

# THE UNIVERSITY OF WISCONSIN DAIRY BRAIN: THE FUTURE OF DAIRY MANAGEMENT DECISIONS BASED ON BIG DATA ANALYTICS

Victor E. Cabrera, Michael Ferris, and Heather White  
University of Wisconsin-Madison  
vcabrera@wisc.edu

## TAKE HOME MESSAGES

- Continuous analyses of integrated dairy farm data streams can provide novel insights for improved decision-making
- The University of Wisconsin Dairy Brain project addresses dairy farm data integration and their value-added applications as a continuous data-driven decision-making engine
- The Dairy Brain's Agricultural Data Hub (**AgDH**) is designed to ingest dairy data from multiple sources, establish their relationships, connect them, and make the data consistent and available from a single source
- The Dairy Brain's analytical modules encompass near real-time algorithms capable of performing longitudinal historical analyses, forecasting future performance, and prescribing the best course of action in a continuous loop through decision support tools connected to integrated dairy farm data
- We demonstrate the Dairy Brain's value proposition with two practical prescriptive applications. One addresses herd's nutritional accuracy improvement. The other deals with maximizing genetic progress and profitability through reproductive management.

## INTRODUCTION

Dairy farmers have embraced technologies that generate data streams from the dairy farming precision agriculture revolution (Bewley and Russell, 2010) that are difficult to manage because the resulting data ecosystem is large and complex (Ferris et al., 2020). When integrated, these data streams can uncover new insights and therefore improve decision-making and farm management (Liberati and Zappavigna, 2009). Permanent analyses of integrated data pipelines can help identify feeding and health problems (Arcidiacono et al., 2017), optimize reproduction performance (LeRoy et al., 2018), and improve the sustainability of the overall system (Wathes et al., 2008).

Near real-time integration of milking, feeding, reproduction, behavior, and many other dairy farm record-keeping systems, however, is challenging (van der Weerd and Boer, 2015). The lack of data integration and its subsequent interpretation and value-added utilization delays optimal actions, increases mistakes, and hides improvement opportunities, which ultimately decreases profitability and threatens the sustainability of modern dairy farm systems (Cabrera et al., 2020). Integrated big data

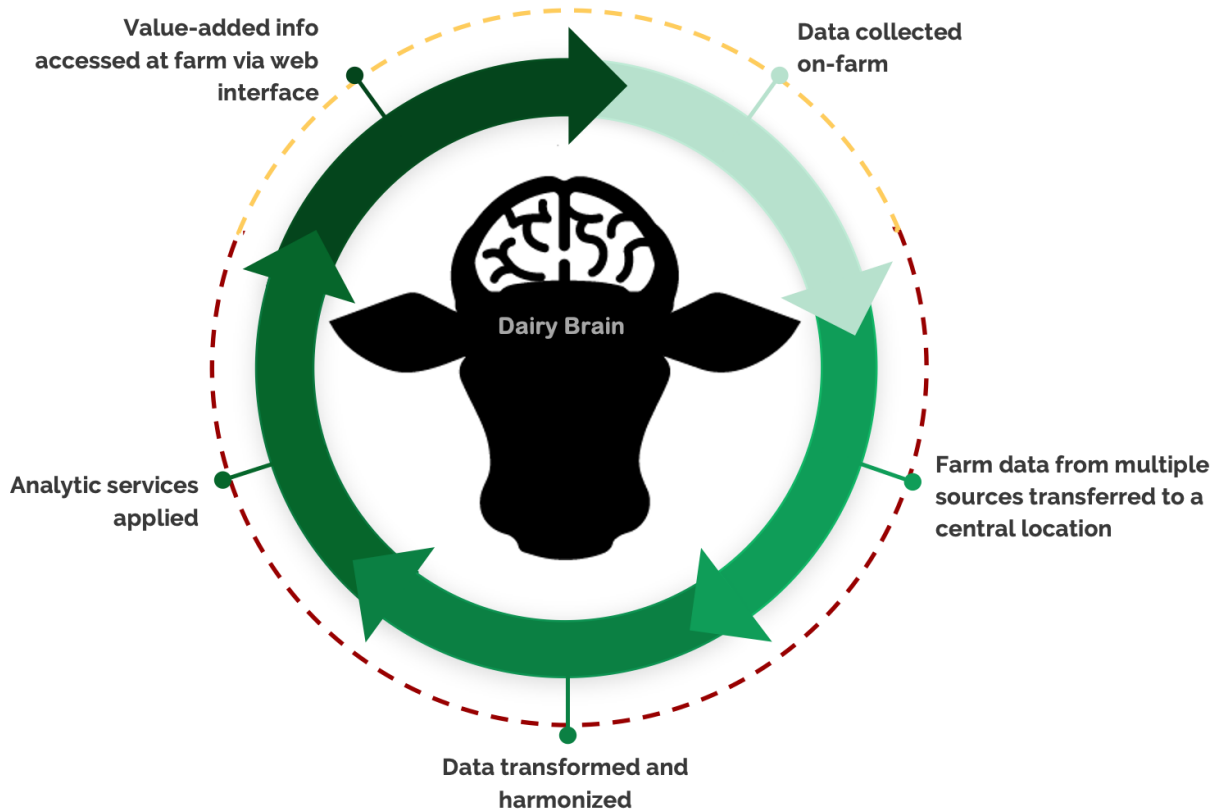
analytics (such as data mining and machine learning techniques (Morota et al., 2018)) are necessary to ameliorate the combined biological, economic, and environmental uncertainties inherent to the production system (Wolfert et al., 2017), but they require management of the quality, heterogeneity, and transformations of the data (Hashem et al., 2015).

Any attempt to develop an interconnected system of data and services requires considerations and standards: dairy farm data tends to be “dirty” and data cleaning is not a priority because of a lack of a clearly defined value proposition (Ferris et al., 2020). Data must be acquired and edited in a reliable fashion, passed to the analytical suite with a consistent structure, and then used for descriptive, predictive, and (or) prescriptive analytics (Ferris et al., 2020). Decisions could be strategic such as breeding or genetic progress protocols; tactical such as diet formulation or vaccination schedules; or operational such as health or hormonal treatments. Depending on the nature of the problem, these decisions are needed at different specified timeframes (e.g., hourly, daily, weekly, monthly, or yearly).

## BACKGROUND

The University of Wisconsin Dairy Brain project (<https://DairyBrain.wisc.edu>) engages a transdisciplinary team of more than 15 dairy, data, and computer scientists working in close

collaboration with progressive dairymen, extension educators, and allied industry tackling the issue of dairy farm data integration and their value-added applications (Figure 1).



**Figure 1.** Core concept of the University of Wisconsin-Madison Dairy Brain, a continuous data-driven decision-making engine.

As such, the UW Dairy Brain encompasses four main objectives: (1) Nurture a Coordinated Innovation Network (CIN) to shape the data service development. (2) Create an Agricultural Data Hub (AgDH) to gather, integrate, and disseminate multiple data streams relevant to dairy operations. (3) Build the Dairy Brain as a suite of analytical modules that leverages the aggregation service to provide dairy management insights as an exemplar of connected ecosystems services. (4) Design and execute an innovative extension program. This paper discusses the development of the AgDH and the Dairy Brain with a vision of the future for dairy management decisions based on permanent data integrated for

predictive and prescriptive data analytics and includes the description of two analytical applications within the Dairy Brain domain.

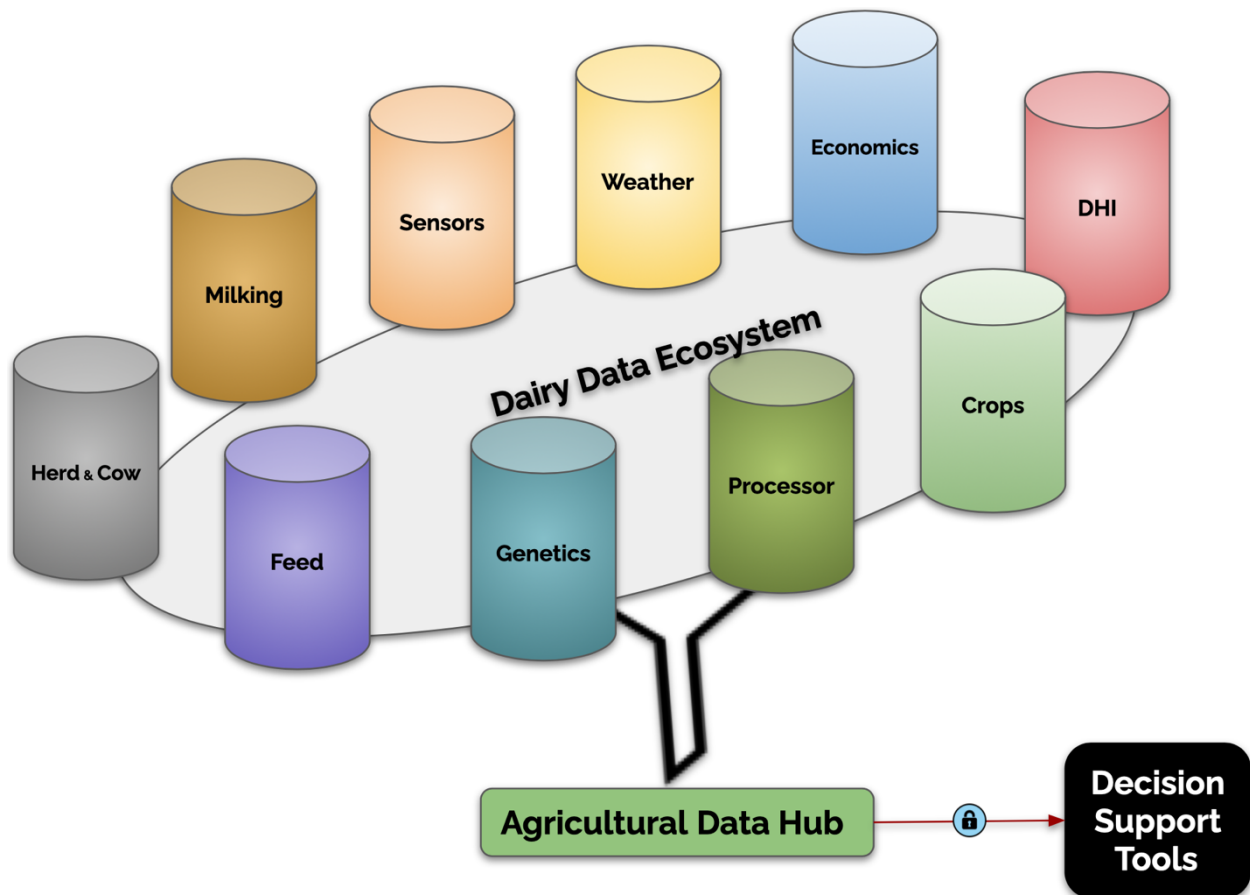
## THE AGRICULTURAL DATA HUB (AgDH)

Although the dairy farm data we are collecting at the moment is not “big data” in the scientific sense, it is handled with the principles and methodologies grounded in the big data literature (Wolfert et al., 2017; Ferris et al., 2020) because it will eventually become big data when more data streams, longer periods of time, and (or) more farms enroll in the project. The AgDH is designed as a system to ingest dairy data from

multiple sources to establish the relationships among these sources and therefore make the data consistent and available from a single location, an aggregator entity to gather and disseminate multiple harmonized data streams to analytical tools (Ferris et al., 2020). It entails the transformation and joining of data streams as they are being generated and provides functional access to the resulting data. The AgDH collects, harmonizes, and stores information from siloed data streams on dairy farms and provides permissioned programmatic access to dairy-specific data (Figure 2).

### ***Data Processing Flow***

Data pertaining to dairy farms, collected on- or off-farm, are stored in raw heterogeneous formats, which we think are better processed through an extraction, transformation, and loading (ETL) process to prepare the data to be accessed via a programmatic interface (Ferris et al., 2020). Within this process it is important to determine a protocol on how to treat missing, corrupt, or inaccurate (“dirty”) data, which is commonly found in dairy data settings. Dirty data will be primarily handled by advanced statistical techniques that consider conditions under which missing data occurred, such as multiple-imputation, full information maximum likelihood, expectation maximization, and matrix completion (Dong and Pen, 2013).



**Figure 2.** Siloed dairy data streams aggregated and disseminated by the Agricultural Data Hub (AgDH).

In order to make the data consistent, independent of a particular vendor or data collection system, we are creating a universal relational framework with common keys to retrieve data efficiently. This requires entity matching to identify the same individual or object and matching within different data sources (Ferris et al., 2020). Although this might seem trivial with clearly organized data and only a few sources, it becomes a challenging task when disorganized data sources grow. We anticipate the need of machine learning approaches to merge dairy data streams (Mudgal et al., 2018). In addition, dairy data that are recorded as an event (e.g., disease onset, breeding, vaccination) must be handled within a taxonomy for events (Shuetz et al., 2018). We favor a centralized database storage as a data warehouse structure because some of the data can be characterized as ephemeral (stored in the original source for only a short period of time), highly heterogeneous, and likely unstructured (Ferris et al., 2020).

Currently, the data collected from all data streams of the five cooperating farms is permanently stored in the University of Wisconsin-Madison Center for High Throughput Computing server (<http://chtc.cs.wisc.edu/>; Liang et al., 2018). Data from the farms are pulled into files in specific directories of the server, whose structure is determined by the data source (farm, data stream, date). When newly arriving files are sensed, our campus server identifies the data type and applies a predetermined processing pathway to extract the data to a universal database using data-parsing logic specific to software vendor. Data are then transformed by a process of cleaning that involves data mining techniques and language processing scripts to detect and correct or remove corrupt or inaccurate data points. Cleaned data are then harmonized using a relational database with common key variables (entity matching) and with consistent format and structure, regardless of the source. Harmonized data are loaded into the AgDH, which serves as our data warehouse and the main bridge among the farms, research activities, and development of decision support tools. Subsequent on-demand data retrieval is provided as needed (Shuetz et al., 2018).

All programmatic code along with its documentation are safely versioned in our project GitHub repository (<https://github.com/orgs/DairyBrain/>). We favor the Python® programming environment (Rossum and Drake, 2009) as the open source, object-oriented language for both database management and analytic modules development because it supports a large and increasing number of libraries, modules, and packages that promote modularity and code reuse.

### ***Applications Programming Interface***

Our system will link services via application programming interfaces (APIs), which are a modern and widely accepted approach for providing programmatic access to data and services. They make data available via web requests using a representational state transfer (REST) architecture and the well-established JavaScript Object Notation (JSON) format for data transfer. APIs work as a query operation by calling data or services through a standard http request that includes a set of input data. Output data normally comes in JSON format, comma separated values (CSV), or other (eXtensible Markup Language (XML) style. Its benefits include a clear definition of the expected data types retrieved and a highly flexible system that allows for controlled access to different data sources and services for different uses. As such, it is a modular system where analytical tools can be developed independently of each other and the data source. Therefore, it supports parallel development of data and analytical services (Ferris et al., 2020). Our APIs allow fine control of data and services access via a secure encrypted https connection in which users need to be approved and authenticated by the AgDH service. The vision is to make both the AgDH and the Dairy Brain analytical services accessible via APIs, which, however, would remain independent of each other. This setup will allow use of analytical services (Dairy Brain) with external data and/or use data services (AgDH) with external analytical services, if so desired.

### **THE DAIRY BRAIN**

The Dairy Brain encompasses development of analytical modules and decision support tools with constant flow of integrated data. Our vision

is near real-time analytical engines capable of performing longitudinal historical analyses as well as forecasting future performance informed by evolving past performance in a continuous loop (Cabrera et al., 2020). It will provide descriptive, predictive, and prescriptive analytics (Ferris et al., 2020). Descriptive analytics refers to key performance indicators or KPIs such as milk income over feed cost, feed efficiency, disease incidence, or pregnancy rate. Predictive analytics involve forecasts and projections such as the required replacements needed to keep the herd size stable, advance feed purchases for the season, cropland required to apply all the manure during the growing season, or likelihood of mastitis disease on a cow. Prescriptive analytics go one step further evaluating the tradeoffs of alternative course of actions and suggesting the best pathway to reach a goal within farm constraints. Examples include determining the herd demographics that would yield the maximum profit (Ferris et al., 2020), proposing the best breeding, reproduction, and culling protocols to reach the greatest genetic progress, or recommending the best grouping strategy to maximize nutritional feed efficiency and net return (Barrientos-Blanco et al., 2020). Data-informed decisions should follow farmer needs and be in tune with their short-, medium-, and long-term goals. Although some operational or short-term decisions informed by milk income minus feed cost (\$/cow per day) or feed efficiency measured as milk produced over feed costs (lb milk/lb feed per day) require rather trivial calculations, these require (even demand) integrated data streams and a detailed ontology (Ferris et al., 2020). This becomes even more critical when decisions are envisioned for mid- or long-term dimensions applying predictive or prescriptive algorithms for strategic planning, that likely entail additional data streams, greater integration, and more advanced analytics.

There are numerous current and upcoming opportunities for integrated data-driven decision making in diverse dairy farm management areas such as early disease detection (Weigel et al., 2017; Fadul-Pacheco et al., 2019; Delgado et al., 2019); cow-life profitability projections (Cabrera, 2010; 2012); reproductive performance (Giordano et al., 2012); environmental

performance (Liang and Cabrera, 2015); nutritional grouping (Kalantari et al., 2016; Wu et al., 2019; Barrientos-Blanco et al., 2020), among others (Cabrera, 2018). As specific examples, two practical prescriptive applications within the Dairy Brain concept are discussed hereafter.

### ***(1) Improving Nutritional Accuracy***

**Motivation.** Nutritional grouping and providing accurate diets to cows is an effective strategy to control cost, increase revenue, and enhance herd feed efficiency in dairy farms (Kalantari et al., 2016). The underlying argument for nutritional grouping is the fact that arranging animals according to their density (unit/kg diet dry matter) of nutrient requirements (e.g., energy, protein, etc.) results in groups of cows that are more homogeneous in their diet requirements. This will result in less underfed or overfed animals and therefore cost savings and increased productivity (Cabrera and Kalantari, 2016). Underfed animals become underconditioned leading to decreased production and reproductive efficiency (Roche et al., 2013), whereas overfed cows become overconditioned with increased metabolic problems when freshening (Roche et al., 2009). Using only one or a few diets for the whole lactating herd cannot represent the nutritional needs of all the cows in the group, prompting farmers and nutritionists to provide diets with nutrient densities tailored to the needs of the top cow producers in a group and therefore overfeeding a great majority of the cows in the group (Cabrera and Kalantari, 2016). Costs of feed are consequently increased, and metabolic issues are exacerbated without improving productivity. Enhanced nutritional accuracy in a group results in feeding cows closer to their requirements (Cerosaletti and Dewing, 2008), which will increase milk productivity and decrease nutrient excretion (Kalantari et al., 2016). Although farmers group animals in pens following diverse criteria (e.g., reproductive or lactation status), most of them rely on too few diets (Contreras-Govea et al., 2015) or diets not formulated according to requirements (Cabrera and Kalantari, 2016) because they find barriers to provide more diets (Contreras-Govea et al., 2015) and/or they do not have expert systems in place to adjust diets dynamically to the ever changing nutritional requirements of the group (Contreras-

Govea et al., 2015). The on-farm data streams, computer and feeding systems, and grouping protocols available today, coupled with data integration initiatives such as the Dairy Brain project provide an opportunity to facilitate effective application of nutritional grouping and accurate diet formulation on dairy farms (Barrientos-Blanco et al., 2020).

**Background.** Currently, a large dairy farm (i.e., 2,400 lactating cows), engaged with the Dairy Brain project, allocates cows into 14 pens according to parity (primiparous and multiparous) and stage of lactation (fresh, early, peak, and late lactation). For these groups, they formulate and provide 9 diets. In general, diets are for early, mid, and late lactation cows with some differences between primiparous and Multiparous cows. These diets are reformulated 3 or 4 times a year according to feed prices and not according to cow groups' nutrient requirements. In addition, individual cow nutritional requirements are not used as a criterion to allocate cows to pens. Every Tuesday (once a week), cows are moved between pens if their status has changed (e.g., early to mid-lactation), and/or if it is required by the natural flow of animals through pens to avoid over or under crowded pens. The decision to move cows is made by the manager using expertise or experience and a printed list of animals. The farm system currently in place takes time, lots of personnel effort, is prone to errors, and is inconsistent. Above all, it does not take into consideration nutritional requirements as a criterion to allocate cows to pens, and diets are not based on the nutritional requirements of the cows in each pen.

**Data integration.** A prescriptive model was developed to propose a better allocation of cows to pens and to suggest better diets to provide in each pen, which requires continuous data inflow and data integration. At least two main data stream sources need to be connected permanently: the herd management software and the feed management software. In the case study farm, these were Dairy Comp 305 and Feed Comp. Although both are from the same vendor, Valley Ag Software (Tulare, CA), they are independent of each other and are not integrated. Connection of the two data sources is via two

common merge reference variables: pen identification number (Pen ID, No) and observation date (Fdate, yyyy-mm-dd). Cow level data required (from Dairy Comp 305) are Cow ID, days in milk, parity, body weight (BW, kg at date available), milk yield (kg/day), milk protein (%), and milk fat (%). Pen data required (from Feed Comp) are dry matter intake (average kg/cow per day) and diet provided (ingredients, quantity, composition, and price). This data integration system is prepared for automation with minimal or no supervision. The farm uses the nutritional dynamic system software for diet formulation, which is based on the Cornell Net Carbohydrate and Protein System (Van Amburgh et al., 2015). The same software, feed library, and prices are used for diet re-formulation within the proposed prescriptive framework.

**Nutritional requirements.** Cow-level, daily dry matter intake (kg DM), energy,  $NE_L$  (Mcal), and grams of metabolizable protein (**MP**) are calculated based on NRC (2001) as a function of week of lactation, energy corrected milk, and BW. Body weight at the farm is available only once per lactation. This value is used internally to predict daily BW as a function of average herd BW, parity, and stage of lactation (Kalantari et al., 2016).  $NE_L$  and MP are then expressed as a density by dividing them by the dry matter intake, Mcal and g/kg DM, respectively.

**Pen allocation.** Respecting as much as possible the farm grouping protocols and procedures (e.g., cows are allocated to pens primarily by parity and stage of lactation), cows are clustered to pens according to their nutrient density requirements ( $NE_L$  and MP) in an attempt to have greatest homogeneity within groups (closest nutritional requirements) while having greatest heterogeneity between groups. The number of cows per pen is constrained to maximum number of animals allowed in each pen.

**Diet formulation and comparison with current farm practice.** Once cows are allocated weekly in pens, the diets are reformulated according to the pen group requirements following the classical 83<sup>rd</sup> percentile cow on the predicted MP and  $NE_L$  pen requirements (McGilliard et al., 1983). These diets should change every week but being

consistent with current farm practice and for comparison purposes, average diet for an analysis period of 9 weeks was kept constant in each one of the pens. The proposed diet for each pen was the average of the 9-week period. At the end, the proposed framework was different than the current farm framework in (1) cows allocated to each pen and (2) diet provided to each pen.

**Outcomes.** Savings on feed across all pens for all lactating cows adds a value of \$31/cow per year when applying proposed framework compared with current farm practice (Table 1). This is explained by the fact that cows allocated to groups according to nutritional requirements

determine greater nutritional accuracy (i.e., provided diets are closer to the cows' requirement; Table 1). This economic value did not include potential increase in milk productivity, which would likely have an even greater economic benefit (St-Pierre and Thraen, 1999; Kalantari et al., 2016; Wu et al., 2019). This value would also be much greater if diets would change weekly and if cows would be allocated to pens only based on their nutritional requirements. Other benefits not considered, but expected, would be improved herd health and decreased environmental footprint of the dairy operation (Cabrera and Kalantari, 2016).

**Table 1.** Diet cost savings and improved diet accuracy of proposed Dairy Brain prescriptive nutritional accuracy practice

Item	Unit	Farm current practice	Dairy Brain proposed practice	Difference in favor of proposed practice
Diet cost	\$/cow/day	8.91	8.82	-0.09
Diet accuracy index <sup>1</sup> MP	g/kg of DM	0.1876	0.1803	-0.0073
Diet accuracy index <sup>1</sup> NE <sub>L</sub>	Mcal/kg DM	0.2171	0.2134	-0.0037

<sup>1</sup>Diet accuracy index = aggregated difference between the density of diet nutrient content and each cow's nutrient requirement: Better when lower. MP = metabolizable protein. NE<sub>L</sub> = net energy for lactation. DM = dry matter.

In addition, not discussed here, the proposed grouping strategy based on cows' nutritional requirements determines much better and consistent allocation of cows to pens across time with less undercrowded or overcrowded pens and much lower risks of misclassifications of cows to pens (Barrientos-Blanco et al., 2020). This exemplar proves the value added of a practical and feasible application when data streams are integrated and continuously analyzed within a prescriptive domain using the underline concepts of the Dairy Brain.

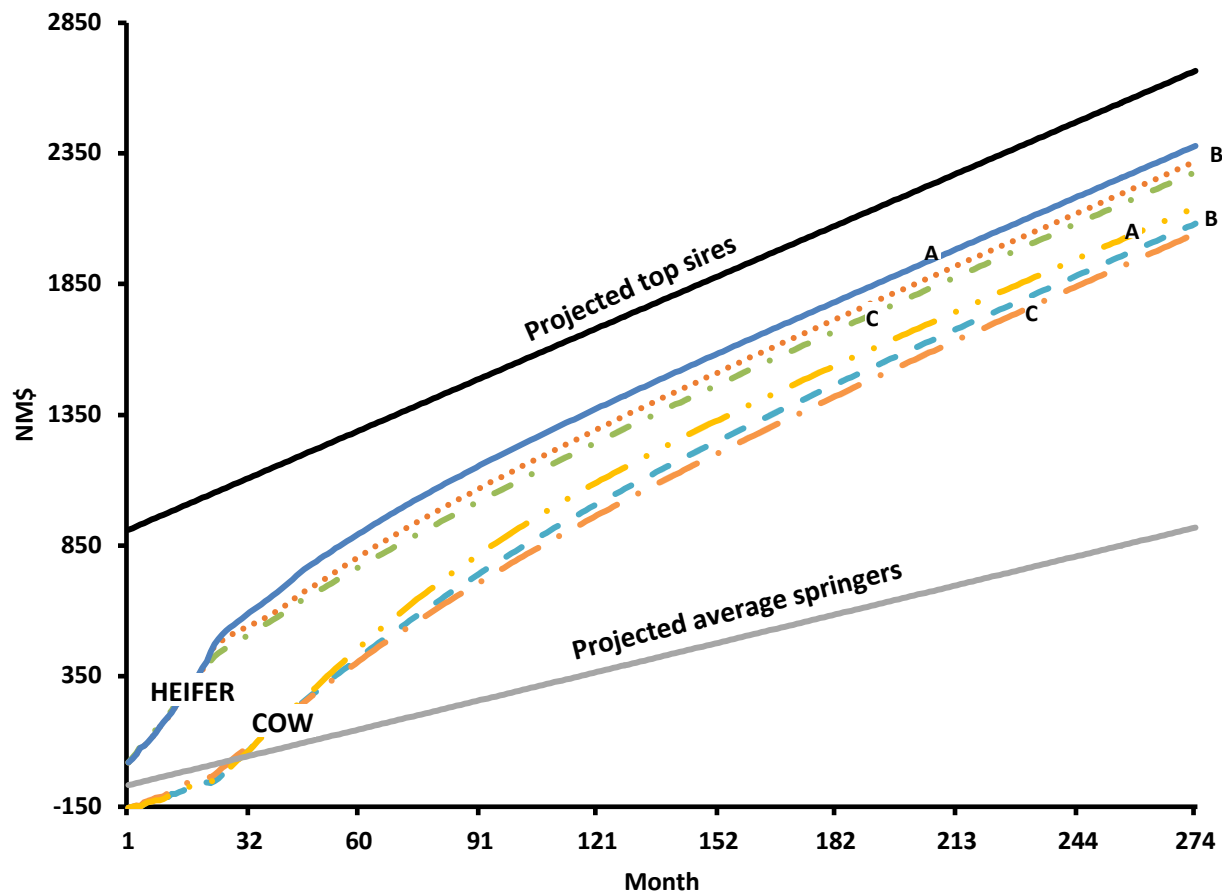
## **(2) Application for Evaluating Reproductive Performance, Genetic Progress, and Culling Protocols**

**Motivation.** Improved reproductive performance causing oversupply of replacements together with increased demand of beef crossbred (dairy x beef) calves prompts the use of sexed and beef semen, which now account for about 40% (~20% sexed

semen + ~20% beef semen) of the dairy breedings in Wisconsin (AgSource, Dairy, Wisconsin, Madison, WI). A logical strategy of breeding superior animals to sexed semen, while using beef semen on inferior animals, increases genetic progress because of intense selection and avoidance of inferior replacements (Ettema et al., 2017) in addition to greatly reducing generation interval. The specific decision is challenging, however, because it needs to be combined with possible reproductive protocols, genetic information, genomic testing, age and status of the cow, culling protocols, among other factors. At the moment, for example, it is unknown which general option would be best for a particular farm: (1) use conventional semen, produce more than required replacement calves, and select to keep only the best required replacements via a genomic test; (2) use a combination of sexed and beef semen; or (3) use sexed, beef, and conventional semen and combine those with

some level of a genomic test to select best calves to keep. More specifically, it is currently completely beyond farmers to determine which animals to enroll in which reproduction protocol, which animals to breed with which semen type, which animals are genomically tested with which test, and/or which animals select out of the herd and when, even though they might have all the information required to answer those questions. Our envisioned prescriptive model would require heavily integrated data from at least genetics/genomics, herd management, feed management, reproductive events, herd health, and culling policies.

**The Vision.** We plan to develop an interactive, dynamic, and highly integrated decision support tool to optimize decisions of breeding, genetic selection, and culling policies for maximum profitability and sustainability. Stochastic lifecycle of individual animals will be simulated according to herd traits. Then, different management strategies will be imposed and compared. This would provide a predictive structure for anticipating long-term results to interventions. An example of possible outcomes related to genetic progress is shown in Figure 3.



**Figure 3.** Lifetime net merit (NM\$) progress for heifers and cows for different breeding protocols starting in May 2020 (month 1) for different management strategies: (A) using sexed semen and beef semen and culling any extra female calves; (B) using conventional semen and culling any extra female calves; and (C) using conventional semen and culling any extra springers. (Li and Cabrera, Unpublished data).

We are interested, however, in providing a prescriptive framework to aid in the decision-making. This would entail machine learning techniques and optimization to suggest the best course of action and adjust it as performance outcomes are analyzed. An example of prescriptive recommendation would be to use sexed semen in heifers for the first three inseminations; in first lactation cows at first breeding, and all other top 20% genetics adult cows. Then, use conventional semen in the remaining heifers and first lactation cows, and beef semen in the remaining breeding-eligible cows.

This breeding protocol would result in approximately 20% more than the required number of replacements produced on farm. Therefore, the need to select this proportion of calves from the young stock herd right after birth. The next recommendation could be to rank all newborn female calves by their parents' genetic average and apply a genomic test on the middle 30% of animals, which will be used to re-rank calves and select the bottom 20% that will be sold. The integrated algorithm would assure this multi-step decision-making protocol will guarantee maximum profitability and best genetic progress at one specific time but always look towards the long-term horizon. The model and decision support tool would be connected to the farm data not only to use historical performance and to project possible future performance, but also to adjust, calibrate, and "learn" from the outcomes on the farm in a seamless connected ecosystem system of data and applications.

#### ACKNOWLEDGMENTS

This project has been supported by the USDA-National Institute for Food and Agriculture (Washington, DC; FACT grant 2018-09256).

#### REFERENCES

Arcidiacono, C., S. M. C. Porto, M. Mancino, and G. Cascone. 2017. Development of a threshold-based classifier for real-time recognition of cow feeding and standing behavioural activities from accelerometer data. *Comput. Electron. Agric.* 134:124–134.

- Barrientos-Blanco, J. A., H. White, R. D. Shaver, and V. E. Cabrera. 2020. Improving nutritional accuracy and economics through multiple ration-grouping strategy. *J. Dairy Sci.* 103:3774-3785.
- Bewley, J. M., and R. A. Russell. 2010. Reasons for slow adoption rates of precision dairy farming technologies: Evidence from a producer survey. Pages 30–31 in *Proc. First North Am. Conf. Precision Dairy Management*, Toronto, Canada. Accessed Oct. 1, 2019. <http://www.precisiondairy.com/proceedings/s1bewley2.pdf>.
- Cabrera, V. E. 2010. A large Markovian linear program for replacement policies to optimize dairy herd net income for diets and nitrogen excretion. *J. Dairy Sci.* 93:394-406.
- Cabrera, V. E. 2012. A simple formulation and solution to the replacement problem: A practical tool to assess the economic cow value, the value of a new pregnancy, and the cost of a pregnancy loss. *J. Dairy Sci.* 95:4683-4698.
- Cabrera, V. E. 2018. Helping dairy farmers to improve economic performance utilizing data-driving decision support tools. *Animal* 12:134-144.
- Cabrera, V. E., J. A. Barrientos, H. Delgado, and L. Fadul-Pacheco. 2020. Real-time continuous decision making using big data on dairy farms. *J. Dairy Sci.* 103:3856–3866.
- Delgado, H., L. Fadul-Pacheco and V. E. Cabrera. 2019. The use of integrated data to identify first-lactation cows at high risk of clinical mastitis. *J. Dairy Sci.* 102: (Suppl.1):M134.
- Ettema, J.F., J.R. Thomasen, L. Hjortø, M. Kargo, S. Østergaard, and A. C. Sørensen. 2017. Economic opportunities for using sexed semen and semen of beef bulls in dairy herds. *J. Dairy Sci.* 100:4161–4171.
- Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction to Quantitative Genetics*. 4th ed. Longman Scientific and Technical, New York, NY.
- Fadul-Pacheco, L., H. Delgado and V. E. Cabrera. 2019. Machine learning algorithms for early prediction of clinical mastitis. *J. Dairy Sci.* 102:(Suppl.1):94.

- Ferris, C.M., A. Christensen, S.R. Wangen. 2020. Symposium review: Dairy Brain—Informing decisions on dairy farms using data analytics. *J. Dairy Sci.* 103:3874–3881.
- Giordano, J. O., A. Kalantari, P. M. Fricke, M. C. Wiltbank, and V. E. Cabrera. 2012. A daily herd Markov-chain model to study the reproductive and economic impact of reproductive programs combining timed artificial insemination and estrous detection. *J. Dairy Sci.* 95:5442–5460.
- Hashem, I. A. T., I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. Ullah Khan. 2015. The rise of “big data” on cloud computing: Review and open research issues. *Inf. Syst.* 47:98–115.
- Kalantari A. S., L. E. Armentano, R. D. Shaver, and V. E. Cabrera. 2016. Economic impact of nutritional grouping in dairy herds. *J. Dairy Sci.* 99:1672–1692.
- LeRoy, C. N. S., J. S. Walton, and S. J. LeBlanc. 2018. Estrous detection intensity and accuracy and optimal timing of insemination with automated activity monitors for dairy cows. *J. Dairy Sci.* 101:1638–1641.
- Liberati, P., and P. Zappavigna. 2009. Improving the automated monitoring of dairy cows by integrating various data acquisition systems. *Comput. Electron. Agric.* 68:62–67.
- Liang, D., and V. E. Cabrera. 2015. Optimizing productivity, herd structure, environmental performance, and profitability of dairy cattle herds. *J. Dairy Sci.* 98:2812–2823.
- Liang, D., H. Delgado, and V. E. Cabrera. 2018. A virtual dairy farm brain. 13th European International Farming System Association Symposium of the Farming and Rural Systems: Farming systems: facing uncertainties and enhancing opportunities. Chania, Crete, Greece, 01-05 July 2018.
- McGilliard, M. L., J. M. Swisher, and R. E. James. 1983. Grouping lactating cows by nutritional-requirements for feeding. *J. Dairy Sci.* 66:1084–1093.
- Madrigal, A. C. 2012. The perfect milk machine: How big data transformed the dairy industry. *The Atlantic*. Accessed Nov. 8, 2019. <https://www.theatlantic.com/technology/archive/2012/05/the-perfect-milk-machine-how-big-data-transformed-the-dairy-industry/256423/>
- Morota, G., R. V. Ventura, F. F. Silva, M. Koyama, and S. C. Fernando. 2018. Big Data Analytics and Precision Animal Agriculture Symposium: Machine learning and data mining advance predictive big data analysis in precision animal agriculture. *J. Anim. Sci.* 96:1540–1550.
- Rossum, G. V., and F. L. Drake. 2009. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.
- van Amburgh, M., E. A. Collao-Saenz, R. J. Higgs, D. A. Ross, E. B. Recktenwald, E. Raffrenato, L. E. Chase, T. R. Overton, J. K. Mills, and A. Foskolos. 2015. The Cornell Net Carbohydrate and Protein System: Updates to the model and evaluation of version 6.5. *J. Dairy Sci.* 98:6361–6380.
- van der Weerd, C., and J. de Boer. 2015. Focusing on behaviour to ensure adoption of big data information services in precision livestock farming. Pages 721–729 in *Proc. 7th European Conference on Precision Livestock Farming*, Milan, Italy. Publications TNO, the Hague, the Netherlands.
- 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.
- Weigel, K. A., R. S. Pralle, H. Adams, K. Cho, C. Do, and H. M. White. 2017. Prediction of whole genome risk for selection and management of hyperketonemia in Holstein dairy cattle. *J. Anim. Breed. Genet.* 134:275–285.
- Wolfert, S., L. Ge, C. Verdouw, and M. J. Bogaardt. 2017. Big data in smart farming—A review. *Agric. Syst.* 153:69–80.
- Wu, Y., D. Liang, R. D. Shaver, and V. E. Cabrera. 2019. An income over feed cost nutritional grouping strategy. *J. Dairy Sci.* 102:4682–4693.