Modeling dairy herd performance for a whole farm system simulation model

By

Manfei Li

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Dairy Science)

at the

UNIVERSITY OF WISCONSIN - MADISON

2021

Date of final oral examination: December 9th, 2020

The dissertation is approved by the following members of the Final Oral Committee:

Victor E. Cabrera, Professor, Dairy and Animal Sciences

Kent Weigel, Professor, Dairy and Animal Sciences

Randy D. Shaver, Professor, Dairy and Animal Sciences

Guilherme J. M. Rosa, Professor, Dairy and Animal Sciences

Kristan F. Reed, Assistant Professor, Animal Science, Cornell University

Abstract

Dairy farms are complex systems comprised of interrelated components and managed with sophisticated procedures. To achieve production and environmental goals, dairy farm decision-makers need to be better informed on the uncertainty of the dairy farm system. A simulation model of the whole farm system might be useful to virtually test management strategies prior to taking operational actions. The Ruminant Farm System model, RuFaS, is being developed as an open-source and highly flexible probabilistic whole-farm simulation model. This thesis introduces a dynamic and stochastic Monte Carlo simulation model of the animal life cycle as a submodule of RuFaS. This RuFaS submodule is intended to provide an extensible framework for the animal and herd simulations, allowing further improvement and expansion. The animal life-cycle submodule simulates the daily growth, production, reproduction, and culling of individual dairy animals and the dynamics of the herd as a group of animals. Each procedure is informed by established scientific literature and industry standards to represent current farm characteristics. A case study examining the use of sexed and beef semen techniques on dairy farms has been conducted using the submodel. The results indicated that switching from conventional semen to a combination of sexed and beef semen was economically advantageous, as the benefits in milk and calf revenue outweighed the additional expenses of breeding and semen.

To obtain more precise production simulation results, we quantified the parameters determining the shape and magnitude of current lactation curves using mixed-effects models on a dataset of over 8 million lactations, each with at least 10 test day recordings. Individual daily targeted milk production estimations for the RuFaS model were to be based on the mixed model projections. Furthermore, the effects of seasonal production changes and long-term trends in milk and milk component productivity were quantified for 17 states with at least 100 thousand lactation records via time-series decomposing and clustering. This time series analysis retrieved trend and seasonality characteristics from state clusters to inform RuFaS estimations of milk and milk component yields.

Acknowledgements

I would like to express my sincere appreciation and gratitude to Dr. Victor. E. Cabrera, my supervisor, for his guidance, support, patience, and wisdom throughout my Ph.D. program. I consider myself extremely fortunate to have Dr. Cabrera as my mentor. I look up to him as a great human being and an excellent scientist. His commitment, vision, and energy have given me the strength to handle research projects and critical life situations.

I would like to express my gratitude to Dr. Kristan F. Reed for her encouragement and mentorship in modeling and writing. She dedicated a lot of time, energy, and patience to our collaboration. Additionally, she serves as a role model for me for being a caring, hard-working, and professional woman in science. Dr. Reed is an expert at directing our RuFaS project, and I am honored and grateful to be a part of the RuFaS team.

I would like to show my thankfulness to Dr. Kent Weigel for his guidance and assistance, especially in getting my hands on the CDCB dataset.

I would like to express my gratefulness to Dr. Guilherme J. M. Rosa for his direction and cooperation on the modeling practices and great recommendations on two courses that were among the most fun and necessary courses I took.

I would like to express my gratitude to Dr. Randy D. Shaver for his advice and support since the early days of my project. He helped me in grasping the concept of bodyweight modeling.

I would like to thank Paul M. Fricke, Dr. Bruce Jones, and Hasan Khatib for their kindness and guidance through my time here at the University of Wisconsin – Madison. Their different perspectives always teach me to understand and learn.

I would like to express my appreciation to the Animal and Dairy Sciences Department staff. Cathy Rook, Nancy Hilmanowski, Debra Schneider, Minh Ngo, and Megan Sippel were very helpful on the paperwork and technical issues.

My profound appreciation also goes to all my lab mates: Dr. Liliana Fadul-Pacheco, Dr. Di Liang, Dr. Hector Delgado, Dr. Veronique Ouellet, Dr. Ramon Novales, Dr. Fan Zhang, as well as Jorge Blanco, Wen Li, and Yijing Gong. I had an ideal environment for study and research with all your generosity and intelligence.

Last but not least, I would like to express my heartfelt appreciation to my family and friends for being the driving force in my life. Thank my parents, Jinlin Li, and Yonghong Wei, for their unwavering love, continuous support, and encouragement throughout my life. When I chose Animal Science as my major in undergrads, I never anticipated I would have so many professional conversations with my parents during my Ph.D. Thank my dear friends, Ruizhe Li, Dr. Beini Lyu, Qian Li, Dr. Yu He, and Dr. Yihe Ma, for the generous support while writing the thesis. I'd like to thank my excellent fourth-floor "coffee buddies", Karen Nielsen, Allison Quick, Dr.Arthur Fernandes, Dr.Ligia Moreira, Dr.Vera Ferreira, Dr.Tiago Passafaro, Ligia Cavanic, Camila Urbano Braz, Dr. Joao Dorea, and Dr. George Shook, for the phenomenon cafe time and Journal Club as well as the fantastic stories about research and life.

I will always treasure these memories of my time as a doctoral student. It's not easy to graduate during a pandemic; I genuinely wish things to return to normal and the world to unite to overcome this difficult period.

Manfei Li,
Madison, Wisconsin

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Researchers, technicians, and farmers' dedication to improving the understanding of the dairy industry and dairy farm operations has brought dairy farm management to a new level. There are more farm management choices and technological innovations than ever before, such as breeding and semen protocols, grouping strategies, and health procedures. However, challenges remain. Further research needs to improve animal welfare and limit environmental impacts while maintaining the dairy industry's sustainable development by providing nutritious food to humans and profitable business to farmers.

Dairy farms are complex systems that include interdependent components with sophisticated management practices. The increasing availability of large amounts of farm data generated by new technologies combined with recent computer modeling and simulation strategies has brought the great potential of better using simulation models to evaluate and assist farm management decision-making. Computer models can evaluate various farm types and management strategies faster and at a much lower cost than field trials. Previous models have been developed to demonstrate farm management practices focused on specific management areas such as culling, reproduction, or feeding. However, these models have limited capacity in capturing the interaction between interrelated components representing the whole dairy farm system. Whole farm models can further evaluate system components' connections, which are not even feasible to investigate through biophysical experiments.

Thus, there is an increasing demand for whole dairy farm system models to fulfill producers, technicians, and researchers to evaluate the interactions among different farm components and further support decision-making. However, only a few whole-farm dairy models exist. These models have limitations when representing modern farms, the latest technologies, and complex management strategies. In addition, these whole dairy farm models focus mainly on nutrient cycles and only include a few herd attributes with simplified animal components. Therefore, it is essential to have a next-generation whole-farm dairy model with a detailed animal submodel capable of evaluating the implementation of recent and upcoming technologies and innovative management strategies on dairy farms.

This thesis aims to contribute to a whole farm system model, the Ruminant Farm System model (RuFaS), by developing a stochastic simulation framework of the dairy herd performance as a submodule, the animal life cycle submodel, which simulates the dairy farm with all its multiple components. Furthermore, two data analysis studies were conducted for improving the representation of individual dairy cow productivity estimation by leveraging a large national database of production records. The first data analysis quantified the parity-, year-, month-, region-, and milking frequency-specific lactation curve parameters via mixed effect models. The second analysis used time-series decomposing and clustering techniques to identify the seasonality and long-term trend of milk and milk component production according to state, parity, and calving date. Results of both data analyses studies are to be implemented into the animal life cycle submodel for individual lactation estimates to improve the predictability of targeted daily production.

And therefore, the literature review presents simulation models and methods, lactation curve modeling, and time-series decomposing and clustering methods. The framework of the herd simulation model, particularly the animal life cycle submodel, is then discussed in depth and illustrated by comparing the herd characteristics and economic outcomes of alternative reproductive management strategies that incorporate the use of sexed and beef semen. Following that, the lactation curve parameters are qualified and discussed in relation to the effects of temporal, geographical, and management aspects. Finally, the seasonality and long-term trend of milk and milk components are evaluated using time series analysis.

1.2 Thesis Outline

This thesis is comprised of 6 chapters. Following this chapter, the remainder of this thesis is presented as follows:

Chapter 2 provides a comprehensive review of the literature on the design, modeling, and processes of developing herd simulation models and modeling techniques of nonlinear curve fitting, mixed effect modeling, and time-series decomposing and clustering.

Chapter 3 describes a Monte Carlo simulation model developed as a submodule, the animal life cycle for a whole farm simulation model (the Ruminant Farm System Model, RuFaS) to simulate individual animal and herd dynamics. The model was used to assess the economic and reproductive performances of alternative reproductive management strategies introducing sexed and beef semen usage.

Chapter 4 includes a two-step approach to analyze lactation curve parameters. The 1^{st} step was animal-parity parameter estimation from individual lactation curve records with a nonlinear least-squares optimization algorithm. The 2^{nd} step was a mixed-effect model analysis of individual parameter estimates of temporal, spatial, and management factors as fixed effects and herd and animal as random effects. Parity-, year-, month-, region -, and milking frequency-specific lactation curve parameters are reported for further applications such as incorporating within the animal life cycle submodule of RuFaS.

Chapter 5 presents a time-series analysis on qualifying the historical trend and seasonal patterns of milk and milk component production by parity and US state. The analysis then classified states and parities with similar trends and seasonal patterns into clusters to adjust expected milk production estimation according to calving date and location in simulation models such as the animal life cycle submodule of RuFaS.

Chapter 6 states a global thesis conclusion and potential future work.

Appendix A has a pseudocode demonstration as part of the animal life cycle model code documentation.

Chapter 2

Literature Review

2.1 Introduction

This thesis focuses on modeling dairy herd performance for a whole farm system simulation model. The modeling process is divided into two: the first is to develop a simulation framework for representing the life cycle characteristics of animals in a model of a whole farm system; the second is to learn from a large lactation record dataset in order to gain a better understanding of lactation patterns and to collect information for improving the precision representation of lactation patterns in our simulation model.

This literature review focuses on modeling methodologies for both components of the thesis; the first section discusses model development, with the primary objective of identifying an appropriate architecture for translating scientific and empirical knowledge into the simulation process. Thus, a review of existing farm system models, herd simulation models, and simulation methods are reported. The essential components relating to herd components in farm systems models and those relating to the life cycle model in herd simulation are described, along with the simulation method to be used, in order to aid in determining the most appropriate modeling strategies to meet our goals of developing an animal life cycle simulation model.

The second section of this literature review introduces and discusses the modeling strategies utilized to exploit the massive dataset of lactation records during the data mining process. These techniques are lactation curve modeling, linear mixed effect modeling, time series decomposition, and time series clustering. These strategies aided the project in extracting data characteristics to improve the representation of milk productivity in the simulation model.

2.2 Farm system models

2.2.1 The system modeling

The general system theory establishes an integrating framework for studying and comprehending complex systems by including components from many disciplines (Dent and Anderson, 1971). In general, a system is described as a collection of items or entities that act and interact in order to accomplish a specific objective (Banks et al., 2009; Velten, 2009). The application of system analysis has been extended to a variety of domains, including technical, biological, organizational, psychological, and agricultural systems (Dekkers, 2017). Within the systems, there are various levels. The dairy farm system model encompasses all systems on a dairy farm, from individual animals through herds to the farm as a whole. Individual animals follow their biological processes, influenced by system management strategies. A herd is comprised of all animals of various statuses that interact with one another, farm equipment, the surrounding environment, and various management practices. On a higher level, the farm system encompasses the herd, the land, the machinery, the crops, the soil, and manure management. The system environment comprises external aspects of the system border, such as the weather, the market, and the farm's social context, that interact directly with internal elements.

2.2.2 The history of the dairy system modeling

For more than a half-century, agricultural system modeling has been developed, applied, and counted with contributions from various scientific disciplines (Jones et al., 2017). Nutritional requirement tables for cattle were established in the United States in 1945 (NRC, 1945), and nutrient requirement tables for ruminants were released in the United Kingdom in 1965 (ARC, 1965). In the 1970s, early herd dynamics simulation models were developed, which were critical for modeling the whole livestock farm, including disease and reproduction components (Freer et al., 1970; IADB, 1975; Davis et al., 1976; ILCA, 1978; Sanders and Cartwright, 1979, Konandreas and Anderson, 1982). In the 1980s, pasture models were proposed (Johnson and Thornley, 1983; Coughenour et al., 1984), as well as dynamic programming modeling of herd replacement decisions (Van Arendonk and Dijkhuizen, 1985). In 1992, the Cornell Net Carbohydrate and Protein System (CNCPS) was launched for ruminants (Russell et al., 1992). In the late 1990s, extensive uses of livestock systems models that integrated individual animal performance, herd dynamics, pasture, and crop models were made (Herrero et al., 1996, 1999, Freer and Donnelly, 1997, Rotz et al., 1999). Global livestock models regarding environmental impact have been created in recent decades across countries (Bouwman et al., 2005; FAO, 2013; Herrero et al., 2013).

Additionally, soil, crop, and water are critical components in integrated farm system models. Around the 1950s, the first computational analyses of plants and soil were produced (van Bavel, 1953; de Wit, 1958). Modeling of soil water balances initially appeared in the 1960s (Slatyer, 1960, 1964; Keig and McAlpine, 1969). Crop forecasting using remote sensing was first reported in the mid-1980s (Ritchie and Otter, 1985). In the 1980s, a soil productivity estimation model influenced by erosion and a soil nitrogen (N) model for forecasting crop responses were developed (Williams et al., 1983). The publication of the first Intergovernmental Panel on Climate Change (IPCC, 1990) assessment report generated a global focus on modeling agriculture's environmental impact, prompting the Australian government, for example, to establish APSIM to model agricultural systems (McCown et al., 1996; Keating et al.,1991, 2003). The European Union funded the System for Environmental and Agricultural Modeling: Connecting European Science and Society (SEAMLESS) project in the late 2000s (van Ittersum et al., 2008), and global livestock models were developed to assess and test strategies for reducing the environmental impacts of livestock production systems, including land use, emissions, economics, and biomass use (Havlik et al., 2014; Cohn et al., 2014, PBL, 2013, Bouwman et al., 2013)

Outside of farm boundaries, the Livestock Revolution study (Delgado et al., 1999) forecast increased consumption in the animal sector because 'as people become wealthier and societies urbanize, they consume more livestock.' Numerous impacts of livestock on agricultural growth are now well known as a result of these studies. However, additional research and holistic evaluation are required to meet the world's expanding food demand while also addressing environmental problems.

2.2.3 The whole-farm system models

A review of the techniques used to develop innovative agricultural production systems at the farm level classified approaches into two categories (Le Gal et al., 2011): those directed toward "design" and those oriented toward "design support." The "design" oriented approaches included studies on creative production system design, while the "design support" oriented methods emphasized the exploration of alternative management practices, activities, and technology. For most of the whole farm models, the concept "design support" methodologies were leveraged. Numerous models have been developed from a variety of dairy farm systems. In order to better understand the performance of herd simulation in relation to the whole farm system and other components in the system, the following models are described here:

i Integrated farm system model (Rotz et al., 2018)

ii Dairy Wise (Schils et al., 2007)

iii Global activity model for evaluating the sustainability of dairy enterprises (Vayssières et al., 2009) iv Dairy cattle production system integrated with forage crop production (Kikuhara et al., 2009)

v Dairy model (Johnson et al., 2008)

vi Grazing plan integrated with Agricultural Production Systems Simulator (Freer et al., 1997)

2.2.3.1 The Integrated Farm System Model (IFSM) was developed as a research tool for evaluating and comparing farming systems' environmental and economic viability. Crop and pasture productivity, feed consumption, animal nutrition, and manure nutrients are simulated for beef (Rotz et al., 2005) or dairy (Rotz et al.,1999, 2013) farms under a variety of meteorological circumstances throughout time. The characteristics of the herd and the feed consumed influence the quantity and nutrient content of the manure produced. Nutrient flows are tracked throughout the farm in order to forecast nutrients released into the environment and soil. A farm-gate life cycle assessment was used to determine the carbon, energy, water, and reactive nitrogen footprints of milk and meat produced (Rotz et al., 2013). Costs of production, income, and economic return are also modeled for each year of meteorological conditions. It typically mimics historical weather over a 25-year period, and the resulting dispersion of annual projections reflects the variability of weather influences.

Six animal groups comprise the IFSM model's animal and herd characteristics component: young stock, heifers, three lactating cow groups, and nonlactating cows. All cow groups are separated into primiparous and multiparous cows based on the culling rate specified by the user. It consists of seven breeds. Each animal group's prospective milk yield, milk fat content, bodyweight change, and fiber ingestive capacity are all defined. Milk yield is predicted using the gamma function, with parameters established for an average Holstein herd producing 9070 kg of milk per year and an average bodyweight of 600 kg during lactation months 2–5. The conventional technique assumes that 15% of cows are in early lactation, 23% are in mid-lactation, 46% are in late lactation, and 15% are not lactating. Animal traits, such as body weight change, are averaged across that period. The lactation cycles are set at 56 weeks, and the calendar year's feed intake and milk production are totaled.

2.2.3.2 DairyWise (Schils et al., 2007) simulates internal and external material and nutrient flows in dairy farming systems and calculates economic characteristics. Changing a single component can have a ripple effect throughout the entire farm system. DairyWise is capable of simulating grazing systems, including grass growth and quality, as well as the management implications of animal intake and excretion. Additionally, it can model the impacts of manure production and composition on grassland. DairyWise is based on agricultural and animal trials conducted in the Netherlands and is therefore region-specific.

The DairyWise approach categorizes dairy cows and young stock separately. The cow's submodel predicts individual cows' feed intake and milk production. The young stock submodel simulates the young stock's feed intake and growth. Cows are fed either TMR diets or roughage mixtures supplemented with concentrates ad libitum. The ratio of feed intake capacity to satiety value is used to compute the feed intake. In this model, the user defines concentrate intake and supplemented roughages during grazing while the model calculates fresh grass. The overall energy intake is divided into four categories: maintenance, growth, pregnancy, and milk production, as well as a two-way flow of mobilization or repletion of body tissue energy reserves. The energy requirements of the various components are determined in relation to production level, parity, and lactation stage. Actual milk production is adjusted daily by comparing the standardized milk production energy uptake to the standardized milk production energy uptake. The feed intake of the young stock is calculated using a typical growth curve. The model was designed to maximize roughage consumption. When roughage alone is insufficient to meet energy requirements, concentrate is added. The cow and young stock models simulate individual animals before merging to form a herd. Each month, a user-defined annual replacement rate is used to determine the culling rates for each animal type and parity.

2.2.3.3 GAMEDE (Vayssières et al., 2009) is a rule-based simulation model evaluated in the context of a stock-flow model. Forages, concentrate feeds, milk, animals, and dung are the major biomasses circulated within the system. There are two sorts of flows: human-induced flows (forage harvesting, animal feeding, manure removal, and manure spreading) and natural-induced flows (grass growth, milk, animal biomass, and manure production, and N volatilization). Biomass cycles are divided into user-driven and naturally occurring flows within the two loops. The farm is an open system where users control biomass imports and exports through the milk and animal sales, feed and fertilizer purchases, and loss due to specific biophysical processes such as nitrogen gas emissions.

GAMEDE models farming using sets of ordinary differential equations. The herd demography module models the herd and animal culling composition using parameters for reproduction, mortality, culling, and replacement. Demography of the herd is critical for assessing forage and feed consumption, milk production, and manure excursion. The herd production module simulates animal weight changes and milk and manure production in response to feeding rations and animal intake during grazing.

2.2.3.4 The dairy cattle production system integrated with forage crop production (Kikuhara et al., 2009) is an optimization model that optimizes the economic and environmental effects of the entire farm. This model may be applied to various Japanese dairy systems by varying the genotype, nutrition, management, and economic variables of the herd and forage crop varieties and area. The model incorporates nutrient requirement prediction, diet formulation optimization, and herd dynamic simulation to optimize the whole farm. The herd dynamic model determined the equilibrium number of cows in each reproductive cycle. It maintains a steady herd size; all replacement heifers are raised on-farm, and all male calves are sold shortly after birth. The growth phase runs from birth to the first conception, the first reproductive cycle runs from conception to the end of the first lactation, and subsequent reproductive cycles run from the end of one lactation to the end of the next lactation. At the conclusion of the final cycle, all cows are culled. The herd composition is determined by the predetermined culling rate for animals with reproduction failures during each cycle. This model suggested that the amount of animal excretions (feces and urine) was utilized to create manure and then applied as organic fertilizer to home-grown farmland.

2.2.3.5 DairyMod (Johnson et al., 2008) is a biophysical pasture simulation model. The model models how pasture growth is utilized by grazing animals. It offers simulations of single and multiple paddocks with a variety of soil types, nutritional status, pasture species, fertilizer, and irrigation management. Johnson (2003) provides a mathematical explanation of DairyMod, Johnson and Thornley (1983) provide the pasture growth model, and Johnson et al. (2012) provide the animal component. Animal growth, metabolism, pregnancy, and lactation are all affected by the amount of energy available from a combination of pasture, concentrate, mixed ration, and forage feed supplies (Johnson, 2012). It is a model of animal growth and metabolism that is energy-driven. Separate models of body protein, water, and fat are used to drive energy simulations. The critical measure of the metabolic condition is animal protein weight, while fat is considered a potential source of metabolic energy. Maintenance and energy required for activity are prioritized over new tissue growth. New fat growth is dependent on the current protein weight and the maximum possible fat fraction of the body mass. Fat reserves can be used to generate energy. The goal is to raise an animal in natural conditions. It is hypothesized that potential intake increases during lactation or pregnancy as a result of physiological changes in the rumen. The growth of the fetus is assumed to be exponential. The simulation begins by calculating the ME necessary for development, maintenance, pregnancy, and nursing. Fat catabolism occurs within the first 95 days after parturition, diverting energy away from fat development and toward milk production. ME is made available to nonlactating animals, and the requirements for growth, maintenance, and pregnancy are compared in order to alter intake, growth, fat catabolism, and weight loss following maximum fat catabolism. Lactation occurs only when available energy exceeds maintenance requirements in response to pasture supply, quality, and lactating animal feed management. Validation of the pasture model was conducted using data from Australia and New Zealand (Cullen et al., 2008). The number of animals, the date of birth, and the duration of nursing are utilized to establish the animal classes that will be simulated (up to 12).

2.2.3.6 The Agricultural Production Systems Simulator (APSIM, Holzworth, et al., 2014) is a modeling framework for a paddock or a farm with a defined protocol (Moore et al., 2007) that connects soil processes, plant, and weather models, as well as management principles (Moore et al., 2014). Additionally, the model incorporates an organic matter model and simulations of the dynamics of N, P, and S. APSIM is a simulation tool for cropping systems and the interplay between cropping and grazing systems. The GRAZPLAN model is integrated into the APSIM to serve as the animal model. GRAZPLAN (Donnelly et al., 2002) incorporates a model of ruminant dynamics (Freer et al., 1997). It was developed in conjunction with the Australian feeding standard (CSIRO, 2007) and has been studied in Canada (Cohen et al., 2003) and China (Donnelly et al., 2005). Sheep, beef, and dairy cattle are all included in the model. The sub-model for energy and protein nutrition is based on Australian dietary guidelines (CSIRO, 2007). The model is intended for ruminants that graze or are fed pasture-based or concentrate-based diets. The herd's status is maintained through a list of physiologically diverse animal groups. Daily computations are made using daily flows of metabolizable energy, rumen-degradable energy, and rumen-degradable protein. They are adaptable to a variety of sheep and cattle breeds by altering the mature weight of the animal. Energy and protein are considered for maintenance (including body temperature), fetal growth, lactation, wool production, and body weight change. The rates of conception and mortality are explicitly modeled as stochastic functions of the animal's genotype, age, weight, and season. The potential for milk production is assessed. Milk production may fall short of its potential for one of two reasons: either ME consumption is insufficient to sustain the potential, or production is constrained by the ability of simulates from calves or milking machines. On a given day of lactation, potential milk output expressed as the milk's ME value for the young is anticipated using Wood's (1969) method for forecasting milk yield as a percentage of potential yield in relation to available ME after maintenance and pregnancy deductions. The connection to APSIM enables flexible management and economic modeling of several paddocks with varying soils and forages.

In general, the herd component of farm system models includes the processes of milk production and feed intake at the animal or herd level; some models may also include changes in body weight (DairyMod) or genetic component (DSIF). The interaction of the animal with the pasture growth module and the soil process is stressed in the grazing model. Other than the herd, the system model frequently incorporates soil, crop, feed supply, manure, weather, economics, and environmental components. Certain models incorporate a greenhouse gas dynamic (DairyMod) or have management criteria (GrazPlan).

2.3 Herd simulation models and methods

2.3.1 Concepts and categories of simulation models

In animal herd modeling, simulation methods and mathematical programming are frequently used as modeling methodologies. They could, however, be different. Mathematical models are frequently used to determine the ideal input-output combination, but simulation approaches are used to study the output by mixing the inputs in ways other than optimization (Velent, 2009). For instance, linear and nonlinear programming is frequently employed in diet formulation to reduce costs or maximize revenue over feed costs. Additionally, mathematical models can be incorporated as a submodel into the simulation process to give optimum inputs during the simulation.

Simulation models can be classified into the following categories (Kalantari, 2016): static or dynamic in terms of time dependency; discrete or continuous in terms of the type of time variables; deterministic or stochastic in terms of event uncertainty; and empirical or mechanical in terms of the system's hierarchy.

Based on output analysis, dynamic simulations can be further categorized as terminating (transient) simulations or steady-state simulations (Rubinstein and Kroese, 2007; Banks et al., 2009). In a terminating simulation, the inputs are utilized to simulate for a particular time or terminate the simulation to investigate the system's evolution through time given its initial state. A steady-state simulation is concerned with the system's long-run qualities. The model begins with an arbitrary state for the system and runs indefinitely until its original conditions do not affect its attributes (Law and Kelton, 2007; Banks et al., 2009).

The distinction between deterministic and stochastic simulation is illustrated by the variability of the input variables, which affects the variability of the output variables. In deterministic models, the input variables are the mean values, but in stochastic models, the input variables are the probability distributions of the variables. Thus, when the variance of input variables is modified, the output of deterministic models remains constant, whereas the output of stochastic models changes proportionately (Sorensen, 1998). A stochastic model contains at least one random input variable. (Reuven et al., 2017). There are two widely used stochastic simulation methods: 1) Markov chain models, which control the transition matrix of the probability distribution of movement from one state to another in each step; and 2) Monte Carlo simulation models, which control discrete events using pseudo-random number generators drawn from appropriate probability distributions related to the events.

2.3.2 Herd simulation models

Numerous simulation models of the dairy herd have been developed to examine various aspects of the dairy farm. For example, health (Allore et al., 1998; Østergaard et al., 2005; Dolecheck et al., 2019); culling management (Kristensen and Thysen, 1991; Kudahl et al., 2011); breeding and genetics (De Vries and Salfer, 2013; Kaniyamattam et al., 2016; Cottle et al., 2018); reproduction (Giordano et al., 2012; Galvão et al., 2013; Pfeiffer et al., 2020); and nutrition (St-Pierre and Thraen, 1999, Kalantari et al., 2016). The following sections provide an overview of many dynamic herd simulation models:

2.3.2.1 SimHerd. SimHerd was one of the first herd simulation models developed, and it is widely used in the literature (Østergaard et al., 1996; Plaizier et al., 1997; Nielsen et al., 2004; Ettema and Østergaard, 2006; Hjort et al., 2015; Ettema et al., 2017; Pfeiffer et al., 2020). SimHerd is a dynamic, stochastic Monte Carlo simulation framework that simulates adults and young stock on a weekly basis to determine the influence of various management practices (Sorensen et al., 1992).

Nine state variables are used to define each animal in the model: age, lactation stage, lactation number, estrus status, pregnancy status, culling choice, milk production potential, milk production, and live weight. Heat detection, pregnancy, fetal death, calf sex and viability, involuntary culling, and death are all discrete events. At the individual-cow level, events are triggered stochastically, and then the herd structure and dynamics are generated. Additionally, the herd structure and production level are influenced by a specific production system or a particular management approach involving decision variables (Sorensen et al., 1992). SimHerd I's first version was introduced in 1992. (Sorensen et al., 1992). Following that, enhanced versions were made available through various investigations. SimHerd II (Østergaard et al., 2000) included a complex for feeding and health production (the milk production and feed intake can be affected by metabolic and reproductive diseases). SimHerd III (Østergaard et al., 2003) included risk variables and consequences of milk fever. SimHerd IV (Østergaard et al., 2005) included somatic cell count and mastitis as additional variables. SimHerd V (Ettema et al., 2011) separated genetic and permanent environment effects on calf milk yield potential and set the calf milk yield potential to the average of its parents.

2.3.2.2 Galvão et al. (2013) conducted a study comparing timed artificial insemination, heat detection, and a combination of synchronization and heat detection reproductive programs using a dynamic, stochastic Monte Carlo simulation model. Daily reproductive status and milk production of individual cows were tracked in accordance with a predefined reproductive protocol and a sequence of events, which included estrus, estrus detection, insemination, diagnostics, and calving. The incidence and detection of estrus, the success of AI, the occurrence of pregnancy loss, calving, involuntary culling, and mortality were all considered stochastic events. Daily milk yield was calculated using Wood's (1967) model with pregnancy-related adjustments. Bodyweight was modeled as a continuous equation of age, days in milk, and days in pregnancy. Dry matter intake was modeled following NRC recommendations. Cows were culled if milk income fell below the cost of production or if they were not pregnant after 450 days in milk. Daily probabilities of death and involuntary culling were used to model the probability of death and involuntary culling. The first insemination was performed using the Presynch-Ovsynch protocol, while the second and subsequent services were performed using Ovsynch. A herd of 1,000 cows was simulated every day for 3,000 days using the Netlogo software to reach steady-state, and then 2,000 days were simulated and utilized to calculate profit. The findings assessed critical parameters affecting the herd's reproductive success, including estrus detection rate, compliance to the reproductive protocol, and milk price. The results indicated that combining TAI and ED for reproductive management could be profitable. However, if a high ED with high accuracy can be achieved using the ED method, or if high compliance with injections can be accomplished with TAI programs, relying solely on just ED or TAI may be more profitable than attempting both. (Galvão et al., 2013).

2.3.2.3 Kalantari et al. (2016) evaluated the implementation of nutritional grouping in commercial dairy herds using a dynamic, stochastic Monte Carlo simulation model. Individual cows were modeled daily after the first parturition using the next-event scheduling method (De Vries, 2001). At initialization and each calving, the stochastic events that could occur to cows throughout each reproductive cycle were arranged. These events included voluntary culling, death, pregnancy, abortion, dry-off, and parturition. Daily updates were made to the cow's simulated state and the nutrient requirements for net energy for lactation and metabolizable protein. The changes in their body weight and body condition score were caused by their consumed energy. The cows were grouped according to their nutrient concentration requirements using the monthly clustering method. When regrouping occurs, cows are rated according to their net energy for lactation and metabolizable protein and then assigned to the same or different nutritional categories. The daily energy balance of each cow in the herd was calculated and tracked using this structure. The income over feed cost analysis of nutritional grouping revealed that most of the economic benefits of nutritional grouping for nursing cows could be met by having two nutritional groups rather than one. While having three groups results in a better profit than having two groups, the additional profit may not be sufficient to offset the management cost of regrouping. The economic benefit mainly was due to increased milk output and lower RUP costs, which were highlighted in times of economic difficulty.

2.3.2.4 Calsamiglia et al. (2018) constructed a dynamic, stochastic mathematical model of a dairy farm. The model is divided into many submodules that address reproductive, feeding, disease, heifer raising, environmental factors, infrastructure, management, and economics. Lactation curves, growth, nutrient requirements, and reproduction are all influenced by environmental and management factors such as heat stress, milking frequency, grouping strategies, disease, and nutrient supplies. Throughout the simulation period, the simulation method was performed every day for each animal (heifers and cows). Each cow on the farm has a theoretical milk yield potential attributed to it. Milk production was estimated using Wood's lactation curve model. Following calving, the cows resumed their estrus cycle. The AI occurred after the user's voluntary waiting period with the estrus detection index for the reproductive program. When milk output fell below a predetermined level, the cow was dried if pregnant or culled if not pregnant. The farm's heifer capacity was set at 60% of adult animals. When the farm's replacement herd was depleted, it purchased ready-to-calve heifers. All males were sold at calving. The energy requirements were determined using NRC data (2001). Environmental circumstances, management practices, and disease incidence all have an effect on growth. A matrix of probable disease impacts was constructed for the most prevalent disorders, a. Users have complete control over the feeding groups as long as each group contains a diet description. Additionally, the model addressed heat stress, which was calculated weekly using the temperature-humidity index. Culling occurred for a variety of reasons, including mortality, disease, and low production. Outputs established the simulation model's correct behavior from many farms and sensitivity assessments to various plausible scenarios.

2.3.3 Events scheduling for Monte Carlo simulation methods

2.3.3.1 General Method

In the agricultural systems with life cycles, the state changes over time with time steps (e.g., daily, monthly, yearly), and the simulation is driven by the occurrence of discrete events with the time variable. The model described in chapter 3 is a discrete-event dynamic system. In particular, the states are changed at a countable number of time points. Those changes are triggered by the execution of simulation events occurring at the corresponding time points. (Reuven et al., 2017). The Monte Carlo simulation method can be used to model discrete-event dynamic systems.

Monte Carlo methods were first used to solve physics problems; in the late 1940s, Stanislaw Ulam invented the modern Monte Carlo method on nuclear weapons projects at the Los Alamos National Laboratory (Metropolis et al., 1949). Then, its usage expanded to other domains. Monte Carlo was named after a gambling house, in which the games of chance connected with the need for random numbers in the simulation (Brandimarte, 2014). Monte Carlo methods are computational algorithms that rely on repeated random draws from probability distributions to compute the distribution of outputs. Monte Carlo simulations model the systems' dynamics and generate different samples from the inputs to estimate a probability or an expectation to the output (Reuven et al., 2017).

2.3.3.2 Steps of Monte Carlo simulation

1. Pseudo-random number generation. A computer is often used to generate random numbers. The most widely used method is the mixed congruential method, which generates a sequence of random numbers by calculating the next one from the last one with an initial seed (Hillier

and Lieberman, 1986). The numbers generated by computer programs should be called pseudo-random numbers (Law and Kelton, 2007). The pseudo-random numbers are useful in developing, debugging, and reducing the results' variance in Monte Carlo simulation models (Banks et al., 2009).

2. Transformation to random variates. Generating random draws from an established distribution (e.g., exponential, uniform, or triangular or empirical) distributions is called random variate generation. The most commonly used technique is called the inverse transform technique (Banks et al., 2009). This technique can be used for all the distributions, but it is not always the most efficient (Banks et al., 2009). When the cumulative distribution function (CDF) is easily invertible, it chooses univariate random variates from continuous distributions (Devroye, 2006). When the CDF is not easily invertible for discrete distributions, it can be described graphically and used to generate random variates.

3. Domain-Specific sample path generation. This step is building a realistic model representation of the underlying system's characteristics in a computer. It uses sets of input parameters to define and initialize the system and a set of decision rules that drive the model's behavior with the parameters (Kristensen et al., 2006).

4. Output analysis. Verification checks whether the conceptual model is correctly implemented in the model. Validation attempts to confirm if the model is an accurate representation of the system (Banks et al., 2009). Objective validation of the model includes statistical tests or visualizations to find the degree of agreement between the model outputs and the reality.

2.4 Lactation curve modeling and mixed-effects model

2.4.1 Lactation curve modeling

Using lactation-curve models to describe and predict milk production is beneficial (Murphy et al., 2014). Lactation curves are mathematical functions that describe milk yield corresponding to days in milk (DIM) during each lactation. The lactation curve models are among the most successfully employed mathematical models in agriculture (Bouallegue and M'Hamdi, 2020). The lactation curve's shape can provide helpful information for dairy farm management. For instance, farm managers can use projections of milk production decline to help them make dry-off and culling decisions.

Several mathematical functions were proposed for the past century with equations differencing in parameter numbers and representing the lactation curve's typical pattern with peak yield, time to peak, and persistency (Brody et al. 1923; Sikka 1950; Wood 1967; Wilmink 1987; Ali and Schaeffer 1987; Guo and Swalve 1995; Quinn, 2005). The parametric models are easy to interpret and fit with actual farm data. With the aim of representing the lactation curve accurately in our simulation model, a parametric model with less mathematical complexity and a strong ability to fit many kinds of curves is desired.

The first published lactation curve dates back to 1923 by Brody et al. (1923). $Y_t = a * exp(-ct)$, Where a is the initial level, and c correlates with the descending phase. By identifying the limits on representing the ascending phase, they modified the equation as $Y_t = a * exp(-bt) - a * exp(-ct)$ (Brody et al., 1924). The adjustment was an improvement but reported underestimates in early lactation and overestimates in later lactation (Cobby and Le Du, 1978). This model was further modified as $Y_t = a - bt - a * exp(-ct)$ (Fischer, 1958), but it was found to overestimate the time to peak (Rowlands et al., 1982). Vujicic and Bacic (1961) proposed $Yt = tc^{-a} exp(-ct)$, which lacks the initial value information. Wood (1967) proposed the incomplete gamma-type function in the ground of the above models and their limitations: $Y_t = a * t^b * exp(-ct)$. It captured both the rising and decay pattern of the lactation curve and later became the most popular lactation curve model. In the model, the relation of the parameter and the shape of the curve is described. Parameter a is the scale factor for initial milk yield, b is the rate factor for the rate of increase in milk yield to peak, and c is the rate factor for the rate of decline in milk yield after the peak. The Wood's model was reported as the best fit with multiple datasets (Rao, 1977; Singh and Bhat, 1978; Morant and Gnanasakthy, 1989; Rashia, 2010), although some limitations of it were also reported (Dijikstra et al., 2010; Macciotta et al., 2011). There were some adjustments to the Wood's model to make it more flexible (Dhanoa, 1981; Jenkins and Ferrel, 1984). Cobby and Le Du (1978) proposed: $Y_t = a - bt - a * exp(-ct)$, and Wilmink (1987) further modified it to a linear model: $Y_t = a + b * exp(-kt) + ct$, where k relates to the peak time and it is helpful for model daily production in early days. Ehrlich (2011) proposed Milkbot: $Y_t = a * (1 - e^{c-t/b}/2) * e^{-dt}$, where a is the scale parameter, b controls the rate to rise, c is the offset of the starting milk yield, and d is the decay parameter.

There are other models such as Ali and Schaeffer (1987) with logarithm-based format: $Y_t = a + b(t/305) + c(t/305)^2 + d(305/t) + e(305/t)^2$. Where a is associated with the peak yield, d, and e with increasing slope, b, and c with the decreasing slope. Also mixed-log-model (Guo and Swalve, 1995): $Y_t = a + b/\sqrt{t} + cln(t)$. Multiple papers have evaluated the goodness of fit for those lactation curve models in each study (Guo and Swalve, 1995; Val-Arreola et al., 2004; Silvestre et al., 2006; Cunha et al., 2010; Korkmaz et al., 2011; Adediran et al., 2012; Lee et al., 2020), and the best model changed from study to study. Dongre et al. (2011) summarized the goodness of fit for 78 studies of lactation curve fitting conducted with various breeds worldwide. They concluded that there isn't the "best model" across studies; however, the gamma type function was the most popular.

2.4.2 Mixed effects models

In 1919, while investigating Mendelian heredity, Ronald Fisher proposed random effects models (Fisher. 1919). One hundred years later, mixed effects models are often employed to analyze clustered data or longitudinal studies in which participants are assessed periodically over time. By including both fixed and random effects, these models enhance the standard linear model. Fixed effects could be thought of as the usual covariates in a standard linear regression model (Brown 2021). These variables are supposed to have a fixed influence on the response variable across the data set. For instance, each linear regression coefficient may be interpreted as the unit rise or reduction in the response variable when the covariate associated with it is raised by one unit. We are interested in estimating the linear influence of the covariate on the response across the population we are examining. On the other hand, random effects could be factors that result in varying response levels for distinct sets of observations. We regard them as random because we cannot quantify these factors or are not interested in predicting their influence on the answer. In general, we aim to understand and estimate the fixed effects in our statistical model, as well as account for the random effects. We are not interested in analyzing individual test participants, but in accounting for their differences and allowing for the correlation of repeated measurements on the same topic. Our findings would be unreliable if we assumed that different observations on the same test subject were independently and identically distributed, as we would in a standard linear model. By using a block diagonal covariance structure, the mixed model automatically adjusts for correlation between data from the same individual. Numerous statistical software for fitting linear mixed effects models has been created. Among the most widely used is lme4 in R, which was released in 2013 (Bates et al., 2013). The lme4 software fits the linear mixed effects model in a frequentist framework by estimating the regression parameters using either maximum likelihood or restricted maximum likelihood.

Mixed-effects modeling enables us to evaluate an observable condition while also accounting for heterogeneity within and among herd and animal. Typically, parity and calving date effects are fixed effects because they are anticipated to behave predictably across different samples of herds and animals. Indeed, milking frequency will be modeled as a fixed effect in chapter 4 because we anticipate that an average link between modality milking frequency and milk yield will appear if we repeat the experiment with a new sample of herds and animals.

2.5 Time series analysis methods

2.5.1 Time series

Time series is a class of temporal data objects which can be easily obtained from chronologically observations along time (Fu, 2011). Time-series data often is large in data size and treated as a whole matrix, other than individual numerical fields, because of its continuous nature (Fu, 2011). Time can be regarded as hours, days, months, years, etc. Time series are widely used in medication usage (Wagner et al., 2002), weather forecasting (Campbell, 2011), air pollution assessment (Peng, 2006), and much more. In dairy-related research, time-series methods have been used for milk production prediction (Macciotta et al., 1999; Codrea et al., 2011; Taye et al., 2020), estrus and disease detection (de Mol et al., 1999), and animal mortality (Nørgaard et al., 1999; Morignat et al., 2015).

2.5.2 Time series decomposing

Time series data can display a wide range of patterns, and it is frequently beneficial to decompose a time series into many components, each reflecting a distinct pattern category. The level, annual trend, seasonal, and random changes are all examples of time-series effects (Cleveland et al., 1990). There are two types of classical decomposition: additive and multiplicative. The additive decomposition is useful when the amplitude of seasonal variations or variation around the trend-cycle is constant with respect to the time series level. A multiplicative decomposition is more appropriate when the seasonal pattern's fluctuation or the variation around the trend-cycle appears to be proportionate to the time series' level.

When there is a long-term shift in the mean level, a trend pattern occurs, and when seasonal elements have an effect on the time series, a seasonal pattern exists. In the additive decomposition, the trend component is first identified using a moving average and then eliminated from the time series; this step eliminates the majority of the autocorrelation in the long-term feature. Following that, the seasonal component was averaged for each year and then centered; this step eliminates seasonal autocorrelations. Finally, we determined the reminder component by subtracting both trend and seasonality.

2.5.3 Time series clustering

Time series might also be used to analyze many time series concurrently in order to determine the similarities and differences between them via clustering. Clustering divides data into homogenous groups in which similarity within groups is maximized and dissimilarity between groups is minimized (Aghabozorgi et al., 2015).

2.5.3.1 Clustering methods. Clustering may aid in the comprehension of the patterns seen within a cluster. There are numerous frequently used methods for clustering: 1) partitioning, or clustering, using k partitions, such as the k-means method; 2) hierarchical, or grouping objects into a tree of clusters; 3) density-based, or growing neighbors until preset thresholds are reached; 4) grid-based, or using the grid search; and 5) model-based, for example, using neural networks. The k-means algorithm and hierarchical clustering are two of the most frequently used clustering algorithms in time series clustering (Aghabozorgi et al., 2015). Knowing the desired number of clusters in the end output is required for the k-means method, but hierarchical clustering does not require previous knowledge of the desired number of clusters.

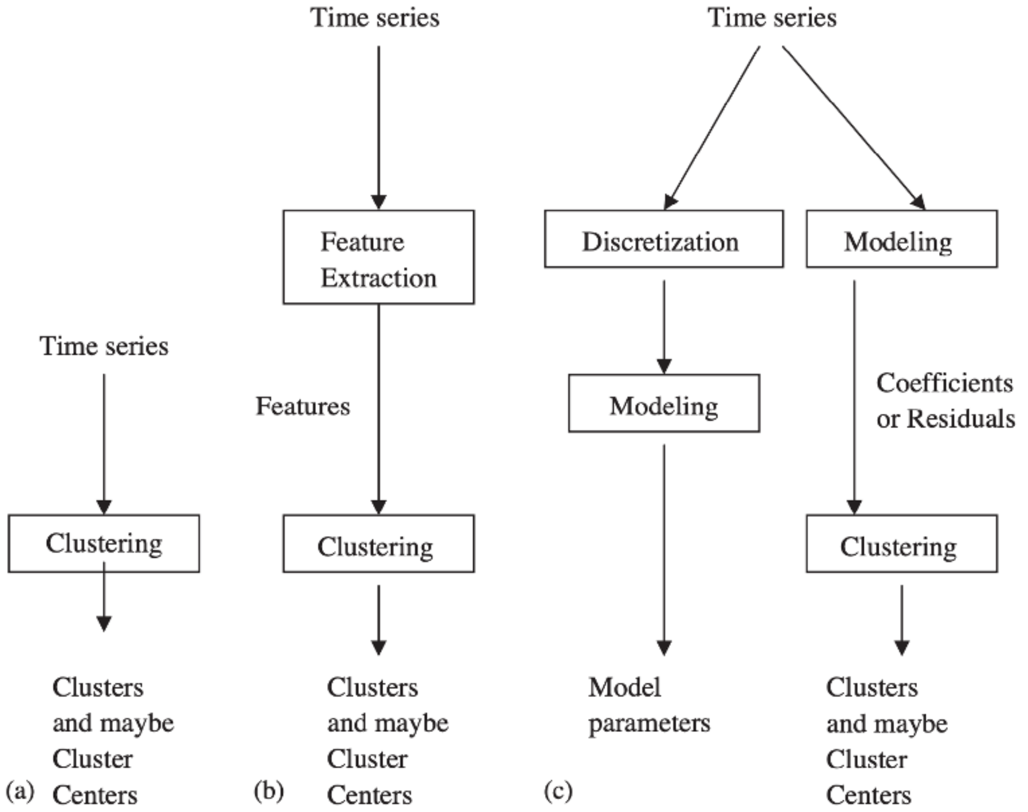


FIGURE 2.1: Three time series clustering approaches: (raw-data-based), (b) feature-based, (c) model-based. Source: Liao (2005)

Liao (2005) organized time series clustering methods into three groups depending upon the clustering (Figure 2.1) is 1) directly with the raw data, 2) indirectly with features extracted from the raw data, or 3) indirectly with models built from the raw data.

2.5.3.2 Hierarchical clustering. The hierarchical clustering (Figure 2.2) procedure begins by selecting a distance function to reflect the similarity or distance between each time series, and then iterates on N clusters (one for each state and parity). At each iteration, determine the two closest clusters using the current matrix of cluster distances between all clusters, then update the list of clusters by merging the two closest clusters, and finally, update the matrix of cluster distances appropriately. The method is repeated until all objects are grouped together.

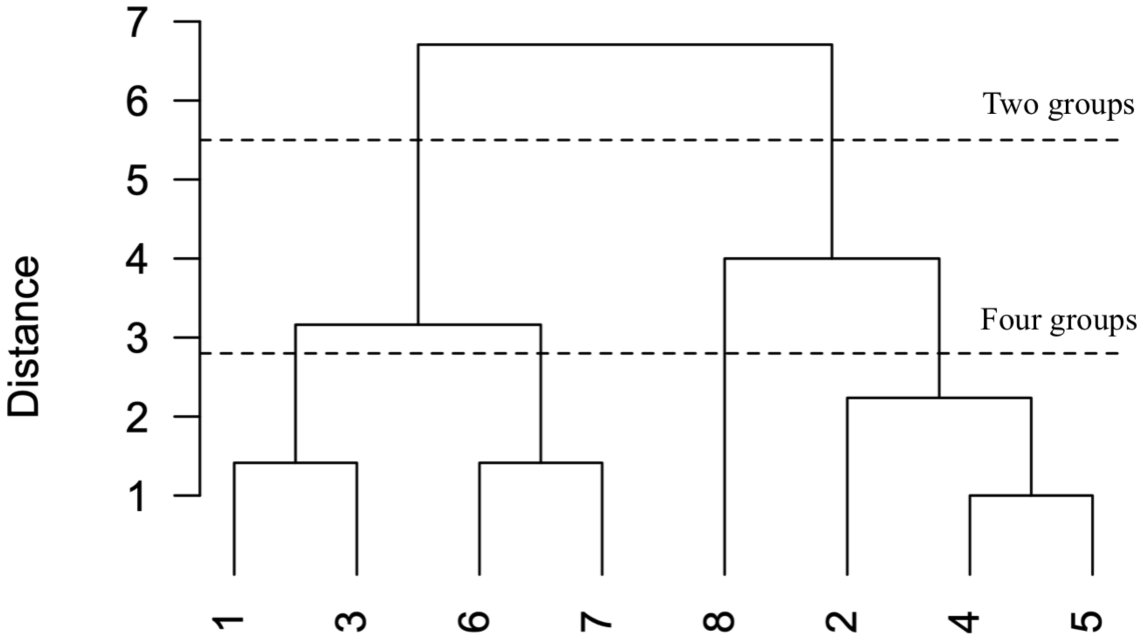
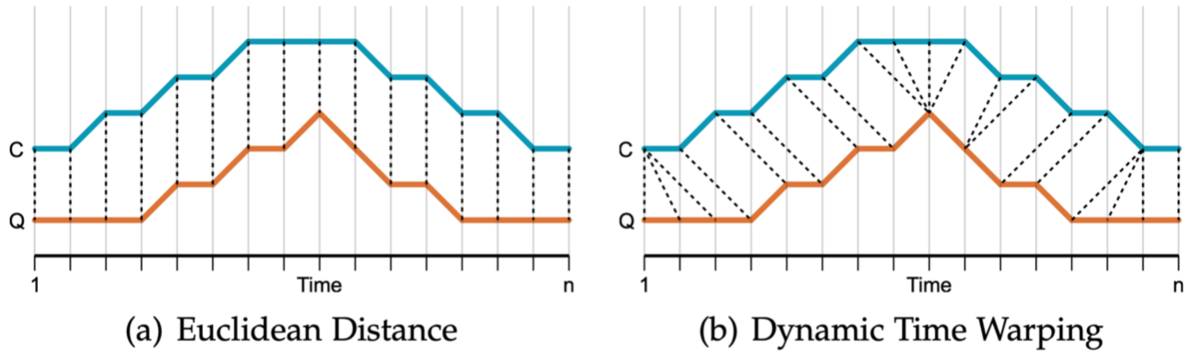


FIGURE 2.2: Dendrogram of hierarchical clustering. Example with eight objects

2.5.3.3 Distance measurement. To ascertain if observations (i.e., time series) in the data are similar, it is critical to use a metric that quantifies this similarity. Apart from Euclidean distance, a prominent approach for measuring time series is Dynamic Time Warping (DTW) (Figure 2.3). DTW compares two temporal sequences that may differ in time or speed and is commonly acknowledged as an efficient metric for time series data (Jeong et al., 2011). In general, this is a



(a) Euclidean Distance (b) Dynamic Time Warping

FIGURE 2.3: Comparison of Euclidean distance with dynamic time warping for distance measurements on times series. Source: Müller (2007)

technique that enables computers to determine the best match between two supplied sequences given specific constraints. Its benefit is that it supports flexible one-to-many mappings; in other words, DTW enables the mapping of a single point in the query sequence to several points in the candidate sequence, or vice versa. DTW was first developed for voice recognition; it was then used to solve a variety of real-world data mining challenges.

2.6 Conclusions

A literature review related to dairy herd simulation and milk production evaluation was presented within this chapter, covering different herd and farm system simulation models, Monte Carlo simulation methods, lactation curve modeling, mixed effects model, and time series decomposing and clustering. Many techniques have been proposed to apply to model dairy herd performance. We chose to develop the animal life cycle submodel as a dynamic and stochastic Monte Carlo simulation model to simulate animal life events on a daily basis as part of the whole farm system model (chapter 3). The mixed effects model approach was used in chapter 4 to investigate fixed and random effects on millions of lactation curve parameters fitted from test-day data. The times series decomposing and clustering methods, including DTW, were used in chapter 5 to analyze the relationship between calving start date and milk and milk components productivity.

References

Adediran, S. A., D. A. Ratkowsky, D. J. Donaghy, and A. E. O. Malau-Aduli (2012). Comparative evaluation of a new lactation curve model for pasture-based Holstein-Friesian dairy cows. In: *Journal of dairy science* 95.9, pp. 5344–5356.

Aghabozorgi, S., A. S. Shirkhorshidi, and T. Y. Wah (2015a). Time-series clustering–a decade review. In: *Information Systems* 53, pp. 16–38.

Ali, T. E. and L. R. Schaeffer (1987). Accounting for covariances among test day milk yields in dairy cows. In: *Canadian Journal of Animal Science* 67, pp.637–644.

Allore, H. G., H. N. Erb, L. W. Schruben, and P. A. Oltenacu (1998). A simulation of strategies to lower bulk tank somatic cell count below 500,000 per milliliter. In: *J. Dairy Sci* 81, 694–702.

ARC (1965). *The Nutrient Requirements of Farm Livestock.* No. 2, Ruminants. HMSOLondon.

Arendonk, Van, J. A. M., and A. A. Dijkhuizen (1985). Studies on the replacement policies in dairy cattle. III. Influence of variation in reproduction and production. In: *Livest. Prod Sci.* 13, 333–349.

Banks, J., J. S. Carson, B. L. Nelson, and D. M. Nicol (2010). *Discrete-event system simulation. 5th Ed.* Upper Saddle River, NJ: Prentice Hall.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2013). Fitting Linear Mixed-Effects Models Using lme4. In: *Journal of Statistical Software* 67.1. URL: `http://www.jstatsoft.org/v67/i01/`.

Bouallegue, M. and N. M'Hamdi (2020). *Lactation in Farm Animals - Biology*. Nutritional Requirements, and Modelization: Physiological Basis. DOI: `10.5772/intechopen.90253`.

Bouwman, A. F., A. H. W. Beusen, J. Griffioen, J. W. Van Groenigen, M. M. Hefting, O. Oenema, P. J. T. M. Van Puijenbroek, S. Seitzinger, C. P. Slomp, and E. Stehfest (2013). Global trends and uncertainties in terrestrial denitrification and N2O emissions. In: *Philos. Trans. R. Soc.* 368.1621, p. 20130112. DOI: `10.1098/rstb.2013.0112`.

Bouwman, A. F., K. W. Van der Hoek, B. Eickhout, and I. Soenario (2005). Exploring changes in world ruminant production systems. In: *Agric. Syst.* 84 (2), 121–153.

Brandimarte, P. (2014). Handbook In Monte Carlo Simulation: Applications In Financial Engineering, Risk Management, and Economics. 1st ed. In: *John & Sons, Hoboken, New Jersey.* Pp. 3–40.

Brody, S., Ragsdale A. C., and C. W. Turner (1924). The relation between the initial rise and the subsequent decline of milk secretion following parturition. In: *J. Gen. Physiol.* 6, pp. 541–545.

Brody, S., C. W. Turner, and Ragsdale A. C. (1923). The rate of decline of milk secretion with the advance of the period lactation. In: *J. Gen. Physiol.* 5, pp. 442–444.

Brown, V. A. (2021). An Introduction to Linear Mixed-Effects Modeling in R. In: *Advances in Methods and Practices in Psychological Science*. DOI: `10.1177/2515245920960351`.

Calsamiglia, S., S. Astiz, J. Baucells, and L. Castillejos (2018a). A stochastic dynamic model of a dairy farm to evaluate the technical and economic performance under different scenarios. In: *J Dairy Sci* 101, 7517–7530. DOI: `10.3168/jds.2017-12980`.

Campbell, S. D. and F. X. Diebold (2005). Weather Forecasting for Weather Derivatives. In: *J Am Stat Assoc* 100, 6–16. DOI: `10.1198/016214504000001051`.

Cleveland, R. B., W. S. Cleveland, J. E. McRae, and I. Terpenning (1990a). STL: A seasonal-trend decomposition. In: *J. Off. Stat* 6.1, pp. 3–73.

Cobby, J. M. and Le Du YLP. (1978). On fitting curves to lactation data. In: *AnimalProduction* 26, pp. 127–133.

Codrea, M. C., S. Højsgaard, and N. C. Friggens (2011). Differential smoothing of time-series measurements to identify disturbances in performance and quantify animal response characteristics: An example using milk yield profiles in dairy cows1. In: *J Anim Sci* 89, 3089–3098. DOI: `10.2527/jas.2010-3753`.

Cohen, R. D. H., J. P. Stevens, A. D. Moore, and J. R. Donnelly (2003). Validating and using the GrassGro decision support tool for a mixed grass/alfalfa pasture in western Canada. In: *Can. J. Animal Sci.* 83, pp. 171–182.

Cottle, D. J., M. Wallace, P. Lonergan, and A. G. Fahey (2018a). Bioeconomics of sexed semen utilization in a high-producing Holstein-Friesian dairy herd. In: *J Dairy Sci* 101, p. 4498. DOI: `10.3168/jds.2017-13172âĂŞ4512`.

Coughenour, M. B., S. J. McMaughton, and L. L. Wallace (1984). Modeling primary production of perennial graminoids - uniting physiological processes and morphometric traits. In: *Ecol Model* 23, 101–134.

CSIRO (2007). *Nutrient Requirements of Domesticated Ruminants*. Tech. rep. CSIRO Publishing.

Cullen, B. R., R. J. Eckard, M. N. Callow, I. R. Johnson, D. F. Chapman, R. P. Rawnsley, S. C. Garcia, T. White, and V. O. Snow (2008). Simulating pasture growth rates in Australian and New Zealand grazing systems. In: *Aust. J. Agric. Res* 59, 761–768.

Cunha, D. D., J. C. Pereira, F. F. Silva, O. F. Campos, J. L. Braga, and J. A. Martuscello (2010). Selection of models of lactation curves to use in milk production simulation systems. In: *Revista Brasileira de Zootecnia* 39, pp. 891–902.

Davis, J. M., T. C. Cartwright, and J. O. Sanders (1976). Alternative beef production systems for Guyana. In: *J. Anim. Sci* 43, p. 235.

De Vries, A. (2001). *Statistical process control charts applied to dairy herd reproduction*. University of Minnesota.

De Vries, A. and J. Salfer (2013). On-Farm Genomics Testing and Dairy Cattle Replacement Decisions.

Dekkers, R. (2017). *Applied Systems Theory.* DOI: `10.1007/978-3-319-57526-1`.

Delgado, C., M. Rosegrant, H. Steinfeld, S. Ehui, and C. Courbois (1999). Livestock to 2020: the next food revolution. In: *Outlook Agric.* 30, 127–29.

Dent, J. B. and J. R. Anderson (1971). Systems, management and Agriculture. In: *Systems Analysis in Agricultural Management*. Ed. by J. B. Dent and J. R. Anderson. Sydney: editors. John Wiley & Sons, pp. 3–14.

Devroye, L. (2006). Chapter 4 Nonuniform Random Variate Generation. In: *Simulation*. Ed. by S. G. Henderson and B. L. Nelson. 83–121: Elsevier.

Dhanoa MS., A (1981). note on an alternative form of the lactation model of Wood. In: *Animal Production* 32, pp. 349–351.

Dijkstra, J., S. Lopez, Dhanoa M. S. Bannink A, E. Kebreab, and et al. Odongo NE (2010b). Evaluation of a mechanistic lactation model using cow, goat and sheep data. In: *The Journal of AgriculturalScience* 148, pp. 249–262.

Dolecheck, K. A., M. W. Overton, T. B. Mark, and J. M. Bewley (2019). Use of a stochastic simulation model to estimate the cost per case of digital dermatitis, sole ulcer, and white line disease by parity group and incidence timing. In: *J Dairy Sci* 102, 715–730. DOI: `10.3168/jds.2018-14901`.

Dongre, V. B., R. S. Gandhi, A. Singh, and A. Gupta (2011). A Brief Review On Lactation Curve Models for Predicting Milk Yield And different Factors Affecting Lactation Curve In Dairy Cattle. In: *International Journal of Agriculture: Research and Review* 1, pp. 6–15.

Donnelly, J. R., M. Freer, L. Salmon, A. D. Moore, R. J. Simpson, H. Dove, and T. P. Bolger (2002). Evolution of the GRAZPLAN decision support tools and adoption by the grazing industry in temperate Australia. In: *Agric. Syst.* 74, pp. 115–139.

Donnelly, J. R., L. Salmon, R. D. H. Cohen, Z. L. Liu, and X. P. Xin (2005). Decision support for temperate grasslands: challenges and pitfalls. In: *Utilisation of Grasslands in Temperate Animal Systems, Proceedings of a Satellite Workshop of the Twentieth International Grasslands Congress*. Ed. by J. J Murphy. Ireland, pp. 105e117: Cork.

Ehrlich, J. L. (2011a). Quantifying shape of lactation curves, and benchmark curves for common dairy breeds and parities. In: *Bovine Practitioner* 45.1, p. 88.

Ettema, J. F., J. R. Thomasen, L. Hjortø, and M. Kargo (2017a). S. Østergaard, and AC. Sørensen Economic opportunities for using sexed semen and semen of beef bulls in dairy herds. In: *J Dairy Sci* 100, 4161–4171. DOI: `10.3168/jds.2016-11333`.

Ettema, J. F. and S. Østergaard. (2006). Economic decision making on prevention and control of clinical lameness in Danish dairy herds. In: *Livest. Sci* 102, 92–106.

Ettema, JF, S Østergaard, and MK Sørensen (2011). Effect of including genetic progress in milk yield on evaluating the use of sexed semen and other reproduction strategies in a dairy herd. In: *Animal* 5.12, pp. 1887–1897.

FAO (2013). *Greenhouse gas emissions from ruminant supply chains.* A Global Life Cycle Assessment. Rome, Italy.: Food and Agriculture Organisation of the United Nations.

Fischer, A. (1958). Research withWürttemberg spotted mountain cows on the shape of the lactation curve and how it may be influenced by non-genetic factors. In: *üchtungskunde* 30, pp. 296–304.

Fisher, R.A (1919). The Correlation between Relatives on the Supposition of Mendelian Inheritance. In: *Transactions of the Royal Society of Edinburgh* 52.2, pp. 399–433.

Freer, M., J. L. Davidson, J. S. Armstrong, and J. R. Donnelly (1970). Simulation of grazing systems. In: *Proceedings of the XI International Grassland Congress*. St. Lucia, Queensland, Australia: University of Queensland Press, 913–917.

Freer, M., A. D. Moore, and J. R. Donnelly (1997). GRAZPLAN: decision support systems for Australian grazing enterprises. I. Overview of the GRAZPLAN project and a description of the Met Access and Lamb Alive DSS. In: *Agric. Syst.* 54, 57–76.

Fu, T. (2011). A review on time series data mining. In: *Eng Appl Artif Intel* 24, 164–181. DOI: `10.1016/j.engappai.2010.09.007`.

Galvão, K. N., P. Federico, A. De Vries, and G. M. Schuenemann (2013). Economic comparison of reproductive programs for dairy herds using estrus detection, timed artificial insemination, or a combination. In: *J. Dairy Sci* 96, 2681–2693.

Giordano, J. O., A. S. Kalantari, P. M. Fricke, M. C. Wiltbank, and V. E. Cabrera (2012a). A daily herd Markov-chain model to study the reproductive and economic impact of reproductive programs combining timed artificial insemination and estrus detection. In: *J. Dairy Sci* 95, 5442–60.

Guo, Z. and H. H. Swalve (1995). Modeling of the lactation curve as a sub-model in the evaluation of test day record. In: *Interbull meeting* 36, pp.52–57. URL: `http://agris.fao.org/agrissearch/search.do?f=1996/SE/SE96003.xml;SE9610776`.

Herrero, M., R. H. Fawcett, and J. B. Dent (1996). *Integrating simulation models to optimize nutrition and management for dairy farms: a methodology*. The Netherlands: Wageiningen Pers., 322–326.

Herrero, M., R. H. Fawcett, and J. B. Dent (1999). Bio-economic evaluation of dairy farm management scenarios using integrated simulation and multiple-criteria models. In: *Agric. Syst.* 62, 149–168.

Herrero, M., P. Havlík, H. Valin, A. Notenbaert, M. C. Rufino, P. K. Thornton, M. Blümmel, F. Weiss, D. Grace, and M. Obersteiner (2013). Biomass use, production, feed efficiencies, and greenhouse gas emissions from global livestock systems. In: *Proc. Natl. Acad. Sci.* 110 (52), 20888–20893.

Hillier, F. S. and G. J. Lieberman (1986). *Introduction to Operations Research, 4th Ed.* Inc., San Francisco, CA, USA: Holden-Day.

Hjortø, L., J. F. Ettema, M. Kargo, and A. C. Sørensen (2015). Genomic testing interacts with reproductive surplus in reducing genetic lag and increasing economic net return. In: *J. Dairy Sci* 98, 646–658.

Holzworth, D. P., N. I. Huth, G. deVoil, E. J. Zurcher, N. I. Herrmann, G. McLean, K. Chenu, E. J. van Oosterom, V. Snow, C. Murphy, A. D. Moore, H. Brown, J. P. M. Whish, S. Verrall, J. Fainges, L. W. Bell, A. S. Peake, P. L. Poulton, Z. Hochman, and P. J. Thorburn (2014). APSIM – Evolution towards a new generation of agricultural systems simulation. In: *Environ Model Softw* 62, 327 –350. DOI: `10.1016/j.envsoft.2014.07.009`.

IADB (1975). *HerdSIM simulation model: user manual*. Vol. Project Analysis Paper No. 2. Washington, D.C.: Economic and Social Development Department, Inter-American Development Bank.

ILCA (1978). *Mathematical modeling of livestock production systems: application of the Texas A & M University beef cattle production model to Botswana*. International Livestock Centre for Africa. Addis Ababa, Ethiopia: Systems Study No. 1.

IPCC, J.T. Houghton, G.J. Jenkins, and J.J. (Eds.) Ephraums (1990). Climate change: the IPCC scientific assessment. In: New York, NY, USA: Cambridge University Press, Cambridge, Great Britain, p. 410.

Jenkins, T. G. and C. L. Ferrell (1984). note on lactation curves of crossbred cows. In: *Animal Production* 39, pp. 479–482.

Jeong, Young-S., M. K. Jeong, and O. A. Omitaomu (2011). Weighted dynamic time warping for time series classification. In: *Pattern Recognition* 44.9, pp. 2231–2240.

Johnson, I. and J. H. M. Thornley (1983). Vegetative crop growth model incorporating leaf area expansion and senescence and applied to grass. In: *Plant Cell Environ* 6, 721–729.

Johnson, I. R., D. F. Chapman, V. O. Snow, R. J. Eckard, A. J. Parsons, M. G. Lambert, and B. R. Cullen (2008a). DairyMod and EcoMod: biophysical pasture-simulation models for Australia and New Zealand. In: *Anim Prod Sci* 48, 621 –631. DOI: `10.1071/ea07133`.

Johnson, I. R., J. France, J. H. M. Thornley, M. J. Bell, and R. J. Eckard (2012). A generic model of growth, energy metabolism, and body composition for cattle and sheep. In: *J Anim Sci* 90, 4741–4751. DOI: `10.2527/jas.2011-5053`.

Johnson, I. R., G. M. Lodge, and R. E. White (2003). The Sustainable Grazing Systems Pasture Model: description, philosophy and application to the SGS National Experiment. In: *Anim Prod Sci* 43, pp. 711–728. DOI: `10.1071/ea02213`.

Jones, J. W., J. M. Antle, B. O. Basso, K. J. Boote, R. T. Conant, I. Foster, H. C. J. Godfray, M. Herrero, R. E. Howitt, S. Janssen, B. A. Keating, R. Munoz-Carpena, C. H. Porter, C. Rosenzweig, and T. R. Wheeler (2017). Brief history of agricultural systems modeling. In: *Agric. Syst.* 155, pp. 240–254. DOI: `10.1016/j.agsy.2016.05.014`.

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

Kaniyamattam, K., M. A. Elzo, J. B. Cole, and A. De Vries (2016). Stochastic dynamic simulation modeling including multitrait genetics to estimate genetic, technical, and financial consequences of dairy farm reproduction and selection strategies. In: *J Dairy Sci* 99, 8187–202. DOI: `10.3168/jds.2016-11136`.

Keating, B. A., P. S. Carberry, G. L. Hammer, M. E. Probert, M. J. Robertson, D. Holzworth, N. I. Huth, J. N. G. Hargreaves, H. Meinke, Z. Hochman, G. McLean, K. Verburg, V. Snow, J. P. Dimes, M. Silburn, E. Wang, S. Brown, K. L. Bristow, S. Asseng, S. Chapman, R. L. McCown, D. M. Freebairn, and C. J. Smith (2003). An overview of APSIM, a model designed for farming systems simulation. In: *Eur. J. Agron.* 18, 267–288.

Keating, B. A., D. C. Godwin, and J. M. Watiki (1991). Optimization of nitrogen inputs under climatic risk. In: *Muchow*. Ed. by Bellamy and R. C. Wallingford, UK, pp. 329–357: Climatic Risk in Crop Production — Models, Management for the Semi-arid Tropics, and Sub-tropics. CAB International.

Keig, G. and J. R. McAlpine (1969). *WATBAL: A computer system for the estimation and analysis of soil moisture regimes from simple climatic data*. Canberra: Tech. Memo. 69/9 CSIRO Division of Land Use Research.

Kikuhara, K., H. Kumagai, and H. Hirooka (2009). Development and Evaluation of a Simulation Model for Dairy Cattle Production Systems Integrated with Forage Crop Production. In: *Asian Austral J Anim* 22, 57–71. DOI: `10.5713/ajas.2009.80098`.

Konandreas, P. A. and F. M. Anderson (1982). *Cattle herd dynamics: An integer and stochastic model for evaluating production alternatives. ILCA Research Report No. 2*. Addis Ababa, Ethiopia: International Livestock Centre for Africa.

Korkmaz, M., F. Üçkardeș, and A. Kaygisiz (2011). Comparison of Wood, Gaines, Parabolic, Hayashi, Dhanno and polynomial models for lactation season curve of Simmental cows. In: *JAPS, Journal of Animal and Plant Sciences* 21.3, pp. 448–458.

Kristensen, A. R., E. Jørgensen, and N. Toft (2006). *Herd management science. Preliminar. University*. The Royal Veterinary and Agricultural.

Kristensen, A. R. and I. Thysen (1991). Ranking of dairy cows for replacement - alternative methods tested by stochastic simulation. In: *ACTA Agric. Scand.* 41, 295–303.

Kudahl, A. B., S. Østergaard, J. T. Sørensen, and S. S. Nielsen (2007). A stochastic model simulating paratuberculosis in a dairy herd. In: *Prev*. Vet. Med. 78: 97–117.

Kumar, M., N.R. Patel, and J. Woo (2002). Clustering seasonality patterns in the presence of errors. In: *Proceedings of KDD 2002*.

Law, A. M. and D. M. Kelton (2007). *Simulation Modeling and Analysis. 4th ed.* New York: McGraw-Hill Higher Education.

Le Gal, P. Y., P. Dugué, G. Faure, and S. Novak (2011). How does research address the design of innovative agricultural production systems at the farm level? A review. In: *Agricultural Systems* 104, 714–728. DOI: `10.1016/j.agsy.2011.07.007`.

Lee, M., S. Lee, J. Park, and S. Seo (2020). Clustering and Characterization of the Lactation Curves of Dairy Cows Using K-Medoids Clustering Algorithm. In: *Animals* 10.8, p. 1348.

Liao, T. W. (2005). Clustering of time series data — a survey. In: *Pattern recognition* 38.11, pp. 1857–1874.

Macciotta, N. P. P., Dimauro C, Rassu S. P. G., R. Steri, and G. Pulina (2011a). The mathematical description of lactation curves in dairy cattle. In: *Italian Journal ofAnimal Science* 10.4, pp. 213–223.

Macciotta, N. P. P., A. Cappio-Borlino, and G. Pulina (2000). Time Series Autoregressive Integrated Moving Average Modeling of Test-Day Milk Yields of Dairy Ewes. In: *J Dairy Sci* 83, 1094–1103. DOI: `10.3168/jds.s0022-0302(00)74974-5`.

Macciotta, N. P. P., A. Nudda, A. Cappio Borlino, G. Pulina, and P. Brandano (1999). Time series analysis of milk yield and somatic cell count of Sarda dairy ewe. In: *Proceedings of the ASPA Congress-Recent Progress in Animal Production Science (Italy)*.

McCown, R. L., G. L. Hammer, J. N. G. Hargreaves, D. P. Holzworth, and D. M. Freebairn (1996). APSIM: a novel software system for model development, model testing and simulation in agricultural systems research. In: *Agric. Syst.* 50, 255–271.

Metropolis, N. and S. Ulam (n.d.). The Monte Carlo Method. 1949. In: *Journal of the American statistical association* 44 (), 335–341.

Mol, R. M. de, A. Keen, G. H. Kroeze, and J. M. F. H. Achten (1999). Description of a detection model for oestrus and diseases in dairy cattle based on time series analysis combined with a Kalman filter. In: *Comput Electron Agr* 22, 171–185. DOI: `10.1016/s0168-1699(99)00016-2`.

Moore, A. D., D. P. Holzworth, N. I. Herrmann, H. E. Brown, P. G. de Voil, V. O. Snow, E. J. Zurcher, and N. I. Huth (2014). Modeling the manager: Representing rule-based management in farming systems simulation models. In: *Environ Mo Softw* 62, 399–410. DOI: `10.1016/j.envsoft.2014.09.001`.

Moore, A. D., D. P. Holzworth, N. I. Herrmann, N. I. Huth, and M. J. Robertson (2007). The Common Modeling Protocol: a hierarchical framework for simulation of agricultural and environmental systems. In: *Agric. Syst.* 95, pp. 37–48.

Morant, S. V. and A. Gnanasakthy (1989). A new approach to the mathematical formulation of lactation curves. In: *Animal Science* 49.2, pp. 151–162.

Morignat, E., E. Gay, J.-L. Vinard, D. Calavas, and V. Hénaux (2015). Quantifying the influence of ambient temperature on dairy and beef cattle mortality in France from a time-series analysis. In: *Environ Res* 140, 524–534. DOI: `10.1016/j.envres.2015.05.001`.

Müller, M. (2007). Dynamic time warping. In: *Information retrieval for music and motion*, pp. 69–84.

Murphy, M. D., M. J. O'Mahony, L. Shalloo, P. French, and J. Upton (2014b). Comparison of modeling techniques for milk-production forecasting. In: *J Dairy Sci* 97, 3352–3363. DOI: `10.3168/jds.2013-7451`.

Nielsen, H. M., A. Groen, J. Pedersen, and P. Berg (2004). Stochastic simulation of economic values and their standard deviations for production and functional traits in dairy cattle under current and future danish production circumstances. Acta Agric. In: *Scand. Sect A- Anim. Sci.* 54, 113–126.

NRC (1945). *Nutrient Requirements of Dairy Cattle.* 7th rev. ed. Washington, DC.: The National Academies Press.

NRC (2001a). *Nutrient Requirements of Dairy Cattle.* 7th rev. ed. Washington, DC.: The National Academies Press.

Nørgaard, N. H., K. M. Lind, and J. F. Agger (1999). Cointegration analysis used in a study of dairy-cow mortality. In: *Prev Vet Med* 42, 99–119. DOI: `10.1016/s0167-5877(99)00068-9`.

Østergaard, S., M. G. G. Chagunda, N. C. Friggens, T. W. Bennedsgaard, and I. C. Klaas. (2005). A stochastic model simulating pathogen-specific mastitis control in a dairy herd. In: *Journal of dairy science* 88.12, pp. 4243–4257.

Østergaard, S, J Hindhede, AR Kristensen, et al. (1996). Technical and economic effects of feeding dairy herds TMR in one group to multiple groups under different herd and management characteristics estimated by stochastic simulation. In: *Livest. Prod. Sci.* 45, pp. 23–33.

Østergaard, S, JT Sørensen, and H Houe (2003). A stochastic model simulating milk fever in a dairy herd. In: *Preventive veterinary medicine* 58.3-4, pp. 125–143.

Østergaard, S., J. T. Sørensen, and A. R. Kristensen (2000). A stochastic model simulating the feeding health-production complex in a dairy herd. In: *J. Dairy Sci* 83, 721–33.

PBL (2013). *Trends in global CO2 emissions: 2013 report*. PBL Netherlands Environmental Assessment Agency Institute for Environment and Sustainability (IES) of the European Commission's Joint Research Centre (JRC) The Hauge.

Peng, R. D, F. Dominici, and T. A. Louis (2006). Model choice in time series studies of air pollution and mortality. In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 169.2, pp. 179–203.

Pfeiffer, J., M. Gandorfer, and J. F. Ettema (2020). Evaluation of activity meters for estrus detection: A stochastic bioeconomic modeling approach. In: *J Dairy Sci* 103, 492–506. DOI: `10.3168/jds.2019-17063`.

Plaizier, J. C., G. J. King, J. C. Dekkers, and K. Lissemore (1997). Estimation of economic values of indices for reproductive performance in dairy herds using computer simulation. In: *J. Dairy Sci* 80, 2775–83.

Quinn, N. (2005). Modeling lactation and liveweight curves in Irish dairy cows. Dublin City University. School of Computing. In: *Available at: [Accessed April* 24, p. 2017. URL: `http://doras.dcu.ie/18167/`.

Rao, M. K. (1977). *Studies on the lactation curves of dairy cattle*. Chandigarh.

Reuven, Y. R. and D. P. Kroese (2017). Simulation and the Monte Carlo Method. In: *Wiley Series in Probability and Statistics*, pp. 91–106. DOI: `10.1002/9781118631980.ch3`.

Ritchie, J. T. and S. Otter (1984). Description and performance of CERES-wheat: a user-oriented wheat yield model. In: *Wheat Yield Project*. Springfield, Missouri, pp. 159–175: ARS-38. National Technical Information Service.

Rotz, C. A., D. R. Buckmaster, and J. W. Comerford (2005). A beef herd model for simulating feed intake, animal performance, and manure excretion in farm systems1. In: *J Anim Sci* 83, 231–242. DOI: `10.2527/2005.831231x`.

Rotz, C. A., M. S. Corson, D. S. Chianese, F. Montes, S. D. Hafner, H. F. Bonifacio, and C. U. Coiner (2018). *The Integrated Farm System Model reference manual, V 4.4*. Tech. rep. USDA-ARS.

Rotz, C. A., B. J. Isenberg, K. R. Stackhouse-Lawson, and E. J. Pollak (2013b). A simulation-based approach for evaluating and comparing the environmental footprints of beef production systems. In: *J Anim Sci* 91, 5427–5437. DOI: `10.2527/jas.2013-6506`.

Rotz, C. A., D. R. Mertens, D. R. Buckmaster, M. S. Allen, and J. H. Harrison (1999a). A Dairy Herd Model for Use in Whole Farm Simulations. In: *J Dairy Sci* 82, 2826–2840. DOI: `10.3168/jds.s0022-0302(99)75541-4`.

Rowlands, G. J. and S.Russel AM Slucey (1982). Comparison of different models of the lactation curve in dairy cattle. In: *AnimalProduction* 35, pp. 135–144.

Rubinstein, R. Y. and D. P. Kroese (2017a). *Simulation and the Monte Carlo Method*. Ser Probab Statistics 91–106: Wiley. DOI: `10.1002/9781118631980.ch3`.

Russell, J. B., J. D. O'Connor, D. G. Fox, P. J. Van Soest, and C. J. Sniffen (1992). A net carbohydrate and protein system for evaluating cattle diets: I Ruminal fermentation. In: *J. Anim. Sci.* 70, 3551–3561.

Sanders, J. O. and T. C. Cartwright (1979). A general cattle production systems model. 2. In: *Procedures used for simulating animal performance* Agric. Syst. 4, 289 –309.

Schils, R. L. M., M. H. A. de Haan, J. G. A. Hemmer, A. van den Pol-van Dasselaar, J. A. de Boer, A. G. Evers, G. Holshof, J. C. van Middelkoop, and R. L. G. Zom (2007b). DairyWise, A Whole-Farm Dairy Model. In: *J Dairy Sci* 90, pp. 5334–5346. DOI: `10.3168/jds.2006-842`.

Sikka LC., A (1950). study of lactation as affected by heredity and environment. In: *The Journal of Dairy Research* 17, pp. 231–252.

Silvestre, A. M. and Colaco J. Petim-Batista F (2006). The accuracy of seven mathematical functions in modelingdairy cattle lactation curves based ontest-day records from varying sample schemes. In: *Journal of Dairy Science* 89, pp. 1813–1821.

Singh, B and PN Bhat (1978). Models of lactation curves for Hariana cattle. In: *Indian journal of animal sciences*.

Slatyer, R. O. (1960). Agricultural climatology of the Yass valley. In: *CSIRO Aust. Div. Land Res. Reg. Surv. Tech.*

Sorensen, J. T. (1998). Modeling and simulation in applied livestock production science. In: *Agricultural Systems modeling and Simulation*. Ed. by R. M. Peart and R. B. Curry. New York, USA, USA. 475–494: editors. Marcel Dekker.

Sorensen, J. T., E. S. Kristensen, and I. Thysen (1992). A stochastic model simulating the dairy herd on a PC. In: *Agricultural Systems* 39, pp. 177–200.

St-Pierre, N. R. and C. S. Thraen (1999). Animal Grouping Strategies, Sources of Variation, and Economic Factors Affecting Nutrient Balance on Dairy Farms. In: *J. Anim. Sci* 77, 72–83.

Taye, B. A., A. A. Alene, A. K. Nega, and B. G. Yirsaw (2020). Time series analysis of cow milk production at Andassa dairy farm, West Gojam Zone, Amhara Region, Ethiopia. In: *Model Earth Syst Environ* 1. DOI: `10.1007/s40808-020-00946-z`.

Val-Arreola, D., E. Kebreab, J. Dijkstra, and J. France (2004). Study of the lactation curve in dairy cattle on farms in central Mexico. In: *Journal of dairy science* 87.11, pp. 3789–3799.

Van Bavel, C. H. M. (1953). A drought criterion and its application in evaluating drought incidence and hazard. In: *J. Agron.* 4, 167–172.

Vayssières, J., F. Guerrin, J.-M. Paillat, and P. Lecomte (2009). GAMEDE: A global activity model for evaluating the sustainability of dairy enterprises Part I – Whole-farm dynamic model. In: *Agr Syst* 101, 128–138. DOI: `10.1016/j.agsy.2009.05.001`.

Velten, K. (2009). *Mathematical Models and simulation*. Wiley.

Vujicic, I. and B. Bacic (1961). Factors Affecting the Shape of the Lactation Curve and Its Mathematical Description. In: *MS Thesis*. University of Wisconsin, Madison.

Wagner, A. K., S. B. Soumerai, F. Zhang, and D. Ross-Degnan (2002). Segmented regression analysis of interrupted time series studies in medication use research. In: *J Clin Pharm Ther* 27, 299–309. DOI: `10.1046/j.1365-2710.2002.00430..`

Williams, J. R., K. G. Renard, and P. T. Dyke (1983). EPIC: a new method for assessing erosion's effect on soil productivity. In: *J. Soil Water Conserv.* 38 (5), 381–383.

Wilmink, J. B. M. (1987). Adjustment of lactation yield for age at calving in relation to level of production. In: *Livestock Production Science* 16. Available at: [Accessed 12, 2014], pp.321–334. URL: `http://www.sciencedirect.com/science/article/pii/0301622687900029`.

Wit, C. T. de (1958). Transpiration and crop yields. In: *Volume 64 of Agricultural research report / Netherlands Volume 59 of Mededeling (Instituut voor Biologisch en Scheikundig Onderzoek va Landbouwgewasses) Verslagen van landbouwkundige onderzoekingen.* Institute of Biological and Chemical Research on Field Crops and Herbage.

Wood, P. D. P. (1967a). Algebraic Model of the Lactation Curve in Cattle. In: *Nature* 216, 164–165. DOI: `10.1038/216164a0`.

Wood, P. D. P. (1969). Factors affecting the shape of the lactation curve in cattle. In: *Anim Sci* 11, 307–316. DOI: `10.1017/s0003356100026945`.

Chapter 3

A Stochastic Animal Life Cycle Simulation Submodel for a Whole Dairy Farm System Model

3.1 Interpretive Summary

A Stochastic Animal Life Cycle Simulation Submodel for a Whole Dairy Farm System Model. By *Li, Cabrera, and Reed.* A whole dairy herd simulation model has the potential to help researchers and farmers virtually test management strategies before taking operational actions. An open-source, highly flexible, probabilistic simulation model populated with on-farm customized inputs, accommodated to management practices, and connected with other farm components empowers decision-makers to anticipate management strategies' outcomes in the short, medium, and long term. This paper introduces the animal life cycle module as part of the Ruminant Farm Systems model.

M. Li, * V. E. Cabrera, *1 K. F. Reed, †

* Department of Dairy Science, University of Wisconsin-Madison, Madison WI 53705

† Department of Animal Science, Cornell University, Ithaca, NY 14850

3.2 Abstract

We constructed a Monte Carlo simulation of dairy cattle life cycles and integrated it with submodels representing other farm components to create the Ruminant Farm Systems model (RuFaS), a holistic dairy farm simulation model. The animal life cycle submodel simulates the growth, reproduction, production, and culling of individual dairy cattle animals and integrates the outcomes of individual animals to represent daily herd dynamics. The herd dynamics are then provided and transferred to other submodels for further modeling and output generation. The submodel was designed to have a highly extensible structure for animal and herd simulation, allowing for future modification and expansion. Additionally, when the animal life cycle submodel is combined with other farm components in RuFaS, such as animal nutrition, manure management, soil nutrient cycling, and crop growth and storage, it provides a holistic approach for evaluating the overall economic and productive performance of dairy farm systems, as well as their environmental impact. The animal life cycle submodel is demonstrated in this article through a case study comparing the economic and reproductive consequences of employing simply conventional semen, sexed and beef semen, or a combination of sexed, conventional, and beef semen. The case study results indicate that transitioning from conventional semen to a combination of sexed and beef semen is economically advantageous, since the improvements in milk and calf revenue outweigh the additional expenses of breeding and semen. Selecting lower productive and reproductive performance cows to beef semen increased the herd's relative productivity level, bringing higher milk income. Also, the higher price of crossbred calves (e.g., \$225) compared to Holsteins calves (e.g., \$57.5 for males and \$45 for females) increased the profitability of the beef semen strategy. In herds with good reproductive performance, employing conventional, sexed, and beef semen resulted in comparable net returns to using only sexed and beef semen. This occurred because breeding lower fertility cows with conventional semen resulted in a higher pregnancy rate, shorter generation interval, and more calves, all of which offset the negative effect of lower production and fewer crossbred calves when conventional semen was incorporated into a sexed and beef semen strategy. Whereas utilizing sexed and beef semen had the highest net return when the herd's reproductive performance was moderate.

Keywords: dairy farm simulation, RuFaS, beef semen

3.3 Introduction

Using integrated simulation models to evaluate farm management decisions over time with farm-specific data is more feasible and cost-effective than field trials (Rotz et al., 2015; Calsamiglia et al., 2018). Integrated farm simulation models can address the entwinement of dairy farms' multiple objectives pertaining to energy efficiency, optimal production, cost reduction, care for animals, and environmental stewardship. Numerous existing simulation models emphasized specific management areas such as reproduction (Giordano et al., 2012), nutritional grouping strategies (Kalantari et al., 2016), replacement policies (Cabrera, 2012), diet formulation (Vaillant and Baldinger, 2016), soil and water (Arnold et al., 1998), or grazing herds (Baudracco et al., 2012). Few, however, explored the critical relationships between these many components. Concentrating on discrete components of farm systems restricted their application and obscured downstream consequences at the whole-farm level. Additionally, the majority of existing integrated models (Schils et al., 2007; Johnson et al., 2008; Rotz et al., 2015) do not modify model inputs and structure systematically to reflect specific farm features or researcher goals. An integrated simulation model based on contemporary, modular computing approaches is required (Kebreab et al., 2019, Hansen et al., 2021).

A national group of researchers from several universities and research institutions in the United States is developing the Ruminant Farm System model (RuFaS) as a framework for modeling integrated dairy farm systems. The model simulates soil nutrient cycling, crop growth and harvest, feed storage, ration formulation, animal life cycles events, herd management, and manure management over a period of several weather years to investigate management decisions at the farm level (Kebreab et al., 2019). The RuFaS model is written in Python (Rossum and Drake, 2009), a widely used object-oriented programming language renowned for its ease of use and human-language-like syntax. Python was rated as the most popular programming language in 2019 for engineering and applied sciences (Cass, 2019) and is not platform-specific. Its design and popularity ensure the software's continued development and maintenance, as well as the language's widespread acceptance among researchers and programmers. Apart from the programming language, RuFaS employs an agile development methodology, an iterative, adaptive workflow strategy (Winter, 2014), to ensure that the model meets the needs of our diverse stakeholders. It adheres to rigorous documentation standards in order to facilitate intelligent application and future adaptability and expansion. The model code and documentation will be entirely open source to ensure transparency, adaptability, and accessibility for other interested researchers, developers, private companies, policymakers, and educators.

All RuFaS submodels are being developed in parallel. A detailed animal life cycle submodel is critical for establishing information flow in animal management and production representation. In particular, the model structure that estimates the number of animals and their daily life events is required to estimate nutrient requirements and milk production for formulating diets, allocating animals to pens, and predicting manure excretion and composition. Additionally, consistent with modular development practices, the animal life cycle submodel can be used independently to evaluate herd management decisions. In this study, we describe the development of the animal life cycle submodel and its application to assess the use of the combination of sexed and beef semen with different reproductive protocols and selection criteria.

The use of sexed semen in the US dairy industry increased from 8.78% in 2015 to 16% at the start of 2019 (Nehls, 2019). The use of beef semen on dairies, which is typically combined with sexed semen to balance heifers needed for the herd, also increased from 1.84% in 2015 to 16% in 2019 (Nehls, 2019). The combination of sexed and beef semen can maintain herd size with superior replacements (Weigel, 2004), bring additional income from calf trade, and lead to shorter generation intervals (Cottle et al., 2018).

Therefore, the objectives of this study are twofold:

- develop a stochastic model to simulate the growth, reproductive events, expected milk production, and herd dynamics of a dairy cow herd, which will function as the life cycle.
- apply the life cycle submodel to assess the economic and reproductive performance of using different types of semen selection strategies.

3.4 Materials and Methods

Although self-contained, the animal life cycle submodel is built within a whole-farm simulation model, the RuFaS (Li et al., 2018; Kebreab et al., 2019, Figure 3.1). The animal life-cycle submodel takes information from user inputs and passes information of simulated animals such as herd structure, animal life events, and expected production parameters to other modules and model outputs (Figure 3.2). The animal life cycle submodel simulates growth, production, disease incidence, culling, and reproduction according to breed, reproduction protocols, production curves, and culling risk.

3.4.1 Simulation Framework

We developed a Monte Carlo simulation model of the dairy cattle life cycle in Python to achieve the simulation objectives. The animal life cycle uses a Monte Carlo process (Rubinstein and Kroese, 2017) to simulate individual animals and herd dynamics by representing the occurrence of life events such as weaning, breeding, calving, milking, and drying in each animal as a

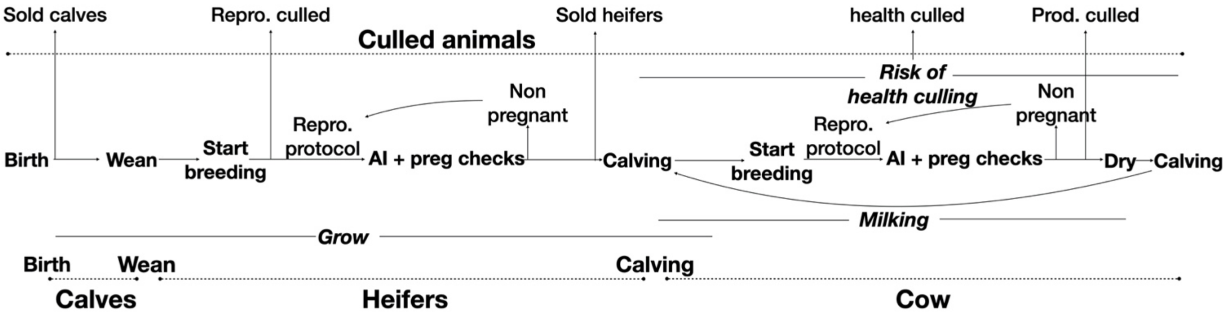


FIGURE 3.1: Flowchart of female dairy cattle animal life cycle simulation model including critical events. There are four major animal stages: calves, heifers, cows, and culled animals

stochastic process. To allow a high degree of customization, the user could modify all parameters through an input file. We selected parameter default values from scientific publications and industry standards to reduce the burden of information required from the user but allowed for adjustment. The model simulates individual animal life events on a daily time-step from calving until removal from the herd. Each animal's life events were scheduled according to (1) animal biological characteristics such as breed and parity, (2) herd-level management practices such as semen type and reproductive protocol choice. Each day, events were triggered according to stochastic schedules that were determined through random draws from probability distributions constructed from user inputs and default values.

We used a two-step Monte Carlo method to determine the life of each animal: (1) compare a random draw from the uniform distribution $U(0,1)$ to the probability of an event occurring; and (2) select a random draw from a known distribution of animal attributes and assign the value to the instantiation of an individual animal. Probability distributions and the corresponding events are given in Tables 3.1, 3.2, and 3.3. The type of distributions used were: (1) Normal Gaussian distributions: $N(\mu,\sigma)$ where μ is the distribution mean, σ is the standard deviation, and random draws are limited within 2 standard deviations. (2) Empirical distributions: $F_n(X) = \frac{1}{n}\sum_{i=1}^{n} 1_{x_i \leq x}$,

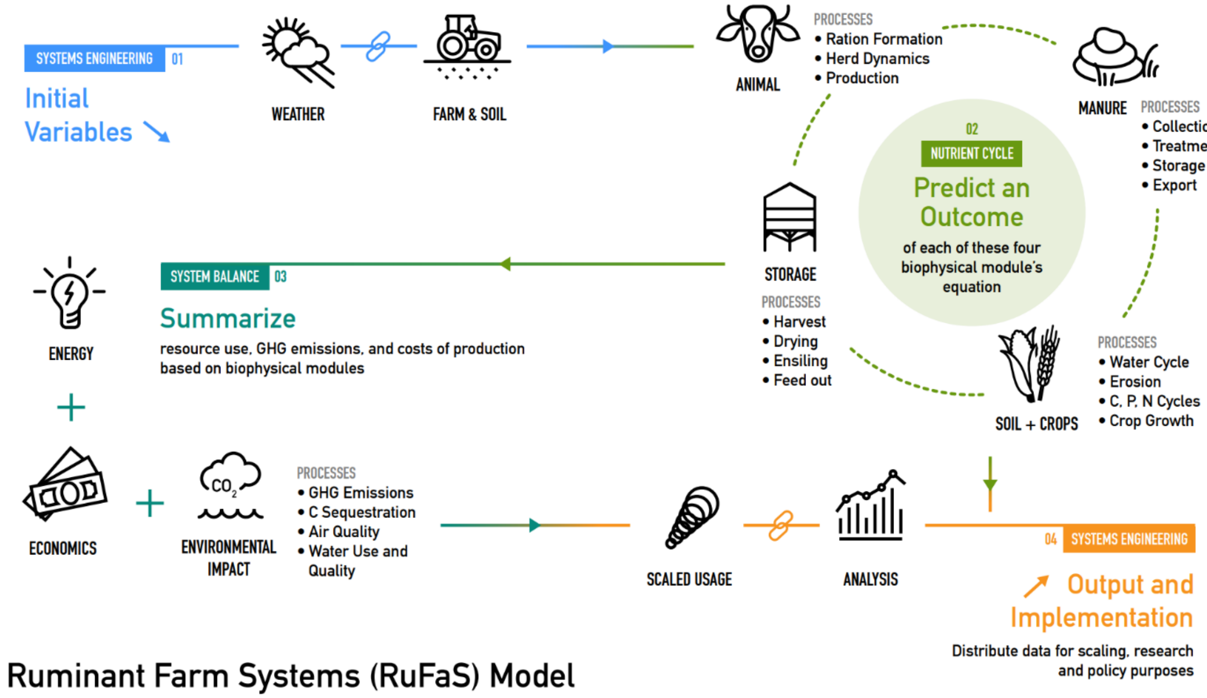


FIGURE 3.2: Model schematic of Ruminant Farm System (RuFaS) model, including the four biophysical modules of Animal, Manure, Soil, and Crop, and storage, connects to represent a whole farm system (from RuFaS executive summary). The life cycle submodel introduced by this study is a submodel of the Animal module of RuFaS that deals with herd dynamics.

where x_i is the observation from the sample and $1_{x_1 \leq x}$ is the indicator of occurring. For example, the calving day for an individual cow is scheduled on the day of successful conception by a random draw from a normal distribution for gestation length. When an event occurs, we append the record of that event, which consists of the simulation date, animal age, and event name, to a list which is an animal class attribute. We aggregate individual animal information into herd statistics and update animal and herd information on a daily basis.

3.4.2 Herd Structure

There are 6 classes of animals in the simulation process (Figure 3.1): calves, heifer I, heifer

II, heifer III, cows, and culled animals. Each class simulates different biological and management processes appropriate for the animal's life stage. Individual animals within the herd are represented by instances of an animal class and are defined by attributes specific to each class. Common attributes among classes include a unique identification code, breed gender, birth weight, target mature BW, current BW, and a list of life events. The choice of breed causes the model to select a set of breed-related parameters that include milk production curve parameters and birth BW. We currently include parameter sets for the 2 most common US breeds: Holstein and Jersey; however, users could define their own breed by defining separate sets of parameters and associating them with their breeds' names.

When an animal graduates from one class to another, these attributes are carried over to initiate a new instance of the elder animal class that contains the characteristics and history of the same animal.

Calves. An instantiation of the Calf class begins when a calf is born and ends on the user-defined weaning day. At birth, 4 animal attributes are selected according to their Monte Carlo process: gender, birth weight, target mature BW, and whether or not the calf is sold (Table 3.2). The semen type used on the dam determines gender by comparing a draw from $U(0,1)$ with the proportion of female calves produced by that semen. A random draw from a normal distribution according to the breed of the calf (Dhakal et al., 2013) defines a calf's birth weight. The target mature BW is generated by drawing from a normal distribution (Schubert et al., 2018), which is also parameterized according to the breed or user input. The calf is either sold immediately or kept on-farm. Male calves and calves produced by beef semen are sold once born. Depending on the farm's goals and needs, Some female calves could be sold. Sales are controlled by manipulating the female calf keeping rate. Processes simulated in the Calf class include growth. Methods for simulating growth for all classes are described below.

Heifers. The first heifer class (Heifer I) starts after weaning and ends when the animal begins her reproductive cycles. The second heifer class (Heifer II) begins when the animal enters a reproduction protocol and ends when she reaches a user-defined day in pregnancy that identifies the beginning of the pre-fresh period. The third heifer class (Heifer III) begins on the first day of the pre-fresh period and ends on the day of her first calving. In the heifer I class, growth is the only life-cycle process simulated. Growth, breeding, conception, and pregnancy development are simulated in the Heifer II class. And growth, pregnancy development, and calving are simulated in the Heifer III class. An instance of a Heifer I animal becomes a Heifer II animal and enters a breeding protocol when the first of the two following events occur: the animal's age reaches a user-defined breeding time (e.g., 400 d), or the animal's BW reaches a user-defined breeding weight (e.g., 550 kg). The breeding process follows the reproduction protocol assigned to the heifer and is simulated as described below. Each breeding protocol includes a conception rate, which determines the probability of the breeding program's success and the start of pregnancy. Pregnancy diagnosis and pregnancy reconfirmation diagnoses follow the insemination. If a heifer is not pregnant at a certain age (e.g., 650 d), the heifer is culled due to reproductive failure. At a user-defined day in pregnancy (e.g., 1 month before calving), the heifer either enters the Heifer III class or is sold as a replacement.

Cows. The Cow class starts the day after the first calving, at the beginning of the first lactation. Lactation is initiated on the first day. The Cow class continues until she is removed from the herd. The life cycle processes in the Cow class include growth until the start of the 3rd parity or attainment of the target mature BW. The BW changes due to lactation, breeding, conception, pregnancy development, and calving. Once the cow reaches the end of the voluntary waiting period in each lactation, the assigned breeding protocol, subsequent inseminations, and pregnancy diagnoses follow as described below. After a successful pregnancy and calving, the next lactation and reproduction program continues.

3.4.3 BW Changes

The growth section is driven by mature BW, which is randomly selected at birth from a normal distribution. Growth rates are calculated as ADG following the recommendation of the target growth system from the Nutrient Requirements for Dairy Cattle (NRC, 2001). Bodyweight changes due to pregnancy and lactation were adapted from the NRC (2001) and Korver et al. (1985). The growth rates, conceptus weight increase, and change due to lactation serve are summed up as the target BW change for the ration formulation submodel.

For calves, BW change is modeled as a linear daily increase. The ADG is determined by setting a target weight at weaning as twice the birth weight (DCHA, 2012; Soberon et al., 2012)

$$calf\,ADG(kg/d) = \frac{birth\ weight * 2 - birth\ weight}{age\ at\ weaning - 0} \tag{3.1}$$

For heifers, BW change due to growth is also modeled as a linear daily increase based on BW targets at key life stages (NRC, 2001)

$$heifer\,ADG(kg/d) = \begin{cases} \frac{0.55*MBW-BW}{target\ age\ at\ first\ pregnancy\ -\ days\ born} & \text{non-pregnant heifer} \\ \frac{0.82*MBW-BW}{gestation\ length\ -DIP} & \text{pregnant heifer} \end{cases} \tag{3.2}$$

Where MBW = Mature BW and DIP = days in pregnancy. After conception, conceptus growth is added to the BW change for pregnant heifers (Korver et al., 1985).

$$conceptus\,growth(kg/d) = \begin{cases} 0 & \text{if } DIP < 50 \\ 3 * conceptus\ parameter^3 * (DIP - 50)^2 & \text{if } DIP > 50 \\ -\ total\ conceptus\ weight & \text{if } DIP = gestation\ length \end{cases} \tag{3.3}$$

where

$$total\ conceptus\ weight(kg) = (0.0148 * gestation\ length - 2.408) * calf\ birth\ weight$$

$$conceptus\ parameter = \frac{total\ conceptus\ weight^{\frac{1}{3}}}{gestation\ length} - 50$$

Where DIP = days in pregnancy. Total conceptus weight is calculated from the gestation length and the calf birth weight, adjusted from NRC (2001).

For cows, simulated BW change is decomposed into 3 sources: growth, conceptus, and lactation. The ADG represents the growth rate until the start of the third lactation (NRC, 2001):

$$cow\,ADG(kg/d) = \begin{cases} \frac{(0.92-0.82)*MBW}{average\ calving\ interval} & \text{if } parity = 1 \text{ and non-pregnant} \\ \frac{0.92*MBW-BW}{gestation\ length-DIP} & \text{if } parity = 1 \text{ and pregnant} \\ \frac{(1-0.92)*MBW}{average\ calving\ interval} & \text{if } parity = 2 \text{ and non-pregnant} \\ \frac{MBW-BW}{gestation\ length-DIP} & \text{if } parity = 2 \text{ and pregnant} \\ 0 & \text{else} \end{cases} \tag{3.4}$$

Where MBW = Mature BW and DIP = days in pregnancy. Conceptus growth for cows is modeled the same as for heifers, and the accumulated conceptus weight is subtracted from the simulated cow's BW record at calving. The BW change due to tissue mobilization and deposition during lactation is modeled as an adaptation of the model presented by (Galvão et al., 2013):

$$Lactation\, BW\ change(kg/d) = \begin{cases} -\frac{P_1}{P_2} * exp(1 - \frac{DIM}{P_2}) + \frac{P_1}{P_2^2} * DIM * exp(1 - \frac{DIM}{P_2}) & \text{Lactating cow} \\ \frac{P_1 * \frac{DIM}{P_2} * exp(1 - \frac{DIM\ when\ dry}{P_2})}{gestation\ length - DIP\ when\ dry} & \text{Dry cow} \end{cases} \tag{3.5}$$

Where P_1 is the maximum decrease of BW during lactation, P_2 is the time during the lactation with minimum BW. Where P_1 is set to 20 and 40 for the first and later lactation cows, and P_2 is set to 65 and 75 for the first and later lactation cows (Galvão et al., 2013).

3.4.4 Reproduction

Options for reproduction programs are those suggested by the Dairy Cattle Reproduction Council (DCRC, 2018) and experts in the field, which can be selected and applied to nulliparous heifers and lactating dairy cows, as desired (Figure 3.3).

Heifer breeding eligibility and estrus cycling start when the simulated animal's age or BW reaches a user-defined threshold. A random draw of estrus cycle length distributions for heifers (De Vries, 2006) determines estrous cycle length. Then, a heifer's 1st breeding is simulated using estrous detection (ED), timed artificial insemination (TAI), or synch-ED. For the ED protocol, the model simulates the spontaneous estrus cycle. On the day of scheduled estrus, a random draw from a uniform distribution $U(0,1)$ is compared to the estrous detection rate to determine whether the estrus was detected. If that random drawing is smaller, AI is scheduled on that day with the pregnancy per AI (P/AI) of natural estrus defined by the user. Otherwise, the next estrous cycle starts. Hormonal treatments that simulate the estrous cycle's synchronization on the prescribed

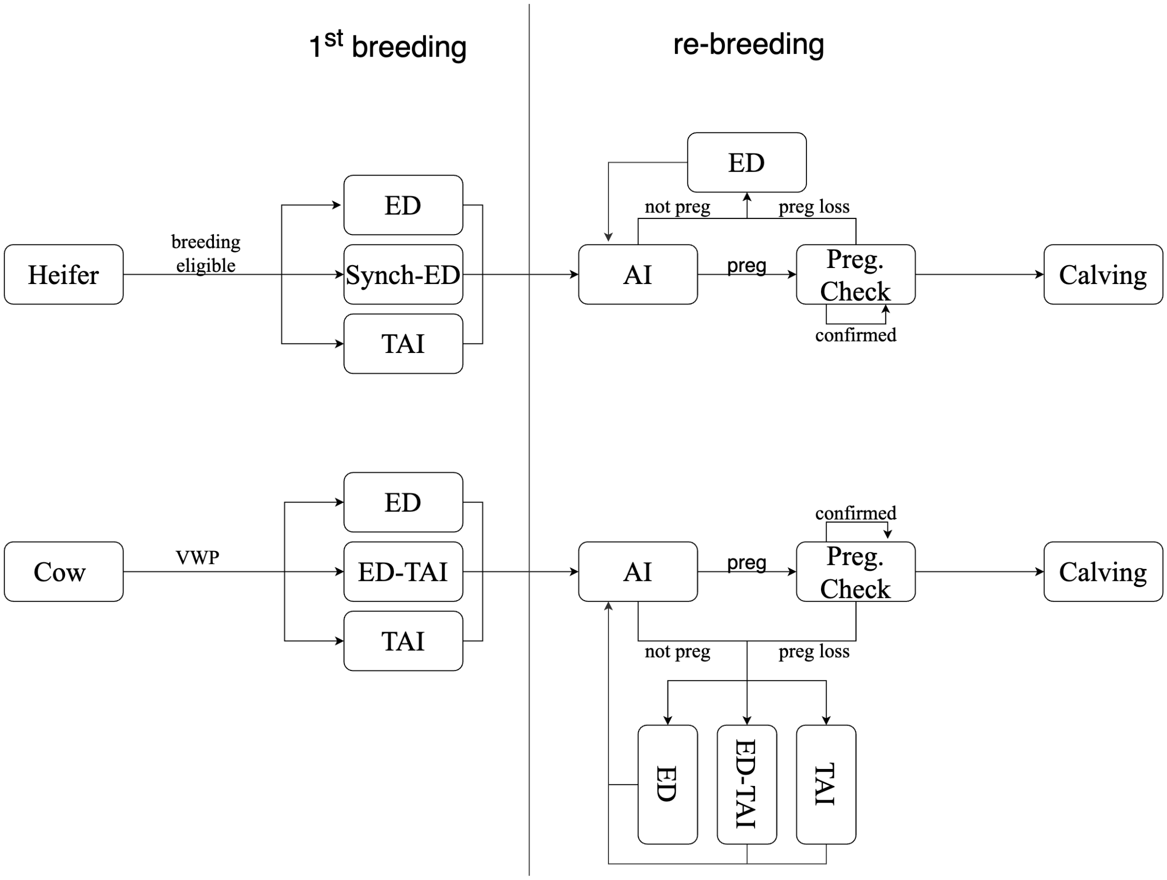


FIGURE 3.3: Schematic representation of the reproductive program simulation in the animal life cycle sub-model of Ruminant Farm Systems model (RuFaS). ED: estrus detection method; TAI: timed-AI method; Preg.: pregnancy

date from the start of the synchronization program is modeled and recorded for the timed AI programs. The date of the AI and the corresponding P/AI are modeled and recorded according to the assigned protocol recommendations. In the synch-ED program, hormone treatment dates are recorded similarly to the TAI program, and estrus is simulated by a protocol-specific estrous cycle after hormonal injection. Simulation of estrous activity detection occurs similarly to the ED program. The TAI protocol starts after a precisely defined time of not showing estrus. For later breedings, also defined by the user, the ED program can be applied to all the heifers who did not impregnate or had a pregnancy loss.

Adult cows return to assigned breeding protocol with spontaneous estrous cycle or at the scheduled time for the first hormone injection of a synchronization protocol during the first and subsequent lactations to be bred after the voluntary waiting period (VWP). Then, cows' 1st insemination after calving is simulated using 3 types of reproduction programs: ED, TAI, or estrous detection combined with TAI (ED-TAI). In the ED program, each estrus day of that cycle is determined after the start of a spontaneous return to estrus after calving with a random draw from the user input estrous cycle. If a cow is not detected in estrus or not inseminated to detection of estrus for a defined number of cycles, the cow can be enrolled in a TAI program. The estrous cycle length is defined similarly to the heifer ED breeding program with distributions that designate the estrous cycle's length. In the TAI programs, dates of hormonal treatment injections are recorded, and the AI day is simulated according to the expected date determined by the protocol with a defined P/AI. ED methods are combined with TAI programs in the ED-TAI program, where ED processes are modeled after the second hormonal treatment and before TAI starts with a scheduled estrous cycle. For ED-TAI programs, estrous detection could also be combined with any TAI resynch protocols. For subsequent breedings, cows can be submitted to any of those 3 reproductive programs based on the user's choice.

Studies on the dynamic process during ovulation and hormones circulation during the estrous cycle built the foundation of most hormone-based programs (Sartori et al., 2017). Proportions of the heifers and cows which do not follow the ovulation cycle in the herd influence the overall estrous detection rate and P/AI. The estrous detection rate and P/AI for the same program vary greatly across herds because of the extensive herd anovulation prevalence from different studies with different measurement methods (5 to 45%, Walsh et al., 2006; 7.3 to 41.7%, Bamber et al., 2009). To account for this phenomenon, we defined an anovulation rate variable in the simulation to represent the percentage of anovular cows in the herd to adjust the estrous detection rate of ED programs and P/AI of all programs for cyclic and anovular cows.

Pregnancy checks for both heifers and cows are scheduled several times after AI. There are typically 3 pregnancy checks. First at 32 d after breeding, which, if confirmed positive, is followed at 60 and 200 d (Wiltbank et al., 2016), but these can be adjusted according to farm practices. Between TAI and pregnancy checks, pregnancy losses are modeled to occur if the random draws from a $U(0,1)$ are smaller than the chance of pregnancy lost between AI and 1st pregnancy check, between 1st and 2nd pregnancy check, and from 2nd pregnancy check and 3rd pregnancy check.

3.4.5 Lactation

A non-linear model that describes the lactation curve estimates daily milk production based on each animal's DIM and a randomly selected set of parameters defining their lactation curve. The estimated milk production feeds into the ration formulation sub-model as the target milk production defines nutrient requirements. We have updated state and breed-specific parameters for the lactation curve model described by Wood (1980) from data provided by the Council of Dairy Cattle Breeding (Li et al., 2019). Breed, parity, and state-specific lactation parameters mean and standard deviations are stored as input options that the user can select. A random draw from each parameter distribution is collected at the beginning of the lactation. The target milk production is updated daily according to DIM. Milk production variability at each DIM is added to the ration-adjusted actual milk production by a stochastic number drawn from $N(0,0.76)$ for primiparous cows and $N(0,0.9)$ for multiparous cows (Nielsen et al., 2010).

3.4.6 Culling

Each day, every animal has a chance to be culled, and after an animal is voluntarily or involuntarily culled, we remove her from the herd simulation. We store the number of animals that are culled, their attributes, and culling reason for summarization in the model output.

A culling event is simulated due to reproductive failure, disease (Pinedo et al., 2010), and death. Suppose the Monte Carlo process results in an instance of the Heifer II class that is not pregnant by the heifer reproductive culling age threshold defined in the input. In that case, the animal is culled for reproduction failure. If an instance of a Cow class does not become pregnant by a maximum DIM to breed, a 'do not breed' attribute is attached to the animal's record. When the milk production of an instance of a Cow with a 'do not breed' attribute drops lower than a user-defined threshold, the animal's record is removed from the herd and marked as reproductive culling.

In addition to reproductive culling, we simulate six health culling reasons: lameness, injury, mastitis, udder problems, other diseases, or unknown causes (Kalantari et al., 2016). For each reason, culling rates are defined by parity-based annualized live culling and death rate reported by Pinedo et al. (2010) and adjusted by Kalantari et al. (2016, Table 3.1). At the beginning of each lactation, the probability of culling in that parity is compared with a random number from $U(0,1)$. If the cow is scheduled to be culled in that parity, another random number is drawn from $U(0,1)$ compared to the thresholds set by Kalantari et al. (2016) to determine the culling reason. The third draw from $U(0,1)$ is compared to the corresponding distributions of a 530-d length lactation to schedule the culling date through an inverse transformation of an empirical cumulated distribution function for the probability of culling occurring each day. The mortality event of each parity is programmed in the same way as the culling for health reasons: at the beginning of each parity, a random $U(0,1)$ draw is compared with the likelihood of death during that parity. Another random draw determines the DIM of the occurrence of this culling event. The threshold to determine the death culling and mortality are shown in Table 3.1.

TABLE 3.1: Thresholds for mortality and health culling events scheduling using next-event scheduling implemented on the animal life cycle submodel of RuFaS model

[1]Culling reason	[2]Thresholds
Cow mortality	
Parity 1	3.9 %
Parity 2	5.6 %
Parity 3	8.5 %
Parity >3	11.7 %
Cow health culling	
Parity 1	16.9 %
Parity 2	23.3 %
Parity 3	30.1 %
Parity >3	40.8 %
Health culling reasons	
Lameness	16.33 %
Injury	28.83 %
Mastitis	24.39 %
Udder problems	6.45 %
Other diseases	13.91 %
Unknown causes	10.09 %

[1] Annualized death and live culling rates and frequency distributions are based on cow parity from Pinedo et al. (2010).

[2] Next-event scheduling approach (De Vries, 2001; Kalantari et al., 2016) schedules the stochastic events that could happen to the cow at the beginning of each parity. a random draw from $U(0,1)$ is compared to the threshold probability to determine whether that event would happen in that parity. Another random draw from $U(0,1)$ is compared to the corresponding distributions of a 530 d length lactation to schedule the culling date.

3.4.7 Herd Initialization

At the first use of the animal life cycle submodel, a database of animals is generated as an initialization database. The initialization database is created through the regular simulation process but simulating a large number of animals. Each animal's core status information and life history at the end of the simulation are stored in the database. The goal of using an initialization database is to reduce the time for the simulation to reach a steady state by initializing the herd with animals of different ages, stages in pregnancy, DIM, and parities. The stochastic simulation starts by randomly selecting the user-defined number of animals in each Animal Class (Calf, Heifer I, Heifer II, Heifer III, Cow) from the initialization database. Each of the individual animals and the herd update with daily steps as described above.

3.4.8 Computer Implementation

A modular structure was used to develop each state of animals and events. The modular structure in the code means the system is separated into many modules with independence and dependence. Each module has functions on a specific aspect and is also connected to represent the herd simulation. The modularized design allows smooth and flexible connections with the other RuFaS modules (Kebreab et al., 2019). For example, calves' ration formulation can be connected with the calf states without interacting with other states like heifer or cow, if not desired. Not only the structure design of the program affects the coding process, but also the approach plays a role in the models' features on flexibility and easiness of interpretation. We use an iterative and incremental development approach (Larman and Basili, 2003) as a type of agile practice to implement the conceptual model to a computer program. Moreover, the model benefits from Python's dynamic semantics characteristics (Rossum. and Drake, 2009), including its supported modules and packages. Overall, our standard developing process promotes program modularity and code reusability.

The animal life cycle submodel uses 5 types of files. They are: (1) The initialization database to generate a file to store animal life information for simulation initialization; (2) The animal base file to handle universal information of each animal like ID and breed; (3) The calf, heifer, cow, and culled files to simulate life events of animals at each stage; (4) The herd simulation file to manage herd-level management decisions for each animal, like semen type choices or reproduction program selection; And (5) the input and output files management.

Working as a team of developing the whole farm RuFaS model requires a consistent workflow of programming processes: (1) Standardizing code documentation, including flowcharts and pseudocodes, and using uniform naming conventions. (2) Branching models for all programmers to be aligned and avoid possible conflicts. And (3) having a code review organism and regular meetings. Each code branch is reviewed by at least 2 other programmers and one subject matter specialist to ensure the code fits the intention and provides the expected output. The development routine ensures the RuFaS model serves as a functional framework for further development and expansion.

For the animal life cycle submodel within the RuFaS, the input data could come from the user, herd management software, or the nutritional model. And its output data could be passed to the herd management submodel, nutrition submodel, economic module, water module, or energy module (Kebreab et al., 2019).

3.4.9 Case Study

We illustrate the functionality of the life cycle submodule with a case study evaluating the strategies of using beef and sexed semen in addition to conventional semen.

Background. The use of beef semen with sexed dairy semen has become a popular management choice in the dairy industry in recent years (Ettema et al., 2017). On the one hand, improved reproductive performance, causing an oversupply of heifers (Weigel, 2019), has increased the heifer rearing burden on farms. And on the other hand, the beef cattle demand projected to remain high in the next ten years (USDA Agricultural Projections to 2029, 2020) promises favorable crossbred prices in the foreseeable future. Moreover, the combination of sexed and beef semen could increase the selection intensity and shorten the genetic lag in herd genetic progress (Hjortø et al., 2015). Under those conditions, the practice of using sexed semen on heifers and reproductive or productively superior cows while using beef semen for the rest of the cows is a promising long-term dairy farm management strategy (Li and Cabrera, 2019). The assessment over the combinations of sexed, conventional, and beef semen programs remains unclear. We used our animal life cycle submodel to evaluate those strategies.

Scenarios. We designed a study with a total of 8 scenarios, including 3 semen programs, each with 2 levels of reproductive performance (high and moderate) and a baseline scenario with conventional semen only at the high (C-h) and moderate (C-m) reproductive performance. To simulate the rearing of 5% more heifers than remain in the herd, we set to keep 86% (for C-h) and 88% (for C-m) female calves born on-farm via the female calf keeping rate in the input for the baseline scenarios. The remaining female Holstein calves and all male calves were sold. We set case-study programs based on expert knowledge of common practices in Wisconsin farms that use sexed semen on younger animals and earlier breedings for greater genetic potential and cost-efficiency (Weigel, 2004). We define the experimental scenarios in Table 3.4: For the 3 semen combination programs studied (6 scenarios), we set: SB-h and SB-m to use sexed semen on all heifers, top 45% 1st lactation cows and top 10% (SB-h) or 15% (SB-m) 2nd lactation cows, and beef semen on other cows; SCB-h and SCB-m to use sexed semen on all heifers, top 25% 1st lactation cows, conventional semen on next top 25% (SCB-h) or 30% (SCB-m) 1st lactation cows and top 35% (SCB-h) or 40% (SCB-m) 2nd lactation cows, and beef semen on all other cows; SCB-h2 and SCB-m2 to use sexed semen on all heifers, top 25% 1st lactation cows for first 3 inseminations, conventional semen on next top 25% (SCB-h2) or 30% (SCB-m2) 1st lactation cows and top 35% (SCB-h2) or 40% (SCB-m2) 2nd lactation cows, and beef semen on all other cows. In these 6 scenarios, we kept and raised all female Holstein calves born on the farm. All male Holstein calves and all crossbred calves were sold. We developed an arbitrary and randomly assigned attribute that modified each animal's baseline production level and determined the ranking system for reproductive decision-making. At the beginning of the simulation following each animal is assigned a random draw from $N(1.0, 0.1)$ within the range of $(0.8, 1.2)$ to indicate the animal's relative production level amongst the original herd. A 1,000-cow Holstein herd in Wisconsin was simulated with 100 replications for each scenario for 7 yr. The outputs compared in the case study represent summaries of the final year of the simulation.

General simulation settings for the case study. All user inputs for the case study scenarios are listed in Table 3.2 and were aligned with the framework described in the development of the submodule. The only exception was the BW that was simulated here using a triangular distribution $T(550, 700, 1000)$ (where a is the lower limit, b is the upper limit, c is the mode, and $a \leq b$, $a \leq c \leq b$. Kalantari et al., 2016) instead of a normal distribution as above described. This was because that the case study was conducted before the mature body weight in default settings was updated with the normal distribution. An analysis of using either distribution (normal or triangular) in a test simulation analysis indicated that it would not make significant differences in the results because the means of those two distributions were not far from each other, and in the simulation, the heifer reproductive program starting day was set as age other than BW of the heifer.

The parameters that differed between the case study scenarios include the conception rate of the first insemination and the female calf keeping rate. Sexed semen produced 90% female calves (DeJarnette et al., 2009), whereas conventional only 47% (Río et al., 2006). Female Holstein calf birth weight distribution followed N (42.9, 6) kg (Linden et al., 2009). The wean day was set at 60 d born (Soberon et al., 2012). The ADG for pre-weaning calves was set at one-sixtieth of the weaning weight minus their birth weights. The cost of raising calves was calculated as \$5.5 per calf per day, \$2.2 per heifer per day for non-pregnant heifers, \$3.19 per heifer per day for pregnant

TABLE 3.2: Benchmarks and distributions of events used in the Monte Carlo simulation process

Event description	Distribution	Threshold	Reference
Stochastic variables			
Reproductive:			
Stillbirth rate, %	U (0,1)^2	0.065	Bicalho et al., 2008
Female calf rate. sexed semen, %	U (0,1)	0.9	DeJarnette et al., 2009
Female calf rate, conventional semen, %	U (0,1)	0.47	Río et al., 2006
Female calf keeping rate, baseline scenarios, %	U (0,1)	0.86/ 0.88	Case study
Female calf keeping rate, other scenarios, %	U (0,1)	1	
Heifer estrus cycle^2	N (21,2.5)	NA	De Vries, 2006
First Ovulation after calving^2	N (23,6)	NA	Schubert et al., 2018
Heifer estrus cycle^2	N (21,4)	NA	Oltenacu et al., 1980
Gestation length, d	N (278, 6)	NA	Kalantari et al., 2016
Conception rate of the first insemination^3, %			
Heifer with conventional and beef semen	U (0,1)	0.60 /0.55	Case study
Heifer with sexed semen	U (0,1)	0.48 /0.44	
Cow with conventional and beef semen	U (0,1)	0.55 /0.50	
Cow with sexed semen	U (0,1)	0.44 /0.40	
Pregnancy lost in 3 periods between pregnancy checks, %	U (0,1)	0.02, 0.096, 0.17	Galvão et al., 2013
Bodyweight:			
Holstein calf birth weight, kg	N (43.9, 1)	NA^4	Linden et al., 2009
Mature body weight, kg	N (740.1, 73.5)^5	NA	Schubert et al., 2018
P_1 for 1st/ later lactation	20/40		
P_2 for 1st/ later lactation	45/60		
Management:			
Female calf keeping rate, %	U (0,1)	1	Case study
Variation in daily milk yield, primiparous, kg	N (0,0.76)	NA	Nielsen et al., 2010
Variation in daily milk yield, multiparous, kg	N (0,0.90)	NA	Nielsen et al., 2010
Deterministic variables			
Wean day, d	NA^6	60	Soberon et al., 2011
Breeding start, d	NA	380	Dawod and Elbaz, 2020
Age for heifer reproductive failure, d	NA	670	Giordano et al., 2012
Pregnancy checks subsequent to AI, d	NA	32, 91, 200	Galvão et al., 2013
DIM for the start of the dry period, d	NA	218	Case study
DIM for cows marked as do-not-breed, d	NA	300	Giordano et al., 2012
P_1 for 1st/ later lactation BW change	NA	20/40	Galvão et al., 2013.
P_2 for 1st/ later lactation BW change	NA	45/60	
Milk production when the do-not-breed cow is culled for reproductive failure, kg/d	NA	22	Case study
DIM when cow TAI protocol started, d	NA	45	
DIM for first insemination post-calving, d	NA	72	
DIM for the start of the dry period, d	NA	218	

[1] A random draw from U (0,1) is compared to the threshold probability of an event occurring to simulate if that event occurs or not.

[2] Those variables related to the estrus cycle are used when either ED or ED-TAI method is selected for the reproduction program.

[3] Conception rates set for Double-Ovsynch protocol with high/moderate reproductive performance.

[4] A random draw from the threshold distribution of animal attributes and assigning that value to the instantiation of an individual animal.

[5] The triangular distribution T (550,700,1000)
(Kalantari et al., 2016) was used in this case study

[6] For deterministic variables, the event occurs on the threshold scheduled time.

heifers (Tranel, 2019). Breeding started at 380 d born (Dawod and Elbaz, 2020). The heifer's reproduction program selected was TAI with the 5dCG2P protocol (DCRC, 2018, shown in Figure 3.4(A)) with a 60% or 55% conception rate (CR) for the first insemination for high and moderate reproductive performance, respectively. After the first AI, each additional insemination had a 2.6 percentage points lower CR (Galvão et al., 2013). The gestation length of heifers followed an $N(278,6)$ d (Kalantari et al., 2016).

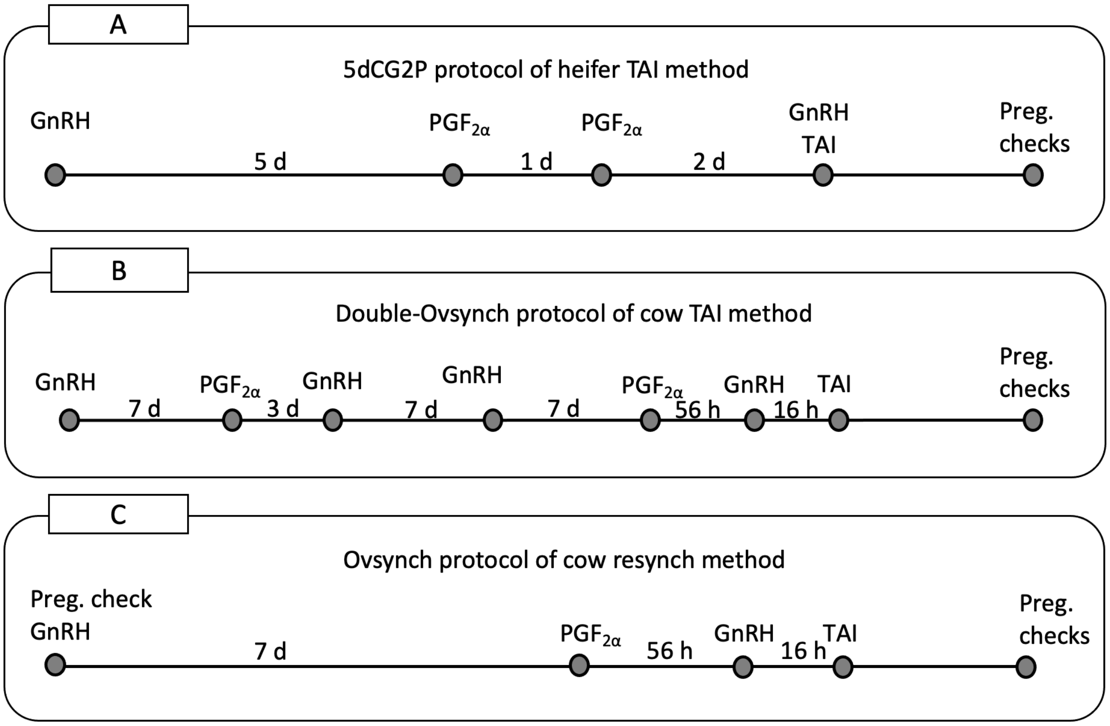


FIGURE 3.4: Schematic diagram of the reproduction protocols used in the case study. A: heifer TAI program, 5-d CIDR-Synch with GnRH and 2 PGF; B: cow TAI program, Double Ovsynch; C: cow resynch program, Ovsynch 56. Protocols suggested for heifers and cows by Dairy Cattle Reproduction Council: (2018) https://www.dcrcouncil.org/protocols/

In the Cow Class, reproduction protocols start at 45 DIM with the Double Ovsynch protocol and OvSynch-56 after pregnancy check for resynch (Figure 3.4(B); DCBC, 2018). We simulated the first insemination after calving on DIM = 72 and set the CR for 1st lactation cows to 55% for the first insemination of conventional and beef semen. As with heifers, we simulate each subsequent insemination with a 2.6 percentage point drop in CR (Galvão et al., 2013). We also decrease the CR at each lactation with a decrease of 5% for the 2nd lactation and 10% for each subsequent lactation. We set sexed semen to have a CR that is 80% of the CR of beef and conventional semen (Ettema et al., 2017). If cows do not conceive after calving, we simulate the OvSynch protocol as resynch. Culling time affects the structure and economics of a herd.

We calculate the reproductive performance during the simulation. The 21-d service and 21-d conception rates are calculated based on the last 15 periods of the 21-d (315 d) before the simulation ends. The 21-d pregnancy rate is defined here as the product of the service rate and conception rate. The service rate measures the percentage of eligible animals inseminated, whereas the conception rate measures the pregnancy by insemination ratio.

Milk production follows Wood's model (Wood, 1967) to predict milk yield according to days in milk at t days after parturition multiplied by the cow-specific production level:

$$y = \left(at^{b}e^{-ct}\right) * P_i \tag{3.6}$$

In Eq. 3.6, parameter a is the scale factor for initial milk yield, b is a rate of increase for the increase in milk yield to peak, and c is the rate of decrease for the decline in milk yield after the peak, and is the production level of that cow. Parameters of a, b, and c estimated for Wisconsin from the year 2016 shown in Table 3.3 (Li et al., 2019) were used to calculate the target milk production in the simulated cows. Cows were dried at 218 DIP (60 d from the average gestation length of 278 days). Therefore, the simulated dry period for each cow varied as a function of the randomly selected gestation duration.

TABLE 3.3: Parameters of Wood's model for the lactation curve for WI, 2016, used in the case study.

	Wood's fitted parameters[1]		
Parity	a (kg/d)	b	c
1st	16.13 ± 0.28	0.235 ± 0.0046	0.0019 ± 3.77^{-5}
2nd	23.61 ± 0.54	0.227 ± 0.0064	0.0032 ± 5.82^{-5}
>2nd	23.81 ± 0.51	0.244 ± 0.0060	0.0036 ± 5.54^{-5}

[1] Parameters of the fitted lactation curve from Wisconsin data of the year 2016 in a previous study (Li et al., 2019)

We recorded economic values for the last year (365 d) of the 7 yr simulation (with 100 replications). We used the default diet for the adult group, which on average, cost \$0.175/ kg of DM. A daily \$2.5 per cow per day was also included as a fixed cost on-farm for operation and management (Giordano et al., 2012). We calculated the TAI program cost based on the price of GnRH at \$2.4/dose and the price of PGF at \$2.65/dose for the reproduction cost. Labor cost was set at \$0.25/ injection of hormones, \$5/ insemination, \$3/ pregnancy check (Galvão et al., 2013). The semen cost was \$35/insemination for sexed semen, and \$15/insemination for beef or conventional semen, and the calf price was \$225/head for crossbred, \$57.5/head for male Holstein, and \$45/head female Holstein (Li and Cabrera, 2019). Culled heifers had a value of \$104/head at any age (USDA-AMS, 2020). The springer price was set at \$1310/head for selling and buying, liveweight salvage value at \$1.39/kg, and the milk price at \$0.37/kg (USDA-NASS, 2020). The targeted milk production was considered actual production. The economic value of switching from conventional semen to sexed and beef semen combinations was measured in terms of annual Net Return (NR) as the difference between the income and cost for each comparison. All values were the average across 100 iterations of the last 365 d of the 7 yr simulation. The income sources of the NR were milk, salvage, springer sales, and calf sales. Costs included feed, breeding (including hormones, AI, pregnancy checks, and associated labor), semen, rearing cost of young animals, and other fixed costs. The crossbred calf price varies a lot from day to day based on the market situation. We also included a post-simulation breakeven analysis to investigate which scenarios should be preferred in different market situations. We maintained Holstein's calf prices in the breakeven analysis and varied the crossbred calf price to determine the minimum viable crossbred calf price.

3.5 Results and Discussion

3.5.1 Model output demonstration

The animal life cycle submodel is programmed with a structure suitable for future expansions and manipulations by allowing individual animal and herd information flows within and across modules. Most simulation models have been designed to evaluate a specific topic or management practice emphasizing the simulation process and outputs on that topic. For example, (Giordano et al., 2012; Galvão et al., 2013) assessed the reproductive strategies of herds through different whole-farm simulations with focuses on reproductive indications and overall farm economics as outputs. Likewise, Kalantari et al. (2016) addressed nutritional grouping applications in herd simulation, which gave diets and income over feed cost the most attention.

In contrast, the animal life cycle submodel of the RuFaS model is designed to serve multiple functions that include the investigation into the herd and life-cycle management as demonstrated by the case study, including tracking of information required by other submodules. For example, daily updates of individual animal BW, growth, target milk production, and stage in pregnancy are needed by the nutrition submodule for ration formulation and manure excretion. Daily animal purchases and sales can assist in tracking nutrient balances and records on the number of hormone injections can support economic analyses. In addition, simulated outcomes from the life-cycle submodule are available at multiple scales, including the whole herd's level, animal class, pen, and individual animal. Individual life stories can be followed, making it possible to evaluate precise management decisions.

Herd dynamics and steady state. Figure 3.5 depicts the herd structure output from the CH scenario input for the case study. It shows the number of animals in each group during the 7-year simulation period. The number of cows represents the number of simulated animals in the Cow class between their first calving and the day they leave the herd, including milking and dry cows. As expected, the total number of cows oscillates around 1000 because we set the target herd size at 1000. The total number of cows is divided into the sub-groups of pregnant, open, milking, and dry cows. The number of young animals, including the number of calves and heifers, oscillates around 820 animals (82% of the adult herd) and maintains sufficient replacements on-farm. The number of young animals reflects replacement policy and reproductive performance (Pennsylvania State extension, 2020). This number aligns with calculating the number of young animals raised on-farm sufficient for the herd compared with the number calculated using the percentage of heifers raised, calving interval, calf mortality rate, and age of calving (803, (Heinrichs et al., 2017). Our animal life cycle model took a shorter time (600 days of the simulation) to reach a steady-state than the Galvão et al. (2013) model in which a similar graph identified steady-state within 1500-2000 d of simulation. The shorter time needed to reach the steady state is attributed to the herd initialization process.

More herd-level variables can be easily accessed with the herd report output for the animal life cycle submodel. Full access to monitoring the simulation process and detailed output ensures understanding the simulation in many ways. Changes in various output variables can qualify the effect of changing input decisions spontaneously. If something unexpected occurs, the user can go deeper into the code and drill down the possible causes.

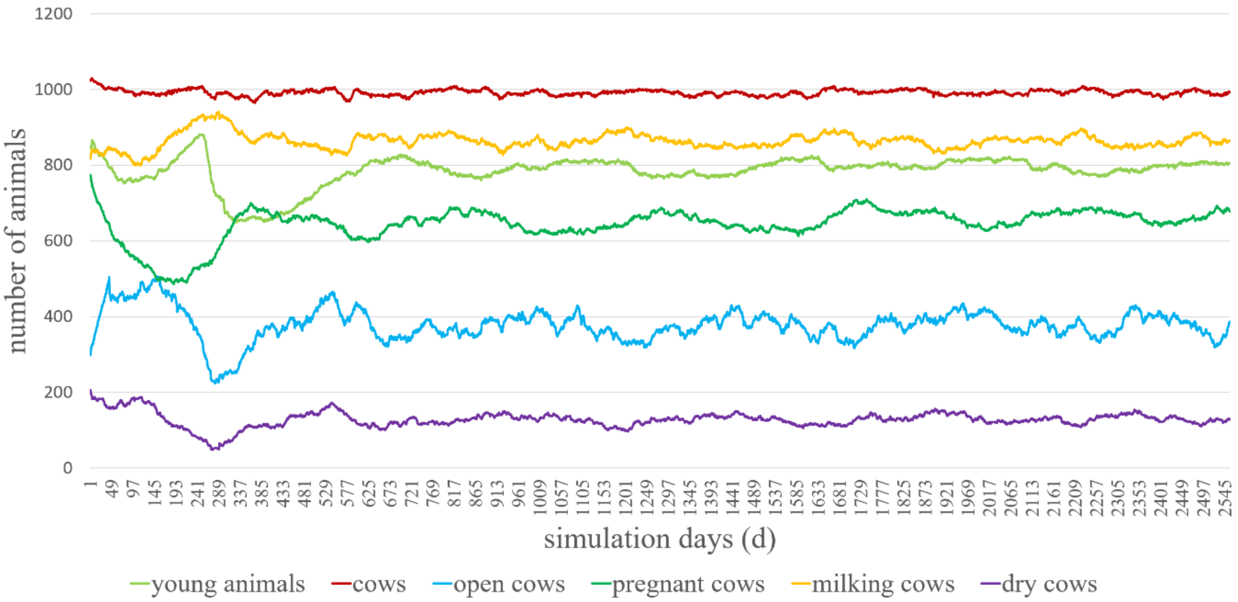


FIGURE 3.5: The herd structure along the simulation time of one simulation of 1000-cow herd of the C-h scenario (conventional semen used in the high reproductive herd) in a 7-year run.

Individual report. The animal life cycle submodel stores and reports individual animal life stories, often unavailable from other models. Figure 3.6(a) and (b) show examples of life story events, BW history, and production records for 2 animals randomly picked from the simulation. The animal life story is shown in Figure 3.6(a); she entered the herd from the initialization database and was in the 3rd lactation, 21 days after initiation of her 4th pregnancy when the simulation started. Due to lameness, she left the herd during her 5th lactation at 151 DIM. The animal represented by the life story displayed in Figure 3.6(b) was born on the farm at 29 d after the start of the simulation, conceived as a heifer at 388 d of age, passed all pregnancy checks, calved, and started her first lactation at 664 d age, after 276 d of gestation. After the first calving, it took her 72 d to conceive again and successfully carried her second pregnancy and had second calving. However, during her second lactation, she got pregnant at the 3rd insemination and lost that pregnancy between the 1st and 2nd pregnancy checks. Although she returned to the Ovsynch 56 after abortion, she was still not pregnant in the next AI and was marked as ’do-not-breed’ at 300 DIM. She was then culled for reproductive failure at the age of 1802 d and 411 DIM into her 2nd lactation when milk production dropped below 22kg/d.

Detailed reproductive events are also shown in Figure 3.6(b) with records of hormones applications, AI, and pregnancy checks. The BW represents target values that depend solely on the life stage and the cow-specific mature body weight. The milk production describes the target milk production and daily production variation.

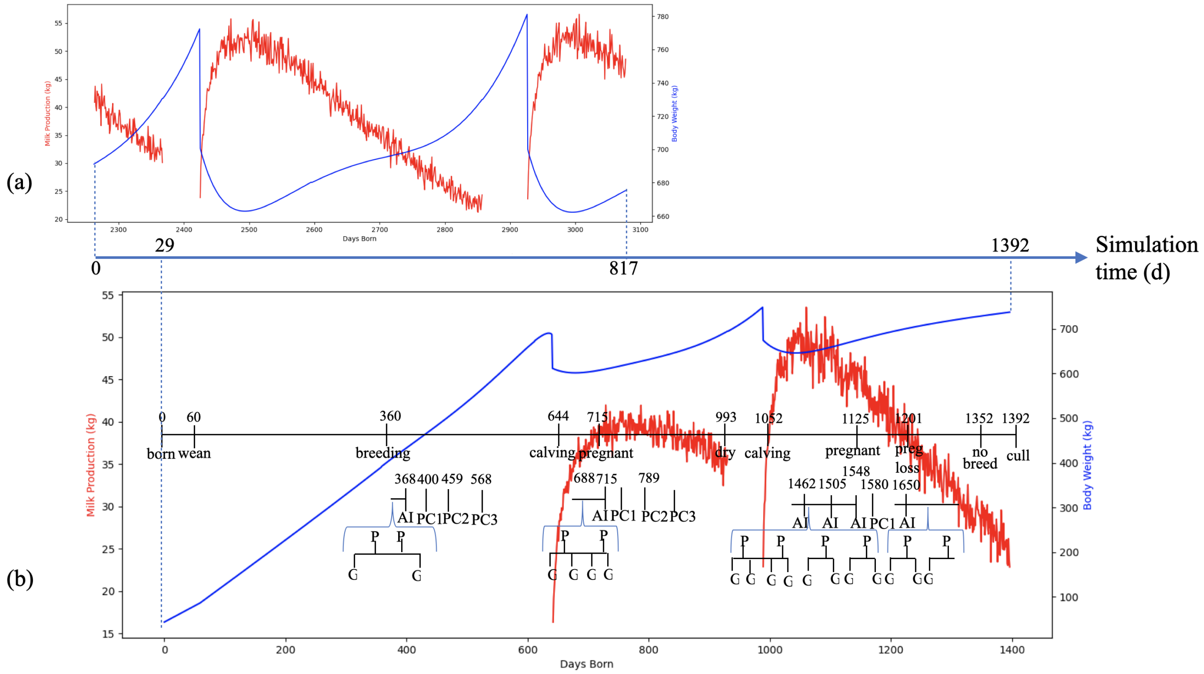


FIGURE 3.6: Individual cow production, BW history, and event records of two simulated cows in one simulation of 1000-cow herd of the C-h scenario (conventional semen used in the high reproductive herd) in a 7-year run.
PC = pregnancy check, P: injection of PGF, G: injection of GnRH

Simulation of individual animals on a daily time-step in the life cycle submodel expands the opportunity to represent a wider diversity of management practices. For instance, nutritional grouping and improved nutritional accuracy components (Barrientos-Blanco et al., 2020) are under development to improve feed efficiency using the animal model. The model makes it possible to formulate closer-to-requirements diets. It also helps to further incorporate the modeling process with more precise information associated with individual animals, such as genetic makeup (Martin and Sauvant, 2010) or grazing behavior (Johnson et al., 2008). The simulation structure empowers users to apply management practices to selected proportions of the animals. As illustrated in the case study, we applied combinations of semen types to selected groups of animals.

3.5.2 Case study evaluation

The use of beef semen with sexed dairy semen has become a popular management choice in the dairy industry in recent years (Ettema et al., 2017). However, this management strategy might not be optimally applied because the lower fertility of sexed semen may challenge herd size maintenance so that the extra income from crossbred calves may not always compensate for the extra cost for the semen (Weigel, 2004). However, recent industry trends may make this management strategy viable. Improved reproductive performance has led to an oversupply of heifers (Weigel, 2019). In addition, beef cattle demand is projected to remain high in the next ten years (USDA Agricultural Projections to 2029, 2020), which promises favorable crossbred prices in the foreseeable future. Moreover, the combination of sexed and beef semen is expected to increase selection intensity, shorten genetic lag, and increase genetic progress (Hjortø et al., 2015). Preliminary results from our group suggest that under these expected conditions, the practice of using sexed semen on heifers and reproductive or productively superior cows while using beef semen for the rest of the herd is a promising long-term dairy farm management strategy (Li and Cabrera, 2019). We used the flexibility of the life-cycle submodule in the RuFaS environment to investigate this management option further.

Economic Values. The financial results of the case study are summarized in Table 3.4. All scenarios that used sexed and beef semen had a better NR than the baseline scenario that used only conventional semen. The scenarios SB-h, SCB-h, and SCB-h2 had a higher NR than C-h ($160, $157, $156 greater NR per cow per yr, respectively). The scenarios SB-m, SCB-m, SCB-m2 had higher NR than C-m ($152, $129, $133 greater per cow per yr, respectively). The additional income from milk and crossbred calf sales when using sexed and beef semen exceeded the extra cost of breeding, semen, and rearing associated with sexed semen. As expected, more replacements were needed, with a lower conception rate associated with sexed semen (Table 3.5). The selection of breeding cows with lower reproductive and productive performance to beef semen increased the herd's relative productivity level, bringing higher milk income. Meanwhile, the higher price of crossbred calves ($225) compared with Holsteins calves ($57.5 for males and $45 for females) benefited from selling more crossbred calves.

Using a combination of conventional, sexed, and beef semen in the high reproduction herd (SCB-h, SCB-h2) showed similar NR outcomes to the SB-h scenario. Breeding lower fertility cows (2nd lactation and 4th and later inseminations in 1st lactation) with conventional semen yielded a higher 21-d average pregnancy rate (27.4% for SCB-h and 27.5% for SCB-h2 vs. 26.4% for SB-h), slightly shorter calving interval (Table 3.5), and more calves sold (Table 3.6). These reproductive improvements compensated economically for reduced income from selling fewer crossbred calves. When the reproductive performance was moderate, the SB-m scenario had a much higher NR than the other scenarios. At lower production levels, the advantage of having lower semen cost did not compensate for reduced milk income and fewer crossbred calves.

Break-even analysis of crossbred calf price. We used the number of young animals for each simulation (Table 3.6) to calculate income from selling calves with changing crossbred-calf prices. When the crossbred calf price went down to $171/head from $225/head, the NR of the SB-h and

TABLE 3.4: Semen used in each scenario of the case study for the 7-year simulations

Scenarios[1]	Sexed semen on[2]			Conventional semen on[2]	
	Heifers	1st lactation cows	2nd lactation cows	1st lactation cows	2nd lactation cows
SB-h	All	Top 45%	Top 10%	-	-
SCB-h	All	Top 25%	-	Next top 25%	Top 35%
SCB-h2	All	Top 25% for first 3 inseminations	-	Following inseminations, next top 25%	Top 35%
SB-m	All	Top 45%	Top 15%	-	-
SCB-m	All	Top 25%	-	Next top 30%	Top 40%
SCB-m2	All	Top 25% for first 3 inseminations	-	Following inseminations, next top 30%	Top 40%

[1] The baseline scenarios were noted as C-h and C-m, conventional semen used on all heifers and cows. SB-h = sexed, beef semen, high reproduction; SCB-h = sexed, conventional, beef semen, high reproduction; SB-m = sexed, beef semen, moderate reproduction; SCB-m = sexed, conventional, beef semen, moderate reproduction; For the high reproduction scenarios (C-h, SB-h, SCB-h, SCB-h2): conception rate was set at 60% for 1st insemination heifers, 55% for 1st insemination of 1st lactation cows. The moderate reproductive level (C-m, SB-m, SCB-m, SCB-m2) conception rate was set at 55% for 1st insemination on heifers, 50% for 1st lactation 1st insemination cows.

[2] Top refers to each animal's production level ranking amongst the cows at the same lactation. Beef semen was used on the rest of the eligible cows.

the SCB-h2 scenarios were equivalent and $130/cow per yr more than the NR of C-h. When the crossbred calf price was $150/head, the NR of the SB-h scenario was no longer greater than that of the SCB-h scenario, but it was still $118/cow per yr more than the NR of the C-h scenario. When the crossbred calf price was set to 0, the SB-h scenario NR was still $31/cow per yr higher than the C-h scenario. In the moderate reproductive performance group, the NR of the SB-m scenario was always higher than the NR of the C-m, SCB-m, and SCB-m2 scenarios. The advantage gained by higher milk production played a significant role in the SB-m scenario. Even when the crossbred calf price was decreased to 0, the SB-m scenario NR still had $5/cow per yr more than the NR of the SCB-m and SCB-m2 scenarios and $29/cow per yr more than the NR of the C-m scenario. For

TABLE 3.5: The net return of baseline and studied reproductive scenarios of the case study during the last year of the 7-yr simulation

Scenarios[1]	Income ($/cow/yr)				Cost ($/cow/yr)[2]				Net Return ($/cow/yr)
	Milk	Slaughter	Heifer	Calf	Feed	Breeding	Semen	Rearing	
C-h	4,673	332	70	36	1,353	52	35	890	1,867
SB-h	4,807	334	71	133	1,352	56	62	931	2,027
SCB-h	4,805	335	73	122	1,351	55	57	930	2,023
SCB-h2	4,807	331	71	123	1,354	56	58	924	2,024
C-m	4,664	337	70	35	1,350	55	38	905	1,840
SB-m	4,802	340	73	126	1,349	59	68	956	1,992
SCB-m	4,789	340	72	112	1,347	59	62	962	1,969
SCB-m2	4,788	338	72	115	1,350	59	62	951	1,973

[1] Scenarios were: C-h, C-m: conventional semen used for all heifers and cows; SB-h: sexed semen used on all heifers, top 45% for 1st lactation cows, and top 10% on 2nd lactation cows, beef semen used on other cows; SCB-h: sexed semen used on all heifers, top 25% 1st lactation cows; conventional semen used on next top 25% for 1st lactation cows and top 35% for second lactation cows; beef semen used for other cows; SB-m: 5% more on 2nd comparing to SB-h; SCB-m: 5% more on the conventional semen used for both 1st and 2nd lactation cows than SCB-h. SCB-h2, SCB-m2: limited sexed semen used on 1st lactation cows on first 3 inseminations, and conventional semen used for later inseminations compared to SCB-h, SCB-m. The high reproductive level (C-h, SB-h, SCB-h, SCB-h2): conception rate was set at 60% for 1st insemination on heifers, 55% for 1st insemination of 1st lactation cows. The moderate reproductive level (C-m, SB-m, SCB-m, SCB-m2) conception rate was set at 55% for 1st insemination on heifers, 50% for 1st lactation 1st insemination cows.

[2] Cost also included fixed cost, which was $913/ cow per year across the scenarios.

the high reproductive performance scenarios, including conventional semen in addition to sexed and beef semen, it was more profitable when the crossbred price went lower than $171/head. These results show that the benefit of switching from conventional semen to a combination of beef and sexed semen is sensitive to the crossbred calf price and the farm's reproductive performance, in agreement with the findings in previous studies (McCullock et al., 2013; Mur-Novales and Cabrera, 2017).

Herd structure and reproductive performance. The herd structures at the end of the simulation

TABLE 3.6: Herd dynamics [mean (SD)] of the case study scenarios during the last year of the 7-year simulation.

Scenarios[1]	21-day Pregnancy Rate (%)	Estimated Milk Production (kg/cow/yr)	Average Parity	Average calving Interval (d)	Average Days In Milk (d)	Average Culling Rate (%)
C-h	28.4 (0.8)	12,630.0 (70.6)	2.36 (0.05)	397.1 (2.5)	172.5 (3.6)	38.4 (1.9)
SB-h	26.4 (0.9)	12,992.6 (69.7)	2.34 (0.05)	399.7 (2.7)	175.0 (3.7)	38.5 (1.6)
SCB-h	27.4 (1.0)	12,985.8 (86.9)	2.35 (0.05)	398.4 (2.8)	173.7 (4.2)	38.6 (2.2)
SCB-h2	27.5 (1.0)	12,933.0 (73.5)	2.36 (0.05)	398.8 (1.9)	173.7 (3.8)	38.1 (1.7)
C-m	24.5 (0.9)	12,606.3 (95.4)	2.30 (0.05)	402.2 (2.5)	177.3 (3.5)	38.7 (1.7)
SB-m	22.5 (0.8)	12,978.7 (72.2)	2.28 (0.05)	405.5 (2.8)	180.4 (4.3)	39.0 (1.7)
SCB-m	23.6 (0.9)	12942.2 (81.2)	2.29 (0.05)	403.8 (2.5)	178.1 (4.1)	39.1 (1.7)
SCB-m2	23.7 (0.8)	12940.4 (80.5)	2.30 (0.05)	403.7 (2.6)	178.0 (4.5)	38.8 (1.7)

[1] Scenarios were: C-h, C-m: conventional semen used for all heifers and cows; SB-h: sexed semen used on all heifers, top 45% for 1st lactation cows, and top 10% on 2nd lactation cows, beef semen used on other cows; SCB-h: sexed semen used on all heifers, top 25% 1st lactation cows; conventional semen used on next top 25% for 1st lactation cows and top 35% for second lactation cows; beef semen used for other cows; SB-m: 5% more on 2nd comparing to SB-h; SCB-m: 5% more on the conventional semen used for both 1st and 2nd lactation cows than SCB-h. SCB-h2, SCB-m2: limited sexed semen used on 1st lactation cows on first 3 inseminations, and conventional semen used for later inseminations comparing to SCB-h, SCB-m. The high reproductive level (C-h, SB-h, SCB-h, SCB-h2): conception rate was set at 60% for 1st insemination on heifers, 55% for 1st insemination of 1st lactation cows. The moderate reproductive level (C-m, SB-m, SCB-m, SCB-m2) conception rate was set at 55% for 1st insemination on heifers, 50% for 1st lactation 1st insemination cows.

for all scenarios are shown in Figure 3.7. All scenarios had a similar herd size and number of milking cows with a slight variance. We designed the simulation structure to maintain a stable herd size. As expected, the number of pregnant cows for high productive performance scenarios (C-h, SB-h, SCB-h, and SCB-h2) are greater than that of moderate reproductive performance scenarios. Within the same reproductive performance, the C-h scenario has the fewest young animals amongst all scenarios. The C-h scenario also has the highest pregnancy rate for heifers, which led to fewer heifer reproduction-failure cullings and fewer young animals. In the SB-h scenario, the milking herd was composed of slightly younger animals (more 1st lactation and fewer later lactation cows). This reflects a slightly more intensive culling, mostly due to reproductive failures.

The scenarios with high reproductive performance had a higher 21-day pregnancy rate (27.4% vs. 23.5%), shorter average calving interval (398.5 d vs. 403.8 d), and shorter average DIM (173.7 d vs. 178.5 d) than the scenarios with moderate reproductive performance.

Culling summary. Figure 3.7(c) illustrates the summaries of culling reasons for the recorded simulation year. The distributions of death and health culling reasons are similar between all scenarios. The reproductive culling rate, however, did vary across scenarios. In general, the scenarios with higher use of sexed semen had higher reproductive failure culling due to the lower conception rate of sexed semen.

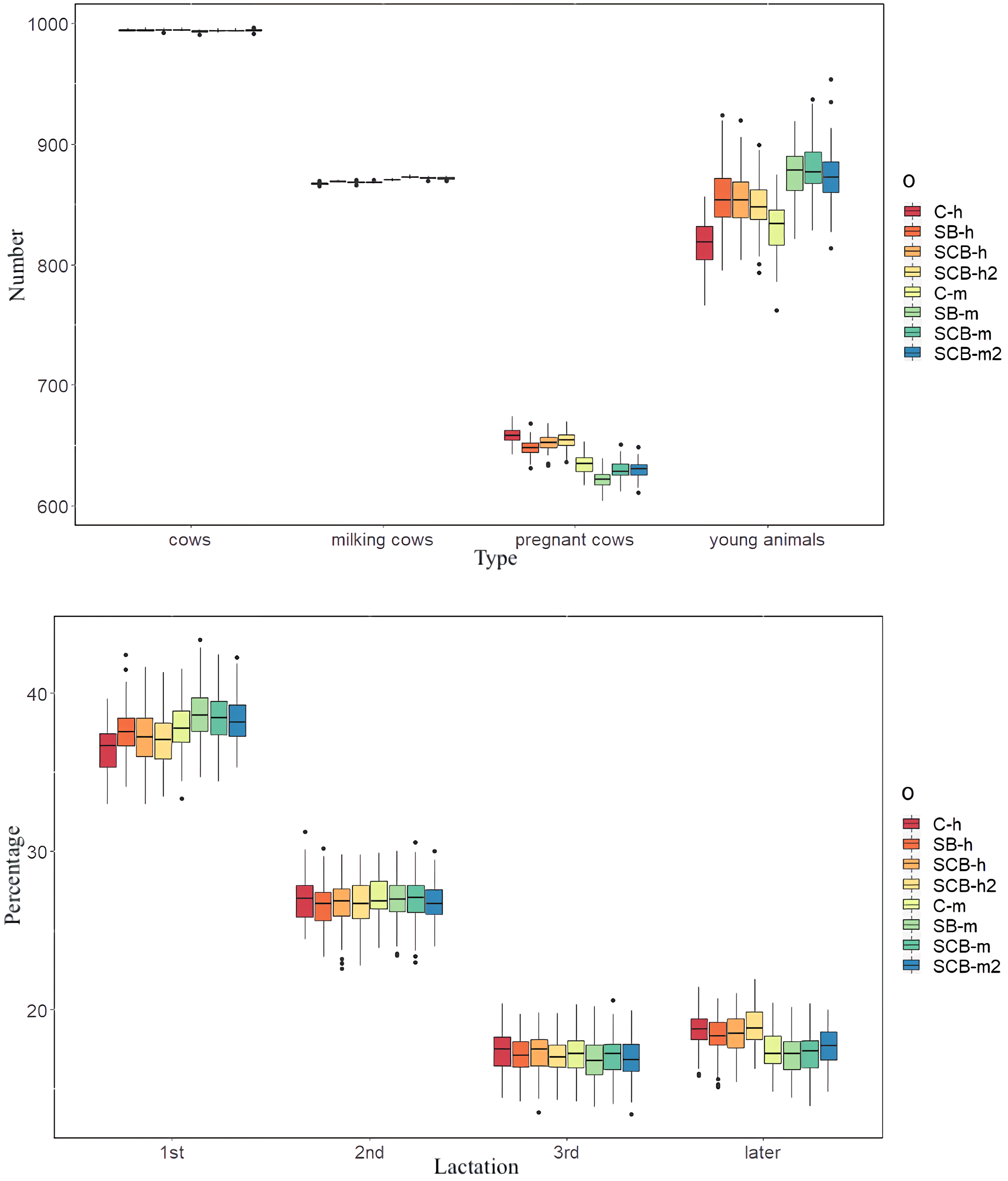
The destinations of calves born on the farm during the last 365 d of the 7-yr simulation for all scenarios are shown in Table 3.6. Compared with the C-h and C-m, more of the calves raised in other scenarios came from heifers or young lactation cows, which could be shortening the generation interval in the herd (Sørensen et al., 2011) and promoting genetic progress (Ettema et al., 2017).

3.5.3 Limitations and future steps

The animal life cycle submodel described here is the foundation of a dairy herd simulation, enabling future additions and connections with other RuFaS modules. We used the animal life cycle submodel to compare the use of conventional semen with a combination of sexed and beef semen. One limitation of the case study is that we did not include a formal representation of genetic progress. We used the simplified, arbitrary production levels that were randomly assigned to the initial herd and assumed the production level was inherited from the dam to the calf. The evaluation of semen use would be more valuable if an accurate representation of genetic inheritance and progress is included in the model for both dam and sire (Hjortø et al., 2015; Ettema et al., 2017).

3.6 Conclusions

The animal life cycle submodel in the RuFaS model was developed as a stochastic simulation model of the dairy herd. The submodel's input can be adjusted based on research needs and actual farm characteristics. Its output delivers multi-level information from the simulation, including herd dynamics, reproductive performance, economic indications, individual animal life stories, and culling summaries. We used the submodel to assess the economic benefit of using different combinations of sexed and beef semen compared to using conventional semen only. Results suggest that using a combination of sexed and beef semen is economically beneficial. When a herd's reproductive performance is high, using sexed semen combined with conventional semen resulted in similar economic outcomes to only using sexed and beef semen. When the herd reproductive performance was moderate, using sexed and beef semen had a higher NR to combining the 3 types of semen. A breakeven analysis revealed that the benefit of using sexed and beef semen would remain, in the high reproductive farms, if the beef crossbred calves' price is about 3 times higher than that of Holstein male calves. The life cycle submodule described here is part of a flexible, open-source whole-farm modeling framework built for continuous adaptation by the research community to solve real-life problems for farmers and researchers.


1000
900
800
700
600
Number
cows
milking cows
pregnant cows
young animals
Type
O
C-h
SB-h
SCB-h
SCB-h2
C-m
SB-m
SCB-m
SCB-m2
40
30
20
Percentage
1st
2nd
3rd
later
Lactation
O
C-h
SB-h
SCB-h
SCB-h2
C-m
SB-m
SCB-m
SCB-m2

(A)

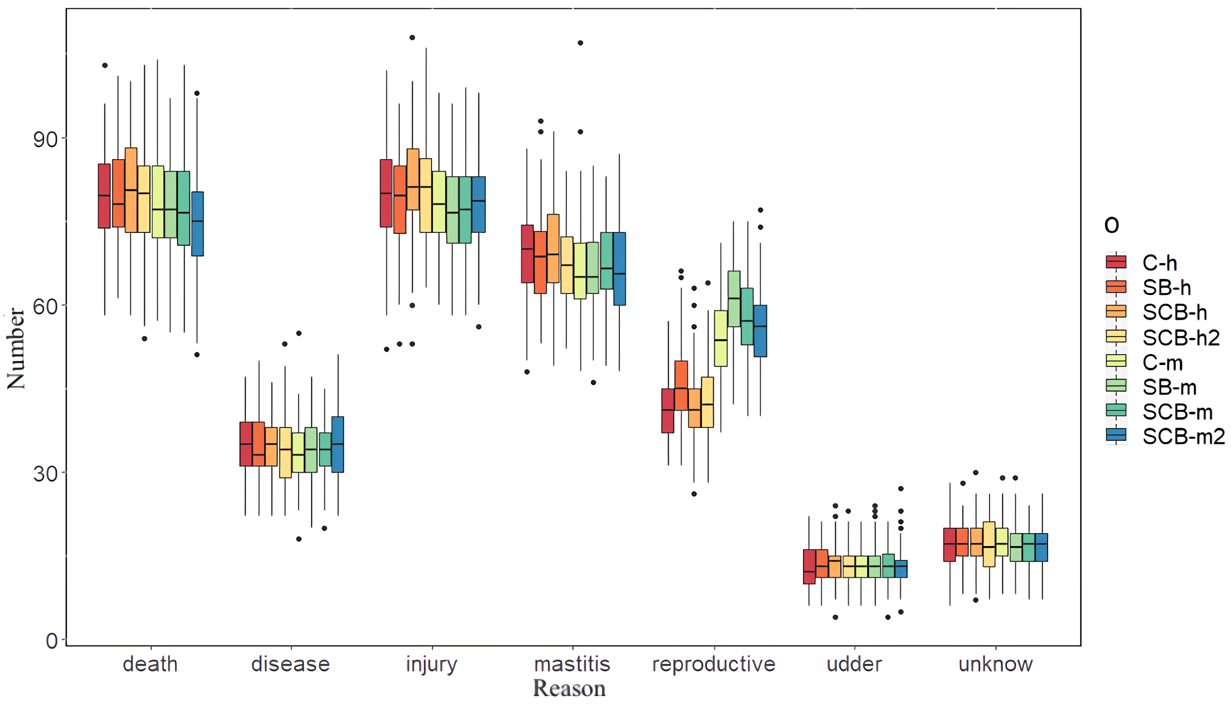


FIGURE 3.7: Box-and-whisker plot (median, first and third percentiles, range) of herd structure numbers (top), percentages (medium), and the number of animals culled for each reason (bottom) during the last year of the 7-year simulation with 100 iterations for each study scenario. Scenarios were: C-h, C-m: conventional semen used for all heifers and cows; SB-h: sexed semen used on all heifers, top 45% for 1st lactation cows, and top 10% on 2nd lactation cows, beef semen used on other cows; SCB-h: sexed semen used on all heifers, top 25% 1st lactation cows; conventional semen used on next top 25% for 1st lactation cows and top 35% for second lactation cows; beef semen used for other cows; SB-m: 5% more on 2nd comparing to SB-h; SCB-m: 5% more on the conventional semen used for both 1st and 2nd lactation cows than SCB-h. SCB-h2, SCB-m2: limited sexed semen used on 1st lactation cows on first 3 inseminations, and conventional semen used for later inseminations compared to SCB-h, SCB-m. The high reproductive level (C-h, SB-h, SCB-h, SCB-h2): conception rate was set at 60% for 1st insemination on heifers, 55% for 1st insemination of 1st lactation cows. The moderate reproductive level (C-m, SB-m, SCB-m, SCB-m2) conception rate was set at 55% for 1st insemination on heifers, 50% for 1st lactation 1st insemination cows.

References

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams (1998). LARGE AREA HYDROLOGIC MODELING AND ASSESSMENT PART I: MODEL DEVELOPMENT. In: *J Am Water Resour As* 34, pp. 73–89. DOI: `10.1111/j.1752-1688.1998.tb05961.x`.

Barrientos-Blanco, J. A., H. White, R. D. Shaver, and V. E. Cabrera (2020). Improving nutritional accuracy and economics through a multiple ration-grouping strategy. In: *J Dairy Sci* 103, pp. 3774–3785. DOI: `10.3168/jds.2019-17608`.

Baudracco, J., N. Lopez-Villalobos, C. W. Holmes, E. A. Comeron, K. A. Macdonald, and T. N. Barry (2012). e-Dairy: a dynamic and stochastic whole-farm model that predicts biophysical and economic performance of grazing dairy systems. In: *Animal Int J Animal Biosci* 7, pp. 870–8. DOI: `10.1017/s1751731112002376`.

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. In: *J Dairy Sci* 95, pp. 4683–4698. DOI: `10.3168/jds.2011-5214`.

Calsamiglia, S., S. Astiz, J. Baucells, and L. Castillejos (2018b). A stochastic dynamic model of a dairy farm to evaluate the technical and economic performance under different scenarios. In: *J Dairy Sci* 101, pp. 7517–7530. DOI: `10.3168/jds.2017-12980`.

Cottle, D. J., M. Wallace, P. Lonergan, and A. G. Fahey (2018b). Bioeconomics of sexed semen utilization in a high-producing Holstein-Friesian dairy herd. In: *J Dairy Sci* 101, pp. 4498–4512. DOI: `10.3168/jds.2017-13172`.

Dawod, A. and H. T. Elbaz (2020). Effect of sexed semen, puberty and breeding ages on fertility of Holstein dairy heifers treated with double Ovsynch protocol. In: *Trop Anim Health Pro*, pp. 1–6. DOI: `10.1007/s11250-020-02306-6`.

DCHA (2012). Production and Performance Standards Established for Holstein Calves, from Birth to 6 Months of Age, across the United States. Accessed. eprint: `GoldStandardsI.`. URL: `http://calfandheifer.org/gold_standards/index.php`.

DCRC (2018). Dairy Reproduction Protocols. Accessed. URL: `https://www.dcrcouncil.org/protocols/`.

De Vries, A., M. B. Crane, J. A. Bartolome, P. Melendez, C. A. Risco, and L. F. Archbald (2006). Economic Comparison of Timed Artificial Insemination and Exogenous Progesterone as Treatments for Ovarian Cysts 1. In: *J Dairy Sci* 89, 3028–3037. DOI: `10.3168/jds.s0022-0302(06)72576-0`.

DeJarnette, J. M., R. L. Nebel, and C. E. Marshall (2009). Evaluating the success of sex-sorted semen in US dairy herds from on farm records. In: *Theriogenology* 71, pp. 49–58. DOI: `10.1016/j.theriogenology.2008.09.042`.

Dhakal, K., C. Maltecca, J. P. Cassady, G. Baloche, C. M. Williams, and S. P. Washburn (2013). Calf birth weight, gestation length, calving ease, and neonatal calf mortality in Holstein, Jersey, and crossbred cows in a pasture system. In: *J Dairy Sci* 96, pp. 690–698. DOI: `10.3168/jds.2012-5817`.

Ettema, J. F., J. R. Thomasen, L. Hjortø M. Kargo, S. Østergaard, and A.C. Sørensen. (2017b). Economic opportunities for using sexed semen and semen of beef bulls in dairy herds. In: *J Dairy Sci* 100, pp. 4161–4171. DOI: `10.3168/jds.2016-11333`.

Extenion, Penn State (2020). Calculating Heifer Numbers. Accessed May 21, 2020. URL: `https://extension.unh.edu/resource/calculating-heifer-numbers`.

Galvão, K. N., P. Federico, A. De Vries, and G. M. Schuenemann (2013). Economic comparison of reproductive programs for dairy herds using estrus detection, timed artificial insemination, or a combination. In: *J Dairy Sci* 96, pp. 2681–2693. DOI: `10.3168/jds.2012-5982`.

Giordano, J. O., A. S. Kalantari, P. M. Fricke, M. C. Wiltbank, and V. E. Cabrera (2012b). A daily herd Markov-chain model to study the reproductive and economic impact of reproductive programs combining timed artificial insemination and estrus detection. In: *J Dairy Sci* 95, pp. 5442–5460. DOI: `10.3168/jds.2011-4972`.

Hansen, T., L. M. Li, J. Li, C. J. Vankerhove, M. A. Sotirova, J. M. Tricarico, V. E. Cabrera, E. Kebreab, and K. F. Reed (2021a). The Ruminant Farm Systems Animal Module: A Biophysical Description of Animal Management. In: *Animals* 11.5, p. 1373.

Heinrichs, A. J., G. I. Zanton, G. J. Lascano, and C. M. Jones (2017). A 100-Year Review: A century of dairy heifer research. In: *J Dairy Sci* 100, pp. 10173–10188. DOI: `10.3168/jds.2017-12998`.

Hjortø, L., J. F. Ettema, M. Kargo, and A. C. Sørensen (2015). Genomic testing interacts with reproductive surplus in reducing genetic lag and increasing economic net return. In: *J Dairy Sci* 98, pp. 646–658. DOI: `10.3168/jds.2014-8401`.

Johnson, I. R., D. F. Chapman, V. O. Snow, R. J. Eckard, A. J. Parsons, M. G. Lambert, and B. R. Cullen (2008b). DairyMod and EcoMod: biophysical pasture-simulation models for Australia and New Zealand. In: *Anim Prod Sci* 48, pp. 621–631. DOI: `10.1071/ea07133`.

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

Kebreab, E., K. F. Reed, V. E. Cabrera, P. A. Vadas, G. Thoma, and J. M. Tricarico (2019a). A new modeling environment for integrated dairy system management. In: *Animal Frontiers* 9.2, pp. 25–32. DOI: `10.1093/af/vfz004`.

Korver, S., J. A. M. van Arendonk, and W. J. Koops (1985). A function for live-weight change between two calvings in dairy cattle. In: *Anim Sci* 40, pp. 233–241. DOI: `10.1017/s0003356100025332`.

Larman, C. and V. R. Basili (2003). Iterative and incremental developments. a brief history. In: *Computer* 36, pp. 47–56. DOI: `10.1109/mc.2003.1204375`.

Li, M., V. E. Cabrera, and K. F. Reed (2018c). Comparison of Holstein and Jersey milk production with a new stochastic animal model. In: *Journal of Dairy Science 101* suppl.2, p. 79.

Li, M., V. E. Cabrera, and K. F. Reed (2019). Updating Holstein and Jersey lactation curve parameters for the Rumination Farm System Model (RuFaS). In: *Journal of Dairy Science* Vol. 102, Suppl, 1:54.

Li, W. and V. E. Cabrera (2019). *Dairy x Beef: Fad or Sustainable Future? Page Proc*. Pittsburgh, PA: Dairy Cattle Reproduction Council Conf.

Linden, T. C., R. C. Bicalho, and D. V. Nydam (2009). Calf birth weight and its association with calf and cow survivability, disease incidence, reproductive performance, and milk production. In: *J Dairy Sci* 92, pp. 2580–2588. DOI: `10.3168/jds.2008-1603`.

Martin, O. and D. Sauvant (2010). A teleonomic model describing performance (body, milk and intake) during growth and over repeated reproductive cycles throughout the lifespan of dairy cattle. 1. In: *Trajectories of life function priorities and genetic scaling* Animal 4, pp. 2030–2047. DOI: `10.1017/s1751731110001357`.

McCullock, K., D. L. K. Hoag, J. Parsons, M. Lacy, G. E. Seidel, and W. Wailes (2013). Factors affecting economics of using sexed semen in dairy cattle. In: *J Dairy Sci* 96, pp. 6366–6377. DOI: `10.3168/jds.2013-6672`.

Mur-Novales, R. M. and V. E. Cabrera (2017). *WHAT TYPE OF SEMEN SHOULD I USE*. Reno, NV: Page. Proceedings Dairy Cattle Reproduction Council Annual Convention.

Nehls, N. (2019). Technological Advances and Economics: How Farm Management Is Changing Because of It. In: *Accessed*. URL: `https://dairy.agsource.com/2019/05/16/technological-advances-and-economics-how-farm-management-is-changing-because-of-it/`.

Nielsen, P. P., G. Pettersson, K. M. Svennersten-Sjaunja, and L. Norell (2010). Technical note: Variation in daily milk yield calculations for dairy cows milked in an automatic milking system. In: *J Dairy Sci* 93, pp. 1069–1073. DOI: `10.3168/jds.2009-2419`.

NRC (2001b). *Nutrient Requirements of Dairy Cattle*. 7th rev. ed. Washington, DC.: The National Academies Press.

Pinedo, P. J., A. De Vries, and D. W. Webb (2010). Dynamics of culling risk with disposal codes reported by Dairy Herd Improvement dairy herds. In: *J Dairy Sci* 93, pp. 2250–61. DOI: `10.3168/jds.2009-2572`.

Rossum., G. V. and F. L. Drake (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Rotz, C. A., M. S. Corson, D. S. Chianese, F. Montes, S. D. Hafner, H. F. Bonifacio, and C. U. Coiner (2015a). *The Integrated Farm System Model Reference Manual*. Vol. version 4.

Rubinstein, R. Y. and D. P. Kroese (2017b). *Simulation and the Monte Carlo Method*. Ser Probab Statistics 91–106: Wiley. DOI: `10.1002/9781118631980.ch3`.

Río, N. S. del, B. W. Kirkpatrick, and P. M. Fricke (2006). Observed frequency of monozygotic twinning in Holstein dairy cattle. In: *Theriogenology* 66, pp. 1292–1299. DOI: `10.1016/j.theriogenology.2006.04.013`.

Schils, R. L. M., M. H. A. de Haan, J. G. A. Hemmer, A. van den Pol-van Dasselaar, J. A. de Boer, A. G. Evers, G. Holshof, J. C. van Middelkoop, and R. L. G. Zom (2007c). DairyWise, A Whole-Farm Dairy Model. In: *J Dairy Sci* 90, pp. 5334–5346. DOI: `10.3168/jds.2006-842`.

Schils, R. L. M., M. H. A. de Haan, J. G. A. Hemmer, A. van den P. Dasselaar, J. A. de Boer, A. G. Evers, G. Holshof, J. C. van Middelkoop, and R. L. G. Zom (2007d). DairyWise, A Whole-Farm Dairy Model. In: *J Dairy Sci* 90, 5334–5346. DOI: `10.3168/jds.2006-842`.

Soberon, F., E. Raffrenato, R. W. Everett, and M. E. V. Amburgh (2012). Preweaning milk replacer intake and effects on long-term productivity of dairy calves. In: *J Dairy Sci* 95, 783–793. DOI: `10.3168/jds.2011-4391`.

Sørensen, M. K., J. Voergaard, L. D. Pedersen, P. Berg, and A. C. Sørensen (2011). Genetic gain in dairy cattle populations is increased using sexed semen in commercial herds: Sexed semen in dairy cattle breeding. In: *J Anim Breed Genet* 128, 267–275. DOI: `10.1111/j.1439-0388.2011.00924.x`.

USDA (2020). Agricultural Projections to 2029. In: *Accessed.* URL: `https://www.ers.usda.gov/webdocs/outlooks/95912/oce-2020-1.pdf?v=9508.8`.

Vaillant, J. and L. Baldinger (2016). Application note: An open-source JavaScript library to simulate dairy cows and young stock, their growth, requirements and diets. In: *Comput Electron Agr* 120, 7–9. DOI: `10.1016/j.compag.2015.11.005`.

Weigel, K. A. (2004). Exploring the Role of Sexed Semen in Dairy Production Systems. In: *J Dairy Sci* 87, E120–E130. DOI: `10.3168/jds.s0022-0302(04)70067-3`.

Winter, R. J. (2014). Agile Software Development: Principles, Patterns, and Practices. In: *Perform Improv* 53, 43–46. DOI: `10.1002/pfi.21408`.

Chapter 4

Leveraging a large database of test day milk production records to investigate the effects of temporal, spatial, and management factors on lactation curves of US Holstein cattle

M. Li,* G. J. M. Rosa,* K. F. Reed,[1] **V. E. Cabrera***

* Department of Animal and Dairy Sciences, University of Wisconsin-Madison, Madison WI 53705

† Department of Animal Science, Cornell University, 272 Morrison Hall, Ithaca, NY 14850;

4.1 Abstract

The Wood's model was fit to an extensive database of test-day milk production records of US Holstein cows. Individual-specific parameters' estimates were further analyzed to investigate the effects of temporal factors: calving year (2006 to 2016), calving month (January to December); spatial factor: farm region (Northeast, Appalachian, Southeast, Lake States, Corn Belt, Delta states, Northern plains, Southern plains, Mountain, West Coast, Wisconsin, Pennsylvania, or New York); and management factors: milking frequency (2 or 3 times a day), age at first calving for 1^{st} lactation), and parity (1^{st}, 2^{nd}, and later lactations) on lactation curve shape. The analysis approach included 2 steps: (1) individual animal-parity parameter estimation with a non-linear least-squares optimization algorithm, and (2) mixed effects model analysis of parameter estimates on 8,595,413 sets of individual lactation curve parameters of temporal, spatial, and management factors as the fixed effects, and herd and animal as random effects. Overall, significant effects of all factors were found for the Wood's parameters and 305-d milk yield. Results showed parity was the most significant fixed effect on parameter a, parameter b, and 305-d milk yield; the month was on parameter b for all-lactations models. The month was the most significate fixed effect on all parameters for 1^{st} lactation models, in which age at calving, year, and milking frequency accounted for a higher proportion of the variance than month on 305-d milk yield. All parameter estimates and 305-d milk yield increased as parity increased; parameter *a* and 305-d milk yield rose as year and milking frequency increased; however, parameters *b* and *c* decreased. Month-to-month estimations showed the lowest values for parameter *a* in September, parameter *b* in May, parameter *c* in June, and 305-d milk yield in July. The results suggested that the random effects enhanced the model's prediction ability, while the fixed effects effectively identified the influence of each studied factors on lactation curve parameters and 305-d milk yield estimates. In conclusion, lactation curve parameter estimates from the mixed effects models well described lactation curves in terms of temporal, spatial, and management factors.

Keywords: Lactation curve, production, Holstein

4.2 Introduction

Lactation curves are expressed as mathematical equations that enable milk yield predictions at the farm or individual level. Parametric models such as Wood's model (Wood, 1967) have been widely applied to describe the behavior of milk production during the lactation period (Bouallegue and M'Hamdi, 2020). Wood's model is an incomplete gamma function for fitting the observed daily milk yield to a parameterized curve. The physiologies behind this shape were explained to be related to cell division and cellular differentiation (Knight, 2000; Capuco et al., 2001) or gene expressions (Atashi et al., 2019).

Understanding the shape of the lactation curve provides insight for making wise management decisions and implementing effective production strategies for farms. The lactation curve's shape is described as having an ascending phase leading up to the peak and a descending phase following the peak with lactations curve parameters. Parameters of lactation curve showed the milk productivity across the lactation and proved to be valuable for yield monitoring, early diseases detection, feed formulation, breeding selection. They have been included into a variety of farm management tools (De Vries, 2006). The use of lactation-curve models to predict target milk yield was also a practical method to describe cow and herd milk production without previous milk production data (Rotz et al., 2013; Calsamiglia et al., 2018) with a small number of parameters (Macciotta et al., 2011). Fitted lactation curve parameters can assist the farm management decision-making process. For example, it can be used to represent baseline estimations in simulation models (Ben Abdelkrim et al., 2019), which can then be adjusted to account for the cow-specific environmental or genetic impacts such as health issues, diet changes, or genetic progress.

However, the lactation curve parameters are affected by the cumulative effect of several factors. For example, the geographical region where the cow is located (Cole et al., 2011), the season of the year when the lactation starts (Torshizi, 2016), milking frequency (Stelwagen, 2001), or age at first calving for 1^{st} lactation (Atashi et al., 2019) have all been reported to alter the shape and scale of lactation curves. Thus, obtaining precise lactation curve parameters for practical use remains a challenge.

Consequently, we aim to:

1. analyze the effect of calving year (2006 to 2016), calving month (January to December), farm region (Northeast, Appalachian, Southeast, Lake States, Corn Belt, Delta states, Northern plains, Southern plains, Mountain, West Coast, Wisconsin, Pennsylvania, or New York), milking frequency (2 or 3 times per day), parity (1^{st}, 2^{nd}, or later lactations), and age at calving (for 1^{st} lactation) on lactation curve parameter estimates;

2. estimate lactation curve variances among herds and cows;

3. characterize parity, year, month, region, and milking frequency-specific lactation curve parameters to be used in applications like simulation models.

4.3 Materials and Methods

4.3.1 Data

The dataset used in this study was provided by the US Council on Dairy Cattle Breeding (`https://www.uscdcb.com/`) and contained individual test-day records for 10.15 million lactations from US Holstein cows located in 45 states from January 2006 to December 2016. Each lactation was composed of at least 10 test-day records and the calving date associated with the beginning of the lactation.

4.3.2 Lactation-specific curve fitting

Lactation curves are mathematical equations that describe milk yield as a function of days in milk (DIM; a.k.a., days postpartum). The Wood's model (Wood, 1967), the most commonly used parametric model for describing lactation curves (Dijkstra et al., 2010; Bouallegue and M'hamdi, 2019), is a gamma function with 3 parameters (a, b, and c) for the milk production as

$$y = at^b e^{-ct} + \varepsilon \tag{4.1}$$

where y is the observed daily milk yield (kg), t is the number of days postpartum (or DIM), parameter a is the scale factor for the production level, b indices the growth rate in milk yield until peak production, c describes the decline rate after the peak, and ε is the random residual (Wood, 1967; Wood, 1972). We used non-linear least square method to fit each set of individual lactation test-day records to Wood's function using the nlme v3.1-152 package (Pinheiro et al., 2021) in R v4.0.3 (R Core Team, 2020) and RStudio v1.3.1093 (RStudio Team, 2020). As the result of the non-linear fitting procedure, we collected the lactation-specific estimats for a, b, c, and the residual sum of squares (RSS, a measure of goodness of fit) as the non-linear fitting process output. Due to potential test-day data errors and outliers, some of the resulting parameter estimates did not follow the typical shape of lactation curves. The following criteria were used to classify estimated lactation curves as atypical:

1. negative values for the b or c parameters, which would reverse the shape of the typical lactation curve (Wood,1976; Macciotta, 2011);
2. the peak of production occurring after 300 DIM;
3. RSS greater than 1000, which was found to produce unreliable fitting results;
4. parameters a, b, and c being greater than 3 standard deviations from their means, $log(a)$, $log(b)$, and $log(c)$ being less than -5 standard deviations, or 305-d milk yield being greater than 5 standard deviations from their respective means.

We excluded such atypical lactation curves, resulting in 8,599,526 lactations retained for further analysis. The cumulative 305-d milk yield, commonly used to compare lactation performance, was calculated as:

$$a \int_{1}^{305} t^{b} e^{-ct} dt \tag{4.2}$$

Which represents the sum of estimated daily production during the first 305 days of lactation. Figure 4.1 shows the distributions of lactation-specific model parameters and 305-d milk yield values for the subset of lactations retained for further analyses.

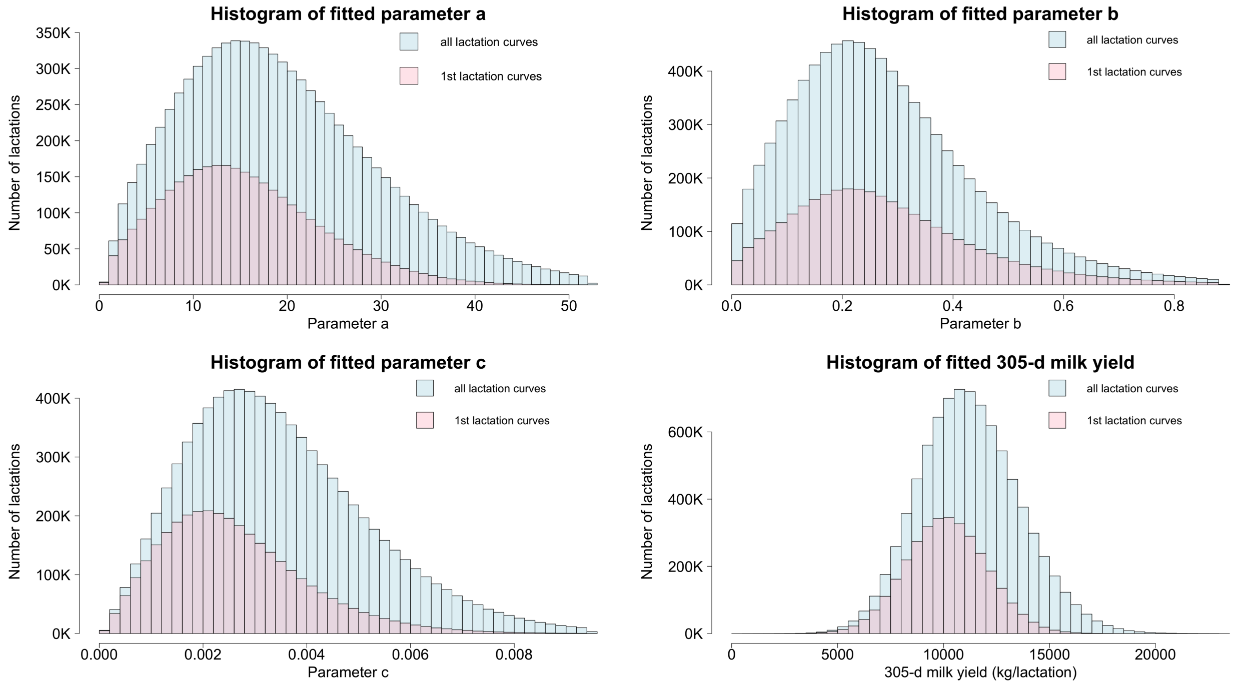


FIGURE 4.1: Histograms of fitted individual lactation curve parameters *a*, *b*, *c*, and 305-d milk yield for all-lactations models and the 1^{st} lactation models

4.3.3 Linear mixed effects models of lactation curve parameters and 305-d milk yield

We used parameter estimates from fitted individual lactation curves to investigate the effects of different factors on the lactation curve shape and 305-d milk yield, including spatial (farm region), temporal (calving year, and calving month), and management (milking frequency, age at calving for 1^{st} lactation, and parity) components. We categorized the 45 states available in our dataset into 10 regions in accordance with the USDA convention (USDA-ERS, 2021) to explore the location effect. The regions included West Coast, Mountain, Northern Plains, Southern Plains, Lake States, Corn Belt, Delta States, Appalachian, Southeast, and Northeast. We then separated records from Wisconsin, Pennsylvania, and New York (the 3 states with the most lactation records in our dataset) from the Lake States and the Northeast (Figure 4.2, Table 4.1). Some states on the

West Coast and Mountain regions that are large dairy production states (e.g., California and Idaho; USDA-ERS, 2021) were not separated from their region because they were not as well represented in our dataset.

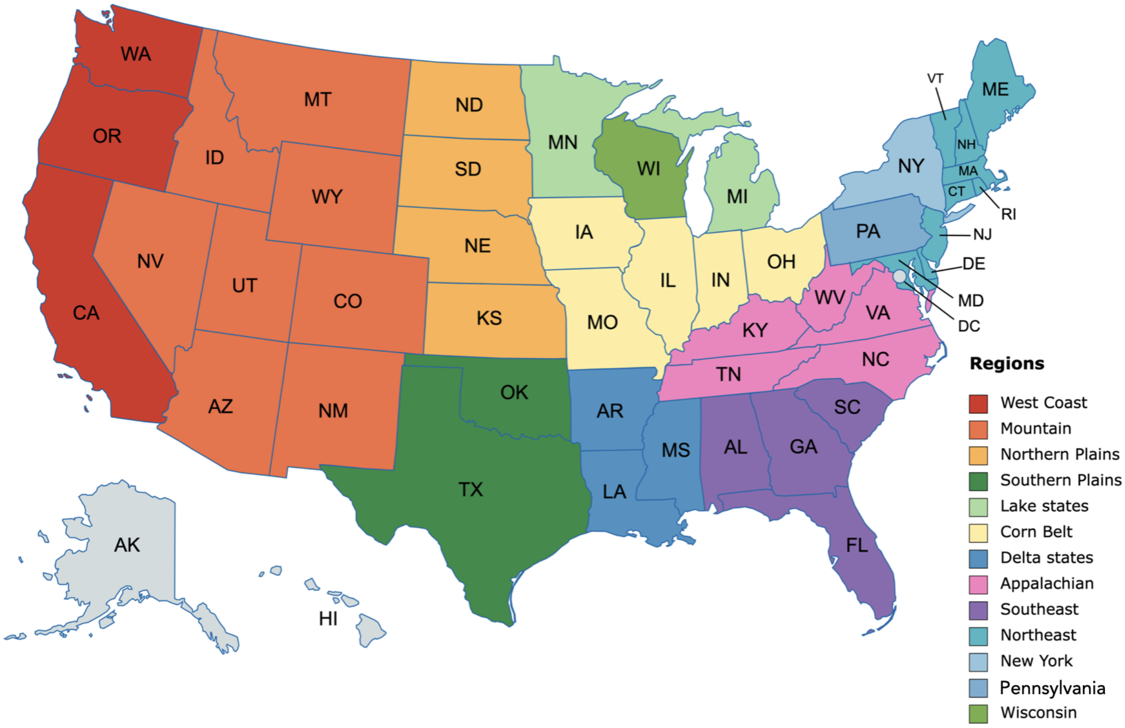


FIGURE 4.2: Map with different regions in the study

Linear mixed effects models were used to account for correlated data associated with numerous lactations by the same animal and lactation within the same herd. To ensure that the herd was treated as a random variable in our analysis, we further excluded 4,426 lactations from herds with fewer than 5 lactations, remaining 8,595,100 lactations (from a total of 21,056 herds) and their corresponding parameter estimations.

To assess the impacts of spatial, temporal, and management factors on the lactation curve parameters (*a*, *b*, *c*) and 305-d milk yield, we used the following mixed effects model:

$$y_{ijklmnpq} = \mu + Parity_i + Year_j + Month_k + Region_l + Milkingfreq_m + Animal_n + herd_p + \varepsilon_{ijklmnpq} \tag{4.3}$$

where $y_{ijklmnpq}$ is the response variable (lactation curve parameter *a*, *b*, *c*, or 305-d total milk yield) related to the q^{th} lactation of the n^{th} animal record in p^{th} herd from the ith parity (3 groups, 1^{st}, 2^{nd}, and 3^{rd} and later lactations), in the j^{th} calving year (11 years, from 2006 to 2016), the k^{th} month of calving (12 months, from January to December), the l^{th} region (13 levels, as described above), and at the m^{th} milking frequency (2 levels, 2 or 3 times a day). The components $Parity_i + Year_j + Month_k + Region_l + Milkingfreq_m$ and refer to the fixed effects of parity, calving year, calving month, US region where the herd was located, and daily milking frequency, respectively. In addition, $Herd_p$ ($n = 21,056$) was the random effect of the herd, and $Animal_m$ ($n = 5,375,200$) was the random effect of the animal. Lastly, $\varepsilon_{ijklmnpq}$ represented the residual term, with expectation 0 and variance equal to σ^2. Table 4.1 reported the number of lactations for each factor level. Equation 4.3 was fit to all parameter estimates from the selected subset of individual lactation curves (8,599,526 lactations).

We fit the following model separately for 1^{st} lactation curves only:

$$y_{ijklmnp} = \mu + AgeAtCalving_i + Year_j + Month_k + Region_l + Milkingfreq_m + Herd_n + \varepsilon_{ijklmnq} \tag{4.4}$$

The model described in Equation 4.4 was fit to parameters from curves of the 1^{st} lactation only (3,324,803 lactations) and was similar to the model for all lactations except that we replaced the effect of parity by the effect of age at calving. The corresponding variable, $AgeAtCalving_i$, was a continuous variable with a median of 746 d, a mean value at 764 d, and a standard deviation of 84 d (Figure 4.3). This model included only herd as a random effect ($n = 20,509$).

TABLE 4.1: Number of lactations in each level of factors in the linear mixed model for all-lactations models.

Milking frequency		Parity		Year		Month		Region[1]	
2/d	4,772,772	1	3,324,803	2006	710,337	Jan	737,857	WI	2,366,999
3/d	3,822,328	2	2,425,790	2007	775,778	Feb	627,922	PA	1,418,826
		>=3	2,844,507	2008	773,511	Mar	661,400	NY	1,217,061
				2009	754,638	Apr	567,234	Lake	1,148,390
				2010	782,829	May	608,170	Corn Belt	905,891
				2011	777,104	Jun	668,246	Northeast	386,526
				2012	783,874	Jul	737,861	Appalachian	347,466
				2013	797,748	Aug	758,073	Northern Plains	226,274
				2014	820,567	Sep	812,506	Southeast	191,887
				2015	804,117	Oct	820,951	Southern Plains	129,999
				2016	814,597	Nov	784,688	Mountain	127,054
						Dec	811,192	West Coast	99,077
								Delta	29,650

[1] Region: Following the USDA convention https://www.ers.usda.gov/data-products/dairy-data/

We conducted the linear mixed effects model analyses using R (R Core Team, 2020) and lme4 (Bates et al., 2015). We investigated the effect of each factor on each response variable by examining the average fitted parameter estimates for each factor level and the ANOVA tables for the linear mixed effects models. Additionally, we used likelihood ratio tests to test the significance of including random effects. Furthermore, we report the coefficients of mixed effects models using parity-, year-, month-, region -, and milking frequency-specific lactation curve parameters for estimating for further applications.

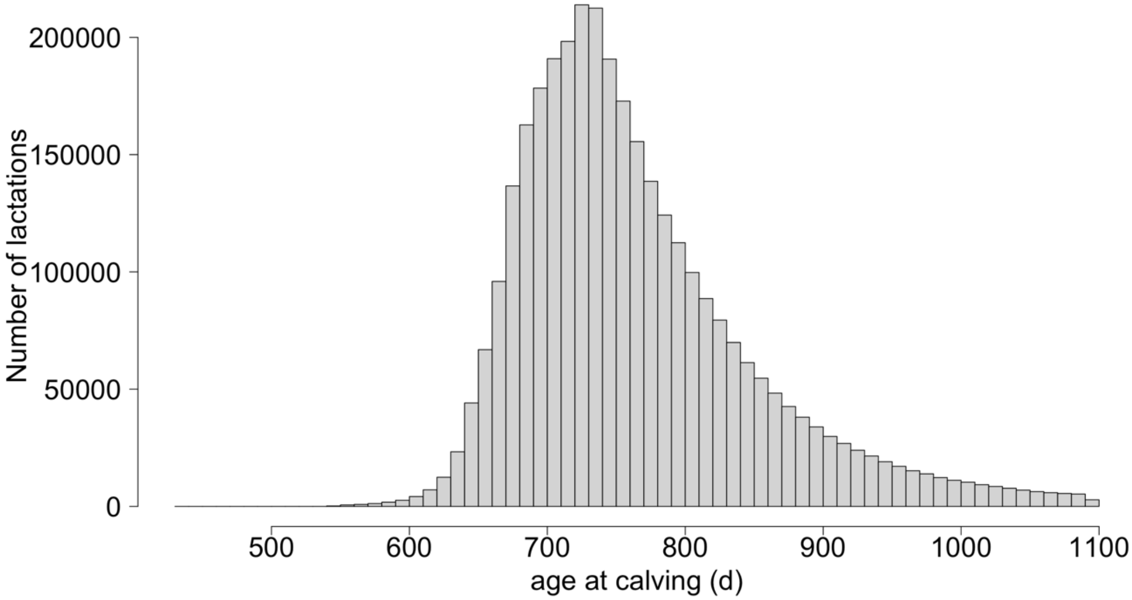


FIGURE 4.3: Histogram of number of lactations for the age at 1^{st} calving

4.4 Results

Visual inspection of residual plots did not reveal any apparent deviations from homoscedasticity or normality. The fixed effects of parity (or age at calving), year and month of calving, region, and milking frequency were statistically significant ($P < 0.05$) on both models (all-lactations and 1^{st} lactation only), except for the effect of milking frequency on the 1^{st} lactation curve model of parameter b (Table 4.2 and 4.3). The random effects of herd and animal were statistically significant for all-lactations models and herd for 1^{st} lactation models from likelihood ratio tests ($P < 0.05$).

For parameter a, the sum of square portions of the total sum of squares were ranked as: lactation group, month, year, milking frequency, and region; for parameters b and c, the milking frequency had the least sum of squares, and for parameter b, the month had the most considerable sum of squares. Ranking of the proportion of the sum of squares to the total fixed effects sum of squares for 305-d milk yield was similar to that for parameter *a* except that the year was ranked higher than the month (Table 4.2). The 1^{st} lactation mixed effects models included the age at first calving, and this explanatory variable explained the largest proportion of the variability in 305-d milk yield and parameter *a* (Table 4.3).

TABLE 4.2: The residual sum of squares and proportion of variance of the fixed effects in mixed models on the lactation curve parameters and 305-d milk yield for all-lactations models

Fixed effects[a]	DF	Parameter a		Parameter b		Parameter c		305-d milk yield	
		Sum sq	Proportion of variance[b] (%)	Sum sq	Proportion of variance[b] (%)	Sum sq	Proportion of variance[b] (%)	Sum sq	Proportion of variance[b] (%)
Lactation group	2	$7.2x10^7$	90.32	927.4	33.82	3.49	91.82	$5.11x10^{12}$	93.20
Year	10	$9.3x10^5$	1.17	178.8	6.52	0.03	0.79	$1.44x10^{11}$	2.63
Month	11	$6.1x10^6$	7.65	1627.0	59.33	0.28	7.37	$1.26x10^{11}$	2.30
Region	12	$8.0x10^4$	0.10	8.0	0.29	0.0006	0.02	$8.51x10^8$	0.02
Milking frequency	1	$6.1x10^5$	0.77	0.9	0.03	0.0002	0.01	$1.02x10^{11}$	1.86

[a] $p < 0.0001$ for all F tests

[b] Proportion of total fixed effects variance

Tables 4.4 and 4.5 summarized the estimates and standard errors for each coefficient in the mixed models. To reconstruct estimated lactation curves for specific categories, we used the mean of each parameter as the baseline for each coefficient estimate (Table 4.4 and 4.5), which can be adjusted by adding the mean by all corresponding coefficients. In general, when the number of lactation number increases, an increasing tendency could be observed (exception: 1^{st} lactation has a greater *b* value than 2^{nd} lactation). Later years had higher values of Wood's parameter *a* and

TABLE 4.3: The residual sum of squares and proportion of variance of the fixed effects in mixed models on the lactation curve parameters and 305-d milk yield for 1^{st} lactation models.

Fixed effects[a]	DF	Parameter a		Parameter b		Parameter c		305-d milk yield	
		Sum sq	Proportion of variance[b] (%)	Sum sq	Proportion of variance[b] (%)	Sum sq	Proportion of variance[b] (%)	Sum sq	Proportion of variance[b] (%)
Year	10	$3.5x10^5$	9.05	138	11.85	0.015	10.89	$7.7x10^{10}$	29.80
Month	11	$2.6x10^6$	67.22	1002	86.01	0.12	87.10	$2.0x10^{10}$	7.74
Region	12	$4.8x10^4$	1.24	9.0	0.77	0.0007	0.51	$1.4x10^9$	0.54
Milking frequency	1	$2.4x10^5$	6.20	0.02	–[c]	0.00008	0.06[d]	$6.9x10^{10}$	26.70
Age at calving	1	$6.3x10^5$	16.29	16	1.37	0.002	1.45	$9.1x10^{10}$	35.22

[a] p < 0.0001 for all F tests if not specifically stated

[b] Proportion of total fixed effects variance

[c] p-value of this F test is 0.3

[d] p-value of this F test is 0.03

305-d milk yield and lower values for parameters b and c than earlier years. Month and region had different patterns among parameters and 305-d milk yield. In the 1^{st} lactation model, each day of increase in age at calving resulted in an increase of 0.0064 on the parameter a, 0.00000033 on parameter c, and 2.47 kg/lactation on 305-d milk yield, and a decrease of 0.000032 on parameter b. For example, to get the parameter a for the 2^{nd} lactation that began in August 2010 in a PA herd milking $2X/d$, one could extract the relevant coefficients from Table 4.4 and add them as following:

$$a = 19.9 + 2.16 + (-0.11) + (-0.96) + 1.15 + (-0.74) = 21.4 \tag{4.5}$$

TABLE 4.4: Estimate (standard error) of the coefficients of variables in the mixed models of lactation curve parameters and 305-d milk yield for all-lactations models.

Fixed effects variables	Value of category	Parameter *a*	Parameter *b* ($x10^{-2}$)	Parameter *c* ($x10^{-4}$)	305-d Milk yield (kg)
Mean		19.90(0.06)	24.70(0.10)	33.76(0.10)	10,210(31.60)
Lactation group	1	-4.18(0.01)	-0.37(0.01)	-9.31(0.01)	-1,276(0.68)
	2	2.16(0.01)	-1.20(0.01)	2.66(0.01)	355.7(0.68)
	3	2.02(0.01)	1.57(0.01)	6.65(0.01)	920.3(0.68)
Year	2006	-0.37(0.01)	0.72(0.02)	0.83(0.02)	-102.0(1.82)
	2007	-0.59(0.01)	1.00(0.02)	1.23(0.02)	-159.8(1.68)
	2008	-0.31(0.01)	0.47(0.02)	0.98(0.02)	-190.0(1.61)
	2009	-0.24(0.01)	0.24(0.02)	0.60(0.02)	-146.2(1.59)
	2010	-0.11(0.01)	-0.14(0.02)	0.31(0.02)	-214.9(1.56)
	2011	0.10(0.01)	-0.58(0.02)	-0.56(0.02)	-150.3(1.56)
	2012	0.33(0.01)	-0.71(0.02)	-0.83(0.02)	7.1(1.55)
	2013	0.27(0.01)	-0.51(0.02)	-0.73(0.02)	40.7(1.56)
	2014	0.12(0.01)	0.07(0.02)	-0.37(0.02)	225.3(1.57)
	2015	0.28(0.01)	-0.12(0.02)	-0.68(0.02)	322.9(1.63)
	2016	0.52(0.01)	-0.44(0.02)	-0.78(0.02)	367.2(1.66)
Month	January	-0.46(0.01)	1.81(0.02)	3.13(0.02)	126.3(1.69)
	February	0.18(0.01)	0.76(0.02)	2.43(0.02)	120.9(1.83)
	March	1.05(0.01)	-0.77(0.02)	1.04(0.02)	92.7(1.77)
	April	1.58(0.01)	-2.03(0.02)	-0.56(0.02)	46.6(1.90)
	May	1.49 (0.01)	-2.47(0.02)	-1.95(0.02)	-33.6(1.84)

	June	0.74(0.01)	-2.01(0.02)	-2.75(0.02)	-153.8(1.75)
	July	-0.41(0.01)	-0.81(0.02)	-2.68(0.02)	-282.2(1.68)
	August	-0.96(0.01)	0.11(0.02)	-2.06(0.02)	-261.4(1.66)
	September	-1.08(0.01)	0.78(0.02)	-1.08(0.02)	-97.1(1.61)
	October	-0.85(0.01)	1.20(0.02)	0.27(0.02)	83.5(1.60)
	November	-0.63(0.01)	1.45(0.02)	1.51(0.02)	177.2(1.64)
	December	-0.65(0.01)	1.98(0.02)	2.70(0.02)	180.9(1.68)
Region	Appalachian	-0.22(0.10)	-0.04(0.17)	-0.89(0.16)	-60.4(54.80)
	Corn belt	0.55(0.08)	-0.58(0.13)	-1.12(0.12)	192.9(40.90)
	Delta	-2.56(0.25)	0.59(0.41)	1.47(0.40)	-1,442.0(128.20)
	Lake	0.61(0.08)	-0.40(0.12)	-0.64(0.12)	288.5(39.80)
	Mountain	-0.96(0.36)	3.13(0.59)	1.50(0.57)	555.8(194.60)
	Northeast	1.04(0.10)	-1.99(0.16)	-1.13(0.16)	-80.9(52.60)
	Northern plains	-0.26(0.13)	0.19(0.21)	-0.79(0.20)	50.3(68.30)
	NY	0.67(0.08)	-1.21(0.13)	-0.45(0.13)	-23.3(42.70)
	PA	1.15(0.07)	-0.96(0.12)	0.06(0.11)	292.5(37.60)
	Southeast	-2.00(0.17)	2.59(0.27)	2.60(0.26)	-452.8(88.30)
	Southern plains	-0.51(0.19)	-1.02(0.30)	-0.93(0.29)	-568.2(96.30)
	West coast	1.09(0.47)	0.53(0.76)	0.52(0.73)	759.3(257.30)
	WI	1.4(0.07)	-0.83(0.12)	-0.20(0.10)	488.3(40.50)
Milking frequency	$2X/d$	-0.74(0.01)	0.09(0.01)	0.15(0.01)	-378.8(1.44)
	$3X/d$	0.74(0.01)	-0.09(0.01)	-0.15(0.01)	378.8(1.40)

TABLE 4.5: Estimate (standard error) of the coefficients of variables in the mixed models of lactation curve parameters and 305-d milk yield for 1^{st} lactation models.

Fixed effects variables	Value of category	Parameter a	Parameter b ($x10^{-2}$)	Parameter c ($x10^{-4}$)	305-d Milk yield(kg)
Mean		10.99(0.07)	26.89(0.15)	23.00(0.12)	7,170(30.4)
Year	2006	-0.48(0.02)	1.16(0.03)	0.99(0.03)	-64.5(2.7)
	2007	-0.66(0.01)	1.37(0.03)	1.29(0.02)	-134.3(2.6)
	2008	-0.33(0.01)	0.57(0.03)	0.82(0.02)	-142.9(2.5)
	2009	-0.13(0.01)	0.08(0.03)	0.45(0.02)	-126.7(2.5)
	2010	-0.03(0.01)	-0.21(0.03)	0.34(0.02)	-179.5(2.5)
	2011	0.13(0.01)	-0.63(0.03)	-0.36(0.02)	-139.1(2.4)
	2012	0.30(0.01)	-0.78(0.03)	-0.71(0.02)	6.2(2.4)
	2013	0.16(0.01)	-0.42(0.03)	-0.57(0.02)	39.6(2.4)
	2014	0.05(0.01)	0.06(0.03)	-0.44(0.02)	176.9(2.5)
	2015	0.34(0.01)	-0.33(0.03)	-0.78(0.02)	272.0(2.5)
	2016	0.64(0.01)	-0.94(0.03)	-1.03(0.02)	292.3(2.7)
Month	January	-0.42(0.10)	1.70(0.03)	2.77(0.02)	0.6(2.5)
	February	0.11(0.07)	0.50(0.03)	1.78(0.03)	-10.5(2.7)
	March	0.90(0.25)	-1.10(0.03)	0.45(0.02)	-10.2(2.7)
	April	1.41(0.07)	-2.40(0.03)	-1.04(0.03)	1.1(2.7)
	May	1.46(0.34)	-2.90(0.03)	-2.23(0.03)	-11.1(2.6)
	June	0.92(0.10)	-2.40(0.03)	-2.80(0.03)	-56.0(2.6)
	July	-0.15(0.12)	-0.90(0.03)	-2.47(0.03)	-125.9(2.5)
	August	-0.87(0.08)	0.52(0.03)	-1.60(0.03)	-102.8(2.4)

	September	-1.15(0.07)	1.50(0.03)	-0.32(0.02)	-4.2(2.5)
	October	-0.98(0.16)	1.90(0.03)	0.99(0.02)	109.2(2.4)
	November	-0.69(0.18)	1.80(0.03)	1.94(0.02)	131.3(2.5)
	December	-0.54(0.10)	1.80(0.03)	2.53(0.02)	78.5(2.5)
Region	Appalachian	-0.13(0.10)	-0.01(0.20)	-0.68(0.15)	-42.4(50.3)
	Corn belt	0.63(0.07)	-0.59(0.15)	-1.19(0.11)	263.0(37.5)
	Delta	-2.03(0.25)	0.73(0.50)	2.63(0.40)	-1,391(119.4)
	Lake	0.63(0.07)	-0.44(0.14)	-0.86(0.11)	321.9(36.3)
	Mountain	-1.32(0.34)	3.60(0.67)	1.12(0.52)	589.5(180.0)
	Northeast	1.03(0.10)	-2.40(0.19)	-1.18(0.15)	-135.8(48.1)
	Northern plains	-0.15(0.12)	0.15(0.24)	-1.17(0.19)	73.9(62.1)
	NY	0.74(0.08)	-1.38(0.15)	-0.35(0.12)	-18.0(39.1)
	PA	1.22(0.07)	-1.38(0.13)	-0.12(0.10)	280.6(34.4)
	Southeast	-1.58(0.16)	2.67(0.31)	1.69(0.25)	-382.7(80.9)
	Southern plains	-0.53(0.18)	-1.00(0.35)	-0.32(0.28)	-626.8(88.3)
	West coast	0.29(0.43)	0.88(0.85)	-0.21(0.66)	574.8(233.3)
	WI	1.20(0.07)	-0.84(0.14)	0.64(0.11)	493.0(35.5)
Milking frequency	$2X/d$	-0.69(0.01)	-	0.039(0.02)	-399.5(2.1)
	$3X/d$	0.69(0.01)	-	-0.039(0.02)	399.5(2.1)
Age at calving	Continuous	0.0064 (0.000061)	-0.0032 (0.00012)	0.0033 (0.00011)	2.47(0.01)

Figure 4.4 shows the lactation curves of mean parameter values from each parity and milking

frequency. Since $2X/d$ milking had a lower level, a faster rise and decline, the peak time between different milking frequencies did not differ much compared with $3X/d$ milking. Moreover, the lactation shapes of 1^{st} lactation were flatter than the others and intersected with the 2^{nd} lactation curves. Figure 4.5 included the error bar for parameters a, b and c within 3 standard deviations across year, month, and region. The year and month ranges of parameters were modest; however the regional ranges were substantial, particularly in regions with fewer data.

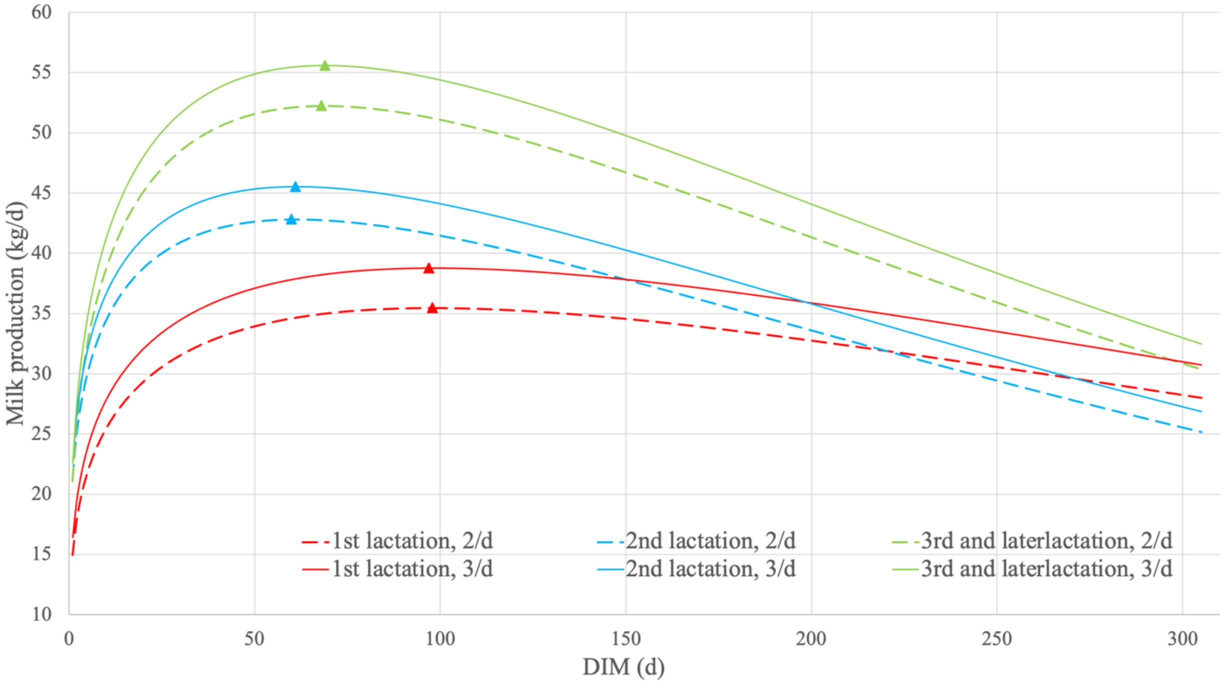


FIGURE 4.4: Lactation curves are plotted according to the estimated mean of the lactation curve parameters for each lactation group and milking frequency. Triangles indicate the peak.

Tables 4.6 and 4.7 contains estimates and standard errors for the variances of the random effects. The proportion of total variance explained by the herd was similar to the proportion explained by the animal for the parameters a (8.0% and 7.1%) and b (9.2% and 6.9%), however, the proportion explained by the animal for parameter c was significantly greater (18.0%) than the proportion

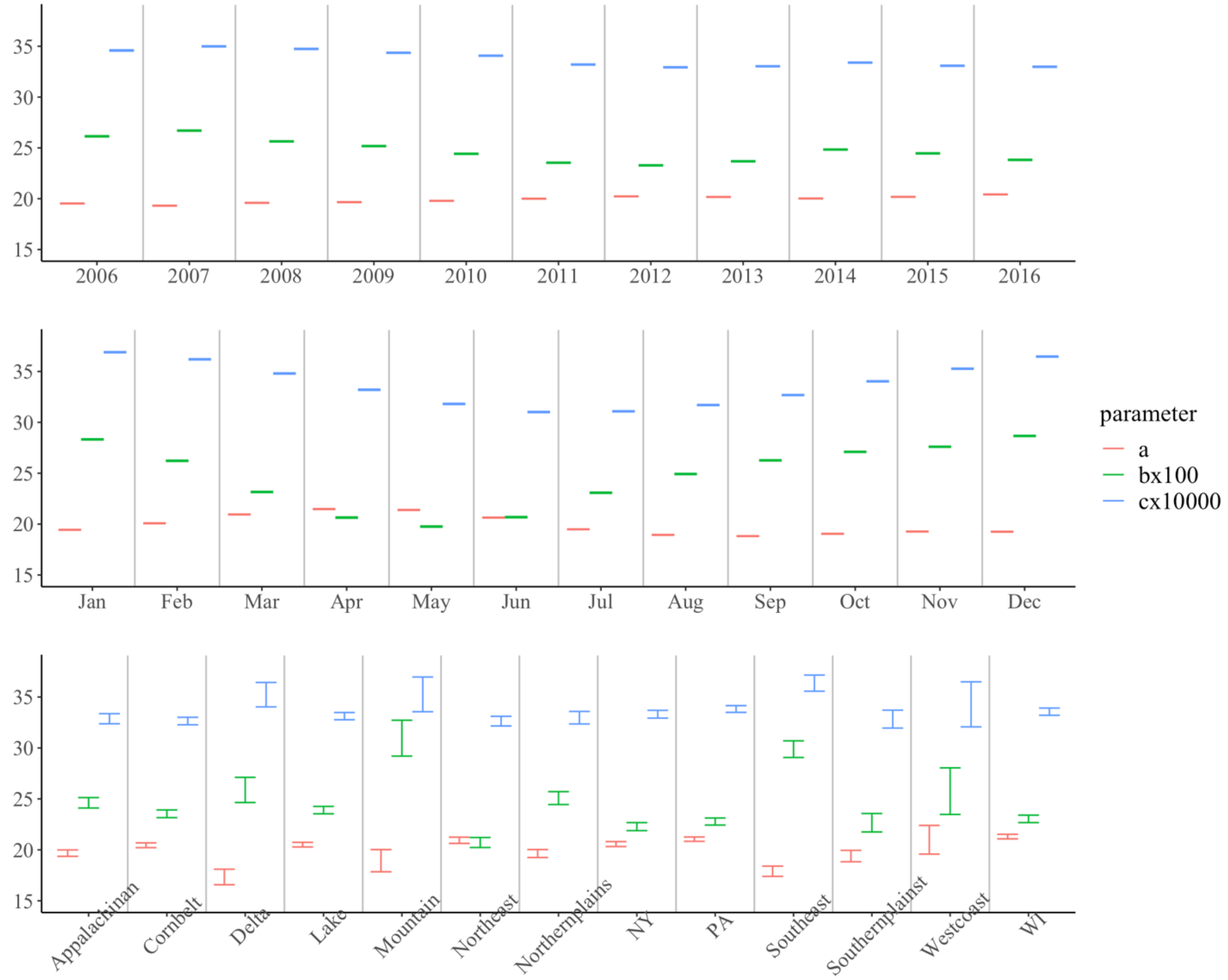


FIGURE 4.5: Plots of error bar within 3 standard deviations of lactation curves parameters estimated from the mixed model for each year (top), month (middle), and region (bottom)

explained by herd (6.7%). For 305-d milk yield, herd effect (44.8%) explained a greater proportion of the variance than animal (25.8%) in all-lactations models (Table 4.6). Herd explained 8.5%, 8.0%, 7.5%, and 46.5% of total variance in the 1^{st} lactation models for parameter a, b, c, and 305-d milk yield, respectively (Table 4.7).

TABLE 4.6: Variance and proportion of the total variance of random effects in mixed models on lactation curve parameters and 305-d milk yield for all-lactations models

Random effects	Animal[a]		Herd[a]		Residual	
	Variance	Proportion of the total variance (%)	Variance	Proportion of the total variance (%)	Variance	Proportion of the total variance (%)
Parameter a	7.75	8.0	6.83	7.1	81.8	84.9
Parameter b	0.0024	9.2	0.0018	6.9	0.022	84.0
Parameter c	$4.47x10^{-7}$	18.0	$1.67x10^{-7}$	6.7	$1.87x10^{-6}$	75.3
305-d milk yield	1,236,877	25.8	2,144,236	44.8	1,404,257	29.3

[a] $p < 0.0001$ for all likelihood ratio test of random effects

TABLE 4.7: Variance and proportion of the total variance of random effects in mixed models on lactation curve parameters and 305-d milk yield for 1st lactations models

Random effects	Herd[a]		Residual	
	Variance	Proportion of the total variance (%)	Variance	Proportion of the total variance (%)
Parameter a	5.39	8.5	58.18	91.5
Parameter b	0.0020	8.0	0.023	92.0
Parameter c	$1.23x10^{-7}$	6.5	$1.76x10^{-6}$	93.5
305-d milk yield	1,684,589	46.5	1,941,789	53.5

[a] $p < 0.0001$ for all likelihood ratio test of random effects

4.5 Discussion

Parameter *a* is the scale factor representing initial milk yield, parameter *b* is the rate factor representing the rate of increase in milk yield to peak, and parameter *c* is the rate factor representing the rate of decline in milk yield after the peak (Wood, 1967). Understanding the behavior of factors affecting the lactation curve shape can help making better informed farm management decisions.

Fixed effects factors. Lactation groups with different parity accounted for 90%, 34%, 92%, and 93% of the total variance of fixed effects on parameters a, b, c, and the 305-d milk yield, respectively. As a result of the considerable lactation group effects on parameter a and c, even lactations with comparable production levels in different lactation groups exhibit substantial differences in the shape of the estimated lactation curves, particularly during the decay period of lactation. According to our findings and those of Ehrlich (2011), it is critical to distinguish the lactation curve parameters for the 2^{nd} lactation to utilize more energy to growth-related metabolization. According to Figure 4.4, after 200 DIM for $3X/d$ (and 222 DIM for $2X/d$), the milk yield of the 1^{st} lactation would be greater than that of 2^{nd} lactation. This information could be used for reproductive management purposes, such as shortening the open period for 2^{nd} lactation cows or extending the length of the 1^{st} lactation, which could be more beneficial to total farm productivity. When estimating lactation curve parameters, the distinction among lactation groups of the lactation curve parameters should be emphasized.

Other than lactation groups, the fixed effect of calving month accounted for 8%, 59%, 7%, and 2% of the total variance of fixed effects on parameters a, b, c, and the 305-d milk yield, respectively, in all-lactations models, and 67%, 86%, 87%, 7.4% respectively, in 1^{st} lactation models. The month's considerable impact on parameter b that the calving season had an effect on the rate of milk yield progression in the early days across all lactation groups. Also, the effects of the month on parameters were more remarkable than it on the 305-d milk yield, suggesting that even when the total milk yields were on the same level across months, the shape of the lactations, namely raising and decaying rates, the initial and peak yield, or the peak time, could be different. Thus, knowing the lactation curve parameters for a particular calving month would help in estimating milk yield and provide more precise information for nutrition, reproduction, and facility management decisions. Specifically, parameter a was greater than the mean of all months from February to June, parameter b from August to February, and parameter c from October to March. Due to the different rates of the parameters b and c, which determine the peak time (b/c) of lactation curves, the average peak time in August (78 DIM) was the latest and the earliest was in April (68 DIM). Apart from the parameters, as Torshizi (2016) and Dedkova and Nemcova (2003) stated, winter calving led to a higher total milk production, which corroborated our findings of the highest 305-day milk production and rate to peak (parameter b). Lactations started during hot seasons, namely May to September, have lower-than-average estimates for the 305-d milk yield (July decreases 3% from the average), which might indicate the influence of heat stress.

Between 2006 and 2016, the estimated 305-day milk yield in the mixed model outcomes increased from 10,108 kg/yr to 10,577 kg/yr (+4.6%), which reflected the change in milk production due to the change in calving year. Capper and Cady (2019) reported that the national mean annual milk yield per cow increased from 9,164 kg/yr to 10,406 kg/yr (+13.6%) from 2007 to 2017, whereas our data prior to the mixed models (individual non-linear fitting results) showed an increase from 10,747 kg/yr to 11,812 kg/yr (+9.9%) over the same 11 years period. Over the years, the increasing milk yield trend (increasing parameter a and 305-d milk yield, decreasing parameter b and c, and later peak time) has been interpreted as a response to genetic selection and management practice advancements.

Cows milked $3X/d$ had a higher milk production than $2X/d$, an increase of on average of 2.2 kg/d and 2.5 kg/d for the 1^{st} lactation model and all-lactations model, respectively, which was consistent with previous studies (Hart et al., 2013 reported 2.5 kg/d and 3.2 kg/d for 1^{st} and 2^{nd} and later lactations, respectively; Erdman and Varner, 1995 summarized it from the literature as 3.5kg/d for all lactations). The magnitude of the changes from our model (7.7% for all-lactations model, 11.8% for 1^{st} lactation models) was less than 15% reported by Smith et al. (2002) or 12% for total and 14% for 1^{st} lactation reported by Gisi et al. (1986). It is possible that the increased production over the last decade had an impact on dominator's base production level; in other words, the effect of increasing milking frequency may not be proportionate to the production as suggested by Erdman and Varner (1995). Also, Smith et al. (2002) reported results from an analysis of a DHI dataset in which 7.1% $3X/d$ of the herd in the year 2000, compared to our dataset in which 44.5% of the herd milked $3X/d$ day during 11-year period from 2006 to 2016. A more balanced number of records controlled other confounding factors associated with production levels, such as management levels. The peak time (b/c) and the overall shape between milking $3X/d$ and $2X/d$ was not much different (Figure 4.4), while peak and 305-d production improved constantly across lactation groups.

As Cole et al. (2009, 2011) suggested, the geographical influence had an effect on milk production and lactation curve parameters. According to our findings (Figure 4.5), New York and Pennsylvania exhibited similar parameter ranges. They were closer to Wisconsin than the rest of the Northeast; the Northeast in particular had a smaller parameter b. Wisconsin, on the other hand, had a higher parameter a and a lower parameter b than the other Lake states. Delta states and the Southeast, Mountain, and West Coast had higher parameters b and c, with the longest error bars indicating large regional variance. Generally, a greater rate of increase and decay than in other regions. The West Coast had the highest parameter a and a large b, which resulted in a greater lactation peak production among those 4 regions. Parameters from such regions with a low standard deviation would be more dependable in characterizing the region's general lactation curve. Furthermore, while all variables included in the mixed model were significantly affected, the proportion of variance controlled by region was the lowest for parameter a and 305-d milk yield, followed by parameter b and then parameter c. This could be partly related to the substantial variation within each region, which limited the region factor's exploratory power.

In the 1^{st} lactation mixed model, age at calving explained for 16.3%, 1.4%, 1.4%, and 22.9% of total variance for parameter *a*, *b*, *c*, and 305-d milk yield, respectively. The associations were statistically significant ($p < 0.005$), which agreed with Torshizi (2016). Increased age at calving led to increased 305-d milk yield, the value of parameters *a* and *c*, and decreased the value of parameter *b*, similar to what Atashi et al. (2019) reported with a few exceptions on parameters *a* and *b*. A positive relationship between 305-d milk yield and age at first calving was consistent with a previous study indicating a positive relationship between age at calving and lactation milk yield (Ettema and Santos, 2004, Albarran-Portillo and Pollott, 2011). One more day of age at first calving resulted in a 2.47 kg/lactation increase, similar to what was reported by Berry and Cromie (2009) reported, indicating that 1-month increase in age at first calving resulted in a 55.5 kg/lactation increase. Our findings of an increase in parameter *c* with increasing age at calving supported the argument stated by Muir et al. (2004) that the lactation curve persistency for calving at a younger age was higher than calving at a later age. According to Teke and Murat (2013), the increased 305-d milk yield related to the increased age at first calving which allows for additional time for mammary gland development. The relationships between age at first calving and 305-d milk yield was positive for 18-26 months and negative thereafter (Niloforooshan and Edris, 2004; Torshizi, 2016). In other cases, only negative effects have been reported (Bewley et al., 2001).

Random effects factors. In our study, including animals as a random variable resulted in an improved fit, as measured by the AIC and P-value of likelihood ratio tests. It was consistent with research on non-linear mixed models of lactation curve fitting when fitting lactation curves for the purpose of making predictions (Piccardi et al., 2017), which concluded that incorporating random deviations enables the model to accurately represent typical cows by averaging the random effects. Ehrlich (2013) stated that a considerable difference in parameters may be found between herds, indicating that uncontrolled genetic or environmental factors, including random effect herds, worked to reduce the herd's variance. As Macciotta et al. (2005) noted, individual lactation curves exhibit a high degree of diversity in parameter estimations, including random effects. Animal influence explained 8.0%-18.0% of the variance in parameters and 25.8% of the variance in 305-d milk yield in our all-lactations models. On all-lactations and 1^{st} lactation models, the herd effect controlled 6.7%-7.1% and 6.5%-8.5% variance in parameters and 44.8% and 46.5% variance in 305-d milk yield, respectively (Table 4.6 and 4.7). In all-lactation models, the animal had a greater influence on parameter values, whereas the herd had a greater influence on the 305-day milk yield. The range of herd effects as a proportion of total variance was similar in both models, indicating that there was considerable variation between herds in our dataset.

Applications. Our mixed model's estimated lactation curve parameters could be applied in a variety of ways. For instance, dairy farm simulation models in which daily milk production should be estimated in order to model body weight change, calculate nutrient requirements, formulate diets, and account for income. While milk production from dairy cattle increased by 11% over the 10-year period from 2011 to 2020 (USDA-NASS, 2020), the majority of dairy farm simulation models available today employ lactation curve characteristics from the late 1990s and early 2000s. For example, the Integrated Farm System Model (IFSM; Rotz et al., 2013) utilized lactation curve parameters from 1999 (Rotz et al., 1999), Calsamiglia et al. (2018) implemented lactation curve parameters from 1967 (Wood, 1967), and Silva-Villacorta et al. (2016) incorporated lactation curve parameters from 2006. Other lactation curve fitting analyses have been developed for specific farms or herds using more recent data (Murphy et al., 2014; Masia et al., 2020); however, due to the small number of animals included, inferences from those works were limited to the conditions of the data used for model fitting. Furthermore, when characterizing lactation curves, including additional factors that can modify the lactation curve parameters enables model users to tailor the application to accurately represent their interested conditions.

The findings of this study may contribute to a more accurate representation of current animal performance in a dairy farm system model, such as the Ruminant Farm System Model (RuFaS, Li et al., 2018; Kebreab et al., 2019; Hansen et al., 2021). The RuFaS model is a daily time-step, process-based, whole-farm simulation model. Accurate modeling of each animal's targeted daily milk yield is crucial for the model to reflect nutrient flows and overall farm performance. The results of the mixed models will be integrated into the RuFaS model to more precisely model lactation performance and to more properly represent the simulated farms.

4.6 Conclusions

We found that all temporal (calving year and month), spatial (region of the farm), and management factors (milking frequency, and age at calving for 1^{st} lactations, parity) had statistically significant effect on the lactation shape and 305-d milk yield (milking frequency on parameter b was the only exception). By include herd and animal as random effects in the mixed effects models, we were able to control within-herd and animal variability and provided better fits than reduced models. Additionally, the fixed effects estimates provided parity-, year-, month-, region-, and milking frequency- specific lactation curve parameters for future applications. The mixed model's estimated lactation curve parameters well characterized lactation curves in terms of temporal, spatial, and management factors.

References

Abdelkrim, Ben, L. Puillet A., P. Gomes, and O. Martin (2019). *Lactation curve model with explicit representation of perturbations as a phenotyping tool for dairy livestock precision farming*. PCI Animal Science.

Albarran-Portilloa, B. and G. E. Pollottb (2011). Environmental factors affecting lactation curve parameters in the United Kingdom's commercial dairy herds. In: *Arch Med Vet* 43, pp. 145–153.

Atashi, H., M. Salavati, J. De Koster, J. Ehrlich, M. Crowe, G. Opsomer, and M. Hostens (2019). Genome-wide association for milk production and lactation curve parameters in Holstein dairy cows. In: *Journal of Animal Breeding and Genetics* 137, pp. 292–304. DOI: `10.1111/jbg.12442`.

Bates, D., M. Machler, B. Bolker, and S. Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. In: *Journal of Statistical Software* 67.1, pp. 1–48.

Berry, D. and A. Cromie (2009). Associations between age at first calving and subsequent performance in Irish spring calving Holstein–Friesian dairy cows. In: *Livest. Sci* 123, pp. 44–54.

Bewley, J., R. W. Palmer., and D. B. Jackson-Smith (2001). Modeling milk production and labor efficiency in modernized Wisconsin dairy herds. In: *J. Dairy Sci* 84, pp. 705–716.

Bouallegue, M. and N. M'Hamdi (2020). Mathematical modeling of lactation curves: A review of parametric model in Lactation in Farm Animals: Biology, Physiological Basis, Nutritional Requirements, and Modelization. In: ed. by N. M'Hamdi. IntechOpen Limited.

Calsamiglia S., and Astiz S., J. Baucells, and L. Castillejos (2018). A stochastic dynamic model of a dairy farm to evaluate the technical and economic performance under different scenarios. In: *J. Dairy Sci* 101.29803414, pp. 7517–7530.

Capper, J. L. and R. A. Cady (2019a). The effects of improved performance in the US dairy cattle industry on environmental impacts between 2007 and 2017. In: *J Anim. Sci* 98. DOI: `10.1093/jas/skz291`.

Capuco, A. V., D. L. Wood, R. Baldwin, K. McLeod, and M. J. Paape (2001). Mammary cell number, proliferation, and apoptosis during abovine lactation: Relation to milk production and effects of bST. In: *J. Dairy Sci* 84, pp. 2177–2187.

Cole, J. B., D. J. Null, and P. M. VanRaden (2009). Best prediction of yields for long lactations. In: *J. Dairy Sci* 92, pp. 1796–1810.

Cole, J. B., D. J. Null, and A. De Vries (2011a). Short communication: Best prediction of 305-day lactation yields with regional and seasonal effects. In: *J. Dairy Sci* 94, pp. 1601–1604.

Dedkova, L. and E. Nemcova (2003). Factors affecting the shape of lactation curves of Holstein cow cows in Czech Republic. In: *Czech J Anim Sci* 48.10, 395–402.

Dijkstra, J., S. López, A. Bannink, M. S. Dhanoa, E. Kebreab, N. E. Odongo, and J. France (2010a). Evaluation of a mechanistic lactation model using cow, goat and sheep data. In: *Journal of Agricultural Science* 148.3, pp. 249–262.

Ehrlich, J. L. (2011b). Quantifying shape of lactation curves, and benchmark curves for common dairy breeds and parities. In: *Bovine Practitioner* 45.1, p. 88.

Ehrlich, J. L. (2013). Quantifying inter-group variability in lactation curve shape and magnitude with the MilkBot lactation model. In: *PeerJ* 1.e54.

Erdman, R. A. and M. Varner (1995). Fixed yield responses to increased milking frequency. In: *J. Dairy Sci* 78, pp. 1199–1203.

Ettema, J. F. and J. E. Santos (2004). Impact of age at calving on lactation, reproduction, health, and income in first-parity Holsteins on commercial farms. In: *J. Dairy Sci* 87, pp. 2730–2742.

Gisi, D. D., E. J. DePeters, and C. L. Pelsissier (1986). Three times daily milking of cows in California dairy herds. In: *J. Dairy Sci* 69, pp. 863–868.

Hansen, T., L. M. Li, J. Li, C. J. Vankerhove, M. A. Sotirova, J. M. Tricarico, V. E. Cabrera, E. Kebreab, and K. F. Reed (2021b). The Ruminant Farm Systems Animal Module: A Biophysical Description of Animal Management. In: *Animals* 11.5, p. 1373.

Hart, K. D., B. W. McBride, T. F. Duffield, and T. J. DeVries (2013). Effect of milking frequency on the behavior and productivity of lactating dairy cows. In: *J. Dairy Sci* 96, pp. 6973–6985.

Kebreab, E., K. F. Reed, V. E. Cabrera, P. A. Vadas, G. Thoma, and J. M. Tricarico (2019b). A new modeling environment for integrated dairy system management. In: *Animal Frontiers* 9.2, pp. 25–32. DOI: `10.1093/af/vfz004`.

Knight, C. H. (2000). The importance of cell division in udder development and lactation. In: *Livest. Prod. Sci* 66, pp. 169–176.

Li, M., V. E. Cabrera, and K. Reed (2018a). Comparison of Holstein and Jersey milk production with a new stochastic model. In: *J. Dairy Sci* 100(Suppl. 2): M-207.

Macciotta, N. P. P., Dimauro C., S. P. G. Rassu, R. Steri, and G. Pulina (2011b). The mathematical description of lactation curves in dairy cattle. In: *Ital. J. Anim. Sci* 10.e51.

Macciotta, N. P. P., D. Vicario, and A. Cappio-Borlino (2005). Detection of different shapes of lactation curve for milk yield in dairy cattle by empirical mathematical models. In: *J. Dairy Sci* 88, pp. 1178–1191.

Masía, F. M., NA. Lyons, M. Piccardi, M. Balzarini, R. C. Hovey, and S. C. Garcia (2020). Modeling variability of the lactation curves of cows in automated milking systems. In: *J. Dairy Sci* 103, pp. 8189–8196.

Muir, B. L., J. Fatehi, and L. R. Schaeffer (2004). Genetic relationships between persistency and reproductive performance in first-lactation Canadian Holsteins. In: *J Dairy Sci* 87, pp. 3029–3037.

Murphy, M. D., M. J. O'Mahony, L. Shalloo, P. French, and J. Upton (2014a). Comparison of modeling techniques for milk-production forecasting. In: *J. Dairy Sci* 97, pp. 3352–3363.

Niloforooshan, M. A. and Edriss MA. (2004). Effect of age at first calving on some productive and longevity traits in Iranian Holsteins of the Isfahan province. In: *J Dairy Sci* 87, pp. 2130–5.

Piccardi, M., R. Macchiavelli, A. C. Funes, G. A. Bo, and M. Bal-zarini (2017). Fitting milk production curves through non-linear mixed models. In: *J. Dairy Res* 84, pp. 146–153.

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, and R-Core-Team (2021). *nlme: Linear and Nonlinear Mixed Effects Models*. Tech. rep. URL: `https://CRAN.R-project.org/package=nlme`.

R-Core-Team (2020a). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: `https://www.R-project.org/`.

Rotz, C. A., M. S. Corson, D. S. Chianese, F. Montes, S. D. Hafner, and C. U. Coiner (2013a). *The Integrated Farm System Model*. Washington, DC: USDA-ARS.

Rotz, c. a., D. R. Mertens, D. R. Buckmaster, M. S. Allen, and J. H. Harrison (1999b). A Dairy Herd Model for Use in Whole Farm Simulations. In: *J. Dairy Sci* 82, 2826–2840.

RStudio-Team (2020). *RStudio: Integrated Development for R*. PBC, Boston, MA: RStudio. URL: `http://www.rstudio.com/`.

Silva-Villacorta, D., N. Lopez-Villalobos, H. T. Blair, R. E. Hickson, and MacGibbon A. K. (2016). A stochastic farm model to simulate dairy farms and the segregation of cows to produce milk with different concentrations of unsaturated fatty acids. In: *Comput Electron Agr.* 125, pp. 29–39.

Silvestre, A. M., F. Petim-Batista, and J. Colaço (2006). The Accuracy of Seven Mathematical Functions in Modeling Dairy Cattle Lactation Curves Based on Test-Day Records From Varying Sample Schemes. In: *J. Dairy Sci* 89, pp. 1813–1821.

Smith, J. W., L. O. Ely, W. M. Graves, and W. D. Gilson (2002). Effect of milking frequency on DHI performance measures. In: *J. Dairy Sci* 85, pp. 3526–3533.

Stelwagen, K. (2001). Effect of milking frequency on mammary function and shape of the lactation curve. In: *J. Dairy Sci* 84(E Suppl.) E204–E211.

Teke, B. and H. Murat (2013). Effect of age at first calving on first lactation milk yield, lifetime milk yield and lifetime in Turkish Holsteins of the Mediterranean region in Turkey. In: *Bulgarian Journal of Agricultural Science* 19.5, pp. 1126–1129.

Tekerli, M., Z. Akinci, I. Dogan, and A. Akcan (2000). Factors Affecting the Shape of Lactation Curves of Holstein Cows from the Balikesir Province of Turkey. In: *J. Dairy Sci* 83, pp. 1381–1386.

Torshizi, M. E. (2016). Effects of season and age at first calving on genetic and phenotypic characteristics of lactation curve parameters in Holstein cows. In: *J. Anim. Sci. Technol* 58, p. 8.

USDA-ARS (2020). Milk cows and production by State and region (Annual). URL: `https://www.ers.usda.gov/data-products/dairy-data/`.

USDA-NASS (May 2020a). Agricultural Marketing Service, and Economic Research Service calculations. URL: `https://downloads.usda.library.cornell.edu/usda-esmis/files/c821gj76b/jm2159057/qv33sh51q/agpr0520.pdf`.

Wood, P. D. P. (1967b). Algebraic model of the lactation curve in cattle. In: *Nature* 216, pp. 164–165.

Wood, P. D. P. (1972). A note on seasonal fluctuations in milk production. In: *Animal Production* 15, pp. 89–92.

Wood, P. D. P. (1976). Algebraic models of the lactation curves for milk, fat and protein production, with estimates of seasonal variation. In: *Animal Production* 22, pp. 35–40.

Chapter 5

A time-series analysis of milk productivity changes in US dairy states

M. Li, * K. F. Reed, † V. E. Cabrera, *

* Department of Animal and Dairy Sciences, University of Wisconsin-Madison, Madison WI 53705

† Department of Animal Science, Cornell University, 272 Morrison Hall, Ithaca, NY 14850

5.1 Abstract

Dairy cow productivity continues to improve in the United States; however production levels vary according to geographic location and seasonal patterns. It is necessary to characterize milk productivity trends and patterns in order to forecast milk production in certain US states or regions. The purpose of this study was to:

1. quantify historical trends and seasonal patterns of milk and milk components production by US states;
2. classify states with similar trends and seasonal patterns into clusters;
3. predict expected milk and milk components production according to calving date and location for use in simulation models.

Our dataset contained 9.18 million lactations (300 < DIM < 600) from 5.61 million Holstein cows from January 2006 to December 2016 distributed in 17 states. Each record included total milk, fat, and protein yield during the lactation. We used time series decomposition to obtain an annual trend and each state's seasonal pattern in milk productivity. Then, we classified states with agglomerative hierarchical clustering into smaller groups according to 2 methods:

1. dynamic time wrapping on raw time series

2. time-series clustering on extracted features of trend and seasonality from the decomposition

Results showed distinguishable trends and seasonality for all states and parity. The clusters and cluster centroid pattern showed a general upward trend for all yields (ECM, milk, fat, and protein) and a steady trend for fat and protein percent in all states except Texas. We also found a larger seasonality amplitude for all yields (ECM, milk, fat, and protein) from higher lactation numbers and a similar amplitude for fat and protein percent across lactation numbers. The results could be used for advising breeding programs like having more heifers calving during late June to early August according to the achieve better fat percent in their production. Furthermore, the trend and seasonality pattern can be used to adjust the production level of farms simulated in a specific state, year, and season.

Keywords: milk yield, calving season, time series, cluster

5.2 Introduction

Dairy cows in the US increased milk productivity by 14% over the 11 years from 2006 to 2016 (USDA-NASS, 2020). Capper and Cady (2019) described the connection between improvement in US dairy cattle's performance and reduction in the number of dairy cows needed to produce the same amount of saleable energy-corrected milk, which reduced resource use, nutrient extraction, and greenhouse gas emissions per unit of milk. Also, the average milk production per cow varied greatly across different US states (USDA-ERS, 2021). Moreover, the seasonal pattern of milk yield and milk component percent had been well recognized (Salfer et al., 2019, Ferreira et al., 2020). Accurate projections of milk production by location, year, and season would benefit farm and market level simulation models that inform management and policy decisions by providing an accurate representation of lactation performance. One example is the Ruminant Farm

Systems (RuFaS) model (Kebreab et al., 2019; Hansen et al., 2021), which is a stochastic and modular dairy farm simulation model under development by our group that simulates the dairy farm nutrient cycle from the animal to the manure, soil, and crops. In RuFaS, stochastic lactation curve parameters inform predictions of milk production for each animal on a daily timestep. We previously characterized lactation curve parameters fitted to lactation records from 2016 as a set of baseline production curves for use in RuFaS (Li et al., 2019). However, we plan to incorporate yearly trends and seasonal effects on milk and components yields to reflect temporal changes more accurately in milk production (Cole et al., 2011; Ferreira et al., 2020). To our knowledge, calving season and year have not been included in methods for predicting milk production in the most herd or whole-farm simulation models (Schils et al., 2007; Rotz et al., 2015). Therefore, the objectives of this study are:

1. accurately quantify trends and seasonal patterns of milk and milk components production by state;
2. classify states with similar trends and seasonal patterns into clusters;
3. summarize the general pattern for each cluster for further applications in simulation models.

The milk and milk components change along time of calving date are formed as time series data format. Time series consisted of a list of numbers (variables) and their dependent time-sequential information when those numbers were recorded (Hyndman and Athanasopoulos, 2018). Time-series studies analyzed and modeled the dependence between time and the value of the variable of interest (Box and Jenkins, 1976). Time-series effects include the level, yearly trend, seasonal, and random variations (Cleveland et al., 1990). The trend pattern exists when there is a long-term change in the mean level. Similarly, a seasonal pattern exists when the time series has a cyclic feature that repeatedly occurs with a specific time length: week, month, or year. Time series decomposition was widely used to identify trends and seasonal patterns (Hyndman and Athanasopoulos, 2018). The original time series and decomposed components of time series could be clustered into homogeneous groups where the with-in-group similarity is maximized, and the between-group-object dissimilarity is maximized (Aghabozorgi et al., 2015).

5.3 Materials and Methods

5.3.1 Dataset

The Council on Dairy Cattle Breeding (https://www.uscdcb.com/) provided a large dataset containing lactation records with 7 key variables: milk yield, fat yield, protein yield, lactation starting date, parity, US state, and lactation length. The lactation records were from 17 states (each with at least 100,000 lactation records, Table 5.1) over 11 years (January 2006 to December 2016). We removed records from the original dataset with extreme (outside the 99th percentile) milk, fat, and protein yields and those with lactation lengths less than 300 or greater than 600 days. The resulting dataset used in this study consisted of 9,184,086 million lactation records from 5,606,351 million Holsteins. Although geographically diverse, the data did not represent the geographic distribution of milk production in the US. Most records were from Wisconsin (28.4%), Pennsylvania (15.2%), and New York (14.4%), and the dataset did not include records from Washington and Idaho. The parity distribution was 38.6% from 1st lactation, 28.1% from 2nd lactation, and 33.3% from 3rd and later (3rd +) lactations cows (Table 5.1). For each week, the mean number of records was 176,617, with a standard deviation of 12,286. We calculated fat percent, protein percent, and

ECM (IFCN, 2010):

$$ECM = milk\_yield * \frac{0.383 * fat\_percent + 0.242 * protein\_percent + 0.7832}{3.1138} \quad (5.1)$$

from the milk, fat, and protein yield. We then averaged the lactation-length yield for milk, fat, protein, fat percent, protein percent, and ECM by the week of the year when that lactation started. Boxplots of lactation length show similar distributions across years, weeks, states, and lactation numbers (Figure 5.1), indicating effects from lactation length would not interfere with our analysis result.

TABLE 5.1: Number of records by lactation group, year, and state

Parity		Year		State	
1	3,546,421	2006	766,576	Wisconsin	2,608,304
2	2,583,315	2007	840,300	Pennsylvania	1,392,502
>=3	3,054,350	2008	827,239	New York	1,323,249
		2009	817,003	Minnesota	969,995
		2010	850,101	Michigan	689,397
		2011	850,025	Iowa	375,355
		2012	839,301	Ohio	372,010
		2013	859,855	Virginia	244,003
		2014	856,587	Vermont	181,537
		2015	836,653	Texas	176,241
		2016	840,446	Illinois	160,519
				Indiana	131,192
				Florida	123,980
				South Dakoda	118,473
				Massachusetts	109,463
				California	107,676
				Georgia	100,190

5.3.2 Time Series Decomposition

To better assess and describe the features contained in ECM, milk, fat, and protein yield,

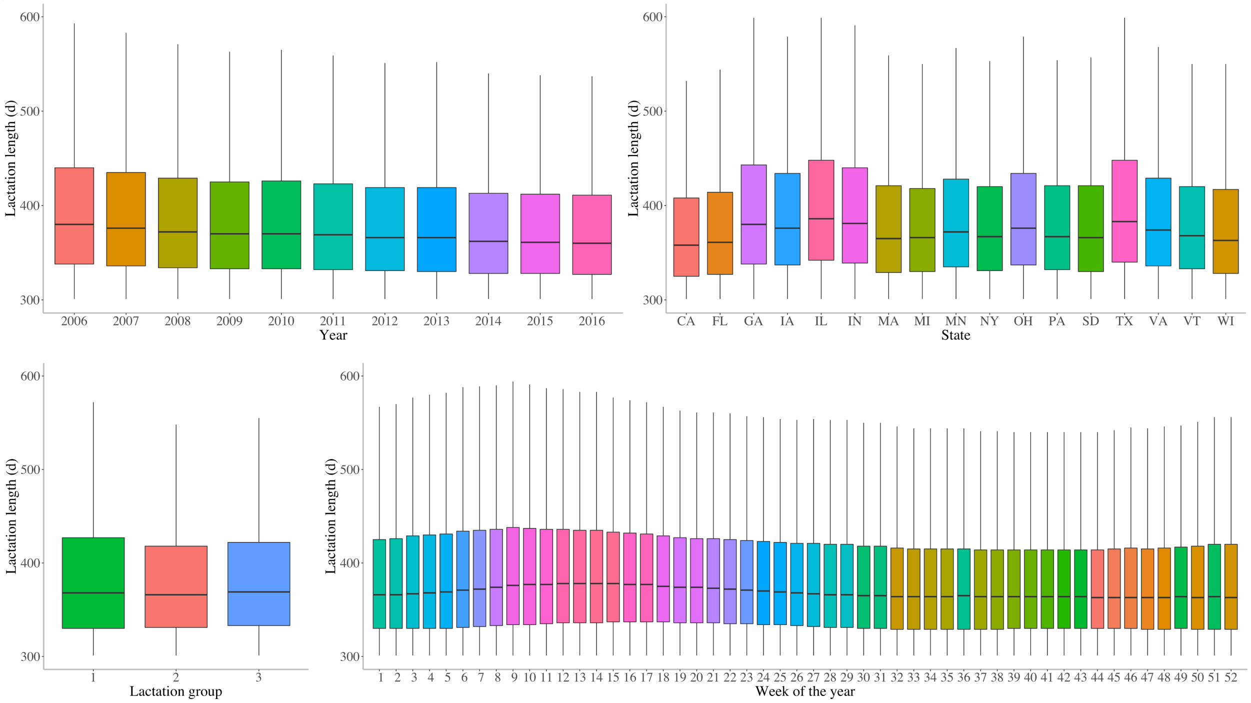


FIGURE 5.1: Boxplots of lactation length across years, states, lactation groups (1st, 2nd, 3rd +), and week of the year when lactation started

and fat and protein percent, we decomposed the time series of each response with an additive model. The trend component was often identified using a moving average, then removing the trend component from the time series removed most of the autocorrelation from the long-term feature. Next, the average seasonal component was described by the centered average curve across each cyclic period. Then removing the seasonality component further removed the autocorrelations due to seasonality. Last, the remainder component was left after removing both trend and seasonality. The decomposing method was applied to each time series, i.e.

$$Y_t = S_t + T_t + R_t \tag{5.2}$$

where Y_t, S_t, T_t, and R_t were the response, the seasonal, annual trend, and residual terms, respectively, at time t. We used R (R Core Team, 2020) to decompose each time series. We conducted the analysis for all 17 states, for each of the 6 response variables (milk yield, protein yield, fat yield, ECM yield, protein percent, and fat percent), and for the 3 parity groups (1st, 2nd, and 3rd + lactations) which resulted in 306 time series decompositions. We measured the strength of trend and seasonality components following Wang, Smith, & Hyndman (2006):

$$F_t = max(0, 1 - \frac{Var(R_t)}{Var(T_t + R_t)}) \tag{5.3}$$

$$F_s = max(0, 1 - \frac{Var(R_t)}{Var(S_t + R_t)}) \tag{5.4}$$

Where F_t was the strength of the trend and F_s was the strength of the seasonality. When the strength of the trend or seasonality is strong, the seasonally adjusted or detrended data have more variation than the remainder component and, therefore, F_t or F_s would be closer to 1. The strength indicated how well the decomposition process identified trends and seasonal components. The closer to 1 the strength was, the more significant the trend or seasonal pattern was identified from the remainder. The hypotheses of having the same strength of trends and seasonal patterns were tested across lactation groups.

We calculated the variance of each decomposition component compared with the variance associated with the original time-series data to obtain the relative importance of that component. The ratios of variances were calculated following Lafare et al. (2016) as:

$$Ratio_{trend} = \frac{Var(T_t)}{Var(Original)} \tag{5.5}$$

$$Ratio_{seasonal} = \frac{Var(S_t)}{Var(Original)} \tag{5.6}$$

Where $Ratio_{trend}$, and $Ratio_{seasonal}$ are the ratio of trend and seasonal variance with the variance of the original time series.

5.3.3 Time Series Clustering

We applied agglomerative hierarchical clustering, which started the process with each cluster and sequentially grouped the algorithm based on the between-cluster dissimilarities in each step (Hastie et al., 2009). Pairwise distance matrices of time series pairs were used to conduct nested hierarchies of similar time series. We used the dtwclust library in R (Sarda-Espinosa A, 2019) to conduct the hierarchical clustering with Ward's minimal variance method (Murtagh and Legendre, 2014) to minimize the total variance to classify all models.

To achieve our goal of understanding the similarity of production features across states, we used cluster distance of both the raw time series data and the converted time series from the coefficients of trend and seasonal from the decompositions. We calculate the distance between the time series of the raw data using dynamic time warping (DTW) distance and the trend and seasonality components from decompositions using Euclidean distance.

We clustered 3 parity groups (1st, 2nd, and 3rd + lactations) in 17 states in each of 3 types of time series (raw, trend, and seasonal) for each of the 6 variables (milk yield, protein yield, fat yield, ECM yield, protein percent, and fat percent). That meant for each variable and time series type; we had 51 time series to be clustered. Then, time series of 17 states that included only 3rd + lactations were clustered for each type of time series and variable. The analysis of all lactations groups aims to investigate the patterns across parity and states, and the analysis of 3rd + lactations aimed to distinguish state classification further. At the end and in total, we had 6 variables (milk yield, protein yield, fat yield, ECM yield, protein percent, and fat percent) $\times$ 3 data types (raw data, trend, and seasonal) × 2 lactation type analyses (all parity and only 3rd + lactations) = 36 hierarchical clusters.

5.3.4 Optimal cluster number identification

After building the hierarchical output trees, we measured the clustering performance and identified the optimal number of clusters. To find the optimal number of clusters for both groups, we used three metrics: the Silhouette index (based on the ratio of the average distance of its cluster and the nearest cluster, Silhouette, 1987), the Calinski-Harabasz index (the ratio of between cluster dispersion and inter-cluster dispersion, Calinski and Harabasz,1974), and the fusion level graph (visual observation of fusion level graph plotted against the heights of linkage).

Silhouette index. The Silhouette index measured the average relative ratio of each object's distance to the other objects within its cluster and the average distance to the objects in the closest cluster. The index was defined as:

$$\bar{S} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{b(i)-a(i)}{max\{a(i),b(i)\}}\right) \tag{5.7}$$

Where i is ith object (state); $a(i)$ is the average distance of object i and other objects in the same cluster; $b(i)$ is the average distance of the object i to the nearest cluster other than its cluster. For the clusters of all lactations, $n = 51$; for 3^{rd}+ lactations, $n = 17$. The Silhouette index of cluster numbers from 2 to 10 was obtained for each interested variable. The higher Silhouette index means the clusters are denser and better separated.

Calinski-Harabasz index. The Calinski-Harabasz index was based on the degree of dispersion between clusters. It was calculated as the ratio of overall between cluster variance and the overall inter-cluster variance. The index was defined as:

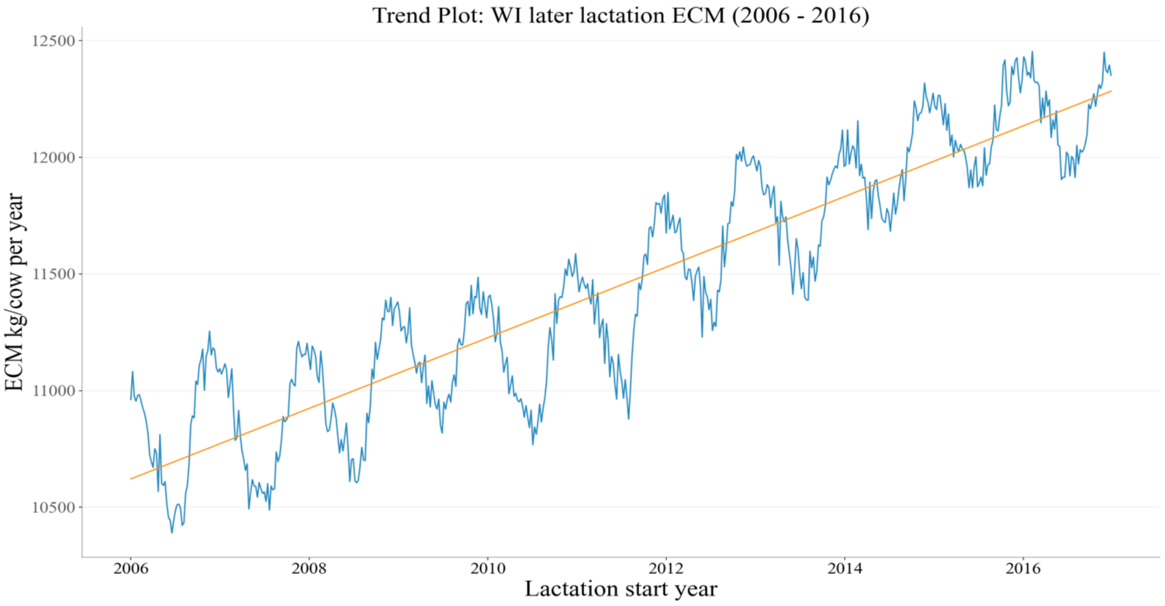
$$CH_K = \frac{(N-K)\sum_{i=1}^{K} n_i ||m_i - m||^2}{(K-1)\sum_{i=1}^{K}\sum_{x\in c_i} ||x - m_i||^2} \tag{5.8}$$

Where K was the corresponding number of clusters, N was the number of objects, m_i was the centroid of cluster i, m was the overall mean, n_i was the number of points in cluster i, x was one object, and c_i was the set of points in cluster i. The Silhouette index of cluster numbers from 3 to 10 was obtained for each interested variable. The higher the Calinski-Harabasz index, the better the performance. Each of the 3 metrics (Silhouette index, Calinski-Harabasz index, and the fusion level graph) of each cluster suggested an optimal number for that cluster and was counted as a vote; the most voted number of clusters was selected.

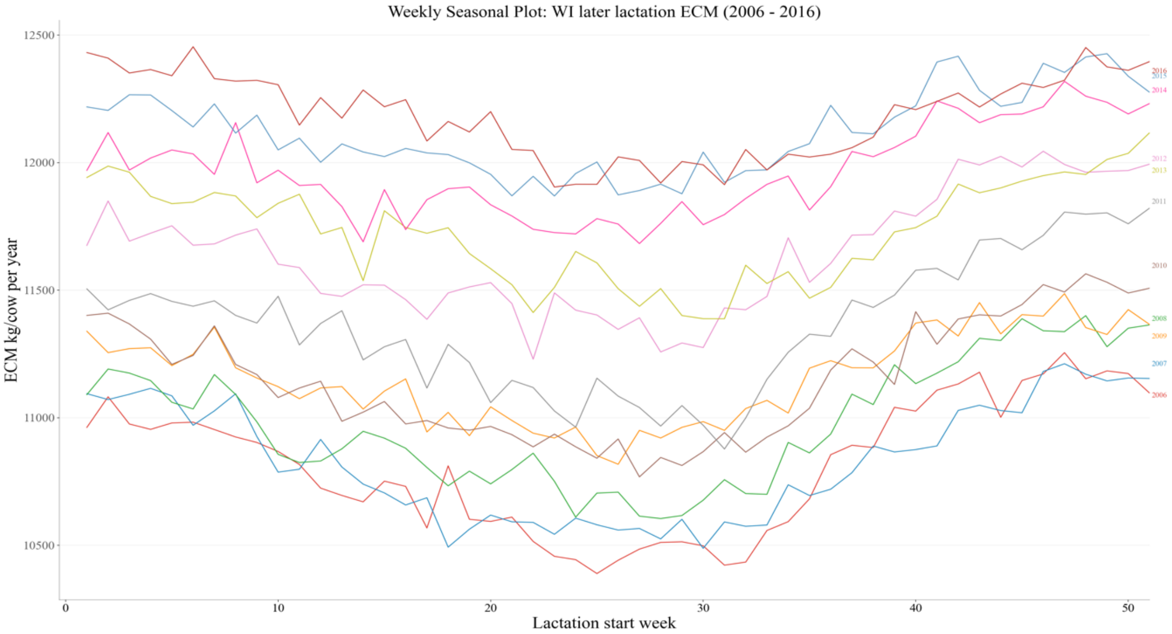
5.4 Results and discussion

5.4.1 Time series decomposition

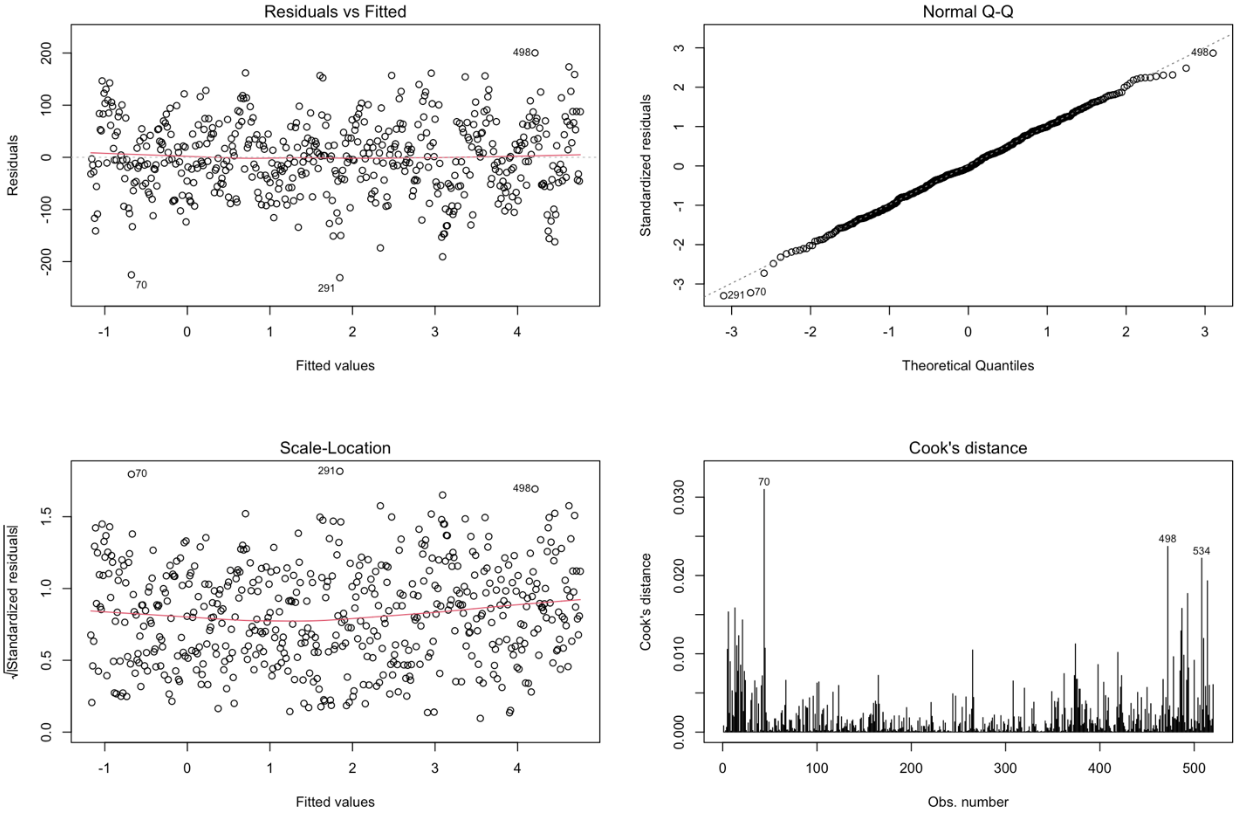
All 306 time-series decomposed in this study resulted in non-zero estimates for the yearly trend and seasonality components. We used Wisconsin's 3rd + lactations ECM yield as an illustrative example (Figure 5.2). Its yearly trend, seasonality, and residual plots after removing the yearly trend and seasonality of the original time-series are shown in Figure 5.2. First, the trend showed linearity, and the average annual yield increased by 1.15% per year over the 11 years studied (2006-2016) (Figure 5.2(A)). Second, seasonal patterns resulted in a production drop during the summer months and an increase in winter (Figure 5.2(B)). Third, the residual plots showed satisfaction with the normality assumption of the residuals (Figure 5.2(C)).


Trend Plot: WI later lactation ECM (2006 - 2016)
ECM kg/cow per year
12500
12000
11500
11000
10500
2006
2008
2010
2012
2014
2016
Lactation start year

(A)


Weekly Seasonal Plot: WI later lactation ECM (2006 - 2016)
ECM kg/cow per year
12500
12000
11500
11000
10500
0
10
20
30
40
50
Lactation start week
2016
2015
2014
2012
2013
2011
2010
2008
2009
2007
2006

(B)



(C)

FIGURE 5.2: Time series components represent energy corrected milk (ECM) for 3rd+ lactations in Wisconsin. ECM was calculated from lactation milk yield, protein percent, and fat percent of records for lactations starting from Jan 2006 to Dec 2016. 843,998 ECM of 3rd+ lactations in Wisconsin was averaged over each week. Time series components are represented as (a) trend, as fitted linear in the orange line with the blue line as raw data; (b) seasonality, with repeated patterns of raw data from each year, represented in one-year time scope; and (c) remainders, tested for normality assumption of remainders after removing trend and seasonality components.

The trend component reflected the long-term change of all yields (ECM, milk, fat, and protein) or component percentage over the time spanned by the series. The management level, new technologies, and breed selections often affected trend. The seasonal component reflected a cyclical process in which the feature across different seasons varies among states because of the impact of seasonal factors. The remainder (noise) component was likely to represent the short-term impacts on the production system.

All decompositions showed non-zero strength for trend and seasonality (Table 5.2). The trends' strength ranged from 0.14 (for 2nd lactation protein percent in Vermont) to 0.98 (for protein yield in Wisconsin), with an average of 0.68. The strength of seasonality ranged from 0.09 (for 1st lactation protein yield in Florida) to 0.92 (for 3rd + lactations protein percent in Wisconsin) with an average of 0.42. The average strength of trend across states for ECM, milk yield, fat yield, and protein yield varied from 0.73 to 0.78, and for fat and protein percent, it was 0.53 and 0.50, respectively. The strength for the trend of fat and protein percent was weaker than all yields (ECM, milk, fat, and protein) because the component percentage had a minor long-term change tendency compared with the yield. The average strength for seasonality across states for ECM, milk yield, fat yield, and protein yield varied from 0.37 to 0.47, and for fat and protein percent, from 0.28 to 0.58, respectively. The fat percent had a weaker seasonality than all yields (ECM, milk, fat, and protein), but protein percent had a stronger seasonality than all yields (ECM, milk, fat, and protein). This indicated that the seasonal effect of protein percent was more apparent when separated from the remainder.

Across lactation groups, the average strength for the trend for each variable was not significantly different $(p > 0.58)$ from each other. This meant the long-term trend is not different for all parities. In contrast, the average strength for seasonality of all yields (ECM, milk, fat, and protein) was significantly different $(p < 0.05)$ across lactation groups. The higher lactation numbers had a more substantial effect of seasonality on each yield, consistent with the results reported by Salfer et al. (2019). However, the average strength for the seasonality component of fat percent was smaller for later lactations (0.33 for 1st lactation, 0.27 for 2nd lactation, and 0.22 for 3rd + lactations). The average strength for the seasonality component of protein percent did not vary across lactation groups and had the greatest strength for seasonality (0.57-0.59).

The trends and seasonality components from Wisconsin, Minnesota, and New York, the states with the most data, were the strongest (first quantile) among states. The reduced variance of the remainders could be explained through the averaging step of getting the value of each week in the times series. The more data used in the averaging process, the better population characteristics were represented. This removed residual variability was making the estimation of the trend and seasonal pattern across those states stronger. Conversely, Texas and Georgia had the weakest (last quantile) trend and seasonality components, which could result from fewer data points available for averaging.

TABLE 5.2: Strength for trend and seasonality for each time-series decomposition

State	Strength of trend[2]						Strength of seasonality[2]					
lactation[1]	ECM	MY	FY	PY	F%	P%	ECM	MY	FY	PY	F%	P%
CA1	0.30	0.29	0.29	0.23	0.10	0.27	0.87	0.82	0.86	0.89	0.52	0.55
CA2	0.35	0.33	0.35	0.25	0.13	0.28	0.87	0.83	0.83	0.90	0.34	0.43
CA3	0.41	0.42	0.39	0.33	0.16	0.31	0.86	0.83	0.81	0.89	0.42	0.51
FL1	0.12	0.16	0.13	0.09	0.12	0.40	0.80	0.77	0.76	0.84	0.40	0.51
FL2	0.27	0.34	0.24	0.21	0.14	0.37	0.76	0.73	0.72	0.81	0.41	0.51
FL3	0.40	0.45	0.37	0.31	0.12	0.37	0.73	0.70	0.67	0.78	0.26	0.53
GA1	0.11	0.14	0.11	0.09	0.18	0.39	0.71	0.67	0.71	0.70	0.27	0.18
GA2	0.19	0.28	0.15	0.18	0.22	0.37	0.73	0.70	0.71	0.71	0.21	0.16
GA3	0.32	0.34	0.31	0.26	0.14	0.29	0.75	0.69	0.75	0.74	0.26	0.22
IA1	0.34	0.40	0.30	0.30	0.36	0.61	0.87	0.74	0.90	0.86	0.80	0.71
IA2	0.50	0.58	0.46	0.38	0.34	0.69	0.86	0.76	0.88	0.85	0.72	0.64
IA3	0.63	0.62	0.63	0.55	0.19	0.66	0.85	0.75	0.87	0.85	0.77	0.79
IL1	0.15	0.18	0.18	0.12	0.30	0.56	0.72	0.63	0.73	0.71	0.33	0.35

IL2	0.33	0.41	0.30	0.23	0.27	0.61	0.69	0.63	0.68	0.67	0.26	0.29
IL3	0.45	0.48	0.45	0.33	0.15	0.61	0.67	0.59	0.66	0.69	0.22	0.43
IN1	0.28	0.28	0.29	0.23	0.18	0.49	0.71	0.61	0.71	0.73	0.31	0.41
IN2	0.27	0.29	0.27	0.19	0.19	0.42	0.69	0.67	0.64	0.72	0.15	0.21
IN3	0.45	0.41	0.46	0.38	0.19	0.42	0.76	0.69	0.75	0.77	0.32	0.25
MA1	0.25	0.25	0.25	0.23	0.22	0.41	0.48	0.41	0.52	0.44	0.16	0.56
MA2	0.34	0.34	0.34	0.27	0.24	0.44	0.50	0.48	0.50	0.45	0.15	0.48
MA3	0.42	0.44	0.43	0.32	0.14	0.46	0.57	0.58	0.55	0.54	0.18	0.42
MI1	0.46	0.57	0.38	0.39	0.40	0.70	0.82	0.79	0.77	0.89	0.52	0.90
MI2	0.55	0.63	0.48	0.45	0.23	0.75	0.79	0.79	0.65	0.89	0.50	0.78
MI3	0.75	0.79	0.70	0.71	0.30	0.82	0.88	0.89	0.78	0.94	0.65	0.84
MN1	0.37	0.54	0.26	0.41	0.59	0.83	0.95	0.90	0.96	0.95	0.90	0.89
MN2	0.63	0.72	0.62	0.42	0.43	0.89	0.93	0.87	0.94	0.93	0.84	0.86
MN3	0.80	0.83	0.79	0.72	0.36	0.89	0.93	0.89	0.92	0.94	0.78	0.86
NY1	0.38	0.53	0.33	0.31	0.64	0.79	0.95	0.91	0.96	0.94	0.90	0.75
NY2	0.58	0.67	0.56	0.39	0.51	0.84	0.95	0.90	0.96	0.93	0.86	0.58
NY3	0.82	0.82	0.82	0.74	0.41	0.86	0.96	0.94	0.97	0.95	0.82	0.48
OH1	0.28	0.38	0.24	0.22	0.32	0.61	0.79	0.68	0.80	0.79	0.60	0.63
OH2	0.48	0.52	0.46	0.33	0.18	0.63	0.71	0.61	0.70	0.71	0.48	0.47
OH3	0.69	0.68	0.68	0.61	0.14	0.66	0.76	0.65	0.76	0.78	0.49	0.61
PA1	0.60	0.66	0.55	0.61	0.65	0.86	0.91	0.84	0.92	0.87	0.79	0.33
PA2	0.76	0.77	0.75	0.67	0.53	0.86	0.92	0.86	0.93	0.89	0.74	0.29
PA3	0.90	0.90	0.89	0.87	0.40	0.89	0.94	0.91	0.94	0.93	0.68	0.36

SD1	0.21	0.22	0.19	0.19	0.17	0.35	0.81	0.70	0.85	0.76	0.68	0.48
SD2	0.23	0.23	0.22	0.18	0.10	0.29	0.78	0.69	0.79	0.76	0.55	0.50
SD3	0.36	0.36	0.35	0.30	0.12	0.36	0.79	0.73	0.76	0.79	0.39	0.38
TX1	0.31	0.26	0.30	0.33	0.10	0.43	0.39	0.56	0.50	0.47	0.83	0.49
TX2	0.19	0.19	0.20	0.15	0.14	0.38	0.24	0.43	0.39	0.36	0.80	0.35
TX3	0.31	0.31	0.32	0.22	0.09	0.33	0.20	0.47	0.39	0.36	0.83	0.44
VA1	0.31	0.36	0.30	0.20	0.27	0.56	0.76	0.74	0.75	0.73	0.36	0.33
VA2	0.51	0.55	0.49	0.40	0.28	0.58	0.68	0.65	0.68	0.65	0.36	0.24
VA3	0.67	0.69	0.65	0.60	0.26	0.67	0.67	0.61	0.66	0.68	0.40	0.40
VT1	0.16	0.23	0.14	0.12	0.22	0.48	0.72	0.63	0.75	0.70	0.47	0.27
VT2	0.38	0.42	0.36	0.28	0.15	0.50	0.76	0.68	0.76	0.74	0.35	0.14
VT3	0.51	0.50	0.50	0.43	0.11	0.51	0.81	0.75	0.81	0.80	0.29	0.15
WI1	0.51	0.67	0.42	0.50	0.76	0.89	0.98	0.97	0.98	0.98	0.90	0.94
WI2	0.73	0.79	0.71	0.58	0.58	0.89	0.98	0.96	0.97	0.98	0.85	0.90
WI3	0.88	0.90	0.87	0.85	0.52	0.92	0.98	0.97	0.97	0.98	0.80	0.92

[a] For 3^{rd}+, marked as 3.

[b] ECM=Energy corrected milk, MY = milk yield, FY = fat yield, PY = protein yield, F% = fat percent, P% = protein percent

After measuring the strength of the decomposed components, we further investigated the ratio between the variance of trend or seasonal effect and the variance of original time series, which further showed the relative importance of the trend or seasonal component. The comparison between the ratio of the trend or seasonal variance to the original time series variance is shown in Figure 5.3. These graphs examined the relationship between the trend and seasonality associated with each time series and its characteristics. Most of the time series were characterized by a trend component that explains most of the variability of the original data. The seasonality displayed a smaller amplitude in most of the time series except for the protein percent, which indicated the long-term change on trend for protein percent was mild compared to the stable seasonal pattern shown for most of the time series. The seasonal change in protein percent could be because of the "endogenous annual rhythm" which controls milk synthesis (Salfer et al., 2019). Also, the trend component accounted for the majority of the time-series variance in states like Wisconsin, Minnesota, and California for all the yield-related variables, an indication of the constant improvement in production level in the past years (USDA-ERS, 2021). On the other hand, the seasonality components were a larger portion of the time series variance from 3rd + lactations in Virginia, Ohio, and Pennsylvania. Salfer et al. (2019) reported annual rhythm patterns of milk yield in 10 of 11 herds, and fat and protein concentration in all 11 studied herds in Pennsylvania. Although southern states reported the most seasonal changes in milk production (Ferreira et al., 2020), the homogeneity of the seasonal pattern across herds could be a possible reason for the higher proportion of variance explained by the seasonal component for Virginia, Ohio, and Pennsylvania.

Moreover, from the ratio of trend and seasonal contrast plot (Figure 5.3), the dots from the same state were not apart in terms of trend ratio, whereas for the seasonal ratio, later lactations often had a higher seasonal ratio except for the fat percent. This means the long-term pattern is more likely to explain the proportion of variance at the same scope for all lactations of one state. The seasonal pattern related to the yield variables explained more of the proportion of variance for later lactations. The earlier the lactation, the larger the variance ratio between seasonal component and original time series was for the fat percent, indicating a more sensitive reaction to seasonal factors for earlier lactations in terms of fat percent. On the other hand, for the protein percent, the seasonal component explained more variance than the trend component, revealing that the protein percent had only minor changes throughout the years and was more sensitive to the cyclic factors from the seasonality than the other factors influencing long term changes. These findings could help, for example, to advise management decisions such as taking reproductive measurements to limit the number of cows calving during those low of the calving time to achieve higher protein percent in states like Pennsylvania and New York or breeding heifers to calving in those weeks that led to higher fat percent for Michigan, Illinois, Pennsylvania, and Wisconsin.

5.4.2 Time series clustering

To further investigate the similarity and dissimilarity among states, we performed time series clustering on raw data and the trend and seasonality time series produced through the decomposition analysis. The clustering approach enabled the partition of variable components into clusters based on distance measurements. Time series clustered into the same group were considered similar. The DTW distance, which was broadly used in the time series clustering process, was used for raw data clustering; it measured the similarities in the shape of two time series by enabling a one-to-many match between the two series (Izakian et al., 2015). The warping functions in DTW remapped the time indices dynamically for the pair of time series and found the optimal alignment that minimizes the cross-distance matrix between two time series (Giorgino, 2009). Hierarchical clustering did not require information of number clusters a priori, leaving us to define a suitable number of groups for each variable. Based on the clustering dendrogram, Silhouette index, Calinski-Harabasz index, and the fusion level graph, we picked 5 as the optimal number of clusters for all parity groups and 3 for the 3rd + lactations.

For the clusters on raw data (Figure 5.4(A)), ECM, milk yield, fat yield, and protein yield had similar patterns for classification in which the 1st lactations were mainly grouped into two clusters relatively independent from 2nd and 3rd + lactations. A visual assessment of the 1st lactation

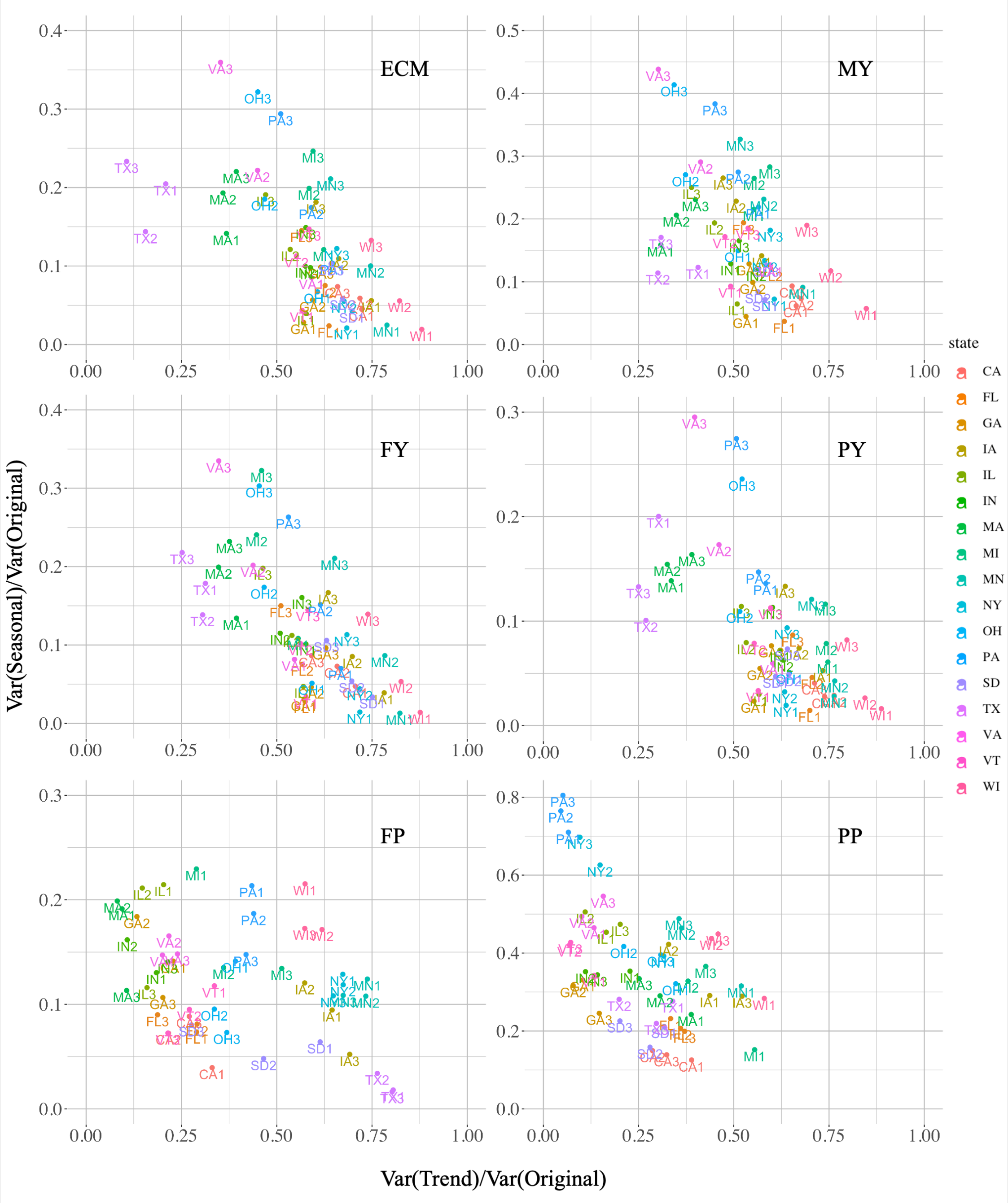


FIGURE 5.3: Graphical comparison of the ratios of variance of trend or seasonality against the variance of original time series for each variable. **MY** = milk yield; **FY** = fat yield; **PY** = protein yield; **FP** = fat percent; **PP** = protein percent

clusters suggested a lower production level and less variability. The 2nd and 3rd + lactations of the same state were commonly grouped into the same cluster, and the centroid patterns of the two major 2nd and 3rd + lactations clusters showed a difference between the production levels. The fact that the 1st lactation time series were away from the later lactations could be explained by the production level difference. On average, the production of 1st lactation and 2nd lactations were 85.4% and 98.3% of the 3rd lactations in our dataset. According to the extent of centroid pattern for each cluster, milk yield showed more extent than ECM, and fat yield shows more variate than protein yield, vertically. The centroid pattern for fat percent and protein percent was different from all yields (ECM, milk, fat, and protein) and each other. For fat percent, regardless of lactation number, the time series of the same state were grouped into the same cluster. This indicates the factors influencing the fat percent worked on the geographic level regardless of the parity effect. Further looking at each group, the raw data fat percent (cluster 5) only included all lactations from Texas, and the centroid patter of this cluster was distinct from the other clusters, which had a rise in the early years and a drop in the late years where the others generally had a slightly upward trend for those years. For protein percent, the 1st and 2nd lactation from the same state were often grouped into the same cluster, and the 3rd + lactations were separated. Interestingly, those states clustered into the same group (cluster 5) with clear upward trends of milk yield, like Wisconsin, California, South Dakota, Michigan, and Iowa, did not change much in the long term of protein percent. Furthermore, in centroid pattern, fat percent showed more variate and less cyclic than protein percent in general, which meant the protein percent was less sensitive to other factors other than seasonal factors.

Although the lactation lengths were shown not statistically different across years and states (figure 5.1), there seemed to have some trend like later years, and California, Wisconsin, Michigan had shorter lactation lengths than others. While even later years and those states had a shorter lactation length, they had higher production than earlier years and other states. This shows the difference in lactation length didn't bring bias to our analysis.

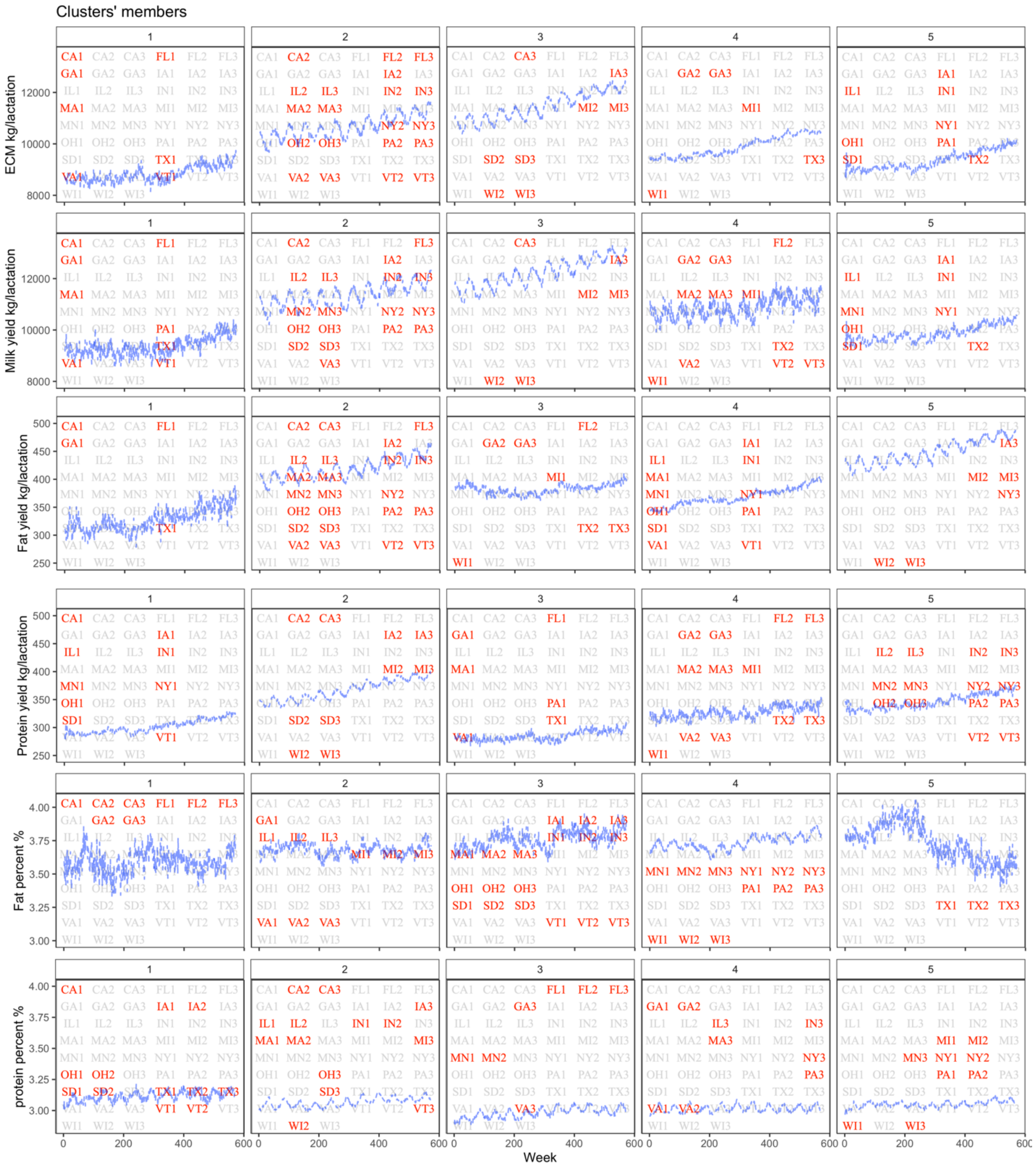
For clusters of trends (Figure 5.4(B)), the centroid patterns were like the raw data in terms of having 2nd and 3rd + lactations from the same state into the same cluster for all yields (ECM, milk, fat, and protein), all lactation from the same state into the same cluster for fat percent, and 1st and 2nd lactations of the same state into the same cluster. The 3rd + lactations from Wisconsin, Michigan, California, South Dakota showed the highest level, and the 1st lactation from California, Georgia, Florida, Massachusetts, Texas showed the lowest level of all yields (ECM, milk, fat, and protein). Moreover, the upward trend can be found in most yields (ECM, milk, fat, and protein) cluster centroid patterns. In contrast, the fat and protein percent trends were relatively flat, and Texas even had a downward trend for fat percent (cluster 4) and protein percent (cluster 5). The fat percent showed most states, except for Texas (cluster 4) and Vermont (cluster 5), had a drop on about week 180 (Jun 2009). Florida showed the lowest fat percent (cluster 1) and protein percent (cluster 3) level, whereas California had the lowest fat percent level (cluster 1), but the highest level for protein percent (cluster 1) for 1st and 2nd lactations.

For clusters on the seasonal pattern (Figure 5.4 (C)), all the yield clusters have one primary cluster for each of 1st, 2nd, and 3rd + lactations. For milk yields, the slope of the seasonal centroid pattern was more significant for later lactation clusters. The valley of the centroid patterns happened during summer weeks (mainly week 30 which was the last week of July for 2nd lactation (cluster 2), and week 28 or the 2nd week of July for 3rd lactation (cluster 4)). The peak appeared in week 4 (late Jan) for 2nd lactation and week 49 - 50 (mid-Dec) for 3rd + lactations. For the 1st lactation (cluster 1), the milk yield centroid pattern did not vary much before August; it increased for 5 weeks before dropping. Virginia, 2nd lactation of Illinois, and 2nd, 3rd + lactations of Georgia (cluster 5) showed a different milk yield centroid pattern, which before the valley, the yield did not drop until May. Whereas the fat yield in some groups, especially in the groups with 3rd + lactations (cluster 5), had a valley before week 30 (late July), and protein yield for 3rd + lactations (cluster 3) had a valley between late Jun and early July. Those valleys of components happened earlier than the milk yield, which caused an earlier valley for the ECM. The fat and protein percent, in which the 2nd and 3rd + lactations from the same state were clustered together, had slopes across all lactations that do not appear different from each other. Interestingly, the peaks of the centroid patterns for fat percent happen during summer weeks (particularly early August) and a little after the peak of the protein percent. The fat percent centroid patterns with 2nd and 3rd + lactations (cluster 5) of Minnesota, Wisconsin, Michigan, New York, and Pennsylvania had a relatively flat fat percent before May and raised to the peak in mid-August then dropped until the end of December where for 1st lactation, the peak patio was reached in late Jun. The seasonal pattern for 2nd and 3rd + lactations (cluster 1 and cluster 2) of California, Florida, Illinois, Indiana, Iowa, Ohio, Virginia, Georgia, Massachusetts were not as clear as the others; generally, they had valleys in spring. The cluster of protein percent with mainly 1st lactations (cluster 2) showed a clear seasonal pattern with peak time during July and August. Moreover, protein percent had a solid and clear seasonal pattern across states and lactations, with peaks around late July and early August, declining steadily from November to March or April, then increased.

The average DIM in this dataset was 386, which means the lactation production would likely include production time across the whole year and 2-4 weeks more the following year after calving. Pennsylvania had an average DIM at 384, which included the production of one year plus 2.7 weeks. When comparing all Pennsylvania seasonal curves with results in Salfer et al. (2019, Figure 5.4), the lowest points in our curves would occur earlier than in their curves. The curve for 1st lactation did not follow the same pattern because the 1st lactation variance had large randomness.

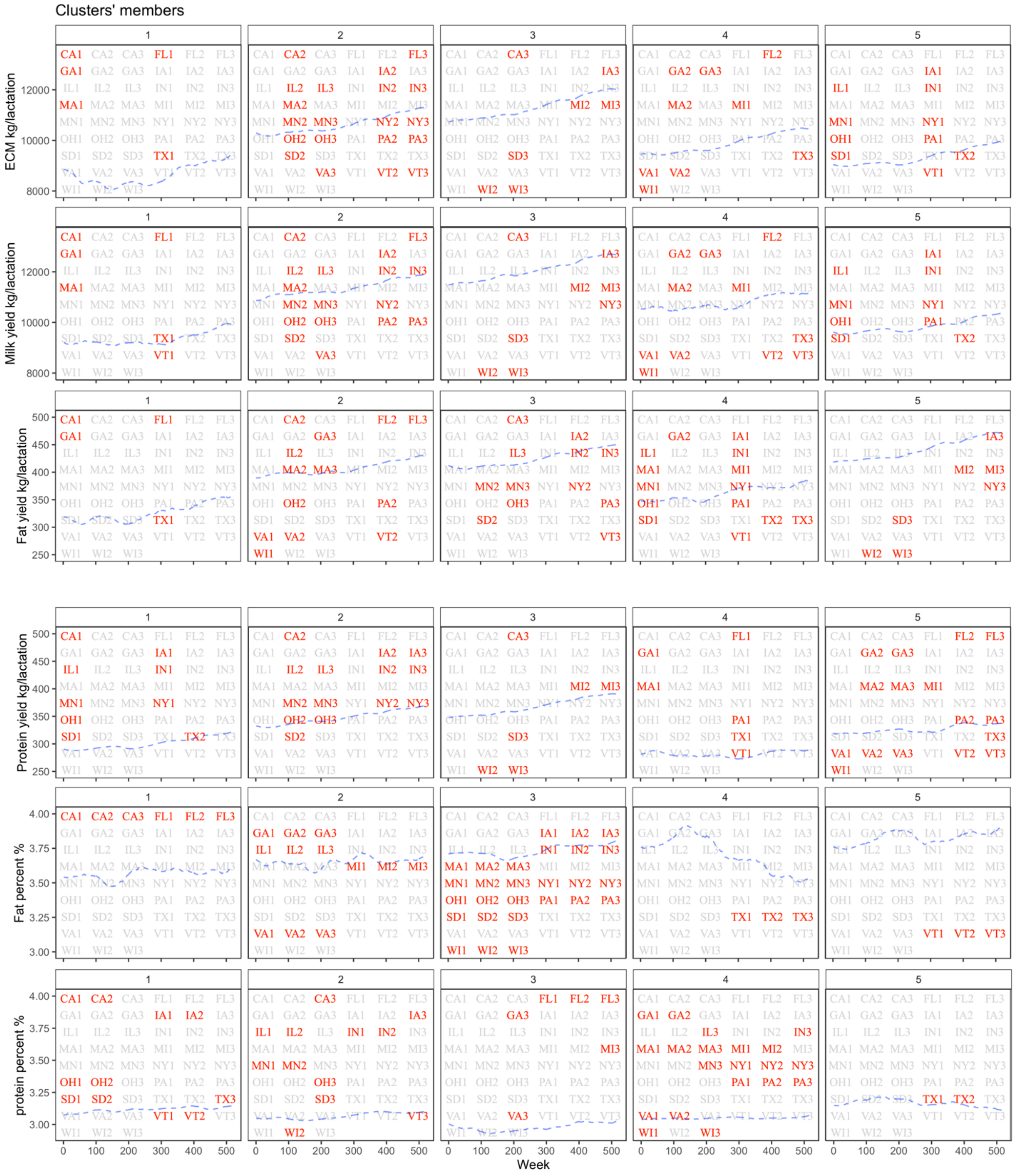
Also, the fat and protein percent curves were not showing the relationship between our lactation-level percentage and their daily percentage representation. Understanding the seasonal pattern would help choose the optimal time for breeding to control calving time to help the farm manage lost from lower milk or milk component yield based on the management goals. In general, first-time calving time would not result in much difference in the production of 1st lactation, except calving in November. Farms that value the milk production more than components production in the same group of states with New York should avoid calving at week 30 and have more calving during the wintertime for cows with parity greater than 2. In contrast, calving between November and May would not lead to much difference in the final production for cows in Georgia.

The milk component could be affected by several factors such as breed and genotype, the difference in feed and feeding pattern, the impact from the season, and climate conditions such as heat or cold stress. Much research related to lactation patterns (Vieira-Neto et al., 2017; Atashi et al., 2020) had the 2nd lactation treated in the same group as the 3rd + lactations. Our cluster result showed that the 2nd lactation productions were not clustered into the same group with the 3rd + lactations in the same state, especially for protein percent.

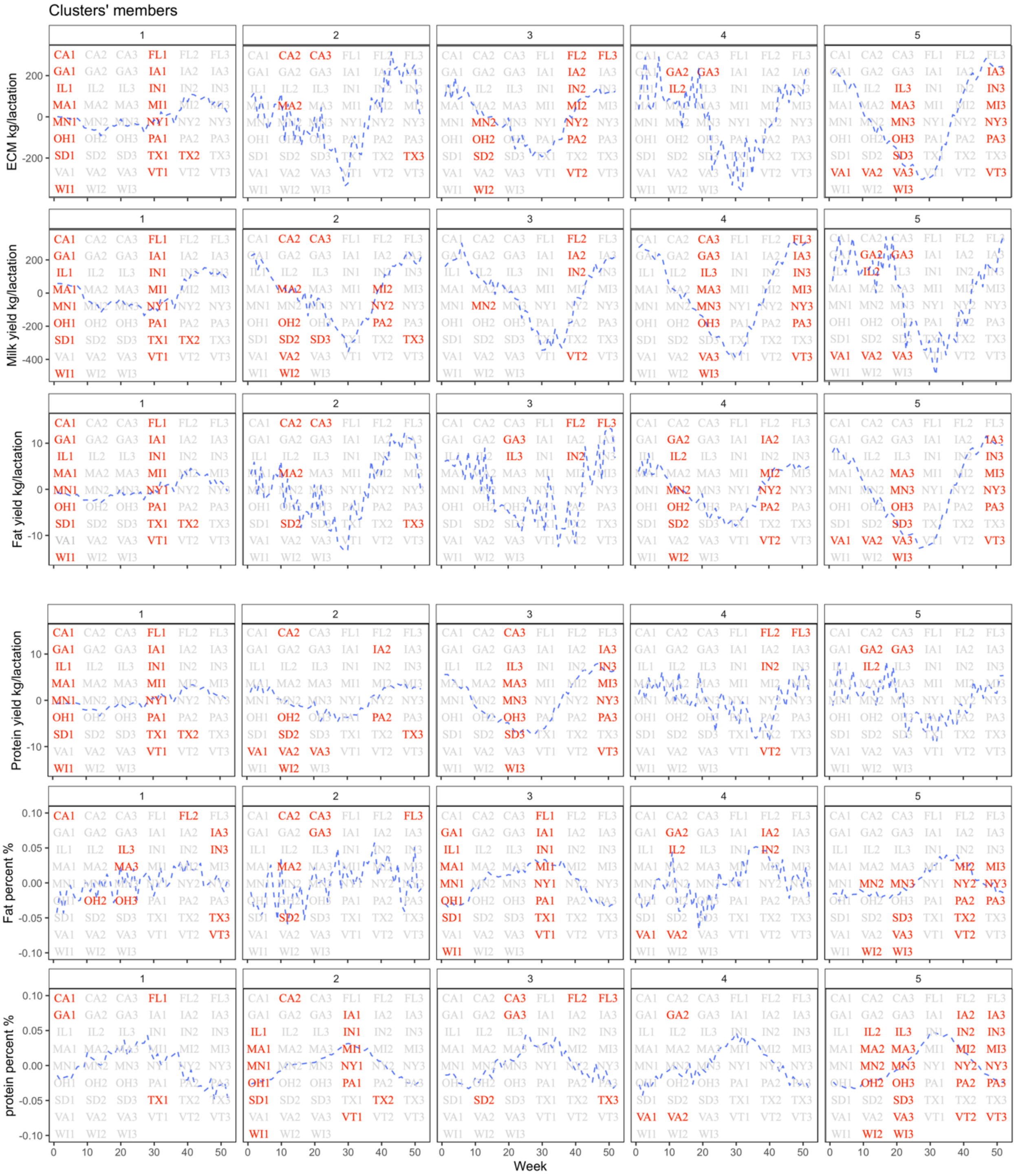


(A)

Geographic representations of clusters for 3rd + lactations were represented in Figure 5.5.


Clusters' members
1
2
3
4
5
ECM kg/lactation
Milk yield kg/lactation
Fat yield kg/lactation
Protein yield kg/lactation
Fat percent %
protein percent %
8000
10000
12000
250
300
350
400
450
500
3.00
3.25
3.50
3.75
4.00
0
100
200
300
400
500
Week
CA1 CA2 CA3 FL1 FL2 FL3
GA1 GA2 GA3 IA1 IA2 IA3
IL1 IL2 IL3 IN1 IN2 IN3
MA1 MA2 MA3 MI1 MI2 MI3
MN1 MN2 MN3 NY1 NY2 NY3
OH1 OH2 OH3 PA1 PA2 PA3
SD1 SD2 SD3 TX1 TX2 TX3
VA1 VA2 VA3 VT1 VT2 VT3
WI1 WI2 WI3

(B)

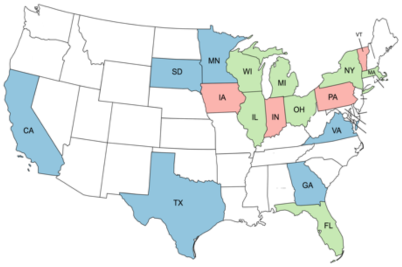


(C)

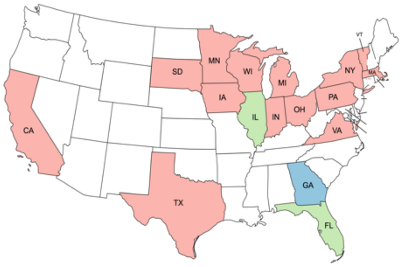
FIGURE 5.4: Cluster groups and group centroids pattern representation of (a) raw data, (b) trend, and (c) seasonal. Clustered states are shown in red, with the respective curve of the centroid pattern of each cluster shown in blue.

We followed USDA-ERS (2021) regions on US dairy states to group our data into 8 different states/regions. Those included: 1) Lake: Michigan, Wisconsin, Minnesota; 2) Corn Belt: Iowa, Illinois, Indiana, Ohio; 3) Northeast: Pennsylvania, New York, Vermont, Massachusetts; 4) Appalachian: Virginia; 5) Southeast: Georgia, Florida; 6) Southern Plains: Texas; 7) Northern plains: South Dakota; and 8) West Coast: California. From our cluster results, the states in the Lake region did not usually cluster into the same group. Similarly, the states in the Southeast, Northeast, and Corn Belt regions also did not often cluster together, suggesting improvements to how states were grouped to characterize milk yield or milk components. Management practices focusing on adopting the seasonality could be customized for farms. There were more similarities among yield clusters than percent clusters. The seasonality clusters showed the most apparent spatial patterns, concluding that the state's geographic location largely affects the seasonality of protein percentage. Understanding the seasonal pattern of milk yield and milk components in each state could inform farm management decisions like delaying breeding to avoid low reproductive performance or calving at hot periods (DeJarnette et al., 2007) in certain states. Furthermore, the difference in seasonal patterns among states could address the value of having a more precise evaluation of the productivity of each region for simulation models.

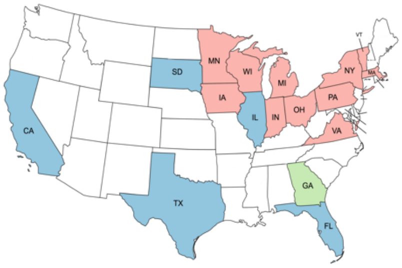
Seasonal clusters on 3rd and later lactations



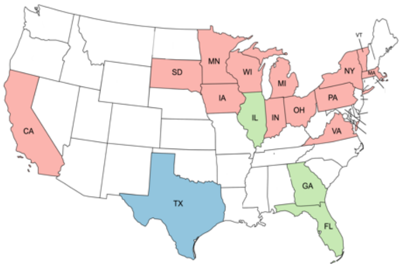
Milk yield



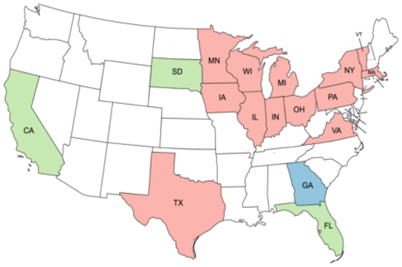
Fat yield



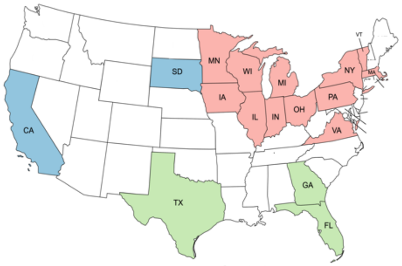
Protein yield



ECM

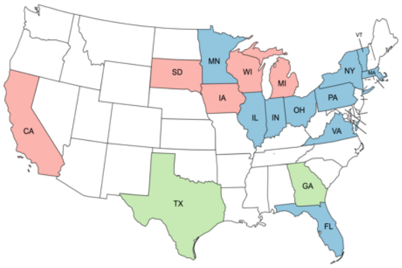


Fat%

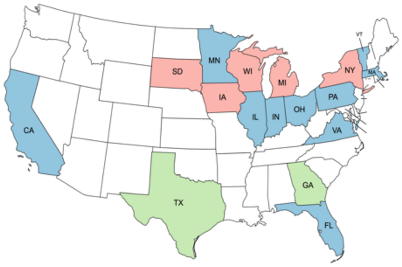


Protein%

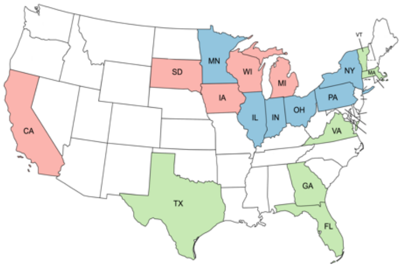
Raw data clusters on 3rd and later lactations



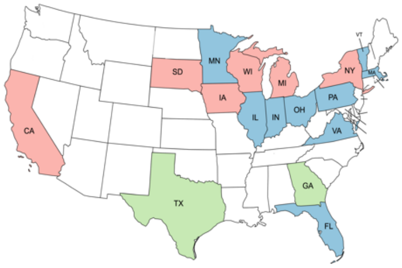
Milk yield



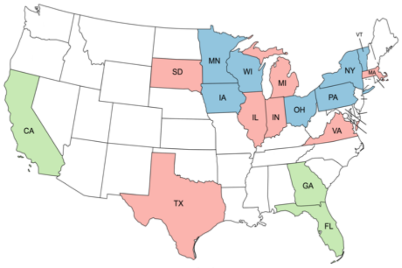
Fat yield



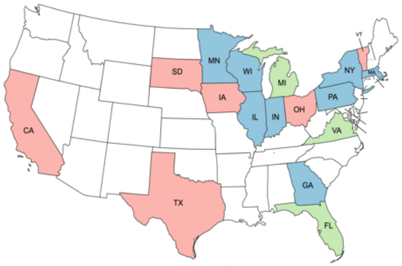
Protein yield



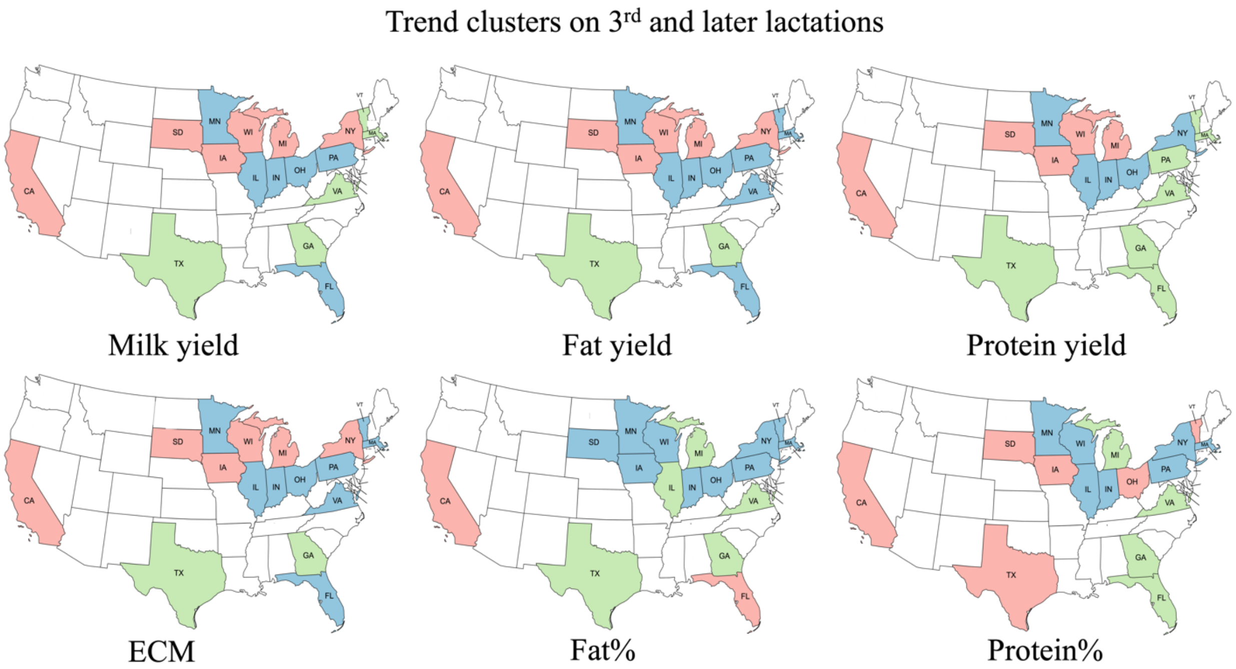
ECM



Fat%



Protein%


Trend clusters on 3rd and later lactations

Milk yield

Fat yield

Protein yield

ECM

Fat%

Protein%

FIGURE 5.5: Geographic representation for clusters of 3rd + lactations on raw data (top), seasonal (middle), and trend (bottom). States of the same color belong to the same cluster.

Modifications by season and trend of expected production would be implemented into the RuFaS model in at least two ways. First, results presented here will be used for simulating a general farm when the farm to be simulated does not have historical lactation records by including seasonal adjustment on the production level. In such a situation, the predicted value of the milk, fat, and protein yield reported here would be used to adjust the scale of the lactation curve according to the lactation start time to correct the expected productivity. The other means of implementing the time series decomposition methods in the RuFaS model would take place when farm historical lactation records are available. In such a situation, farm-specific data could be decomposed and applied to generate farm-specific estimated lactation curves for animal and herd simulations of that farm in the future using the RuFaS model.

5.5 Conclusions

Protein percent had the strongest seasonality component. Time series of all yields (ECM, milk, fat, and protein) were largely influenced by the long-term effect of the trend. This was noticeable when looking at the strength, ratio of trend and seasonality variance, and cluster results. Cluster results showed distinct patterns of the long-term trend and seasonality among lactations from different states. Analysis of the trend showed a general increasing pattern across the years for most parities. In contrast, fat and protein percent had a flat trend being less variable across years. Time series decomposition and clustering distinguished yields and percentages according to parity and location that have different patterns. Clustering within time series provided a practical framework to distinguish seasonal and trend patterns across time as an essential prior knowledge when production estimates are needed. The improved classification of the seasonality and the decomposition parameters can assist with herd and cow management decisions demonstrating the importance of the different seasonal patterns of the production variables according to geographical location and parity.

References

Aghabozorgi, S., A. S. Shirkhorshidi, and T. Y. Wah (2015b). Time-series clustering–a decade review. In: *Inf. Systems* 53, pp. 16–38.

Box, G. E. P. and G. M. Jenkins (1976). *Time Series Analysis: Forecasting and Control*. Revised ed. San Francisco: Holden-Day.

Calinski, T. and J. Harabasz (1974). A dendrite method for cluster analysis. In: *Comm. in Statistics* 3.1, pp. 1–27.

Capper, J. L. and R. A. Cady (2019b). The effects of improved performance in the US dairy cattle industry on environmental impacts between 2007 and 2017. In: *J Anim Sci* 98.

Center, IFCN Dairy Research (2010). *Dairy Report*. Kiel, Germany: International Farm Comparison Network.

Cleveland, R. B., W. S. Cleveland, J. E. McRae, and I. J. Terpenning (1990b). STL: A seasonal-trend decomposition procedure based on loess. In: *Journal of Official Statistics* 6.1, pp. 3–33.

Cole, J. B., D. J. Null, and A. De Vries (2011b). Short communication: Best prediction of 305-day lactation yields with regional and seasonal effects. In: *J Dairy Sci* 94, pp. 1601–1604.

Ferreira, F. C., J. S. Clay, and A. De Vries (2020). Distribution of seasonality of calving patterns and milk production in dairy herds across the United States. In: *J Dairy Sci* 103, pp. 8161–8173.

Giorgino, T. (2009). Computing and visualizing dynamic time warping alignments in R: the dtw package. In: *Journal of Statistical Software* 31, pp. 1–24.

Hansen, T. L., M. Li, J. Li, C. J. Vankerhove, M. A. Sotirova, J. M. Tricarico, V. E. Cabrera, E. Kebreab, and K. F. Reed (2021c). The Ruminant Farm Systems Animal Module: A Biophysical Description of Animal Management. In: *Animals* 11.5, p. 1373.

Hastie, T., R. Tibshirani, and J Friedman (2009). *The elements of statistical learning: data mining, inference and prediction*. Vol. 2nd ed. Berlin: Springer.

Hyndman, R. J. and G. Athanasopoulos (2018). Forecasting: principles and practice. In: *OTexts: Melbourne, Australia* 2nd edition. URL: `OTexts.com/fpp2`.

Izakian, H. and W. Pedrycz (2015). Jamal Iqbal. Fuzzy clustering of time series data using dynamic time warping distance. In: *Eng. Appl. Artif. Intell.* 39, pp. 235–244.

Kebreab, E., K. F. Reed, V. E. Cabrera, P. A. Vadas, G. Thoma, and J. M. Tricarico (2019c). A new modeling environment for integrated dairy system management. In: *Animal Frontiers* 9.2, pp. 25–32. DOI: `10.1093/af/vfz004`.

Lafare, A. E. A., D. W. Peach, and A. G. Hughes (2016). Use of seasonal trend de- composition to understand groundwater behaviour in the Permo- Triassic Sandstone aquifer, Eden Valley, UK. In: *Hydrogeol J* 24, pp. 141–158.

Li, M., V. E. Cabrera, and K. Reed (2018b). Comparison of Holstein and Jersey milk production with a new stochastic model. In: *J. Dairy Sci* 100(Suppl. 2): M-207.

Murtagh, F. and P. Legendre (2014). Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? In: *Journal of Classification* 31.3, pp. 274–295.

R-Core-Team (2020b). *R: A language and environment for statistical computing*. Vienna, Austria.: R Foundation for Statistical Computing. URL: `https://www.R-project.org/`.

Rotz, C. A., M. S. Corson, D. S. Chianese, F. Montes, S. D. Hafner, H. F. Bonifacio, and C. U. Coiner (2015b). The Integrated Farm System Model Reference Manual. In: 4.

Rousseeuw, P. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. In: *J. Comput. Appl. Math.* 20.1, pp. 53–65.

Salfer, I. J., C. D. Dechow, and K. J. Harvatine (2019). Annual rhythms of milk and milk fat and protein production in dairy cattle in the United States. In: *J. Dairy Sci* 102, pp. 742–753.

Sarda-Espinosa, A. (2019). dtwclust: Time Series Clustering Along with Optimizations for the Dynamic Time Warping Distance. In: *R package version 5.5.6*. URL: `https://CRAN.Rproject.org/package=dtwclust`.

Schils, R., L. M., M. H. A. de Haan, J. G. A. Hemmer, A. van den P. Dasselaar, J. A. de Boer, A. G. Evers, G. Holshof, J. C. van Middelkoop, and R. L. G. Zom (2007a). DairyWise, A Whole-Farm Dairy Model. In: *J Dairy Sci* 90, pp. 5334–5346. DOI: `10.3168/jds.2006-842`.

USDA-ARS (n.d.). *Milk cows and production by state and region (Annual)*. URL: `https://www.ers.usda.gov/data-products/dairy-data/`.

USDA-NASS (May 2020b). Agricultural Marketing Service, and Economic Research Service calculations. 29. URL: `https://downloads.usda.library.cornell.edu/usda-esmis/files/c821gj76b/jm2159057/qv33sh51q/agpr0520.pdf`.

Wang, X., K. A. Smith, and R. J. Hyndman (2006). Characteristic-based clustering for time series data. In: *Data Mining and Knowledge Discovery* 13.3, 335–364.

Chapter 6

Conclusion

This thesis focuses on developing an animal life cycle submodel of a whole-farm systems model to simulate individual animal life stories and herd dynamics. Researchers, farm managers, private companies, and policymakers desired a well-documented and integrated software that could simulate farm performance from a holistic perspective of the dairy farm. Developing such a tool is challenging. The working process requires massive efforts such as project management, programming architecture, identifying and prioritizing the needs of potential users, translating research findings and experts' ideas into simulation process, standardizing and documenting the coding rationale, actively maintaining the code, and filling the gaps of essential information for a more accurate representation of simulations.

This thesis first proposes a programming framework to represent animal attributes and herd characteristics with a statistic, infinite, dynamic Monte Carlo simulation model. This framework resulted in an essential component of a whole farm system model, the animal life cycle submodel, which achieved the purpose of:

1. representing the life stage development of animals with bodyweight change, milking performance, reproductive programs, and culling incidences;
2. summarizing herd level representations of herd dynamic, production, and reproductive performance;

A case study using sexed and beef semen demonstrated the functions of the animal life cycle simulation submodel.

This thesis then reported two ways of analyzing a large database of lactation records. The results of those analyses are to be implemented into the animal life cycle simulation submodel as precise lactation curve parameters to represent milking performance. We used a mixed effects model to investigate the effects of fixed and random factors on individually fitted lactation curve parameters and 305-day milk yields. The qualified parity-, year-, month-, region-, and milking frequency-specific lactation curve parameters and 305-day milk yield can be used as an input matrix for the lactation curve parameter identification in the animal life cycle submodel. When a farm is simulated, the input herd's average 305-day milk yield would be compared to the corresponding 305-day milk yield in the resulting matrix according to the region and milk frequency to adjust the overall production. Then specific lactation curve parameters will be applied to animals' parity and milking according to the year and month when the lactation starts. Time-series decomposing and clustering methods are used to identify the long-term trend and seasonality of milk yield and milk component yield/ percentage across parity and state. The expected seasonality and trend features can inform management decisions like reproductive program selection. Furthermore, the clustered state-parity groups and their trend and seasonality features can adjust the overall production on specific states and parity beyond the calving start time.

The animal life cycle submodel framework is flexible and powerful. It can incorporate a future expansion to include more animal attributes such as genetics and disease, and it is designed to interact with other submodels and models within a whole farm simulation model such as the Ruminant Farm Systems mode (RuFaS).

Appendix A

Terms, Definitions, and Pseudocode

Animal life cycle: This is a simulating process that simulates individual animals' life events and herd dynamics. This Monte Carlo stochastic simulation simulates the growth, production, reproduction, and culling of individual dairy cattle animals and overall herd dynamics daily. The herd dynamics, an aggregation of individual animals, is then transferred to other submodels for further modeling. This simulation aims to provide animal and herd information to other submodels and modules with animal and herd dynamics.

Flow of information

A.1 Animal information is carried in the **animal** base file

A.2 Animal events record life stories of animals with the **animal event** file

A.3 Animal daily update: animal goes through life stages from **Calf** (from the day of birth until weaning), **Heifer I** (from weaning until the breeding process starts), **Heifer II** (from breeding until close to calving), **Heifer III** (close to calving until calving), **Cow** (from the beginning of the first calving to culling)

A.4 Animal database is generated and stored with the **animal initialization** file

A.5 Groups of animals from each stage is integrated in the **life cycle** file

Function numbering convention: **(A.1A.A1.1)**: A: animal module in RuFaS model, 1A: life cycle submodel in the animal module, A1: 1st type of information flow in lifecycle submodel, 1:1st function in the code

Monte Carlo Simulation: The life cycle module represents the variation of animal performance through Monte Carlo Simulation methods. (Reuven et al., 2016). Monte Carlo simulation relies on random draws from probability distributions rather than fixed values. In this model, two main strategies are used:

1. Compare a random draw from $U(0,1)\{\rho_{Uniform}\}$ to the probability of an event's occurrence to simulate if that event occurs or not. (RandomUni in this document representing a random draw from $U(0,1)\{\rho_{Uniform}\}$)

2. Select a random draw from a known distribution of animal attributes and assign that value to the instantiation of an individual animal. ($Random \sim$ in this document representing a random draw from the following distribution) Probability distributions used are:

    a. Normal Gaussian: $N(\mu, \sigma)$ where μ is the distribution means, and σ is the standard deviation.

    b. Empirical: $F(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{x_i \leq x}$: where x is the observations from the sample and an indicator function equal to if and otherwise.

A.1 Animal Base

Implemented in `animal_base.py`

Implement standard parameters to store each individual's data during the simulation process. It initializes:

- Animal identity info, which is constant for each animal and carried through its life: Animal ID, breed, birth; date, mature body weight (potential genetic information)
- Animal status info, which updates animal status such as culling and breeding constraints;
- Animal historical records, including bodyweight and pen
- Animal output variables, such as nutrient requirements, nutrient balance, manure excretion, ration formulation settings, etc.

(A.1A.A1.1)

A.2 Animal Events

Implemented in `animal_events.py`

Implement functions for adding events and searching for events. Works with `animal_event_constants.py` (string list of all events) and serves as animal history recording. Animal events are events like birth, reproduction actions, dry, and culling that happens to the animals according to the scheduled time, always accompanies a change in status.

- Add event function: add animal events to the animal's records with animal age, actual simulation day, and event description. Events include:
  1) Milestones of the animal lives could lead to changes in animal life status in the simulation, like birth, wean, artificial inseminations (AI), calving, do not breed, dry, and culling reasons. For example, AI would happen at the AI-day determined by the reproductive program at the start day of that reproductive program, and if it gets pregnant at this AI, the gestation length, preg check dates, and calving date would be scheduled; otherwise, the following reproductive events would be scheduled.
  2) Records of animal lives history, which record events related to the indications of the animal description (like entering the herd, mature bodyweight reached)

(A.1A.A2.1)

- Get the most recent event function: fetch the most recent age for specific events to identify animals' status.

(A.1A.A2.2)

A.3 Individual Animal Daily Update

Individual animals update daily with the event scheduling process. Variables like daily body weight change estimates, nutrient requirement calculations for all stages, and targeted production for lactating cows are calculated and updated every day. At the same time, other variables such as pan allocation, reproductive events, and culling are updated upon the scheduled status change date.

1. Calves

Implemented in `calf.py`

Simulation of calves begins when the calf is born on Day 1 and ends on weaning day. A calf's birth is simulated by initializing an instance of a calf class object on day 1. Five main attributes of the calf are determined at birth: (1) if the calf was stillborn, (2) the calf's gender, (3) the calf's birth weight, (4) if it was sold as a calf, and (5) mature bodyweight of this animal.

Input:

Stillbirth rate: the rate of stillbirth happens

Male calf rate: the rate of male calf based on semen type used to the dam

Keep the female calf rate: the rate of female calf kept on the farm

Birth weight distribution: breed-specific normal distribution with average and standard deviation

Mature body weight distribution: breed-specific normal distribution

Wean day: the time before weaning for calves

```
1.1 initial calf
  1.1.1 Stillbirth
  Determine if the calf is a stillbirth
```

```
If $RandomUni$ < stillbirth rate:
  Still Birth = True
Else:
  Still Birth = False
```

(A.1A.A3.1)

1.1.2 Gender

Determine the gender of the calf

```
If $RandomUni$ < female calf rate:
  Gender = male
Else:
  Gender = female
```

(A.1A.A3.2)

1.1.3 Sold as calf

Determine whether to sell the calf; all male calves and some female calves are sold as a calf

```
If gender = male or gender = female and
    $RandomUni$ < keep female calf rate:
  Sold = True
Else:
  Sold = False
```

(A.1A.A3.3)

1.1.4 Birth weight

Determine the birthweight of the calf based on breed-specific distributions

Birth weight = $Random \sim$ N (avg, std) of birth weight distribution

(A.1A.A3.4)

1.1.5 Mature body weight

Determine the target mature body weight

Mature body weight = $Random \sim$ N (avg, std) of mature body distribution

(A.1A.A3.5)

1.2 update calf

Calves are updated daily

1.2.1 bodyweight change

After birth, a calf's growth is stimulated daily until the day they are weaned. Bodyweight changes every day based on the targeted average daily gain (ADG), targets to double the birth weight by wean.

$Bodyweight_i$ = $Bodyweight_{i-1}$ + calf targeted ADG

Where calf targeted ADG = birth weight/wean day

(A.1A.A3.6)

1.2.2 Wean

Calves are weaned as scheduled, leaving the calf stage and entering the heifer stage.

```
  If days born = wean day:
    Wean day = True
```

(A.1A.A3.7)

2. Heifers I

Implemented in `heiferI.py`

Heifer I simulation starts when the calf is weaned and ends when the heifer starts breeding programs.

Input:

Target heifer pregnant day: the target age of the heifer's first pregnancy

Breeding start day for heifers: the schedule of time for the heifers to start the breeding program

2.1 Nonpregnant heifer body weight change

Heifers are targeted to grow 55% of their mature body weight by pregnancy. A targeted heifer pregnant time is set through the user input. The ADG of nonpregnant heifers is calculated based on the target heifer's pregnancy time and body weight.

Nonpregnant heifer ADG (kg/d) = $\frac{0.55*MBW-BW}{target\ \ age\ \ at\ \ first\ \ pregnancy\ -\ days\ \ born}$

Where MBW is mature bodyweight (0.96*mature bodyweight) and BW is bodyweight (0.96*bodyweight)

(A.1A.A3.8)

2.2 update heiferI

```
Heifers are updated on a daily basis
  2.2.1 bodyweight change
  $Bodyweight_i$ = $Bodyweight_{i-1}$ + non pregnant heifer ADG
                                                        (A.1A.A3.9)
  2.2.2 move to the next stage
  Heifers are moved to the next stage as scheduled
  If days born = Breeding start day for heifers:
    Second stage = True
                                                        (A.1A.A3.10)
```

3. Heifers II

Implemented in `heiferII.py`

Heifer II simulation starts when heifers enter into breeding programs and proceed through the first pregnancy until calving. Three processes are simulated during this animal class: (1) growth, (2) breeding, (3) pregnancy detection

Input:

Gestation length distribution: normal distribution with average and standard deviation

Heifer reproduction method: heifer reproduction method set from the input file

Prefresh day: the time before first calving when the heifer is moved to the next stage

Heifer reproductive cull time: the age of heifer to be culled when not pregnant

Heifer estrus cycle: the estrus cycle of heifers in the ED program

Estrus detection rate: the chance of the estrus to be detected

Estrus service rate: the chance of a detected estrus to be serviced

ED conception rate: the conception rate of the estrus detection method

TAI method for heifer: the user input of protocol picked for the TAI method

5dCGP conception rate: the conception rate when apply 5dCGP protocol

5dCGP conception rate: the conception rate when apply 5dCG2P protocol

User define TAI length: the length of the user define TAI protocol

User define conception rate: the conception rate when apply User define TAI protocol

Synch ED method: the user input of the protocol picked for the synch ED method

Pregnancy check day 1/ 2/ 3: days in pregnancy of doing pregnant checks

Pregnancy loss rate 1/ 2/ 3: pregnancy loss rate between pregnant checks

3.1 pregnant heifer body weight change

3.1.1 average daily gain of the pregnant heifer

Heifers are targeted to grow 82% of their mature body weight by first calving. A gestation length is set at conception. The ADG of pregnant heifers is calculated based on the target heifer's body weight, gestation length, and days in pregnancy.

Pregnant heifer ADG (kg/d) = $\frac{0.82*MBW-BW}{gestation\ \ length\ \ -DIP}$

Where MBW is mature bodyweight (0.96*mature bodyweight) and BW is bodyweight (0.96*bodyweight)

(A.1A.A3.11)

3.1.2 conceptus growth of pregnant heifer

The conceptus weight change along gestation

When DIP < 50,

conceptus growth = 0;

When 50 < DIP < gestation length,

conceptus growth = $3*conceptus\ \ parameter^3*(DIP-50)^2$

When DIP = gestation length,

conceptus growth = - total conceptus weight

Where $total\ conceptus\ weight\ (kg) =\ 0.0148\ *\ gestation\ length\ -\ 2.408\ *\ calf\ birth\ weight$

$conceptus\ parameter\ =\ total\ conceptus\ weight^{\frac{1}{3}}\ /\ gestation\ length\ -\ 50$

(A.1A.A3.12)

3.2 update heiferII

Heifers are updated daily

3.2.1 bodyweight change

From the start of breeding until close to calving, the Heifer II class is increasing its body weight in addition to weight gain for pregnancy.

When DIP = 0 and bodyweight < mature body weight,

$Bodyweight_i$ = $Bodyweight_{i-1}$ + non pregnant heifer ADG

When DIP > 0 and bodyweight < mature body weight,

$Bodyweight_i$ = $Bodyweight_{i-1}$ + pregnant heifer ADG

When bodyweight = mature body weight,

$Bodyweight_i$ = mature bodyweight

(A.1A.A3.13)

3.2.2 breeding program assignment

Breeding program is assigned to heifers when they reach breeding start day.

If days born >= breeding start day for heifers,

reproduction method = Heifer reproduction method

Pregnancy updates follow the reproduction method

(A.1A.A3.14)

3.2.3 move to the next stage

Heifers are moved to the next stage as scheduled

```
When DIP = gestation length - prefresh day,
  Third stage = True
```

(A.1A.A3.15)

3.2.4 heifer reproduction culling

Heifer who failed in reproductive are culled after certain age

```
When days born > heifer repro cull time,
  Cull stage = True
```

(A.1A.A3.16)

3.3 heifer reproductive programs

There are three methods recommended by the Dairy Cattle Reproduction Council (DCRC) implemented in the code, they are (1) estrus detection (ED), (2) timed artificial insemination (TAI), and (3) synchronized estrus detection (synch-ED).

3.3.1 estrus detection method (ED)

It does not require the use of hormones but is based on watching for animal expression of heat. To represent this, the model simulates estrus occurrence and the probability of detecting the estrus event.

3.3.1.1 estrus day determination

Determines the estrus day of each cycle for natural estrus after *breeding start day, detection failure, or AI failure.*

```
Estrus day = estrus start date + $Random \sim$ N (avg, std) of
            heifer estrus cycle
```

Where estrus start date is the day the heifer first enters or re-enters the breeding program after failed estrus detection, failed conception, or an aborted pregnancy. In the case of re-entry, the start date is set equal to the last day of estrus if estrus detection or conception failed; otherwise, it is set to the abortion day for a failed pregnancy.

(A.1A.A3.17)

3.3.1.2 estrus detection update

Update estrus, estrus detection, service upon detection with conception rate specifically related to this method.

```
When days born = estrus day:
  Estrus count += 1
  If $RandomUni$ < estrus detection rate
    Estrus Detection = True,
    if $RandomUni$ < estrus service rate
      AI day = estrus day +1
      conception rate = estrus conception rate
    Else, return estrus
```

(A.1A.A3.18)

3.3.2 Timed Artificial Insemination (TAI)

Uses a series of hormonal injections to synchronize the estrus cycle and perform AI to heifers at the set time from the protocol's start.

3.3.2.1 Program start day determination

Determines the program start day of each synchronized cycle TAI program after breeding start day and AI failure.

For heifers entering the breeding program for the first time:

TAI program start day for heifer = breeding start day

For heifers that did not conceive or lost a pregnancy, restart the TAI program:

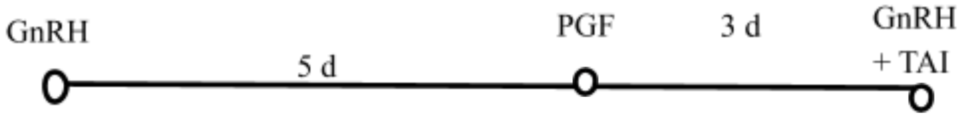
TAI program start day for heifer = abortion day + 1

(A.1A.A3.19)

3.3.2.2 Heifer TAI programs

There are 3 options for TAI programs for heifers. They are 5dCGP, 5dCG2P, and a user-defined option. Each program consists of a series of hormonal injections and has an associated cost and conception rate.

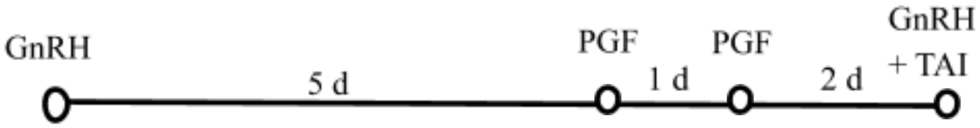
5dCGP:



AI day = tai program start day for heifer + 8

Conception rate = 5dCGP conception rate

5dCG2P:



```
AI day = tai program start day for heifer + 8
Conception rate = 5dCG2P conception rate
```

Define user-defined tai program:

Variables from input: Tai program length, Defined conception rate

```
AI day = tai program start day for heifer + 8
Conception rate = 5dCG2P conception rate
```

(A.1A.A3.20)

3.3.2.3 TAI update

Assign TAI programs to the heifer

```
When days born = TAI program start day for heifer
  Call TAI method for heifer
```

(A.1A.A3.21)

3.3.3 Synch - estrus detection (synch-ED)

Uses hormones to synchronize groups of animals in their cycles and then watches for heat expression after.

3.3.3.1 Synch ED estrus day determination

Determines the estrus day of each cycle for synched estrus after breeding start day or conception failure. Estrus day is determined by adding a normal random variable to the start day.

Synch estrus day = estrus start date + *Random* $\sim$ N (avg, std) of estrus cycle

Synch estrus day < program length

(A.1A.A3.21)

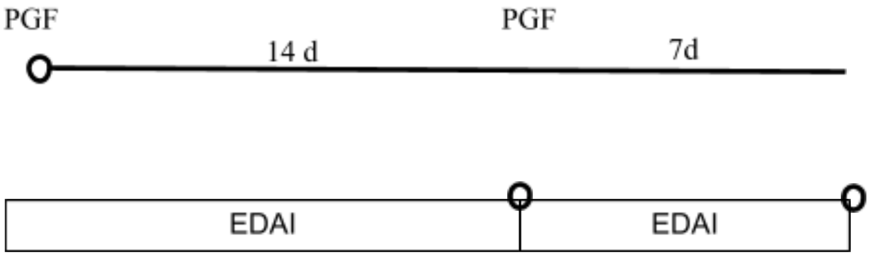
Where estrus start date is the day of breed start day for heifers, the stop day is the last day of the program length (allowed estrus period), and abortion day. And the estrus cycle distribution is method-specific. The stop day of the last estrus period is defined in the protocol; without detected estrus, the heifer will return to the next round of synch ED. For heifers who experienced abortion and identified in pregnancy checks, the synch restart at the pregnancy check day marked as abortion day.

3.3.3.2 synch ED programs

There are two Synch ED programs for a heifer. They are 2P and CP. Estrus happens on estrus day, with detection rate, whether this estrus is detected can be determined.

2P:

two injections of PGF, watch for estrus after each injection, and inseminate upon heat detection.

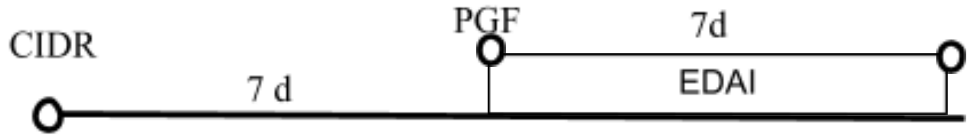


```
When days born = synch-ED program start day for heifer: inject PGF.
  Synch ED estrus day = synch ED estrus day with estrus cycle
      $Random \sim$ N (5,3), and program length = 14
When days born = Synch ED estrus day:
  Estrus count += 1
  if $RandomUni$ < estrus detection rate
    Estrus Detection = True,
    if RandomUni  < estrus service rate
      AI day = estrus day +1
      conception rate = estrus conception rate
  Else:
    inject PGF
    Synch ED estrus day = synch ED estrus day with
        estrus cycle $Random \sim$ N (3,2),
        and program length = 7
    If days born = Synch ED estrus day:
      Estrus count += 1
      if $RandomUni$ < estrus detection rate
        Estrus Detection = True
        if $RandomUni$ < estrus service rate
          AI day = estrus day +1
          conception rate = estrus
                  conception rate
      Else:
```

```
        synch ED stop day =  Synch ED
                  estrus day + 7
                                                    (A.1A.A3.23)
```

CP:

CIDR implements for 7 days, upon removal, inject PGF, then watch for estrus after injection with N (3,2), and inseminate upon heat detection



```
Days born = synch-ED program start day for heifer: implement CIDR
Days born = synch-ED program start day for heifer + 7: inject PGF
Synch ED estrus day = synch ED estrus day with estrus cycle
      $Random \sim$ N (3,2), and program length = 7
When days born = Synch estrus day:
  Estrus count += 1
  if $RandomUni$ < estrus detection rate
    Estrus Detection = True
    if $RandomUni$ < estrus service rate
      AI day = estrus day +1
      conception rate = estrus conception rate
  Else:
```

```
synch ED stop day = Synch ED estrus day + 7
```

(A.1A.A3.24)

3.3.3.3 synch ED update

Assign synch ED programs to the heifer

```
When days born = TAI program start day for heifer
  Call synch ED method for heifer
```

(A.1A.A3.25)

3.4 Heifer pregnancy update

On the day of AI, the success of conception is determined. There are 3 subsequent pregnancy checks at designated time points after AI, at which time the continued success of the pregnancy is determined.

3.4.1 Conception

```
When days born = AI day:
  If $RandomUni$ < conception rate:
    Preg = True
    Days in preg = 1
    Gestation length = $Random \sim$ N (avg, std) of
          heifer gestation length
  Else:
    Preg = False
```

(A.1A.A3.26)

3.4.2 Pregnancy checks

If the heifer's age (days born) reaches the AI date + preg *check day 1/ 2/ 3, compare a random U (0,1) draw with the pregnancy loss rate 1/ 2/ 3 set in the input to determine whether the pregnancy is lost. The reproductive program will resume on the pregnancy check day.*

```
if heifer pregnant = True:
  When days born = AI day + pregnancy check day 1:
    If RandomUni > preg loss rate 1:
      Days in preg = preg check day 1
    Else:
      Preg = False
      Abortion day = days born
                                                    (A.1A.A3.27)
If heifer pregnant = True:
  When days born = AI day + pregnancy check day 2:
    If RandomUni > preg loss rate 2:
      Days in preg = preg check day 2
    Else:
      Preg = False
      Abortion day = days born
                                                    (A.1A.A3.28)
If heifer pregnant = True:
  When days born = AI day + pregnancy check day 3:
    If RandomUni > preg loss rate 3:
      Days in preg = preg check day 3
    Else:
```

```
      Preg = False
      Abortion day = days born
                                                        (A.1A.A3.29)
  3.4.3 Heifer repro culling
  If the heifer does not have a successful conception or
pregnancy, and age (days born) exceeds heifer repro cull time,
set at 650 days, this heifer is culled
  If heifer pregnant = false and days born > heifer repro cull time:
    Cull stage = True
                                                        (A.1A.A3.30)
```

4. Heifers III

Implemented in `heiferIII.py`

The Heifer III class simulates heifers from close to calving until the start of milking.

No breeding is involved. Heifers continue to grow, and the ranking of replacements can happen here.

```
4.1 update heiferIII
  Heifers are updated on a daily basis
  4.1.1 bodyweight change
  The Heifer III class is increasing its body weight with ADG in addition
to weight gain for pregnancy.
  When bodyweight < mature body weight,
    $Bodyweight_i$ = $Bodyweight_{i-1}$ + pregnant heifer ADG
  When bodyweight = mature body weight,
```

```
  $Bodyweight_i$ = mature bodyweight
                                                        (A.1A.A3.31)
4.1.2 move to the next stage
Heifers are moved to the next stage when reach the end of gestation
When DIP = gestation length,
  Cow stage = True
                                                        (A.1A.A3.32)
```

5. Cow

Implemented in `cow.py`

Simulate cows when they start lactating until they leave the herd. With calving, lactating, breeding, and culling.

Input:

Cow repro method: the reproductive method used for the cows

Semen type: semen used in the inseminations

Days in pregnancy when dry: days of scheduled dry time for lactating cows

Lactation curve: lactation curve models for an estimate the targeted milk production

Repro cull time: days in pregnancy when marked as do not breed and terminated breeding practices

Cull milk production: milk production threshold for low production culling

Cow estrus cycle: the estrus cycle of cows in the ED program

Estrus detection rate: the chance of the estrus to be detected

Estrus service rate: the chance of a detected estrus to be serviced

ED conception rate: the conception rate of the estrus detection method

TAI method for cow: the user input of protocol picked for the TAI method

Ovsynch56 conception rate: the conception rate when applying Ovsynch56 protocol

Ovsynch48 conception rate: the conception rate when applying Ovsynch48 protocol

Cosynch72 conception rate: the conception rate when applying Cosynch72 protocol

Cosynch5d conception rate: the conception rate when applying Cosynch5d protocol

User define TAI length: the length of the user define TAI protocol

User define conception rate: the conception rate when apply User define TAI protocol

Cow resynch protocol: the resynch protocol picked for the TAI method

Voluntary waiting period: time after calving before starting breeding

Tai program start day: days in milk for the ED-TAI program to start the TAI program

Conception rate decrease: conception rate decrease by this much percentage points for each AI after the first

Gestation length distribution: the average and std length for the gestation period

Lactation curve parameters distribution: the average and std for lactation curve parameters

Pregnancy check day 1/ 2/ 3: days in pregnancy of doing pregnant checks

Pregnancy loss rate 1/ 2/ 3: pregnancy loss rate between pregnant checks

Thresholds of culling reasons:

Lameness	Injury	Mastitis	Disease	Udder	Unknown
0 - 16.33%	16.33% - 45.16%	45.16% - 69.55%	69.55% - 83.46%	83.46% - 89.91%	89.91% - 1

Thresholds of this culling time:

Days	0	5	15	45	90	135	180	225	270	330	380	430	280	530
Mastitis CDF	0	0.06	0.12	0.19	0.30	0.43	0.56	0.68	0.78	0.85	0.90	0.94	0.97	1
Lameness CDF	0	0.03	0.08	0.16	0.25	0.36	0.48	0.59	0.69	0.78	0.85	0.90	0.95	1
Injury CDF	0	0.08	0.18	0.28	0.38	0.47	0.56	0.64	0.71	0.78	0.85	0.90	0.95	1
Disease CDF	0	0.04	0.12	0.24	0.34	0.42	0.50	0.57	0.64	0.72	0.81	0.89	0.95	1
Udder CDF	0	0.12	0.24	0.33	0.41	0.48	0.55	0.62	0.68	0.76	0.82	0.89	0.95	1
Unknown CDF	0	0.05	0.11	0.18	0.27	0.37	0.45	0.54	0.62	0.70	0.77	0.84	0.92	1

```
5.1 Update status
  For cow's status to be updated on a daily basis.
  5.1.1 Calving
  When days in preg reaches the expected gestation length, calving happens, pregnancy terminates, lactation starts, and the estrus cycle restarts.
  When preg = True and days in preg = gestation length,
    Calves += 1
    Milking = True; Days in milk initialized as 0
    Preg = False; Days in preg = 0; Gestation length = 0
    Call restart estrus, if repro program involves ED
    Lactation curve parameters are selected based on breed and parity
        specific distributions ($Random \sim$ N (avg, std) of
        parameters l, m, n)
                                                      (A.1A.A3.33)
  5.1.2 Breeding protocols assign to cows
```

Breeding protocol for this lactation will be selected among ED, ED-TAI, or TAI.

```
If repro program == 'ED', call ed update
If repro program == 'ED-TAI', call ed-tai update
If repro program == 'TAI', call tai update
```

(A.1A.A3.34)

5.1.3 Bodyweight change

Cow BW change is composed of 3 sources: growth, conceptus, and lactation. The growth rate is represented by ADG until the end of the 2nd lactation. Conceptus growth for cows is modeled the same as for heifers. The BW change due to tissue mobilization is modeled with a parametric equation of DIM.

$$Bodyweight_i = Bodyweight_{i-1} + \text{average daily gain} + \text{conceptus growth} + \text{lactation body weight change} \quad \text{(A.1A.A3.35)}$$

5.2 Lactating

When the cow starts lactating, she produces milk from the calving day until dry or culled according to the assigned lactation curve model.

5.2.1 Milking update

Daily expected milk production updates every day from calving to dry, according to breed and parity specific parameters with wood's curve model, settled at the lactation (Three wood's model parameters: wood_l, wood_m, wood_n). For each breed (Holstein, Jersey), each parity group (1st, 2nd, and 3+). The targeted lactation curve is set at the beginning of each *lactation, and the variance can be added after adjusted by the nutrition submodel. The single animal accumulated milk production is calculated here.*

```
If milking = True:
  Estimated daily milk production = $lt^{m}e^{-nt}$
  Single accumulated milk production += estimated daily milk production
                                                              (A.1A.A3.36)
```

5.2.2 Dry

Cows are dry when they reach a dry off time before the end of gestation length. They stop milking and often change their pen and diet.

```
When days in preg = days in preg when dry,
  Milking = False
  Days in milk = 0
  Estimated daily milk production = 0
                                                              (A.1A.A3.37)
```

5.3 cow reproductive programs

When the cow passes the voluntary waiting period (VWP) and before AI. It is under breeding programs from one of TAI, ED, or ED-TAI (suggested by DCRC 2018) assigned to her. The voluntary waiting period is the period after pregnancy that the breeders choose not to breed the cow even if it enters estrus again.

5.3.1 Estrus detection method (ED)

It does not require the use of hormones but is based on watching for animal expression of heat. To represent this, the model simulates estrus *occurrence and the probability of detecting the estrus event.*

5.3.1.1 Estrus Day Determination

Determines the estrus day of each cycle for nature estrus after calving, detection failure, and AI failure.

```
Estrus day = estrus start date + $Random \sim$ N (avg, std) of
          cow estrus cycle
Where for the first estrus cycle occurs after calving:
  restart estrus function
  Estrus start day = days born when days in milk = 1
  estrus cycle follows estrus cycle return
For the first estrus cycle occurs after calving:
  return estrus function
  Start day = estrus day
  estrus cycle follows estrus cycle cow
For the first estrus cycle occurs after calving:
  after ai estrus function
  Start day = estrus day
  estrus cycle follows the estrus cycle cow
For the first estrus cycle occurs after calving:
  after abortion estrus fuction
  Start day = estrus day
  estrus cycle follows estrus cycle cow
                                                   (A.1A.A3.38)
```

5.3.1.2 Estrus detection update

Update estrus, estrus detection, service upon detection with conception rate specifically related to this method.

```
When days born = estrus day:
If days in milk >1 and days in milk < VWP:
  Call later estrus
Else: Estrus count += 1
  If $RandomUni$ < estrus detection rate:
    Estrus Detection = True,
    if $RandomUni$ < estrus service rate:
      AI day = estrus day +1
      conception rate = estrus conception rate
  Else, return estrus
```

(A.1A.A3.39)

5.3.2 Timed Artificial Insemination (TAI)

Uses a series of hormonal injections to synchronize the estrus cycle and perform AI to cows at the set time from the start of the protocol.

5.3.2.1 Program start day determination

Determines the program start day of each synchronized cycle TAI program after the voluntary waiting period and pregnancy checks.

```
  Tai program start day for cow = tai protocol start date
After passing the voluntary waiting period:
  Tai program start day for cow = call presynch function
For the first tai after preg check:
```

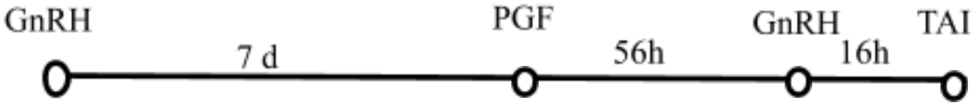
Tai program start day for cow = call resynch function

(A.1A.A3.40)

5.3.2.2 TAI programs

We have reproduction programs for cows from DCBC recommendations. For TAI programs, there are OvSynch56, OvSynch48, CoSynch72, 5dCoSynch, and user defined.

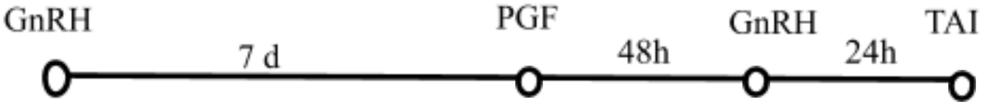
OvSynch56:



AI day = tai program start day for cows + 10

Conception rate = OvSynch56 conception rate

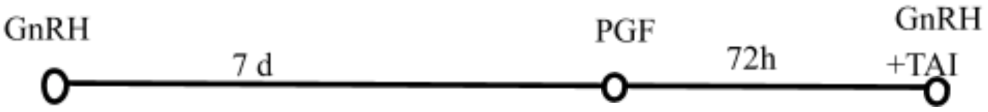
OvSynch48:



AI day = tai program start day for cows + 10

Conception rate = OvSynch48 conception rate

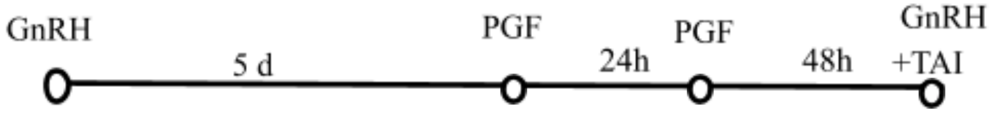
CoSynch72:



AI day = tai program start day for cows + 10

Conception rate = CoSynch72 conception rate

5dCoSynch:



AI day = tai program start day for cows + 8

Conception rate = 5dCoSynch conception rate

(A.1A.A3.41)

5.3.3 TAI presynch programs

Presynch protocols that could be connected with TAI programs to help with the conception rate of the TAI programs. The protocols are Presynch, Double Ovsynch, and G6G.

5.3.3.1 Program start day determination

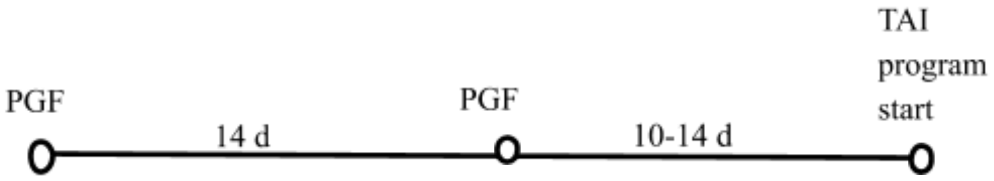
Presynch starts when the cow reaches the end of the VWP

```
If days in milk = vwp:
  Presynch start day = days born
```

(A.1A.A3.42)

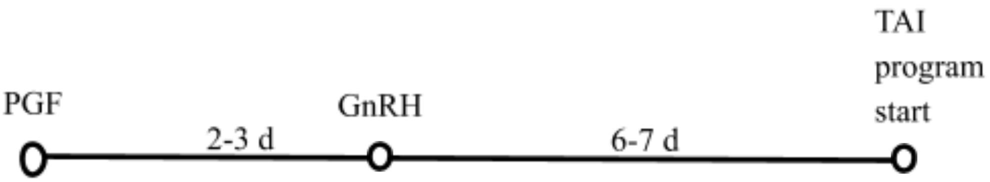
5.3.3.2 Presynch protocols

Presynch:



Double Ovsynch:

G6G:



(A.1A.A3.43)

5.3.4 TAI resynch programs

There are 3 ways of resynch after diagnosis open at preg check. They are called resynch, and they are TAI after Pregnancy Check, TAI before Pregnancy Check, and PGF at Pregnancy Check. The conception rate would decrease at the rate of reducing.

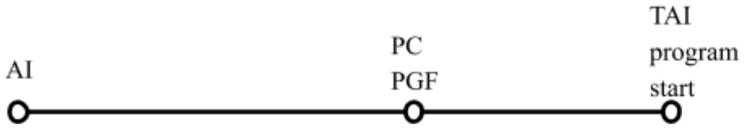
5.3.4.1 TAI after Pregnancy Check:

```
Tai start day for cow = abortion day +1
Conception rate -= conception rate decrease
```

5.3.4.2 TAI before Pregnancy Check:

```
Tai start day for cow = abortion day -6
Conception rate -= conception rate decrease
If tai method for cow is one of 'OvSynch 56', 'OvSynch 48',
            or 'CoSynch 72.':
  Event (inject PGF)
```

5.3.4.3 PGF at Pregnancy Check:



```
Event (inject PGF)
Tai start day for cow = abortion day +8
```

```
  Conception rate -= conception rate decrease
                                                        (A.1A.A3.44)
```

5.3.5 Estrous Detection - Timed Artificial Insemination (ED-TAI)

In the ED-TAI programs, ED methods are combined with TAI programs. ED happens during the resynch stage; the ED process will be happening before TAI starts, with natural estrus.

5.3.5.1 Estrus detection update

Update estrus, estrus detection, service upon detection with conception rate specifically related to this method. Estrus detection only happens when days in milk pass VWP and before tai start time.

Estrus happens on estrus day, with detection rate, whether this estrus is detected can be determined.

```
When days born = estrus day and days in milk < tai program start day
  Estrus count += 1
  If days in milk >1 and days in milk < VWP:
    Call later estrus
  Else:
    If RandomUni < estrus detection rate:
      Estrus Detection = True,
      if RandomUni < estrus service rate:
        AI day = estrus day +1
        conception rate = estrus conception rate
    Else:
```

```
    call return estrus function
```

(A.1A.A3.45)

If estrus is not detected during the detecting time, tai program starts at user-determined days in milk.

```
If days in milk = tai program start the day and ai day = 0:
  Estrus day = 0
  call determine tai program day function
  call tai method update according to user input
```

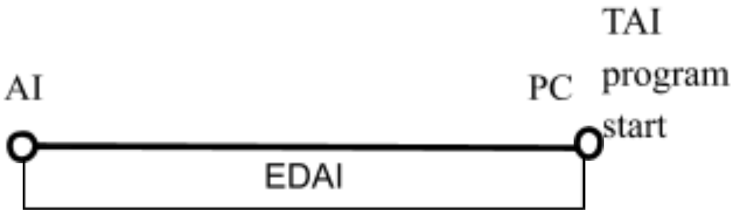
(A.1A.A3.46)

When TAI programs start, the process would be like in the tai method, included 4 methods recommended by CDCB (OvSynch56, OvSynch48, CoSynch72, and 5dCoSynch) and user-defined.

(A.1A.A3.47)

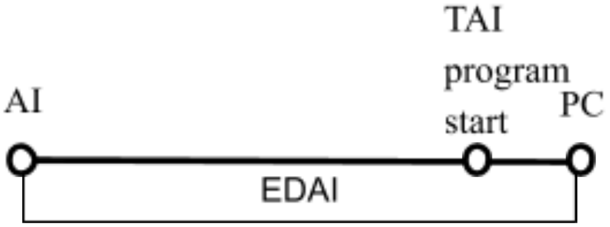
In the ED-TAI method, resynch process mix with estrus detection:

5.3.5.2 TAI after Pregnancy Check:



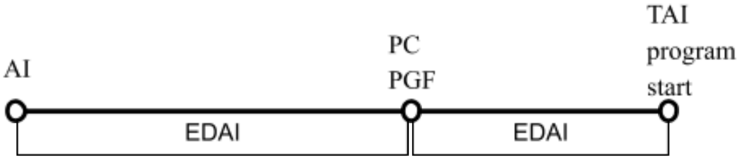
```
Tai program start day for cow = abortion day +1
Conception rate -= conception rate decrease
```

5.3.5.3 TAI before Pregnancy Check:



Tai program start day for cow = abortion day -6

Conception rate -= conception rate decrease



Tai program start day for cow = abortion day +8

Conception rate -= conception rate decrease

(A.1A.A3.48)

5.3.6 adjust conception rate

Adjust conception rate based on the parity of the cow

If lactation = 1, conception rate = conception rate,

If lactation = 2, conception rate = conception rate - 0.05,

If lactation > 2, conception rate = conception rate - 0.1.

(A.1A.A3.49)

5.3.7 open cow update

Assign breeding method for open cows after spotted open at pregnancy checks

```
If repro program = ED: call after abortion estrus,
If repro program = TAI: call tai program after preg check,
If repro program = ED: call resynch ed tai.
```

(A.1A.A3.50)

5.4 Cow pregnancy update

On the day of AI, the success of conception is determined. There are 3 subsequent pregnancy checks at designated time points after AI, at which time the continued success of the pregnancy is determined.

5.4.1 Conception

```
When days born = AI day:
  If $RandomUni$ < conception rate:
    Preg = True
    Days in preg = 1
    Gestation length = $Random \sim$ N (avg, std) of cow
              gestation length
  Else:
    Preg = False
```

(A.1A.A3.51)

5.4.2 Pregnancy checks

If the cow's age (days born) reaches the AI date + preg check day 1/2/3, *compare a random U (0,1) draw with the pregnancy loss rate 1/2/3 set in the input to determine whether pregnancy is lost. The reproductive program will resume on the pregnancy check day.*

```
If cow pregnant = True:
  When days born = AI day + pregnancy check day 1:
    If RandomUni > preg loss rate 1:
      Days in preg = preg check day 1
    Else:
      Preg = False
      Abortion day = days born
      Call open cow update
                                                        (A.1A.A3.52)
If cow pregnant = True
    When days born = AI day + pregnancy check day 2:
    If RandomUni > preg loss rate 2:
      Days in preg = preg check day 2
    Else:
      Preg = False
      Abortion day = days born
      Call open cow update
                                                        (A.1A.A3.53)
If cow pregnant = True
    When days born = AI day + pregnancy check day 3:
    If RandomUni > preg loss rate 3:
```

```
    Days in preg = preg check day 3
  Else:
    Preg = False
    Abortion day = days born
    open cow update
```

(A.1A.A3.54)

5.5 Culling

5.5.1 Cull update

Record culling time and reason:

```
If days born = future cull date:
  Cull = True
```

(A.1A.A3.55)

5.5.2 Low production Culling

If the cow is unable to be pregnant (marked as do not breed), and her milk production drops lower than 'cull milk production' from the input, it will be culled for low production.

5.5.2.1 do not breed

The cow is marked as do not breed when not pregnant for a period determined by the user input

```
If preg = False and days in milk > do not breed time:
  Do not breed = True
```

(A.1A.A3.56)

5.5.2.2 cull for low production

Do not breed cows are culled when their milk production is below the threshold

```
If do not breed and days in milk > 80, and estimated daily
     milk produced <  cull milk production:
  Future cull date = day born
  Cull reason = 'low production'
```

(A.1A.A3.57)

5.5.3 Health issue Culling

The reason and date of culling a cow for health reasons other than reproduction and production. It has a different culling rate for each parity and six reasons for health culling with a different probability. First, it will determine whether this cow will be culled for health reasons in this parity. If it is culled in this parity, then the culling reason for this culled is determined by a random draw. If the reason is specified, then an empirical CDF is used to determine which day it will be culled in this parity.

5.5.3.1 determine whether culled because of health issue in this parity based on the parity cull probability from the input

```
cull rate for this parity = parity cull probability
If RandomUni < cull rate for this parity
  Cull reason random = RandomUni
  If cull reason random is between thresholds of two
          health culling reasons:
    Cull reason = health culling reason
```

Where Cull reason = lameness/ Injury/ Mastitis/ Disease/ Udder/ Unknow

(A.1A.A3.58)

```
Cull time random = $RandomUni$
If cull time random is between thresholds of two culling
      probabilities thresholds of this culling reason:
  Future cull date = $x_{i-1} + a_i * (R - c_{i-1})$
Where put date random to the CDF function of the reason
get $c_{i-1}, c_i$ (the upper and lower number around R), $x_{i-1}, x_i$ (days,
correspond to $c_{i-1}, c_i$),
compute $a_i$. Empirical CDF: $a_i = \frac{x_i - x_{i-1}}{c_i - c_{i-1}}$
Compute the cull day of this cow $X = x_{i-1} + a_i * (R - c_{i-1})$
```

(A.1A.A3.59)

5.6 Cow body weight change

5.6.1 average daily gain

Cows are targeted to grow to their mature body weight until the end of the 2nd lactation. A calving interval is in either from the user input or the herd average. The ADG of cows is calculated based on the target of reaching 0.92 of the mature body weight by the end of the 1st lactation, and full mature body weight at the end of the 2nd lactation.

$$cow\,ADG(kg/d) = \begin{cases} \frac{(0.92-0.82)*MBW}{average\ calving\ interval} & \text{if } parity = 1 \text{ and non-pregnant} \\ \frac{0.92*MBW-BW}{gestation\ length-DIP} & \text{if } parity = 1 \text{ and pregnant} \\ \frac{(1-0.92)*MBW}{average\ calving\ interval} & \text{if } parity = 2 \text{ and non-pregnant} \\ \frac{MBW-BW}{gestation\ length-DIP} & \text{if } parity = 2 \text{ and pregnant} \\ 0 & \text{else} \end{cases} \tag{A.1}$$

Where MBW is mature bodyweight (0.96*mature bodyweight)

(A.1A.A3.60)

5.6.2 conceptus growth

The conceptus weight change along gestation, the same as for heifers

When DIP < 50,

conceptus growth = 0;

When 50 < DIP < gestation length,

conceptus growth = $3*conceptus\ \ parameter^3*(DIP-50)^2$

When DIP = gestation length,

conceptus growth = - total conceptus weight

Where $total\ \ conceptus\ \ weight\ = \ 0.0148\ *\ gestation\ \ length\ -\ 2.408\ *\ calf\ \ birth\ \ weight$

$conceptus\ \ parameter\ = \ total\ \ conceptus\ \ weight^{\frac{1}{3}}\ /\ gestation\ \ length\ -\ 50$

(A.1A.A3.61)

5.6.3 Lactation body weight change

Lactation related body weight changes are counting the tissue change due to lactation. It is only considered for lactating cows and is assumed to be 0 for dry cows.

$$Lactation\ BW\ change(kg/d) = \begin{cases} -\frac{P_1}{P_2} * exp(1 - \frac{DIM}{P_2}) + \frac{P_1}{P_2^2} * DIM * exp(1 - \frac{DIM}{P_2}) & \text{Milking cow} \\ \frac{P_1 * \frac{DIM}{P_2} * exp(1 - \frac{DIM\ when\ dry}{P_2})}{gestation\ length - DIP\ when\ dry} & \text{Dry cow} \end{cases} \tag{A.2}$$

`Where` P_1 `is the the maximum decrease of BW during lactation,` P_2 `is the time during the lactation with minimum BW.`

`(A.1A.A3.62)`

A.4 Animal Initialization

Implemented in `animal_initilization.py`

A database of animals in different statuses and replacement market is generated as an initialization database in this file. The initialization database simulation is simulated similarly to the regular life cycle simulation process, but with an extended period and with a large number of animals.

The core status information and its life history of each animal are stored in the database at the end of the simulation process.

```
1.  Initialize animals
For days in range of simulating days
  For calf in calves: calf update
  If wean = true:
    remove(calf)
    append(heiferI)
  For heiferI in heiferIs:
    If second stage = true:
      remove(heiferI)
```

```
      append(heiferII)
      Update heifer reproductive program
  For heiferII in heiferIIs:
    If third stage = true,
      remove(heiferII)
      append(heiferIII)
  For heiferIII in heiferIIIs:
    If cow stage = true,
      remove(heiferIII)
      append(cow)
      Update cow reproductive program
  If simulation day > replacement not recording length
    append(replacement)
  For cow in cows:
    If culled = true,
      remove(cow)
    If new born:
      append(calf)
                                                            (A.1A.A4.1)
```

```
2. Write information to the database
For the last day of simulation:
  Write animal information of each group to sqlite file
                                                            (A.1A.A4.2)
```

A.5 Life Cycle

Implemented in `life_cycle.py`

Simulate the herd with an initial number of each animal group. Drive animal daily updates.

Inputs:

calf num: number of animals in the calf when the initial herd

heiferI num: number of animals in heiferI when the initial herd

heiferII num: number of animals in heiferII when the initial herd

heiferIII num: number of animals in heiferIII when the initial herd

Cow num: number of animals in cow when the initial herd

Replace num: number of animals in the replacement market when the initial herd

Herd num: number of animals in the cow for the simulation

Herd init: whether to have herd initialization running

breed: the breed of the simulating herd

`1. Initialize herd`

Initiate herd with herd number, calf number, heifers number, cow number, replacement number, and simulation length in days in this function as the starting stage of the simulation.

`Animals in each group are initial by random drawings from the initial herd to meet the number of animals needed for the initial herd.`

`(A.1A.A5.1)`

`2. Daily update`

This function updates the herd and individual animals with a daily step. Each animal group is updated sequentially, starting with calves *and ending with cows. If an animal reaches the point in their life-cycle when they are ready to move to a different animal class, they are added to the next animal class and removed from their previous animal class. For example, when a calf reaches weaning, that instance of the calf class is removed, and a new instance of the HeiferI class is initiated.*

Reproduction methods are assigned to heifers and cows when they move to the heifer II and cow group.

If the number of the heifers is more than the herd needs, the heifers in the heifer III group would be sold. While if the number of heifers III is less than the herd needs, a replacement would be bought from the market.

(A.1A.A5.2)