample of animals (Martin & Bateson 1993; Forstmeir *et al.* 2017; Bruneel *et al.* 2012).

UML algorithms cannot recover information about behaviors that were never expressed, and so are also not immune to the biasing effects of non-independence between animals. These tools can, however, provide model-free tests of association that may serve as a sanity check for statistical inferences when degrees of freedom may be uncertain. We explore this option here by again combining modern clustering tools with a flexible information theoretic approach to pattern detection (Hsieh *et al.* 2018). First, independent clustering trees were used to subdivide the herd based on queuing records and each of the cow attributed variable. The resulting categorical variables were then used to form contingency tables between queue subgroups and each of the candidate predictor variables. If no relationship existed between these two axes, then a cow belonging to a given row category based on

queue records would be just as likely to belong to any of the column categories based on cow attribute and vice versa. If instead an underlying biological mechanism was present linking these axes, then cows within a range of cow attribute values would be spread unevenly among queue subgroups. Such heterogeneity in cell counts was quantified by calculating a weighted mutual conditional entropy (MCE) value across first the rows and then the columns of the contingency table and averaging the results, which reflected the amount of mutual information shared between the two variables. To determine if the observed MCE value was significantly smaller than would be expected from random fluctuations in the sample, row and column classifiers were randomly permuted across cows to remove any underlying bivariate relationship and MCE recalculated. This randomization procedure was repeated over 2000 iterations, and the observed entropy value compared to the resulting empirical CDF to produce a p-value for the significance of the bivariate association. Mutual conditional entropy tests were performed for all significant or marginally significant linear effects for both regression models. While the number of clusters used to discretize the cow attribute and queue records may be specified *a priori* provided strong biological reasoning or empirical evidence, mutual conditional entropy tests were here performed on a grid from two clusters up to the highest visible granularity of the corresponding clustering tree, and the optimal metaparameter values selected by minimizing the average marginal rank.

Exploring Associations between Sensor and Queue Records

Previous studies seeking to identify factors that predict an animal's parlor entry position have focused primarily on biological drivers of queueing behavior related to productivity, health, and traditional measures of fitness such as age and size (Danielson 2012). As this herd was also fitted with ear tag accelerometers, it is here also possible to explore relationships between queue position and behavioral patterns displayed between milkings. Due to the size of these datasets, however, this small step beyond the bounds of the existing literature constitutes a considerable leap in statistical complexity

within a linear modeling framework. A multivariate mixed model that considers all observations from either dataset would exceed the capacity of many solvers (Snijders & Bosker 2004). A simpler approach to exploring this relationship might therefore be to compress the information available in parlor entry records into a grouping variable and then attempt to identify differences in the various home pen behaviors across the resulting subsections of the herd.

We implement this strategy here by using the *nlme* package to fit linear mixed models, with cow fit as a random intercept, against each of the five behaviors recorded by the CowManager platform (non-activity, activity, high activity, rumination, eating) and also average body temperature (Pineiro *et al.* 2018). To avoid the risk of anomalous behaviors that might skew model inferences, only cows with no recorded health events were used. Hour of the day was fit as a categorical variable to capture cyclical patterns. Days on trial was also fit as a categorical fixed effect to allow for non-smooth longitudinal changes in behaviors due weather and also the shift to pasture. Finally, queue groups were determined by arbitrarily dividing the herd into quartiles based on median entry position. The resulting categorical variable was then fit as both a main effect and an interaction effect against both cyclic and longitudinal time effects. Due to the size of the model, temporal correlation and heterogeneous variance models both exceeded the capacity of this package to converge. Comparisons of the cyclic and longitudinal trends in behavioral patterns between queue groups were made using the plotting utility available in the *emmeans* package (Lenth 2019), with the complete results provided in the supplemental materials.

While linear models provide an expedient means to statistically evaluate targeted experimental hypotheses, the more open-ended approach to knowledge discovery provided by UML algorithms may offer an advantage in exploratory data analysis problems such as this. We explore the utility of this alternative strategy here by again employing a mutual conditional entropy (MCE) test to identify significant associations between these two behavioral axes (Hsieh *et al.* 2018). The flexibility of

hierarchical clustering tools allows this technique to be directly extended from the previous section, which compared repeated measures of queue position against a univariate covariate, to accommodate both high dimensional datasets. For each parameter recorded by the CowManager platform, this model free test of association was performed on the complete sensor record, on subsets of the records corresponding to each of the three lounging periods (morning, afternoon, and night), and finally on a subset of the records where observations from all three lounging periods had been aggregated. As in the previous section, the number of clusters used to discretize queue and sensor data were evaluated on a grid, here from tree depths 2-10. To characterize the divergent behavioral patterns across queue groups identified by significant tests of association, tube plots were created by plotting each within-day subgroup median on a circular grid and then stacking rings to form a tube using the 3D plotting tools in the *plotly* package (Sievert 2018).

Results & Discussion

Quantifying Degree of Randomness

Looking first at the entropy calculations for each segment of the queue visualized in Figure 1, it is clear that all parlor entry positions are not stochastically equivalent. The same animals are seen consistently at the very front and back of the queue, such that the resulting entropy values are far lower than would be seen with a purely random queueing process. Moving towards the middle of the queue, however, there is progressively less consistency in the animals present across milkings, such that the observed entropy values approach a random process. Looking next at the stochasticity demonstrated by each individual cow in Figure 2, we see there is again a clear gradient. Cows with median entry quantiles at the front and rear of the herd again show far greater consistency in their entry positions. As their median quantile position moves towards the center of the herd they become more variable in

their entry positions over the observation window. This gradient is seen using both entropy and variance as estimators of stochasticity, but is more visually distinct using entropy estimates. While discretizing an intrinsically continuous parameter results in a loss of information, we see here that this sacrifice has excluded extraneous noise in the system to bring the underlying stochastic pattern into clearer resolution. This data thus highlights the potential upside of amending entropy estimates to the traditional cadre of summary statistics, particularly when working with outcome variables that are prone to extreme or anomalous values.

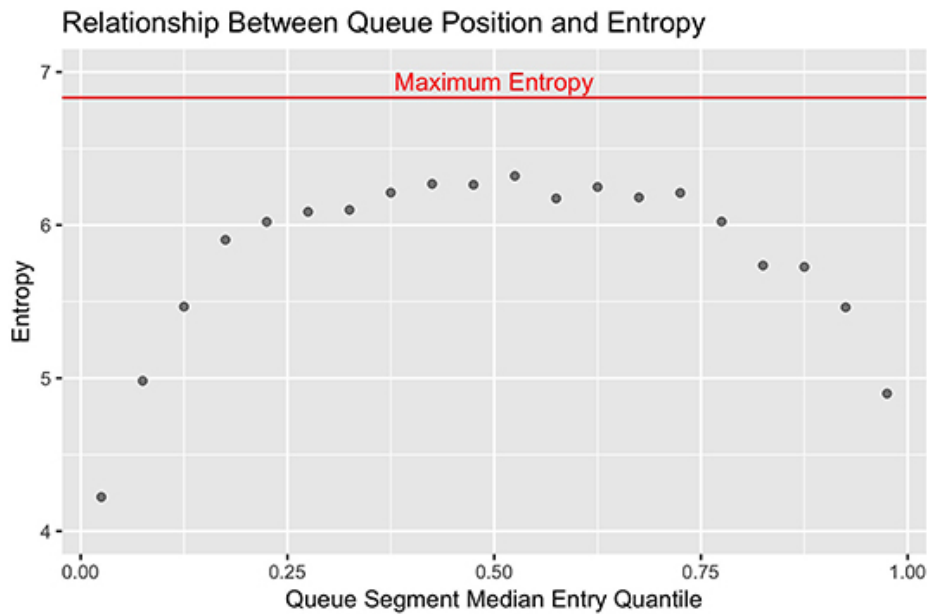


Figure 1: Entropy estimates from observed (red) and randomly permuted (blue) datasets are plotted against the median quantile value for the corresponding segment of the queue. The front and end of the queue are the least stochastic, but all sections of the queue demonstrate lower entropy than with purely random queue formation.

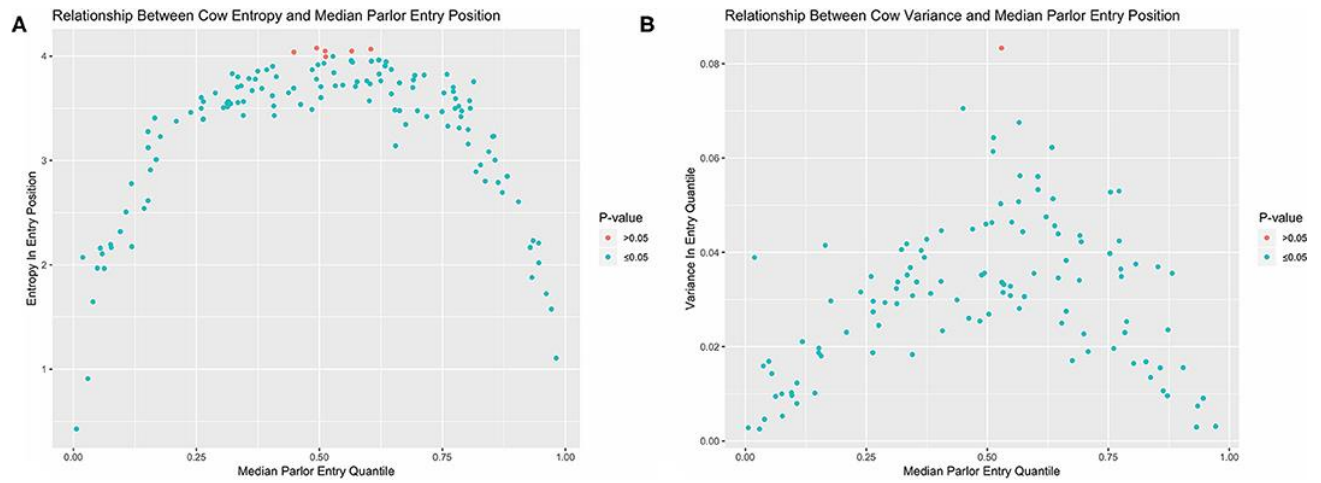


Figure 2: Observed entropy (A) and variance (B) values for each cow plotted against their corresponding median entry position. Cows with the greatest consistency in parlor entry position are found at the front and end of the queue. With both estimators, nearly all cows in the herd demonstrate less stochasticity in parlor entry position than completely random queue formation as determined by permutation test.

In examining the results of the permutation tests, nearly all animals demonstrated significantly less stochasticity in their entry positions at the standard $\alpha = 0.05$ significance level as compared with a completely randomized queueing process. Only 3 cows out of 114 overlapped with the empirical distribution of entropy estimates under a randomized queueing pattern, and only 1 cow overlapped when variance was used as the estimator of stochasticity. This suggests that nearly all animals in the herd might contribute some information about the underlying nonrandom patterns in queue formation to subsequent analyses; however, the amount of information they contribute may not be equal, as there is considerable heterogeneity between cows. Of greater concern, this heterogeneity is systematic, as there are no cows showing high consistency in entry quantile in the center of the queue. If this pattern is not driven by variability in the underlying predictors of queue position, but instead reflects either an underlying behavioral mechanism or something even more fundamental to this system such as the inherent domain constraint (Beggs *et al.* 2018), this could lead to inaccurate statistical inferences. To

avoid such risks, these simple visualizations provide clear evidence that a nontrivial variance model should be incorporated into the model specification phase to accommodate the heterogeneous variance structures in this dataset.

Finally, the insights gleaned from these entropy-based visualization techniques agree well with the prior literature. Previous studies have repeatedly determined milk order records to be significantly more consistent than would be expected from a random queuing process using an array of correlation and regression-based approaches (Kilgour & Scott 1959; Dietrich *et al.* 1965; Soffie 1976; Grasso *et al.* 2007; Polikarpus *et al.* 2015; Beggs *et al.* 2018). Fewer papers, however, have explored differences in the consistency of entry positions between animals. Gadbury (1975) observed that only a subset of his herd seemed to demonstrate clear preferences for parlor entry positions. Such preferences do not appear to have been constrained to the front or back of the queue, however, as Gadbury (1975) also reported animals with a preference for the middle of the queue. In a more recent analysis with large commercial herds, however, Beggs *et al.* (2018) reported a nearly identical parabolic relationship between mean entry quantile and variance. With clear and consistent evidence of nonrandom patterns having been recovered from this dataset, further investigation of the behavioral mechanisms that might give rise to such heterogeneity in milk order records was clearly warranted.

Visualization of Inter-Animal Relationships

Visual inspection of the scree plot produced from PCA analysis revealed only one significant dimension was recovered from the original 80-dimensional dataset. To visualize the resulting projections, the first two principal components were plotted (Figure 3:A). Cows appeared evenly spaced along the first principal axis with no clear gaps between observations. In two dimensions points also appeared randomly scattered with no clear clustering. Thus, the PCA results revealed no compelling visual evidence of social cohesion. The color encoding further revealed that the first

principal component conveyed information about the center of each cow's entry quantile observations. As this was the only significant dimension, this may suggest that a linear model to predict variations in central moment would be a reasonable representation of this dataset. This feature of the dataset was not, however, self-evident in the geometric relationships between data points revealed by the PCA projection, and thus might have been overlooked without specification of color encoding by median entry quantile value *a priori*.

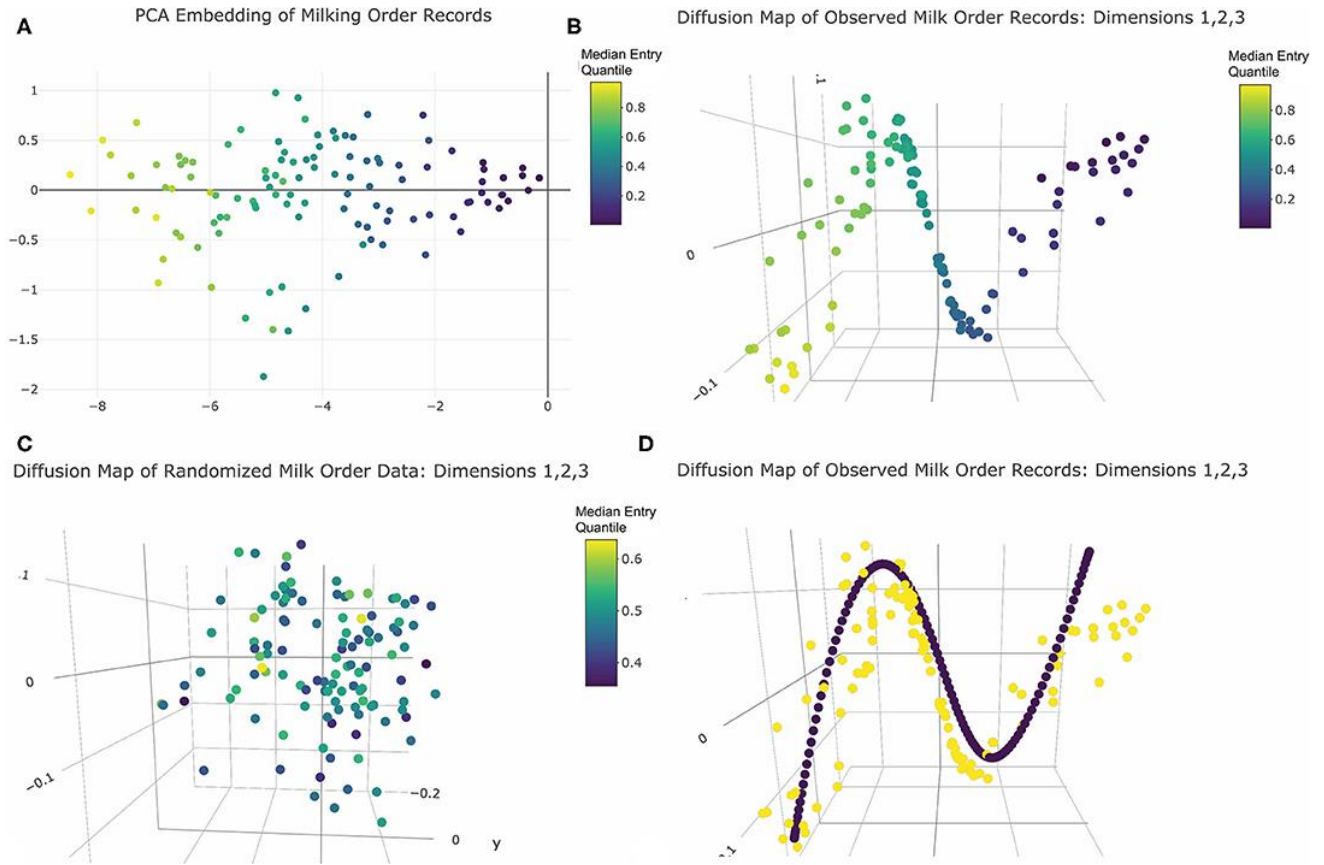


Figure 3: Dimension reduction of milk order records colored by median entry quantile value: (A) PCA embedding returns one principal axis reflecting each cows central moment of entry position. (B) The first three significant axes of diffusion map embeddings reveals a clearly linear underlying geometry reflecting central moment. (C) Diffusion map embeddding of data from the simulation of purely random queue formation recovers no distinctive geometry. (D) Harmonic series (purple points) imposed over observed cow records (yellow points) reveal curvilinear geometry is a embedding artifact.

Evaluation of eigenvalues returned by the diffusion map embedding identified five significant dimensions. The 3D visualizations of these axes in Figure 3:B and provided in supplemental materials revealed quite clearly the underlying linear geometry of this dataset. Color encodings showed that the relative positions of animals along this narrow geometric band were determined by median entry quantile, further reinforcing that central moment was the most defining feature of this dataset. As with the PCA results, cows appeared fairly evenly spread along this linear object, with no clear clustering to suggest social cohesion amongst large or temporally persistent subgroups. Comparing these results with the embedding of the permuted queue records (Figure 3:C), no clear geometric features were recovered from data simulated from a purely random queueing strategy. This reinforced that the linearity of the observed records was not simply an artifact of the physical linearization of cows as they enter the parlor single-file, but a reflection of a consistent pattern in queue formation that might be driven by some underlying behavioral or biological mechanism.

While the diffusion map embeddings convey a clear linear geometry, there was also unexplained curvature in the band along which cows were projected. This proved not to be an inherent feature of the data itself but a harmonic artifact imposed by the spectral value decomposition of the graph Laplacian used to deduce the shape of the underlying network between cows (Spielman 2009). Such a mathematical operation has several physical interpretations. One is that an singular value decomposition (SVD) of the Laplacian is akin to walking around an object in the dark and striking with a mallet at many points across its surface so that the quality of the resulting sounds can be used to discern its shape (Fisher 1966). The linear geometry of this dataset forms a “rope-like” network (Spielman 2009). When the SVD decomposition “strikes” such a network to assess the quality of sound produced, it responds like a plucked guitar string. As a result, each axis of the subsequent embedding contains an element of the harmonic series, producing the curvature seen in these milk order visualizations. Fortunately, this artifact can be described by closed form equations (Spielman 2009)

and imposed onto the data to aid in discerning authentic geometric features of the data (Figure 3:C). Thus, while diffusion map did provide a clearer geometric representation of the inherent linearity of this dataset than PCA, this dataset also reinforces that modern manifold learning techniques are also not infallible in recovering the underlying geometry of high dimensional data. While such embedding techniques may provide helpful insights into the underlying structure of large datasets, a conservative approach to visual interpretation of such results is still warranted.

Characterization of Temporal Dynamics

Independent visualization of parlor entry records from each individual cow (see Supplemental Materials) revealed that the majority of animals in this sample were surprisingly stationary in their queueing position. Animals that frequented the front and end of the queue, being more consistent in their entry position, provided clearer visual evidence for a lack of temporal trend. Cows in the middle of the queue showed far greater variability in their entry positions, making it more difficult to visually discern temporal trend from stochastic fluctuations. Only two animals were identified as having a clearly visible trend: cow 13467, who had no recorded health events, and cow 13826, who was diagnosed with metritis during the enrollment phase early in the trial. Both cows showed similar trajectories, starting nearer the end of the herd and moving progressively forward towards the front, but neither change in queue position coincided with the shift to overnight pasture access.

This consistency in queue position was further reflected in a clear linear association between median entry quantiles from overnight pen and pasture subperiods (see Figure 4). A slightly wider spread was discernable amongst cows occupying the middle ranks, but for the majority of animals, median entry quantile values did not change more than ± 0.2 . Among the handful of animals demonstrating a more extreme shift, these jumps tended to be in the forward direction towards the head of the queue. Overall, fewer extreme shifts were seen in this dataset than in a similar bivariate means

plot provided in Beggs *et al.* (2018), though this may simply be a reflection of the longer subperiods over which median entry positions were assessed. Correlations between these values were also quite high, with a Pearson correlation estimate of 0.91 ($p < 2.2e-16$) and a Kendal Tau estimate of 0.74 ($p < 2.2e-16$). These values are, as expected, higher than the estimates of consistency reported for individual milk order samples (Grasso *et al.* 2007; Berry & McCarthy 2012), but on par with results using subperiod averages on similar time scales (Soffie 1976; Beggs *et al.* 2018; Polikarpus *et al.* 2015). Given the extreme shift in management routine spanning these two subperiods, however, this level of stability in parlor entry positions was an unexpected result. Such resilience to changes in overnight housing environment and the subsequent distance traversed to access the parlor could suggest that milking order is largely determined in the crowd pen, a result supported by early observations by Soffie (1976), who reported little correlation between the order of cows exiting the home pen and entering the parlor past the first few animals.

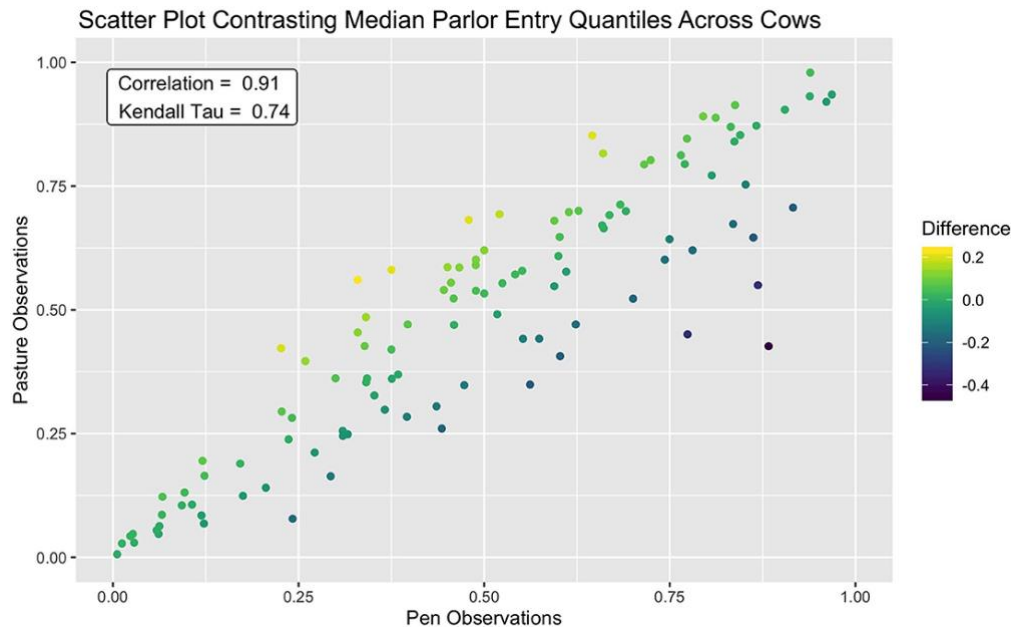


Figure 4: Scatter plot showing strong agreement of median parlor entry quantiles between the pen and pasture subperiods. A small portion of cows which demonstrated a larger shift in entry position moved forward in the queue after the shift to pasture from pen.

Collective assessment of entry quantile records using data mechanics visualizations did, however, reveal additional temporal features not identified using independent visualizations of cow records or collective assessment of aggregate records. The first and perhaps most surprising insight was that, with finer granularity in number clusters applied to the temporal (row) axis, data mechanics identified several days with anomalous queuing patterns. In Figure 5, a total of 8 column clusters are imposed without any social stratification on the subset of cows with no health events. If these records were completely stationary with no temporal effects, we would expect days to be randomly partitioned into these eight categories. Instead four days are isolated from the remaining observations. Days 85 and 91 are separated into clusters of size $n = 1$, and 89 and 91 are also isolated into their own cluster of size $n = 2$. Looking from left to right along the heat map to identify temporal heterogeneity, it is easy to see that on these observation days animals typically occupying the extremes of the queue appear to have been pushed towards the center and animals typically found in the center of the herd were either pushed towards the extremes or inverted their tendency to stay towards the front or end of this middle section of the queue. While some of the entry quantile values encompassed by these observation days would likely be identified as outliers for individual cows, other values would likely be deemed irregular but not worthy of exclusion. These clustering results, on the other hand, suggest that either transient environmental or internal social factors have disrupted the entire herd and caused them to collectively respond with highly irregular queuing patterns. As the row axis is stratified to allow for nonhomogeneous temporal responses across subsets of animals, these same days are consistently isolated from the remainder of the dataset, reinforcing that these observations constitute an outlier that should be excluded from any downstream analyses.

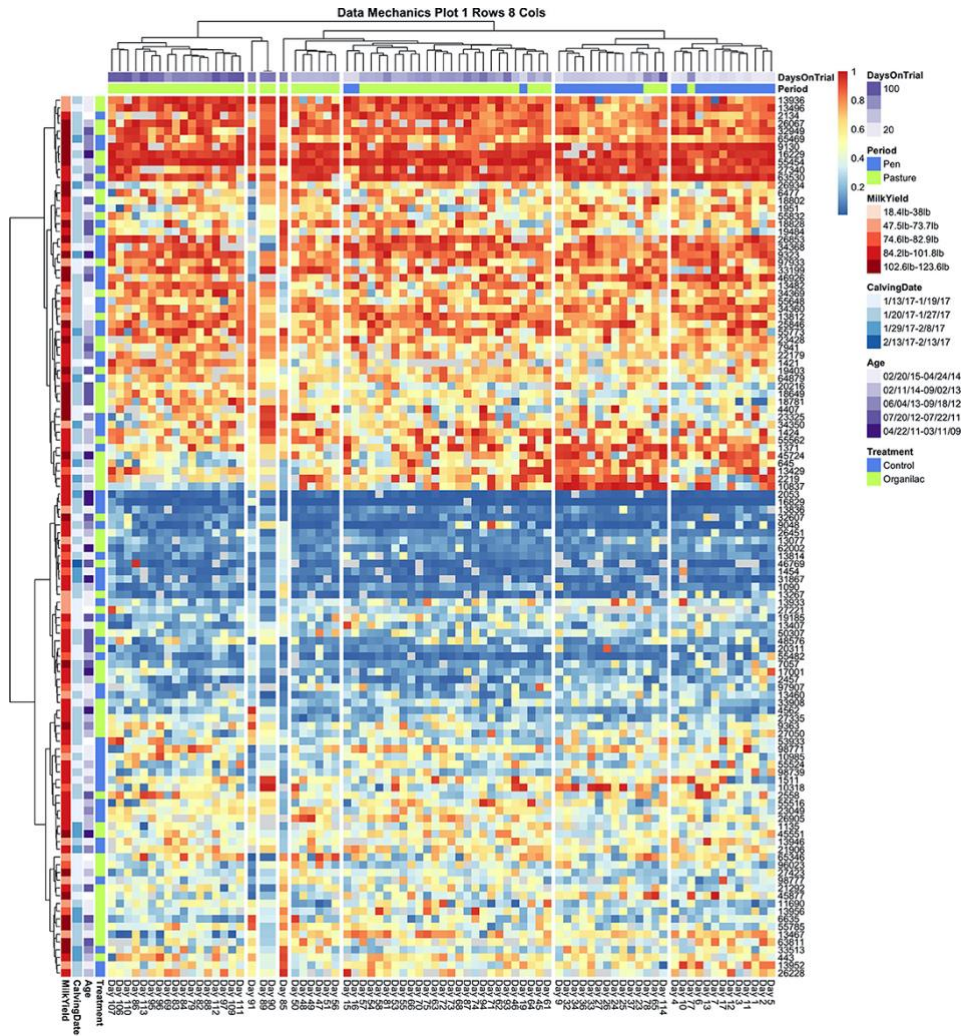


Figure 5: Data Mechanics visualization of cows with no recorded health events. Clustering along the temporal axis has isolated four days of milk order observations in middle of the pasture subperiod with anomalous queuing patterns, which can be viewed as irregularities in color values scanning from left to right. These days remain isolated in diffusion mappings which also allow for social stratification along the row axis (see supplemental materials), suggesting that these observations likely constitute outliers. Color annotations along the column axis reveal observations from the pen and pasture subperiods remain fairly distinct. No clear patterns or gradients are seen on the row color annotations for cow attributes, even though the heat map itself clearly reflects a gradient along the column axis driven by individual differences in queue entry position. Progressive clustering of the row axis did not bring any additional patterns in cow attribute variables into clear resolution.

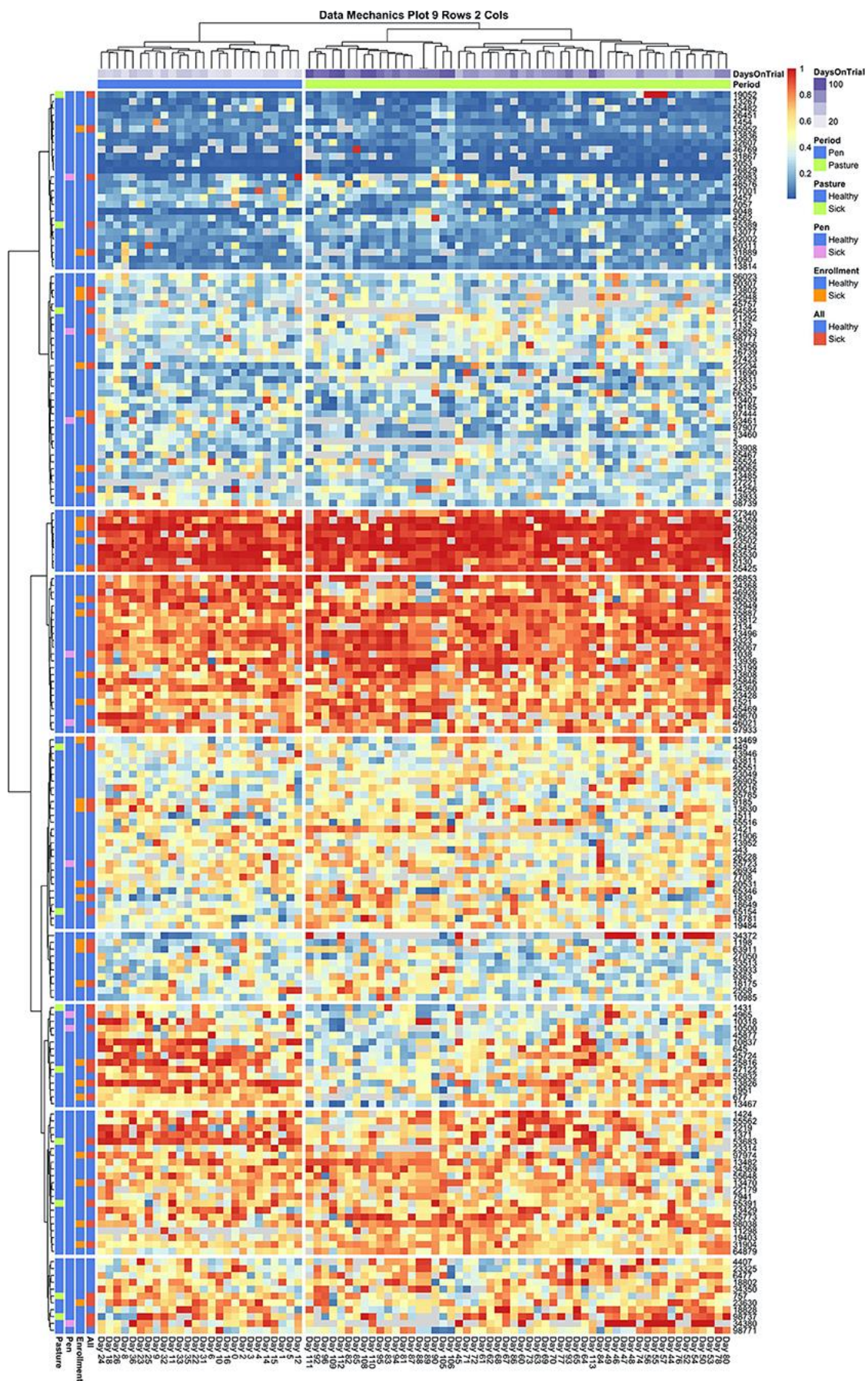


Figure 6: Data Mechanics visualization of all cows. Color annotations of the column axis reveal that coarser clustering along the temporal axis has revealed a perfect division of observations between pen and pasture subperiods. Scanning from left to right within the heat map, cows occupying the front and rear of the queue appear consistent in entry quantile values. Cows in the center of the herd appear to be the source of this temporal heterogeneity, as evidence by systematic changes in color along the column axis. In the cluster of animals starting with cow 1431, there is a progressive shift from the rear of the herd in the pen period towards the front of the herd in the pasture. As row color annotations reveal not all these animals have recorded health events on record, this pattern likely cannot be explained away by anomalous behaviors from acute or chronic illness.

Looking next at the coarser stratifications of the temporal axis, we also see that pen and pasture observations are not equally dispersed among the column clusters. As the animal (row) axis is more finely stratified to allow for social heterogeneity within the herd, the source of the temporal heterogeneity between these two environments comes into resolution. In Figure 6, which contains entry quantile observations on both sick and healthy animals, pen and pasture observations are perfectly stratified across only two column clusters. Looking at the subsets of animals who consistently entered at the front and rear of the herd, entry quantile values appear quite homogenous in color between the two temporal clusters. Scanning from left to right among the subgroups of animals that frequented the center of the queue, on the other hand, systematic fluctuations in daily entry quantile values can be seen even without finer temporal stratification. This pattern is clearest in the cluster which contains both cow 13467 and cow 13826 – the two animals identified by independent inspection of cow entry quantile plots to show evidence of non-stationarity. In this subgroup, cows showed a tendency to frequent the latter half of the queue when coming to the milking parlor from the home pen, but during the pasture period showed progressively greater proclivity to enter in the front half of the queue. Where this shift is the most uniform in the latter half of the pasture period, we also see a compensatory pattern

in the subgroup directly above, where cows shifted from nearer the front to the back half of the queue. Whether these results reflect the coordinated movement of relatively small social subgroups or just a common response to environmental conditions is impossible to say from this data alone. These results do make it clear, however, that not only are the cows occupying the center of the queue less consistently in their entry position, they are also less stable in their entry pattern. Further, these visualizations underscore that these divergent dynamics in the pen and pasture subperiods cannot be captured by a simple fixed effect term. The simplest option would be to drop from the analysis the animals that show the strongest nonstationary patterns. With such a large group, this would still leave ample observations to maintain statistical power, but could risk biasing the subsequent inferences. Alternatively, by specifying a heterogeneous variance model between animals, as was deemed necessary in the original entropy plots, the influence of these cows on the fitted model may be reduced sufficiently that deviations from the assumption of stationarity in this subgroup might not unduly destabilize the final model.

Finally, some preliminary insights can be gleaned from the cow attributes added to the row margins of both heat maps. In Figure 6, animals with documented health events appear fairly evenly dispersed across subsections of the queue. A slightly lower rate of illness might be attributed to animals that consistently occupied the very front of the queue, and perhaps a marginally higher rate of transition diseases was seen in the animals at the very rear of the queue, but these patterns appear subtle at best and thus likely not the only determinant of queue position. This result was somewhat surprising, as previous research has suggested that sick animals tend to populate the rear of the queue (Rathore 1982; Grasso *et al.* 2007; Berry & McCarthy 2012; Polikarpus *et al.* 2015). If this previously reported trend is driven by a reluctance among animals in the acute phases of a disease to move, it is possible that the daily health checks prescribed in this experimental trial succeeded in identifying and removing sick animals from the herd sufficiently early that this behavioral mechanism was not at play in this dataset.

This might suggest that inclusion of these additional animals into subsequent analyses might not unduly bias subsequent behavioral inferences. Of perhaps greater concern to subsequent modeling is the lack of clear color gradients among cows attribute values across the queue, which could indicate that underlying associations may either be weak or that there are complex interaction effects creating a nonuniform trend.

Linear Analysis of Cow Attributes

For both the full dataset and the subset of healthy animals, likelihood ratio tests revealed the heterogeneous variance model allowing for differing degrees of variability in queue position across cows to be a costly but necessary model component ($p < 0.0001$). With the model fit to cows with no recorded health events, significant linear associations were recovered for two fixed effects. Cows with higher peak milk yields demonstrated a tendency to enter nearer the rear of the queue ($\hat{B} = 0.14$, $F_{1,96} = 9.58$, $p = 0.003$). A significant interaction term revealed this trend was further amplified for older cows ($\hat{B} = 0.07$, $F_{1,96} = 6.11$, $p = 0.015$). No other terms approached significance for this dataset. With the model fit to all cows that attended at least 50% of recorded milkings, no predictors were significant at the $\alpha = 0.05$ cutoff. Peak yield remained marginally significant ($\hat{B} = 0.06$, $F_{1,152} = 2.93$, $p = 0.089$), as did the interaction term between peak yield and cow age ($\hat{B} = 0.04$, $F_{1,152} = 2.92$, $p = 0.090$). With this larger dataset, however, cow age also demonstrated a marginally significant trend, indicating older cows tended to be nearer the front of the queue ($\hat{B} = -0.07$, $F_{1,152} = 3.84$, $p = 0.052$).

In contrasting the results of these two models, the loss of significant association between entry quantile values and peak yield with the addition of sick animals is perhaps not surprising. If a disease challenge early in the trial curtailed peak lactation in these cows but did not cause chronically deficient production, then the 95-th quantile value of milk yield used here to estimate peak lactation level may

not adequately reflected the overall productivity of these animals across the duration of this extended trial, obscuring the underlying biological mechanism. The emergence of a nearly significant association between entry quantile and age with the addition sick animals, however, is more difficult to explain. Given that peak yield and age are highly correlated biological parameters ($r_{all} = 0.66$, $r_{healthy} = 0.70$), this sample may simply contain too few older cows with low productivity levels by which to disentangle the positive association with peak yield from the negative association with cow age. Alternatively, if a diseased state permanently alters a cow's queueing pattern and if risk of health complications in turn varies with age, then health status may be a lurking variable masquerading as an age effect. In either case, a relatively small number of animals may be unduly influencing statistical inferences.

Visual examination of predicted queue positions plotted against age and peak yield for both the full dataset (see Figure 7) and healthy subset (see Supplemental) seem to confirm these misgivings. Looking first at age, the first lactation heifers, being evenly spread across the center of the queue, cannot be driving this linear effect. Among the multiparous animals, the five cows seen consistently in the front of the queue are indeed among the oldest in the herd, but if this handful of animals and their corresponding queue positions are ignored, a clear gradient is not visible among the remaining cows. Results of the mutual conditional entropy tests confirm this suspicion. For the disease free subset the MCE test confirms the insignificant association found in the linear model ($p_{2,2} = 0.103$). For the full dataset, where the linear effect is marginally significant, the MCE test does not ($p_{7,2} = 0.305$). This suggests that either that age effect is only discernable after adjusting for peak yield or that the association is not robust.

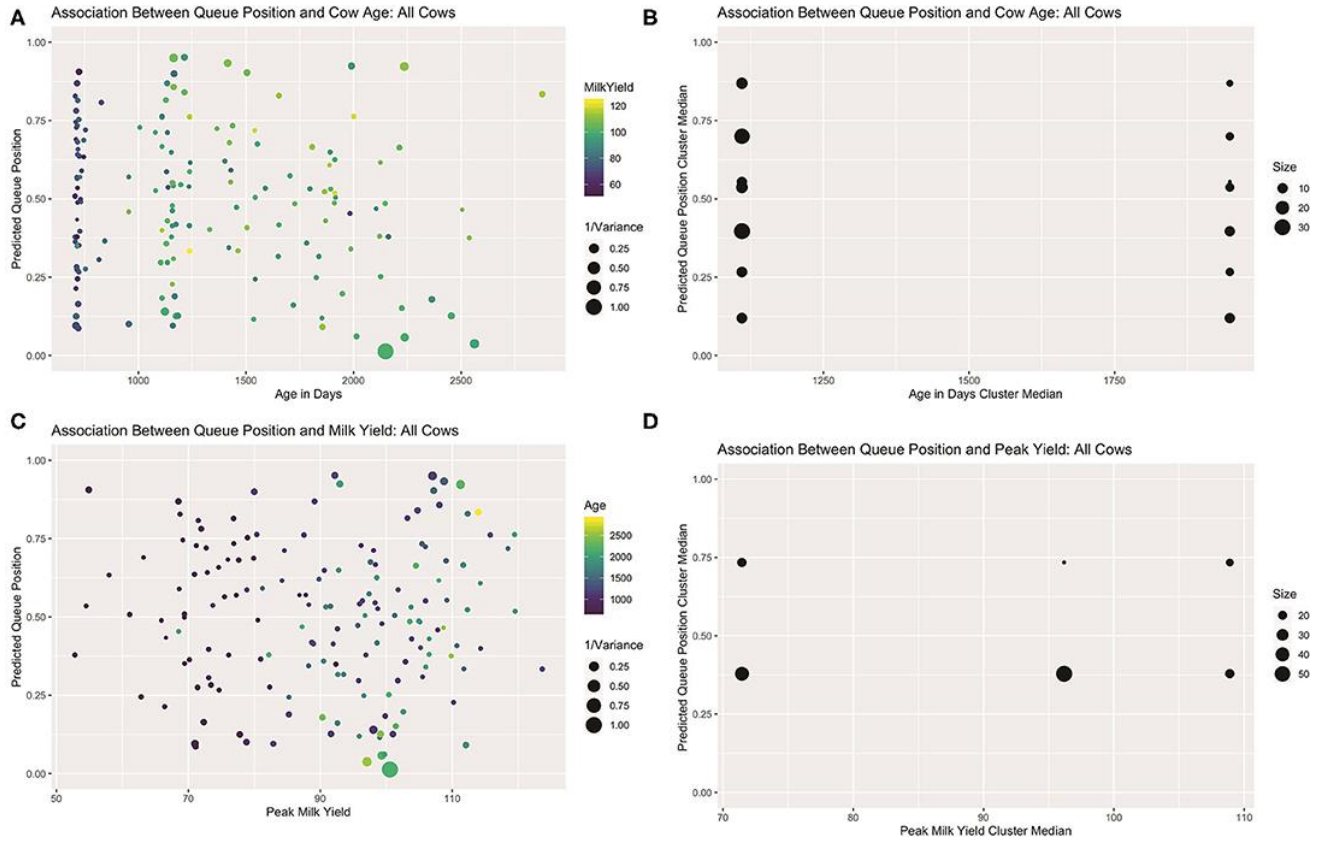


Figure 7: Visualizations for all cows that attended >50% of observed milkings. Scatter plots of predicted queue position against cow age (A) and peak yield (B), scaled for heterogeneity in repeated samples of entry quantiles. For both attributes, linear trends are only readily discerned among multiparous animals. Sparsity in parity 3+ animals might skew the calculation of age effect as a continuous variable. (B) Visualization of contingency table results derived from independent clusterings used in MCE test of bivariate association. Age effect is lost with empirical discretization (C). A nonlinear trend is recovered with peak yield (D).

Looking next at peak yield, a clear global trend could not be discerned. Among the lower-yielding cows, a group comprised predominantly of heifers, a linear trend is difficult to discern, but amongst older cows a slight positive gradient is perhaps perceptible. Results of the mutual conditional entropy tests not only confirmed this trend among the healthy animals ($p_{3,5} = 0.036$), but also within the full dataset ($p_{2,3} = 0.012$) where the linear effect was only suggestive. Visualization of the

contingency table for this later result revealed no distinctive pattern among the lowest and highest yielding clusters, but a nearly perfect division of roughly 50 moderate-yielding cows into the leading queue cluster. This result suggests that the MCE tests may also be used in mixed modeling analyses to recover nonrandom patterns that are not well-represented by linear trends. Such a nonlinear trend here could belie more complex interaction effects between these or other unmeasured biological drivers of queue position. Alternatively, a multilevel model may be necessary to disentangle complex hierarchical relationships between the drivers of position preference and a cow's ability to assert that preference.

Contextualizing these results within the existing base of literature underscores the inconsistency in drivers of queuing behaviors. With respect to milk yield, several studies have found no significant association (Soffie 1976; Polikarpus *et al.* 2015), but among those that have, most have reported high yielding cows frequent the front of the queue. Differences in motivation to obtain feed might explain this result. In early studies, cows were offered concentrate in the milking parlor, which may have increased the motivation of high yielding animals with greater energy deficiencies to enter the parlor (Gadbury 1975; Rathore 1982). In more recent work, cows may have been motivated to access limited feed bunk space on commercial dairies (Berry & McCarthy 2012) or to obtain prime pasture (Scott *et al.* 2014). In this study, as all animals were locked following milking to facilitate feeding treatments and health checks, cows would have had ample access silage regardless of queue position. Alternatively, Rathore (1982) suggested greater intermammary pressure might motivate high yielding animals to be milked earlier. As this herd was milked three times daily, however, this biological driver may also have been attenuated. Indeed, among modern studies with herds milked thrice daily, Polikarpus *et al.* (2015) found no significant correlation and Grasso *et al.* (2007) also found high yielding cows frequented the rear of the queue. Ultimately, as yield is influenced by a wide range of health and management factors, any number of confounding variables might be implicated in explaining this somewhat unexpected result. In this study a significant linear association between age

and entry position was not found. Recent work by Berry and McCarthy (2012), which identified a nonlinear trend across parity, and by Grasso *et al.* (2007), which highlighting significant interactions of parity with other biological drivers of queue position, suggests that a linear effect may not adequately capture the underlying biological relationship. A larger and more structured sample may be necessary to bring more complex age dynamics into clearer resolution.

Exploring Associations between Sensor and Queue Records

Visual inspection of means plots produced from mixed model analysis of sensor records recovered only a handful of statistically significant differences between queue quartiles when hour and day effects were assessed individually, but several global trends were still readily visible. With respect to minutes recorded as active, the 1st-3rd queue quartiles were visually indistinguishable in their cyclical behavioral patterns, but cows in the fourth queue quartile were consistently more active, particularly during the night and morning lounging period. With respect to longitudinal trends across days, fourth queue quartile animals were again more active across the observation window, whereas cows in the first queue quartile were consistently the least active. These patterns were somewhat mirrored in the longitudinal and cyclical analysis of high activity minutes, but the pattern was both less distinct and less consistent. No clear qualitative insights could be drawn for cyclical or longitudinal patterns in non-activity. Cyclical patterns in minutes spent eating were not seen overnight or in the afternoon, but first queue quartile cows may have spent slightly more time eating after the morning milking. Longitudinal analysis of eating patterns suggested cows in the fourth queue quartile spent relatively less time eating, whereas the cows in the first and second queue quartile consistently spent more time at the bunk. This contrasted with longitudinal results for minutes spent ruminating, where the cows in the second queue quartile were consistently low. No clear distinctions between groups were recovered in cyclical rumination patterns. Temperature patterns were, surprisingly, the most visually distinct of all the sensor

parameters. Cows in the first queue quartile were consistently lower in body temperature in both the longitudinal and cyclical time dimensions as compared with the remainder of the herd.

	All	Lounging	Morning	Afternoon	Night
Non-activity	<0.001 _{7,5}	0.048 _{2,5}	0.432 _{8,7}	0.002 _{2,5}	0.132 _{2,6}
Activity	<0.001 _{3,8}	0.006 _{2,8}	0.038 _{2,11}	<0.001 _{2,3}	0.033 _{2,9}
High Activity	0.052 _{2,9}	0.028 _{2,2}	0.272 _{10,9}	0.306 _{3,4}	0.014 _{2,4}
Eating	0.004 _{7,6}	0.234 _{7,5}	0.188 _{6,4}	0.066 _{2,4}	0.212 _{9,2}
Rumination	0.021 _{2,4}	0.05 _{9,3}	0.325 _{7,2}	0.083 _{10,3}	0.152 _{8,10}
Temperature	<0.001 _{5,10}	0.004 _{5,5}	0.022 _{7,3}	0.006 _{3,5}	0.015 _{5,3}

Table 2: P-values generated from mutual conditional entropy tests comparing queue records to sensor logs. Subscripts represent the number of clusters used to discretize the row variable (queue records) and column variable (sensor records). For example, in the test for morning non-activity, cows were assigned to 8 clusters using queue records and 7 clusters using sensor logs. Activity and temperature data demonstrated the strongest association with queue records. The afternoon lounging period produced the strongest associations between queue and all sensor dimensions save for high activity, which showed the strongest distinction overnight.

While the preceding analyses revealed few statistically significant differences at individual time points, collective analysis of days and subsets of the 24-hour management cycle would undoubtedly return statistically significant differences for the broader qualitative trends visually identified via mean plots. Within a linear modeling framework, however, this constitutes no small task. For all of the above models, Wald's tests revealed Group-by-Hour interactions effects to be highly significant components of the model ($p < 0.0001$). Group-by-date interaction effects were also significant for activity, high activity, and temperature models ($p < 0.05$). This suggests that these models should not be simplified to a single cyclical or longitudinal trend, which would allow overall differences between groups to be tested via a single group intercept term. Targeted hypotheses

comparing comprehensive trends between groups would instead require formulation of linear contrasts – a daunting task with so many fixed effects terms used to accommodate the high sampling frequency and extended observation period of this dataset. Further, as with the linear models with cow attributes, behavioral synchronization due to social cohesion or compensatory use of physical resources in the pen could again create non-independence between animals in such sensor records. Any such issues in estimation of model degrees of freedom, compounded with the inability to fit behaviorally and empirically compelling correlation and variance models, would only serve to further confound the estimation of appropriate p-values from these models.

Fortunately, the qualitative trends identified via the preceding means plots largely aligned with the significant bivariate associations identified by mutual conditional entropy tests summarized in Table 2. Activity again proved to be the most distinctive behavioral axis. Significant associations were identified for all three lounging periods when analyzed both independently and in aggregate, with the afternoon lounging periods being the most distinct. High activity also showed a significant relation to queue records, but this association may have been driven predominantly by overnight lounging period. Whereas no clear qualitative patterns were identified for non-activity data via the means plots, a significant association with queue records was identified during the afternoon lounging period. A highly significant relationship was identified for time spent eating for the full sensor record, but given that time budgets recorded by this platform were segmented somewhat arbitrarily at the start of each hour, this result may simply reflect a lag in the arrival of cows to the feed bunk after exiting the parlor. Significant associations were not found during the lounging periods at the standard $\alpha = 0.05$ cutoff, though records from the afternoon lounging period approached significance. These results were mirrored in rumination patterns, where again no significant association was recovered, but the afternoon lounging period approached significance. Finally, as with the linear modeling results, temperature proved highly distinct between queue subgroups for all subperiods.

Visual inspection of tube plots produced with median queue subgroup values again yielded insights comparable to the linear modeling results (Figure 8). Based on the results of the MCE tests, all behavioral axes cows were clustered into two subgroups based on queueing records, with Group 1 cows consisting of 80 animals at the front of the queue, and Group 2 cows constituting the 34 animals in the rear. Tube plots of minutes spent active revealed Group 2 cows to be more active across all three lounging periods. This pattern was the most consistent in the morning and overnight lounging periods, though this difference was ultimately quite subtle and seldom constituted more than a few minutes. In the afternoon subperiod there was evidence of several periods with anomalously high activity levels, most of which occurred post-pasture access. The significant association recovered for minutes spent highly active in the overnight subperiod appeared to be largely driven by increased activity immediately following the evening milking, which could reflect divergent home pen behaviors, but might also have been driven by delays in milking. To complement these results for active and high active minutes, the significant association for afternoon non-activity records appears to have been driven by increased non-activity among the Group 1 cows during the three hours immediately preceding the night milking. As anticipated, differences in time spent eating were largely restricted to the 2-3 hours immediately following milking. Cows only lingered at the feed bunk during the morning lounging period, where median eating times for Group 1 cows were perhaps slightly higher. Similarly, differences in rumination also appeared restricted to time periods immediately following milking, with no clear differences seen during the lounging period with this coarse stratification of animals. Finally, as with the mean plots, body temperature values again produced surprisingly distinctive results. More finely segmented into five queue groups by the mutual conditional entropy test, the tube plots proved a slightly cumbersome means of comparing temperature records, but a clear visual distinction could still be made between the Group 2 animals and the remainder of the herd. For all three lounging periods, this relatively small cluster of 17 cows that constituted the very front of the milking queue demonstrated lower median body temperature values, a distinction seen most clearly at night.

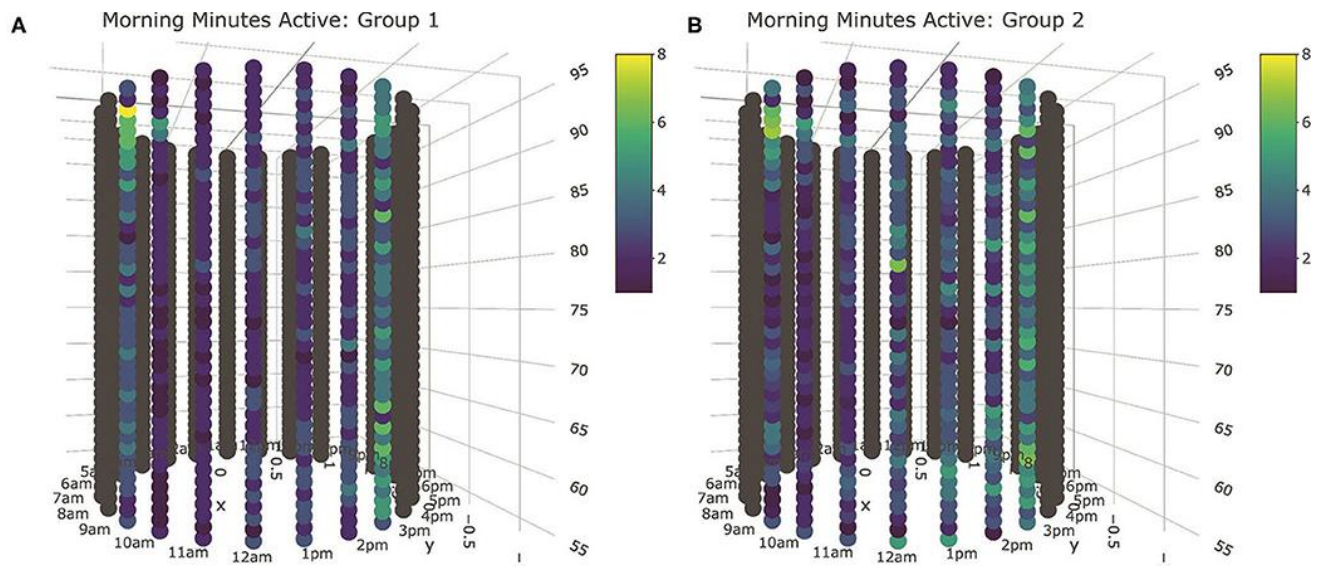


Figure 8: Tube plot visualizations for minutes spent active during the morning lounging period. Based on the results of the MCE tests, all behavioral axes cows were clustered into two subgroups based on queueing records, with Group 1 cows (A) consisting of 80 animals at the front of the queue, and Group 2 cows (B) constituting the 34 animals in the rear. Cyclical patterns can be seen around the diameter of the tube, and longitudinal patterns observed along its length. The median activity level of Group 2 animals, consisting of 34 animals in the rear of the queue, is heightened during the middle of the lounging period for the duration of the trial. Interactive 3D plots for all sensor output fields are provided in Supplemental Materials.

The strong agreement between the results of these two analytical pipelines suggests that UML and conventional linear modeling approaches could be used interchangeably or in concert to glean preliminary insights from exploratory analyses of large sensor-based datasets that may inform future hypothesis-driven studies. Perhaps the most surprising result of these analyses, that cows frequenting the back of the queue are consistently more active between milkings, may indeed warrant further exploration. In much of the prior literature, health challenges that impede movement (lameness, subclinical mastitis, etc.) have been identified as the main driver of delayed entry into the parlor (Gadbury 1975; Flower 2006; Grasso *et al.* 2007; Polikarpus *et al.* 2015). In fact, this mechanism is so well-established that it has even been proposed that milk order records might be incorporated into genetic evaluations to improve estimates of health traits (Berry & McCarthy 2012). As these analyses were run on the subset of animals with no recorded health events, however, it is possible that this dataset has brought other behavioral mechanisms into focus.

One potential explanation for these results might be a dominance gradient. Previous studies have found that animals of low social status frequent the rear of the herd in voluntary movements (Beilharz and Mylrea 1963; Kilgour and Scott 1959), and social dominance is known to impact resource access in spatially constricted conditions (Wierenga 1990; Alm & Moller 2004) such as those found at the entrance to the milking parlor. If low dominance animals are in turn also forced to wait longer or walk farther to access resources in the home pen, this could potentially explain the increased activity levels of animals found in the rear of the queue. While the early literature has found the relationship between dominance value and milking order to be tenuous at best (Gadbury 1975; Rathore 1982; Soffie 1976; Dickson *et al.* 1967), it is possible that such social mechanisms may have been confounded by health status, with linear analyses of limited sample size failing to disentangle these mechanisms in non-disaggregated data. Alternatively, in more recent analyses in automated milking systems, where dominance has proven highly correlated with milking order (Olofsson 2000), greater

attention has also been paid to “avoiders” – animals that seem to actively avoid social interactions and therefore occupy no clear position in the herd hierarchy (Danielson 2012). On this farm, where resources are not severely restricted and animals are frequently remixed, energetic investments in a dominance hierarchy may offer few returns (Pagel & Dawkins 1996). Such a behavioral strategy might also explain why it is high-yielding multiparous cows and not heifers that occupy the end of the queue. Both these hypotheses are ultimately purely speculative interpretations of these exploratory results; however, if proposals to incorporate milk order records into genetic indices are progressed, any correlations between queueing position and consistent individual differences in home pen behaviors likely warrant closer inspection to mitigate the risk of unintended and potentially deleterious selection pressures.

Conclusions

As with previous studies of milk order records, these analyses perhaps raise more questions than answers. As dairy record management systems grow to accommodate an ever wider range data streams, perhaps future work considering more herds from a wider range of management strategies will succeed in further untangling the complex web of explanatory variables at the individual, herd, and farm levels that drive variation in queueing patterns. This dataset has, none the less, demonstrated the utility of unsupervised machine learning tools in ethological studies using sensor platforms to study larger groups of animals over extended periods of time. While these analyses recovered no evidence of social cohesion amongst large or temporally consistent subgroups, information theoretic approaches succeeded in clarifying the underlying pattern of heterogeneity in error variance between animals and also demonstrated an advantage in recovering evidence of non-uniform patterns in temporal nonstationary over basic EDA tools. After incorporating these insights into the structure of subsequent linear models, these model-free tools then showed some capacity to confirm inferential results where probabilistic assumptions were not strictly met, as well as an aptitude for recovering significant

associations not captured by a simple linear effect. This flexible clustering-based approach to identifying significant bivariate associations was then easily extended to accommodate two high dimensional behavioral axes, providing equivalent insights to more computationally taxing mixed effect models. While UML approaches are by no means infallible, as seen here with artifacts produced by the spectral embeddings, these analyses have demonstrated that such tools can add value at every stage of the standard hypothesis-driven linear analysis pipeline, and may even offer an advantages over model-based approaches in early-stage exploratory projects. While many new methodological developments are doubtless on the ethological horizon, we hope this algorithmic toolset will provide a meaningful step forwards to meet the challenges of a future defined by ever larger and more complex data.

References

- Alm, A., Möller, J., n.d. Welfare of dairy cows in AMS and conventional loose housing – differences in behaviour and the hormones oxytocin and cortisol between cows high or low in social rank. Swedish University of Agricultural Sciences, Department of Animal Nutrition and Management, Uppsala.
- Banhazi, T.M., Lehr, H., Black, J.L., Crabtree, H., Schofield, P., Tschärke, M., Berckmans, D., 2012. Precision Livestock Farming: An international review of scientific and commercial aspects. *Int J Agric & Biol Eng* 5, 1–9.
- Barter, E., Gross, T., 2019. Manifold Cities: Social variables of urban areas in the UK. *Proc. R. Soc. A* 475, 20180615. <https://doi.org/10.1098/rspa.2018.0615>
- Beggs, D.S., Jongman, E.C., Hemsworth, P.H., Fisher, A.D., 2018. Short communication: Milking order consistency of dairy cows in large Australian herds. *Journal of Dairy Science* 101, 603–608. <https://doi.org/10.3168/jds.2017-12748>
- Beilharz, R.G., Mylrea, P.J., 1963. Social position and movement orders of dairy heifers. *Animal Behaviour* 11, 529–533. [https://doi.org/10.1016/0003-3472\(63\)90275-6](https://doi.org/10.1016/0003-3472(63)90275-6)
- Berry, D.P., McCarthy, J., 2012. Genetic and non-genetic factors associated with milking order in lactating dairy cows. *Applied Animal Behaviour Science* 136, 15–19. <https://doi.org/10.1016/j.applanim.2011.11.012>
- Biro, P.A., Dingemanse, N.J., 2009. Sampling bias resulting from animal personality. *Trends in Ecology & Evolution* 24, 66–67. <https://doi.org/10.1016/j.tree.2008.11.001>

- Bruneel, H., Maertens, T., Steyaert, B., Claeys, D., Fiems, D., Walraevens, J., 2012. Analysis of a Two-Class FCFS Queueing System with Interclass Correlation, in: Al-Begain, K., Fiems, D., Vincent, J.-M. (Eds.), *Analytical and Stochastic Modeling Techniques and Applications*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 32–46. https://doi.org/10.1007/978-3-642-30782-9_3
- Coifman, R.R., Lafon, S., Lee, A.B., Maggioni, M., Nadler, B., Warner, F., Zucker, S.W., 2005. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the National Academy of Sciences* 102, 7426–7431. <https://doi.org/10.1073/pnas.0500334102>
- Danielsson, T., 2004. The effect of social rank on milking and feeding behaviour in automatic milking system for dairy cows (Master's Thesis). Swedish University of Agricultural Sciences, Department of Animal Nutrition and Management, Uppsala.
- Dawkins, M.S., 2007. *Observing Animal Behaviour: Design and Analysis of Quantitative Data*. Oxford University Press, Oxford, UK.
- Dickson, D.P., Wieckert, D.A., Barr, G.R., 1967. Social Relationship of Dairy Cows in a Feed Lot. *Behav* 29, 195–203. <https://doi.org/10.1163/156853967X00118>
- Dietrich, J.P., Snyder, W.W., Meadows, C.E., Albright, J.L., 1965. Rank order in dairy cows. *Anim. Zool.* 5.
- Ellen, E., van der Sluis, M., Siegford, J., Guzhva, O., Toscano, M., Bennewitz, J., van der Zande, L., van der Eijk, J., de Haas, E., Norton, T., Piette, D., Tetens, J., de Klerk, B., Visser, B., Rodenburg, T., 2019. Review of Sensor Technologies in Animal Breeding: Phenotyping

Behaviors of Laying Hens to Select Against Feather Pecking. *Animals* 9, 1–21.

<https://doi.org/10.3390/ani9030108>

Fisher, M.E., 1966. On hearing the shape of a drum. *Journal of Combinatorial Theory* 1, 105–125.

[https://doi.org/10.1016/S0021-9800\(66\)80008-X](https://doi.org/10.1016/S0021-9800(66)80008-X)

Flower, F.C., Sanderson, D.J., Weary, D.M., 2006. Effects of Milking on Dairy Cow Gait. *Journal of Dairy Science* 89, 2084–2089. [https://doi.org/10.3168/jds.S0022-0302\(06\)72278-0](https://doi.org/10.3168/jds.S0022-0302(06)72278-0)

Forstmeier, W., Wagenmakers, E.-J., Parker, T.H., 2017. Detecting and avoiding likely false-positive findings - a practical guide. *Biol Rev* 92, 1941–1968. <https://doi.org/10.1111/brv.12315>

Gadbury, J.C., 1975. Some preliminary field observations on the order of entry of cows into herringbone parlours. *Applied Animal Ethology* 1, 275–281. [https://doi.org/10.1016/0304-3762\(75\)90020-6](https://doi.org/10.1016/0304-3762(75)90020-6)

Gareth, J., Witten, D., Hastie, T., Tibshirani, R., 2013. *An Introduction to Statistical Learning*, Springer Texts in Statistics. Springer New York, New York, NY. <https://doi.org/10.1007/978-1-4614-7138-7>

Grasso, F., De Rosa, G., Napolitano, F., Di Francia, A., Bordi, A., 2007. Entrance order and side preference of dairy cows in the milking parlour. *Italian Journal of Animal Science* 6, 187–194. <https://doi.org/10.4081/ijas.2007.187>

Guan, J., Hsieh, F., 2018. Coupling Geometry on Binary Bipartite Networks: Hypotheses Testing on Pattern Geometry and Nestedness. *Front. Appl. Math. Stat.* 4, 38. <https://doi.org/10.3389/fams.2018.00038>

- Halachmi, I., Guarino, M., 2016. Editorial: Precision livestock farming: a ‘per animal’ approach using advanced monitoring technologies. *Animal* 10, 1482–1483.
<https://doi.org/10.1017/S1751731116001142>
- Hsieh, F., Liu, S.-Y., Hsieh, Y.-C., McCowan, B., 2018. From patterned response dependency to structured covariate dependency: Entropy based categorical-pattern-matching. *PLoS ONE* 13, 1–28. <https://doi.org/10.1371/journal.pone.0198253>
- Izenman, A.J., 2008. *Modern Multivariate Statistical Techniques*, Springer Texts in Statistics. Springer New York, New York, NY. <https://doi.org/10.1007/978-0-387-78189-1>
- Kilgour, R., Scott, T.H., 1959. Leadership In A Herd Of Dairy Cows. *Proceedings of the New Zealand Society of Animal Production* 19, 36–43.
- Kirby, M., 2001. *Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns*. John Wiley & Sons, Inc, New York, USA.
- Kolde, R., 2019. pheatmap: Pretty Heatmaps.
- Lenth, R., 2019. emmeans: Estimated Marginal Means, aka Least-Squares Means.
- Lin, B., He, X., Ye, J., 2015. A geometric viewpoint of manifold learning. *Appl Inform* 2, 1–12.
<https://doi.org/10.1186/s40535-015-0006-6>
- Manriquez, D., Chen, L., Albornoz, G., Velez, J., Pinedo, P., 2018. Case Study: Assessment of human-conditioned sorting behavior in dairy cows in farm research trials. *The Professional Animal Scientist* 34, 664–670. <https://doi.org/10.15232/pas.2018-01749>

- Manriquez, D., Chen, L., Melendez, P., Pinedo, P., 2019. The effect of an organic rumen-protected fat supplement on performance, metabolic status, and health of dairy cows. *BMC Vet Res* 15, 1–14. <https://doi.org/10.1186/s12917-019-2199-8>
- Martin, P., Bateson, P., 2007. *Measuring Behavior: An Introductory Guide*, 3rd ed. Cambridge University Press, Cambridge UK.
- McCowan, B., Beisner, B., Bliss-Moreau, E., Vandeleest, J., Jin, J., Hannibal, D., Hsieh, F., 2016. Connections Matter: Social Networks and Lifespan Health in Primate Translational Models. *Front. Psychol.* 7. <https://doi.org/10.3389/fpsyg.2016.00433>
- Neuwirth, E., 2014. *RColorBrewer: ColorBrewer Palettes*.
- Olofsson, J., 2000. *Feed Availability and Its Effects on Intake, Production and Behaviour in Dairy Cows (Doctoral Thesis)*. Swedish University of Agricultural Sciences, Uppsala.
- Pagel, M., Dawkins, M.S., 1997. Peck orders and group size in laying hens: “futures contracts” for non-aggression. *Behavioural Processes* 40, 13–25. [https://doi.org/10.1016/S0376-6357\(96\)00761-9](https://doi.org/10.1016/S0376-6357(96)00761-9)
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, 2018. *nlme: Linear and Nonlinear Mixed Effects Models*.
- Pinheiro, J.C., Bates, D.M., 2000. *Mixed-Effects Models in S and S-PLUS*, Statistics and Computing. Springer-Verlag, New York USA.
- Polikarpus, A., Kaart, T., Mootse, H., De Rosa, G., Arney, D., 2015. Influences of various factors on cows’ entrance order into the milking parlour. *Applied Animal Behaviour Science* 166, 20–24. <https://doi.org/10.1016/j.applanim.2015.02.016>

- R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rathore, A.K., 1982. Order of cow entry at milking and its relationships with milk yield and consistency of the order. *Applied Animal Ethology* 8, 45–52. [https://doi.org/10.1016/0304-3762\(82\)90131-6](https://doi.org/10.1016/0304-3762(82)90131-6)
- Reinhardt, V., 1982. Movement Orders and Leadership in a Semi-Wild Cattle Herd. *Behav* 83, 251–264. <https://doi.org/10.1163/156853983X00183>
- Rosser, G., Fletcher, A.G., Maini, P.K., Baker, R.E., 2013. The effect of sampling rate on observed statistics in a correlated random walk. *J. R. Soc. Interface* 10, 20130273. <https://doi.org/10.1098/rsif.2013.0273>
- Saul, L.K., Roweis, S.T., 2003. Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds. *Journal of Machine Learning Research* 4, 37.
- Schein, M.W., Fohrman, M.H., 1955. Social dominance relationships in a herd of dairy cattle. *The British Journal of Animal Behaviour* 3, 45–55. [https://doi.org/10.1016/S0950-5601\(55\)80012-3](https://doi.org/10.1016/S0950-5601(55)80012-3)
- Scott, B.A., Camacho, A., Golder, H., Molfino, J., Kerrisk, K.L., Lean, I., Garcia, S.C., Chaves, A.V., Hall, E., Clark, C.E.F., 2014. The nutritive value of pasture ingested by dairy cows varies within a herd. *Proceedings of the 5th Australasian Dairy Science Symposium* 343–346.
- Shannon, C.E., 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal* 27, 379–423.
- Sievert, C., 2018. plotly for R.

Snijders, T.A.B., Bosker, R.J., 2012. Multilevel Analysis: An Introduction to Basic & Advanced Multilevel Modeling, 2nd ed. SAGE Publications Ltd, London UK.

Soffié, M., Thinès, G., De Marneffe, G., 1976. Relation between milking order and dominance value in a group of dairy cows. *Applied Animal Ethology* 2, 271–276. [https://doi.org/10.1016/0304-3762\(76\)90060-2](https://doi.org/10.1016/0304-3762(76)90060-2)

Spielman, D.A., 2009. Spectral Graph Theory: The Laplacian.

Valletta, J.J., Torney, C., Kings, M., Thornton, A., Madden, J., 2017. Applications of machine learning in animal behaviour studies. *Animal Behaviour* 124, 203–220. <https://doi.org/10.1016/j.anbehav.2016.12.005>

Wathes, C.M., Kristensen, H.H., Aerts, J.-M., Berckmans, D., 2008. Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall? *Computers and Electronics in Agriculture* 64, 2–10. <https://doi.org/10.1016/j.compag.2008.05.005>

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wierenga, H.K., 1990. Social dominance in dairy cattle and the influences of housing and management. *Applied Animal Behaviour Science* 27, 201–229. [https://doi.org/10.1016/0168-1591\(90\)90057-K](https://doi.org/10.1016/0168-1591(90)90057-K)

CHAPTER 2

*Livestock Informatics Toolkit: A Case Study in Visually Characterizing Complex
Behavioral Patterns across Multiple Sensor Platforms, Using Novel Unsupervised
Machine Learning and Information Theoretic Approaches*

Published in Sensors (2022, 22:1)

Introduction

Precision Livestock Farming (PLF) technologies produce prodigious amounts of data (Stygar *et al.* 2021). While the behaviors encoded by such sensors are often much simpler than those that can be quantified by a human observer, the measurement granularity and perseverance provided by these technologies create new opportunities to study complex behavioral patterns across time and in a wider range of contexts. Observations collected on a single animal over extended observation windows at high sampling frequencies can, however, contain a range of complex temporal patterns – cyclicity, non-stationarity, autocorrelation, etc (Pineiro and Bates 2000). Further, when sensors are applied to large heterogenous groups of animals housed socially in spatially restricted environments, recorded behaviors may also contain complex interdependencies between animals at the dyadic, triadic, clique, and herd levels (Cooper *et al.* 2008; McCowan *et al.* 2016; Farine 2017). Failing to accommodate all these complex structural and stochastic features in a conventional model-based approach to

statistical inference risks returning spurious insights into the underlying behavioral dynamics. Developing such a model with a single PLF data stream can be challenging. Provided multiple data streams, however, the logistical challenges presented by model-based analytical frameworks can rapidly compound, creating significant barriers to cross-sensor inferences and thereby impeding researchers from extracting more holistic behavioral inferences from increasingly data-rich farm environments.

Unsupervised Machine Learning (UML) tools may provide a more flexible and forgiving approach to knowledge discovery in the context of large sensor datasets (Valletta *et al.* 2017; McVey *et al.* 2020). Such algorithms excel at identifying and characterizing complex non-random behavioral patterns lying beneath the stochastic surface of a dataset, while often employing relatively few structural assumptions about the data (Kirby 2001; Hastie *et al.* 2009; Gareth *et al.* 2013). Hierarchical clustering-based techniques offer an intuitive and highly adaptable approach to visualizing high dimensional datasets that is particularly well-suited to exploratory data analysis (Kirby 2001; Cooper *et al.* 2008; Gareth *et al.* 2013; McCowan *et al.* 2016; Valletta *et al.* 2017; McVey *et al.* 2020). Indeed, by reducing the complex behavioral signals present in a sensor dataset into a series of discrete clusters, such algorithms may be viewed as an empirical extension of classical ethological techniques. Discrete data, however, can be challenging to work with in most frequentist and even many Bayesian frameworks. Estimators based on information entropy, on the other hand, are purpose-made to quantify uncertainty in discretely encoded data without knowledge of the underlying distribution, and

thus naturally complement hierarchical clustering-based algorithms (Kirby 2001; MacKay 2003; Hastie *et al.* 2009; McVey *et al.* 2020; Hsieh *et al.* 2021).

Clustering algorithms, by virtue of their incredible flexibility, have successfully been applied to a range of PLF data streams (Schwager *et al.* 2007; Brenninkmeyer *et al.* 2016; Adamczyk *et al.* 2017; Dutta *et al.* 2015; Lee *et al.* 2020; McVey *et al.* 2020; Xu *et al.* 2020). In our previous work, we've highlighted the utility of hierarchical clustering-based approaches in leveraging the behavioral co-dependencies of cows housed socially in large groups in a production environment in order to recover complex temporal patterns in behavior (McVey *et al.* 2020). In these analyses, data mechanics algorithms were able to recover complex nonstationarity in the order in which cows entered the milking parlor. Some of these changes in queuing patterns could attributed to the shift to spring pasture access, but other transient and persistent shifts in entry order recovered in these encodings that may have been driven by environmental factors not experimentally recorded (Guan and Hsieh 2018; Hsieh *et al.* 2018; McVey *et al.* 2020). Entropy-based nonparametric permutation tests were also successful in recovering preliminary evidence of significant nonlinear associations between encodings of entry order patterns and activity patterns recorded using ear tag accelerometers. In this paper we will explore how novel ensemble simulation techniques (Hsieh *et al.* 2021) that emulate and adjust for the complex sources of error in PLF data streams may be used to produced more balanced encodings of multi-dimensional behavioral data. We also introduce a new dendrogram pruning algorithm that is able to efficiently repurpose these same ensemble simulations to ensure that that the power of hierarchical clustering tools do not exceed the

resolution of the sensor. Finally, we demonstrate the utility of information decomposition techniques within our existing nonparametric mutual information testing framework to better facilitate visual characterization of complex behavioral patterns across sensor data sets that might be overlooked in more conventional model-based analyses.

Materials and Methods

Description of Data

To demonstrate the efficacy of our analytical approach, data was repurposed from a feed trial assessing the impact of an organic fat supplement on cow health and productivity through the first 150 days of lactation. All animal handling and experimental protocols were approved by the Colorado State University Institution of Animal Care and Use Committee (Protocol ID: 16-6704AA). The study ran from January through July in 2017 on a USDA Certified Organic dairy in Northern Colorado, enrolling a total of 200 cows over a 1.5 month period into a mixed-parity herd of animals with predominantly Holstein genetics. Cows were maintained in a closed herd in an open-sided free stall barn, stocked at roughly half capacity with respect to both feed bunk spaces and stalls. Cows had free access to an adjacent outdoor dry lot while in their home pen, and beginning in April were moved onto pasture at night to comply with Organic grazing standards. Cows were milked three times a day, with free access to TMR between milkings, and were head locked each morning to facilitate data collection and daily health checks. For more details on feed trial protocols see Manriquez *et al.* (2018) and Manriquez *et al.* (2019).

In addition to standard production and health assessments, behavioral data was also obtained from several PLF data streams (Hsieh *et al.* 2018). Milking order, or the sequence in which cows enter the parlor to be milked, is automatically recorded as metadata in all modern RFID-equipped milking systems. Study cows were here milked in a DelPro™ rotary parlor (DeLaval, Tumba, Sweden). At each morning milking, raw milking logs were exported from the parlor software, and the data processed to extract the single-file order that cows entered the rotary (Manriquez *et al.* 2019). A total of 80 milk order records – 26 recorded while cows remained overnight in a freestall barn, and 54 following the transition to overnight access to spring pasture – were used to create discrete encodings for parlor entry patterns via data mechanics clustering (see McVey *et al.* 2020 for further analytical details). The dendrograms summarizing the distribution of cow entry order patterns and subsequent heatmap visualizations will here be subjected to further analysis without modifications to the previously reported encodings.

Animals enrolled in this feed trial were also fitted with a CowManager ear tag accelerometer (Agis Automatisering BV, Harmelen, Netherlands). This commercial sensor platform, while designed and optimized for disease and heat detection, also provides hourly time budget estimates for total time (min) engaged in five mutually exclusive discrete behaviors - eating, rumination, non-activity, activity, and high activity (Bikkey *et al.* 2014; Pereira *et al.* 2018). Time budget data was collected on all animals for a contiguous period of 65 days (1560 hours), with the observation window beginning shortly after trial enrollment was completed on February 17th, and ending on April 23rd when the grazing season commenced and cows were

moved overnight beyond the range of the receiver antennae. After dropping cows that were removed prematurely from the observation herd due to acute clinical illness, as well as several cows with persistent receiver failure, complete sensor records were available for 179 animals. In order to more fully focus on the logistical challenges in encoding and characterizing the complex multivariate dynamics of this system, we have here chosen to compress this data over the time axis to consider only the overall time budgets of these cows, and will leave explorations of the longitudinal and cyclical complexity of this dataset for future work.

Improving Empirical Encodings of Overall Time Budget Through Simulation

Regardless of its original distribution, data can always be coarsened into a discrete variable (Agresti 2013). For complex or poorly defined systems, where appropriate cutoffs (binning rules) cannot be inferred *a priori*, an empirically-determined encoding may provide a more flexible and comprehensive approach to discretizing the underlying behavioral signals. One algorithm that provides a model-free approach to pattern encoding within the larger cannon of UML tools is hierarchical clustering. This approach employs a bottom-up agglomerative strategy to group observational units into discrete clusters of variable sizes, progressively building a coherent picture of herd-level global structures from the similarities in behavioral patterns observed between pairs of individuals (Hastie *et al.* 2009; Gareth *et al.* 2013). This series of progressive pairings can be expressed graphically in the form of a dendrogram, which serves as a 2D representation of the data's geometric distribution in its

higher dimensional measurement space, and can subsequently be used in data visualizations to highlight the most prominent structural features of a dataset (Hsieh *et al.* 2018).

The efficacy of any hierarchical clustering scheme, however, is largely contingent on the adequacy of the estimator used to quantitatively express the pairwise dissimilarity between observational units (Hastie *et al.* 2009). The Euclidean distance (L2 norm) is the default estimator used in most applications of this algorithm (Hastie *et al.* 2009; Gareth *et al.* 2013; Shirkhorshidi *et al.* 2015), including much of the previous work in precision livestock applications (Adamczyk *et al.* 2017; Schwager *et al.* 2007). The L2 norm is appropriate for many measurement systems where variance is reasonably uniform across a continuous domain of support. Time budget data, however, is distributed multinomially, and as such has significant domain constraints (Agresti 2013). Put more simply, we know that the minutes logged for each behavior must sum to an hour. So, if a cow has ruminated for 60 minutes, then there can be no uncertainty in the remaining axes, because we know these values must be zero. These domain constraints impose co-dependencies between the behavioral axes that become stronger as observations shift towards the boundaries of the distribution's support, which in turn warp the intrinsic variability of each axis contingent upon their location within the domain.

This statistical tedium also has some fairly intuitive behavioral implications. Suppose we have two cows, Betty and Bessy, who respectively spend 13 and 14 hours a day ruminating. How “different” are these values? Since both cows are exceeding rumination rates needed to sustain a healthy metabolism, we would not anticipate that this difference would have a significant biological impact on these animals, and may ultimately be explained by relatively

trivial behavioral fluctuations. Now suppose instead that we have two other cows, Daisy and Delilah, who respectively spend only 3 and 4 hours a day ruminating. Given that both these cows now well below the normal threshold for this behavior, this one-hour difference may have significant biological impacts. With a simple L2 norm, however, these two pairs would be given equivalent dissimilarity estimates for this behavioral axis, and so clearly a better estimator is needed.

Relative entropy, also referred to as the Kullback–Leibler divergence, is a classic information theoretic metric specifically designed to contrast discrete probability distributions, and thus a natural candidate for analysis of time budget data (MacKay 2003). For any two distributions that utilize the same alphabet of $k = 1 \dots K$ categorical features (i.e. use the same ethogram), relative entropy can be calculated using Equation 1, and converted to a symmetric distance measure using Equation 2. By utilizing the proportion of time that an animal invests in each behavior as both a nominal and relative value, this estimator is able to adjust the relative difference between cows by the absolute position of each observation relative to the boundary of the domain.

$$D_{KL}(P||Q) = \sum_k P(k) \log \frac{P(k)}{Q(k)} \quad (1)$$

P = normalized time budget vector for Cow A

Q = normalized time budget vector for Cow B

$$D_{KL}(P, Q) = D_{KL}(P||Q) + D_{KL}(Q||P) \quad (2)$$

Domain constraints are not, however, the only stochastic feature that need be accommodated when working with time budget data. There is also the measurement error attributable to the sensor itself. Returning to the previous example, suppose that we also know that our rumination records are only accurate to ± 1 hr. Is it then still appropriate to give more weight to the one-hour difference between Daisy and Delilah than between Betty and Betsy? Since both observations are within the bounds of error, attempting to enhance the underlying biological signal may instead only succeed in amplifying measurement noise. A closed form estimator, however, may not be readily generalizable to the wide range of measurement error models encountered with PLF sensors. We therefore propose that a simulation-based approach may offer a more flexible means of accounting for measurement error in dissimilarity estimates (Hsieh *et al.* 2021).

The LIT package provides a built-in simulation utility for time budget data that seeks to mimic the stochastic error structure of the original data while still preserving the underlying behavioral signal (Hsieh *et al.* 2021). Data is provided as a tensor, with cow indexed on the first axis, time indexed on the second, and the component behaviors on the final axis. The count data at each cow-by-time index is then used to redraw a simulated datapoint from one of three optional distributions (Agresti 2013). In the first, the user may sample directly from a multinomial distribution centered around the normalized observed count vector. This model assumes that measurement error should shrink as a cow dedicates larger proportions of an observation window to specific behaviors, and intrinsically prevents estimates from being

generated outside the domain of support. Variance can be under-estimated at the extremes of the domain, however, if the probability for a behavior is non-negligible but the observed count is zero due to under-sampling. This issue may be addressed in sampling option two, where samples are redrawn from a Multivariate Beta Distribution (MBD), also known as a Dirichlet distribution, again parameterized using the normalized observed count. While this sampling strategy slightly biases the simulation towards the center of the distribution, it prevents under-sampling at the extremes of the domain. Finally, users may combine these sampling strategies in sampling option three, wherein the probability vector used to parameterize the multinomial is drawn first from the Dirichlet, in order to further increase the uncertainty in the simulated data. After simulation has been completed by redrawing samples at the finest level of temporal granularity supported by the sensor, the data can then be conditionally or fully aggregated along temporal axis as required for downstream analysis as a time budget.

This simulation routine was used to create an ensemble of $B = 500$ simulated overall time budget matrices that mimicked the stochasticity attributable to a reasonable approximation of the measurement error of the sensor. Stored as a tensor with replication on the last axis, the variance of the ensemble of simulations could then be easily calculated for each combination of cow index and behavioral axis. If the underlying simulation strategy is a reasonable representation of the noise in the sensor, then these variance terms will then serve as a sufficient approximation of the relative uncertainty in each data point. We propose that that this information can then be incorporated into the calculation of dissimilarity estimates by serving

as penalty terms in the calculation of an ensemble-weighted distance estimator defined in Equation 3.

$$D_{EW}(P, Q) = \sum_k \frac{(P_{obs}(k) - Q_{obs}(k))^2}{\sigma_{P^*(k)}^2 + \sigma_{Q^*(k)}^2} \quad (3)$$

$\sigma_{P^*(k)}^2 = \text{Variance of ensemble of simulated values for Cow A for behavior } k$

$\sigma_{Q^*(k)}^2 = \text{Variance of ensemble of simulated values for Cow Q for behavior } k$

The rescaling strategy employed in our proposed dissimilarity estimator is strongly inspired by traditional Analysis of Variance (ANOVA) techniques, thereby providing several insights into its anticipated behavior. First, because the simulations were generated using the multinomial or one of its analogs, we can infer that these penalty terms will not be homogenous across the domain of support, but should shrink as observations approach the boundary. This will allow the ensemble weighted distance estimator to emulate the rescaling dynamic achieved with the KL distance, but here rescaling at the extremes of the domain will ultimately be bounded by our simulated measurement error so as not to exceed the precision of the sensor. Second, because we have here emulated measurement error in our simulation using sampling uncertainty, the central limit theorem will apply (Gareth *et al.* 2013). Thus, we can anticipate that as the number of observations per animals increases, the impact of measurement error on our inferences will shrink, allowing progressively more subtle differences between animals to

come into resolution. Taking this property to its limit, however, can it be said that with enough observation minutes the differences between cows can be inferred with near certainty? That intuition, of course, is at odds with our characterization of a dairy herd as a complex system, and highlights an additional stochastic element that must be accommodated – the behavioral plasticity of the cows themselves in response to changes in the production environment (Tucker *et al.* 2021).

Given the extended observation window of this particular data set, it would be possible to recalculate time budget conditional on day of observation, and then use the variance in daily time budget along each behavioral axis as a penalty term. Such estimates would collectively reflect heterogeneity in variance attributed domain constraints, measurement error, and behavioral plasticity. Such an approach would not, however, be feasible for datasets collected over shorter time intervals with fewer replications or in applications with behavioral responses where there is no clear hierarchy in the temporal structure of the same. We therefore propose that our stochastic simulation model can be extended to also provide a generalizable means to approximate the uncertainty of the underlying behavioral signal.

As before, the measurement error was simulated by redrawing samples at the finest temporal granularity provided by the sensor. Prior to compression along the temporal axis, however, a random subsample of observations days was selected across all cows, and only these values used to calculate overall time budget. If all cows demonstrated comparable levels of consistency in their daily time budgets, then reducing the effective sample size of our simulated data sets through a subsampling routine would increase the ensemble variance

estimates. This in turn would make our approximation of measurement error hyper-conservative, but this increase would be uniform across all cows. If instead some cows were less consistent in their time budgets across days, then the sampling error imposed by the subsampling routine would be greater, resulting in a larger ensemble variance estimate. Thus, we would expect a stronger penalty to be applied to cows whose demonstrated greater plasticity in their behavioral response to both transient and persistent changes in the production environment. For small datasets with limited number of replications, the number of subsamples could be set quite close to the size of the complete sample, and would thus emulate a jackknife approach to variance estimation (Efron 1993; Hastie *et al.* 2009; Gareth *et al.* 2013). For larger datasets, however, the subsample size could be set smaller to make the resulting ensemble variance estimates progressively more sensitive to the uncertainty in the underlying behavioral signal.

To evaluate the empirical performance of these dissimilarity estimators, distance matrices were calculated for the 177 cows with complete CowManager time budget records. Euclidean distance and KL Distance were calculated using base R utilities, with speed up options utilizing the *Rfast* package (R Core Team 2018; Pepadakis *et al.* 2020). An ensemble-weighted dissimilarity matrix was first calculated using simulated values accounting only for measurement error using the most conservative joint Dirichlet-Multinomial sampling scheme, hereafter referred to as Noise-Penalized Distance. A second ensemble-weighted dissimilarity matrix was then calculated using the same sampling scheme for measurement noise but aggregated over a 14-day subsample to account for behavioral plasticity in daily time budgets,

hereafter referred to as Plasticity-Penalized Distance. The LIT package provides users a clustering visualization utility, which converts dissimilarity matrices into a dendrogram using the *hclust* utility in base R with default Ward D2 linkage (R Core Team 2018) and the generates heatmap visualizations of the resulting clustering results using the *pheatmap* package (Kolde 2019). Heatmaps were generated on a grid of cluster values from $k = 1 \dots 10$ for each of the four dissimilarity estimators, with complete results provided in Supplemental materials, and the results for $k = 10$ clusters compared in Figure 2. The LIT package also provides users a plotting utility to visually contrast the broader patterns between behavioral encodings. Outputs from the clustering utility are passed in to create a contingency matrix generated using *ggplot2* with cells colored by their corresponding cell count (Wickham 2016). The heatmap visualizations for each encoding are then added to the row and column margins of the contingency matrix using the *ggpubr* package (Kassambara 2020), and arranged such that each row cluster in either heatmap matches the order of the contingency matrix reading either up-down or left-to-right, allowing for direct and detailed visual comparison of the discretized behavioral patterns. Comparisons between the Noise-Penalized and Plasticity-Penalized encodings are provided in Figure 3.

Improving Tree Pruning Decisions Through Simulation

An optimal encoding strategy seeks to minimize the loss of relevant information by retaining as much of the underlying deterministic signal as possible while hemorrhaging only noise (Agresti 2013). In a hierarchical clustering framework, this is achieved by pruning the

dendrogram built from the dissimilarity matrix where the branches cease to represent differences in the underlying signal. Standard pruning strategies typically either 1) to allow the user to provide a dissimilarity cutoff, below which value all further branches are grouped into the same bin, or 2) allow users to extract the first K branches of the tree (Hastie *et al.* 2009; Gareth *et al.* 2013). As with the default Euclidean distance dissimilarity estimator, this approach may be appropriate for datasets with relatively homogenous variance structures. For data drawn from intrinsically heterogenous distributions, however, the branch lengths cannot be directly compared across the domain of support, making globally-defined pruning rules a suboptimal strategy for analysis of time budget data.

More fundamentally, a homogenous pruning strategy may be too simplistic for many PLF sensor datasets, for which the underlying signal often represents a complex composite of behavioral mechanisms that operate at multiple scales. While some environmental factors might be expected to have an impact on cattle behaviors that is fairly uniform across the herd, other factors might elicit responses that differ in magnitude for different subgroups within the larger populations, or even become isolated within smaller social cliques. For example, we might expect the number of times cows are moved each day for milking will place similar constraints on the time left to lie down across all animals, but overstocking with respect to stall spaces might have a much larger magnitude of impact on the lying patterns of subordinate heifers than the more dominant older cows (Tucker *et al.* 2021). In such a complex system, we would expect the heterogeneity imposed by the underlying biological signal to differ in scale across the dataset. Subsequently, in attempting to employ a global cutoff decision to encode

information for such a dataset, we would always be faced with the difficult decision to either ignore the subtler behavioral patterns present in some branches of the tree, or else allow noise to contaminate our encoding of other branches with intrinsically coarser behavioral patterns.

While all the components that contribute to the signal in a complex livestock system might be difficult to anticipate *a priori*, we propose that a more dynamic pruning algorithm might still be achieved by again employing flexible simulation-based approaches to emulate the comparably simpler sources of uncertainty. If each branch of the dendrogram is viewed as a pairwise contrast between two groups of animals, then we need only to determine whether the bifurcation under inspection represents a difference in the underlying signal can be reliably distinguished from noise. If it can, then the two groups should be split in the final encoding to capture this feature of the data's distribution. If a branch falls below the intrinsic resolution of the data, however, then the branch may be pruned so that all animals are placed into the same cluster with no loss of meaningful information. By implementing such a branch-level test recursively, we can gradually work our way down the tree with locally-defined pruning decisions.

To evaluate the reliability of the behavioral signal encoded at each bifurcation of the tree (Rand 1971), our branch test utility utilizes two mimicries. The first set of simulations are generated under the alternative hypothesis that assumes a branch contains an underlying deterministic signal that is only partially obscured by stochastic noise. Here then we can simply repurpose the ensemble of simulated data sets used previously to calculate the ensemble weighted dissimilarity metrics by mimicking the uncertainty in the observed data. The second

set of simulations are generated under the null hypothesis that a given branch contains only noise. As the null implies that animals demonstrate equivalent patterns of behavior within the resolution of the sample, this mimicry can be generated quite efficiently using a standard bootstrapping routine (Efron 1993), wherein time budgets simulated under the alternative are unconditionally resampled from amongst all animals in a given branch. HClustering is then performed independently on each data mimicry in either ensemble and the first k branches are extracted to create an ensemble of discrete encodings.

Under the alternative hypothesis, a strong signal should produce a robust tree structure such that, even after the addition of simulated noise, the resulting encoding would still closely mirror that of the original observed data. As the stochastic component of a dataset becomes stronger relative to the signal, these bifurcation points will become progressively less stable and the subsequent encodings less reliably aligned with the original data. When the signal falls below the resolution of the data, the tree structures of the simulated data would then seldom match that of the original data, and so would become poorly distinguished from encodings generated under the null with no signal component. We propose that mutual information, which can be calculated without any additional distributional assumptions, can be used to quantify the similarity between of the observed data and each mimicked dataset, and subsequently used to determine if simulations under the alternative are distinguishable (MacKay 2003). A bifurcation was here determined to be significant if less than 5% of the MI values calculated for data sets simulated under the alternative hypothesis fell below the 95th quantile of MI values calculated for data simulated under the null. If a bifurcation was instead deemed insignificant,

the branch would be pruned and all cows within it assigned to the same cluster in subsequent encodings.

In evaluating the significance of a bifurcation, it seems intuitive that a $k = 2$ binary encoding should be utilized. For complex systems subject to the influence of multiple competing drivers of behavioral responses, a false negative result can occur with this parameterization if the addition of stochastic noise perturbs the order in which two significant mechanisms with similar magnitudes of impact are bifurcated. Such trivial destabilizations of the tree structures can be readily identified in visualizations of the distributions of MI values calculated against simulations under the alternative, as the “flip flopping” between bifurcation points produces clear evidence of multimodality (see Figure 1). To circumvent this issue, the LIT package provides users the option to re-test any bifurcations deemed insignificant using a binary encoding with a more granular discretization ($k > 2$). This effectively allows the algorithm to “look down the branch” to absorb any irrelevant flip-flopping between competing signals, thereby preventing spurious over-pruning that would hemorrhage information on significant behavioral patterns from the final encoding.

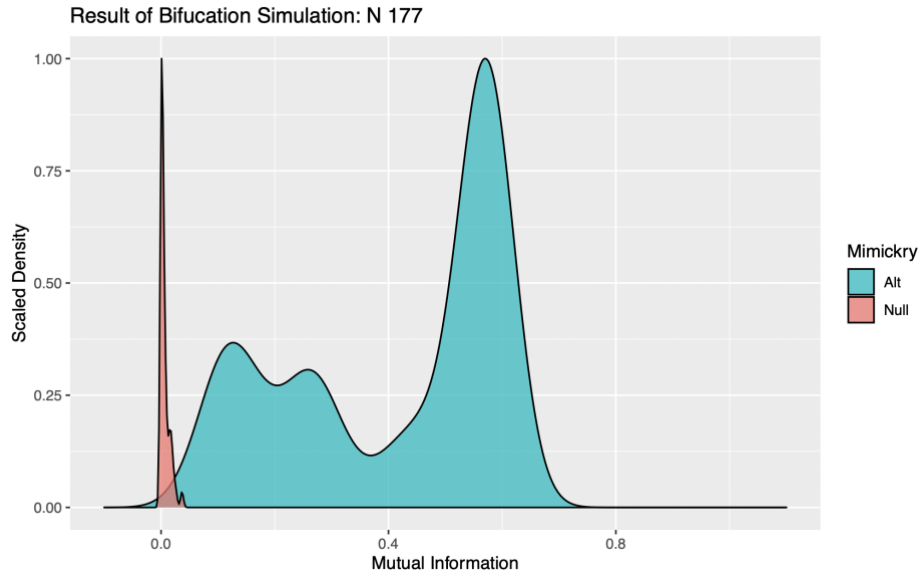


Figure 1: Visualization of testBranch results for the first bifurcation of the Euclidean distance time budget dendrogram cut using the noise-penalized ensemble of data mimickries. In simulations under the alternative hypothesis, the addition of noise intended to mimic measurement error has destabilized the tree, causing it to “flip flop” between first isolating cows with more moderate time budgets from animals at the two extremes of the tradeoff between eating and ruminating. While both branches are distinguishable from measurement error, this ambiguity in bifurcation order has produced bimodality in the distribution of mutual information estimates against the encoding for the observed data. Retesting with more clusters allows the algorithm to “look down the branch” to produce better separation between encodings under the null and the alternative and thereby avoid spurious over-pruning.

Full results for the application of our ensemble cut algorithm to dendrograms generated using each of the four dissimilarity estimators discussed in the previous section, using both the noise-penalized and plasticity-penalized encodings, are provided in Supplemental Materials.

Results for application of the ensemble cut algorithm applied dendrograms generated from the noise- and plasticity-penalized ensemble weighted distance metrics, the noise-penalized and plasticity-penalized encodings respectively, are summarized in Figure 4.

An Information Theoretic Framework for Cross-Sensor Inferences

Equipped with an appropriate encoding to discretely represent the heterogeneity in overall time budgets within this herd, and provided the encoding of longitudinal patterns in parlor entry position from previous work with this data set, a potential question to ask would be: How does a cow's time budget, which is largely determined by her behaviors in the home pen, relate to her behavior in the milking queue? There are a number of nonparametric and parametric techniques available to evaluate the overall strength of association between two discrete variables by evaluating the distribution of animals in the joint encoding (Agresti 2013). There is, however, perhaps greater practical utility in characterizing low and high points within the joint encodings, which would provide more detailed insights into the tradeoffs between specific behavioral patterns recovered from the data streams in these distinct farm contexts. Towards this end, information theory offers a more comprehensive approach to decomposing the stochasticity within discretely encoded variables, and thus may provide a more holistic approach to evaluating both the global and local features of a joint encoding, while employing few structural assumptions (MacKay 2003).

First, to evaluate the strength of overall relationship between two discretized behavioral responses, the LIT package provides users a permutation-based bivariate testing utility that

uses the mutual information estimator to quantify the amount information entropy that is redundant between the two encodings (MacKay 2003; McVey *et al.* 2020). We can anticipate, however, that the efficacy of this test in recovering significant relationships between the underlying biological signals will be affected by the resolutions of the encodings. Suppose that a single latent biological factor impacts the behavioral responses collected by both PLF data streams, creating informational redundancy between the two encodings. If we cut the trees above the intrinsic magnitude of its impact on a given behavior, its influence may be overlooked and mutual information under-estimated. On the other hand, if we prune the tree far below the magnitude of its impact, our inferences can lose power as bins sizes in the joint encoding become progressively smaller, weakening the empirical estimation of the joint probability distribution and thereby increasing estimation error in the MI estimator. The resolution of our encodings must, therefore, be optimized to match the dynamics of the system, or a false negative result may be returned. To further complicate matters, however, we cannot necessarily assume that the magnitude of impact of a given latent factor will be uniform across behaviors, nor should we expect in a complex farm environment that behaviors will be influenced by a single latent factor.

To overcome this logistical challenge without falling back on dubious *a priori* assumptions, the LIT package implements mutual information-based permutation tests on a grid, varying the cluster resolutions across both behavioral axes (McVey *et al.* 2020). Under the null hypothesis that no significant bivariate relationship exists between data streams, cow ID labels are randomly permuted within each tree, preserving the marginal distribution of the

data along each axis but destroying any latent bivariate relationships. These permuted trees are then cut and the mutual information of the joint encoding estimated for each combination of cluster counts on the grid. A p-value is then generated by comparing the observed MI value of the joint encoding at each grid point against the corresponding distribution of MI values simulated under the null. Just as a scientist varies the focus of a microscope to bring microbes of different size into resolution, we can expect that geometric features of the joint probability distribution imposed by latent deterministic variables that vary in scale of impact will come into and fall out resolution as these meta-parameters are varied across the grid of cluster counts. To help the user visually identify where such features have come into resolution, the LIT package also returns a heatmap visualization of the observed MI value for each grid point that is centered and scaled relative to the distribution of MI values under the null. For behavioral measurements subject to the influence of multiple biological and environmental factors operating simultaneously, this exhaustive approach to parameterization enables users not only to build a more complete picture of a complex behavioral system, but may also provide insights into the hierarchy of these behavioral responses.

Unfortunately, as the resolution of the encodings is increased, MI estimates not only become less precise, but they may also become less accurate. Bias is introduced when empirical estimates of the joint probability distribution become so granular (i.e. – a high number of bins relative to the total sample size) that regions with low but nonzero probabilities go unsampled. These zero-count bins cause the total entropy calculated from the empirical joint probability distribution to be under-estimated, which in turn causes the relative amount of redundant

information to be over-estimated. While the magnitude of this bias is partially dependent on the total sample size, it is also contingent on the structure of the joint probability distribution itself, namely the number of low-probability cells. Given that the joint probability distribution under the null, which is randomly permuted to intentionally remove any nonrandom features in the sample, can be expected to have a more uniform distribution of probability than the observed dataset, we can anticipate that the magnitude of the bias may differ between these two distributions as the sample becomes more granular, preventing MI estimates from being directly comparable. To overcome this issue, the LIT package by default provides entropy estimators based on the ML frequency estimates, but allows users to select from a range of bias-corrected frequency estimates available in the *entropy* package (Hausser and Strimmer 2021). Based on the simulation work by Hausser and Strimmer (2009) the JS “shrink” estimator was here used to conduct bias-corrected mutual information permutation tests (Hausser and Strimmer 2009).

Not only can the impact of latent factors on either behavioral measure differ in magnitude, but we can also anticipate that responses may differ in both strength and direction for different subgroups within the herd. Such nonlinear dynamics are easily captured in a model-free MI test, but further inspection of the contingency table is needed to fully characterize such complex bivariate relationships between sensor outputs. If either marginal encoding has roughly the same number of observations in each bin, then the cell counts in the joint contingency table can be directly compared, as under the null we would expect each cell to be equiprobable. For empirically defined encodings, however, bin sizes can vary significantly to better capture the

underlying geometry of the univariate data distribution. Such differences in marginal probabilities prevent the raw cell counts from being directly compared. To better identify which cells in an empirically-defined joint encoding are driving a significant overall relationship between two data streams, mutual information can be decomposed into pointwise mutual information (PMI) values (Johnson 1999). The LIT package provides users the option in the *compareEncodings* plotting utility to color cells in the joint contingency table by PMI estimate, to better facilitate direct visual comparisons of the encodings. To further enhance visualizations of the joint probability distribution that significantly differ from expected cell counts under the null, users may also specify a probability threshold above which PMI values should not be displayed, which is determined here by simulating PMI estimates under the null by redrawing from a multinomial distribution using the outer product of the marginal distributions.

Bivariate tree tests were applied to the time budget encodings using both the noise- and plasticity-penalized dissimilarity metrics, and pruned using the more conservative plasticity-penalized mimicry, and the encoding of parlor entry order data produced using data mechanics clustering in previous work (McVey *et al.* 2020). A 2:10 x 2:10 grid was used to determine the optimal resolution for the bivariate relationship, with the optimal metaparameters used to create visualizations of the joint encoding, wherein pointwise mutual information values were used to color cell counts that were significant at the $\alpha = 0.05$ significance level. To further explore latent factors that might explain significant associations between entry position and time budgets, bivariate tree tests and pointwise mutual information tests were also applied separately to the encodings of both PLF data streams and health records.

Results & Discussion

Improving Empirical Encodings of Overall Time Budget Through Simulation

Figure 2 provides a visual comparison of time budget encodings for the four candidate dissimilarity metrics. In each heatmap visualization, individual cows are arranged along the row axis and the mutually exclusive behaviors that comprise the overall time budget are ordered along the columns. Each cell within the heatmap is subsequently colored to reflect the proportion of time that a given cow is recorded by the accelerometer system to engage in a specific behavior over the observation window. Few cows dedicated more than half of their time to any one behavioral axis, which is not surprising, given that total lying time in this system is split between the nonactive and rumination axes (Tucker *et al.* 2021). Time recorded as eating and time recorded as ruminating were the highest magnitude behavioral axes, but time spent eating demonstrated far greater range and heterogeneity. Time spent nonactive was lower in overall magnitude but still showed a fair amount of heterogeneity across cows. The active and highly active axes, however, were both quite low in magnitude and generally demonstrated less systematic heterogeneity across the herd. The order of cows along the row axis in each heatmap is determined by the dendrogram calculated for each dissimilarity matrix. The dendrogram can be interpreted as an approximate 2D representation of the distribution of the cows with the 5D multinomial space of the time budget, and thus serves to bring out in the heatmap systematic differences in time budget across the herd. Gaps have here been added between rows to indicate branches that have been pruned, such that all cows within a given branch receive the same discrete value in the final time budget encoding.

A cursory appraisal of all four encodings summarized in Figure 2 reveals that, regardless of the dissimilarity metric utilized, there is a considerable amount of heterogeneity in the distribution of overall time budgets across this herd. Looking more closely at the clustering tree produced from the unweighted Euclidean dissimilarity metric in Figure 2A, we see that the higher magnitude eating and rumination axis entirely dominated the first handful of bifurcations of the dendrogram. Even for users not accustomed to reading dendrograms, this dynamic is clearly animated by parsing through the grid of heatmap visualizations provided by the *encodePlot* utility (see Supplement). Heterogeneity in the moderate-magnitude nonactivity appears to have been largely ignored in the first half-dozen bifurcations, with the first 10 clusters extracted from this dendrogram being ultimately quite variable in the nonactivity response. Nor is there clear evidence that either activity axes influenced the first 10 bifurcations of this tree. This dynamic is almost certainly attributable to the lack of intrinsic scaling with this estimator. While a behavioral axis that represents a larger proportion of a cow's time investments may warrant additional consideration, these results clearly demonstrate that the Euclidean norm does so to nearly the complete exclusion of lower-magnitude behavioral axes that might still convey important ethological information. The Euclidean distance heatmap is also annotated on the row axis with a number of auxiliary data fields for each cow, which included: age (birth date), calving date, an estimate of peak lactation, nutrition supplementation treatment, and health status during the observation window (see Supplemental Materials for details on the encoding of these auxiliary cow attribute variables). A cursory visual inspection reveals that most clusters appear to be fairly homogenous with respect to cow age, tenure in

the pen, and feed supplementation status. Sick cows, however, appear to be slightly over-represented in some groups, namely the smaller branches representing the more extreme time budget tradeoffs.

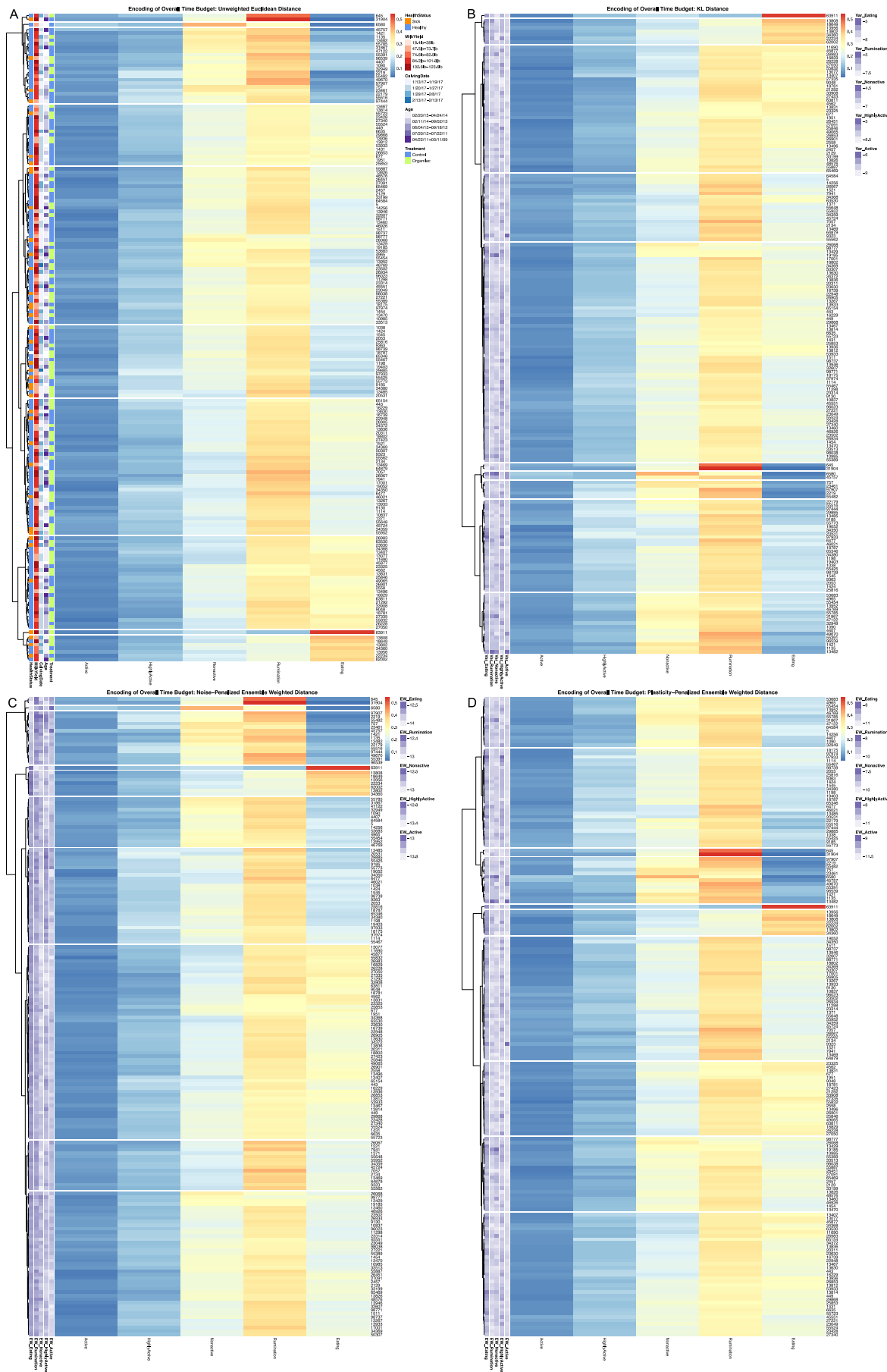


Figure 2: Comparison of overall time budget encodings derived from different dissimilarity metrics. In each heatmap cows are arranged along the row axis, and the mutually exclusive behaviors along the column axis, such that each cell is colored to represent the proportion of time that a given cow is recorded engaging in a specific behavior. Row gaps have been added within each heatmap to reflect the first 10 branches of the corresponding dendrogram, which here are numerically indexed reading from top to bottom A) Euclidean norm encoding with row annotations representing cow-level attributes. B) KL Divergence encoding with row annotations representing log-scaled variance in observed daily time budgets. C) Noise-penalized ensemble-weighted Euclidean distance encoding with row annotations representing the log-scaled ensemble variances. D) Plasticity-penalized ensemble-weighted Euclidean distance encoding with row annotations representing the log-scaled ensemble variances. See Supplemental Materials for full-scale versions of these images.

Looking next at the hierarchical clustering results visualized in Figure 2b, the KL distance seems to have provided a slightly more holistic encoding of the data that better balances the input across the five behavioral axes. Again, extremes in eating and rumination drive the first few bifurcations of the tree structure, but tradeoffs between time spent eating and nonactive are considered much earlier in the bifurcation decisions within this tree. Some systematic heterogeneity is also revealed across the herd in high activity axis, despite its lower magnitude. Unfortunately, the KL distance also appears to have over-stratified cows whose time budgets lie at the extremes. In particular, the cows with extremely low time spent eating

(clusters 6-8) have been divided into clusters that are likely too small and narrowly defined to facilitate cross-sensor inferences in downstream analysis, and thus may obscure important behavioral dynamics in this dataset. The KL distance heatmap is also annotated on the row axis with the variance in observed daily time budgets for each behavioral axis. Given that time budgets have here been normalized and expressed as proportions, the resulting variance terms were quite small in magnitude (less than zero), and so have here been re-expressed on log scale, where a more negative value represents a smaller relative magnitude of variation. That all five axes ranged several orders of magnitude in these variance estimates reveals that there was a significant amount of variability in the relative plasticity of time budgets across days. Visual appraisal revealed very little systematic patterns in this heteroskedasticity across clusters, however, suggesting that differences in relative plasticity in daily time budget observations may be attributed more to the individual than to any specific pattern in overall time budget.

The noise-penalized ensemble weighted distance, visualized in Figure 2C, displays clustering dynamics that fall somewhere in between the two extremes of Figure 2A and 2B. Time spent eating and ruminating still dominate bifurcations nearer the root of the tree, as with the unweighted Euclidean distance, pulling off the most extreme tradeoffs between these axes without over-cutting the tree as with the KL distance. In the later branches of the tree, however, cows with more moderate time budgets are divided with greater input from the nonactive and highly active axes. While the ensemble-rescaled estimator does appear to have succeeded in curbing the rescaling of dissimilarity estimates at the extremes of the distribution, the noise penalized ensemble distance did still bifurcate several cows with anomalously high values in

the eating, rumination, and nonactive axes into their own clusters of size $n = 1$. While isolating these animals into their own branches will effectively exclude them from cross-sensor inferences in downstream analysis, this encoding may still be appropriate if these datapoints represent authentic outliers that cannot be explained by typical variation in the sensor system. The heatmap was also annotated on the row axis with the ensemble variance terms used to penalize the squared distance estimates. We see that, as anticipated, the magnitude of error in the noise-penalized ensemble variance terms is substantially smaller than the observed variance in observed daily time budgets, confirming that, with so many samples over an extended observation window, measurement error is not contributing substantially to the overall uncertainty in observed time budgets. Closer appraisal of the clear systematic differences in these ensemble variance terms observed across clusters, however, confirms that these penalty terms appear to be effectively mimicking the intrinsic heteroskedasticity in this multinomial sampling space.

Ensemble variances calculated for each cow via the plasticity-penalized simulation routine closely matched ($R \geq 0.99$) the variances in observed daily time budget estimates for all five time budget axes, thereby validating the efficacy of the jackknifing routine. Figure 3 directly contrasts the first 10 clusters extracted from the dendrograms generated by the noise- and plasticity-penalized ensemble weighted distance measures. In this visualization, clusters are numbered in each heatmap from top to bottom, and so directly align with the row and column indices of the contingency table. For example, we can easily confirm from this graphic that the first three cows constituting the first two clusters in the Noise-Penalized heatmap are

the same cows isolated into the third and fourth clusters in the Plasticity-Penalized heatmaps – a determination that can be easily confirmed by zooming into this high-definition rendering to compare Cow ID values. Further comparisons reveals that cluster designations for cows with extremely high time spent eating, extremely low time spent ruminating, and relatively low time spent nonactive (clusters 5 and 6 in the noise- and plasticity-penalized encodings respectively) are virtually identical. In the plasticity-penalized dendrogram, the extremely low eating time cluster (cluster 3) is shrunk by just a few animals compared to the noise-penalized encoding (cluster 4). Additionally, after penalizing for behavioral consistency, the cow with the highest time spent nonactive in the sample (cow 6580) was not isolated as a outlier. This bifurcation was instead shifted to the cows with more moderate time budgets (Kirby 2001; Hastie *et al.* 2009; Gareth *et al.* 2013; McVey *et al.* 2020), serving to better distinguish between cows with relatively high and only moderate times spent eating. The plasticity-penalized dissimilarity estimator was also notably more generous in assigning cows to the cluster characterized by slightly higher rates of rumination while all other axes remained relatively low (cluster 7), and appeared to place greater emphasis on the nonactive axis to determine the remaining clusters. In spite of these differences, both ensemble weighted dissimilarity metrics have succeeded in producing encodings that provide a more holistic and balanced description of this dataset, and ultimately serve to better visualize heterogeneity in the tradeoffs between all five behavioral axes.

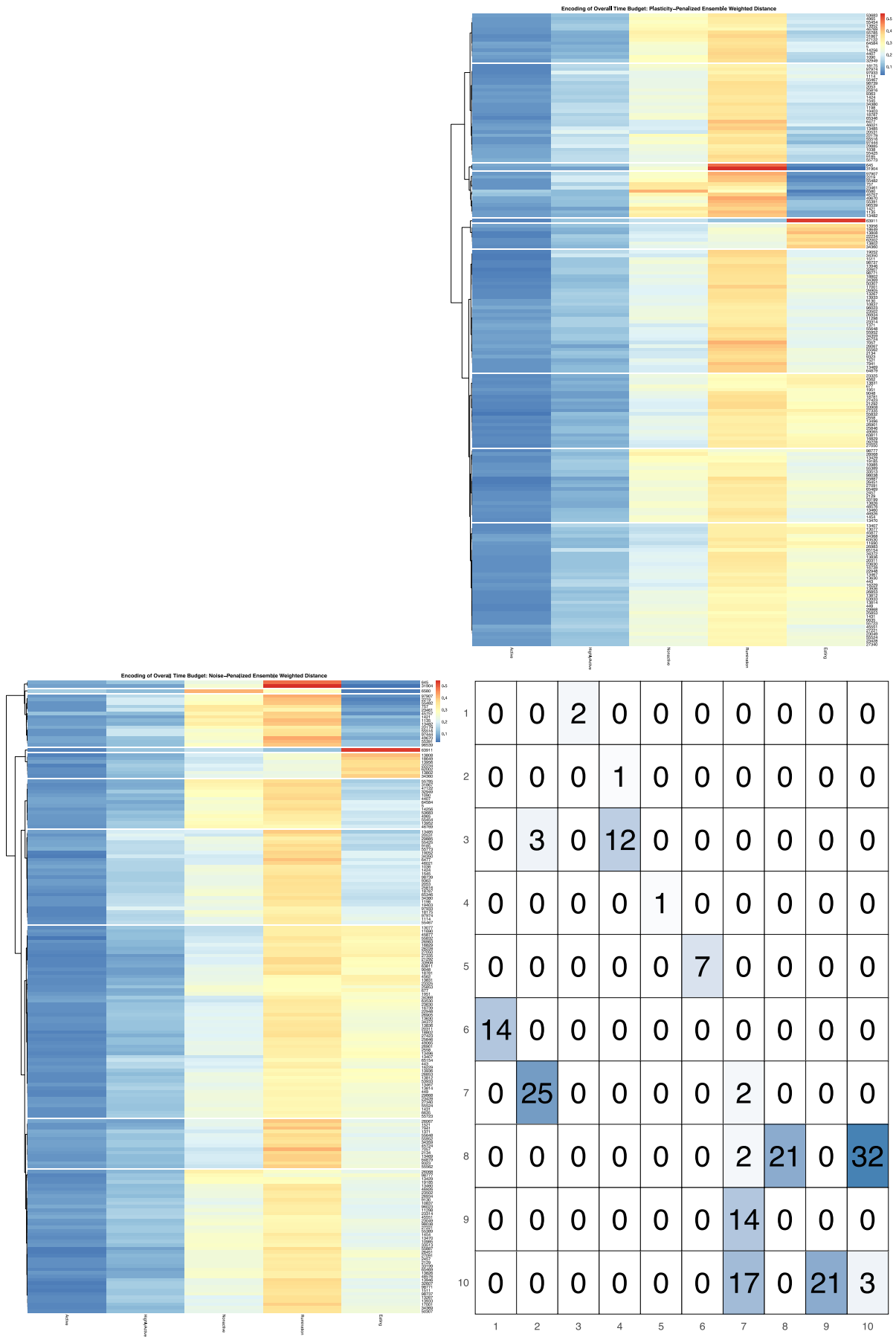


Figure 3: Visualization produced using the compareEncoding utility. The noise-penalized encoding is represented on the row axis and the plasticity-penalized encoding is represented on the column axis. Clusters in either heatmap are numbered from top to bottom, and so align directly with the corresponding row and column margins of the contingency table reading up-down and left-right respectively. Cell counts show that these two encodings are quite similar at the extremes of the time budget distribution, but differ slightly in cutoffs amongst the more moderate time budget clusters. See Supplemental Materials for larger versions of these images.

Improving Tree Pruning Decisions Through Simulation

For all dendrograms pruned using the ensemble of simulations that accounted only for measurement noise, an extremely fine-grained encoding was returned. A total of 39 clusters were returned for the unweighted Euclidean distance, 31 for the KL distance, and 38 clusters for the noise-penalized dissimilarity metric. In Figure 3, the heatmap visualization of the noise-penalized encodings helps to illustrate just how far down each branch the pruning algorithm was able to penetrate before the signal was lost to simulated measurement error. In fact, amongst the first dozen bifurcations in this dendrogram, the only branch not validated was that which would have isolated the cow with the highest observed time spent eating (cow 63911) into her own branch. This result is not necessarily surprising, given the extended observational period over which sensor records were recorded. With over 1500 minutes of observation for each cow, even in using a relatively conservative simulation strategy that very likely over-estimated the noise intrinsic to this sensor, we should expect by the CLT that the standard error

attributable to measurement error would ultimately be quite small after averaging over so many sampled timepoints. Subsequently, these results reinforce that the sensors themselves should impose few limitations on downstream inferences for this dataset, and that inconsistencies in the environment and the animals themselves should be the true limiting factor for the resolution of this encoding.

As expected, the dendrograms pruned with the ensemble of simulations that accounted for both measurement error and longitudinal consistency of the underlying behavioral pattern produced encodings that were far more granular. A total of 13 clusters were returned for the unweighted Euclidean metric, 17 for KL distance, and 14 for both the noise-penalized and the plasticity-penalized dissimilarity metrics. In Figure 5B, the heatmap visualization of these pruning results for the plasticity-penalized dissimilarity metric reveal an encoding that is coarser but ultimately quite well balanced, with the pruning heights modulated to produce cluster sizes that were reasonably uniform across the domain of support. Closer inspection revealed that this final encoding largely matched the order of bifurcations in the original tree, except that this pruning strategy left no animals isolated in anomalous clusters. It should be noted, however, that the granularity of this encoding is not entirely intrinsic to this system, but was dependent on the size of the subsample used to calculate the overall time budget in each simulation. While we can expect cows that were more inconsistent in their daily time budgets to be subjected to a stronger penalty with this estimator due to relatively higher rates of sampling error imposed by the subsampling routine, we can also anticipate that the overall scale of the sampling error imposed on all cows should grow as the size of the subsample is

reduced. This would in turn modulate how quickly the underlying behavioral signals would be drowned out by simulated noise within the tree. This suggest that, for larger samples where a greater range of subsample sizes can be utilized, this simulation value can also be treated as a meta-parameter to tune the granularity of the final encoding. Given that the plasticity-penalized mimicry was created for this data set by subsampling only 14 out of 65 observation days, the resolution achieved in the pruned encodings for all four dissimilarity metrics reinforces that this herd was overall fairly consistent in their daily time budgets, and that this data set will support fairly detailed inferences against a strong underlying behavioral pattern.

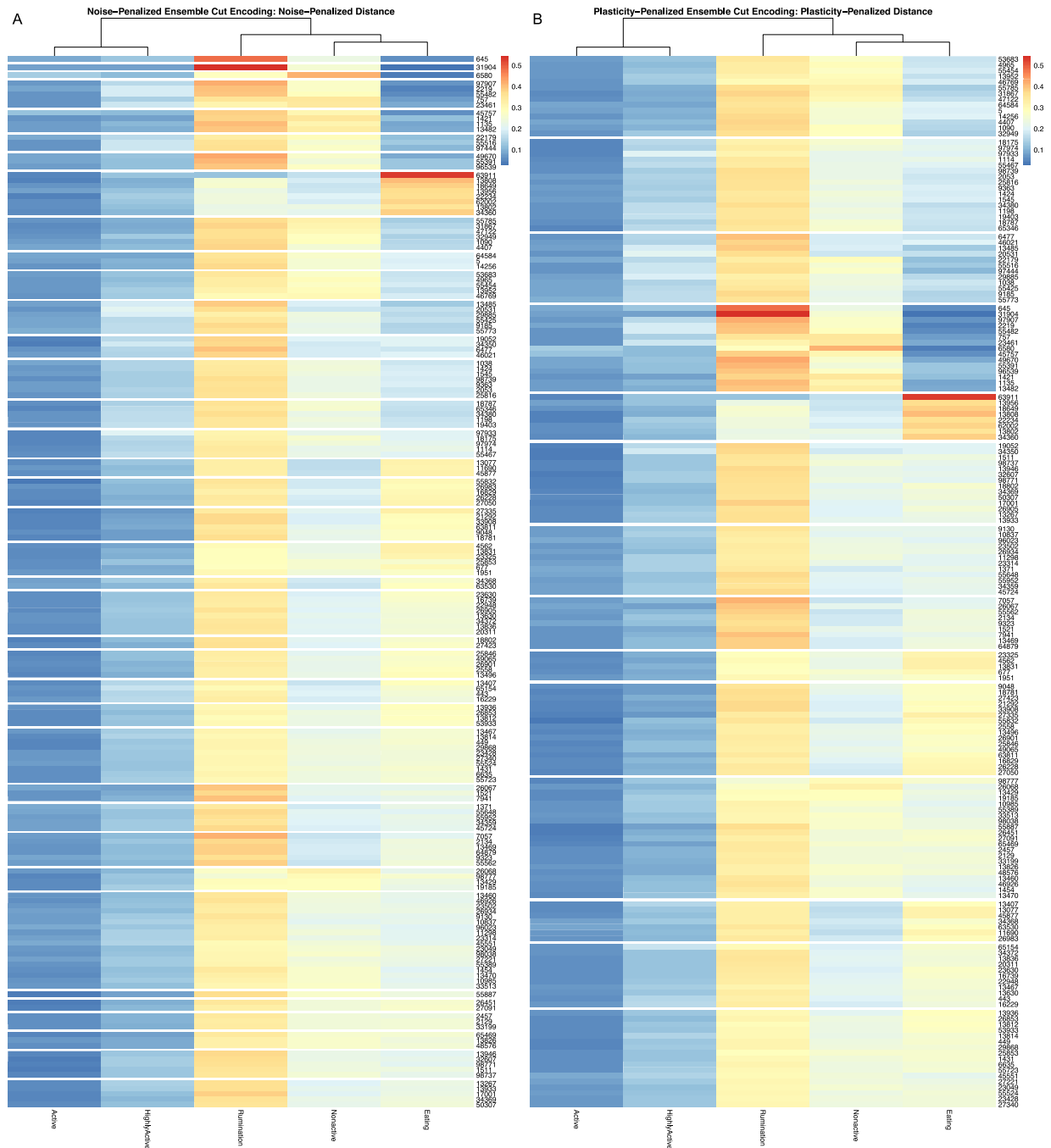


Figure 4: Encodings produced by the cuttreeEnsemble algorithm. A) Dendrogram produced by the noise-penalized ensemble weighted dissimilarity metric cut using the noise-penalized data mimicry. The extremely fine encoding with 38 stochastically validated clusters demonstrates that, with so many recorded observations over this extended observation window, the accuracy of the sensor itself should impose few constraints on our behavioral inferences. B) Dendrogram

produced by the plasticity-penalized ensemble weighted dissimilarity metric cut using the plasticity-penalized data mimicry. A courser encoding is returned when uncertainty in time budget observations attributable to the behavioral plasticity of the animal itself is taken into consideration

An Information Theoretic Framework for Cross-Sensor Inferences

Encodings of the overall time budgets produced using both the noise and plasticity-penalized dissimilarity estimators, wherein both were pruned using the more conservative plasticity-penalized ensemble, produced similar behavioral insights when compared against longitudinal patterns in parlor entry position across the herd. For the bivariate analyses run with encodings for all 177 cows with complete records, highly significant associations with entry order were recovered for both the noise-penalized ($p = 0.006$) and plasticity-penalized ($p = 0.005$) time budget encodings. The bivariate relationship was optimized for both time budget encodings with a five-cluster encodings of entry order patterns. The noise-penalized encoding produced the strongest associations with entry order with seven time budget clusters, whereas the plasticity-penalized encoding performed better with a finer encoding of nine clusters, the key difference being the degree of stratification among animals with the most moderate time budgets.

Visualization of the contingency tables for the optimized encodings colored by their PMI estimates revealed that the significant overall association between the two data streams was driven predominantly by animals in the latter half of the milking queue. Figure 5 displays the

results for the noise-penalized encoding. We see first that cows that entered consistently at the very rear of the queue (cluster 1) were significantly over-represented in the time budget cluster characterized by moderate time spent eating, low time nonactive, and high rates of rumination (cluster 4). Cows that entered nearer the back of the queue (cluster 2), just ahead of the cows that consistently brought up the rear, were also over-represented in the same time budget cluster – a trend that was statistically significant for the plasticity-penalized encoding but only marginally significant for the noise-penalized encodings. In fact, very few animals that entered in the front half of the queue were found to have this time budget pattern, with cows entering just behind the leaders being significantly under-represented in this time budget cluster. One potential interpretation of this pattern might be that, if these cows are prioritizing time investments in rumination, then this strategy may include hanging back towards the later part of the queue, where they may be able to chew their cud while avoiding the more serious contention for parlor entry position. Further analysis that could facilitate visualization of the cyclical patterns in this time budget data would be needed, however, to confirm this suspicion, and will be left for future work.

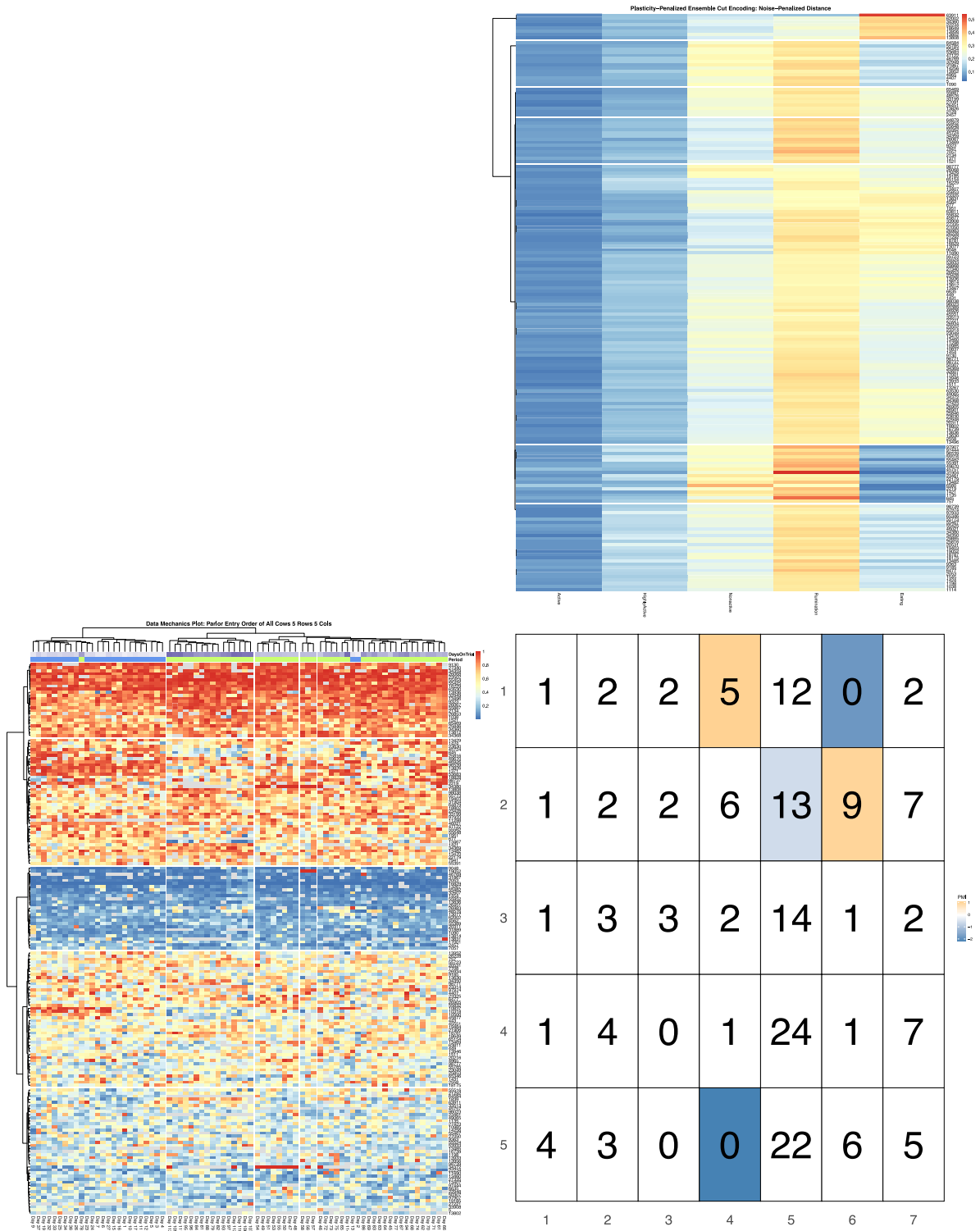


Figure 5: Visualization produced using the `compareEncoding` utility with cells colored by pointwise mutual information estimates significant at the $\alpha = 0.05$ significance level after simulations using multinomial resampling. Data mechanics encoding of parlor entry position

is presented to the row margin of the contingency table, wherein the heatmap contains row annotations representing days on trial and the observation period, such that the pen period corresponds with the observation window of the overall time budget. The noise-penalized encoding of overall time budget is represented on the column axis of the contingency table. Pointwise mutual information values reveal that significant MI test between these two encodings is driven predominantly by behavioral patterns amongst cows in the latter half of the milking queue.

While this more moderate tradeoff between rumination and nonactivity demonstrated a fairly straightforward and progressive trend across the milking queue, which might readily have been captured by a linear model, more complex dynamics were found for the time budget cluster characterized by extremely low time spent eating and high time spent ruminating and nonactive (cluster 6). Cows that consistently entered at the very end of the queue were significantly under-represented in this extreme time budget, while the cows that entered just ahead of them were significantly over-represented. While an extreme tradeoff in eating and ruminating might be explained by issues with sensor placement, that such cows are not evenly dispersed across the herd may instead indicate a biological driver. Health status naturally comes to mind with such an extreme time budget, and indeed several previous studies have reported higher rates of health complications amongst animals in the latter part of the milking queue (Gadbury 1975; Rathore 1982; Berry *et al.* 2012), but health status alone would not necessarily explain the inversion in association pattern between these two adjacent queue groups.

Previous analyses of milk order records have also revealed that, while cows in general tend to be more consistent in their parlor entry order than would be expected in a purely random system, cows at both the front and rear of the herd tend to be particularly persistent their queuing position (Soffie *et al.* 1976; Rathore 1982; Gadbury 1975l; Berry and McCarthy 2012; Beggs *et al.* 2018). It remains unclear in analyses of milk order alone, however, to what degree this pattern is attributable to the cows themselves, and any broader behavioral strategies (syndromes) that they may have adopted, and how much is driven by the natural domain constraints intrinsic to this measurement system (Beggs *et al.* 2018; McVey *et al.* 2020). In older observational studies of movement patterns of semi-feral cattle, it has been noted that cow herds appear to be “led” from both the front and the rear of the queue (Kilgour and Scott 1959; Reinhardt 1982). One interpretation may then be that sick cows, who cannot maintain a normal time budget, may also be pushed back by competition for entry position in the milking queue, but they cannot be pushed behind this small group of cows that may be “leading from the rear”, effectively serving as a “caboose” for the longer train of animals as they move between locations to ensure that stragglers are not left behind. While this behavioral pattern has not been reported in previous experimental studies, it would also not be surprising that such a nonlinear dynamic might be overlooked in analyses relying on linear modeling methods. Indeed, competition between these two behavioral mechanisms as part of a more complex behavioral system may explain why relationships between parlor entry position, home pen behaviors, and health status have proven particularly difficult to reliably establish in previous work (Soffie *et al.* 1976; Cooper *et al.* 2008; Berry *et al.* 2012).

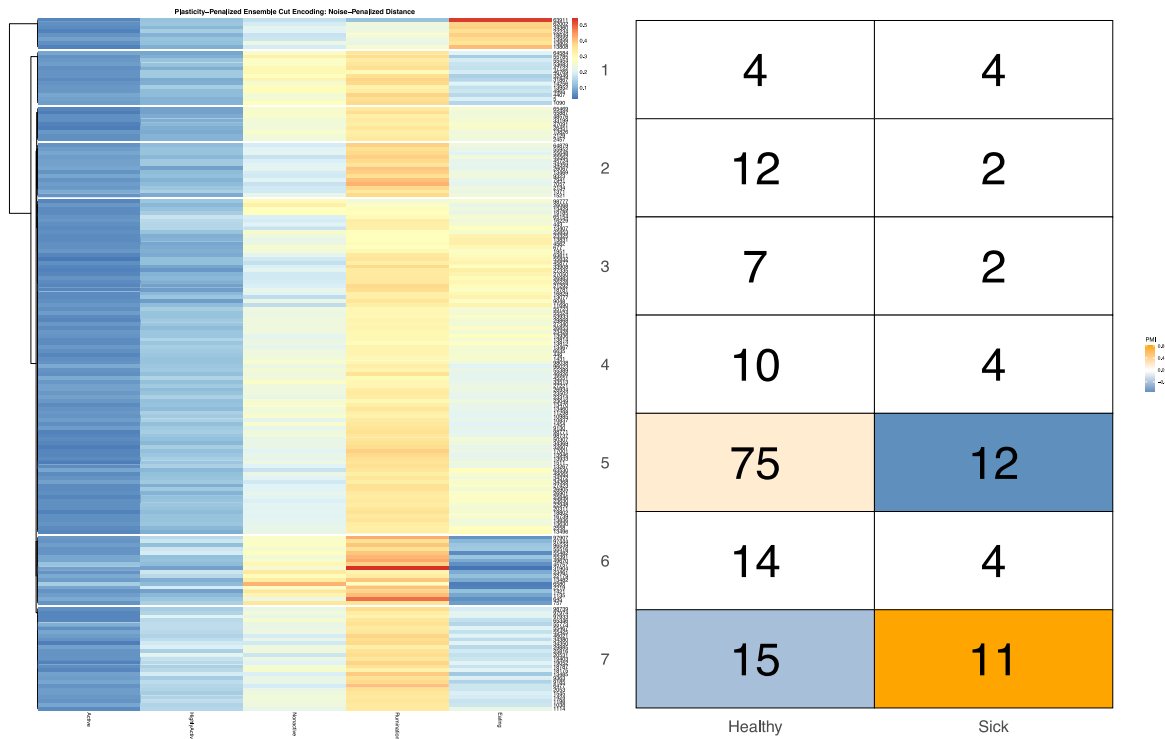


Figure 6: Visualization produced using the `compareEncoding` utility with cells colored by pointwise mutual information estimates significant at the $\alpha = 0.05$ significance level after simulations using multinomial resampling. Cows with the most moderate time budgets were over-represented among animals with no recorded health events, while sick cows were over-represented in the overall time budget cluster characterized by relatively low time spend eating and low-to-moderate amounts of time spent nonactive.

Follow-up bivariate tests with health records confirmed that cows with confirmed health complications were indeed over-represented in the later third of the milking queue (see Supplement). Mutual information and PMI values did not, however, reveal a significant link between health status and a finer stratification of these late-entering cows. This result, which considers only confirmed cases of acute illness, may however be under-powered if this

behavioral relationship is also influenced by subclinical illnesses not reflected in these health records. More perplexingly, in bivariate analyses with either encoding of overall time budget, cows recorded with acute illness were not found to be significantly over-represented among the time budget cluster with extremely low rates of eating. In Figure 7 we see that sick cows were in fact only significantly over-represented in the time budget cluster characterized by relatively low time spent eating, moderate nonactivity, and elevated rates of high activity – an association that was significant for both the noise and plasticity penalized time budget encodings. Interestingly, in repeating the bivariate analyses using an encoding of entry order patterns fit only to animals with no recorded health events, the same time budget cluster with elevated rates of illness was also found to be over-represented amongst cows entering near the end of the queue just in front of the “caboose cows” (see Figure 8). Conversely, in this analysis absent animals with clinical disease, animals entering at the very rear of the herd were shown to be over-represented in the time budget cluster that was perhaps best-characterized as demonstrating the most balanced time investments across all five behavioral axes, while cows entering just ahead of the caboose cows were under-represented in this moderate time budget cluster. These results may add weight to the suspicion that ambiguities between clinical and subclinical illness may be obscuring the role of latent health status as at least one key biological link between home pen and milking queue behavior.

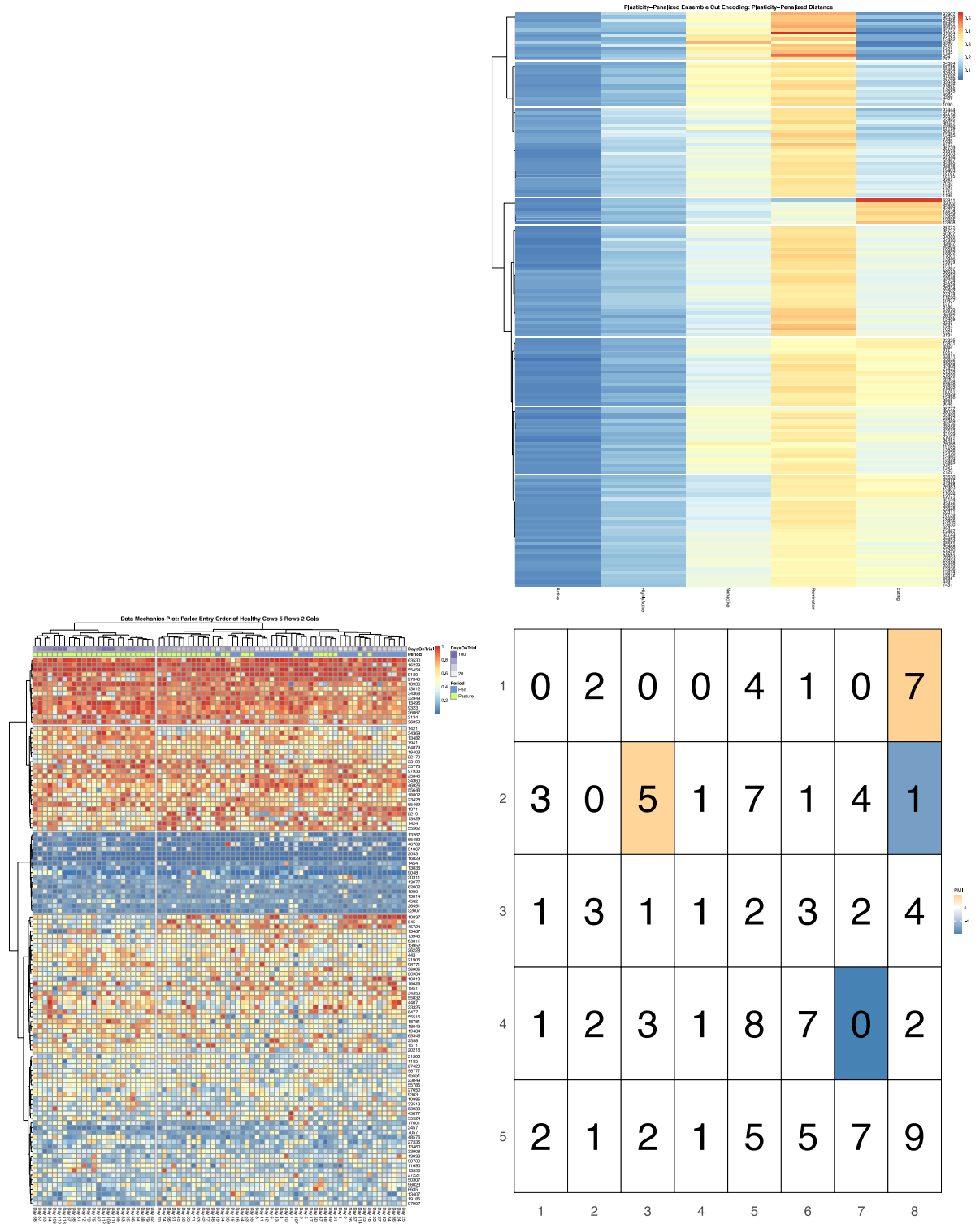


Figure 7: Visualization produced using the compareEncoding utility with cells colored by pointwise mutual information estimates significant at the $\alpha = 0.05$ significance level after simulations using multinomial resampling. Data mechanics encoding of time budget data using

only cows with no recorded health events is represented on the row axis, and the plasticity-penalized encoding of overall time budget is represented on the column axis. Among cows with no acute illness, cows at the very end of the queue are now over represented in the time budget cluster characterized with fairly high time spent eating (cluster 8). Cows entering just ahead of them are not only under-represented in this high eating time cluster, but are also over-represented in the cluster with relatively low eating time cluster with low-to-moderate nonactivity (cluster 3) that was independently associated with higher rates of clinical illness.

Conclusion

Time budgets provide a convenient and intuitive means of quantitatively summarizing the behavioral tradeoffs of animals, but multinomial-distributed data present a number of analytical challenges. The results of this analytical case study have highlighted how a novel simulation-based approach may be employed to simultaneously accommodate both the codependency structures fundamental to multivariate-distributed data formats and the complex multi-faceted sources of measurement uncertainty that may be encountered across a broader range of PLF data streams. While such simulations may be more computationally expensive than closed-form estimators, we have demonstrated that an ensemble of data mimics can be efficiently repurposed throughout the analytical pipeline to improve not only the visualization of these behavioral tradeoffs, but also the compression of such information into robust empirically-defined discrete encodings. It should be noted, however, that the utility of these novel clustering techniques are not restricted to time budget data. The ensemble-penalized dissimilarity

estimator and ensemble cut algorithm that we have introduced in this case study are both fundamentally nonparametric. This means that their implementation is in no way intrinsically restricted to any particular class of data. Subsequently, the choices that a user makes in constructing an appropriate error simulation model are restricted only by their own creativity, allowing this analytical framework to be easily generalized to a much wider array of PLF data streams and the wider array of complex error structures that they have to offer.

Additionally, while discrete data is typically seen as an impediment to statistical analysis in most model-based approaches, we hope that this analytical case study has served to demonstrate the comparable ease with which insights may be extracted from encoded data when an information theoretic approach is employed. For large, structurally complex, and often informationally redundant PLF data streams, an efficient encoding may be far easier to achieve than a comprehensive model that can fully accommodate the temporal dynamics of behavioral responses in dynamic farm environments. This may be especially true for data sets where all the factors driving such behavioral responses are not measurable. By avoiding entirely any form of least-squared optimization utilized in most model-based approaches, we've shown that an entirely model-free approach is able to recover nonlinear dynamics between entry order and overall time budget that likely would have been overlooked if an assumption of linearity had been employed. While more formal model-based inferences may be warranted for further analysis of the underlying causes of this relationship, the exploratory analysis tools provided by the LIT pipeline have undoubtedly served to create a more comprehensive picture of the complex behavioral dynamics hiding within these two underutilized data streams.

References

- Adamczyk, K., Cywicka, D., Herbut, P., Trześniowska, E., 2017. The application of cluster analysis methods in assessment of daily physical activity of dairy cows milked in the Voluntary Milking System. *Computers and Electronics in Agriculture* 141, 65–72.
<https://doi.org/10.1016/j.compag.2017.07.007>
- Agresti, A., 2013. *Categorical Data Analysis*, 3rd ed, Wiley Series in Probability and Statistics. John Wiley & Sons, Inc, Hoboken NJ.
- Beggs, D.S., Jongman, E.C., Hemsworth, P.H., Fisher, A.D., 2018. Short communication: Milking order consistency of dairy cows in large Australian herds. *Journal of Dairy Science* 101, 603–608. <https://doi.org/10.3168/jds.2017-12748>
- Berry, D.P., McCarthy, J., 2012. Genetic and non-genetic factors associated with milking order in lactating dairy cows. *Applied Animal Behaviour Science* 136, 15–19.
<https://doi.org/10.1016/j.applanim.2011.11.012>
- Bikker, J.P., van Laar, H., Rump, P., Doorenbos, J., van Meurs, K., Griffioen, G.M., Dijkstra, J., 2014. Technical note: Evaluation of an ear-attached movement sensor to record cow feeding behavior and activity. *Journal of Dairy Science* 97, 2974–2979.
<https://doi.org/10.3168/jds.2013-7560>
- Brenninkmeyer, C., Dippel, S., Brinkmann, J., March, S., Winckler, C., Knierim, U., 2016. Investigating integument alterations in cubicle housed dairy cows: which types and locations can be combined? *Animal* 10, 342–348.
<https://doi.org/10.1017/S1751731115001032>

- Clouard, C., Resmond, R., Prunier, A., Tallet, C., Merlot, E., n.d. Exploration of early social behaviors and social styles in relation to individual. *Scientific Reports* 15.
- Cooper, M.D., Arney, D.R., Webb, C.R., Phillips, C.J., 2008. Interactions between housed dairy cows during feeding, lying, and standing. *Journal of Veterinary Behavior* 3, 218–227. <https://doi.org/10.1016/j.jveb.2007.09.005>
- Dutta, R., Smith, D., Rawnsley, R., Bishop-Hurley, G., Hills, J., Timms, G., Henry, D., 2015. Dynamic cattle behavioural classification using supervised ensemble classifiers. *Computers and Electronics in Agriculture* 111, 18–28. <https://doi.org/10.1016/j.compag.2014.12.002>
- Efron, B., Tibshirani, R.J., 1993. *An Introduction to the Bootstrap*. Springer US, Boston, MA. <https://doi.org/10.1007/978-1-4899-4541-9>
- Farine, D.R., 2017. A guide to null models for animal social network analysis. *Methods Ecol Evol* 8, 1309–1320. <https://doi.org/10.1111/2041-210X.12772>
- Gadbury, J.C., 1975. Some preliminary field observations on the order of entry of cows into herringbone parlours. *Applied Animal Ethology* 1, 275–281. [https://doi.org/10.1016/0304-3762\(75\)90020-6](https://doi.org/10.1016/0304-3762(75)90020-6)
- Gareth, J., Witten, D., Hastie, T., Tibshirani, R., 2013. *An Introduction to Statistical Learning*, Springer Texts in Statistics. Springer New York, New York, NY. <https://doi.org/10.1007/978-1-4614-7138-7>

- Guan, J., Hsieh, F., 2018. Coupling Geometry on Binary Bipartite Networks: Hypotheses Testing on Pattern Geometry and Nestedness. *Front. Appl. Math. Stat.* 4, 38. <https://doi.org/10.3389/fams.2018.00038>
- Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, Springer Series in Statistics. Springer New York, New York, NY. <https://doi.org/10.1007/978-0-387-84858-7>
- Hausser, J., Strimmer, K., 2021. entropy: Estimation of Entropy, Mutual Information and Related Quantities, 1.3.0. ed.
- Hausser, J., Strimmer, K., n.d. Entropy Inference and the James-Stein Estimator, with Application to Nonlinear Gene Association Networks 16.
- Hou, S., Cheng, X., Shi, L., Zhang, S., 2020. Study on Individual Behavior of Dairy Cows Based on Activity Data and Clustering, in: Proceedings of the 2020 2nd International Conference on Robotics, Intelligent Control and Artificial Intelligence. ACM, Shanghai China, pp. 210–216. <https://doi.org/10.1145/3438872.3439082>
- Hsieh, F., Chou, E.P., Chen, T.-L., 2021. Mimicking Complexity of Structured Data Matrix's Information Content: Categorical Exploratory Data Analysis. *Entropy* 23, 594. <https://doi.org/10.3390/e23050594>
- Hsieh, F., Liu, S.-Y., Hsieh, Y.-C., McCowan, B., 2018. From patterned response dependency to structured covariate dependency: Entropy based categorical-pattern-matching. *PLoS ONE* 13, 1–28. <https://doi.org/10.1371/journal.pone.0198253>

- Johnson, M., 1999. Confidence intervals on likelihood estimates for estimating association strength.
- Kassambara, A., 2020. ggpubr: “ggplot2” Based Publication Ready Plots, 0.4.0. ed.
- Kilgour, R., Scott, T.H., 1959. Leadership In A Herd Of Dairy Cows. Proceedings of the New Zealand Society of Animal Production 19, 36–43.
- Kirby, M., 2001. Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns. John Wiley & Sons, Inc, New York, USA.
- Kolde, R., 2019. pheatmap: Pretty Heatmaps.
- Lee, M., Lee, S., Park, J., Seo, S., 2020. Clustering and Characterization of the Lactation Curves of Dairy Cows Using K-Medoids Clustering Algorithm. Animals 10, 1348. <https://doi.org/10.3390/ani10081348>
- MacKay, D.J.C., 2003. Information Theory, Inference, and Learning Algorithms, 4th ed. Cambridge University Press, Cambridge UK.
- Manriquez, D., Chen, L., Albornoz, G., Velez, J., Pinedo, P., 2018. Case Study: Assessment of human-conditioned sorting behavior in dairy cows in farm research trials. The Professional Animal Scientist 34, 664–670. <https://doi.org/10.15232/pas.2018-01749>
- Manriquez, D., Chen, L., Melendez, P., Pinedo, P., 2019. The effect of an organic rumen-protected fat supplement on performance, metabolic status, and health of dairy cows. BMC Vet Res 15, 1–14. <https://doi.org/10.1186/s12917-019-2199-8>

- McCowan, B., Beisner, B., Bliss-Moreau, E., Vandeleest, J., Jin, J., Hannibal, D., Hsieh, F., 2016. Connections Matter: Social Networks and Lifespan Health in Primate Translational Models. *Front. Psychol.* 7. <https://doi.org/10.3389/fpsyg.2016.00433>
- McVey, C., Hsieh, F., Manriquez, D., Pinedo, P., Horback, K., 2020. Mind the Queue: A Case Study in Visualizing Heterogeneous Behavioral Patterns in Livestock Sensor Data Using Unsupervised Machine Learning Techniques. *Frontiers in Veterinary Science* 7, 20. <https://doi.org/10.3389/fvets.2020.00523>
- Papadakis, M., Tsagris, M., Dimitriadis, M., Fafalios, S., Tsamardinos, I., Fasiolo, M., Borboudakis, G., Burkardt, J., Zou, C., Lakiotaki, K., Chatzipantsiou, C., 2020. Rfast: A Collection of Efficient and Extremely Fast R Functions, 1.9.9. ed.
- Pereira, G.M., Heins, B.J., Endres, M.I., 2018. Technical note: Validation of an ear-tag accelerometer sensor to determine rumination, eating, and activity behaviors of grazing dairy cattle. *Journal of Dairy Science* 101, 2492–2495. <https://doi.org/10.3168/jds.2016-12534>
- Pinheiro, J.C., Bates, D.M., 2000. *Mixed-Effects Models in S and S-PLUS*, Statistics and Computing. Springer-Verlag, New York USA.
- R Core Team, 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rand, W.M., 1971. Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association* 66, 846–850. <https://doi.org/10.1080/01621459.1971.10482356>

- Rathore, A.K., 1982. Order of cow entry at milking and its relationships with milk yield and consistency of the order. *Applied Animal Ethology* 8, 45–52.
[https://doi.org/10.1016/0304-3762\(82\)90131-6](https://doi.org/10.1016/0304-3762(82)90131-6)
- Reinhardt, V., 1982. Movement Orders and Leadership in a Semi-Wild Cattle Herd. *Behav* 83, 251–264. <https://doi.org/10.1163/156853983X00183>
- Schwager, M., Anderson, D.M., Butler, Z., Rus, D., 2007. Robust classification of animal tracking data. *Computers and Electronics in Agriculture* 56, 46–59.
<https://doi.org/10.1016/j.compag.2007.01.002>
- Shirkhorshidi, A.S., Aghabozorgi, S., Wah, T.Y., 2015. A Comparison Study on Similarity and Dissimilarity Measures in Clustering Continuous Data. *PLoS ONE* 10, e0144059.
<https://doi.org/10.1371/journal.pone.0144059>
- Soffié, M., Thinès, G., De Marneffe, G., 1976. Relation between milking order and dominance value in a group of dairy cows. *Applied Animal Ethology* 2, 271–276.
[https://doi.org/10.1016/0304-3762\(76\)90060-2](https://doi.org/10.1016/0304-3762(76)90060-2)
- Stygar, A.H., Gómez, Y., Berteselli, G.V., Dalla Costa, E., Canali, E., Niemi, J.K., Llonch, P., Pastell, M., 2021. A Systematic Review on Commercially Available and Validated Sensor Technologies for Welfare Assessment of Dairy Cattle. *Front. Vet. Sci.* 8, 634338. <https://doi.org/10.3389/fvets.2021.634338>
- Tucker, C.B., Jensen, M.B., de Passillé, A.M., Hänninen, L., Rushen, J., 2021. Invited review: Lying time and the welfare of dairy cows. *Journal of Dairy Science* 104, 20–46.
<https://doi.org/10.3168/jds.2019-18074>

Valletta, J.J., Torney, C., Kings, M., Thornton, A., Madden, J., 2017. Applications of machine learning in animal behaviour studies. *Animal Behaviour* 124, 203–220.

<https://doi.org/10.1016/j.anbehav.2016.12.005>

Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Xu, H., Li, S., Lee, C., Ni, W., Abbott, D., Johnson, M., Lea, J.M., Yuan, J., Campbell, D.L.M., 2020a. Analysis of Cattle Social Transitional Behaviour: Attraction and Repulsion. *Sensors* 20, 5340. <https://doi.org/10.3390/s20185340>

Xu, H., Li, S., Lee, C., Ni, W., Abbott, D., Johnson, M., Lea, J.M., Yuan, J., Campbell, D.L.M., 2020b. Analysis of Cattle Social Transitional Behaviour: Attraction and Repulsion. *Sensors* 20, 5340. <https://doi.org/10.3390/s20185340>

CHAPTER 3

Tensor Mechanics: A case study in recovering systematic temporal patterns from multivariate data streams by employing unsupervised machine learning approaches in the analysis of daily time budget records

Introduction

Precision Livestock Farming (PLF) technologies create new opportunities to record the behaviors of large groups of socially housed animals in complex working farm environments. The datasets produced by such technologies, however, often contain measurements collected more frequently and over longer observation windows than is possible with direct human observation, resulting in complex structural features that in turn present new logistical challenges in extracting usable knowledge from these rich data streams (Valleta *et al.* 2017; McVey *et al.* 2021). In previous work, we introduced the concept of ensemble mimics, wherein minimally parametric simulation techniques can be integrated into standard hierarchical clustering pipelines in order to account for complex nested sources of measurement error often encountered in PLF applications. In analyses of overall time budget data, this not only allowed us to account for heterogeneous variance in sensor noise, but also to penalize animals whose behavioral responses were more variable over the observation window. This subsequently produced encodings that gave relatively more weight to cows with the most clearly defined behavioral responses that were pervasive across environmental contexts (McVey *et al.* 2021). In previous work with milking order records extracted from metadata produced by RFID-equipped rotary milking parlors, however, we also

demonstrated that interesting ethological insights can be gleaned by recovering systematic temporal variations in behavioral responses using a data mechanics clustering approach, even when the environmental factors eliciting changes in queueing patterns were not recorded and only the behavioral artifacts remained in the response data (McVey *et al.* 2020). In light of these collective results, an algorithmic framework is needed to disaggregate overall time budgets to determine if variability that was previously penalized using ensemble mimicry approaches might in fact be hiding systematic heterogeneity in daily time budgets that might provide further insights into the behavioral responses of this herd of cows across management contexts.

In previous work with entry order records, we showed that data mechanics can be conceptualized as a form of nonlinear PCA, and thus used as a discrete but more readily explainable alternative to manifold embedding algorithms (Hsieh *et al.* 2018 ; McVey *et al.* 2020; Hsieh *et al.* 2021). Its efficacy has been demonstrated for information compression for multivariate response data (McCowan *et al.* 2016; Hsieh *et al.* 2018) and for repeated measures of univariate response (McVey *et al.* 2020), but daily time budget records are a multivariate repeated measure. While it would be possible to flatten daily time budget records in order to analyze this dataset as a 2D matrix using the standard data mechanics analysis pipeline, this would afford no control or differentiation in weighting the behavioral and temporal axes in the final encodings. Forcing the algorithm to learn, from the data itself, fundamental structural features that could easily be specified *a priori*, is inherently inefficient. This chapter will therefore explore how data mechanics algorithms can be extended, not only to incorporate information about measurement error from new ensemble mimicry techniques, but also to utilize tensor constructs to more naturally accommodate the complex temporal structures often encountered in PLF data streams. The result will be an algorithmic framework that will allow ethologists to leverage simultaneously the multivariate and temporal richness of similar PLF data streams, in order to provide an approach to

model-free knowledge discovery that is not only more statistically powerful, but will also provide more wholistic ethological inferences.

Description of Data

To explore the efficacy this proposed analytical framework, data was repurposed from a feed trial conducted on a working dairy farm assessing the efficacy of an organic fat supplement on cow health during early lactation (Manriquez *et al.* 2019). All animal handling and experimental protocols were approved by the Colorado State University Institution of Animal Care and Use Committee (Protocol ID: 16-6704AA). The study was conducted in 2017, from January through July, on a USDA Certified Organic dairy in Northern Colorado. A total of 200 cows were enrolled over a 1.5-month period into a mixed-parity herd. Cows were housed in a closed herd in an open-sided free-stall barn that was stocked at roughly half capacity with respect to both feed bunk spaces and stalls, with free access to an adjacent outdoor dry lot. The grazing season began in April, at which point cows were moved onto pasture each night to comply with organic grazing standards. Cows had free access to TMR between milkings while in their home pen, and were head locked each morning to facilitate daily health checks and heat detection. For more details on feed trial protocols, the housing environment, and the day to day management of animals, see Manriquez *et al.* (2018) and Manriquez *et al.* (2019).

In order to facilitate direct comparisons with previously reported work, daily time budget records were calculated for this methods paper using the same CowManager ear tag accelerometer (Agis Automatisering BV, Harmelen, Netherlands) records used previously to analyze overall time budgets (McVey *et al.* 2021). Data scrubbing decisions used to exclude incomplete records were identical to those reported previously to calculate overall time budgets, with the only appreciable difference in data wrangling being that here hourly time budgets were aggregated conditional on

day of observation. The resulting data was formatted as a three-dimensional tensor, with cows ($n = 177$) on the first axis, time (65 days after enrollment period and prior to pasture access) on the second axis, and mutually exclusive behaviors (eating, ruminating, nonactive, active, highly active) on the third axis (Bikker *et al.* 2014; Pereira *et al.* 2018). In order to more intuitively represent this data as the proportion of time devoted to each behavior, the total minutes a given cow was recorded as engaging in a given behavior on a given day was normalized by the total minutes that a given cow was recorded on a given day of observation.

Creating an Ensemble of Data Mimicries using Simulated Noise

An ensemble of data mimicries was created for this observed data tensor using the *simTimeBudget* utility previously developed for the LIT package. Observational error attributable to the precision of the sensor was again simulated by stochastically resampling each observed hourly time budget using the joint Dirichlet-Multinomial sampling strategy (McVey *et al.* 2021). The resulting mimicry of the raw sensor data was then conditionally aggregated by date of observation to create a four-dimensional tensor formatted identical to the observed data tensor but with simulation number on the fourth axis to comply with LIT package formatting standards. Ensemble variance estimates were calculated over the simulation axis for each combination of cow, day, and behavior indices. As with overall time budget, these ensemble variance estimates could subsequently be used to scale dissimilarity estimates for the observed data to account for heterogeneity in multinomial-formatted data both within and between behavioral axes.

In previous analysis of overall time budgets, jackknife resampling could be performed within-animal to nonparametrically estimate the reliability of the underlying behavioral signal by leveraging information in the temporal subsamples used to create the aggregate record (McVey *et al.* 2021). With daily time budget records, however, this strategy could not be employed, as

homogeneity of hourly time budgets could not be assumed due to fluctuations in behaviors imposed by the management schedule and the circadian rhythms of the animals themselves (McVey *et al.* 2021). Thus, while the behavioral axis could be rescaled using the precision-penalized ensemble weighted distances, systematic differences between days of observation was here accommodated by the empirically-driven iterative re-weighting of the time axis at the heart of the data mechanics algorithm (Guan and Fushing 2018).

Encoding Daily Time Budgets

Algorithmic Approach

In order to encode daily time budgets, the basic data mechanics algorithm was extended to a three-dimensional tensor. As before, column clusters were used to reweight dissimilarity matrices calculated for row observations, and row clusters were used to re-weight dissimilarity matrices calculated for column observations, allowing structural information to be shared between the two axes (Guan and Fushing 2018). To accommodate a multivariate response, dissimilarity values were aggregated over the third axis at each matrix index prior to aggregation over the remaining matrix index for which the dissimilarity matrix was calculated. The efficacy of three dissimilarity estimators were explored. The first was a simple unweighted Euclidean distance (L2 norm), which is functionally equivalent to the standard two-dimensional implementation of the data mechanics algorithm, except that in the tensor implementation all behaviors for a given time budget observation are forced into the same cluster. The second dissimilarity estimator considered was the Kullback-Leibler Divergence (KLD) Distance, which is the sum of the asymmetric relative entropy estimates for any two probability distributions vectors that sum to one. Here KLD-distance was calculated at each index for the first two axes and prior to aggregation over the axis for which the dissimilarity matrix was computed. The third and final dissimilarity estimator explored was

the ensemble-weighted Euclidean distance, wherein the squared distance between observed values were normalized by the sum of the ensemble variance estimates calculated over all simulated datasets at the corresponding tensor indices. The dissimilarity matrices calculated over the subset of row or column indices were reweighted iteratively using the cluster results of the opposite matrix axes until either cluster sets became stable or a maximum of ten iterations were reached (Guan and Fushing 2018). For all three dissimilarity estimators, clustering results were here calculated on a grid of metaparameter values: from one to ten cow clusters and one to six day clusters. Full details on the implementation of the *tensorMechanics* algorithm are provided in Supplemental Materials.

To visualize the final dendrograms produce from each tensor mechanics optimization, the *pheatmap* package was used to create heatmaps wherein cows were arranged along the row axis, days along the column axis, and cells colored to represent the proportion of time that a given cow dedicated to a given behavior on a given day (Kolde 2019). To help identify temporal patterns captured by these clustering results, the column axis was annotated in purple with the day on trial (DoT) that each set of records were recorded. In order to visualize patterns recovered across all five behavioral axes simultaneously, the *ggpubr* package was used to create a composite image of the final heatmaps created using equivalent 0 to 1 scales to facilitate direct visual comparisons of behavioral investments (Kassambara 2020).

Clustering Results

Visualizations for final clustering results for all three dissimilarity estimators for all candidate metaparameter values are provided in Supplemental Materials. Comparisons of dissimilarity estimators for encodings of daily time budgets largely mirrored the clustering dynamics found with encodings of overall time budgets. Tensor mechanics encodings using the

standard unweighted Euclidean norm, which employed no rescaling of behavioral axes, over-emphasized patterns in high frequency behaviors such as eating and rumination, recovering very little systematic differences in any of the activity axes across days or animals. KLD distance, on the other hand, produced a more balanced encoding across the five behavioral axes, but again was prone to over-estimate dissimilarity at the extremes of the time budget distribution, which resulted in a number of animals with extremely low or high eating times being classified as outlier in clusters containing only one animal, which would effectively remove these animals from consideration in downstream analyses of bivariate associations using these clustering results.

The encoding created using the ensemble-weighted Euclidean distance is visualized in Figure 1. This dissimilarity estimator provided arguably the most balanced representation of heterogeneity in daily time budgets across all five behavioral axes without over-pruning at the extremes of the distribution. Perhaps the most striking feature of this visualization, however, is the remarkable consistency in daily time budget records across this observational window. As with overall time budget, eating time remains the primary driver of difference between animals in this encoding. At the extremes of these eating time budgets, there are few systematic differences in clusters across the temporal (column) axis, nor even much variability in observations within clusters, suggesting that neither transient environmental fluctuations or more persistent changes in management or the biology of these animals had much influence on the time investments of these individuals with more extreme behavioral strategies.

Among cows with more moderate eating times, there is certainly more variability within clusters, but systematic differences in cells across the temporal axis are still surprisingly subtle. In contrasting what systematic patterns are apparent against the time indices of these records, it appears that time spent eating was slightly elevated during roughly the first 30-60 days on trial relative to the remainder of the observation window, which would have encompassed the transition

period for many of these animals and the earliest (pre-peak) stages of lactation for nearly all cows enrolled on this trial. This result is counterintuitive, as we would expect appetites to be suppressed immediately following calving and gradually increase throughout this observational window; however, it has been shown previously that time spent eating is not always well correlated with feed intake, as cows can compensate for reduced time invested in mastication at initial ingestion with increases in time masticating during subsequent rumination (Huzzy *et al.* 2005; Miedema *et al.* 2011; Beauchemin 2018). Therefore, it is also possible that this early surge in time spent at the feed bunk might represent increase in feed sorting behaviors, which might represent latent behavioral strategies cows employ to cope with this period of negative energy balance, or it could simply reflect palatability issues related to the nutritional supplement or some other element of the total mixed ration.

It is also interesting to note that, amongst animals with moderate eating times, the drop off in eating times seems to correspond with a slight increase in time spent highly active observed across all cow clusters. While this temporal subperiod would likely correspond to the return to estrus for some animals in this herd (Reith and Hoy 2018), the pervasiveness of this shift would seem more easily attributable to some latent biological or managerial shift, such as an increase in appetite during peak milk that led to a greater number of trips to the feed bunk between milkings (Beauchemin 2018). It is also interesting to note just how little temporal heterogeneity is found among the nonactive and rumination axes. As these cows were understocked with respect to bunk space during this observation window, this may indicate that conditions in this pen support good lying time irrespective of fluctuations in the management environment (Tucker *et al.* 2021).

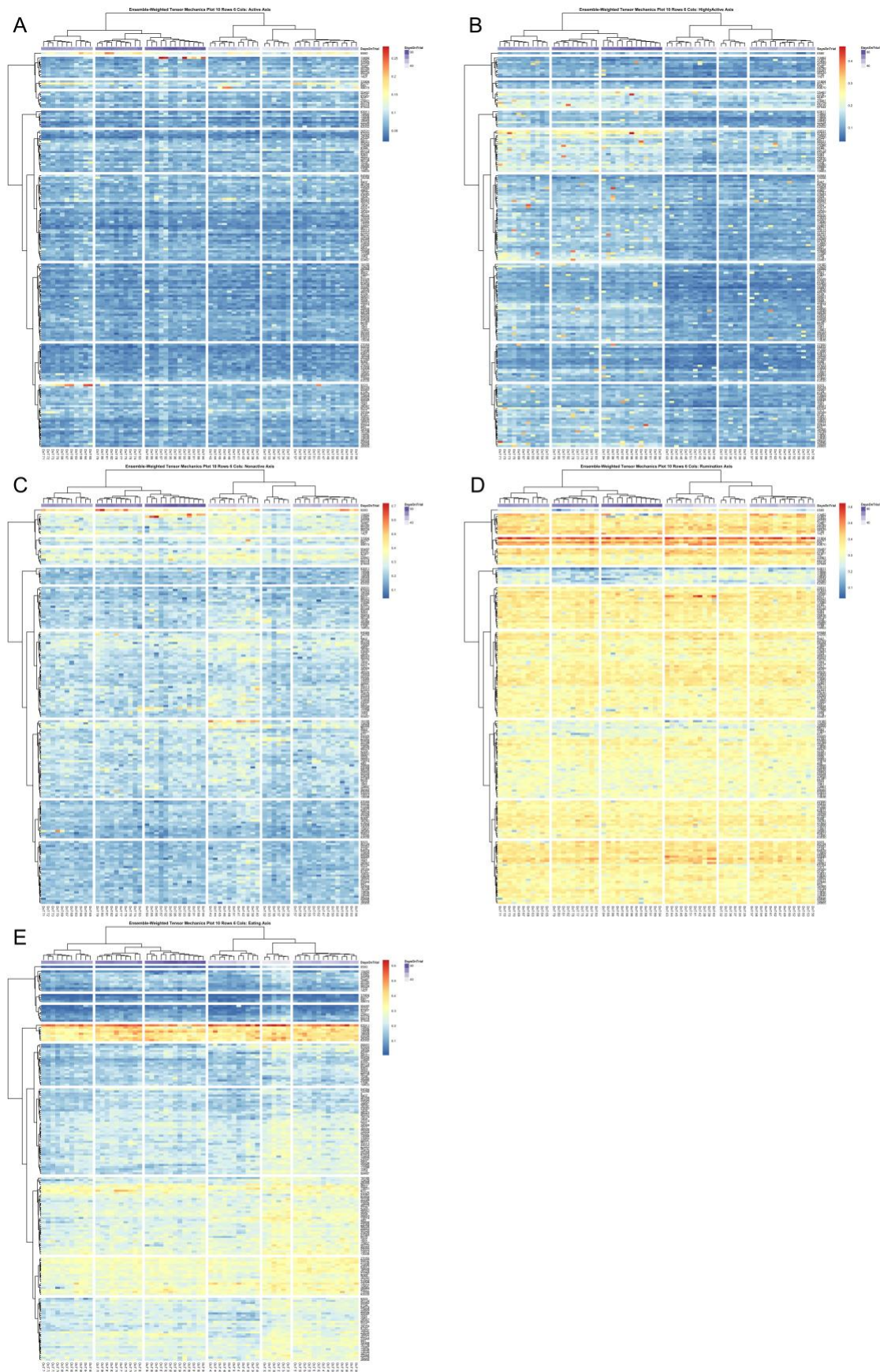


Figure 1: Heatmap visualization of tensor mechanics clustering results created using the ensemble-weighted distance. Cows are indexed on the row axis, day on trial that observations were recorded are indexed on the column axis, and each cell is colored by the proportion of time that was dedicated to a given behavior by a given cow on a given day. Patterns in each behavioral axis are visualized in distinct heatmaps arranged as active (A – top left), highly active (B – top right), nonactive (C – middle left), rumination (D – middle right), and eating (E – bottom). Note that all heatmaps use the same dendrogram orderings for row and columns and equivalent cell color scaling to facilitate direct visual comparisons. From these results we see that there is far greater heterogeneity in time budgets across animals than over days. Some temporal heterogeneity is found in minutes spent highly active (cows are more active later in the trial), and among animals with moderate eating times (spending more time eating earlier in the trial), but overall all cows are remarkably consistent in their time budget during this observational window. See Supplemental Materials for larger high-resolution versions of these images, wherein full details can be readily visualized.

In comparing the results of the tensor mechanics clustering with the results for data aggregated to an overall time budget, which are visualized in Figure 2, we see that the two encodings are in close agreement. The contingency table reveals that cluster assignments are nearly identical between the two data sets for the coarser branches of the dendrogram nearer the trunk, which reflect differences among more extreme time budgets, and differ only slightly in the cut off values established between branches representing more subtle differences in time budgets. Given the temporal consistency in this daily time budget data, this result is not necessarily surprising, as there is little additional information or complexity to be recovered from disaggregating this information across days for this herd. Accommodation of temporal heterogeneity in the tensor

mechanics encoding appears to have largely only served to more finely distinguish between animals with low eating times. Closer inspection reveals that these distinctions appear to have been made based on differences in tradeoffs between the nonactive and highly active behavioral axes across the early and later phases of this observation window. As a result, the tensor mechanics encoding produces a coarser encoding of animals with more moderate time budgets. This has caused some of the more moderate overall time budget clusters to be consolidated in the tensor mechanics encoding, with some ambiguity in the resulting cutoffs that appears again to be driven by greater weight being placed on how the eating-nonactivity tradeoff and the eating-highly active tradeoff shifts over the observation window.

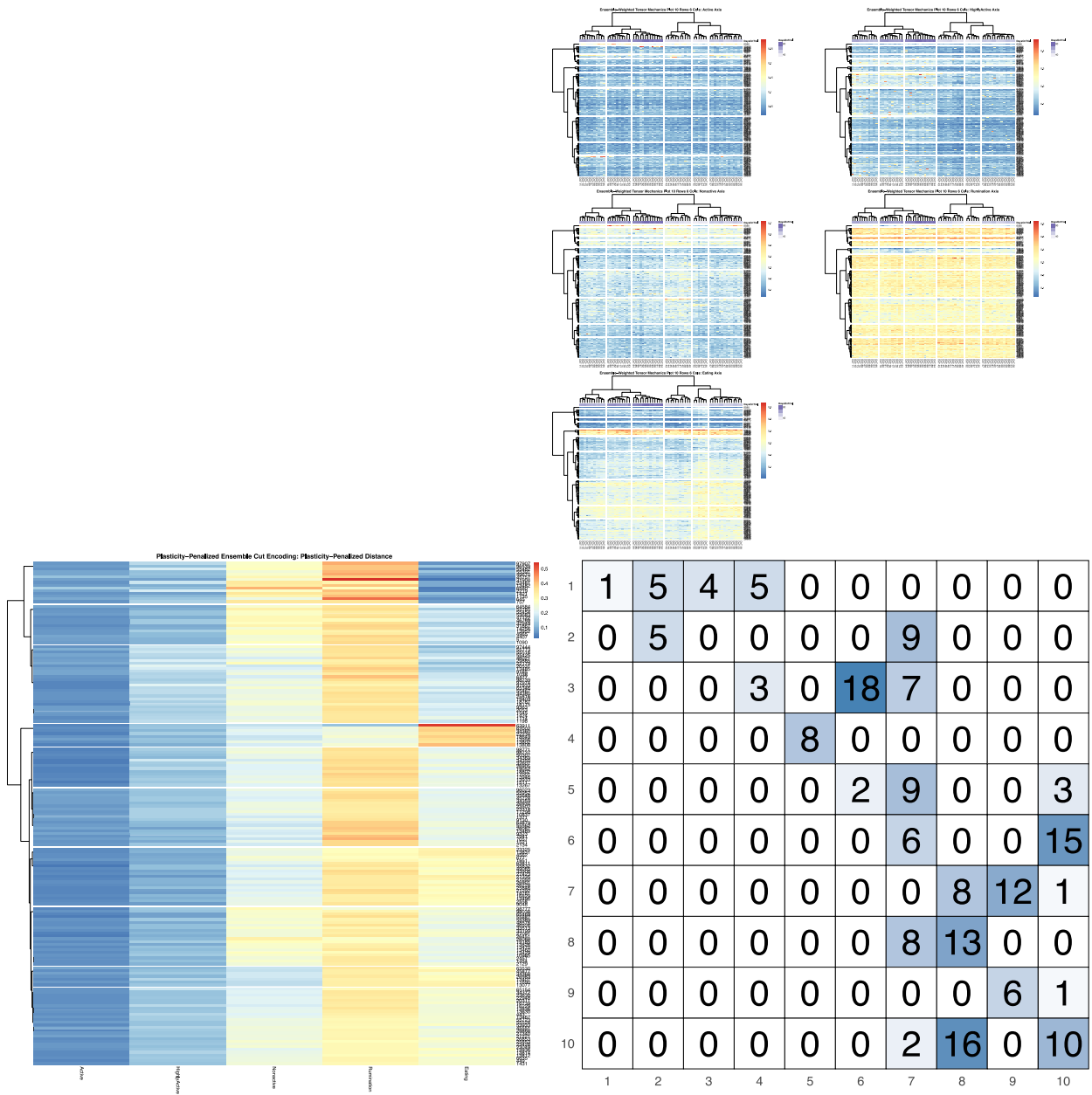


Figure 2: Comparison of encodings created for overall time budgets (bottom left) and daily time budgets (upper right) using ensemble-weighted distance. A contingency table (bottom right) shows the counts of the joint encodings, wherein clusters are numbered from the top to the bottom of each respective heatmap, with the overall time budget clusters arranged on the row axis and the daily time budget clusters arranged on the column axis. From this we see that cows with extremely low eating time are more finely differentiated in the daily time budget encoding, which has placed

more emphasis on heterogeneity in the activity and rumination axes. As a result several of the time budgets with more moderate time spent eating that are differentiated in the overall time budget encoding are combined in the daily time budget encoding, but this again seems to have been driven by greater weight placed on the activity axes than by systematic temporal heterogeneity recovered across daily eating time records. See Supplemental Materials for larger high-resolution versions of these images, wherein full details can be readily visualized.

Bivariate Associations

To determine if the temporal heterogeneity found in daily time budget records would modify bivariate associations found in previous analyses between overall time budgets and other farm data streams, bivariate tree tests were conducted using the tensor mechanics results for all three dissimilarity estimators. The first set of tests conducted explored if patterns in daily time budgets differed between animals fed control diets and those whose TMR rations were amended with the Organilac fat supplement. In prior analyses with overall time budgets, no significant bivariate associations were recovered using the bivariate tree test framework – a result that was not necessarily surprising given that control and treatment animals were housed together throughout the duration of the trial except while head locked for the morning feeding and herd check. Significant bivariate associations ($\alpha \leq 0.05$) were, however, recovered between treatment group and all three dissimilarity estimators used to create tensor mechanics encodings. Visual characterizations of these relationships, which can be found in Supplemental Materials, were created using the *compareEncoding* utility, wherein contingency table cells were colored by pointwise mutual information estimates that were deemed statistically significant by simulation against the null using multinomial resampling (McVey *et al.* 2021). The relationship recovered using the ensemble-weighted Euclidean distance is visualized in Figure 3. Organilac cows were

significantly over represented among daily time budgets that were consistent throughout the observation window and characterized by relatively (but not extremely) high time spent eating, moderate rates of rumination, and low nonactivity. Cows in the control group, on the other hand, were over-represented among daily time budgets that were characterized by time spent eating that varied between moderately-low to moderately-high (slightly elevated during the first half of the observation window), rumination rates that were consistently relatively low, elevated rates of nonactivity during the first half of the observation window, and elevated rates of high activity during the second half of the observation period. Collectively, these results might suggest that, amongst cows with more moderate time budgets, control animals may have struggled in the early stages of this study that encompassed much of the transition milk period, whereas fat-supplemented cows were able to maintain a robust and well balanced time budget throughout this early lactation period (Beauchemin 2018). Alternatively, Organilac cows might simply have spent more time eating throughout the first phase of this research trial in order to sort through their TMR ration in order to avoid the fat supplement. Direct observation of feed bunk behaviors would ultimately be needed to positively determine the underlying mechanism for this behavioral pattern among this subset of animals.

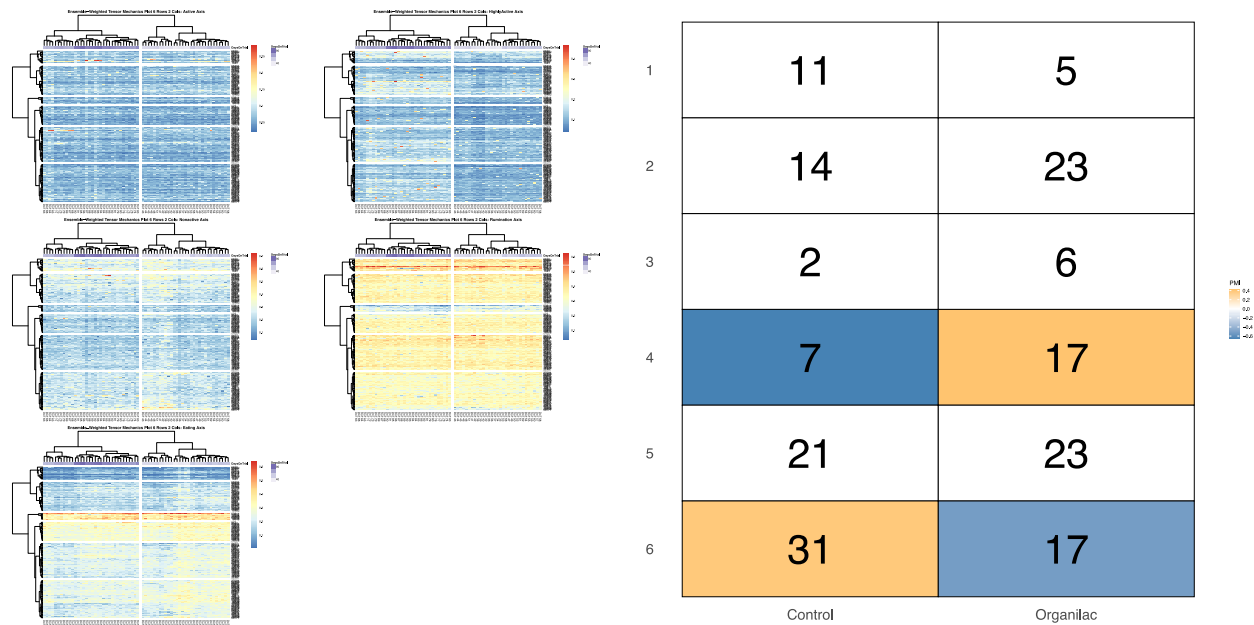


Figure 3: Visualization of significant bivariate relationship between diet treatment group and daily time budgets encoded using the ensemble-weighted dissimilarity estimator (left). The contingency table for this joint encoding (right), colored using significant pointwise mutual information values, reveals that cows on the Organilac treatment diet were significantly over represented among daily time budgets with consistently (but not extremely) high time spent eating. Alternatively, control cows were significantly over represented among animals with more moderate daily time budgets that varied between moderately-low to moderately-high time spent eating, slightly low time spent ruminating, and some temporal tradeoffs between time spent active and nonactive during the early and later phases of this observation window. See Supplemental Materials for larger high-resolution versions of these images, wherein full details can be readily visualized.

Age of cow and milk production were highly correlated throughout the first few lactations (parities 1, 2, and 3+) in this herd. In order to account for this informational redundancy, age in days at start of trial and the 95th quantile of daily milk yield (a model-free estimate of peak milk yield) were here scaled to uniform variance and jointly encoded using the *ecodePlot* utility (McVey *et al.* 2021). Visualizations of this encoding can be found in Supplemental Materials. Bivariate tree tests, however, returned no significant associations between this age-yield encoding and any of the tensor plot encodings of daily time budgets. This result differed slightly from analyses of overall time budgets, wherein encodings that accounted for heterogeneity in sensor error and plasticity in daily response also returned no significant association, but encodings that accounted only for sensor noise found that heifers were significantly under represented among the most moderate time (and also the most variable) time budgets (McVey *et al.* 2021). Collectively, these results suggest that plasticity in daily time budgets might be obscuring links to age and productivity, but that this relationship cannot be brought into resolutions by accounting for systematic responses to environmental factors alone. This could occur if there were heterogeneity in the lag time between an environmental stressor and the behavioral response, or if there are more complex interactions occurring between transient environmental variables and more persistent internal biological states. If this in fact the case, then this result may suggest that an encoding strategy capturing only the variability in the response, and not systematic fluctuations, may in fact be a more effective means of pinning down this bivariate association.

Bivariate tree tests against health status, wherein cows were recorded as “sick” if any acute clinical health complication occurred within the observation window and “healthy” otherwise, recovered significant associations with encodings of daily time budgets created using both the unweighted and the ensemble-weighted Euclidean distance estimator. Visualizations of these associations are provided in Supplemental Materials. As with overall time budgets, these results

revealed that cows with no health complications were over represented among the most moderate time budgets while sick cows were over represented among cows with low time spent eating. Overall, however, these associations were stronger with encodings of overall time budgets than with daily time budgets. Bivariate associations were also evaluated against health records disaggregated into five broad categories: healthy, mammary infection and injury, hoof and leg lameness, infections of the reproductive tract, and digestive and metabolic diseases. Significant associations were recovered for all three dissimilarity estimators used to encode daily time budgets. Visualizations of these associations can again be found in Supplemental Materials. Again, these results mirror the bivariate patterns recovered for overall time budgets, with no clear evidence that temporal dynamics notably modified these associations. Figure 4 displays the strongest association recovered for the ensemble-weighted encoding, wherein the optimal metaparameter combination for the tensor mechanics encoding employed no re-weighting along the temporal axis. As with overall time budgets, mammary infections were over-represented among cows with consistently extremely high time spent eating, while digestive and metabolic diseases were over-represented among low eating time budgets (McVey *et al.* 2021). Infections of the reproductive tract were also again over-represented among the most moderate time budgets, but in this encoding with daily time budgets, it is more visually apparent that these same time budgets are also the most variable over time. As these observations occurred fairly early in the lactation period, it seems unlikely that temporary perturbations in behavioral patterns attributable to estrus behaviors could fully explain so much variability in eating times (Beauchemin 2018; Reith and Hoy 2018; Heringstad and Wethal 2022).

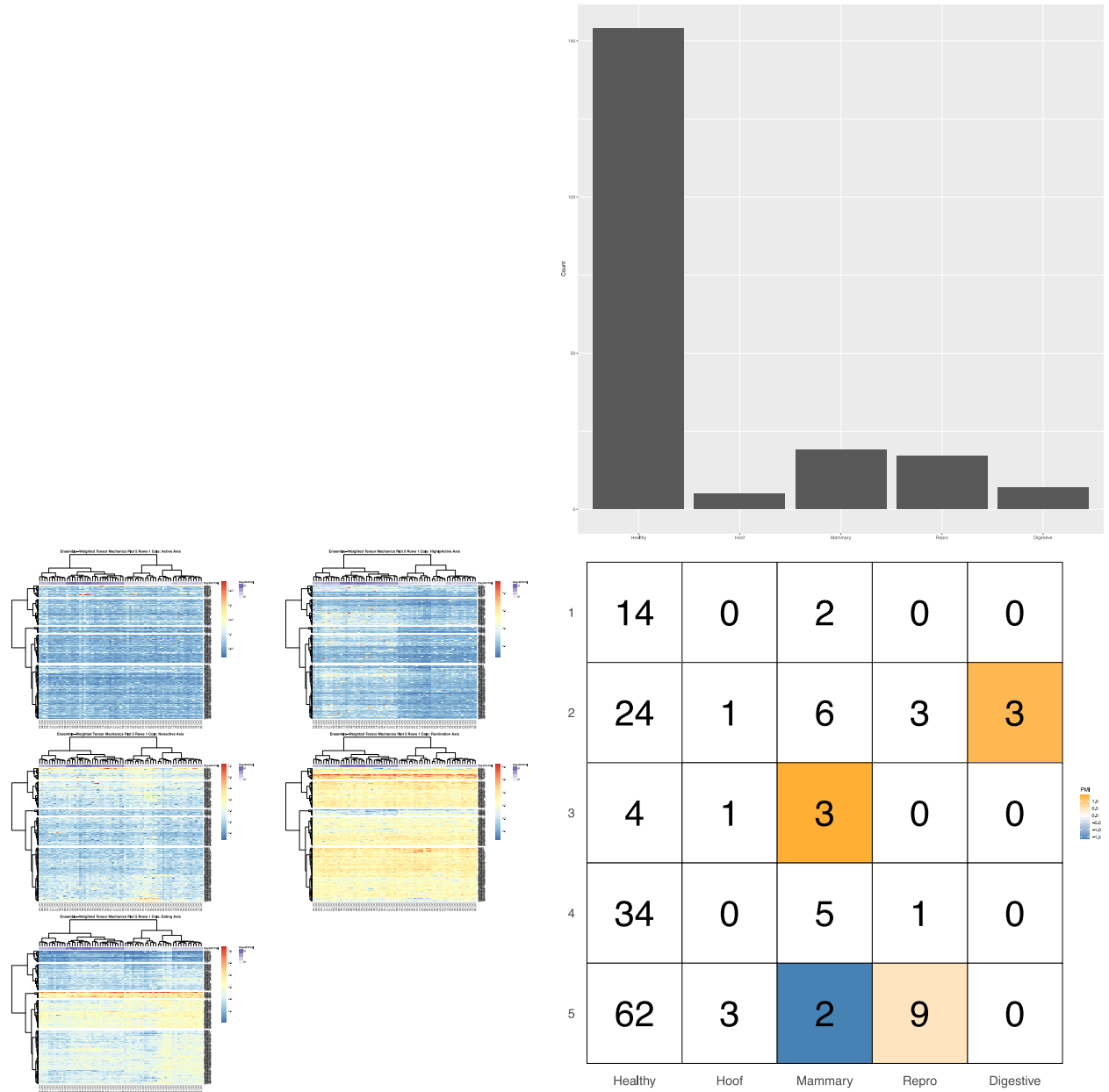


Figure 4: Visualization of significant bivariate relationship between health diagnoses (upper right) and daily time budgets encoded using the ensemble-weighted dissimilarity estimator (bottom left). The contingency table for this joint encoding (bottom right), colored using significant pointwise mutual information values, reveals that cows with digestive and metabolic complications were significantly over represented among time budgets with low time spent eating, while mammary infections were significantly over-represented among time budgets with extremely high

time spent eating. Cows with infections of the reproductive tract are significantly over represented among time budgets with moderate but also variable time spent eating. See Supplemental Materials for larger high-resolution versions of these images, wherein full details can be readily visualized.

Finally, significant bivariate associations were recovered between patterns in milking order and encodings of daily time budget using all three dissimilarity estimators. Visualizations of these bivariate patterns are provided for all encodings in Supplemental Materials. As a whole, the bivariate associations recovered against encodings created using overall time budget were stronger than for encodings using daily time budgets (McVey *et al.* 2021). Results for the unweighted Euclidean distance and the KLD distance largely mirrored patterns recovered using overall time budgets, with extremely low eating times being under represented among cows entering at the end of the queue and over represented among animals just ahead of these caboose animals. Visualizations of the bivariate association recovered against the ensemble-weighted Euclidean distance encoding are provided in Figure 5. Here the pattern between time budgets and cows that enter nearer the rear of the milking queue are brought into slightly higher resolution. This visualization shows that it is cows that consistently enter at the end of the queue, but not the last handful of cows, that are over represented among time budgets with moderate-to-high time spent eating. Time budgets with consistently low time spent eating, on the other hand, are significantly over represented among cows that routinely jumped between the middle and rear of the milking queue. Collectively, these results confirm that a later milking position does not preclude a cow from investing a significant amount of time at the feed bunk, though it should be noted that these results may not extrapolate to larger milking groups with longer waiting times in the milk parlor (Beggs *et al.* 2018). As systematic temporal patterns in daily time budgets do not, however, appear

to play a significant role in further distinguishing these bivariate patterns, future work might consider if these asynchronous perturbations in queuing patterns and time budgets might be more directly (temporally) linked within individual, and consider latent biological causes of these isolated behavioral responses.

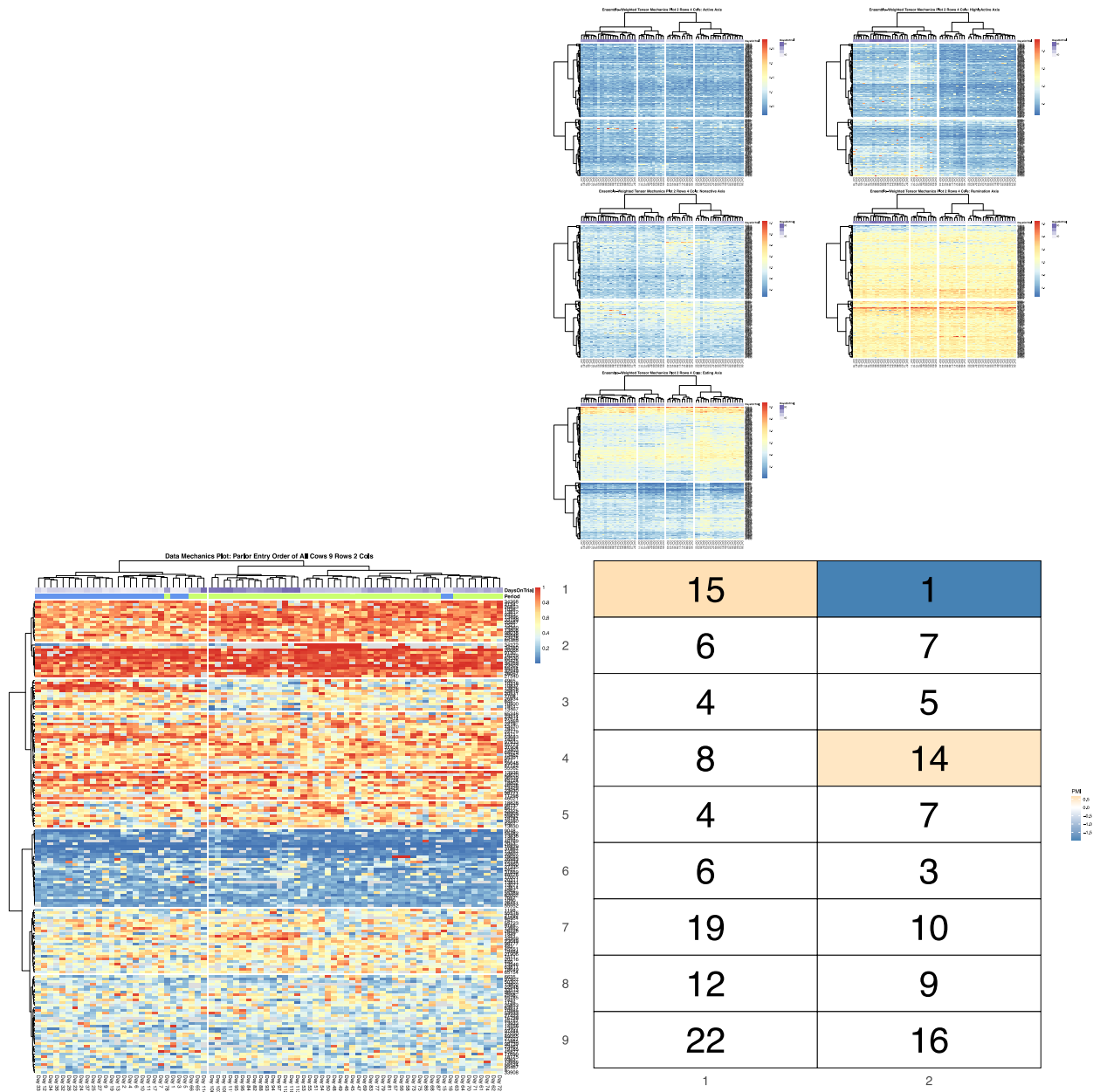


Figure 5: Visualization of significant bivariate relationship between patterns in milking order (bottom left) and daily time budgets encoded using the ensemble-weighted dissimilarity estimator (upper right). The contingency table for this joint encoding (bottom right), colored using significant pointwise mutual information values, reveals that consistently entered at the rear of the milking queue, but just ahead of the last handful of animals, were nearly all found to have time budgets with good time spent eating, while cows that were generally found nearer the middle of

the milking queue but routinely fell back towards the rear were significantly over represented among time budget with low eating times. See Supplemental Materials for larger high-resolution versions of these images, wherein full details can be readily visualized.

Conclusions

Drawing complete and holistic quantitative inferences is always challenging when working with multivariate data streams. When a temporal component is also present, information compression via data aggregation can simplify the analytical pipeline, but a considerable amount of behavioral complexity can be lost. By extending the data mechanics analytical pipeline to multidimensional use cases via a tensor-based approach, we've demonstrated that it is possible to simultaneously probe large PLF data streams for systematic changes in multiple behavioral axes over extended observation windows, even when the causes of those changes are not fully recorded. In analyses of CowManager rear tag accelerometer records available in the Organilac dataset, tensor mechanics analyses revealed that cows in this experimental herd were remarkably consistent in their overall time budgets. Never the less, these algorithms were still able to capture subtle shifts in the tradeoffs between the eating and nonactive axes and the eating and highly active axes as cows progressed in their lactations, even though such temporal patterns were only found amongst a subset of the herd. While the resulting encodings did not differ substantially from overall time budgets, these analyses helped to better visualize differences between animals in the relative plasticity of their behavioral responses. While tensor mechanics algorithms are designed to extract systematic changes over time by leveraging collective responses of animals to a shared environment, these results have served to highlight the need to look beyond the mean value and more fully explore the richness of PLF datasets, which can provide repeated measurements on individuals at a scale never before possible (Reith *et al.* 2018; Brito *et al.* 2020; Stygar *et al.* 2021).

Future work should consider how to further leverage these approaches to not only explore patterns in such plastic responses that can be extracted by leveraging group level information, but also variability attributable only to the individual, and subsequently any complex interactions that may occur between these systematic and fleeting perturbations in behavioral responses.

References

- Beauchemin, K.A., 2018. Invited review: Current perspectives on eating and rumination activity in dairy cows. *Journal of Dairy Science* 101, 4762–4784. <https://doi.org/10.3168/jds.2017-13706>
- Beggs, D.S., Jongman, E.C., Hemsworth, P.E., Fisher, A.D., 2018. Implications of prolonged milking time on time budgets and lying behavior of cows in large pasture-based dairy herds. *Journal of Dairy Science* 101, 10391–10397. <https://doi.org/10.3168/jds.2018-15049>
- Bikker, J.P., van Laar, H., Rump, P., Doorenbos, J., van Meurs, K., Griffioen, G.M., Dijkstra, J., 2014. Technical note: Evaluation of an ear-attached movement sensor to record cow feeding behavior and activity. *Journal of Dairy Science* 97, 2974–2979. <https://doi.org/10.3168/jds.2013-7560>
- Brito, L.F., Oliveira, H.R., McConn, B.R., Schinckel, A.P., Arrazola, A., Marchant-Forde, J.N., Johnson, J.S., 2020. Large-Scale Phenotyping of Livestock Welfare in Commercial Production Systems: A New Frontier in Animal Breeding. *Front. Genet.* 11, 793. <https://doi.org/10.3389/fgene.2020.00793>
- Guan, J., Hsieh, F., 2018. Coupling Geometry on Binary Bipartite Networks: Hypotheses Testing on Pattern Geometry and Nestedness. *Front. Appl. Math. Stat.* 4, 38. <https://doi.org/10.3389/fams.2018.00038>

- Heringstad, B., Wethal, K.B., 2022. Cow activity measurements can be used to define new fertility traits for use in genetic evaluation. JDS Communications S2666910222000941. <https://doi.org/10.3168/jdsc.2022-0251>
- Hsieh, F., Chou, E.P., Chen, T.-L., 2021. Mimicking Complexity of Structured Data Matrix's Information Content: Categorical Exploratory Data Analysis. Entropy 23, 594. <https://doi.org/10.3390/e23050594>
- Hsieh, F., Liu, S.-Y., Hsieh, Y.-C., McCowan, B., 2018. From patterned response dependency to structured covariate dependency: Entropy based categorical-pattern-matching. PLoS ONE 13, 1–28. <https://doi.org/10.1371/journal.pone.0198253>
- Huzzey, J.M., von Keyserlingk, M.A.G., Weary, D.M., 2005. Changes in Feeding, Drinking, and Standing Behavior of Dairy Cows During the Transition Period. Journal of Dairy Science 88, 2454–2461. [https://doi.org/10.3168/jds.S0022-0302\(05\)72923-4](https://doi.org/10.3168/jds.S0022-0302(05)72923-4)
- Kassambara, A., 2020. ggpubr: “ggplot2” Based Publication Ready Plots, 0.4.0. ed.
- Kolde, R., 2019. pheatmap: Pretty Heatmaps.
- Manriquez, D., Chen, L., Melendez, P., Pinedo, P., 2019. The effect of an organic rumen-protected fat supplement on performance, metabolic status, and health of dairy cows. BMC Vet Res 15, 1–14. <https://doi.org/10.1186/s12917-019-2199-8>
- Manriquez, D., Chen, L., Albornoz, G., Velez, J., Pinedo, P., 2018. Case Study: Assessment of human-conditioned sorting behavior in dairy cows in farm research trials. The Professional Animal Scientist 34, 664–670. <https://doi.org/10.15232/pas.2018-01749>
- McCowan, B., Beisner, B., Bliss-Moreau, E., Vandeleest, J., Jin, J., Hannibal, D., Hsieh, F., 2016. Connections Matter: Social Networks and Lifespan Health in Primate Translational Models. Front. Psychol. 7. <https://doi.org/10.3389/fpsyg.2016.00433>

- McVey, C., Hsieh, F., Manriquez, D., Pinedo, P., Horback, K., 2021. Livestock Informatics Toolkit: A Case Study in Visually Characterizing Complex Behavioral Patterns across Multiple Sensor Platforms, Using Novel Unsupervised Machine Learning and Information Theoretic Approaches. *Sensors* 22, 1. <https://doi.org/10.3390/s22010001>
- McVey, C., Hsieh, F., Manriquez, D., Pinedo, P., Horback, K., 2020. Mind the Queue: A Case Study in Visualizing Heterogeneous Behavioral Patterns in Livestock Sensor Data Using Unsupervised Machine Learning Techniques. *Front. Vet. Sci.* 7, 523. <https://doi.org/10.3389/fvets.2020.00523>
- Miedema, H.M., Cockram, M.S., Dwyer, C.M., Macrae, A.I., 2011. Changes in the behaviour of dairy cows during the 24h before normal calving compared with behaviour during late pregnancy. *Applied Animal Behaviour Science* 131, 8–14. <https://doi.org/10.1016/j.applanim.2011.01.012>
- Pereira, G.M., Heins, B.J., Endres, M.I., 2018. Technical note: Validation of an ear-tag accelerometer sensor to determine rumination, eating, and activity behaviors of grazing dairy cattle. *Journal of Dairy Science* 101, 2492–2495. <https://doi.org/10.3168/jds.2016-12534>
- Reith, S., Hoy, S., 2018. Review: Behavioral signs of estrus and the potential of fully automated systems for detection of estrus in dairy cattle. *Animal* 12, 398–407. <https://doi.org/10.1017/S1751731117001975>
- Stygar, A.H., Gómez, Y., Berteselli, G.V., Dalla Costa, E., Canali, E., Niemi, J.K., Llonch, P., Pastell, M., 2021. A Systematic Review on Commercially Available and Validated Sensor Technologies for Welfare Assessment of Dairy Cattle. *Front. Vet. Sci.* 8, 634338. <https://doi.org/10.3389/fvets.2021.634338>

Tucker, C.B., Jensen, M.B., de Passillé, A.M., Hänninen, L., Rushen, J., 2021. Invited review: Lying time and the welfare of dairy cows. *Journal of Dairy Science* 104, 20–46.

<https://doi.org/10.3168/jds.2019-18074>

Valletta, J.J., Torney, C., Kings, M., Thornton, A., Madden, J., 2017. Applications of machine learning in animal behaviour studies. *Animal Behaviour* 124, 203–220.

<https://doi.org/10.1016/j.anbehav.2016.12.005>

CHAPTER 4

Model-free approaches to quantifying plasticity and resiliency in PLF data streams using unsupervised machine learning and information theoretic approaches

Introduction

Farmers have used data to breed better cows for generations – long before the term “big data” was originally coined. Milk yield is a key factor in determining farm profitability, and so it has become a primary target for breed improvement programs. Subsequently, a substantial volume of research has been dedicated to optimally quantify the characteristics of projected lactation curves. As a result of this work, modern dairy cattle are on average more than twice as productive today as they were when the first herd improvement program were put into place fifty years ago, making dairy farms more efficient and more profitable (Brito *et al.* 2021). Unfortunately, with these gains in milk productivity have come losses in other traits that are inherently more difficult to quantify such as fertility, health outcomes, and longevity (Brito *et al.* 2021). Over the last decade, with the help of expanded data collection efforts and genomic selection technology, some genetic progress has already been realized (Migior *et al.* 2021). Health and fertility outcomes are, however, often time consuming and expensive to quantify, and so there are fundamental inefficiencies in collecting data and developing selection indices for the wide range of health complications that dairy cattle may fall victim to during their productive lives (Brito *et al.* 2021; Migior *et al.* 2021).

In light of these challenges, there is growing interest in developing methods to select for cattle that are more generally robust across a broader range of contexts (Scheffer *et al.* 2018; Berghof *et al.* 2019; Brito *et al.* 2021). Resiliency is broadly defined as the ability of an animal to cope with fluctuations in their environment with minimal disruption to their internal state of homeostasis (Scheffer *et al.* 2018). It is hoped that breeding for such robustness would not only serve to indirectly improve a range of health and fertility outcomes, but also help to prepare cattle for the uncertain future that is promised by climate change (Berghof *et al.* 2019; Brito *et al.* 2021). Resilience, however, is a latent trait that cannot be directly measured. One strategy that has been proposed to assess its value is to analyze repeatedly measured traits under variable environmental conditions. In conventional scientific literature, such data would typically be analyzed as a reaction norm, wherein the changes in a response trait would be directly compared against changes in an environmental gradient, and the features of the resulting curve extracted via a mixed effect model to quantify differences in relative plasticity between individuals (Ghalambor *et al.* 2010; Dingemanse and Araya-Ajoy 2015). Unfortunately, this algorithmic framework can rapidly become impractical beyond the realm of controlled experimental settings, as the number of variables that can fluctuate in a working farm environment can rapidly exceed the capacity to record such environmental measures, none the less model their joint impact (Garcia-Baccino *et al.* 2021).

While environmental variables are required to capture a reaction norm using estimators of first moment (centrality or expected value), estimators of higher moments (variance, skew, kurtosis) may be used to capture this information in an unsupervised context where only the marginal distribution of response traits is available (Scheffer *et al.* 2018; Poppe *et al.* 2020). In this framework, the more plastic an animal's response to fluctuations in the environment, the greater the variability in the response. While this analytical approach has the advantage of

circumventing the need to record environmental variables, several logistical barriers can be anticipated. The first is that measures of higher moments need progressively larger sample sizes to bring into resolution more subtle distributional features. With the widespread adoption of PLF sensor technologies, however, acquiring large samples of repeated measurements over extended observation windows can be done cheaply and with minimal effort (Berghof *et al.* 2019; Poppe *et al.* 20202). The second logistical barrier to implementing to this approach is that estimators of higher moments are often quite sensitive to anomalous records and heterogenous measurement error (Zoubir *et al.* 2012). Unfortunately, both of these stochastic features are commonly found in PLF data streams. Subsequently, disentangling variability due to individual differences in plasticity and to sensor noise poses a significant analytical challenge that must be addressed if these technologies are ultimately to be brought to bear in developing better measures of robustness.

One indicator of resilience that has been proposed as a potential selection target in dairy cattle is daily milk yield, or more specifically, the variability in milk yield records about the projected lactation curve (Elgersma *et al.* 2018; Berghof *et al.* 2019; Poppe *et al.* 2020; Moncur *et al.* 2021; Abdelkrim *et al.* 2021). While not a direct measure of health or fertility, such data streams are ubiquitous among dairies with RFID-equipped parlors, and so any metric derived from such data has considerable potential to realize genetic progress in selecting for greater resiliency. This is particularly true on commercial dairies that have not implemented intensive genetic improvement programs, where greater investments in genomic selection technologies or expanded health records may offer minimal returns. Exploring the utility of daily milk yield estimates from archival lactation records has been the focus of a number of research paper published over the last five years. These studies have found estimators of higher moments in residual milk yield to be low to moderately heritable, with variance being genetically correlated to health, fertility, and longevity traits used in conventional breed improvement programs (Elgersma *et al.* 2018; Poppe *et al.* 2020;

Poppe *et al.* 2021; Moncur *et al.* 2021). As such studies have relied on genetically evaluated herds, these records may be cleaner than those found on commercial dairies who perhaps stand to benefit the most from this type of analysis.

The goal of this study will be to explore a wider range of information compression techniques to quantify plasticity in daily yield records, absent any information about environmental conditions and with minimal preprocessing of data. The performance of these algorithms will focus on 1) the resilience of these estimators in the face of complex error structures, and 2) their ability to produce wholistic encodings of the geometric features of marginal residual yield curves that can be readily related back to other indicators of biological and behavioral resiliency.

Description of Data

To explore the efficacy of these analytical strategies when faced with the complexity of lactation data acquired from a working farm environment, data was repurposed from a feed trial assessing the effect of an organic fat supplement on cow health and fertility (Manriquez *et al.* 2019). All animal handling and experimental protocols were approved by the Colorado State University Institution of Animal Care and Use Committee (Protocol ID: 16-6704AA). The study ran from January through July in 2017 on a USDA Certified Organic dairy in Northern Colorado. A total of 200 cows were enrolled over a 1.5-month period into a mixed-parity herd of animals with predominantly Holstein genetics. Cows were maintained in a closed herd in an open-sided free-stall barn that was stocked at roughly half capacity with respect to both feed bunk spaces and stalls. Cows had free access to an adjacent outdoor dry lot while in their home pen, and beginning in April were moved onto pasture at night, to comply with organic grazing standards. Cows had free access to TMR between milkings while in their home pen, and were head locked each morning

to facilitate data collection and daily health checks. For more details on feed trial protocols, the housing environment, and the day to day management of animals, see Manriquez *et al.* (2018) and Manriquez *et al.* (2019).

Cows were milked three times daily in a DelPro™ rotary parlor (DeLaval, Tumba, Sweden). Daily lactation records were available for each study cow over roughly the first 150 days, the exact observation window being contingent on an individual's enrollment date, with no animals recorded through less than 110 and no more than 172 days in milk (see Supplemental Materials). Animals with an excessive number of missing records, ostensibly attributable either to extended removal from the milking herd due to acute health complications or chronic registration failures in the RFID reader, were excluded from analyses. This resulted in the removal of 17 animals from the dataset, of which six had no recorded health events, and the remainder diagnosed with a diverse range of transition diseases (see Supplemental Materials). This left a total of 185 cows with at least 75 days of recorded milkings. Additionally, all records were dropped around daylight savings due to apparent errors in milking date attributions. See Supplemental Materials for complete details on all data scrubbing and exclusion procedures.

Two additional PLF data streams, the subjects of previously published analytical case studies, were also available for this dataset. First, milking order, or the sequence in which cows enter the parlor to be milked, is automatically recorded as metadata in all modern RFID-equipped milking systems. For each morning milking, the single-file entry positions of cows were extracted from raw RFID milk logs exported from the parlor software. A total of 80 milk order records—26 recorded while cows remained overnight in a free-stall barn, and 54 following the transition to overnight access to spring pasture—were used to create discrete encodings for parlor entry patterns via data mechanics clustering. See McVey *et al.* 2021 for further analytical details. Additionally, animals enrolled in this feed trial were also fitted with a CowManager ear tag accelerometer (Agis

Automatisering BV, Harmelen, The Netherlands). This commercial sensor platform provides hourly time budget estimates for total time (min) engaged in five mutually exclusive discrete behaviors – eating, rumination, non-activity, activity, and high activity. Time budget data was collected on all animals for a contiguous period of 65 days (1560 h). The observation window began on February 17, shortly after trial enrollment was completed, and ended on April 23, when the grazing season commenced and cows were moved overnight beyond the range of the receiver antennae. Overall time budgets were calculated from this dataset and encoded via both a precision and a plasticity-penalized ensemble-weighted distance clustering framework. See McVey *et al.* 2021 for further analytical details. The dendrograms summarizing patterns in the distribution of cow entry order and overall time budget data, and subsequent heatmap visualizations, will here be subjected to further analysis without modifications to the previously reported encodings.

Fitting Lactation Curves

In order to calculate estimates of residual milk yields, an appropriate lactation curve first had to be fit to each cow's lactation records. In previous work, this has been accomplished via a number of robust regression techniques (Elgersma *et al.* 2018; Poppe *et al.* 2020; Poppe *et al.* 2021; Abdelkrim *et al.* 2021). Due to the high rate of outliers and missing milkings present in this dataset, it was determined that a model-based approach was needed to provide additional structure where observations for some animals became sparse. Subsequently, a Wilmink lactation curve fit via quantile regression was deemed the most appropriate approach (Wilmink 1987; Poppe *et al.* 2020). In preliminary analyses, however, it was discovered that the rate of milking outliers, some of which were presumably caused not only by low yield days but also by premature claw kickoffs, varied considerably between cows. This variation did not appear to unduly influence the shape of

the estimated curve, but was found to bias downward the final projected curve for high-outlier animals. Such biases in the estimated curve would in turn shift the center of the subsequent residual yield distribution, making it more difficult to directly compare the value of the de-trended yield observations.

To remedy this issue, outliers needed to be thresholded out without the aid of milking duration records. This was done by first fitting a preliminary Wilmink lactation curve to the 90th quantile ($\tau = 0.90$) of the data using the *quantreg* package in R (Koenker 2018). In order to also accommodate variations in the time of peak yield observed in this herd, the Wilmink's curve was fit on a grid of k values ranging from 0.01 to 0.6, with the best fitting model retained (Wilmink 1987). Once this preliminary fit was obtained, milkings were flagged as outliers if they were greater than 10% above or less than 20% below the predicted yield value for that observation date, as visualized in Figure 1. Visualizations of thresholding curves for the complete dataset are provided in Supplemental Materials. Once outliers were excluded, the Wilmink's curve was refit to the 70th quantile ($\tau = 0.70$) of the data, to allow for skew among inlier values (Poppe *et al.* 2020). Visualizations of the finalized curves fit to each cow are provided in Supplemental Materials.

From these fitted models, data matrices were compiled across all animals for optimal lactation curve parameters and projected curve values. For any observation dates at the extremes of the curve with projected values below 1, these estimates were deemed to be unduly influenced by extrapolation error and artificially set to 1. Additionally, due to extrapolation errors and high rates of missing observations, all yield records from the first week of lactation were dropped from subsequent analyses (Poppe *et al.* 2020). Finally, normalized residual yield estimates were here calculated as the difference between observed and projected yield, rescaled by the value of the projected yield. This normalization was applied to circumvent issues with heterogeneity in raw

residual yield estimates attributable to the overall level of milk production observed in previous work (Elgersma *et al.* 2018; Poppe *et al.* 2020), thereby allowing the geometric features of the resulting marginal residual yield distributions to be directly compared across cows independent of individual differences in productivity

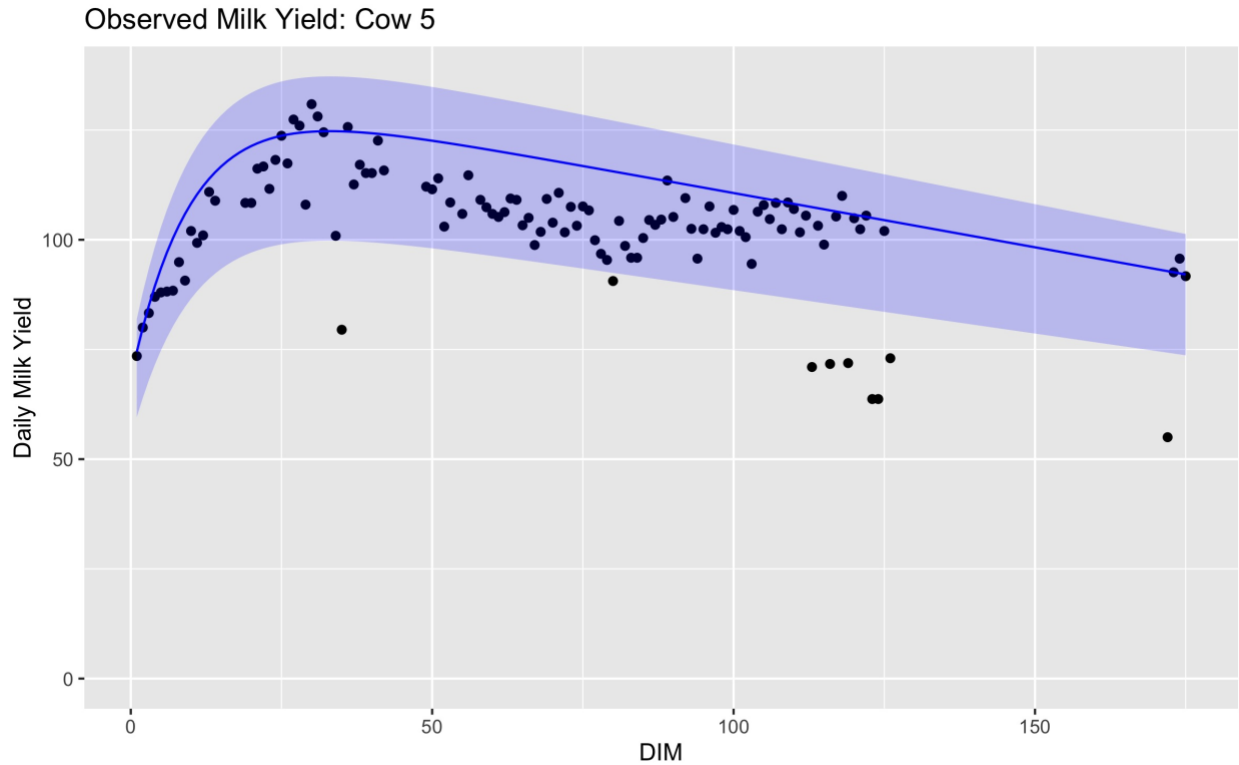


Figure 1: Example of a preliminary thresholding plot. The blue line represents the Wilmink's lactation curve fit to the 90th quantile of recorded daily yield. All points that fall within the blue bands are designated as “inliers” and used to fit the final lactation curve.

Creating an Ensemble of Data Mimicries using Simulated Noise

In order to reliably capture individual differences in the marginal distributions of residual milk yields, any analytical framework developed to analyze these estimates should strive to distinguish the intrinsic variability in this biological response from measurement error. There are, however, multiple co-dependent sources of noise interacting within this data stream. Such complex error structures are difficult to accommodate within frequentist statistical frameworks, where typically only a single residual error distribution may be assumed. Multiple sources of nested error might be accommodated within a Bayesian framework, but such a model would necessitate a number of parametric assumptions to be made, and could struggle to converge if error structures become too deeply layered (Christensen *et al.* 2011; McVey *et al.* 2021). To circumvent such logistical challenges, a simulation-based approach was here utilized to capture the complexity of the error structures in this data stream while employing the fewest parametric assumptions possible (McVey *et al.* 2021).

In previous research with the CowManager ear tag accelerometer records also found within this dataset, we've demonstrated that simple nonparametric and minimally parametric approaches to stochastic simulation can easily be strung together to effectively mimic the multi-faceted and often non-independent sources of errors commonly found in PLF data streams (McVey *et al.* 2021). This simulated error can then be re-infused into the observed data in order to create an ensemble of synthetic datasets that mimic both the signal and noise structures of the original data (Hsieh *et al.* 2021). This ensemble of data mimicries can then be utilized within a hierarchical clustering framework, not only to provide better dissimilarity estimates, which in turn provides a more authentic dendrogram representation of the underlying distribution of the data within the feature space, but also to prune the resulting clustering tree to ensure that all branches represent biological signals that are statistically distinguishable from noise (McVey *et al.* 2021).

To mimic the noise present in the marginal distributions of residual yield estimates for each cow, three sources of error were simulated. The first source of error was measurement error attributable to the milk reader itself. This was simulated by uniformly jittering the observed milk yield by $\pm 2\%$ of the observed value, which is the accepted range of error for a calibrated parlor (Lefcourt 1999). The second source of error is bias that is attributable to uncertainty in the fitted lactation curve used to calculate residual yield value. Instability in these estimated curves were here simulated by implementing a jackknife subsampling routine on the jittered observed values before refitting the lactation curve using the same regression routine as implemented on the observed values (Givens and Hoeting 2013). For each jackknife refitting, only 90% of the original sample was retained. To ensure that this jackknifing routine did not leave sections of the lactation curve under-sample, observations were here subsampled conditionally within the subperiods for early lactation ($\text{DIM} \leq 30$), peak lactation ($30 < \text{DIM} \leq 60$), and late lactation ($\text{DIM} > 60$). Using this simulation routine, a total of 500 simulated residual yield datasets were generated from the original observed lactation records. The third and final source of error impacting the geometry of the marginal residual yield distribution was then sampling error itself. To mimic this source of instability in distributional properties, a bootstrap resampling routine was here implemented to create a tensor of 500 simulated marginal yield distributions from the jackknife resampled residual yield estimates. In order to capture not only the impact of outliers but also differences in the sampling density between cows, each bootstrap distribution was resampled with the number of observations recorded in the original sample with no missing values included (Givens and Hoeting 2013).

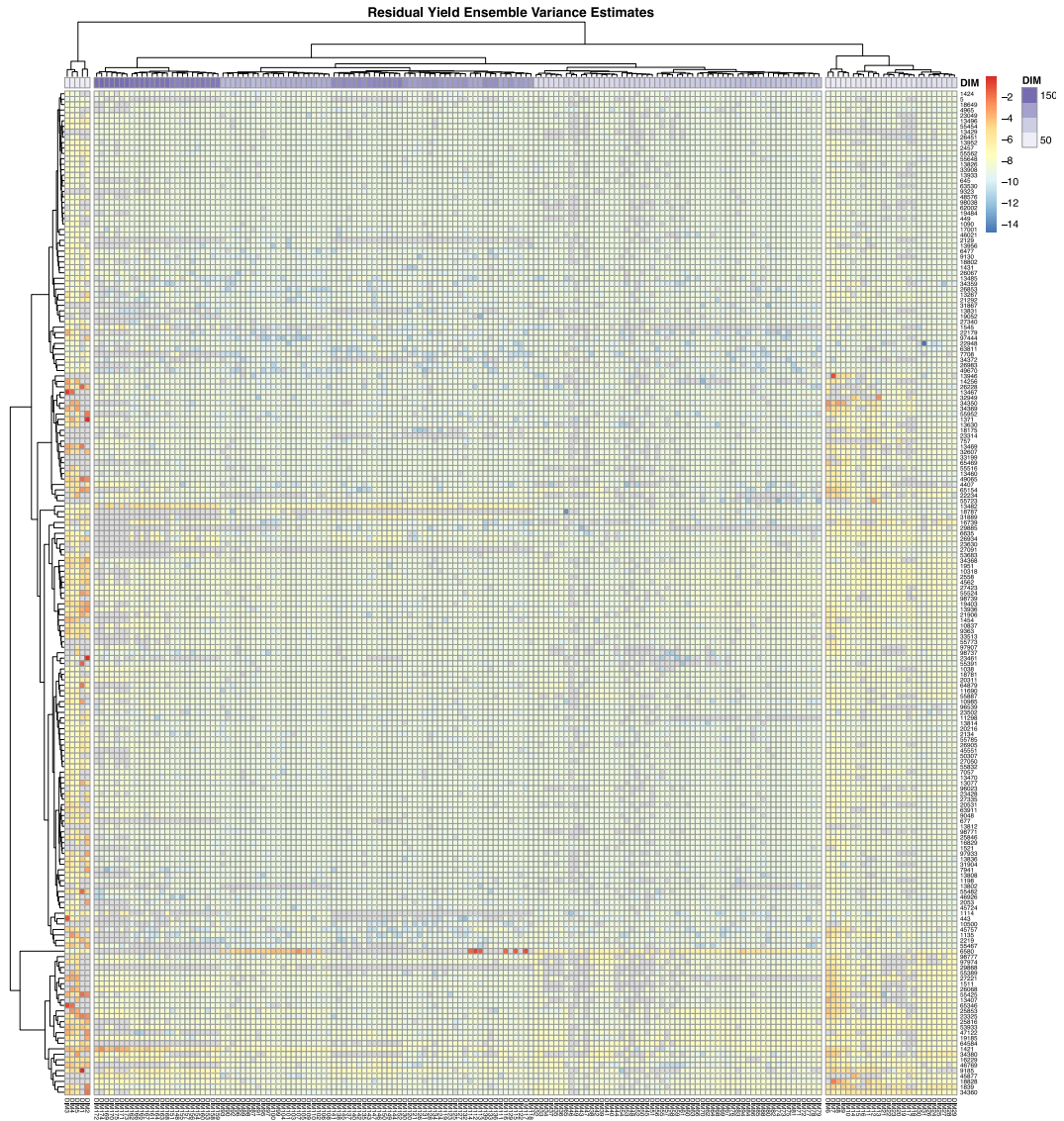


Figure 2: Heatmap visualization of the log-ensemble variance estimates calculated over all simulations for each observation. Individual cows are indexed on the rows and days in milk are indexed on the columns. This visualization reveals that ensemble variance estimates are relatively high for many animals during the first few days in lactation, which is likely attributable to extrapolation error due to the intrinsic uncertainty in fit of the lactation curve at the boundary of the domain. For some cows, ensemble variance may remain somewhat elevated through the transition period, but most observations during the persistence period are estimated with a relatively high degree of confidence due to good stability in the fitted lactation curve.

From this tensor of simulated residual yield values, the reliability of each lactation record was estimated using two distinct information compression strategies. First, ensemble variance estimates were extracted for each observation of each cow over all simulation indices. These values are visualized as a heatmap in Figure 2, revealing that residual estimates were quite unstable for approximately the first five days of lactation for most animals. This result is not surprising given that such values at the edges of the domain may suffer from extrapolation error – an issue often compounded by patchy data during the first week in milk (Poppe et al. 2020). This instability extended into the observation window approaching peak lactation for some animals, but residual estimates for most cows become quite stable during the persistence phase of the curve. The second information compression strategy employed a quality control framework, and simply recorded the proportion of simulated residual yield values that fell within ± 3 units of standard error from the observed residual, where standard error was estimated as the median of the ensemble variance estimates across all time points for a given cow.

Information Compression

Equipped with a reasonable mimicry of the various sources of noise and bias in this dataset, the next step in this analytical pipeline was to compress the information contained over many replications of lactation records down into an encoding that could succinctly convey details about the geometric features of each animal’s marginal residual yield distribution.

Summary Statistics

In previous publications, standard summary statistics – mean, variance, skew, and kurtosis – have been used to capture the key geometric features of residual yield distributions that vary between individuals (Elgersma et al. 2018; Poppe et al. 2020; Poppe et al. 2021; Kok et al. 2021;

Moncur *et al.* 2021). Therefore, the first strategy employed to create an encoding of such information was to extract these estimates for each cow from the observed lactation records and also from each of the simulated datasets. Ensemble variance estimates were then calculated for each estimator for each cow using the simulated residual yield data, and these values subsequently used to calculate an ensemble-weighted pairwise dissimilarity matrix, wherein the L2 norm calculated across the summary statistics for the observed residual yield data were normalized by the variance of the corresponding terms in the data mimicry. This ensemble weighting served not only to normalize the scale of each estimator, but also to penalize the relative stability of these calculated values so that estimates that were unstable when subjected to simulated noise would not receive undue influence in the final clustering (McVey *et al.* 2021). Hierarchical clustering using Ward's D2 method was then applied to this dissimilarity matrix to create a dendrogram representation of how distributional features varied across this sample herd. The ensemble mimicry of summary statistics was then used again in the *cutreeEnsemble* algorithm in order to adaptively prune the resulting dendrogram, such that all clusters in the final encoding represented heterogenous patterns in the observed data that could be statistically distinguished from the simulated noise in the dataset (McVey *et al.* 2021).

The results of this unsupervised clustering pipeline are visualized via a heatmap in Figure 3, wherein the estimators are centered and scaled to provide more direct visual comparisons of relative scale. A total of eight statistically validated clusters were recovered using the *cutreeEnsemble* utility, which indicates that there is a fair amount of heterogeneity in the distributional characteristics of these marginal residual yield distributions that can be recovered from this noisy data stream. Visual appraisal of these clusters reveals that estimates of mean seem to have the largest influence on the structure of this dendrogram. This result may in part be explained by the ensemble-reweighting used to generate this clustering tree, as comparisons of

observed against simulated estimator values confirmed that mean estimates were more stable than those of higher moments when subjected to simulated noise (see Supplemental Materials), and so were given relatively more weight in the dissimilarity estimates. Visual inspection of this heatmap also reveals, however, that estimates of mean and variance are here quite negatively correlated. For highly skewed data, this result is not necessarily surprising, except that we also see here that mean estimates are not strictly redundant relative to estimates of higher moments. One potential explanation for this result might then be that the multimodality present in this data due to kickoffs could be confounding the information compression provided by these summary statistics, which are generally better suited to simpler unimodal distributions (Ram *et al.* 2017).

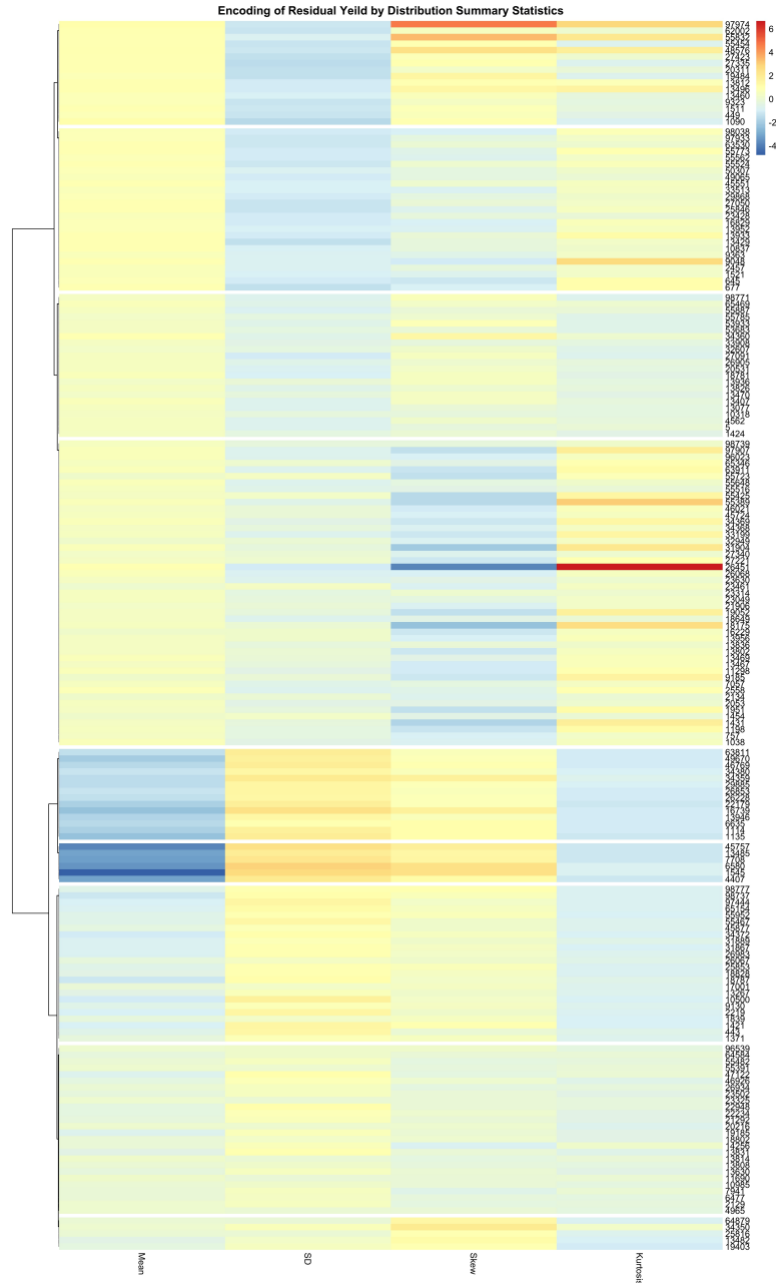


Figure 3: Heatmap visualization of encoding of marginal residual yield distributions based on summary statistics. Cows are arranged on the row axis and summary statistics – mean, standard deviation, skew, and kurtosis – are arranged along the column axis, having been centered and scaled to facilitate comparisons of relative magnitude. Mean and variance estimates are negatively correlated and drive the structure of the dendrogram, with skew and kurtosis providing less reliable but also less redundant information about underlying distributional features.

Earth Movers Distance

In the previous encoding scheme, information compression occurred not only in the hierarchical clustering algorithm but also in transforming the lactation records into a set of summary statistics prior to the calculation of the pairwise dissimilarity matrix. In order to streamline these computations and isolate all information compression within the clustering routine itself, a pairwise dissimilarity estimator designed to directly compare distributional characteristics between cows was next considered to create an encoding of residual yield records (Ram *et al.* 2017). Earth Movers Distance, also known as Wasserstein distance, works exactly as its name implies. For any two cows, each residual yield observation can be visualized as a box that sits along a number line at its recorded value. To compare the distribution of these values between animals, Earth Movers Distance calculates the amount of “work” that would be required to push these boxes around until the two distributions became identical. For univariate distributions, these calculations can be performed quickly in $O(n)$ time (i.e. - run time scales linearly with sample size). By default “work” scales linearly with distance and each observation receives equal weighting; however, weights can be modified to adjust for difference in sampling rates and also to reduce the influence that specific observations may have on the final dissimilarity estimate (Schuhmacher *et al.* 2019). Two dissimilarity matrices were here computed using Earth Movers Distance. The first applied no weighting to the lactation records except where necessary to adjust for differences in the total number of records between distributions. In the second, however, each lactation record was reweighted using the previously calculated ensemble weighting values.

Dendrograms were created for both dissimilarity matrices via hierarchical clustering using Ward’s D2 method. Prior to extracting encoded values, both clustering trees were passed through the *cutreeEnsemble* algorithm to prune all branches that could not be statistically distinguished from the simulated noise contained within the ensemble of data mimics (McVey *et al.* 2021). A

total of eight validated clusters were recovered for the unweighted Earth Movers Distance encoding and nine validated clusters were recovered for the weighted Earth Movers Distance encodings. To visualize these results, heatmaps were created such that each row represented the empirical histogram of residual yield records for each individual cow, wherein each cell in the heatmap was colored to represent the proportion of residual yield records for a given cows that fell within a specified range of normalized residual values specified across all cows. These global binning rules were here empirically determined by fitting a possibly gapped histogram over all residual records for all cows in order to extract dynamic bin widths that better captured the complex geometric features of these marginal residual yield distributions (Hsieh and Roy 2018).

A contingency count table, created using the *compareEncoding* utility (McVey *et al.* 2021), provides a visual comparison of the weighted and unweighted encodings in Figure 4. A cursory scan of these results reveal that multimodality caused by kickoffs is clearly captured by this estimator. In fact, differences in the rate of yield outliers between cows seems to be the most influential factor driving differences between clusters of marginal residual yield distributions. The contingency count table reveals that there are few appreciable differences between the weighted and unweighted encodings. Among residual yield distributions with low to moderate rates of yield outliers, there is some ambiguity in the cutoffs between clusters. This may suggest either that distributional characteristics of yield records found closer to the projected yield curve are not sufficiently distinct to create clear cutoff, or perhaps this dissimilarity estimator is undervaluing the information in these more subtle features. Further, the additional validated cluster recovered in the weighted encoding served only to distinguish between curves with moderate rates of outlier records, with both encoding schemes showing similar degrees of granularity amongst the low-outlier animals.

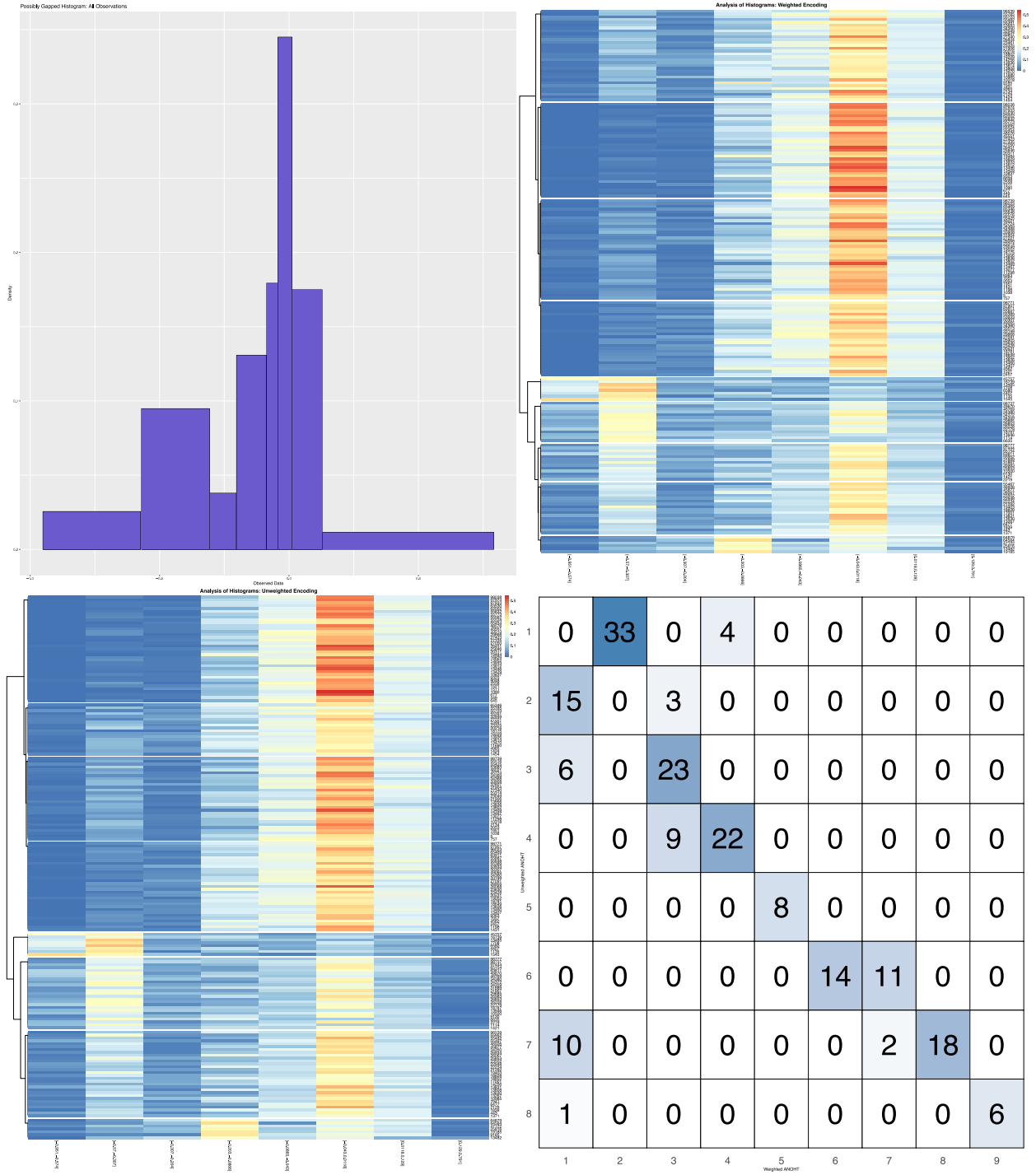


Figure 4: Comparison of marginal residual yield encodings created using the weighted and unweighted Earth Movers Distance. In both heatmap visualizations of either encoding, cows are indexed along the rows, and bins of the possibly gapped histogram (top left) are indexed along the

columns, such that each cell represents the proportion of recorded observation for a given cow that fall within a given histogram bin. The contingency table (bottom right) compares the joint encodings, with the unweighted EM encoding (bottom left) arranged along the row axis, and the weighted EM encoding (top right) arranged along the column axis, wherein clusters for both heatmaps are numbered from top to bottom. Weighting the residual yield values has here not modified the cluster dendrogram significantly, with the encodings differing only slightly in the cutoffs between cows with low rates of outlier yield observations. See Supplemental Materials for larger high-resolution versions of these images, wherein full details can be readily visualized.

Analysis of Histograms

In both of the previous encodings, the dendrograms derived to describe the geometric features of the observed marginal residual yield distributions seemed to place more emphasis on the rates of outlier yield records than the skew in yields from records closer to the projected yield curve. In this application, however, the biological significance of these outlier records, many of which may simply be attributed to premature claw kickoffs (Pope *et al.* 2020), may not scale linearly with the subsequent impact on the empirical measurement. Put more simply: is a 90% reduction in yield on a given day of observation due to claw kickoffs, which might be attributed to a fleeting behavioral disturbance during milking, really three times as biologically meaningful as a 30% drop in milk production on a day where a cow is milked out completely but yield is impaired by an underlying metabolic disturbance? While converting from a continuous to a discrete scale always results in a loss of information (Agresti 2013), it also provides a means to more easily and flexibly accommodate nonlinearity in the response space. In this application, it will not only mitigate any undue leverage that outlier records are exerting on the final encoding,

but also promote a more natural state-space interpretation of these kickoff events that, as a behavioral response, are arguably more fundamentally discrete in nature.

To capture geometric features of marginal residual yield distributions on a discrete scale, an Analysis of Histograms (ANOHT) approach was employed (Hsieh and Roy 2018). Possibly Gapped Histograms were fit globally across all observed lactation records for all cows to establish adaptive binning rules that best represented the underlying geometry of this data set. Using these binning rules, the normalized residual yield records were transformed into histogram matrices, wherein each row corresponded to a given cow, each column corresponded to a histogram bin, and thus each cell represented the proportion of an animal's lactation records that fell within a given range. Histogram matrices were computed for the observed and simulated data. Kullback-Leibler Divergence (KLD) Distance, or the sum of the relative entropies, was used to create dissimilarity matrices contrasting these discretized residual yield distributions between cows (MacKay 2003), and animals were subsequently clustered using Ward's D2 method. The *cutreeEnsemble* utility was then used to prune these trees using the ensemble of simulated histogram matrices (McVey *et al.* 2021). Finally, because results of the ANOHT clustering are dependent on the granularity of the discretized histogram used, this clustering routine and the resulting encoding were returned on a grid of bin count values from two to ten, wherein the results for each bin count were independently pruned.

Results of the ANOHT encoding returned encodings with between 7-10 validated clusters for all bin count parameters considered, with most encodings returning 7 or 8 clusters. This suggests that an undue amount of information was not lost in the preliminary discretization step. A visualization of the ANOHT encoding with seven discrete bins, which returned the finest validated encoding after the *cutreeEnsemble* pruning, is provided in Figure 5. Clusters in this encoding are perhaps marginally more homogenous than visualizations for the Earth Movers

Distance encodings, but in comparing the results for this discrete distribution dissimilarity estimator against those produced with the continuous Earth Movers dissimilarity estimates, the dendrogram structures are not remarkably distinct (see Supplemental Materials). Amongst animals with relatively few outlier yield values, the relative skew in yield records closer to the projective curve again only produces three clusters that can be statistically distinguished from the simulated noise in these measurements. A finer encoding is still returned only by further distinguishing between rates of outliers amongst the remaining animals, as with the continuous distributional dissimilarity value. This result may suggest that it is not the leverage on the continuous scale leading to the finer stratification in these records, but could instead reflect intrinsic differences in the relative confidence in measurements at which distinct distributional features can be resolved. In other words, in a system with both discrete (state-space) and continuous features, it is much easier to establish that an outlier is indeed an outlier relative to measurement error than it is to distinguish between levels of skew in a noisy data stream.

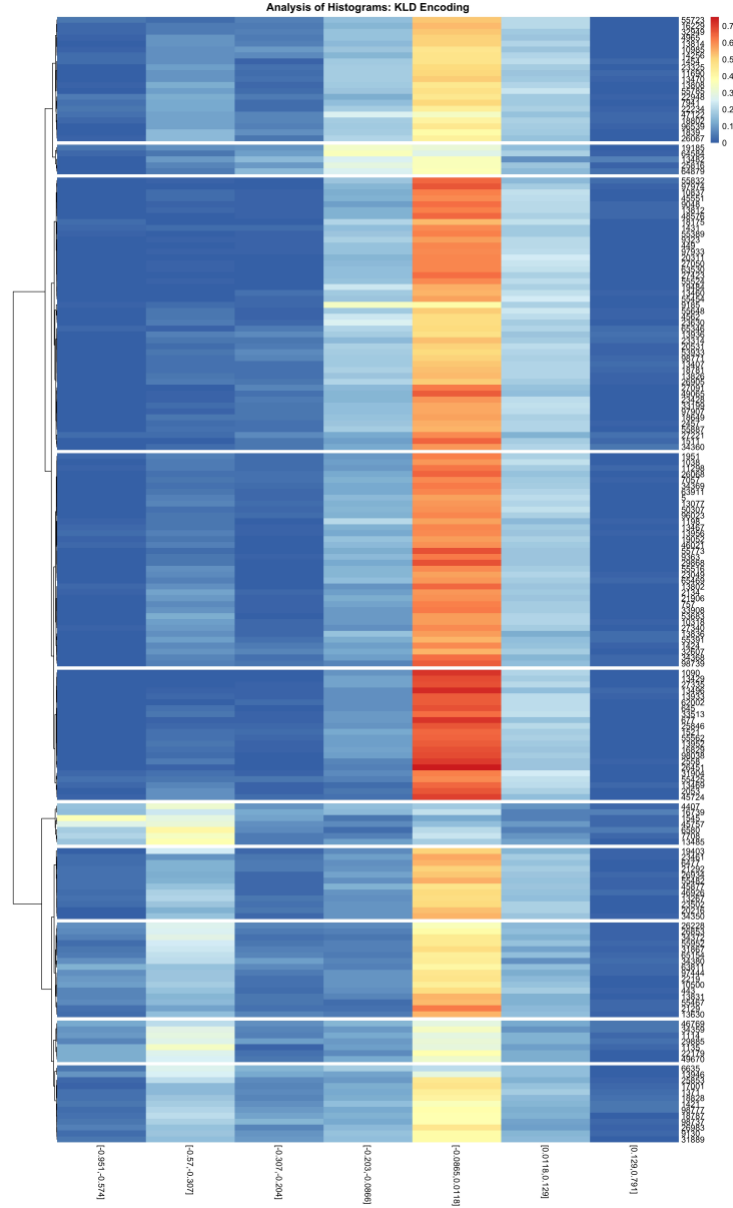


Figure 5: Heatmap visualization of the Analysis of Histograms (ANOHT) encoding for seven histogram bins. Cows are arranged on the row axis and histogram bins along the column, such that each cell represents the proportion of lactation records for a given cow that fell into a given histogram bin. A finer encoding of cows with moderate-to-high rates of outlier yield records is provided with this clustering strategy. Among cows with low outliers, differences in skew among values closer to the projected curve are more homogenous in this encoding by still only coarsely distinguished into relatively large clusters, which likely reflects instability in the measurements.

Entropy Spectra

In the ANOHT analyses using Kullback-Leibler distance, discretization of the marginal residual yield distributions allowed for nonlinearity in the response system with minimal loss of information, but also introduced an additional metaparameter into the encoding scheme: the number of histograms bins used to summarize the data's distribution using the possibly gapped histogram (Hsieh and Roy 2018). While this metaparameter could potentially be optimized to emphasized different structural features of these marginal residual yield distributions that occur at distinct scales of magnitude in the biological response (McVey *et al.* 2021), a data compression scheme capable of encompassing all such information into the final encoding may still be preferable. To explore this possibility, we have developed an entropy spectrum encoding schema, which is loosely inspired by the complexity profiles found in complexity science (Bar-Yam 2004). To implement this algorithm, possibly gapped histograms are fit over all residual yield records for all cows on a grid of bin count values. The resulting binning rules are then used to discretize the continuous residual yield observations, as in ANOHT encoding scheme. From this ensemble of discretized records, an entropy spectrum matrix is created wherein each row corresponds to a given cow and each column to information entropy estimates calculated over all records for a given cow for each of the bin count discretization schemes (MacKay 2003).

Entropy spectra were calculated for the observed and for each of the simulated data sets. To compare the stability of this closed-form entropy estimator to the stability of the closed-form distributional summary statistics, Kendal Tau estimates were computed between values for the observed and each of the simulated datasets. Full results of these stability analyses are provided in Supplemental Materials. Select histograms summarizing these simulation results are provided in Figure 6. All estimators proved to be reasonably reliable, but the relative stability of these values decreased progressively with the granularity of the distributional features measured. It should be

noted, however, that this attenuation was more subtle for the entropy estimates computed from discretized yield records than for the fundamentally continuous summary statistics, which suggests that the discretization step may have served to improve the stability of the information compression in the estimator calculation (Agresti 2013).

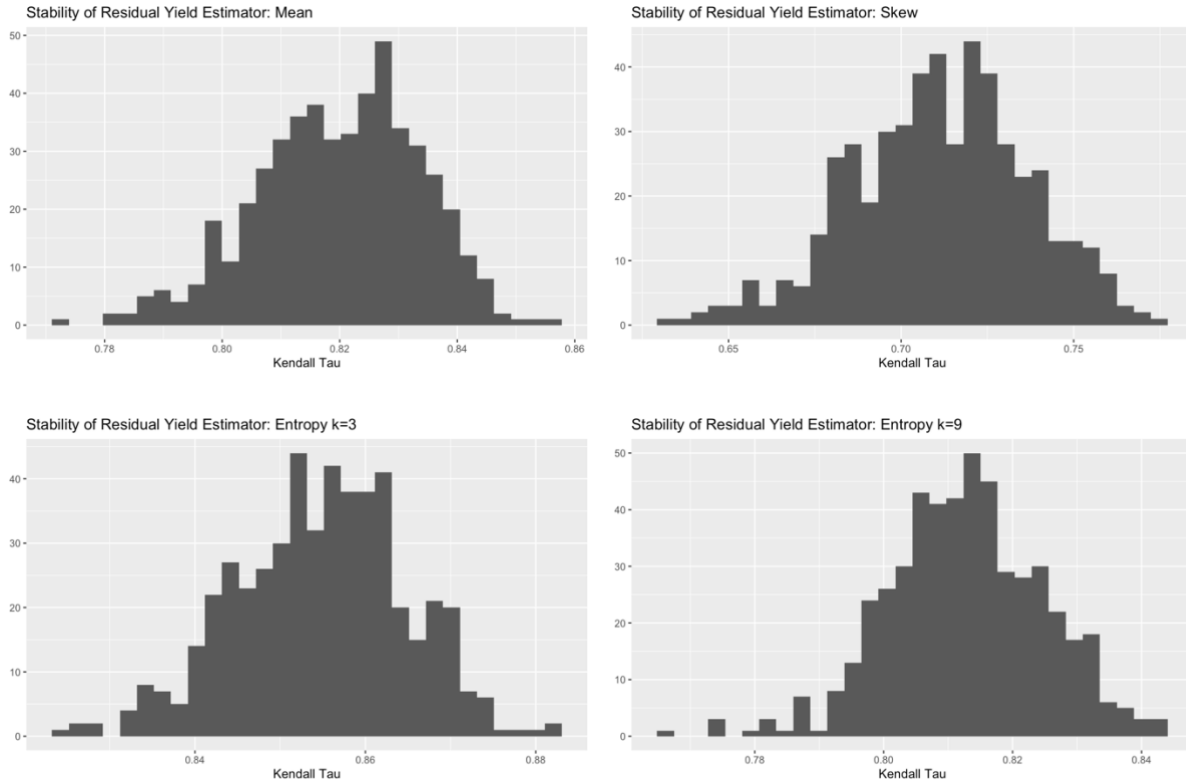


Figure 6: Visual representation of the results of stability analyses. Kendall Tau correlations were calculated against closed form estimators calculated for the observed data set and each data mimicry created using simulated measurement error. Reliability was reasonably high for all estimators, but for the summary statistics using continuous measures of residual yield, the estimates of higher moment (skew and kurtosis) produced notably lower correlations than the estimates of higher moments (mean and variance. For the entropy estimates using discretized yield records, Kendall Tau estimates remained fairly high regardless of encoding granularity.

Additionally, entropy estimates calculated over a grid of bin count were compared against summary statistics estimates using Pearson correlations values for the observed lactation residuals, and the results summarized via a heatmap provided in Figure 7. These results reveal that there are no clear one-to-one comparisons between entropy values and summary statistics. The strongest correlations to skew and kurtosis are recovered with the entropic values for the coarsest encodings, which based on the structure of the corresponding possibly gapped histograms used to create these discretization schemes (see Supplemental Materials), suggests that these summary statistics are quite influenced by outlier rates in these residual yield records. Interestingly, all entropic estimates are at least moderately correlated to estimates of standard deviation, but entropy values from finer encodings are at best weakly correlated to the estimates of higher moments. Given that the binning rules for these encoding schemes provide finer stratification around not only the projected yield curve, but also for more systematic outliers found just below the fitted yield curve that likely correspond to single kick-off milking days, these entropic estimates might be providing a more dynamic estimate of authentic stochasticity in residual yield records that more naturally accommodate the multimodality created by kickoff events without loss of records that would be incurred by tossing out such records.

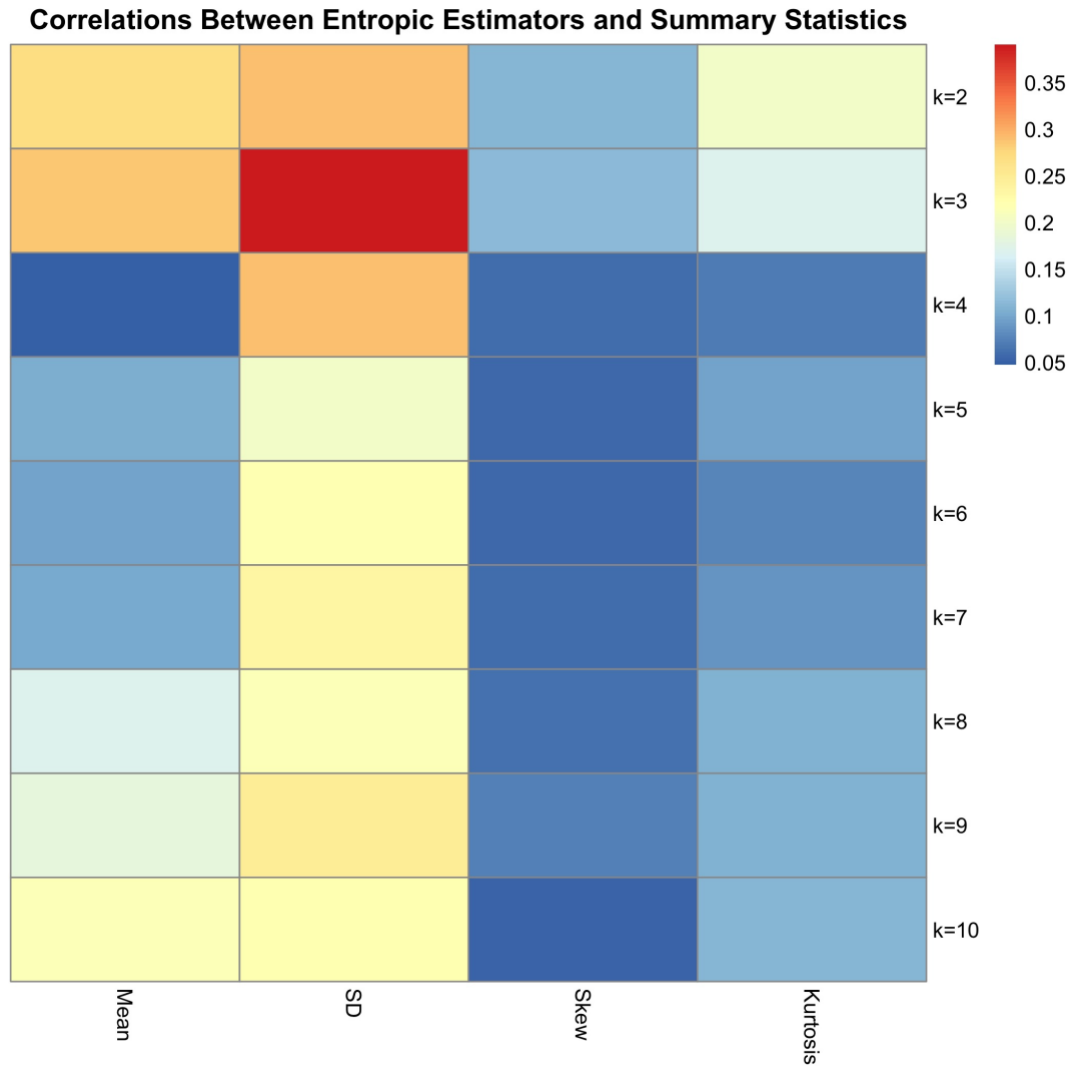


Figure 7: Heatmap visualization of the Pearson correlations calculated for the observed residual yield records between standard summary statistics and entropy estimates for encoding granularities with 2 to 10 clusters. All correlations were low, indicating that no single entropy estimate can be directly (linearly) related back to a single summary statistic. All entropy estimates were moderately correlated to estimates of variance, but here only entropy values from the coarsest encodings were even weakly correlated to estimators of higher moments.

Entropy spectra were clustered by first calculating the pairwise dissimilarity between cows using the ensemble-weighted Euclidean distance and then hierarchically clustering using Ward's D2 method (McVey *et al.* 2021). In calculating the penalty term for the dissimilarity estimate, however, the greater stability of the entropy estimates here created a logistical hurdle, as coarser encodings for some cows were in fact so stable that zero variance was found in the subsequent entropic estimates across the ensemble of simulated datasets after accounting for the various sources of measurement error. To overcome this issue, ensemble penalties were calculated by adding the ensemble variances of a given estimate for a given cow to the square of the global average of each mean estimate calculated across cows, which simultaneously scaled each component of the dissimilarity estimate and penalized for observation uncertainty. While different aspects of the underlying distributional geometry might be expected to be emphasized as the bin count of the underlying encoding scheme is varied, some redundancy in the entropic measures might be anticipated when calculated on a uniform grid. To account for this, two distinct encodings of entropy spectrum were here create: one using just the ensemble penalty, and a second where the ensemble penalty was multiplied by the average column weights returned by a data mechanics optimization on a grid of cluster counts from two to six (Hsieh *et al.* 2018; McVey *et al.* 2020). Finally, both of the resulting dendrograms were subsequently pruned using the *cutreeEnsemble* utility using the ensemble of entropy spectra calculated for the simulated datasets (McVey *et al.* 2021).

A total of six validated clusters were recovered for the ensemble-weighted entropy spectrum encoding, while a total of eight validated clusters were recovered for the encoding reweighted using both the ensemble and data mechanics results. Figure 8 provides a visual comparison of the finalized encodings. This reveals that both encodings are in strong agreement for the coarsest divisions of the data, particularly in distinguishing between distributions with high

relative entropy values at the coarsest levels of discretization that reflect the relative rates of milking outliers. In visually inspecting the resulting histograms, however, it is interesting to note the beyond the entropic estimates recovered from the first handful of grid count values, the complexity of the underlying histograms seems to increase not incrementally, but instead to jump up as specific underlying geometric features seem to come into resolution. The resulting redundancy in the entropic estimates from the high bin count metaparameter values likely explains the greater acuity in encoding reweighted using the data mechanics values.

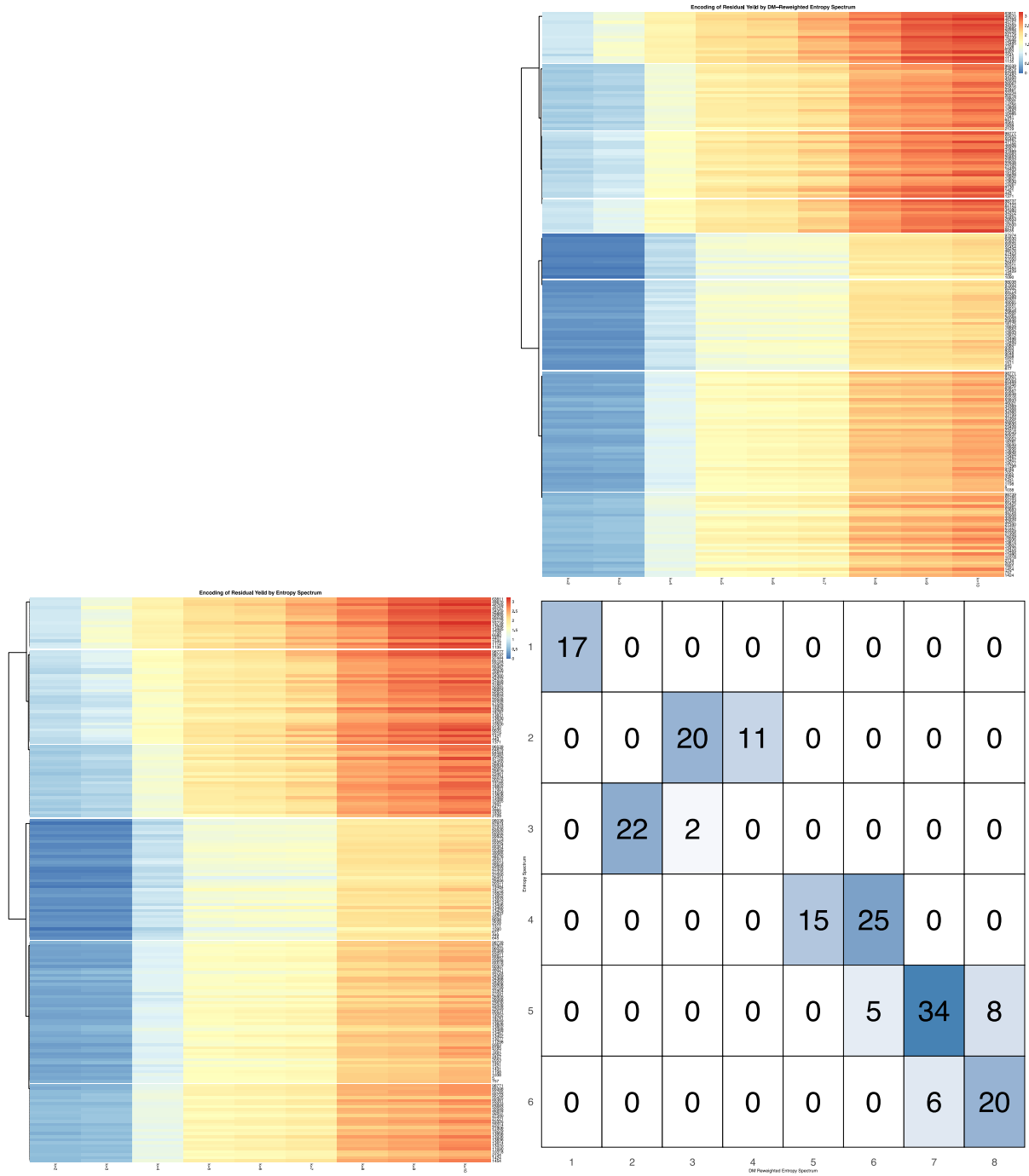


Figure 8: Comparison of entropy spectra encodings created using ensemble-weighted distance (bottom left) and the jointly-weighted distance using both an ensemble penalty and a data mechanics weighting (top right). For both heatmap visualizations, cows are indexed on the row axis, and the column axis indexes entropy estimates calculated using discretized residual yield

records on a grid of 2:10 clusters. The contingency table (bottom right) reveals that these two encodings are closely aligned, but that the data mechanics reweighting helped to further distinguish differences in some clusters by applying greater weight to entropy estimates calculated using the coarser encodings. See Supplemental Materials for larger high-resolution versions of these images, wherein full details can be readily visualized.

Cross-Sensor Inference

As resilience is a latent trait that cannot be directly measured, there is no “ground truth” against which the various encodings of marginal residual yield can be directly contrasted. To create the most holistic picture possible of the relative efficacy these information compression strategies, we instead must compare these encodings against a range of other behavioral and biological responses captured by independent PLF data streams that may also be at least partially influenced by the relative resilience of these animals (Colditz and Hine 2016; Berghof *et al.* 2019; Scheffer *et al.* 2018). In order to capture the widest range of potential bivariate dynamics, significant associations were recovered via a nonparametric permutation-based mutual information testing framework using the *bivarTreeTest* utility available in the LIT package (McVey *et al.* 2021). Any bivariate associations significant at the standard $\alpha = 0.05$ significance level were then further characterized using the *compareEncoding* utility, wherein cells of the resulting contingency table were colored by their pointwise mutual information value only if the observed estimate differed significantly from values recovered under the null using a semi-parametric multinomial testing framework (McVey *et al.* 2021).

	Summary Statistics	Unweighted Earth Movers	Weighted Earth Movers	Kullback-Leibler ANOHT	Unweighted Entropy Spectrum	Weighted Entropy Spectrum
Lactation Number	0.323	0.531	0.659	0.364	0.423	0.423
Health	0.025	0.046	0.068	0.072	0.068	0.072
Diagnosis	0.028	0.040	0.009	0.006	0.046	0.046
DIM at First Service	0.006	0.005	0.002	<0.001	0.001	0.008
DIM Pregnant	0.016	0.092	0.136	0.006	0.044	0.029
Entry Order All Cows	0.092	0.045	0.030	0.007	0.121	0.121
Entry Order Healthy Cows	0.026	0.038	0.178	0.027	0.026	0.026
Precision Penalized Overall Time Budget	0.017	0.028	0.008	0.002	0.034	0.014
Plasticity Penalized Overall Time Budget	0.008	<0.001	0.005	0.001	0.005	0.003

Table 1: P-value results from bivariate tree tests performed between each encoding strategy and a range of performance, health, and behavior records also collected for study animals.

Health Outcomes

Health records are of course the most natural response metric against which to compare potential estimators of resilience. To do so, several data streams were considered. First, no significant associations were found between any of the marginal residual yield encodings and lactation number, which was here discretely encoded for lactations 1, 2, and +3. Thus, while it is still possible that age could modulate the bivariate associations between encodings residual yield

and responses variables from other data streams, it does not on its own appear to drive heterogeneity in the geometric features of residual yield distributions.

Bivariate associations were next assessed using an comprehensive encoding of health status, wherein cows were simply encoded as “sick” if they were identified as suffering from any acute medical condition during the first 150 days of milk, and “healthy” if no such complications were detected. All information compression strategies were found to have either significant or marginally significant associations with this health status encoding. Figure 9 provides a comparison of the summary statistics encoding and overall health status. Cows with the highest mean and lowest variance estimates are significantly under-represented among cows with any recorded health events. These results appear to agree with previously published work on residual milk yield, where only variance estimates provided significant genetic correlations against health traits (Poppe *et al.* 2020; Poppe *et al.* 2021; Kok *et al.* 2021). A similar comparison against the unweighted Earth Movers Distance, which is provided in Supplemental Materials, confirms that this represented cows with the lowest overall rate outlier residual yield estimates, though heterogeneity in skew in values nearer the queue did not appear to influence this bivariate relationship.

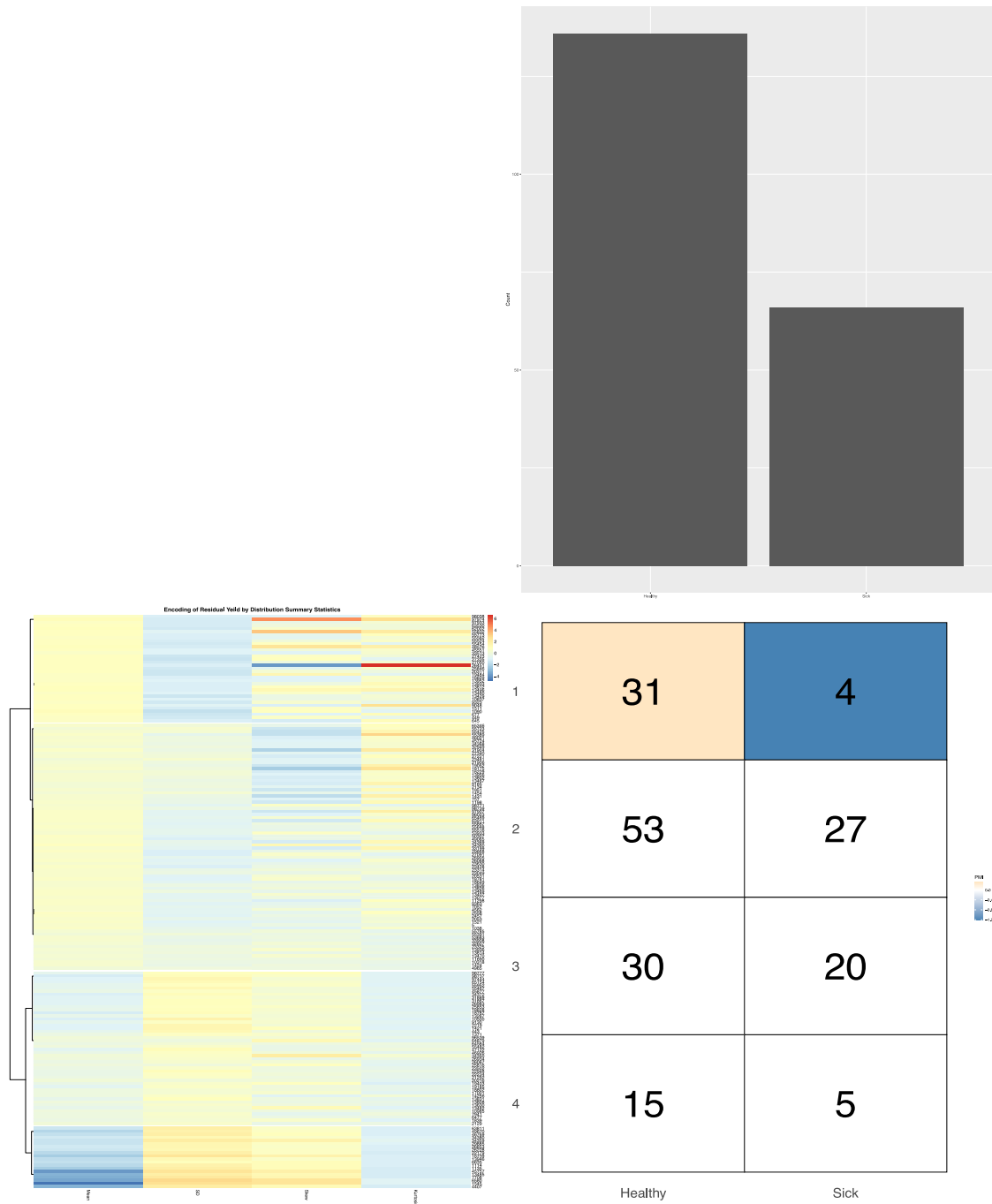


Figure 9: Visualization of the significant bivariate association between overall health status and the summary statistics encoding of marginal residual milk yield. Cows whose residual yield distributions can be characterized by high (near zero) means and low variance are significantly over-represented among cows with no recorded health complications. See Supplemental Materials for larger versions of these images.

While the bivariate associations recovered against a comprehensive encoding for acute health complications are encouraging, a significant amount of detail might be masked by confounding dynamics in the behavioral and biological responses to health complications with distinct physiological etiologies. To explore this possibility, health records were recoded into five broad diagnostic categories: healthy, reproductive (infections of the reproductive tract), mammary (infection or damage to the udder), digestive (metabolic diseases or digestive upset), and hoof (infections or damage to the hoof or legs causing lameness). Disaggregating this data produced significant or highly significant associations with all encodings of marginal residual yield distributions. Contingency table visualizations comparing diagnoses against all information compression techniques are provided in Supplemental Materials, but perhaps the clearest picture of this complex bivariate dynamic is recovered for the weighted Earth Movers Distance encoding, which is visualized in Figure 10. From this we see that cows with moderate to relatively high rates of outlier yield records are significantly over-represented among digestive and hoof complications. It should be noted, however, that for cows with multiple recorded acute health events within this milking interval, only the first recorded health event was encoded. Thus, given that metabolic complications often progress to hoof disease later in the lactation interval, and also that hoof infections may be indicative of subacute metabolic complications that went undetected earlier in the lactation interval, the similarity in these results may reflect shared underlying behavioral or biological dynamics (Manriquez *et al.* 2019). A number of potential biological dynamics might be proposed to link metabolic complications with high rates of compromised yield. Alternatively, a behavioral interpretation of this pattern might be that cows with metabolic complications have sore feet, and that causes them to shift their weight more while being milked, which in turn causes the milk claw to fall off prematurely on a regular basis. Emerging research has recovered preliminary genetic correlations between residual yield estimates and variability in daily step count throughout

the lactation window (Poppe *et al.* 2021). Additional information on milking duration would ultimately be needed, however, to fully disentangle behavioral and biological mechanisms that might drive this association.

No significant associations were recovered between any of the residual milk yield encodings and mammary infections. This null result is surprising, as previous studies have recovered significant genetic correlations between residual yield estimates and measures of udder conformation and health (Poppe *et al.* 2020). This finding, however, may be attributed more to missing records than lack of any interesting biological or behavioral dynamics, as milk dumped from infected utters may have resulted missing values in daily milk records, which would have subsequently been excluded from subsequent analyses. It should be noted, however, that for discrete approaches to analyses of marginal residual yield, missing milkings could be incorporated into these encoding frameworks simply by creating a bin for empty record fields. Future work might explore such opportunities by leveraging metadata from additional data streams, such as farm health or even accelerometer records, to distinguish between missing records that are simply caused by failure of the RFID reader and those attributable to acute health events in order to correct for this potential source of measurement bias in estimating residual yield distributions.

A perhaps more surprising result, however, would be the finding that cows with the lowest rates of outlier residual yield values were in fact significantly over-represented among cows with infections of the reproductive tract, while cows with a low but appreciable rate of low yield outliers were in fact significantly under-represented. A number of behavioral or biological dynamics might explain this result. In previous research, cows with low overall yield have been found to have lower variability in residual milk yield (Poppe *et al.* 2020). As reproductive infections typically occur during the transition period, such infections might reduce the rate of outlier yields if they have longer-term impacts on the overall level of milk production throughout the subsequent lactation.

A larger data set would unfortunately be needed to conclusively identify such complex interactions. Alternatively, it is possible premature kickoffs or transient low production days provide some behavioral or biological benefit to the cow's reproductive system. As cows with reproductive infections are often quite late to return to breeding, it is also possible that this pattern reflects that such animals are not demonstrating estrous behaviors, which are known to result in transient low milk production (Reith and Hoy 2018).

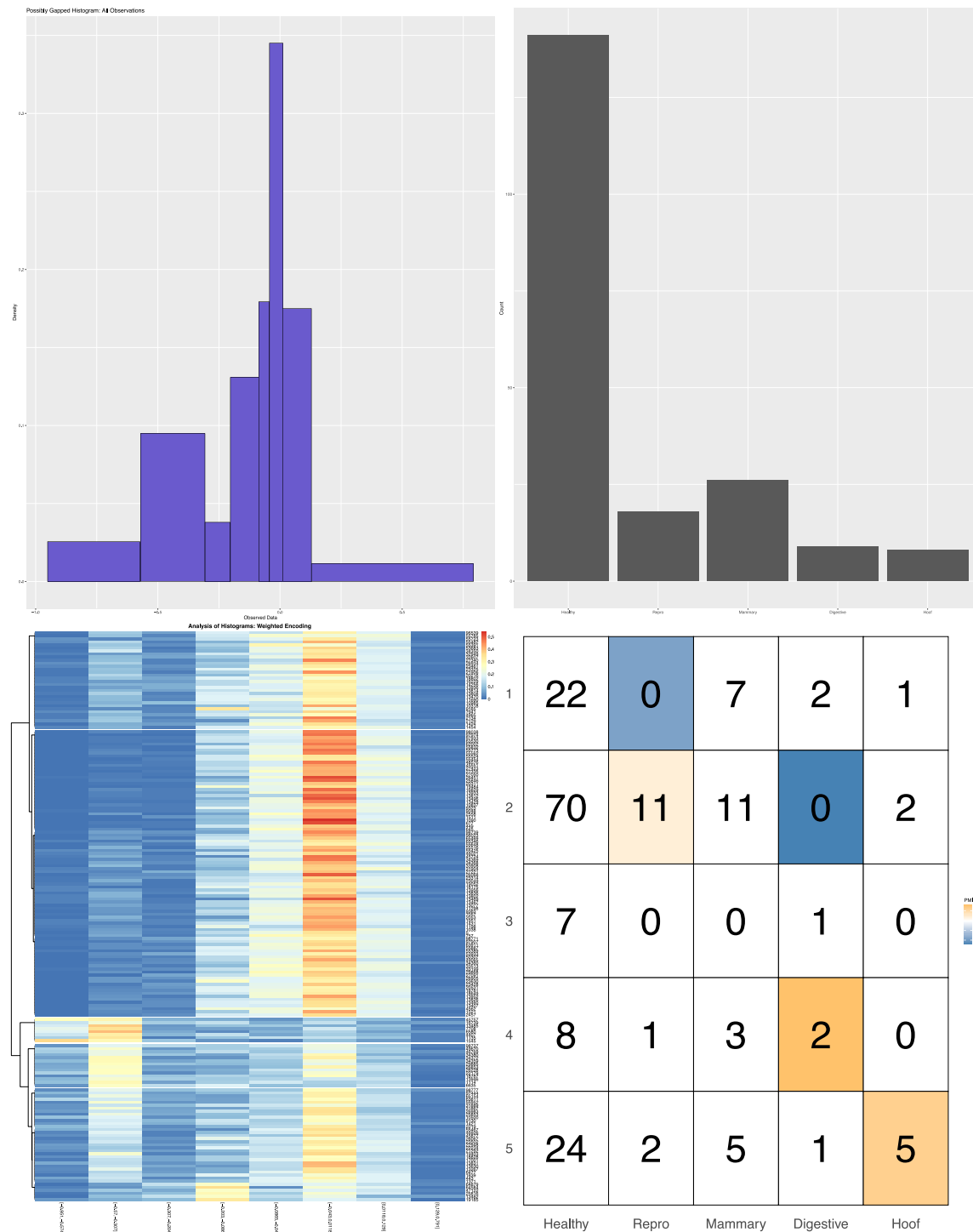


Figure 10: Visualization of the significant bivariate association between health diagnosis and the weighted Earth Movers Distance encoding of marginal residual yield. Cows with moderate to high rates of outlier residual yields are significantly over-represented among cows with digestive or

hoof diseases. Cows with the lowest rate of outlier residual yield records are significantly under-represented among digestive diseases, but significantly over-represented among infections of the reproductive tract. Cows with low but appreciable rates of residual yield are significantly under-represented among cows with reproductive infections. See Supplemental Materials for larger versions of these images.

Reproductive Performance

To explore the relationship between marginal residual yield distributions and reproductive health, two indicators of reproductive performance were extracted from herd records: days in milk at first AI service (DIM-FS), and days in milk at confirmation of pregnancy by ultrasound (DIM-Preg). Both of these values were here encoded using possibly gapped histograms on a grid of bin count values from one to six, wherein bins were numbered from lowest to highest value and cows that never return to their heat cycle were assigned a value of zero (Hsieh and Roy 2018). This discretization strategy allowed for the encoding granularity to be optimized using the *bivarTreeTest* utility, while still capturing any nonlinear dynamics in the underlying survival curve.

Highly significant bivariate associations were recovered between DIM-FS and all encodings of marginal residual milk yield. Visualizations created using the *compareEncodings* utility are provided for all encodings in Supplemental Materials. Figure 11 shows the bivariate association between DIM-FS and the ANOHT encoding created using KLD distance. No significant associations were found between residual yield encodings and cows who never returned estrous, though the relatively low numbers of such animals observed in this trial may simply have not provided enough power for such inferences. Cows that were delayed in returning to estrous

relative to their herd mates (≥ 72 DIM) were significantly over represented not only among cows with moderate-to-high rates of low yield records, but also among cows with almost no yield records deviating significantly from their projected yield curve. In contrast, cows that returned to estrous quickly relative to their herd mates (≤ 70 DIM) were significantly over represented among marginal residual yield curves with low but non-negligible rate of outlier yield records.

This result may lend further evidence to the interpretation of the link between infections of the reproductive tract and patterns in residual yield – that milk production may need to fall off at least occasionally to allow energy to be temporally diverted to estrous behaviors (Reith and Hoy 2018). It is interesting to note, however, that this bivariate relationship was in fact strongest for the ANOHT encoding among the coarser discretization of residual milk curves. In fact, the strongest results are recovered from a four-bin histogram with empirically-determined cutoffs that effectively distinguished between severe underproduction (-0.95 to -0.57 normalized residual), moderate underproduction (-0.57 to -0.20 normalized residual), normal low production (-0.20 to -0.01 normalized residual), and normal high production and overproduction (-0.01 to 0.79 normalized residual). Collectively these results suggest that it is relatively small differences in rates of milking outliers and not more subtle differences in the degree of skew among values closer to the projected yield curve that drive this result. Such an association might therefore be easily lost or modified by data scrubbing prior to analysis of residual yield curves. Thus, further work is likely warranted to delve deeper into the cause of these outlier records – namely kickoffs or metabolic factors – and how this information might be better incorporated into data processing and encoding schemes to help ensure that any selection schemes for improved resilience using residual yield records do not unintentionally penalize energetically expensive but still economically desirable estrous behaviors (Berghof *et al.* 2019).

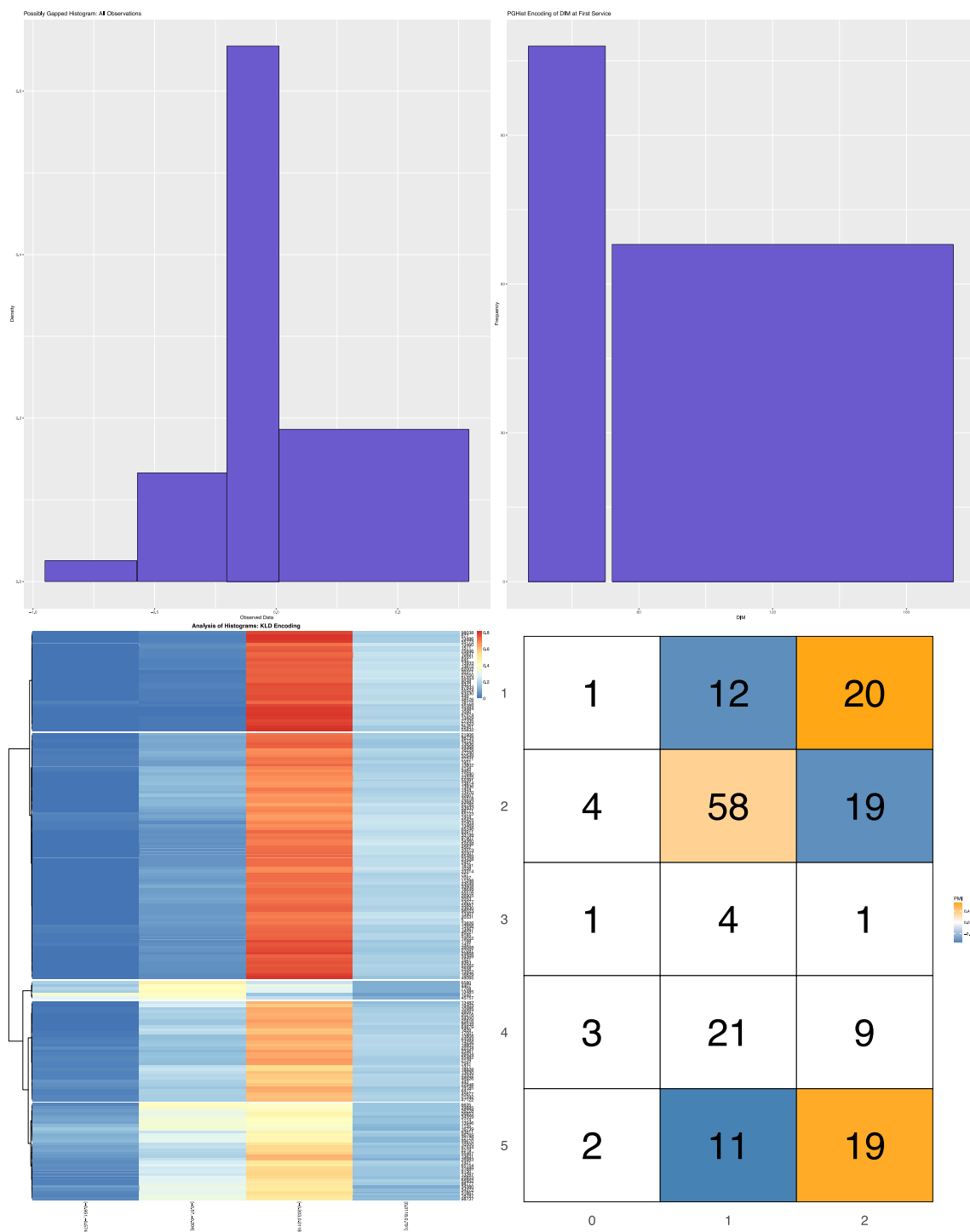


Figure 11: Visualization of the significant bivariate association between days in milk at first service and the ANOHT encoding of marginal residual milk yield with four bins. No significant

associations were recovered for cows that failed to return to estrous (encoded as zero). Cows that were slow to return to estrous (≥ 72 DIM) were significantly over-represented both among cows with moderate-to-high rates of outlier milk yields, but also among residual yield distributions with few appreciable outliers. Conversely, cows quicker to return to estrous (≤ 70 DIM) were significantly over represented among residual yield distributions with low but appreciable rate of outliers. See Supplemental Materials for larger versions of these images.

The strength of bivariate associations with days in milk at confirmed pregnancy was more variable across information encoding schemes, with Earth Movers Distance metrics recovering only marginally significant associations, and the ANOHT encoding using KLD Distance again providing the most highly significant results, which is visualized in Figure 12. Once again, a coarse encoding scheme for residual yield record using just a three-bin histogram was found to provide the strongest association, which further emphasizes the dominant role that outlier rates seem to play in associations with reproductive performance. These results found no significant association between patterns in residual milk yield and cows who failed to breed back during the observation window. Cows with moderate rates of outlier yield observations, however, were significantly over represented among animals that were late to be confirmed in calf (≥ 144 DIM), while cows with the highest rates of underproduction were significantly under-represented among late-bred individuals. This result suggests that cows, if cows with modest rates of yield outliers do in fact hold a biological or behavioral advantage over their peers with respect to their ability to return to estrous, then the underlying mechanism does not appear to also improve conception rates. If these outliers are directly tied to estrous behaviors, then it also reasonable to conclude that this dynamic might only be seen with animals who progress through a larger number of cycles prior to

conception (Berghof *et al.* 2019). The higher conception rates found among cows with quite high rates of outlier yield values, on the other hand, may be more difficult to attribute to a behavior mechanism, but given the consistency of under-production, a potential mechanism to explore might be differences in embryo reabsorption driven by differences in degree of negative energy balance (Esposito *et al.* 2014).

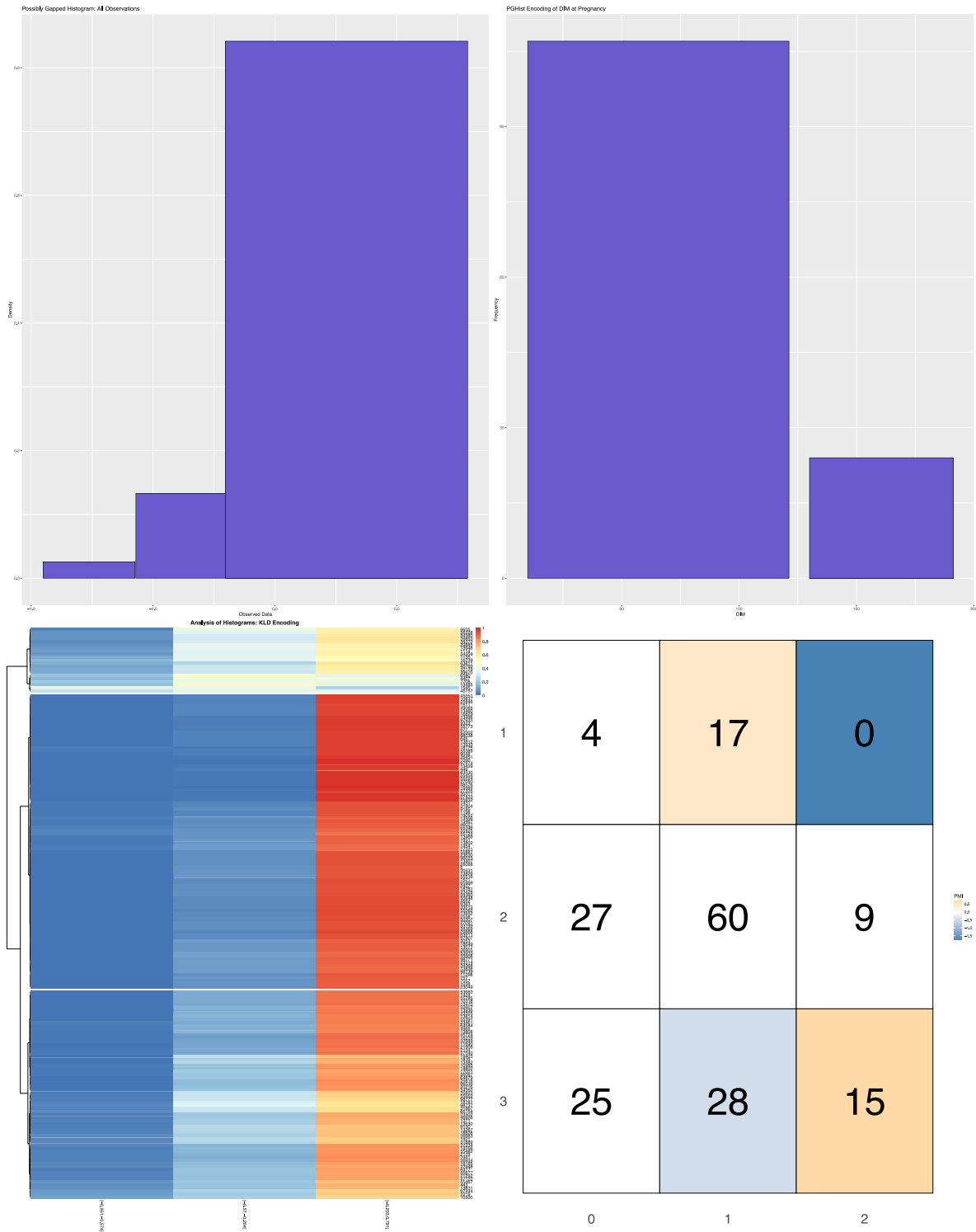


Figure 12: Visualization of the significant bivariate association between days in milk at confirmation of pregnancy via ultrasound and the ANOHT encoding of marginal residual milk yield with three bins. No significant associations were recovered for cows that failed to breed back

(encoded as zero). Residual yield distributions low but appreciable rates of outlier records were here significantly over-represented among cows that were late to breed, while animals with high rates of outlier yield records were significantly underrepresented among these late bred cows. See Supplemental Materials for larger versions of these images.

Parlor Entry Order

Milking order, or the order in which cows arrange themselves as they enter the parlor to be milked, is thought to be influenced by a number of latent biological and behavioral mechanisms: health status, milk yield, energy balance (hunger), dominance status, etc. (Schein and Fohrman 1955; Gadbury 1975; Soffie *et al.* 1976; Rathore 1982; Polikarpus *et al.* 2015; Beggs *et al.* 2018). In previous work with this dataset, RFID milking metadata were used to extract the single file order in which cows entered the rotary parlor from a total of 80 morning milkings. Temporal patterns in cow entry positions were subsequently encoded using data mechanics clustering algorithms (McVey *et al.* 2020). These previously reported encodings will here be used to explore relationships between patterns in marginal residual milk yield and queueing behaviors prior to milking.

Patterns in parlor entry position that were fit only to cows with no recorded health complications produced significant bivariate associations against all encoding of marginal residual milk yield except for the weighted Earth Movers Distance encoding. Contingency table visualizations for all significant associations are provided in Supplemental Materials, but all reveal the same relatively straight forward pattern: cows that entered in the center of the milking queue, and who subsequently showed the most variability in their overall entry position, were significantly over represented among residual yield distributions with moderate to high rates of outlier yield records. If queue position is viewed as an aspect of the milking environment , then

this result may simply reflect that greater variability environmental factors may create more opportunities for animals to display a more varied biological or behavioral response, which underscores that variability in environmental conditions are not necessarily homogeneous across animals even within the same herd in complex commercial farm environments (Scheffer *et al.* 2018; Berghof *et al.* 2019). If queuing position is viewed as an independent behavioral response, however, then this result may raise interesting questions about the generality of individual differences in the plasticity of behavioral responses across contexts (Sih *et al.* 2004; Nilam *et al.* 2017; Scheffer *et al.* 2018).

For all encodings of marginal residual yield, except for those utilizing entropic estimators, significant associations were also recovered for patterns in parlor entry position that were fit to all animals that remained on trial regardless of health status. The clearest bivariate pattern was recovered again for the ANOHT encoding using KLD Distance, which is visualized in Figure 13. This result shows that cows that generally entered nearer the rear of the queue were under-represented among marginal yield distributions with moderate to high rates of outlier records. Additionally, a small group of animals that appeared to randomly jump between queue positions that were generally nearer the front or nearer the rear of the herd were also significantly over-represented among marginal residual yield distributions with low rates of outlier records. These results seem to contrast with the bivariate patterns found among exclusively healthy cows, which suggest that mechanisms linking yield patterns to subacute health status may differ from behavioral or biological responses to acute health issues (Hart 1988). Further analyses with a larger dataset would ultimately be needed to disentangle complex interactions between health status, entry order, and residual yield.

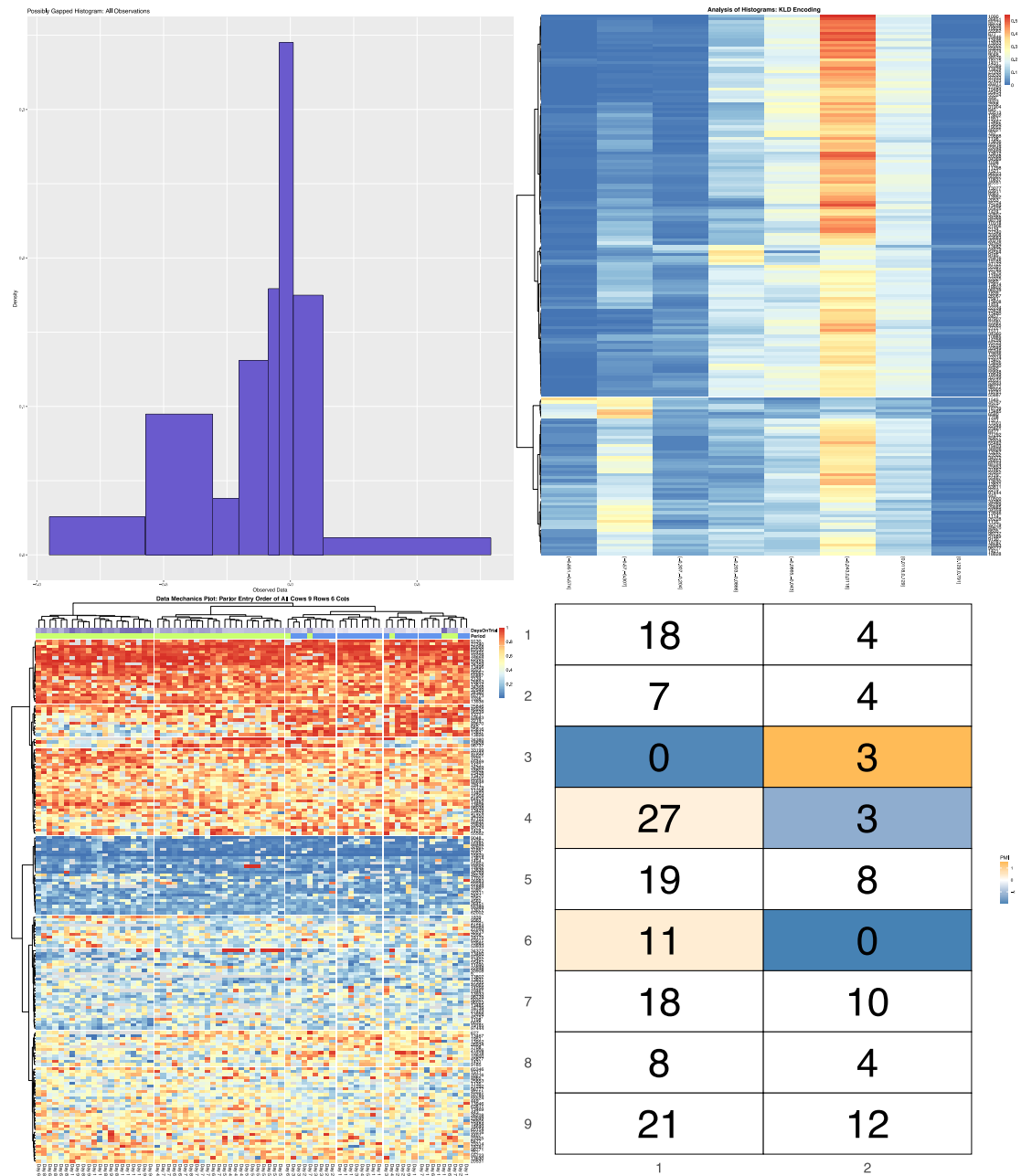


Figure 13: Visualization of the significant bivariate association between ANOHT encoding of marginal residual milk yield with eight bins and patterns in entry order into the milking parlor for all cows on trial. Cows that generally entered towards the rear or the queue, and those that generally entered towards the front of the queue but who still showed high variability in their entry position over time, were both significantly over represented among residual yield curves with low rates or outlier yield records. See Supplemental Materials for larger versions of these images.

Overall Time Budget

Time spent in the milking parlor ultimately constitutes a relatively small portion of a cow's management schedule (Tucker *et al.* 2021). In previous work with this dataset, overall time budgets were calculated for study animals using records from the commercial ear tag accelerometer system that was used in the feed trial to track five mutually exclusive behaviors: eating, ruminating, nonactive, active, and highly active. Heterogeneity within the herd in the temporal tradeoffs made between behavioral axes were subsequently encoded by augmenting conventional hierarchical clustering techniques with insights into the relative reliability of behavioral patterns gleaned by mimicking the sources of noise within the dataset using layered nonparametric simulation techniques (McVey *et al.* 2021). These previously reported encodings are here used to determine if significant bivariate patterns might also exist between patterns in residual milk yield and how a cow spends her time between milkings.

Significant bivariate associations were recovered between all encodings of marginal residual milk yield and encodings of overall time budget penalized to account for heterogeneity in the precision of the sensor. Contingency table visualizations for all encodings are provided in Supplemental Materials. The clearest pattern was again recovered for the ANOHT encoding using the KLD Distance, as shown in Figure 14. These results show that cows that spent a high amount of time eating were significantly under represented among marginal residual milk curves with moderate to high rates of outlier records. On the other hand, marginal residual yield curves with the highest rates of outlier records were significantly over-represented among time budgets with low time spent eating and were significantly under-represented amongst animals with the most moderate time budgets. Interestingly, cows with time budgets characterized by low time spent eating, moderate rates of rumination, and high time spent nonactive were significantly over represented among residual yield curves with low but non-negligible rates of yield outliers and

increased skew about the projected milk curve. The fairly linear link between time spent eating and rate of outliers could easily be explained by differences in energy balance and subsequently appetite attributable to variations in level of aggregate milk production. That this relationship also, however, appears to be modified for cows with increased rates of nonactivity, which can be an indication of latent health issues (Beauchemin 2018), suggests that skew might here provide additional insights into more nuanced differences in latent health and resilience below the acute or clinical level.

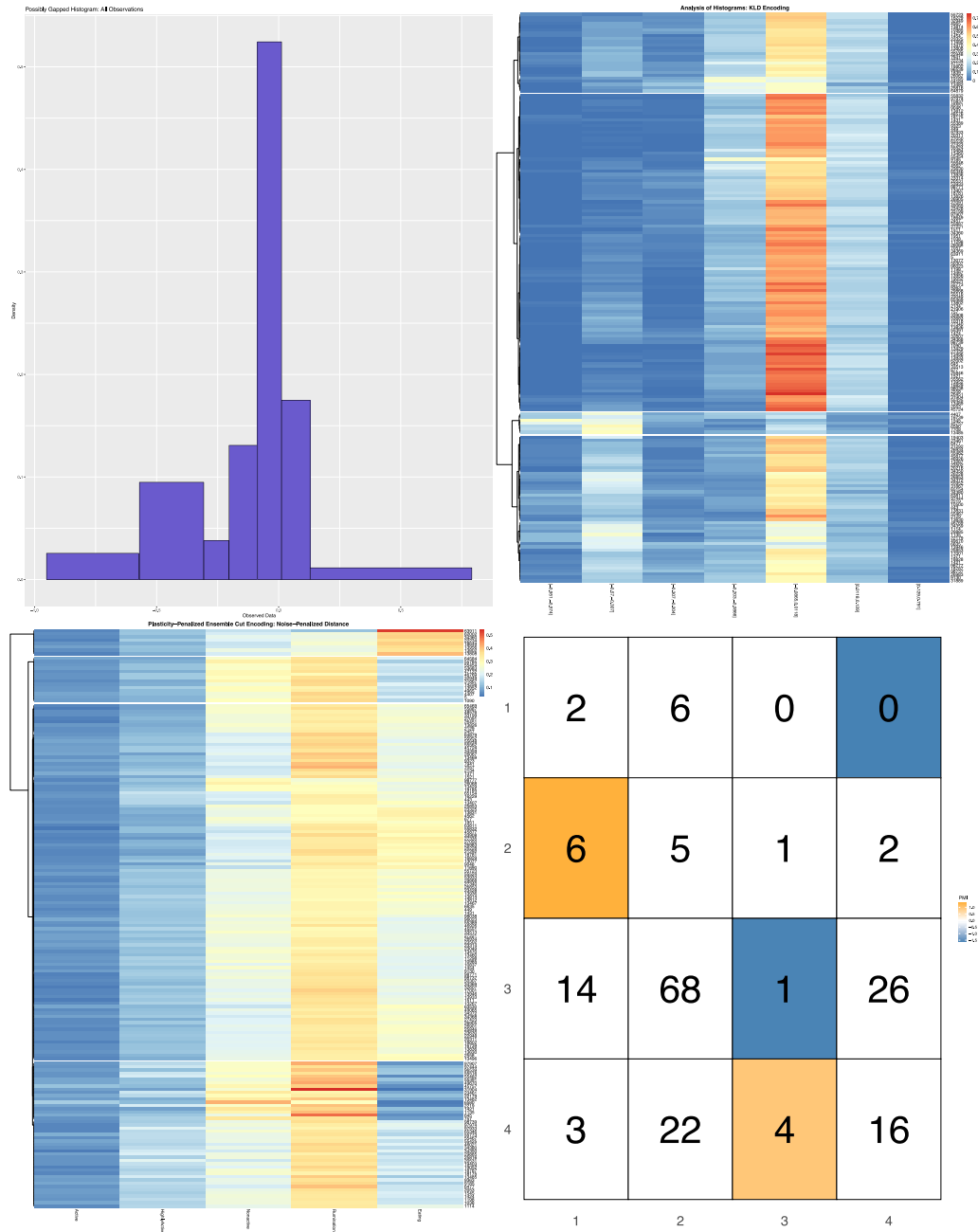


Figure 14: Visualization of the significant bivariate association the ANOHT encoding of marginal residual milk yield with seven bins and encodings of overall time budgets penalized to account for heterogeneity in sensor precision. Cows with high time spent eating were significantly under-represented among residual yield distributions with low-to-moderate rate of yield outliers. Cows with high rates of outlier yields were significantly over represented among cows with low eating

times and under-represented among more moderate time budgets. Cows with low outlier rates but high skew around the projected curve were significantly over-represented among cows with low eating time and high nonactivity. See Supplemental Materials for larger versions of these images.

For all encodings of marginal residual milk yield, bivariate associations became notably more significant when encodings of overall time budget were modified to also penalize the relative plasticity of time budgets demonstrated by animals across observation days. Contingency table visualizations for all encodings are provided in Supplemental Materials. After accounting for consistency in daily time budgets, nearly all these encodings brought into even sharper relief the relationship between low eating time and higher rates of yield outliers. Additional details may be discerned, however, in the relationship between time budgets and the entropy spectrum encoding, visualized in Figure 15. Residual yield curves with the highest rates of yield outliers were significantly over-represented among animals with extremely low time spent eating and under-represented among cows with the most moderate time budgets. It should be noted, however, that in an encoding that penalized variability in time budget between days, this least distinct time budget cluster not only contained cows with the least distinct time budgets but also the most variable. It is therefore also interesting to find that such animals were not only significantly over-represented among residual yield curves with the lowest entropy estimates across all encoding granularities of residual, but also significantly under-represented among animals whose entropy spectra showed low but non-negligible rate of outlier yields, but elevated entropy estimates that distinguished between records closer to the projected curve. This inverted pattern suggests that levels of behavioral and biological plasticity might not be directly correlated (Sih *et al.* 2004), but instead that more complex tradeoffs between outcome measures may exist, though a larger sampling of daily time budgets would likely be needed to more fully characterize this finding.

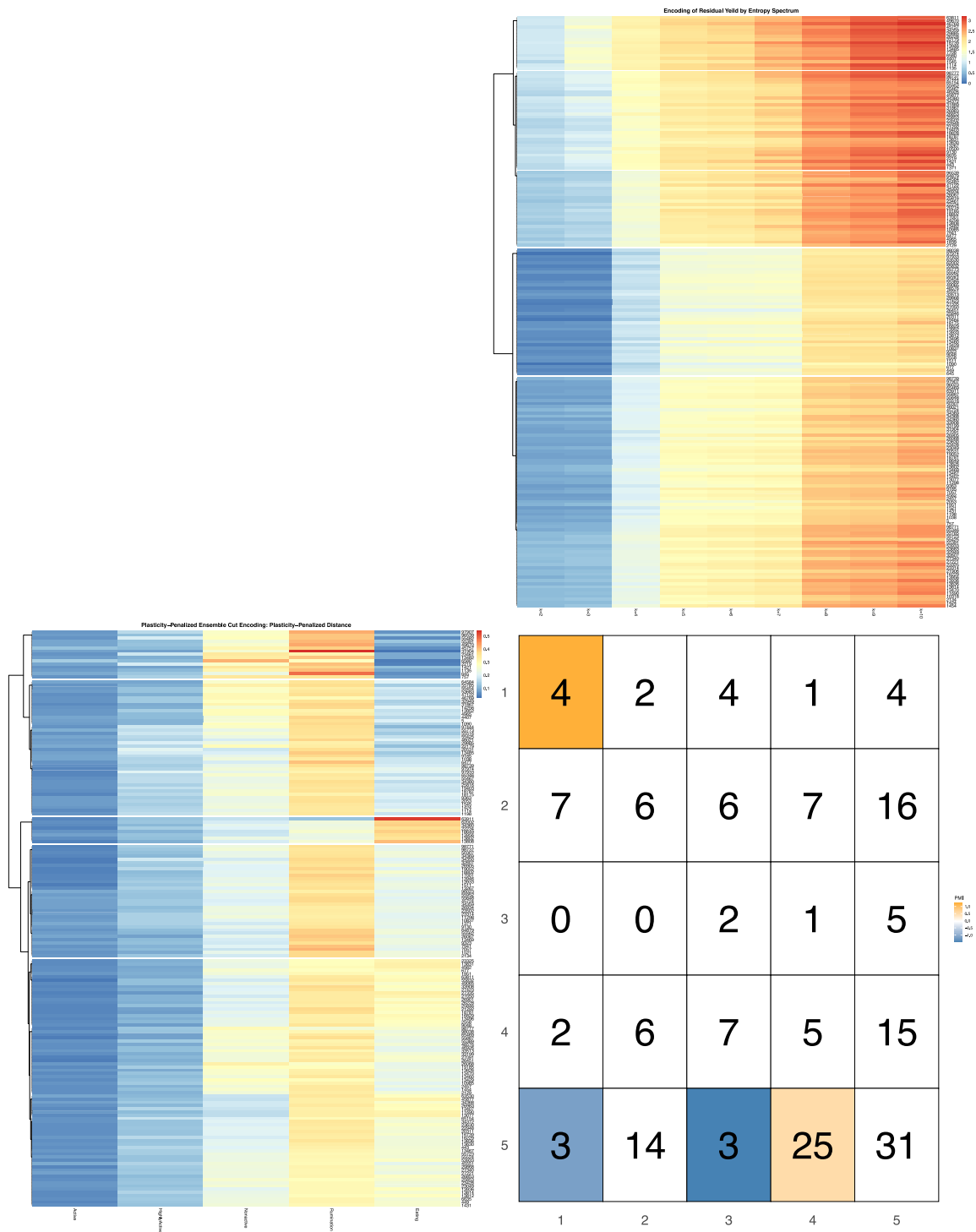


Figure 15: Visualization of the significant bivariate association the ANOHT encoding of marginal residual milk yield with seven bins and encodings of overall time budgets penalized to account for heterogeneity in sensor precision and between-day plasticity. Cows with extremely low eating

times are significantly over represented among the highest entropy curves. Cows with the least distinctive time budgets were over represented among residual yield distributions with the lowest overall entropy, but under represented among the highest entropy curve and the with low entropy among coarser encodings but high entropy for finer record encodings. See Supplemental Materials for larger versions of these images.

Conclusions

The results of this methodological case study confirm that marginal residual milk yield distributions collected from a working farm environment contain a range of complex geometric features that collectively constitute a challenging information compression problem. By layering nonparametric simulations strategies, we have demonstrated how an ensemble mimicry framework might be employed to at least partially disentangle the various sources of measurement error in this noisy data stream from authentic sources of biological variability that might be attributed to individual differences in plasticity and resilience. These results underscore that different distributional features may be estimated with differing degrees of confidence – namely that multimodality related to outliers can be brought into resolution more easily than differences in skew among records closer to the projected lactation curve. Where additional metadata is available, future research should consider the causes of these outliers to provide better insights into the multimodal patterns that such records create in marginal residual yield records, and subsequently the weight that they are afforded in downstream analyses. These results also confirm that individual yield records can vary considerably in their reliability due to intrinsic differences in the confidence with which projected milk yield models can be fit at different stages of lactation, and ensemble mimicries can be used to correct for the systematic bias that these fitting issues create in downstream analyses.

Using these measurement error simulations, all information compression strategies considered here produced comparable levels of validated encoding granularities. This result indicates relatively little information may be hemorrhaged in discretizing residual yield records when an empirically-driven binning rules are applied. It also underscores that, for data streams where we are not left wanting for repeated measurements, there can be considerable advantages in sacrificing the continuity and even the ordinality of a measurement to achieve a more discrete representation that can more flexibly accommodate hidden hierarchical structures and nonlinearity within the measurement space. In analyses of residual yield records, these discrete approaches to information compression ultimately provided encodings that more seamlessly balanced the multimodal feature created by kickoff events, which are likely better attributed to underlying behavioral dynamics and thus naturally more discrete, and more subtle fluctuations in yields about the projected curve, which are more easily attributed to more subtle biological fluctuations that are more naturally considered on a continuous scale. Among the model free encoding schemes, it is also notable that both the ANOHT and entropy spectrum encoding strategies were considerably more computationally efficient than the Earth Movers Distance encoding. Further, all model free encoding schemes were arguably more easily interpreted than the summary statistics encoded for this application with highly multimodal data.

As in previous studies, all encoding of marginal residual milk yield were highly correlated with health and fertility records. Discretization of these distributional features, however, here provided some new evidence of additional nonlinear complexity in these relationships not previously reported using linear modeling-based approaches needed to recover genetic relationships. These results here indicate that further research may be needed to fully disentangle variability in milk yield attributable to resilience to perturbations in the environment and desirable biological perturbations that may be attributed to estrous behaviors. This result suggests that

multiple PLF data streams will likely need to be integrated to create more comprehensive measures of resilience that avoid any unintended selection pressures. In this effort the novel approaches to information compression of plasticity patterns presented in this work arguably offer the most potential, as discrete frameworks offer far more flexible options to expanding resiliency measures to a wider range of PLF data formats – namely data streams that are inherently discrete, nonlinear, or even multivariate in their nature. Future work will consider how these analytical techniques might be used to encode for resiliency in behavioral responses, with careful consideration of preliminary evidence of compensatory relationships between biological and behavioral relationships recovered in these investigations with lactation records.

References

Agresti, A., 2013. Categorical Data Analysis, 3rd ed, Wiley Series in Probability and Statistics.

John Wiley & Sons, Inc, Hoboken NJ.

Bar-Yam, Y., 2004. Multiscale Complexity & Entropy. *Advs. Complex Syst.* 07, 47–63.

<https://doi.org/10.1142/S0219525904000068>

Beauchemin, K.A., 2018. Invited review: Current perspectives on eating and rumination activity in dairy cows. *Journal of Dairy Science* 101, 4762–4784. [https://doi.org/10.3168/jds.2017-](https://doi.org/10.3168/jds.2017-13706)

[13706](https://doi.org/10.3168/jds.2017-13706)

Beggs, D.S., Jongman, E.C., Hemsworth, P.H., Fisher, A.D., 2018. Short communication:

Milking order consistency of dairy cows in large Australian herds. *Journal of Dairy Science* 101, 603–608. <https://doi.org/10.3168/jds.2017-12748>

- Ben Abdelkrim, A., Puillet, L., Gomes, P., Martin, O., 2021. Lactation curve model with explicit representation of perturbations as a phenotyping tool for dairy livestock precision farming. *Animal* 15, 100074. <https://doi.org/10.1016/j.animal.2020.100074>
- Berghof, T.V.L., Poppe, M., Mulder, H.A., 2019. Opportunities to Improve Resilience in Animal Breeding Programs. *Front. Genet.* 9, 692. <https://doi.org/10.3389/fgene.2018.00692>
- Brito, L.F., Bedere, N., Douhard, F., Oliveira, H.R., Arnal, M., Peñagaricano, F., Schinckel, A.P., Baes, C.F., Miglior, F., 2021. Review: Genetic selection of high-yielding dairy cattle toward sustainable farming systems in a rapidly changing world. *Animal* 15, 100292. <https://doi.org/10.1016/j.animal.2021.100292>
- Christensen, R., Johnson, W., Branscum, A., Hanson, T., 2011. *Bayesian Ideas & Data Analysis*, Texts in Statistical Science Series. Taylor and Francis Group, LLC, Boca Raton, USA.
- Colditz, I.G., Hine, B.C., 2016. Resilience in farm animals: biology, management, breeding and implications for animal welfare. *Anim. Prod. Sci.* 56, 1961. <https://doi.org/10.1071/AN15297>
- Computational Statistics, 2013. . John Wiley & Sons, Inc., Hoboken, NJ, USA. <https://doi.org/10.1002/9781118555552.fmatter>
- Dingemanse, N.J., Araya-Ajoy, Y.G., 2015. Interacting personalities: behavioural ecology meets quantitative genetics. *Trends in Ecology & Evolution* 30, 88–97. <https://doi.org/10.1016/j.tree.2014.12.002>
- Elgersma, G.G., de Jong, G., van der Linde, R., Mulder, H.A., 2018. Fluctuations in milk yield are heritable and can be used as a resilience indicator to breed healthy cows. *Journal of Dairy Science* 101, 1240–1250. <https://doi.org/10.3168/jds.2017-13270>
- Esposito, G., Irons, P.C., Webb, E.C., Chapwanya, A., 2014. Interactions between negative energy balance, metabolic diseases, uterine health and immune response in transition dairy

cows. *Animal Reproduction Science* 144, 60–71.

<https://doi.org/10.1016/j.anireprosci.2013.11.007>

Friggens, N.C., Adriaens, I., Boré, R., Cozzi, G., Jurquet, J., Kamphuis, C., Leiber, F., Lora, I., Sakowski, T., Statham, J., De Haas, Y., 2022. Resilience: reference measures based on longer-term consequences are needed to unlock the potential of precision livestock farming technologies for quantifying this trait. *Peer Community Journal* 2, e38.

<https://doi.org/10.24072/pcjournal.136>

Friggens, N.C., Duvaux-Ponter, C., Etienne, M.P., Mary-Huard, T., Schmidely, P., 2016.

Characterizing individual differences in animal responses to a nutritional challenge: Toward improved robustness measures. *Journal of Dairy Science* 99, 2704–2718.

<https://doi.org/10.3168/jds.2015-10162>

Fushing, H., Roy, T., 2018. Complexity of possibly gapped histogram and analysis of histogram.

R. Soc. open sci. 5, 171026. <https://doi.org/10.1098/rsos.171026>

Gadbury, J.C., 1975. Some preliminary field observations on the order of entry of cows into

herringbone parlours. *Applied Animal Ethology* 1, 275–281. [https://doi.org/10.1016/0304-3762\(75\)90020-6](https://doi.org/10.1016/0304-3762(75)90020-6)

Garcia-Baccino, C.A., Marie-Etancelin, C., Tortereau, F., Marcon, D., Weisbecker, J.-L.,

Legarra, A., 2021. Detection of unrecorded environmental challenges in high-frequency recorded traits, and genetic determinism of resilience to challenge, with an application on feed intake in lambs. *Genet Sel Evol* 53, 4. <https://doi.org/10.1186/s12711-020-00595-x>

Ghalambor, C.K., Angeloni, L.M., Carroll, S.P., 2010. Behavior as Phenotypic Plasticity, in:

Evolutionary Behavioral Ecology. pp. 90–107.

Hart, B.L., n.d. *Biological Basis of the Behavior of Sick Animals* 15.

- Hsieh, F., Chou, E.P., Chen, T.-L., 2021. Mimicking Complexity of Structured Data Matrix's Information Content: Categorical Exploratory Data Analysis. *Entropy* 23, 594.
<https://doi.org/10.3390/e23050594>
- Hsieh, F., Liu, S.-Y., Hsieh, Y.-C., McCowan, B., 2018. From patterned response dependency to structured covariate dependency: Entropy based categorical-pattern-matching. *PLoS ONE* 13, 1–28. <https://doi.org/10.1371/journal.pone.0198253>
- Koenker, R., 2018. quantreg: Quantile Regression. R package version 5.36.
- Kok, A., Tsousis, G., Niozas, G., Kemp, B., Kaske, M., van Kneegsel, A.T.M., 2021. Short communication: Variance and autocorrelation of deviations in daily milk yield are related with clinical mastitis in dairy cows. *Animal* 15, 100363.
<https://doi.org/10.1016/j.animal.2021.100363>
- Lefcourt, A.M., 1999. Method to Monitor the Precision of Milk Yields Recorded at Individual Milking Stalls on a Daily Basis. *Journal of Dairy Science* 82, 953–956.
[https://doi.org/10.3168/jds.S0022-0302\(99\)75314-2](https://doi.org/10.3168/jds.S0022-0302(99)75314-2)
- MacKay, D.J.C., 2003. Information Theory, Inference, and Learning Algorithms, 4th ed. Cambridge University Press, Cambridge UK.
- Manriquez, D., Chen, L., Albornoz, G., Velez, J., Pinedo, P., 2018. Case Study: Assessment of human-conditioned sorting behavior in dairy cows in farm research trials. *The Professional Animal Scientist* 34, 664–670. <https://doi.org/10.15232/pas.2018-01749>
- Manriquez, D., Chen, L., Melendez, P., Pinedo, P., 2019. The effect of an organic rumen-protected fat supplement on performance, metabolic status, and health of dairy cows. *BMC Vet Res* 15, 1–14. <https://doi.org/10.1186/s12917-019-2199-8>
- McVey, C., Hsieh, F., Manriquez, D., Pinedo, P., Horback, K., 2021. Livestock Informatics Toolkit: A Case Study in Visually Characterizing Complex Behavioral Patterns across

- Multiple Sensor Platforms, Using Novel Unsupervised Machine Learning and Information Theoretic Approaches. *Sensors* 22, 1. <https://doi.org/10.3390/s22010001>
- McVey, C., Hsieh, F., Manriquez, D., Pinedo, P., Horback, K., 2020. Mind the Queue: A Case Study in Visualizing Heterogeneous Behavioral Patterns in Livestock Sensor Data Using Unsupervised Machine Learning Techniques. *Front. Vet. Sci.* 7, 523. <https://doi.org/10.3389/fvets.2020.00523>
- Miglior, F., Fleming, A., Malchiodi, F., Brito, L.F., Martin, P., Baes, C.F., 2017. A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. *Journal of Dairy Science* 100, 10251–10271. <https://doi.org/10.3168/jds.2017-12968>
- Moncur, V.S., Hardie, L.C., Dechow, C.D., 2021. Genetic analysis of daily milk yield variability in Holstein dairy cattle in an experimental herd. *Livestock Science* 244, 104397. <https://doi.org/10.1016/j.livsci.2021.104397>
- Ouweltjes, W., Spoelstra, M., Ducro, B., de Haas, Y., Kamphuis, C., 2021. A data-driven prediction of lifetime resilience of dairy cows using commercial sensor data collected during first lactation. *Journal of Dairy Science* 104, 11759–11769. <https://doi.org/10.3168/jds.2021-20413>
- Polikarpus, A., Kaart, T., Mootse, H., De Rosa, G., Arney, D., 2015. Influences of various factors on cows' entrance order into the milking parlour. *Applied Animal Behaviour Science* 166, 20–24. <https://doi.org/10.1016/j.applanim.2015.02.016>
- Poppe, M., Bonekamp, G., van Pelt, M.L., Mulder, H.A., 2021a. Genetic analysis of resilience indicators based on milk yield records in different lactations and at different lactation stages. *Journal of Dairy Science* 104, 1967–1981. <https://doi.org/10.3168/jds.2020-19245>

- Poppe, M., Mulder, H.A., van Pelt, M.L., Mullaart, E., Hogeveen, H., Veerkamp, R.F., 2022. Development of resilience indicator traits based on daily step count data for dairy cattle breeding. *Genet Sel Evol* 54, 21. <https://doi.org/10.1186/s12711-022-00713-x>
- Poppe, M., Mulder, H.A., Veerkamp, R.F., 2021b. Validation of resilience indicators by estimating genetic correlations among daughter groups and with yield responses to a heat wave and disturbances at herd level. *Journal of Dairy Science* 104, 8094–8106. <https://doi.org/10.3168/jds.2020-19817>
- Poppe, M., Veerkamp, R.F., van Pelt, M.L., Mulder, H.A., 2020. Exploration of variance, autocorrelation, and skewness of deviations from lactation curves as resilience indicators for breeding. *Journal of Dairy Science* 103, 1667–1684. <https://doi.org/10.3168/jds.2019-17290>
- Ram, N., Benson, L., Brick, T.R., Conroy, D.E., Pincus, A.L., 2017. Behavioral landscapes and earth mover's distance: A new approach for studying individual differences in density distributions. *Journal of Research in Personality* 69, 191–205. <https://doi.org/10.1016/j.jrp.2016.06.010>
- Rathore, A.K., 1982. Order of cow entry at milking and its relationships with milk yield and consistency of the order. *Applied Animal Ethology* 8, 45–52. [https://doi.org/10.1016/0304-3762\(82\)90131-6](https://doi.org/10.1016/0304-3762(82)90131-6)
- Reinhardt, V., 1982. Movement Orders and Leadership in a Semi-Wild Cattle Herd. *Behav* 83, 251–264. <https://doi.org/10.1163/156853983X00183>
- Reith, S., Hoy, S., 2018. Review: Behavioral signs of estrus and the potential of fully automated systems for detection of estrus in dairy cattle. *Animal* 12, 398–407. <https://doi.org/10.1017/S1751731117001975>

- Revilla, M., Friggens, N.C., Broudiscou, L.P., Lemonnier, G., Blanc, F., Ravon, L., Mercat, M.J., Billon, Y., Rogel-Gaillard, C., Le Floch, N., Estellé, J., Muñoz-Tamayo, R., 2019. Towards the quantitative characterisation of piglets' robustness to weaning: a modelling approach. *Animal* 13, 2536–2546. <https://doi.org/10.1017/S1751731119000843>
- Scheffer, M., Bolhuis, J.E., Borsboom, D., Buchman, T.G., Gijzel, S.M.W., Goulson, D., Kammenga, J.E., Kemp, B., van de Leemput, I.A., Levin, S., Martin, C.M., Melis, R.J.F., van Nes, E.H., Romero, L.M., Olde Rikkert, M.G.M., 2018. Quantifying resilience of humans and other animals. *Proc. Natl. Acad. Sci. U.S.A.* 115, 11883–11890. <https://doi.org/10.1073/pnas.1810630115>
- Schein, M.W., Fohrman, M.H., 1955. Social dominance relationships in a herd of dairy cattle. *The British Journal of Animal Behaviour* 3, 45–55. [https://doi.org/10.1016/S0950-5601\(55\)80012-3](https://doi.org/10.1016/S0950-5601(55)80012-3)
- Schuhmacher, D., Bähre, B., Gottschlich, C., Hartmann, V., Heinemann, F., Schmitzer, B., 2019. transport: Computation of Optimal Transport Plans and Wasserstein Distances. R package version 0.12-2.
- Sih, A., Bell, A.M., Johnson, J.C., Ziemba, R.E., 2004. Behavioral Syndromes: An Integrative Overview. *The Quarterly Review of Biology* 79, 241–277. <https://doi.org/10.1086/422893>
- Tucker, C.B., Jensen, M.B., de Passillé, A.M., Hänninen, L., Rushen, J., 2021. Invited review: Lying time and the welfare of dairy cows. *Journal of Dairy Science* 104, 20–46. <https://doi.org/10.3168/jds.2019-18074>
- Wilmink, J.B.M., 1987. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livestock Production Science* 16, 335–348. [https://doi.org/10.1016/0301-6226\(87\)90003-0](https://doi.org/10.1016/0301-6226(87)90003-0)

Zoubir, A.M., Koivunen, V., Chakhchoukh, Y., Muma, M., 2012. Robust Estimation in Signal Processing: A Tutorial-Style Treatment of Fundamental Concepts. IEEE Signal Process. Mag. 29, 61–80. <https://doi.org/10.1109/MSP.2012.2183773>