Applied Animal Science 39:99–116
https://doi.org/10.15232/aas.2022-02335

WELFARE AND BEHAVIOR: *Invited Review*

INVITED REVIEW: Applications of unsupervised machine learning in livestock behavior: Case studies in recovering unanticipated behavioral patterns from precision livestock farming data streams*

Catherine McVey,[1] Fushing Hsieh,[2] Diego Manriquez,[3] PAS, Pablo Pinedo,[3] PAS, and Kristina Horback[1]†
[1]Department of Animal Science, University of California Davis, 95616; [2]Department of Statistics, University of California Davis, 95616; and [3]Department of Animal Science, Colorado State University, Fort Collins 80523

ABSTRACT

Purpose: The objective of this article is to review fundamental differences between model-dependent and model-free approaches to data analysis, and to explore the potential advantages of more open-ended machine learning approaches in recovering complex behavioral patterns from precision livestock farming data streams.

Sources: Case studies using simulated data were designed to mimic a real-world scenario. Data from a feeding trial in an organic dairy were reanalyzed using the Livestock Informatics Toolkit.

Synthesis: Case studies using simulated data are used to demonstrate how incomplete information about the management system can prohibit the development of an appropriate model for information compression, allowing aggregation bias to mask important behavioral indicators of compromised welfare. These hidden behavioral patterns are then recovered using unsupervised machine learning approaches that are able to leverage the intrinsic behavioral codependencies of group-housed animals. This simulated case study is then extended to demonstrate how model-based approaches can also overlook causes of compromised welfare when the link between environmental factors and behavioral responses is strong but nonlinear, whereas model-free information-theoretic tools can easily recover and characterize such complex dynamics. Finally, in an empirical case study with data from a commercial organic dairy, the Livestock Informatics Toolkit is used to recover from milk parlor metadata complex associations between herd age structure, levels of milk production, and order of milking.

Conclusions and Applications: Model-free machine learning algorithms provide a more open-ended approach to knowledge discovery that require fewer up-front assumptions about the management system. This can yield more comprehensive insights into large precision livestock farming data sets now commonly encountered in on-farm research trials and in applied data auditing scenarios.

Key words: unsupervised machine learning, entropy, behavior, welfare, precision livestock farming

INTRODUCTION

Behavioral data has conventionally been collected using highly trained human observers to encode behaviors of interest predefined by detailed ethograms (Dawkins, 2007). Directly involving research personal in the incoming data stream has traditionally served to nurture greater familiarity with the study environment generating behavioral responses, which can aid in the development and interpretation of downstream statistical models, but also places practical limits on the amount of data that can be generated. In recent years precision livestock farming (**PLF**) technologies have become more widely available (Wathes et al., 2008; Banhazi et al., 2012; Stygar et al., 2021). In experimental settings these tools enable animal scientists to collect behavioral data on larger numbers of animals over longer time periods at far greater sampling frequencies. With wider industry adoption, these technologies also create new opportunities to analyze animal behaviors in working farm behaviors, not only for scientific endeavors but also for more practical applications such as auditing and consulting. Data sets of this scale and granularity often contain a range of more complex stochastic features, such as temporal nonstationarity (seasonal and circadian rhythms), autocorrelation, heterogeneous variance structures, nonindependence between experimental units, and more, that can be difficult to accommodate in convention-

---

The authors have not declared any conflicts of interest.

*This invited article resulted from the presentation given at the ARPAS Symposium, “Artificial Intelligence and Machine Learning in Dairy Production Systems,” Kansas City, Missouri, June 2022.

†Corresponding author: kmhorback@ucdavis.edu

al statistical models, particularly when data are collected remotely and cannot be anticipated a priori, which can in turn lead to spurious inferences (Pinheiro and Bates, 2000).

In data science, there is the formal concept of “no free lunch,” meaning that no single algorithm can be expected to perform optimally in all scenarios or applications (Wolpert and Macready, 1997). Unsupervised machine learning (**UML**), however, may provide animal scientists an alternative approach to extracting behavioral patterns from PLF data streams that is better suited to large and complexly structured data sets (Valletta et al., 2017). Whereas conventional statistical analyses, which were originally developed for application in controlled experimental trials, excel at answering targeted hypotheses, UML algorithms are designed to systematically sift through data sets to identify and characterize any and all nonrandom patterns until only noise remains, offering a more flexible and open-ended approach to knowledge discovery (Kirby, 2001; Fushing et al., 2018). The purpose of this article will be to review, through a series of case studies, how UML algorithms differ from the standard statistical approaches, not only in their technical implementation and efficacy but also in their fundamental analytical philosophies. Using simplified simulated data, we will first introduce the fundamental concepts of information compression and loss to functionally distinguish model-dependent from model-free analytical approaches. Using this framework, we will then demonstrate how model-free UML tools may be used to recover and visualize unexpected behavioral patterns in PLF data streams that might be missed by conventional model-based analyses when the environmental factors driving such responses are not recorded. We will then expand upon this simulation to explore how model-free information-theoretic approaches can then be used to contrast the patterns recovered by UML approaches to recover the complex links between behavioral responses and the management factors that elicit them, across multiple data sets. Finally, through an empirical case study with herd records from a commercial dairy, we will demonstrate how these 2 model-free approaches may be seamlessly integrated, using open-source software available in R to analyze patterns in milk parlor metadata with minimal assumptions about the behavioral patterns that may be lurking therein.

SIMULATION CASE STUDY: DATA COMPRESSION AND INFORMATION LOSS

What is the difference between data and information? Although these 2 terms are often used interchangeably in biology, entire subfields of data science are dedicated to solving this riddle (Kirby, 2001; MacKay, 2003). Although mathematicians have developed formal definitions for these terms (Shannon, 1948; MacKay, 2003), we will attempt to illustrate the conceptual differences with a simplified simulated example.

Suppose a researcher 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 1 accelerometer would be good, then 2 accelerometers must be even better. For their pilot study, they attach 2 sensors onto the same leg of a well-behaved cow and record her walking normally down the alleyway in a straight line. If both sensors were calibrated and synchronized properly, then extracting recordings of the vertical axis of acceleration from either sensor might produce a data set 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 sensor, or vice versa? In this simplified example, beneath the simulated random measurement error attributed to each sensor, the 2 signals are virtually identical. Thus, even though this researcher 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 as more akin to a physical resource—something that can be measured in bytes of memory space. Information, however, should be regarded as a more nebulous unit of measure that represents how much can be learned from a data set. In experimental settings, collecting measurements is typically costly, so sampling strategies are often developed to minimize redundancy between data points, such that data and information become functionally equivalent terms (Pinheiro and Bates, 2000). With many commercially marketed sensor technologies, however, the sampling frequency that researchers may use to generate data sets is often dictated to them by the hardware itself, which can result in a considerable amount of redundancy between datapoints (Kirby, 2001; McVey et al., 2020). With such data sets, information compression strategies are often implemented in an effort to reduce the size of a data set without losing any potentially useful information about the system. Returning to the previous example, suppose the researcher realizes they have not requisitioned enough hard drives to store data from both accelerometers, so they decide to record only the average output of the 2 sensors at each time point, as illustrated in Figure 1B. This simple information compression step reduces the size of their data set by half, but no details about the signal have been sacrificed.

Now suppose that the researcher 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 axes of the 2 signals will become inverted, as illustrated in Figure 1C. Subsequently, if the previously validated rolling average filter were applied to this new data set, the signals would here cancel out, leaving only the measurement noise, as shown in Figure 1D. Although this compression error could be remedied quite easily in this simplified simulation, this example still serves to il-

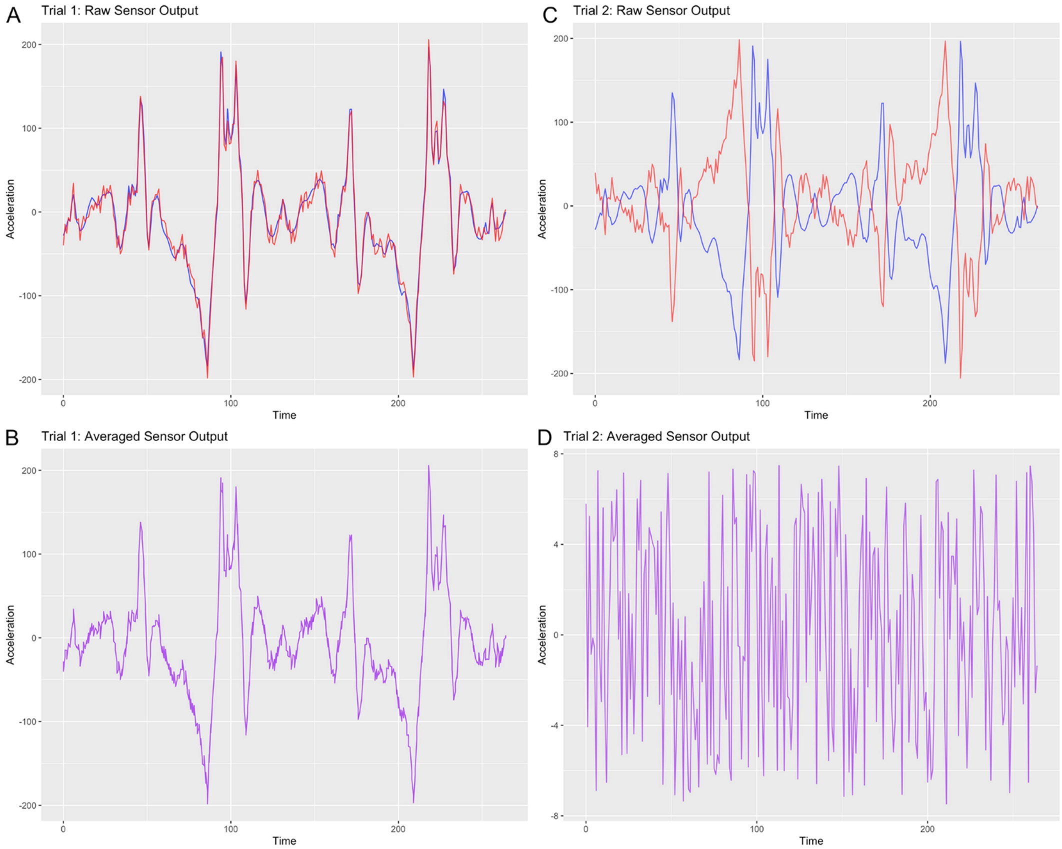


Figure 1. Human gait data collected from the forward axis of a 120-Hz triaxial accelerometer has been modified to help visualize the difference between data and information. Upper left (A): The data collection protocol is implemented flawlessly, such that the red and blue sensors produce nearly identical signals. Lower left (B): Averaging their results at each time point reduces the size of the data set 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. Lower right (D): Assumptions employed by this simple information compression algorithm are now violated. In reducing the data set by half, all information about gait dynamics is lost, the signals having canceled out to leave only measurement noise.

lustrate just how easily biologically relevant patterns can be lost when the assumptions employed in an information compression strategy do not match the realities of the data to which it is applied.

SIMULATION CASE STUDY: 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 simulation. Therefore, to further explore the fundamental differences between model-dependent and model-free approaches to information compression and knowledge discovery, we will next 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 their upcoming welfare audit, the farmer contracts a consultant to analyze data from a commercially marketed 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 whether there is any clear evidence of animals with compromised welfare in 60 d of archival records. Through structured random sampling from Gaussian distributions, a data set that might be realized from such a scenario has been simulated. Full details and reproducible code used to generate this simulation are provided in Supplemental Materials (https://github.com/cgmcvey/ARPAS2023), but readers are encouraged to first consider how the analysis of this data set should be approached without any prior information about its underlying structure.

Suppose the consultant hired is an animal scientist, who, although not necessarily a specialist in welfare and behavior, is fully trained in the 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 techniques and produce the plot shown in Figure 2, wherein each dot represents the proportion of time that a given cow is recorded lying down on a given observation day. A LOESS (LOcally Estimated Scatterplot Smoothing) curve is also

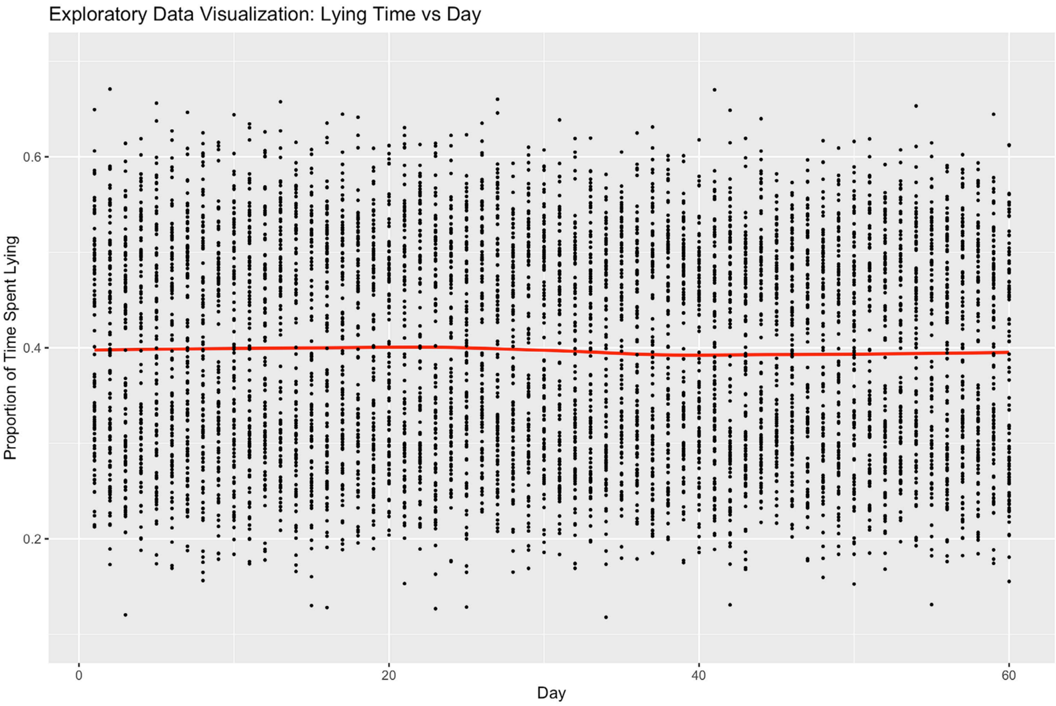


Figure 2. Exploratory data analysis visualization of simulated data set. Each dot represents the proportion of time an individual cow is recorded as lying down on a specific day. The red LOESS (LOcally Estimated Scatterplot Smoothing) fit line represents the average amount of lying time across the herd for a given day.

added to this scatterplot, to reflect the mean lying time for the entire herd at each time point. From this visualization, they glean that a considerable amount of variation between cows exists, 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 is incredibly stable in its average lying time over the observation interval. Based on this visualization, the consultant constructs a mixed-effect model, with cow included as a random effect to account for repeated measures to avoid pseudo-replication, as shown in Equation [1]:

$$Y_{LT} = \beta_1 X_{heifer} + \beta_2 X_{day} + Z_{cow} \qquad [1]$$

where

- Y_{LT} = proportion of day spent lying down by a given cow on a given day;
- $\beta_1 X_{heifer}$ = binary categorical effect (heifer or not heifer);
- $\beta_2 X_{day}$ = linear effect for day of observation;
- Z_{cow} = random effect for consistent individual differences between cows.

Because the data appear 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 (https://github.com/cgmcvey/ARPAS2023). As anticipated from the exploratory data analysis visualizations, the fixed effect for day is neither biologically nor statistically significant ($\hat{B}_2 = -0.0001$, SE = 8.53e −5, T = −1.41). However, the consultant is a bit surprised to find that the difference in average lying time between heifers and cows is also neither statistically or biologically significant ($\hat{B}_1 = -0.0026$, 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 ($\hat{\sigma}_{cow}$ = 2.67e −5) is considerably smaller than the within-cow error term ($\hat{\sigma}_{error}$ = 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, although cows might not get to lie down an adequate amount of time every day, and the average lying time could certainly be improved, there is also no evidence of individual cows in this herd that are consistently critically under-rested.

Now suppose that the milk buyer is dubious of this result, so they send the data to a new consultant who

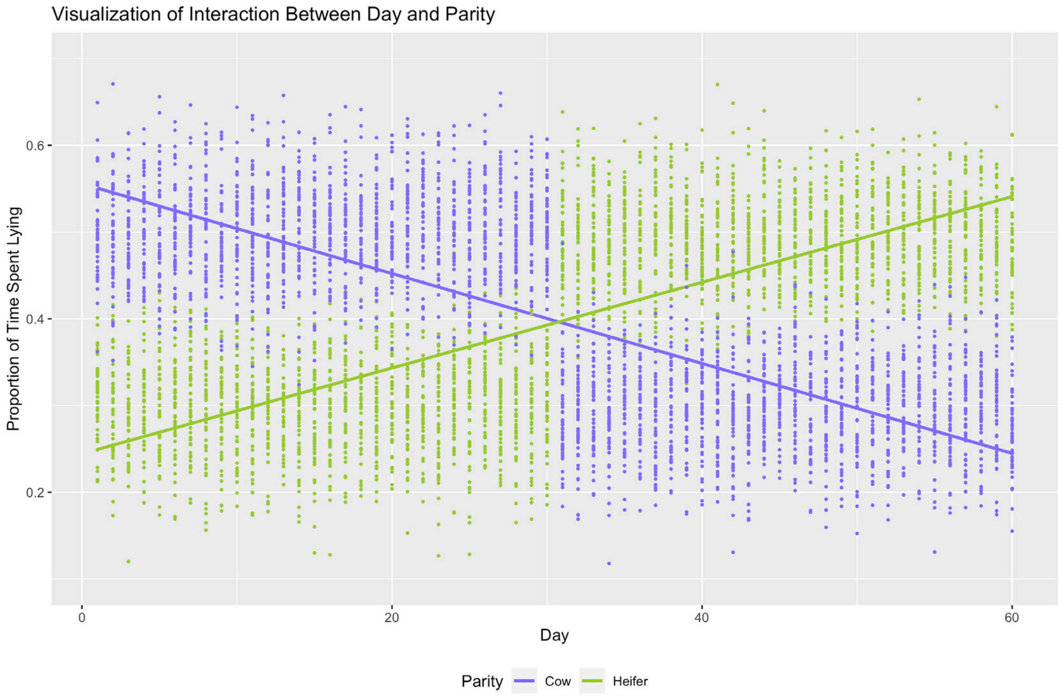


Figure 3. Visualization of the temporal complexity hidden within this data set. Due to the overstocking in this pen, a strong negative correlation is found between subgroups, as cows must compensate for the limited number of freestall spaces. When this physical limitation is removed with the reintroduction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

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. The consultant, suspecting that heifers and cows might react differently to this pasture access, develops a similar model to the first consultant, except that in this set of linear equations, an interaction term is included to allow the temporal dynamics to differ between these subgroups, as shown in Equation [2]:

$$Y_{LT} = \beta_1 X_{heifer} + \beta_2 X_{day} + \beta_3 X_{heifer} \times X_{day} + Z_{cow} \qquad [2]$$

where

- Y_{LT} = proportion of day spent lying down by a given cow on a given day;
- $\beta_1 X_{heifer}$ = binary categorical effect (heifer or not heifer);
- $\beta_2 X_{day}$ = linear effect for day of observation;
- $\beta_3 X_{heifer} \times X_{day}$ = interaction effect between linear effect for day of observation and categorical effect for heifer; and
- Z_{cow} = random effect for consistent individual differences between cows.

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 influence 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 monopolizing the freestall spaces, forcing the heifers to stand; however, once the grazing season starts, the cows leave their freestalls to go graze, and the foot-weary heifers remain in the freestalls.

In the preceding analyses it is important to distinguish between linear models and statistical inferences as 2 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 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 of model-dependent information compression, when the biological system that generates a data set is well understood, this prior knowledge can be infused into an appropriate model to efficiently compress the information contained in such records. However, if we have an incomplete understanding of the system that gives rise to the behavioral responses, a linear model can become an inefficient means of information 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 driven by more complex factors? What if, instead of a single persistent shift in lying patterns after pasture access was granted, the desirability of the freestalls had been influenced by more transient environmental factors such as weather or bedding cleanliness? Alternatively, what if an animal's ability to monopolize these freestall spaces had not been determined by a simple trait such as parity? Resource holding potential is known to be influenced by a range of factors, including size, seniority, energy balance, health status, innate aggressiveness, and more—many of which might not be directly measured in standard production systems (Schein and Fohrman, 1955; Hurd, 2006; Šárová et al., 2013; Hubbard et al., 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 toward the chaos and complexity of commercial farm environments, the more fundamentally challenging it becomes to use model-based approaches for information compression to extract ethological insights from PLF data streams.

Looking beyond the model-dependent tools that dominate conventional experimental statistics, modern machine learning 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 2 general classes (Kirby, 2001; James et al., 2013; Valletta et al., 2017). In supervised machine learning, the user is required to distinguish between explanatory and response variables, but the model that relates these 2 sets of variables is inferred, by varying degrees, from the data itself. Subsequently, this framework lends itself to optimizing the predictive power of a model but may still overlook important patterns in a data set if the factors driving heterogeneity (systematic nonrandomness) in the response variables are not measured and included among the candidate predictor variables. Unsupervised machine learning (**UML**) algorithms, by contrast, 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 data set until only noise remains.

Several algorithmic approaches exist to explore the latent structures (high-dimensional geometry) of large data sets via UML. Neural networks are arguably the most rigorous of such techniques (Goodfellow et al., 2016). Such algorithms work by breaking down large and complex biological phenomena into progressively smaller pieces by passing large data sets through layers of neurons connected by a network architecture. This ground-up approach to learning the key features of a data set, although incredibly powerful, is also incredibly data hungry, with thousands of data points typically required to train deep learning neural architectures. Although unsupervised neural networks may not be as data hungry as their supervised cousins, smaller sample sizes still impose practical limits on the depth of network that may be trained to a data set without risking overfitting. Shallow network architectures, however, can limit the ability of such networks to recover complex (nonlinear) patterns within a data set, with the simplest one-layer networks being able to recover only linear relationships. Future advancements in transfer learning techniques, where the most basic layers of a neural architecture can be borrowed from larger reference databases, may greatly improve the efficiency with which such networks may be trained (Arac et al., 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 may be preferable.

Spectral embedding approaches offer a method of information compression that is suitable for many "wide" problems, where the number of variables in a data set is large relative to the number of observational units (Kirby, 2001). Principal component analysis is the simplest example of this class of algorithms, but more modern embedding approaches are better suited to the complex nonlinear geometric features that can often be found in sensor data. In previous research, however, we have demonstrated that such algorithms are liable to produce embedding artifacts when the space of possible behaviors that may be demonstrated by animals are severely constrained by factors such as spatial restrictions intrinsic to the housing systems or temporal constraints imposed by the management schedule (McVey et al., 2020). For data sets subject to such dynamics, this then leaves clustering algorithms.

Several algorithmic strategies may be employed to compress the complex geometric features of high dimensional data sets into discrete clusters (Valletta et al., 2017; Adamczyk et al., 2017; Rufener et al., 2018; 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 clustering algorithm to implement, and subsequently has seen some adoption in analyses of livestock data. Although 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. Neural networks, such as autoencoders, may be used to create unsupervised clusters and can be parameterized to learn cluster numbers from the data itself but, again, require larger sample sizes to do so. For smaller data sets, hierarchical clustering (**HC**) algorithms may provide a more open-ended approach to developing an optimal discrete encoding of such high-dimensional patterns.

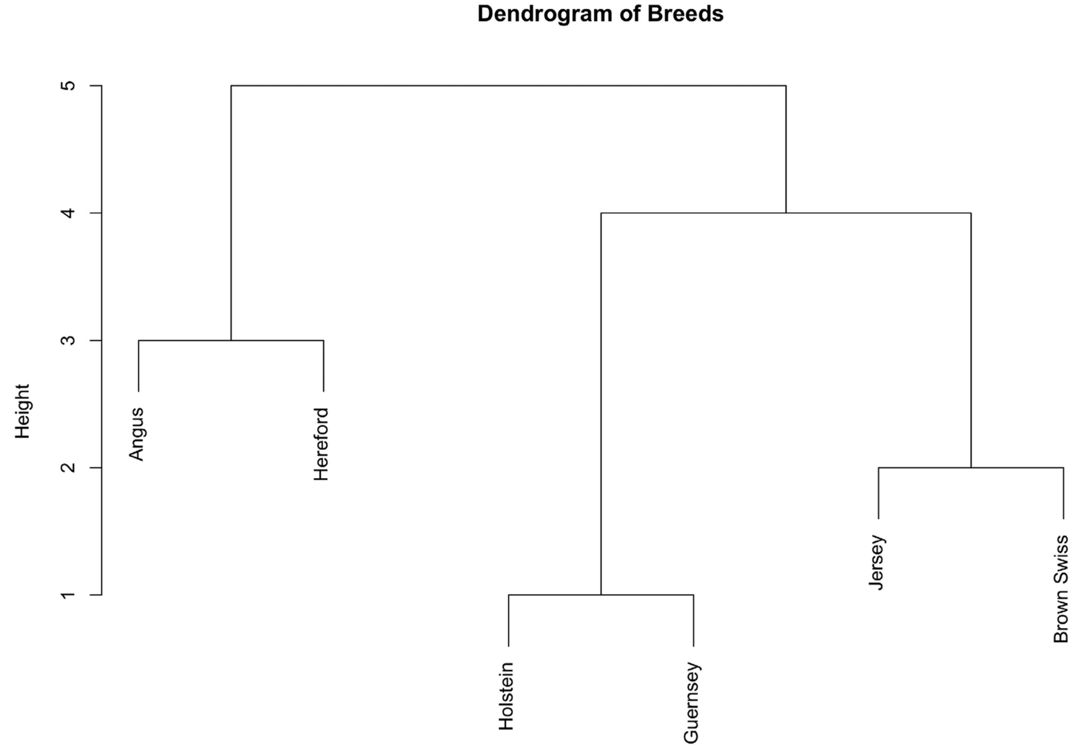


Figure 4. Through sequential agglomerative steps, breeds are progressively paired together based on their relative similarity. The resulting dendrogram provides a 2-dimensional visualization of how these breeds are distributed within a much higher-dimensional space of phenotypic traits.

As its name implies, HC 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 feedstore and pick up a pack of toy cows, each of a different breed. You line them up in front of a toddler and ask which 2 cows look the most alike. They point to the Guernsey and the Holstein, so you pull them forward and stand them together. You ask this question again, and 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 Herford cows are paired together. You ask again, and 2 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, as presented in Figure 4. This schematic, properly called a dendrogram, provides a succinct 2-dimensional representation of how a fairly large number of phenotypic features are distributed within your plastic herd.

Hierarchical clustering algorithms seek to mimic this intuitive agglomerative process using objective mathematical constructs. To demonstrate how this might work, let us return to our previous example of the farmer with overstocked dairy cows. As before, our larger mature cows will monopolize the freestalls 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 data set is to compute a dissimilarity matrix, which is a square symmetric matrix containing quantitative estimates of the dissimilarity between the data vectors for each pair of observational units in the corresponding row and column indices. To cluster cows together with similar lying patterns, we will first calculate a dissimilarity matrix using the Euclidean distance or L2 norm (Equation [3]), which here is just the squared differences in observed lying times between a given pair of cows summed over all observation days. Similarly, to cluster together days wherein the herd demonstrated similar lying patterns in response to environmental factors, we will calculate the Euclidean distance (Equation [3]) 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 (2-dimensional) linkage method, wherein clusters are merged to produce the largest increase in between-group variance at each agglomerative step:

$$\mathrm{d}\left(\vec{\mathbf{p}},\vec{\mathbf{q}}\right)=\left\|\vec{\mathbf{p}}-\vec{\mathbf{q}}\right\|_2=\sqrt{\sum_{i=1}^{k}\left(\vec{\mathbf{p}}_i-\vec{\mathbf{q}}_i\right)^2}, \qquad [3]$$

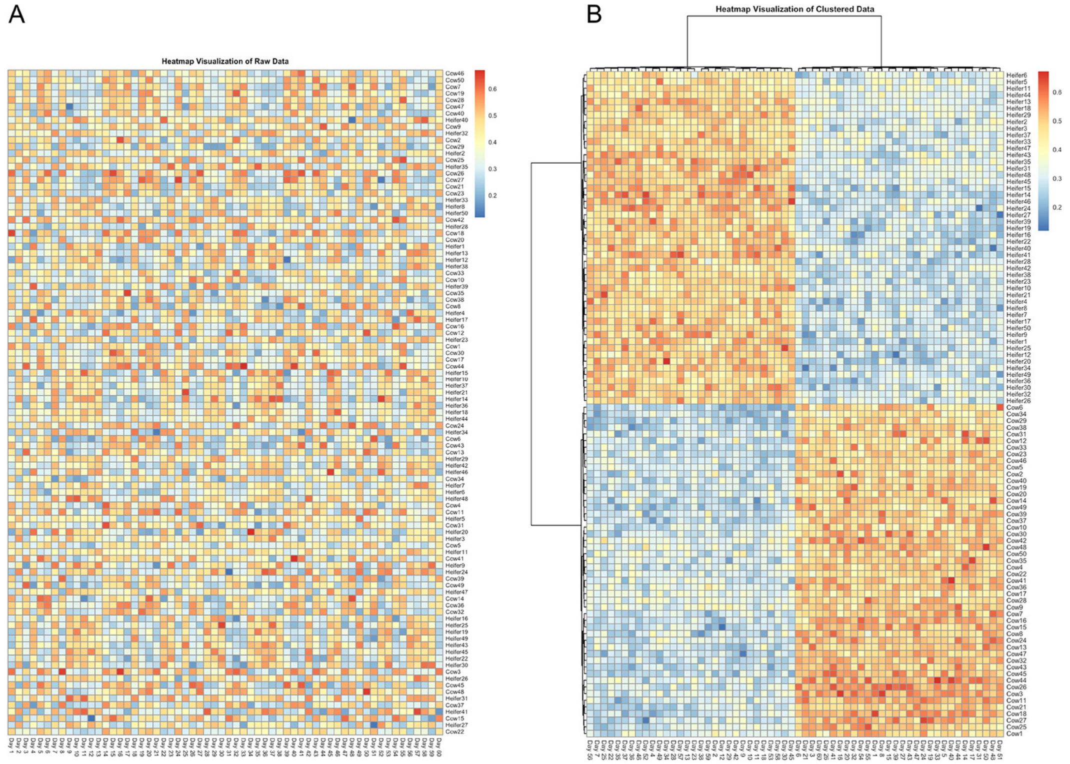


Figure 5. Heatmap visualization of the lying time data matrix. Cows are arranged on the row axis, observation days are arranged on the column axis, and the color of each cell represents the proportion of time that a given cow is recorded as lying down on a given day. In 5A (left) no clear patterns in 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.

where

- $\vec{\mathbf{p}}$ = data vector for observational unit a (animal, day, etc.) with $i = 1, \ldots, k$ recorded variables;
- $\vec{\mathbf{q}}$ = data vector for observational unit b (animal, day, etc.) with $i = 1, \ldots, k$ recorded variables.

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 as 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 that cows and observation days with similar temporal patterns in lying times are grouped together on the row and column axis, respectively. Visualizing these results again, using a heatmap as in Figure 5B, we can now see that this clustering algorithm has captured 2 distinct groups of cows and 2 distinct groups of observation days, and so the inversion in lying patterns within the data are now quite visually striking. Thus, even though this HC 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, a consultant 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.

SIMULATION CASE STUDY: COMPARING MODEL-DEPENDENT AND MODEL-FREE APPROACHES TO IDENTIFYING BIVARIATE ASSOCIATIONS

The preceding example illustrates that, by leveraging intrinsic codependencies 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, a consultant or researcher would typically want to try to pin down the variables eliciting such reactions. In the preceding simulation, the mechanism linking lying time, age, and weather was fairly simple, and so, armed with the insights from the UML analyses, several straightforward linear methods might be employed to probe for the causative variables among farm records. But what if the link between environmental variables and the behavioral responses elicited were more complex? We have 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 us now consider how model-free approaches might be extended to characterize complex behavioral patterns not only within but also across data sets.

Several supervised machine learning approaches are specifically designed to probe for significant associations among 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, 2 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 numbers of candidate predictors are often quite large (James et al., 2013). Even when sufficient sample sizes are available, such techniques inherently prioritize prediction over intuition. Although variable importance estimates may provide some insights into the key drivers of a behavioral response, it can be difficult, using 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 complex bivariate associations between data sets, which naturally compliments UML clustering techniques. Information entropy (H), as calculated using Equation [4], 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 are often called the "vocabulary" or "dictionary," and are here indexed as $k = 1, . . ., K$, wherein K is the total number of discrete levels. For applications in animal behavior, however, this dictionary is typically just the ethogram used. Although logarithmic terms are not often known for their interpretability, here they confer several convenient algebraic properties to entropy estimates, making this estimator intuitive as a relative measure:

$$\mathrm{H}(x) = \sum_{k=1}^{K} \mathrm{P}(X = k)\log_2\left[\frac{1}{\mathrm{P}(X = k)}\right], \qquad [4]$$

where

- $\mathrm{P}(X = k)$ = proportion of observations recorded as a given discrete value (k);
- K = total number of discrete values that variable X may assume.

Suppose a data set were collected wherein each observational unit is coded with 1 of 2 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? Because 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 toward or away from 1 of these 2 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 1 behavior were ever observed, then we could guess the encoded value of any observation in our data set with complete certainty, and so the corresponding entropy estimate would drop to 0. Figure 6 demonstrates how the entropy estimate of a binary variable changes with the symmetry of the distribution of the observed binary 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 used in conventional statistical inference frameworks for linear models.

This model-free framework can subsequently be extended to a multivariate estimator with 2 or more discrete variables. In the bivariate case, the distribution of one variable is compared across each encoded level of the other to decompose, as in Equation [4], the total entropy in the joint encoding into 3 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 2 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 2 sets of encoded data regardless of the underlying dynamic (linear, quadratic, exponential, etc.):

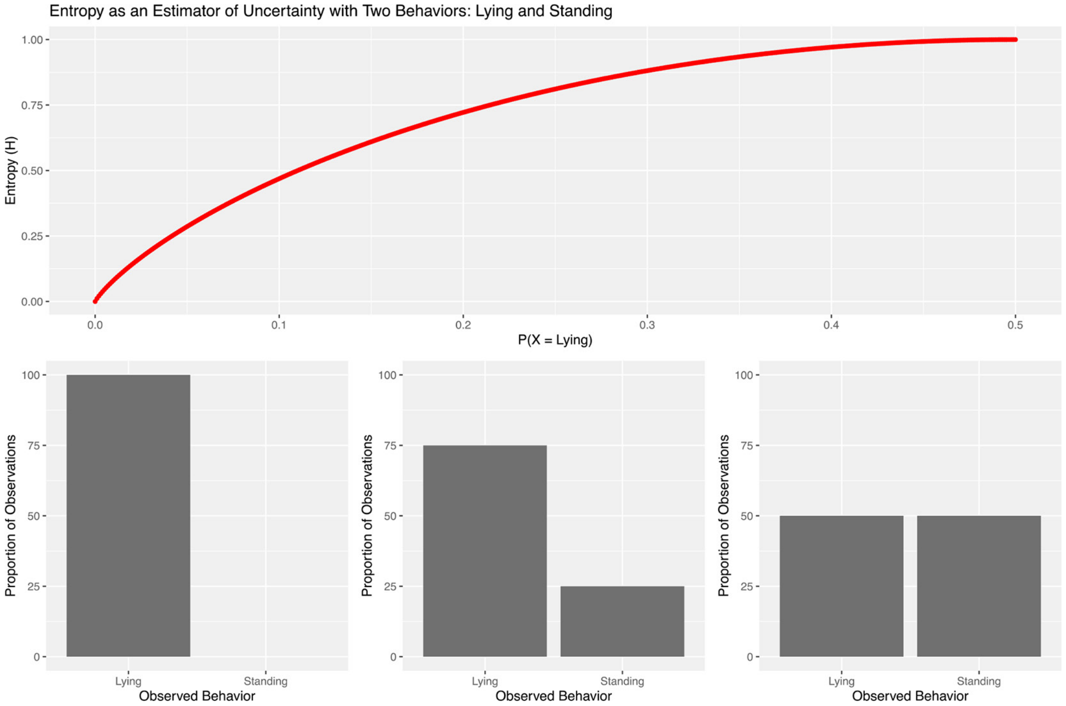


Figure 6. A visualization of how entropy (H) estimates vary with changes in the relative distribution of discrete observations for a simple K = 2 encoding (ethogram). When behaviors "lying" and "standing" are recorded at equal frequencies P(X = "lying") = P(X = "standing") = 0.5, the entropy or uncertainty in this behavioral system is maximized at $\log_2(2) = 1$. P = proportion. As one behavior (lying) is observed more frequently, the entropy of the system progressively decreases. When only one behavior (lying) is observed, then no uncertainty remains in the system and entropy drops to 0.

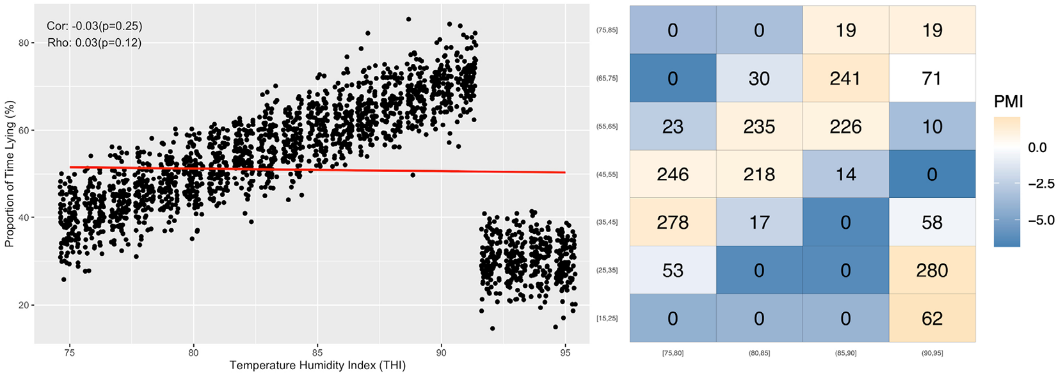
$$\mathrm{H}(X,\ Y) = \mathrm{H}(X|Y) + \mathrm{H}(Y|X) + \mathrm{I}(X,\ Y), \qquad [5]$$

where

- $\mathrm{H}(X,\ Y)$ = total entropy of joint encoding of discrete variables X and Y;
- $\mathrm{H}(X|Y)$ = information about encoding X that could not be gleaned from known value of Y;
- $\mathrm{H}(Y|X)$ = information about encoding Y that could not be gleaned from known value of X;
- $\mathrm{I}(X,\ Y)$ = information that is redundant between encodings X and Y.

To demonstrate the efficacy of an information-theoretic framework for identifying complex bivariate associations over more conventional model-based approaches, let us return to the previous consulting 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 the milk buyer that the issue has been resolved. After several months at this lower stocking rate, the farmer reviews the data and is dismayed to find that there are still days where animals have inadequate lying times. To solve this new problem, the farmer hires yet another data consultant.

Suppose that this new data set was 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 (https://github.com/cgmcvey/ARPAS2023), we have simulated a fairly straightforward but nonlinear dynamic between these 2 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 when 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 used to probe for a significant bivariate association between these 2 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,



	[75,80]	(80,85]	(85,90]	(90,95]
(75,85]	0	0	19	19
(65,75]	0	30	241	71
(55,65]	23	235	226	10
(45,55]	246	218	14	0
(35,45]	278	17	0	58
(25,35]	53	0	0	280
[15,25]	0	0	0	62

Figure 7. Two visualizations of the simulated relationship between temperature-humidity index (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 to move heat more effectively at THI values above 92 THI. This nonlinear dynamic, however, is not captured by a simple linear association assumed in the Pearson correlation test (Cor), which is visualized with the red line imposed over the data, or the monotonically increasing association assumed by the Spearman rank order test (Rho). (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 (PMI), which reflects cells that are significantly over- or underpopulated when compared with 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.

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.

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. Using an arbitrary bin width of 5, this will create a discrete encoding with 4 levels for THI and an encoding with 7 levels for lying time. If we compare the mutual information estimate from the 2 observed encodings against estimates generated from a simple nonconditional permutation of sample indices, as in a standard permutation test (Higgins 2004), 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 2 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 overrepresented relative to the null. From this visualization we can clearly see that the probability of observing a given lying pattern is shifted in different directions based on the level of the THI encoding. Thus, absent any prior intuition or assumptions about the relationship between these 2 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.

EMPIRICAL CASE STUDY: MODEL-FREE KNOWLEDGE DISCOVERY WITH MILK PARLOR METADATA

Although we hope that the previous simulations have served to illustrate the potential advantages of a more open-ended model-free approach to knowledge discovery in large PLF data sets, the behavioral patterns recovered in these simulated examples are fairly simple when compared with the complex web of behavioral responses that might be elicited from the environmental and social stimuli present in a working farm environment. Therefore, to demonstrate the utility of this analytical framework in real-world applications, we will finally consider data collected from the parlor on a working dairy farm using the fully model-free analytical pipeline available in the Livestock Informatics Toolkit, an open-source package available in the R programing environment (McVey et al., 2021).

The Organilac data set was collected in 2017 from a feed trial assessing the influence of an organic fat supplement on cow health and productivity during the first 150 d of lactation. The study ran from January through July on a USDA-certified organic dairy in northern Colorado, and all animal handling and experimental protocols were approved by the Colorado State University Institution of Animal Care and Use Committee (protocol ID: 16-6704AA).

Full details of the experimental design can be found in Manriquez et al. (2018, 2019). A total of 200 cows were enrolled over a 1.5-mo period into a mixed-parity herd of animals with predominantly Holstein genetics. Cows were maintained in a closed herd for the duration of the trial, in an open-sided freestall barn, which was stocked at roughly half capacity with respect to both feedbunk spaces and stalls. Cows had free access to TMR and an adjacent outdoor dry lot when in their home pen, and, beginning in April, cows were moved onto pasture at night in accordance with organic grazing standards. Cows were milked 3 times a day in a radio-frequency identification-equipped rotary parlor (DeLaval), which not only quantified daily milk production weights but also recorded milking order, or the or the sequence in which cows enter the parlor to be milked.

Milking order has been the focus of several research papers since the 1960s (Kilgour and Scott, 1959; Gadbury, 1975; Soffié et al., 1976; Rathore, 1982; Reinhardt, 1983; Berry and McCarthy, 2012; Polikarpus et al., 2015; Beggs et al., 2018; McVey et al., 2020). Almost all of these studies have confirmed that queuing patterns are nonrandom and are surprisingly consistent over time (Beggs et al., 2018), but the factors determining where a cow ends up in the queue have proven difficult to pin down. Although it is quite possible that these dynamics vary between farms based on differences in herd structure and farm environment, inconsistencies in previous results may also be partially attributed to the fact that this behavioral system violates nearly every assumption required to extract statistical inferences from a conventional linear model—namely, independence and homogeneity of variance (McVey et al., 2020). In previous research with entry order records, we have demonstrated the efficacy of an iterative HC approach known as data mechanics in recovering subtle nonstationarity in the queuing positions of a subset of animals within this herd over an extended 80-d observation window (Guan and Hsieh, 2018; McVey et al., 2020).

To illustrate the utility of significant expansions to the analytical tools now available in the LIT package (McVey et al., 2021), we will revisit these results to further explore how queuing behaviors might relate to cow age and milk yield. These 2 variables are typically quite highly correlated, which, in a linear model, can cause variance inflation and model instability (Pinheiro and Bates 2000). However, they are not completely redundant, as there may be biologically relevant information on the productivity of a cow relative to her age. One strategy to address this issue would be to normalize each cow's yield against the average for her lactation cohort. On organic dairies, however, limited access to hormonal synchronization protocols can cause cow age to vary progressively with lactation number, which can reduce the efficacy of this simple normalization scheme. Alternatively, by encoding the information in these 2 variables via HC, this redundancy can be leveraged to establish empirically determined cutoffs.

The results of jointly clustering cow age in days and the 95th daily yield quantile for each cow (a model-free estimate of peak yield) are shown in Figure 8, wherein each variable has been independently normalized to receive equal weights, using the *encodePlot* function in the LIT package (McVey et al., 2021). By examining the branches of the clustering dendrogram used to order the row observations, we see that the first several clusters reflect the redundancy between age and yield. The first cut isolates the heifers into their own cluster, although 2 very young and low-yielding second-lactation animals were also included in this grouping. The second cut served to distinguish the still-growing second-lactation cows from the fully grown cows in parity 4 or more, with similarities in yield being used to divide the third-lactation animals between these 2 groups. All remaining cuts served predominantly to distinguish yield levels within these age clusters. Note that this dynamic can also be easily visualized by parsing through visualizations of each progressive cut provided in Supplemental Materials (https://github.com/cgmcvey/ARPAS2023).

With this information encoded, a mutual information test can be used to evaluate whether a significant bivariate association exists between the encoding of age and yield and the previously reported encodings for queuing patterns. For both these data sets, however, the optimal number of clusters to use to create a discrete encoding is not immediately obvious. If too coarse an encoding is used, valuable information may be lost and significant associations overlooked. On the other hand, if too fine an encoding is used, the mutual information test will become underpowered. To overcome this issue, a bivariate tree testing framework can be employed, wherein mutual information tests are performed for all combinations of encoding granularities, to take full advantage of the geometric information encoded in the clustering trees created from either data set (McVey et al., 2021). Conceptually, this test is analogous to examining a sample under a light microscope when the organisms present are unknown. The test begins at the coarsest resolution, and the focus is slowly sharpened. Just as microbes of different sizes would come into and fall out of resolution in a light microscope as the focus is sharpened, we can expect in a bivariate tree test that behavioral mechanisms that produce associations at different scales will similarly fall into and out of resolution at varying encoding granularities (Fushing et al., 2018; McVey et al., 2020, 2021).

The result of a bivariate tree test applied to the age and yield encoding against the encoding of entry order patterns for all cows is shown in Figure 9. The heatmap displays z-scores comparing the observed mutual information estimate against the randomly permuted tests statistics, revealing that a highly significant bivariate association ($P = 0.001$) emerges with 3 clusters for the age and yield encoding and 6 clusters for the queuing pattern encoding. In this example, the number of clusters used to encode temporal patterns in the entry order data was also varied on

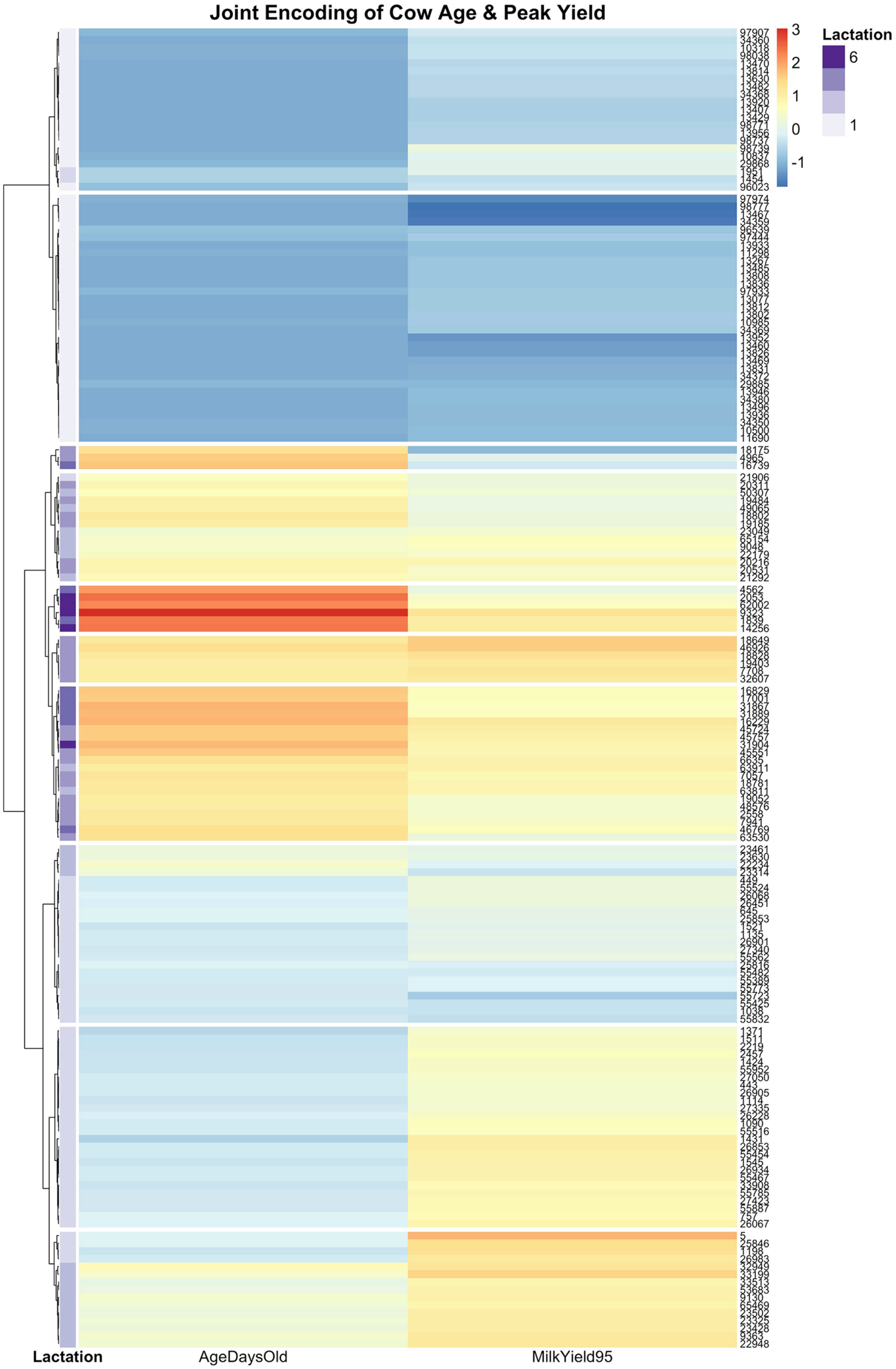


Figure 8. Heatmap visualization of the joint encoding of age in days (AgeDaysOld) and peak yield (95th quantile of daily yield estimates, MilkYield95). Cows are arranged along the row axis (zoom in to read cow ID labels). Age and yield records are arranged along the column axis. Each cell is colored to represent the value of a given record field for a given cow, wherein both age and yield are represented by their normalized z-scores (centered and squared). The row dendrogram reveals that the redundancy between age and yield through the first few lactations drives the first 2 cuts of this clustering tree, but subsequent divisions serve predominantly to distinguish between yield levels within these broader age classes.

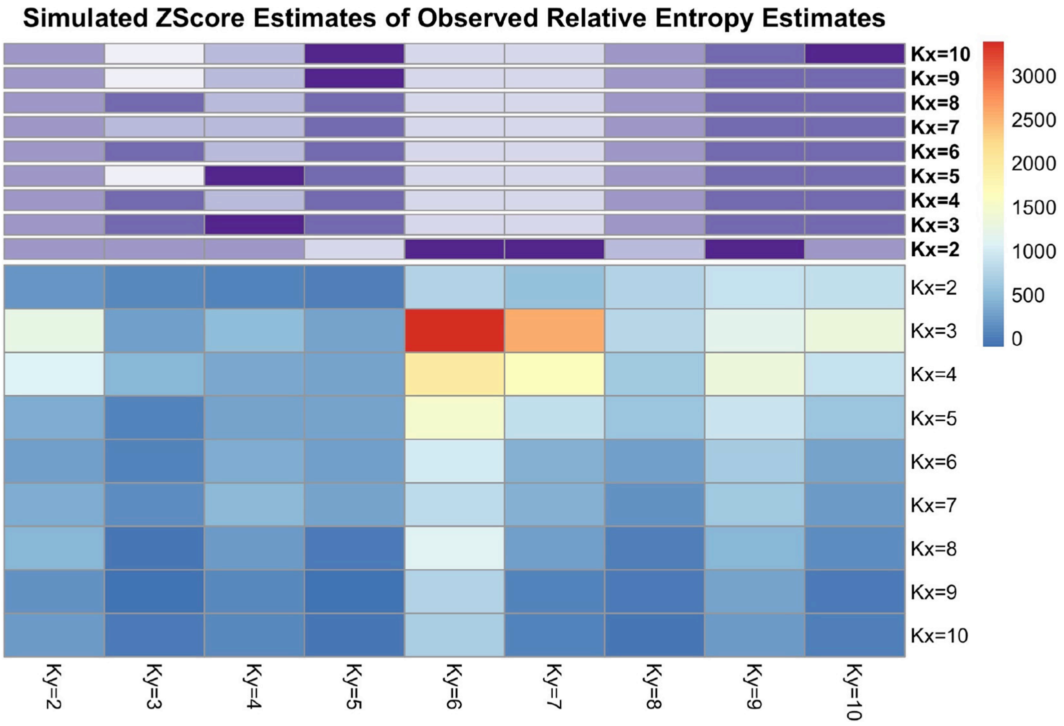
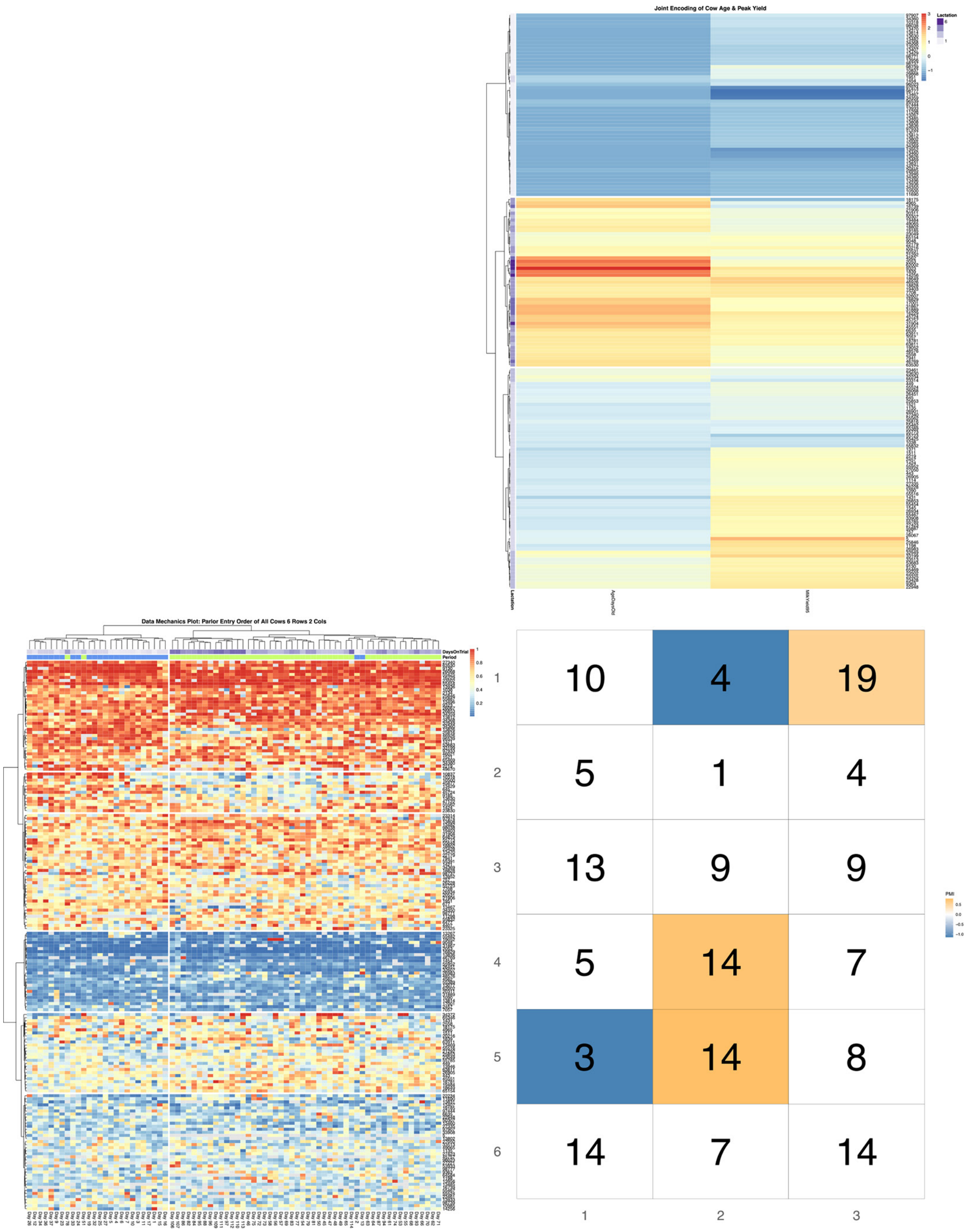


Figure 9. A heatmap visualization of the results of the bivariate tree test. Cell values reflect the difference between the mutual information estimates for the observed data set and those simulated under the null. These results show that a highly significant bivariate association appears with 3 clusters for the age and yield encoding and 6 clusters for the entry order encoding. Annotations along the column margin reflect the number of temporal subperiod clusters using in the entry order encoding that produce the most significant association with age and yield. Kx is the number of clusters for variable x (age), and Ky is the number of clusters for variable y (entry order).

a grid in the data mechanics clustering scheme, and so the purple annotations along the column axis additionally reveals that the strongest bivariate association is consistently found when 2 temporal subperiods are used to encode nonstationarity in the entry positions of cows over this extended observation window. Note that this relationship is stronger than previously reported in preliminary work with a mutual information testing framework wherein age and peak yield were tested against queuing patterns independently without leveraging the redundancy between these 2 variables (McVey et al., 2020).

To characterize this bivariate association, a 2-way contingency table can easily be created using the optimized cluster counts using the *compareEncoding* utility provided in the LIT package (McVey et al., 2021). In this visualization (Figure 10), clusters are numbered for either data set from top to bottom along their corresponding heatmap. Queuing pattern clusters are arranged along the row axis of the contingency table. In this heatmap, individual cows are arranged on the rows and observation days along the column axes, respectively, with each cell in the heatmap representing the relative location of a given cow within the queue on a given day: red toward the back and blue toward the front. The age and yield encoding is arranged along the column axis of the contingency table. Each cell of the contingency table represents the number of cows with a given combination of queue pattern and age yield encoding, and is colored by their pointwise mutual information value only if this observed estimator differs significantly from permutations under the null at the $\alpha = 0.05$ significance level (McVey et al., 2021). Orange cells represent encoding combinations that are significantly overrepresented in this data set when compared with joint marginal probabilities under the null that there is no association, and blue cells are respectively significantly underrepresented.

As shown in Figure 10, within the cluster of cows that consistently entered at the rear of the milking queue (queue cluster no. 1), the oldest cows in this herd (age yield cluster no. 2) are significantly underrepresented, whereas the second and third lactation animals (age yield cluster no. 4) are significantly overrepresented. In contrast, the oldest cows in this herd (age yield cluster no. 2) are significantly overrepresented among cows that consistently entered at the very front of the milking queue (queue cluster no. 4). Several potential behavioral mechanisms could explain this dynamic. Older observational studies describing herd dynamics during movement have previously suggested that ungulate herds are led from both the front and the rear of the herd, so that animals in the front may direct the movement of the herd, while animals at the rear prevent stragglers from falling behind (Reinhardt,



	1	2	3
1	10	4	19
2	5	1	4
3	13	9	9
4	5	14	7
5	3	14	8
6	14	7	14

Figure 10. Visualization of the bivariate association between the encodings for age (AgeDaysOld) and yield (MilkYield95) and for entry order patterns using the optimized encoding granularities. Queuing pattern clusters are arranged on the row axis, and correspond to heatmap clusters numbered from top to bottom. Age and yield encoding are arranged on the column axis, also corresponding to heatmap clusters numbered from top to bottom. The contingency table displays the counts for each joint encoding. Cells are colors by their pointwise mutual information estimate (PMI) if the observed count differs significantly from values permuted under the null at the α = 0.05 significance level, orange being overrepresented and blue being underrepresented.

1983). Previous research with this data set has revealed significant nonlinear associations between these queuing patterns and time budgets that align with this theory, wherein cows that generally enter nearer the rear of the queue were overrepresented among unthrifty animals with low recorded eating times, but that the very rear of the herd was brought up by cows with more balanced time budgets demonstrating moderate eating times (McVey et al., 2021). Given that age and yield are also both correlated with cow size (Schein and Fohrman, 1955), one interpretation of this bivariate pattern might then be that the largest and most experienced animals in this herd are more likely to be found at the front, leading the queue to the parlor, and second- and third-lactation animals with less experience but still sufficient size to occupy higher rungs of the dominance hierarchy may then fill this "caboose cow" role, driving the smaller or less fit animals forward to consolidate the herd (Schein and Fohrman, 1955; Lamb, 1976).

An alternative, if less comprehensive, interpretation of this pattern may be the result of not one but multiple distinct behavioral mechanisms. The oldest cows, by virtue of either their experience or their size, may simply have the greatest capacity to exert their preference (Schein and Fohrman, 1955; Lamb, 1976) to be milked first in jockeying for queue position, whereas second- and third-lactation animals that are more prone to negative energy balance (Berry et al., 2006; Friggens et al., 2007) may be pushed toward the rear of the queue, either as a behavioral mechanism to conserve energy or as a behavioral response to clinical or subclinical health complications. This latter hypothesis may be partially supported by the finding that, when cows with recorded clinical health complications are excluded from these analyses, overrepresentation of older animals at the front of the queue remains a significant finding, but overrepresentation of animals in lactation 2 and 3 at the rear of the queue becomes only marginally significant at the $\alpha = 0.10$ significance level (see the "AgeYield Visualization" directory in Supplemental Materials, https://github.com/cgmcvey/ARPAS2023). Although further research would be needed to decisively determine the causative mechanisms driving this behavioral pattern, it should be noted that this inverted age dynamic would not have been captured with a model-based analysis if a simple linear term were assumed, and this pattern might even have been obscured by the ambiguity in assignment of third-parity animals to a simple categorical term for lactation number.

A bivariate pattern that would be perhaps even more challenging to capture with a model-based approach would be that for queuing pattern cluster no. 5, which is characterized by animals that were generally found nearer the front or middle of the queue when entering the parlor from their home pen but who later fell back nearer the rear of the queue when returning to the parlor from overnight pasture. The oldest cows in this herd (age yield cluster no. 3) were significantly overrepresented in this cluster, whereas the heifers (age yield cluster no. 1) were significantly underrepresented. Though not significant at the $\alpha = 0.05$ significance level, it should also be noted that an inverted age dynamic is also observed for the smaller queue cluster no. 2, which is characterized by animals that entered nearer the rear of the queue when entering the parlor from the home pen but later moved forward in the queue when arriving from the pasture. Although several potential behavioral mechanisms might give rise to this more temporally complex queuing pattern, if we assume that these older cows also prefer being near the front of the queue, as many of their age conspecifics seem to, then evidence that these animals are being pushed backward in the queue by their younger and smaller herdmates may raise some concerns that the fitness and welfare of these animals might have become compromised during the pasture subperiod (O'Connor et al., 2019). Whether pasture access is the cause or simply creates additional opportunities for animals to behaviorally demonstrate latent changes in health state certainly cannot be surmised from this analysis alone, but warrants further study.

CONCLUSIONS AND APPLICATIONS

There is no free lunch in data science, and so no single analytical framework can be expected to best serve the broad spectrum of data sets encountered in animal science (Wolpert and Macready, 1997). Where PLF technologies are employed in 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. Unfortunately, animals are often able to circumvent even the most carefully constructed experimental protocols, and so model-free techniques demonstrated in this article may still add value to model-based analytical pipelines. When implemented in preliminary exploratory data analysis, UML tools can aid in the identification of complex outliers and development of a complete model (McVey et al., 2020). Employed in the visualization of model residuals, UML tools can also be used to visualize latent systematic features of a data set not captured by the fitted model, which can be used to evaluate model sufficiency and potential sources of bias (Enriquez and Hsieh, 2020; McVey et al., 2022).

For PLF data sets collected in uncontrolled farm environments, the preceding simulated and real-world examples have hopefully served to demonstrate the benefits of more open-ended model-free approaches to knowledge discovery. In these case studies we have illustrated how even the leanest UML tools can recover complex and unanticipated behavioral signals in applications in data auditing and consulting, creating additional opportunities for animal scientists to bring their wealth of biological knowledge to bear in analyses of farm records. Through simple modifications to the dissimilarity measures used, UML algorithms can be easily modified to accommodate the wide range of data formats and complex sources of measurement error commonly encountered in working farm environments, such that their application are largely limited only by the creativity of the practitioner (McVey et al., 2021). Although future efforts to extract a wider range of animal phenotypes and welfare metrics from these data streams will likely need to rely to some degree on model-based approaches to facilitate cross-farm comparisons, UML tools can still play a role in helping to develop the most holistic and robust metrics possible from these incredibly rich but still poorly characterized noisy data streams (Ellen et al., 2019; Brito et al., 2020; McVey et al., 2022).

ACKNOWLEDGMENTS

The authors thank Aurora Organic Dairy (Aurora, CO) for facilitating the use of animals and facilities, as well as all the assistance from the farm's personnel, in the creation of the dataset used in the empirical case study with milking order records. The authors also thank Fushing Hsieh and Samuel Loomis (both of the University of California–Davis, Davis, CA) for providing the sensor data used to create the information compression illustration in the introduction of this article. Finally, the authors thank ARPAS for their support in developing and presenting this work at their 2022 Dairy Artificial Intelligence Symposium. Research support was provided to Catherine McVey through the National Science Foundation Graduate Research Fellowship and the Dean's Distinguished Graduate Fellowship from the University of California–Davis College of Agriculture and Environmental Science.

LITERATURE CITED

Adamczyk, K., D. Cywicka, P. Herbut, and E. Trześniowska. 2017. The application of cluster analysis methods in assessment of daily physical activity of dairy cows milked in the Voluntary Milking System. Comput. Electron. Agric. 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 and Sons Inc.

Andersen, P. H., S. Broomé, M. Rashid, J. Lundblad, K. Ask, Z. Li, E. Hernlund, M. Rhodin, and H. Kjellström. 2021. Towards machine recognition of facial expressions of pain in horses. Animals (Basel) 11:1643. https://doi.org/10.3390/ani11061643.

Arac, A., P. Zhao, B. H. Dobkin, S. T. Carmichael, and P. Golshani. 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.

Banhazi, T. M., H. Lehr, J. L. Black, H. Crabtree, P. Schofield, M. Tscharke, and D. Berckmans. 2012. Precision Livestock Farming: An international review of scientific and commercial aspects. Int. J. Agric. Biol. Eng. 5:1–9.

Beggs, D. S., E. C. Jongman, P. H. Hemsworth, and A. D. Fisher. 2018. Short communication: Milking order consistency of dairy cows in large Australian herds. J. Dairy Sci. 101:603–608. https://doi.org/10.3168/jds.2017-12748.

Berry, D. P., and J. McCarthy. 2012. Genetic and non-genetic factors associated with milking order in lactating dairy cows. Appl. Anim. Behav. Sci. 136:15–19. https://doi.org/10.1016/j.applanim.2011.11.012.

Berry, D. P., R. F. Veerkamp, and P. Dillon. 2006. Phenotypic profiles for body weight, body condition score, energy intake, and energy balance across different parities and concentrate feeding levels. Livest. Sci. 104:1–12. https://doi.org/10.1016/j.livsci.2006.02.012.

Brito, L. F., H. R. Oliveira, B. R. McConn, A. P. Schinckel, A. Arrazola, J. N. Marchant-Forde, and J. S. Johnson. 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.

Clouard, C., R. Resmond, A. Prunier, C. Tallet, and E. Merlot. 2022. Exploration of early social behaviors and social styles in relation to individual. Sci. Rep. 12:15. https://doi.org/10.1038/s41598-022-06354-w.

Dawkins, M. S. 2007. Observing Animal Behaviour: Design and Analysis of Quantitative Data. Oxford University Press.

Ellen, E., M. van der Sluis, J. Siegford, O. Guzhva, M. Toscano, J. Bennewitz, L. van der Zande, J. van der Eijk, E. de Haas, T. Norton, D. Piette, J. Tetens, B. de Klerk, B. Visser, and T. Rodenburg. 2019. Review of sensor technologies in animal breeding: Phenotyping behaviors of laying hens to select against feather pecking. Animals (Basel) 9:108. https://doi.org/10.3390/ani9030108.

Enriquez, S., and F. Hsieh. 2020. Categorical exploratory data analysis on goodness-of-fit issues. arXiv 2011:09682. http://arxiv.org/abs/2011.09682.

Franceschini, S., C. Grelet, J. Leblois, N. Gengler, and H. Soyeurt. 2022. Can unsupervised learning methods applied to milk recording big data provide new insights into dairy cow health? J. Dairy Sci. 105:6760–6772. https://doi.org/10.3168/jds.2022-21975.

Friggens, N. C., P. Berg, P. Theilgaard, I. R. Korsgaard, K. L. Ingvartsen, P. Løvendahl, and J. Jensen. 2007. Breed and parity effects on energy balance profiles through lactation: evidence of genetically driven body energy change. J. Dairy Sci. 90:5291–5305. https://doi.org/10.3168/jds.2007-0173.

Fushing, H., S.-Y. Liu, Y.-C. Hsieh, and B. McCowan. 2018. From patterned response dependency to structured covariate dependency: Entropy based categorical-pattern-matching. PLoS One 13:e0198253. https://doi.org/10.1371/journal.pone.0198253.

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

Goodfellow, I., Y. Bengio, and A. Couville. 2016. Deep Learning. The MIT Press.

Guan, J., and F. Hsieh. 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.

Higgins, J. 2004. Introduction to Modern Nonparametric Statistics. Brooks/Cole Cengage Learning.

Hou, S., X. Cheng, L. Shi, and S. Zhang. 2020. Study on individual behavior of dairy cows based on activity data and clustering. Pages 210–216 in Proc. 2020 2nd Int. Conf. Robotics, Intel. Control Artif. Intel. Association for Computing Machinery. https://doi.org/10.1145/3438872.3439082.

Hubbard, A. J., M. J. Foster, and C. L. Daigle. 2021. Social dominance in beef cattle—A scoping review. Appl. Anim. Behav. Sci. 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. J. Theor. Biol. 241:639–648. https://doi.org/10.1016/j.jtbi.2006.01.001.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. An Introduction to Statistical Learning. Springer Texts in Statistics. Springer New York. https://doi.org/10.1007/978-1-4614-7138-7.

Kilgour, R., and T. H. Scott. 1959. Leadership in a herd of dairy cows. Proc. N.Z. Soc. Anim. Prod. 19:36–43.

Kirby, M. 2001. Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns. John Wiley and Sons Inc.

Lamb, R. C. 1976. Relationship between cow behavior patterns and management systems to reduce stress. J. Dairy Sci. 59:1630–1636. 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.

Manriquez, D., L. Chen, G. Albornoz, J. Velez, and P. Pinedo. 2018. Case study: Assessment of human-conditioned sorting behavior in dairy cows in farm research trials. Prof. Anim. Sci. 34:664–670. https://doi.org/10.15232/pas.2018-01749.

Manriquez, D., L. Chen, P. Melendez, and P. Pinedo. 2019. The effect of an organic rumen-protected fat supplement on performance, metabolic status, and health of dairy cows. BMC Vet. Res. 15:450. https://doi.org/10.1186/s12917-019-2199-8.

Mathis, M. W., and A. Mathis. 2020. Deep learning tools for the measurement of animal behavior in neuroscience. Curr. Opin. Neurobiol. 60:1–11. https://doi.org/10.1016/j.conb.2019.10.008.

McVey, C., D. Egger, and P. Pinedo. 2022. Improving the reliability of scale-free image morphometrics in applications with minimally restrained livestock using projective geometry and unsupervised machine learning. Sensors (Basel) 22:8347. https://doi.org/10.3390/s22218347.

McVey, C., F. Hsieh, D. Manriquez, P. Pinedo, and K. Horback. 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.

McVey, C., F. Hsieh, D. Manriquez, P. Pinedo, and K. Horback. 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 (Basel) 22:1. https://doi.org/10.3390/s22010001.

O'Connor, A. H., E. A. M. Bokkers, I. J. M. de Boer, H. Hogeveen, R. Sayers, N. Byrne, E. Ruelle, and L. Shalloo. 2019. Associating cow characteristics with mobility scores in pasture-based dairy cows. J. Dairy Sci. 102:8332–8342. https://doi.org/10.3168/jds.2018-15719.

Pinheiro, J. C., and D. M. Bates. 2000. Mixed-Effects Models in S and S-PLUS, Statistics and Computing. Springer-Verlag.

Polikarpus, A., T. Kaart, H. Mootse, G. De Rosa, and D. Arney. 2015. Influences of various factors on cows' entrance order into the milking parlour. Appl. Anim. Behav. Sci. 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. Appl. Anim. Ethol. 8:45–52. https://doi.org/10.1016/0304-3762(82)90131-6.

Reinhardt, V. 1983. Movement orders and leadership in a semiwild cattle herd. Behaviour 83:251–264. https://doi.org/10.1163/156853983X00183.

Rufener, C., J. Berezowski, F. Maximiano Sousa, Y. Abreu, L. Asher, and M. J. Toscano. 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., M. Špinka, I. Stěhulová, F. Ceacero, M. Šimečková, and R. Kotrba. 2013. Pay respect to the elders: age, more than body mass, determines dominance in female beef cattle. Anim. Behav. 86:1315–1323. https://doi.org/10.1016/j.anbehav.2013.10.002.

Schein, M. W., and M. H. Fohrman. 1955. Social dominance relationships in a herd of dairy cattle. Br. J. Anim. Behav. 3:45–55. https://doi.org/10.1016/S0950-5601(55)80012-3.

Shannon, C. E. 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27:379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Sibanda, T. Z., M. Welch, D. Schneider, M. Kolakshyapati, and I. Ruhnke. 2020. Characterising free-range layer flocks using unsupervised cluster analysis. Animals (Basel) 10:855. https://doi.org/10.3390/ani10050855.

Soffié, M., G. Thinès, and G. De Marneffe. 1976. Relation between milking order and dominance value in a group of dairy cows. Appl. Anim. Ethol. 2:271–276. https://doi.org/10.1016/0304-3762(76)90060-2.

Stygar, A. H., Y. Gómez, G. V. Berteselli, E. Dalla Costa, E. Canali, J. K. Niemi, P. Llonch, and M. Pastell. 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., M. B. Jensen, A. M. de Passillé, L. Hänninen, and J. Rushen. 2021. Invited review: Lying time and the welfare of dairy cows. J. Dairy Sci. 104:20–46. https://doi.org/10.3168/jds.2019-18074.

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

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

Wolpert, D. H., and W. G. Macready. 1997. No free lunch theorems for optimization. IEEE Trans. Evol. Comput. 1:67–82. https://doi.org/10.1109/4235.585893.

Xu, H., S. Li, C. Lee, W. Ni, D. Abbott, M. Johnson, J. M. Lea, J. Yuan, and D. L. M. Campbell. 2020. Analysis of cattle social transitional behaviour: Attraction and repulsion. Sensors (Basel) 20:5340. https://doi.org/10.3390/s20185340.

ORCIDS

Catherine McVey https://orcid.org/0000-0001-5166-3371
Pablo Pinedo https://orcid.org/0000-0001-7111-3377