UNIVERSITÀ DEGLI STUDI DI MILANO

CORSO DI DOTTORATO INDUSTRIALE IN INTERSECTORAL INNOVATION – DOTTORATO INTERSETTORIALE PER L'INNOVAZIONE

XXXVIII ciclo

Area Scienze della vita

Dipartimento di Medicina Veterinaria e Scienze Animali (DIVAS)

INVESTIGATION OF RESILIENCE AND ITS GENETICS THROUGH THE USE OF LONGITUDINAL DATA

(AGRI-09/A: Livestock Systems, Animal Breeding and Genetics)

Dottorando
Andrea Delledonne
Matr. 14033
ORCID n. 0000-0003-3034-3517

Tutor
Prof. Alessandro Bagnato

Co-tutor
Prof. Maria Giuseppina Strillacci

Coordinatore del corso di dottorato
Prof. Ernesto Damiani

A.A.
2024/2025

Making art is difficult. We leave drawings unfinished and stories unwritten. We do work that does not feel like our own. We repeat ourselves. We stop before we have mastered our materials, or continue on long after their potential is exhausted. Often the work we have not done seems more real in our minds than the pieces we have completed. And so questions arise: How does art get done? Why, often, does it not get done? And what is the nature of the difficulties that stop so many who start?

— *Art & Fear, David Bayles & Ted Orland*

Acknowledgments

My sincere gratitude goes to:

Tania, for your love and patience, tears and smiles, for crossing the ocean to reach out to me for a wonderful vacation and for your gifts. Special thanks! ♡

Prof. Alessandro Bagnato and Prof. Maria Giuseppina Strillacci for the opportunity to begin and complete the doctoral program under your guidance.

Prof. Christian Maltecca and Prof. Francesco Tiezzi, for your invaluable support during my time abroad, expertise, leadership, encouragement, patience, cheers and genuine pizza.

Prof. Cristina Sartori and Prof. Luiz F. Brito, acting as external reviewers, for your time, efforts, and constructive comments that have improved these studies and the manuscript.

Carlotta Ferrari, Chiara Punturiero, Francesca Bernini, and Raffaella Milanesi, for sharing this journey with me, for your help, encouragement, and smiles.

Aakilah Hernandez, Chang Xu, Chiara Gini, Hui Wen, Junjian Wang and Simone Callegaro for your welcome and nice lab-time.

Chiara Forti for your fundamental support during the initial phase of my Ph.D.

Chiara Gini, for your support, friendship, and the wonderful memories in Raleigh.

Nicolò Franguelli, for your great friendship, the wonderful times we have shared together and for your hospitality. In love with Aarhus!

Dr. Cristina Ceruti, for your assistance with administrative matters and your enthusiasm and kindness.

Anna B., Anna M., Antonio, Chiara, Dario, Enrico, Giovanni, Ilaria, Ivan L., Ivan R. D., Jasmine, Lorenzo, Luana, Zia and all the people I met in Raleigh for making my time abroad unforgettable.

Chiara Gini and Lorenzo Micalizzi, for the gym brotherhood and the great time in Raleigh.

Ilaria Napolitano: for your enjoyable lifestyle. Best housemate ever!

Lorenzo Micalizzi for your friendship, emotional support, technical assistance, financial help, smiles, candy, and for always being there when I needed you the most. ■

Summary

This thesis investigates resilience in dairy cattle using an integrated approach that combines phenotypic characterization, simulation of perturbations, and genomic analysis. The research was motivated by the need to breed animals capable of maintaining productivity amid the environmental and economic challenges facing modern agriculture. Thanks to close collaboration with a medium-large dairy farm in Northern Italy, we had access to a comprehensive dataset of daily milk records and genotype information, which formed the basis for the analysis.

Over the past twenty years, resilience assessment in livestock has become feasible thanks to new technologies that enable the collection of high-frequency longitudinal data. Resilience indicators are recently defined proxies that combine routinely collected data, such as milk yield records, to capture an animal's response to environmental challenges. Although it has been demonstrated that resilience indicators possess some genetic merit to some extent, their true significance and methodological foundations require further investigation to enhance the understanding of these phenotypes and their application in breeding programs. Lactation curve models aiming to represent the unperturbed state of each individual cow are essential for deriving meaningful resilience indicators; however, their impact on these indicators has not been systematically evaluated.

This thesis addresses these gaps through three main studies: (1) characterization of resilience indicators derived from various lactation curve models; (2) a simulation study evaluating the ability of resilience indicators to capture true genetic resilience; (3) a genome-wide association study (GWAS) identifying genomic regions associated with resilience indicators. Additionally, one further contribution is included: (4) construction of a copy number variation map for the population studied.

The characterization study demonstrated that methodological choices in lactation curve modelling significantly influence phenotypic resilience indicator values and cow rankings. The simulation study revealed that current resilience indicators capture only a fraction of true genetic resilience, with the best model-indicator combinations achieving correlations of 0.15 to 0.25 with simulated resilience breeding values. The genome-wide association study identified genomic regions associated with resilience indicators, confirming the polygenic nature of the trait and highlighting genes involved in immune response, energy metabolism, and tissue integrity. Additionally, the copy number variation map provided a valuable genomic resource for future integrated analyses.

These findings advance our understanding of resilience as a complex, multifaceted trait while providing practical tools for its genetic improvement. The practical implications are

clear: breeding programs should incorporate resilience indicators derived from meaningfully selected lactation curve models, set realistic expectations for genetic progress given the modest correlations with true resilience, and balance selection for resilience alongside production and other traits through appropriately weighted selection indices.

Looking ahead, continued refinement of resilience phenotypes, functional characterization of candidate genes and economic analyses will advance breeding for resilience.

Contents

List of Figures

List of Tables

Introduction

In October 2022, I was assigned to the Ph.D. program in Intersectoral Innovation at the University of Milan. This new program aims to bridge the University with the business world. I was involved in the AGRITECH project, which was funded by the PNRR (Piano Nazionale di Ripresa e Resilienza - National Recovery and Resilience Plan).

Resilience in dairy cattle was the focus of the Task 5.2.2, "Testing of Resistant and Resilient Animals," which aimed to identify resilience phenotypes in cows raised on commercial farms and link them to genetic variation. Since then, resilience has been the word in my mind every day for three years, a chimera to pursue and a nightmare to deal with.[1]

My Ph.D. course started on the 1st of November 2022 and finished on the 31st of October 2025. During this period, I completed the following activities related to the learning and research objectives of the Ph.D. program:

Learning courses provided by the Ph.D. programs of the University of Milan:

- Statistical methods applied to epidemiology and animal sciences - 2 CFU,
- Genomics for ecological and evolutionary studies: from DNA sequencing to data analysis - 3 CFU,
- Cell and tissue culture: from basic principles to advanced technologies - 3 CFU,
- Digital imaging and image integrity in scientific publication - 3 CFU,
- Benessere animale nell'ottica della sostenibilità (translated: Animal welfare from a sustainability perspective) - 3 CFU.

Learning transferable skills courses provided by the University of Milan: these activities included courses on project management, intellectual property in research, presentation skills and scientific writing. In total, I completed 72 hours of transferable skills courses.

Learning courses provided by external academics and entities:

---

[1] The original title of this manuscript was "Niente panico". This means two things: first, the author needed to read it every time he opened the manuscript to get relief from the ironic view; second, the author likes "The Hitchhiker's Guide to the Galaxy" (Adams 1979). Over the course of three complex years, the author learned how important the sense of humor is for dealing with hard things.

- Quantitative genetics and breeding, GN713 - 3 credit hours - North Carolina State University (USA) - instructor: Prof. Christian Maltecca - fall semester 2024/2025;
- Quantitative genetics theory and methods, GN757 - 3 credit hours - North Carolina State University (USA) - instructor: Prof. Zeng Zhao-Bang - fall semester 2024/2025;
- Mixed models - 3 ECTS - Graduate School of Natural Sciences, Aarhus University, instructor: Dr. Rodrigo Labouriau - November 23-December 2, 2023;
- Introduction to the analysis of longitudinal data with R - 1 ECTS - Physalia courses, instructor: Dr. Filippo Biscarini - June 26-29, 2023.

Period abroad: visiting student at North Carolina State University (USA), Department of Animal Science, under the supervision of Prof. Christian Maltecca - August 13, 2024-July 15, 2025.

Congress and conferences participations:

- oral presentation at American Dairy Science Association (ADSA) Annual Meeting 2025 - June 22-25, 2025, Louisville, KY, USA;
- poster presentation at Association for Science and Animal Production 26th congress - June 17-20, 2025, Turin, Italy;
- oral presentation at Associazione Italiana Società Scientifiche Agrarie (AISSA) congress - June 26-27, 2024, Florence, Italy.

The following publications are listed, even those not strictly related to my Ph.D. program:

- Punturiero C, Delledonne A, Ferrari C, Bagnato A and Strillacci MG (2025) Mapping genomic regions affecting resilience traits in a large dairy farm of Holstein cows. Front. Anim. Sci. 6:1627086. `https://doi.org/10.3389/fanim.2025.1627086`.
- Ferrari, C., Punturiero, C., Milanesi, R. et al. Exploring the genetic variability of the PRNP gene at codons 127, 142, 146, 154, 211, 222, and 240 in goats farmed in the Lombardy Region, Italy. Vet Res 55, 99 (2024). `https://doi.org/10.1186/s13567-024-01353-3`.
- Delledonne A, Punturiero C, Ferrari C, Bernini F, Milanesi R, Bagnato A, et al. (2024) Copy number variant scan in more than four thousand Holstein cows bred in Lombardy, Italy. PLoS ONE 19(5): e0303044. `https://doi.org/10.1371/journal.pone.0303044`.
- Bernini, F., Punturiero, C., Vevey, M., Blanchet, V., Milanesi, R., Delledonne, A.,

Bagnato A. & Strillacci, M. G. (2023). Assessing major genes allele frequencies and the genetic diversity of the native Aosta cattle female population. Italian Journal of Animal Science, 22(1), 1008–1022. `https://doi.org/10.1080/1828051X.2023.2259221`.

- Punturiero, C., Milanesi, R., Bernini, F., Delledonne, A., Bagnato, A., & Strillacci, M. G. (2023). Genomic approach to manage genetic variability in dairy farms. Italian Journal of Animal Science, 22(1), 769–783. `https://doi.org/10.1080/1828051X.2023.2243977`.

This thesis contains the research developed during my Ph.D. program. The main goal of the research was to investigate possible phenotypes to objectively measure resilience in dairy cattle and its genetic background. This was possible thanks to the collaboration with a mid-large commercial farm of Holstein cows, which allowed us to continuously update a dataset of milk records and to genotype the entire herd. The farm is located in Northern Italy, in the Po Valley, which is one of the most important agricultural areas in Europe and the most important one in Italy: 35% of all Italian crops and 55% of the Italian livestock production are produced in the Po Valley (Monteleone et al. 2024). The family-run farm is characterized by a high level of technology and innovation, with a strong focus on animal welfare and environmental sustainability, while maintaining high productivity levels. The context outlined above should clarify that the farm is a good example of a modern dairy farm, where the use of technology and innovation is combined with a strong family tradition.

Chapter 1

Literature review

1.1 Resilience

1.1.1 Defining Resilience

The word “resilient” comes from the Latin term *resilire*, meaning “to jump back” or “to rebound” (Online Etymology Dictionary 2025).
The seminal paper of Holling (1973) highlighted the importance of resilience in ecological systems. The author defined two terms that are crucial and can be helpful to understanding: resilience and stability.

Resilience as the:

> “measure of the ability of these systems to absorb changes of state variables, driving variables, and parameters, and still persist.”

Stability as the:

> “ability of a system to return to an equilibrium state after a temporary disturbance. The more rapidly it returns, and with the least fluctuation, the more stable it is.”

In other words, stability minimizes fluctuations mantaining a consistent and unchanging state or returning to it quickly after a minor disturbance, while resilience allows withstanding, adapating and recovering from larger disturbances, even if it means changing the state of the system. The first concept can be regarded as static, the second one as dynamic.Interesting examples were reported by Holling (1973) to illustrate the difference between these two concepts, and I am reporting them here. In a spruce budworm-forest system, spruce populations fluctuate extremely widely between outbreaks, with budworms extremely rare between outbreaks but massively destructive during outbreaks. The result is a system highly unstable (huge fluctuations) but resilient (these systems persist over centuries through these cycles) (Morris 1963). In contrast, Great Lakes fish populations are homogenous systems realtively constant before and after distrubances, but they are easily driven to extinction by fidhing pressure and other changes. This system is stable (little fluctuations) but not resilient (it can be driven to extinction).

It is sufficient to present these two concepts to illustrate that resilience may be often confused with stability or other related concepts such as robustness, resistance, adaptability, and flexibility. For example, Scheffer et al. (2018) reported that definitions of resilience vary across disciplines but consistently relate to the ability of a system to maintain specific functions in the face of change. In the same paper, systemic resilience in animals and humans is defined as the capacity to bounce back to normal functioning after a perturbation.
The terminology surrounding resilience is a complex issue and has likely contributed to confusion regarding this concept. Due to its multifaceted applicability, numerous terms are applicable and there is significant overlap in their meanings. Regarding the specific context of this thesis we can refer to what reported by Berghof et al. (2019b) in their Box 1 that contains a list of definitions for resilience and associated terms.

1.1.2 Resilience across disciplines

The field of resilience encompasses diverse disciplines, including economics, politics, psychology, engineering, environmental sciences, and agriculture. The following terms, related to the broad concept of resilience, are drawn from the resilience dictionary on the Stockholm Resilience Centre (SRC) website (*Resilience dictionary* 2025): social-ecological systems, ecosystem resilience, social resilience, ecosystem services, vulnerability, institutions, complex adaptive systems, response diversity, Anthropocene, natural capital, social capital, regime shifts, adaptive co-management, and blue and green water. This broad spectrum of terms reflects the interdisciplinary nature of resilience studies.

Of practical interest, applications of resilience can be found in the farming sector. It is a complex system where facing environmental and economic challenges can benefit from an understanding of stability, resilience, and their improvement. Scheffer et al. (2018) posited that, from another perspective, the components of a farming system can be regarded as systems in their own right, including humans and animals. Studying resilience should ideally consider all these components, as partially demonstrated in the following sections.

1.2 Why Resilience matters

Resilience, as regarding a wide spectrum of disciplines, is becoming increasingly important in the context of global health, due to its relevance in addressing environmental, social, and economical challenges.

1.2.1 Planetary Boundaries

In 2009, two seminal papers were published (Rockström et al. 2009a; Rockström et al. 2009b) laying the foundation for the concept of planetary boundaries and promoting the idea that researchers, stakeholders across various fields, and the public at large recognize that the planet's resources are being overexploited (Figure 1.1). Since then, a progression of evolutionary developments has been quantified in scenarios outlined in Figure 1.1. Recently, a significant milestone was reached with the quantification of nine planetary boundaries that collectively contribute to maintaining a stable and resilient Earth system (Richardson et al. 2023). However, the authors themselves acknowledged that six boundaries had already been crossed. Furthermore, more recently (September 2025), the Potsdam Institute for Climate Impact Research revealed that a seventh planetary boundary – related to ocean acidification – has also been surpassed (The Lancet Planetary Health 2025).

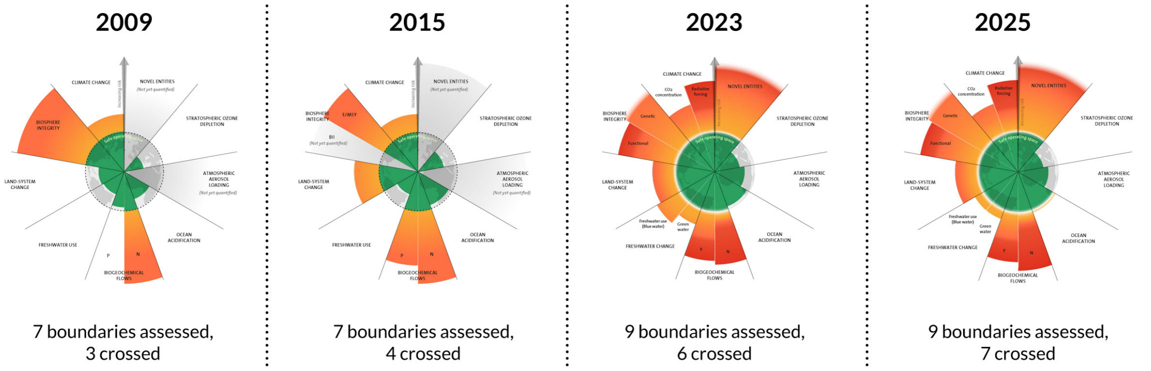


Figure 1.1: The evolution of the planetary boundaries framework. Licensed under CC BY-NC-ND 3.0, downloaded at `https://www.stockholmresilience.org/research/planetary-boundaries.html` on August 28, 2025. If you are interested, while I am writing, the planetary boundaries are continuously monitored and reported at `https://www.planetaryhealthcheck.org/`. Planetary boundaries are science-based limits designed to keep the Earth system within a safe operating space for human development, avoiding crossing critical thresholds that could trigger abrupt or irreversible environmental change.

As a consequence that is included in a looping effect, the crossing of planetary boundaries can lead to environmental changes that challenge our ability to produce food.

1.2.2 Livestock sector and environmental challenges

It is imperative to acknowledge the implications of the livestock sector on the push of planetary boundaries, which include, for those possibly related to resilience, the freshwater resource cycles, biogeochemical cycles of nitrogen and phosphorus, land use (Leip et al. 2015; Bowles et al. 2019; Li et al. 2024). An evaluation of impact of the livestock sector has been approached by FAO proposing a Global Livestock Environmental Assessment Model (GLAM - `https://www.fao.org/gleam/en/`). Extensive literature has been developed for GHG emission and evaluation of CO2 equivalent has been proposed at global dimension (`https://ourworldindata.org/meat-production/`, Nugrahaeningtyas et al. (2024)).

Water resources

Based on Leng et al. (2021), global human freshwater use totals 2,508 km^3/year, remaining below the planetary boundary of 4,000 km^3/year. Livestock water consumption accounts for 270 km^3/year (approximately 11% of total use), growing from 145 km^3/year in 1971. Here, feed crop irrigation represents 90% of livestock water use, with drinking water comprising 10%. While aggregate use stays within boundaries, livestock farming significantly impacts regional water stress, solely causing boundary exceedance in 7% of affected rivers and contributing to 34% of boundary violations. This highlights the critical role of livestock in local water sustainability challenges, a problem that can affect the livestock sector itself (Leng et al. 2021).

A recent illustration of freshwater consumption can be observed in the warning issued by researchers regarding the water usage involved in the production of soybeans and beef in Brazil (Lathuillière et al. 2025). As the authors highlighted, water scarcity problems in Brazil arise primarily in the São Francisco, Parnaíba, Atlântico Sul, and Uruguay river basins, particularly in northeastern regions. The main challenges stem from the competing demands for water resources between the agricultural sector and downstream needs, the necessity of intensive agricultural practices necessitating irrigation, unpermitted construction of reservoirs, and substantial export demands of soybean and beef (to China: 300-461 km^3/year; to EU: 75-88 km^3/year).

Drought and climate extremes

Similarly, droughts are becoming more frequent and intense in many regions worldwide, including Italy. Monteleone et al. (2024) analyzed drought conditions in Italy's Po Valley, identifying severe events in 2003-2004, 2017, and 2022-2023, with the latter being the most severe in 200 years. The primary impacts of these droughts included ecological disruption (36%), pollution (28%), and agricultural losses (16%). Current management strategies focused on agricultural water optimization and forecasting systems. However, there are still significant research gaps concerning the effects on human health and tourism, as well as the development of comprehensive drought policies in this economically critical region (Monteleone et al. 2024).

Farming resilience is crucial as climate change increases the frequency and severity of extreme weather events. Sauer et al. (2024) demonstrated this importance through an analysis of the 2003 Australian drought, which significantly impacted crop productivity, income, and profits. Their framework revealed that resilience requires balancing four capacities: preparedness, absorption, adaptation, and transformation. Most farms performed well in only two of these capacities, highlighting trade-offs among different dimensions of resilience. Key drivers included productivity, technological change, and farmer age, with the drought ultimately prompting structural transformation toward more productive farming operations, underscoring the critical role of resilience in agricultural sustainability (Sauer et al. 2024).

The instances reported above should not be viewed as isolated occurrences; rather, they reflect a global trend that is expected to intensify in the future. The livestock sector is especially vulnerable to these changes, as it depends heavily on natural resources such as water and land use. Enhancing the resilience of both the livestock sector and the animals themselves can help ensure the sustainability of food production in the face of these challenges.

1.2.3 Global food demand and nutrition

Global milk production (81% cow milk, 15% buffalo milk, 4% goat, sheep and camel combined) is projected to increase steadily at 1.8% per annum over the next decade, reaching 146 Mt by 2034 and fueled by larger goats and sheep herds in Africa, by more producing milking cows and buffaloes in Asia, and increasing yields per animal in well developed countries (while here grow in the number of cow is expected to be moderated) (OECD et al. 2025). As milk yield, also the global meat production is expected to increase: projections show it is expected to rise 13% (46 Mt) to 406 Mt by 2034, with increases shared as follows: poultry 21%, sheep 16%, beef 13% and pig 5%; poultry will dominate the expansion of meat production by accounting for the 62% of the additional meet produced, with Asia being the dominant country for growth in both cases (OECD et al. 2025).

Dairy products are notable for their high vitamin content, which includes 13 different vitamins. They are also relatively inexpensive and easy to acquire, depending on the geographical area, level of development, and cultural context. In Western societies, dairy products are a significant source of retinol, calciferol, riboflavin, pantothenic acid, folic acid, and cobalamin; in developing countries, they can play a crucial role in vitamin intake when dietary diversity is limited (Graulet 2014).

While animal-derived foods remain essential to sustain the growing global population, the livestock sector (and its current and prospective structure) is increasingly challenged by societal demands for novel food sources that offer more sustainable and ethical alternatives to conventional animal production. Beckett et al. (2024) analyzed nutritional differences between dairy milk and plant-based alternatives, concluding they are alternatives but no equivalents. Only soy milk provides comparable complete protein; others contain incomplete proteins and lower bioavailability of fortified nutrients like calcium. Dairy milk provides critical nutrients (49% global calcium, 24% vitamin B2) and bioactive

compounds with health benefits, while plant-based alternatives are ultra-processed with 80-90% nutrient loss during processing (Beckett et al. 2024).
Despite environmental claims, nutrient-density comparisons reveal similar CO2 contributions, challenging replacement assumptions (Beckett et al. 2024).

The above evidences remark the importance of animal products in human nutrition, but they also highlight the challenges that the livestock sector is facing in order to meet the increasing demand for food while minimizing its environmental impact. In this context, improving the resilience of livestock can contribute to achieving these goals.

1.2.4 Consumption patterns and sustainability

Global crop demand is projected to increase substantially by 100% to 110% from 2005 to 2050, encompassing direct human food consumption, livestock and fish feed, and post-harvest losses. This growth is driven by rising per capita GDP, which shifts dietary preferences toward greater meat consumption. The low efficiency with which livestock convert crop calories and protein into edible food amplifies this demand for agricultural commodities (Tilman et al. 2011).

Confronting the escalating challenges posed by resources restrictions and climate changes, while it seems absolutely necessary that we need more food, we necessitate even a reevaluation of our consumption patterns, particularly with regard to natural resources utilization and meat production. In developing countries, the practice of raising livestock is a means of improving the nutritional intake of individuals who are malnourished (Parlasca et al. 2022). Conversely, in developed countries, there is a pervasive consumption of meat and dairy products that is detrimental to human and environmental health (Parlasca et al. 2022). Notable are the meat and dairy solids consumption's quantities reported in Figures 1.2 and 1.3 across countries grouped by their income level (OECD et al. 2025). For 2034, values were projected by the mean of the OECD-FAO Aglink-Cosimo model.

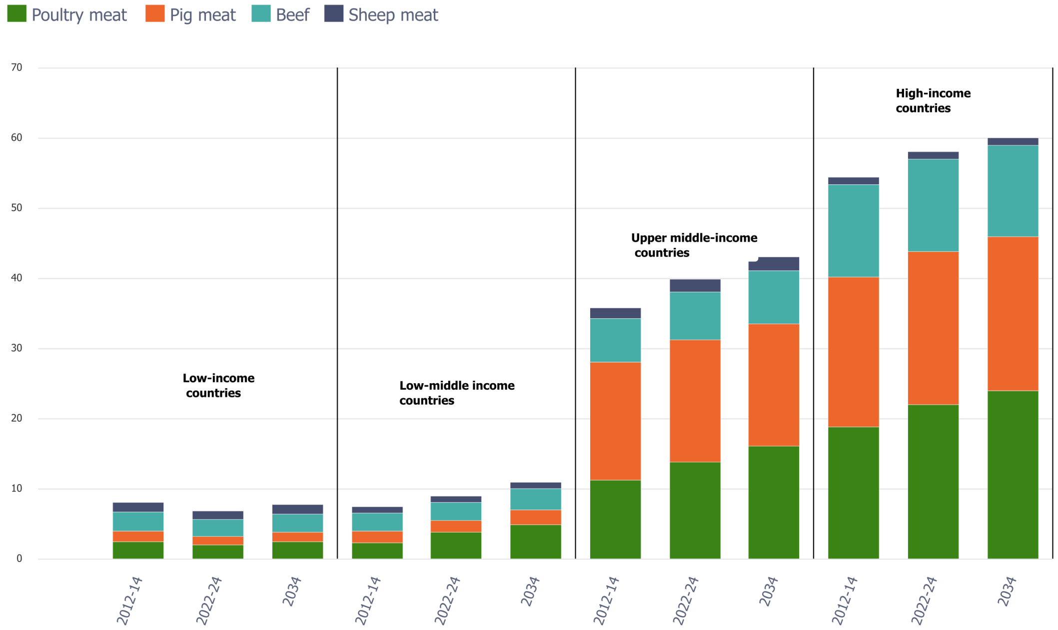


Figure 1.2: *Per capita* meat consumption by income group and meat type (Kilograms/person/year). This is an adaptation of an original work by the OECD and FAO. The opinions expressed and arguments employed in this adaptation should not be reported as representing the official views of the OECD or of its Member countries or FAO. The original reference number of the figure and redundant text were omitted.

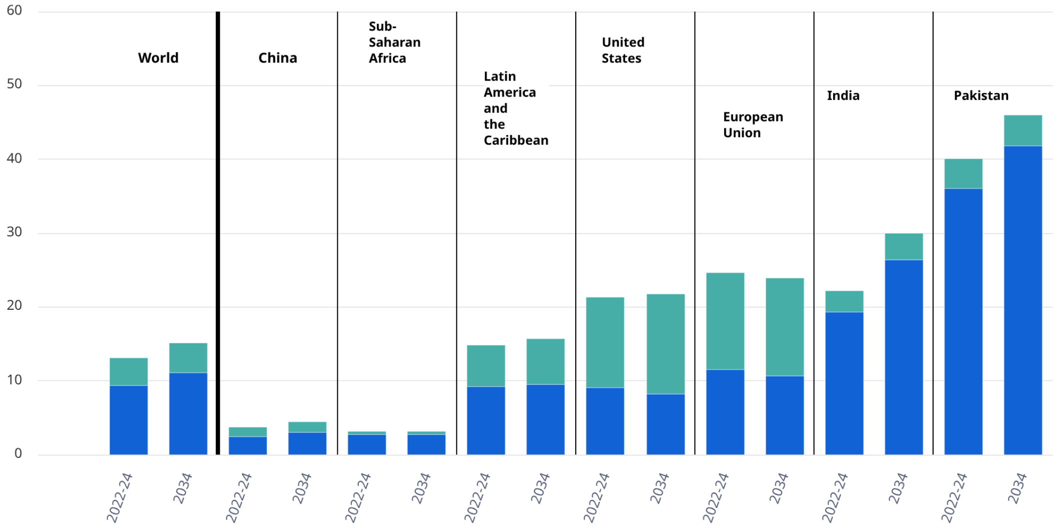


Figure 1.3: *Per capita* consumption of processed (light-blue bars) and fresh (blue bars) dairy products in milk solids (Kilograms/person). Milk solids are calculated by adding the amount of fat and non-fat solids for each product; processed dairy products include butter, cheese, skim milk powder and whole milk powder. This is an adaptation of an original work by the OECD and FAO. The opinions expressed and arguments employed in this adaptation should not be reported as representing the official views of the OECD or of its Member countries or FAO. The original reference number of the figure and redundant text were omitted.

Based on the EAT-Lancet Commission report (Willett et al. 2019), the global food system requires urgent transformation to provide healthy diets for 10 billion people by 2050 while operating within planetary boundaries. This necessitates substantial dietary shifts toward plant-based foods, reducing meat consumption by over 50%, increasing vegetable and legume intake by 100%, improving food production practices and halving food waste globally.

1.2.5 The case for resilience in dairy farming

These and more issues are becoming more relevant in the context of climate change, which is expected to exacerbate the challenges faced by the agricultural sector. Here resilience is one of the goals to pursue in order to maintain the reliability and sustainability of feed and food production. Resilience can indeed be corroborated through the actions of all the subjects of the agricultural sector, in a scheme where the resilience of each actor can enhance the resilience of the whole system.

The previous considerations are valid for the livestock sector and particularly for the dairy sector, which is the focus of this thesis. This sector, specifically, is facing several challenges, including climate change, resource scarcity, and changing consumer preferences. Improving the resilience of dairy cattle towards heat stress, diseases and other stressors can help to ensure the sustainability of dairy production in the face of these challenges (Abera 2024). From an economic perspective, resilience is a desirable trait. It can exert beneficial impacts due to its association with crucial aspects of labor requirements on agricultural establishments. Indeed, the resilience exhibited by these bovines may necessitate a reduction in the amount of attention and care required from farmers, thereby leading to a decrease in labor costs and an enhancement in farm efficiency (Perrin et al. 2020; Kołoszycz et al. 2024).

1.3 Resilience in dairy cattle

1.3.1 Environmental stressors

In addition to previous considerations, dairy cows are exposed to various environmental stressors that can affect their health, productivity, and welfare. Among these stressors, heat stress is one of the most significant challenges faced by dairy cows, especially in the context of climate change (Thornton et al. 2022; Cartwright et al. 2023). As ruminants, cows are particularly sensitive to heat stress because rumen fermentation produces a lot of heat that must be dissipated to maintain homeothermy. Heat stress affects the milk production, reproduction, health, and welfare of dairy cows, especially in tropical and subtropical climates (Oliveira et al. 2025); directly and indirectly (reduced feed intake) lowers milk production in a delayed pattern that can rise 24-48 hours after the stress event, and impacts reproductive performance by reducing estrus expression, conception rates, and increasing embryo mortality (Oliveira et al. 2025).

Another relevant stressor is represented by diseases, which can affect the health and productivity of dairy cows. Common diseases in dairy cattle include mastitis, lameness, respiratory diseases, and metabolic disorders; these diseases can lead to reduced milk production, increased veterinary costs and decreased animal welfare (Huxley 2013; Hogeveen et al. 2019).

Antimicrobial treatments for diseases can lead to the insurgence of antimicrobial resistance (AMR). AMR is a growing concern in cattle farming because can increase animal illness and production costs, and more importantly can affect humans via direct contact or through the food chain (Call et al. 2008). At present, AMR bacteria are found more frequently in conventional farms than organic ones, likely because of a larger use of antibiotic. Conversely, while most antimicrobial use in dairy cattle is for treating mastitis infections, the AMR in mastitis pathogens remains relatively low respect to enteric ones that results to be more keen to develop antimicrobial resistance. According to Call et al. (2008), to reduce AMR in a herd a combination of actions is required: reduced antimicrobial use, improved management and possibly new biological strategies, such as the improvement of cows' resilience to make them more efficient in coping with environmental stressors as infections.

1.3.2 Farming strategies to enhance resilience

Relevant and interesting strategies for enhancing the resilience of cows are those adopted by farmers for everyday herd management, including barn design, the use of cooling systems (such as fans, misters, and sprinklers), nutrition and water management, provision of shade, adjustment of milking times, monitoring of cow behavior and health, farm staff training (Oliveira et al. 2025). To mitigate heat stress in loco these actions can be developed: the adoption of mechanical ventilation systems, either based on positive or negative pressure; and the application of evaporative cooling methods, such as misting, spraying, or the use of evaporative cooling panels. Additional measures comprise the installation of sprinklers in feeding alleys and waiting areas – particularly relevant in Compost Barn systems – and the selection of housing systems appropriate to the specific climatic conditions, taking into account both ambient temperature and humidity. Furthermore, continuous monitoring of environmental variables and animal physiological responses is essential to ensure effective heat stress management (Oliveira et al. 2025). Ultimately, it is important to recognize that no single strategy provides a universal solution; rather, mitigation approaches must be adapted to local climatic conditions, housing design, and operational constraints.
Nutrition is a particularly critical factor. In a review by Min et al. (2019), eight nutritional

strategies were recognized as essential for mitigating detrimental effects of heat stress: the use of dietary fats, fiber, additives, minerals, vitamins, metal and ion buffers, plant extracts, and other anti-stress additives.

1.3.3 Genetic approaches to resilience

Genetic gain is a key focus of breeding programmes due to its potential to improve the resilience of dairy cattle. This represents a particularly significant advantage, as genetic improvement is cumulative and permanent across generations. The identification of genetically superior animals exhibiting an enhanced capacity to withstand environmental stressors is therefore of critical importance for the advancement of sustainable breeding strategies. To achieve this objective, a range of genomic tools and methodologies can be employed, including genome-wide association studies (GWAS) and genomic selection (Silpa et al. 2021).

Over the past century, selection goals in dairy cattle have undergone profound changes, transitioning from a narrow focus on milk production to more balanced breeding objectives. However, this evolution has inadvertently contributed to increased environmental sensitivity in high-producing dairy cows.

At the turn of the 20th century, dairy cattle selection focused almost exclusively on increasing milk production (Miglior et al. 2017). From the 1930s to the 1970s, selection remained solely oriented toward maximizing milk and fat yield, with little consideration for other traits. This production-centric approach was economically justified, as milk yield represented the most financially lucrative trait. However, the long-term consequences of this narrow selection strategy became increasingly apparent as the century progressed. By the 1980s and 1990s, evidence accumulated that intense selection for production had resulted in unfavorable correlated responses in fertility, health, and longevity (Miglior et al. 2017). Recognition of these trade-offs prompted a paradigm shift toward the inclusion of functional traits in selection indices. Modern selection indices now incorporate traits such as reproductive performance, health status, calving ability, longevity, and workability (Miglior et al. 2017; Cole et al. 2018).

Despite these advancements, the intense selection for increased milk yield over the past five decades has resulted in modern high-yielding dairy cows that are substantially more environmentally sensitive than their predecessors (Brito et al. 2021). The unfavorable genetic relationships between production and fitness-related traits can be understood through the lens of resource allocation theory (Rauw et al. 1998). When two biological processes compete for limited resources, selection for one trait inevitably compromises the other. In intensive, high-input production systems with controlled environments, these trade-offs could be managed through superior management and nutrition. However, as production systems diversify and environmental controls become less reliable, high-producing animals demonstrate reduced capacity to maintain multiple phenotypes of interest simultaneously.

Moving forward, sustainable dairy cattle breeding requires balanced selection objectives that simultaneously improve animal health and welfare, productive efficiency, environmental impact, and overall resilience (Brito et al. 2021). This necessitates the measurement and appropriate weighting of all relevant traits in selection indices, including those related to adaptation, robustness, resilience and environmental efficiency.

The assessment of resilience can be conducted using various variables (König et al. 2019), including:

- Physiological (e.g., body temperature, respiratory rate, heart rate)
- Behavioral (e.g., feeding behavior, lying behavior, social interactions)
- Production (e.g., milk yield, milk composition, growth rate)
- Health (e.g., disease incidence, immune response)

These phenotypes can be collected noninvasively and on a large scale, provided that appropriate measurement tools are implemented (König et al. 2019). In this context, resilience phenotyping emerges as a particularly complex challenge. The phenotyping of resilience traits necessitates dense and longitudinal data acquisition, thereby relying on sophisticated technological infrastructures, including sensor-based and automated phenotyping systems, which are not yet widely available to farmers.

Under the umbrella of resilience phenotyping, numerous proxies for resilience have been proposed in the literature, most of which are derived from milk yield data recorded by automatic milking systems (AMS), then are derived from automatic milk feeders (calf resilience, Graham et al. (2024)), automatic feed intake recorders and activity levels (Poppe et al. 2022b). These proxies include measures such as the variance of residuals from expected production levels, autocorrelation of production deviations, and other statistical metrics that capture the consistency and stability of milk yield over time (König et al. 2019; Kašná et al. 2022). Because these proxies are central to the objectives of this thesis, they will be discussed in greater detail in subsequent chapters.

It was demonstrated that incorporating resilience proxies as selection criteria into a genetic index, wherein economic values assigned as 0.3 for milk yield, 0.3 for longevity, 0.2 for udder health, and 0.2 for resilience, could enhance economic profit (labor cost, Berghof et al. (2019b)). This improvement occurred because a reduction in milk yield was counterbalanced by increases in resilience, longevity, and udder health. There is a general trend indicating that resilience tends to be unfavorably associated with milk yield but favorably associated with health-related traits and longevity (Maskal et al. 2024). In this context, greater variability in milk yield (i.e., lower resilience) is typically observed in animals with higher production levels, while animals with lower production levels tend to exhibit more stable milk yield (Berghof et al. 2019b). However, the complex relationship between resilience and milk production is not fully understood and studies which try to disentangle the scale effect (higher phenotype, higher variability) from the true genetic correlation are emerging, as in Mancin et al. (2025).

It is noteworthy that enhancing resilience can be achieved by implementing breeding objectives focused on improving health traits and longevity. These objectives have already been incorporated into the breeding goals of numerous countries (Berghof et al. 2019b). The inclusion of resilience traits in breeding objectives may be particularly relevant for breeding programs in which health and longevity traits are not prioritized, or where selection pressure is primarily directed toward production traits, as commonly observed in swine and poultry species. Furthermore, the integration of resilience traits into selection goals may offer a distinct advantage in production systems equipped with technologies capable of accurately recording resilience-related phenotypes, particularly given that functional and health traits remain difficult to measure routinely (Berghof et al. 2019b). Regarding functional and health traits, it is worth noting that sensors and technologies can be very helpful in monitoring these traits. For example, the use of AMS integrated sensors can monitor udder health through milk electrical conductivity, while activity monitors can track behavioral patterns such as lying, standing, walking, and temperament in AMS robots. Additionally, these technologies can monitor feeding and rumination behaviors, body weight (BW), and udder conformation. Establishing standardized protocols for data collection, preprocessing, and harmonization is essential (Brito et al. 2025).

Recent research has established both the genetic parameters and genomic architecture of resilience proxies in large dairy populations. Regarding genetic parameters, variance-based resilience indicators consistently show moderate heritabilities across studies and populations: 0.10–0.21 for log-transformed variance (LnVar) (Poppe et al. 2021a; Chen et al. 2023; Kessler et al. 2024) and 0.04–0.08 for variance of relative daily milk yields (Poppe et al. 2021a; Kessler et al. 2024). Autocorrelation-based indicators exhibit lower heritabilities (0.01–0.08) (Poppe et al. 2021a; Chen et al. 2023; Kessler et al. 2024), suggesting recovery speed is less heritable than production stability. Genetic correlations across lactations are stronger between lactations 2 and 3 (0.88–0.96) than between lactations 1 and 2 (Chen et al. 2023), indicating resilience in primiparous cows represents a partially distinct trait. Forward prediction studies confirm temporal stability, with genetic correlations of 1.0 between earlier and more recent daughter groups across different herd-year-seasons (Poppe et al. 2021b). Resilience indicators show favorable genetic correlations with health and longevity: LnVar correlates 0.21–0.38 with udder health traits and 0.26–0.30 with metabolic health, while longevity shows negative correlations with variance-based indicators (–0.18 to –0.38) (Poppe et al. 2021a). Fertility traits also demonstrate moderate correlations (0.34–0.54) with variance-based indicators (Mancin et al. 2025). Genome-wide association studies have confirmed the polygenic nature of resilience while identifying specific genomic regions. Chen et al. (2024) identified significant associations across all autosomes in North American Holstein cattle, with notable signals on BTA14 (1.31–3.12 Mb, encompassing DGAT1) and BTA6 (87.8–89.4 Mb) for LnVar. Functional enrichment analyses revealed pathways related to intestinal homeostasis, protein metabolism, aryl hydrocarbon receptor signaling, and butanoate metabolism. Keßler et al. (2025b) extended these findings through across-breed GWAS in German Holstein, Fleckvieh, and Brown Swiss populations, identifying significant SNPs on BTA5, BTA14, BTA2, and BTA8, with many regions overlapping previously identified QTL for production, health, and longevity traits. The selection index approach by Keßler et al. (2025a) demonstrated that combining multiple resilience indicators maximizes correlated response in health traits, with an optimal composite index achieving heritability of 0.18–0.22 and health trait correlations of 0.44–0.51.

Chapter 2

Aims

2.1 Research context

The data comes from a close collaboration with a large dairy farm in the Northern Po valley, providing access to two fundamental set of data information: phenotypes from Automatic Milking Systems (AMS) and genotypes from SNP chips for all cows.
The adoption of AMS has enabled the collection of large volumes of high-frequency data, enabling the derivation of quantitative indicators that characterize dairy cows' ability to cope with environmental challenges–namely, their resilience. Furthermore, the availability of genotypic information for the entire herd across multiple years has provided the genomic foundation necessary to investigate the genetic basis underlying resilience-related phenotypes.

The specific objectives of this research were developed based on the presence of these two key resources.

2.2 Research objectives

2.2.1 Characterization of resilience indicators

New resilience measures have recently been proposed in the literature; however, their effectiveness in capturing resilience is not yet fully understood. Using a dataset from a single large farm – representing a unique environment – with accurately recorded daily milk records from an AMS, it was aimed to evaluate these proposed measures. Additionally, the analyses sought to provide insights into various lactation curve models that can serve as baselines, representing the unperturbed state of the lactation curve (Chapter 3).

2.2.2 Simulation study on resilience proxies

The relationship between resilience proxies and resilience as defined by Colditz et al. (2016) is not fully understood, and it is not clear how well these proxies capture the underlying genetics of resilience. To address partially this gap, I aimed to get insights on the relationshipo between resilience proxies and resilience by simulating quantitative resilience phenotypes based on a precise definition of resilience, and then evaluating the correlation between the simulated resilience phenotypes and the resilience proxies (Chapter 4).

2.2.3 Genomic analysis of resilience

The genetic architecture of four resilience proxies, namely the logarithm of the variance (LnVar), lag-1 autocorrelation (rauto), daily perturbations (DPert) and weighted frequency perturbations (wfPert), was investigated through a genome-wide association study (GWAS), in order to depict associations and functional annotation of SNPs related to these traits (Chapter 5).

2.3 Additional contribution

2.3.1 Copy Number Variants mapping

Different sources of genomic information can be used to get insights into the genetic background of complex traits, such as resilience. Single nucleotide polymorphisms (SNP) are the most used genomic markers in animal breeding. However, other types of genetic variations can be exploited, such as copy number variants (CNV).

CNV are structural variations involved in several biological processes, such as immune response, metabolism and reproduction, therefore involved in resilience. We published the first mid-large scale CNV map in Italian Holstein cows, using more than 4,000 genotyped animals (Chapter 6). This CNV map can be used in future studies to investigate the genetic background of resilience phenotypes.[1]

---

[1] This study was not originally planned as part of the Ph.D. program. It arose from the availability of a dataset comprising more than 4,000 genotyped Holstein cows from commercial farms in Lombardy, Italy, which was collected during the development of the GENORIP project (founded by EAFRD Rural Development Program 2014–2020, Management Autority Regione Lombardia - OP. 16.1.01 Project ID n. 201801062430 –'Operational Group EIP AGRI').

Chapter 3

Characterization of resilience indicators and lactation curve models to detect resilience in dairy cattle

Andrea Delledonne[ab], Christian Maltecca[b], Francesco Tiezzi[c], Chiara Punturiero[a], Maria Giuseppina Strillacci[a], Alessandro Bagnato[a]

[a]*Department of Veterinary Medicine and Animal Sciences, University of Milan LA STATALE, Lodi, Italy;* [b]*Department of Animal Science, North Carolina State University, Raleigh, NC, USA;* [c]*Department of Science and Agricultural, Food, Environmental and Forest Technologies (DAGRI), University of Firenze, Florence, Italy*

This manuscript is the outcome of my research period as visiting Ph.D. student at North Carolina State University and it will be submitted soon, in conjunction with the simulation study (Chapter 4).

3.1 Introduction

Over the past fifteen years, several researchers have focused on capturing resilience using high-throughput data obtained from automatic milking systems (AMS) (Kašná et al. 2022). Longitudinal data have been extensively utilized to develop models (Adriaens et al. 2020; Taghipoor et al. 2023) and proxies (Elgersma et al. 2018; Poppe et al. 2020; Adriaens et al. 2021; Wang et al. 2022a; Chen et al. 2023; Kessler et al. 2024) that serve as tools to assess dairy cattle resilience. Proxies (also known as resilience indicators) are based on a simple concept: measuring and aggregating the deviations between an estimated unperturbed state and the actual state. A key underlying assumption is that the actual state can be observed precisely, whereas the unperturbed state cannot.

Resilience indicators can be divided into two main categories: deviation-based and perturbation-based. The first category (deviation-based) includes several similar indicators that rely on the use of the complete set of deviations, such as the logarithm of the variance (Adriaens et al. 2020; Poppe et al. 2020; Wang et al. 2022a; Chen et al. 2023; Kessler et al. 2024), the area under the curve (AUC) (Zefreh et al. 2024), the autocorrelation (Scheffer et al. 2018; Poppe et al. 2020; Zefreh et al. 2024; Punturiero et al. 2025) and the skewness (Zefreh et al. 2024; Punturiero et al. 2025); the second category (perturbation-based) includes indicators based solely on negative deviations, that is, when the observed phenotypic value of milk yield is lower than the expected value based on the estimated lactation curve. (Adriaens et al. 2020; Chen et al. 2023; Punturiero et al. 2025; Guinan et al. 2025a).

Critical requirements for reliable resilience assessment were outlined in three points by Taghipoor et al. (2023): defining the appropriate expected performance, ensuring sufficient data availability, and personalizing individual analysis. There is no strong consensus on how to define the appropriate expected performance for measuring resilience. Lactation curve models developed by Ali&Schaeffer (Wang et al. 2022a; Chen et al. 2023), Wilmink (Wang et al. 2022a; Chen et al. 2023), Wood (Adriaens et al. 2020; Wang et al. 2022a; Chen et al. 2023; Zefreh et al. 2024) and Nelder (Wang et al. 2022a) have been implemented, along with polynomial models (Poppe et al. 2020; Chen et al. 2023; Zefreh et al. 2024; Punturiero et al. 2025) and spline interpolation (Kessler et al. 2024). Differently, some studies have mitigated the influence of milk yield drops to better fit the unperturbed lactation curve model (Adriaens et al. 2020; Poppe et al. 2020; Wang et al. 2022a; Chen et al. 2023; Zefreh et al. 2024; Kessler et al. 2024; Punturiero et al. 2025; Guinan et al. 2025a).

Recent genetic evaluations have demonstrated that deviation-based indicators exhibit substantially different heritability. LnVar, RMSE, LMS, ΔMY (Guinan et al. 2025a) show moderate heritability estimates (0.13-0.23), making them suitable for breeding applications. In contrast, autocorrelation displays lower heritability values (0.02 to 0.14), while perturbation-based indicators, such as the Weighted Frequency and Daily Perturbation, exhibit low heritability (<0.13) (Poppe et al. 2020; Chen et al. 2023; Kessler et al. 2024; Guinan et al. 2025a; Punturiero et al. 2025). These differences in genetic parameters suggest that the choice of resilience indicator significantly influences the potential for genetic improvement.

Despite the well-established genetic merit of variance-based resilience indicators, the impact of different underlying lactation curve modelling approaches on the consistency of these indicators' calculations and the subsequent ranking of cows remains largely unexplored. In this context, some studies proposed using different models for specific purposes (Ranzato et al. 2024): quantile regression for fast, large-scale applications; iterative Wood for perturbation detection; and the Perturbed Lactation Model (PLM, Ben Abdelkrim et al. (2021a)) as a general framework for accurate perturbation characterization.

In this study, daily milk yield data collected from AMS on a dairy farm in northern Italy were used to calculate resilience indicators. These indicators were based on the entirety of deviations (deviation-based indicators) derived from various models, including classical parametric lactation curves such as Wilmink, Wood, and Ali&Schaeffer, as well as more data-driven models like moving averages and spline functions. The objective was to evaluate how different lactation curve models and resilience indicators perform in capturing resilience and in phenotypically ranking cows according to their resilience indicator values.
The initial hypothesis was that these models would produce different results and rank cows differently within the same resilience indicator. If true, this discrepancy could lead to incorrect conclusions regarding the animals' resilience. To address this, two consecutive steps were undertaken: first, the models were characterized based on how they rank cows within each indicator; second, a principal component analysis (PCA) was performed to identify latent variables and loadings that may reflect the differing importance of the models.

3.2 Materials and methods

3.2.1 Data collection and preprocessing

The data used in this study were collected during the years 2022-2024 in a farm in Northern Italy. The farm is family-run, with almost 500 actual lactating Italian Holstein cows milked via AMS. The daily milk yields for each cow and each lactation were recorded by the AMS and used in the study.

The data were preprocessed to have a dataset with lactations equally represented. The preprocessing steps were the following ones:

- Daily milk yield missing (NA) or equal to zero were removed.
- Outliers three standard deviations above the mean were removed, after visual inspection of the data.
- Only lactations starting before the 15th days in milk (DIM) and ending after the 280th were considered.
- Lactations were terminated at the 305th DIM.
- Lactations with gaps longer than 20 DIM were removed; this was done to retain as much data as possible for the analysis, while still ensuring that the lactation curves were not overly fragmented.

- After estimating lactation curve models but before calculating resilience indicators the first 10 DIM were removed, due to the poor fitting at the initial phase of some models.

The initial dataset consisted of 774,048 observations from 1,161 cows and 2,518 lactations. After the preprocessing steps, the final dataset was composed by 514,131 observations, for a total of 819 cows and 1,423 lactations.

3.2.2 Lactation curve models

Eighteen lactation curve models were implemented to estimate the unperturbed state; all of them were run via the R programming language (R Core Team 2023). They were selected based on the idea of mixing different types, such as classical and parametric or data-driven methods; linear and nonlinear approaches; methods that remove perturbations and those that do not; and different settings when possible, as in the case of quantile polynomial and moving median techniques. For a tabular view see Table 3.1a, which presents the lactation curve models listed using their abbreviations as written throughout the paper.

- Ali&Schaeffer [Ali] (Ali et al. 1987)

  $$\mathrm{Y_t} = p_0 + p_1 \cdot \gamma_t + p_2 \cdot \gamma_t^2 + p_3 \cdot \omega_t + p_4 \cdot \omega_t^2 + e_t,$$

  a regression model where $^{\mathrm{t}}$ is the i^{th} DIM, $\gamma_t = t/30$ and $\omega_t = log305/t$. The parameter p_0 is associated with peak yield, p_1 and p_2 are associated with the decreasing slope of the curve, p_3 and p_4 are associated with the increasing slope of the curve.

- Wilmink (Wilmink 1987)

  $$\mathrm{Y_t} = a + b \cdot t + c \cdot e^{-0.07 \cdot t},$$

  a curve function where t is the i^{th} DIM, the parameter a is associated with level of production, b with production decrease after peak yield, c with production increase towards peak yield and the factor -0.07 for the moment of peak yield (originally 0.05 stand for about 50 days post partum).

- Wood (Wood 1967)

  $$\mathrm{Y_t} = a \cdot t^b \cdot e^{-c \cdot t},$$

  a gamma-type curve where t is the i^{th} DIM, a is the parameter associated with peak yield, b with increasing slope and c with decreasing slope. The a, b and c parameters were estimated via ordinary least squares after linear transformation of the gamma model.

- Nelder (Nelder 1966; Wang et al. 2022a)

  $$\mathrm{Y}_{\mathrm{t}}^{-1} = a + b \cdot t^{-1} + c \cdot t,$$

  an inverse polynomial function.

- Iterated curves: the Ali&Schaeffer [IAli], Wilmink [IWilmink], Wood [IWood] and Nelder [INelder] curves were iterated as described in (Adriaens et al. 2020). In each iteration, daily milk yields less than 85% of the expected milk yield were removed, and the model was re-estimated. Parameters from the final iteration were used to estimate the target trajectory above the original data. The iterative process was run up to twenty times or until the root mean square error (RMSE) between successive iterations was less than 0.1.

- Moving median [MM]: data-driven curve calculates the i^{th} observation as the average of a surrounding moving window of observations, which shifts according to the i^{th} observation being calculated. Three window sizes – 5, 11, and 21 days – were tested. This task was performed using the 'zoo' R package (Zeileis et al. 2005).

- Polynomial regression, third order [Poly3]

  $$\mathrm{Y}_{\mathrm{t}} = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot t^3,$$

  where t is the i^{th} DIM.

- Polynomial regression, fourth order [Poly4]:

  $$\mathrm{Y}_t = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot t^3 + \beta_4 \cdot t^4,$$

  where t is the i^{th} DIM.

- Quantile polynomial regression, fourth order: the mathematical model is equivalent to the Poly4 model but has the distinctive feature of estimating a specific quantile as the expected value of the response variable (Koenker 2005). It was previously implemented with the aim of fitting a model less sensitive to drops in milk yield, thereby better reflecting a true unperturbed state. This task was accomplished using the 'quantreg::rq' function in R (Koenker 2023), testing three quantile orders: 0.50 [Q05], 0.70 [Q07] and 0.80 [Q08].

- Adaptive spline interpolation [ASpline]: an adaptive spline regression that automatically selects the optimal number and position of knots through a penalized likelihood approach, removing unnecessary ones (Goepp et al. 2018). The function was available in Goepp et al. (2022).

- Weighted penalized spline [WSpline]: a penalized spline regression (Ramsay et al. 1997) with weighted data points as developed by Kessler et al. (2024). The weighting algorithm calculates individual weights for each DIM based on the frequency that daily milk yields fall above linear interpolations within sliding windows. Weights influence the spline fitting, thereby creating lactation curves representative of unperturbed production states. The spline interpolation was implemented using the ‘pspline‘ R package (Ramsey et al. 2024) with 5 degrees of freedom. The complete weighting algorithm is detailed in Appendix 1 of Kessler et al. (2024).

3.2.3 Resilience indicators

The resilience indicators, along with their explanations and mathematical formulas, are outlined below and summarized in Table 3.1b. Each resilience indicator was calculated for each cow and each lactation, as many times as there were unperturbed models included in the analysis (n = 18).

- Absolute deviation average

  $$\text{AbsAvg} = \frac{\sum_{t=1}^{n} |x_t|}{n}$$

  where n is the number of observations (number of DIM) and x_t is the deviation at time t (this nomenclature is the same for all indicators).

- Autocorrelation of deviations

  $$\text{AC1} = \frac{\sum_{t=1}^{n-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2},$$

  defined as the correlation between two series of deviations (as time series) that are separated by 1 (lag 1) timepoint (DIM).

- Area under the curve

  $$\text{AUC} = \int_{t_{start}}^{t_{end}} x(t)\, dt \approx \sum_{i=1}^{n-1}(x_{i+1} - x_i)(t_{i+1} - t_i),$$

  where the integration was approximated by trapezoidal rule with the ‘smplot2::sm–auc‘ R function (Min et al. 2021).

- Delta between expected average milk yield and actual deviated milk yield

  $$\Delta MY_i(\%) = \frac{\frac{\sum (yield_i - \hat{yield_i})}{n_i}}{\frac{\sum \hat{yield_i}}{n_i}} \times 100,$$

  as introduced in a recent study to detect perturbations of individual cows at the pen level (Guinan et al. 2025a). The name of the indicator was kept as the original one.

- Logarithm of the variance of the deviations

  $$\mathrm{LnVar} = \log\left(\frac{\sum_{t=1}^{n}(x_t - \bar{x})^2}{n}\right).$$

- Logarithm means square of deviations

  $$\mathrm{LMS} = \log\left(\frac{\sum_{t=1}^{n} x_t^2}{n}\right).$$

- Relative logarithm of the variance of the deviations

  $$\mathrm{LnVaRel} = \frac{\log(\frac{\sum_{t=1}^{n}(x_t - \bar{x})^2}{n})}{\log(\frac{\sum_{t=1}^{n}(x_t - \bar{x})^2}{n} + 1)},$$

  LnVaRel was introduced in (Kessler et al. 2024) as a method to standardize deviations, accounting for the scaling effect – the phenomenon in which the variance of deviations increases as the mean increases (Kessler et al. 2024; Berghof et al. 2019b).

- Root means square error

  $$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(x_t - \bar{x})^2}{n}}.$$

- Skewness of the deviations

  $$\mathrm{Skew} = \frac{\sum_{t=1}^{n}(x_t - \bar{x})^3}{n \cdot s^3},$$

  where s is the standard deviation of the deviations. Skew is the measure of the asymmetry of the deviations' distribution. It is calculated as the third standardized moment of the deviations, thanks to the 'e1071::skewness' R function (Meyer et al. 2023).

Table 3.1: Lactation curve models and resilience indicators included in this study

(a) Lactation curve models grouped by methodological types

Parametric	Iterated refinement	Quantile polynomials	Data driven
Wilmink	Iterated Wilmink [IWilmink]	Fourth degree polynomial, quantile order 0.50 [Q05]	Adaptive spline [ASpline]
Wood	Iterated Wood [IWood]	Fourth degree polynomial, quantile order 0.70 [Q07]	Weighted spline [WSpline]
Ali&Schaeffer [Ali]	Iterated Ali&Schaeffer [IAli]	Fourth degree polynomial, quantile order 0.80 [Q08]	Moving median, window size 5 [MM5]
Nelder	Iterated Nelder [INelder]		Moving median, window size 11 [MM11]
Third degree polynomial [Poly3]			Moving median, window size 21 [MM21]
Fourth degree polynomial [Poly4]			

(b) Resilience indicators

Lag-1 autocorrelation [AC1]
Average of absolute deviations [AbsAvg]
Area under the curve [AUC]
Delta between expected average milk yield and actual deviated milk yield [δMY]
Logarithm of the variance of the deviations [LnVar]
Logarithm of the relative variance of the deviations [LnVaRel]
Logarithm of the mean square of the deviations [LMS]
Root mean square error [RMSE]
Skewness of the deviations [Skew]

3.2.4 Models' characterization

To formally characterize the models, two distinct approaches were employed and evaluated separately. First, the phenotypic rankings of the cows within each indicator, derived from all included models, were compared to assess their consistency. Second, Principal Component Analysis (PCA) was performed on the models to gain insights into the presence of latent components when different models are used to derive the same resilience indicator and to identify the different models' importances.

Ranking the cows

Resilience indicators were adjusted for known systemic effects (see Equation 3.1), and cows were subsequently ranked based on the corrected phenotypic values. Each indicator was calculated separately for each implemented model, resulting in multiple sets of indicator values (a total of 162 (18 models × 9 indicators) combinations were computed). For each indicator, rankings derived from all pairwise combinations of models were compared by grouping the top fifty and bottom fifty positions and then determining the percentage of cows that consistently fell within the same rank distribution zone, which is particularly relevant in the context of animal breeding.
The statistical formula of the linear model implemented was

$$Y = \mu + \beta_j + \gamma_k + \delta_{l(j)} + \zeta_m + \eta_n + \varepsilon_{ijklmn}, \tag{3.1}$$

where Y is the resilience indicator; μ is the grand mean effect; β_j is the effect of the j^{th} level of parity order, from $j = 1$ to $j = 4$; γ_k is the effect of the k^{th} level of the year of calving, from $k = 1$ to $k = 4$; $\delta_{l(j)}$ is the effect of the l^{th} level of month of calving nested within the year of calving; ζ_m is the effect of the m^{th} age at calving, from $m =$ to $m = ...$; η_n is the effect of the n^{th} level (as factor) of estimated breeding value, from $n = 1$ to $n = 3$; ε_{ijklmn} is the residual term. Estimated breeding values were obtained from the GENOCOW service provided by the Italian Holstein Brown and Jersey National Breeder Association (ANAFIBJ).

Associazione Nazionale Allevatori della Razza Frisona, Bruna e Jersey Italiana

Models' Principal Component Analysis

The corrected indicators' values were analyzed using PCA. After standardizing the data by subtracting the mean and dividing by the standard deviation of the trait, eigenvalues and eigenvectors were computed using the 'base::eigen' function applied to the correlation matrix (R Core Team 2023). The new components were then derived from the scaled original variables. PCA was performed on the eighteen lactation curve models within each of the nine resilience indicators (AbsAvg, AC1, AUC, LMS, LnVar, LnVaRel, RMSE, Skew, and δAvg). This analysis aims to provide insights into the presence of latent components when different models are used to derive the same resilience indicator and to identify the different models' importances.

3.3 Results

3.3.1 Consistency of the rankings

The consistency of lactation curve models in ranking the cows varies substantially according to the specific models compared. Table 3.2 presents glimpses of this comparison, showing the first six and last six rows of the complete model combinations dataset. The complete comparison set is available on the supplementary Table S1. To avoid misconceptions, it is important to keep in mind that the results are presented as aggregates across all lactation curves and cows; however, the analyses were performed separately for each lactation and cow before being aggregated.

The degree of similarity between lactation curve models was derived from the averages and standard deviations of the rank agreement of the different resilience indicators. The most consistent model pair is Q07-Q08, achieving a remarkable degree of similarity of 86% ± 8%, followed by model pairs including Q07-WSpline (77% ± 10%) and Q08-WSpline (75% ± 8%), as demonstrated in Table 3.2. These model pairs have high averages and low standard deviations across resilience indicators, diverging from Q05-Q07 (75% ± 25%) or Poly4-Q05 (74% ± 27%), where the average is high but the standard deviation is also high; even the IWood-Wood pair shows a similar pattern (66% ± 24%). The bottom-performing combinations in Table 3.2 reveal the challenges of comparing fundamentally different modelling approaches. Combinations involving ASpline with parametric models (e.g., ASpline-IWilmink: 12% ± 5%) show particularly poor agreement. Similarly, combinations involving moving median models with traditional parametric models demonstrate consistently low agreement across all indicators, with some combinations achieving less than 20% consistency (Table 3.2 and Table S1).

The pattern of between-model similarities demonstrated three distinct performance tiers. High-consistency pairs (similarity $\geq$ 70%) are dominated by closely related polynomial models and WSpline, with Q07-Q08, Q07-WSpline, Q08-WSpline, Q05-Q07, and Poly4-Q05 forming the top tier. These combinations consistently show agreement above 80% for variance-based indicators (AvgAbs, LMS, LnVar, LnVaRel, and RMSE) while maintaining reasonable performance for other indicators. Intermediate-consistency pairs (similarity 40-70%) include parametric model combinations with mathematical differences (e.g., Ali-Poly4: 65% ± 8%, Wilmink-Wood: 53% ± 6%) and some moving median internal comparisons (MM11-MM21: 66% ± 7%). Low-consistency pairs (similarity <20%) are predominantly cross-category comparisons between parametric and non-parametric approaches. The worst-performing combinations include ASpline-IWilmink (12% ± 5%), INelder-MM5 (14% ± 5%), and ASpline-INelder (13% ± 4%) (Table 3.2). Notably, ASpline combinations with parametric models consistently rank among the lowest performers, with AUC agreement often below 20% and overall similarities rarely exceeding 25% (Table S1).

The analysis reveals distinct patterns across different categories of resilience indicators. Variance-based indicators (AvgAbs, LMS, LnVar, LnVaRel, and RMSE) demonstrate remarkably consistent behavior, with the most similar model pairs achieving agreement percentages frequently exceeding 80% and showing a coherent downward trend as model dissimilarity increases (Table S1). For instance, the Q07-Q08 comparison achieves 93% and 96% agreement for LnVar and LnVarRel respectively, with similarly high values for other variance-based indicators (85-87% for LMS, RMSE, and AvgAbs). However, this pattern diverges for AUC, ΔAvg, and to a lesser extent, Skew indicators. While the most consistent model pairs (Q07-Q08, Q07-WSpline, Q08-WSpline) maintain strong agreement for all variance-based indicators, starting from lower position models show lower concordance for AC1, AUC, ΔAvg, and Skew.

Particularly striking is the behavior of AUC and ΔAvg, which show the most erratic patterns across model combinations. Even among highly similar models, these indicators demonstrate substantially lower agreement percentages, with some of the most consistent parametric model pairs showing agreement levels below 50%. For instance, several top-ranking parametric model pairs show dramatically reduced AUC and ΔAvg agreement: Poly4-Q05 shows only 41% agreement for AUC and a mere 15% for ΔAvg while maintaining 88-92% for variance-based indicators. Similarly, Q05-Q07 demonstrates 35% AUC agreement and 28% ΔAvg agreement despite 86-92% agreement for other indicators. The ΔAvg indicator emerges as the most inconsistent metric across nearly all model combinations, showing extreme variability even among closely related models. For example, the Poly4-Q07 combination, which maintains high agreement (82-88%) for variance-based indicators, drops to just 6% for ΔAvg. Similarly, the Poly4-WSpline combination shows only 7% ΔAvg agreement despite maintaining 77-85% agreement across variance-based metrics.

Table 3.2: Each cell represents the percentage of cows ranked within the same tails of the distribution when comparing two rankings of the same indicator but calculated using two different model. The table is sorted based on the degree of similarity, expressed as the mean ± standard deviation of the values in each row.

Model	AC1	AUC	AvgAbs	LMS	LnVar	LnVaRel	RMSE	Skew	ΔAvg	Degree of similarity
Q07-Q08	94	75	85	87	93	96	84	89	71	86 ± 8
Q07-WSpline	81	67	78	83	81	83	86	75	56	77 ± 10
Q08-WSpline	80	68	77	79	82	85	77	72	59	75 ± 8
Q05-Q07	89	35	92	86	91	92	84	74	28	75 ± 25
Poly4-Q05	84	41	88	88	92	91	90	75	15	74 ± 27
Q05-Q08	85	27	81	78	87	92	77	68	18	68 ± 27
IWood-MM5	18	18	17	9	12	13	20	17	11	15 ± 4
IWilmink-MM11	13	19	20	11	11	13	16	22	8	15 ± 5
INelder-MM5	20	19	16	8	7	11	16	17	14	14 ± 5
IWilmink-MM5	17	15	16	9	9	10	16	20	9	13 ± 4
ASpline-INelder	9	17	19	10	9	8	16	16	17	13 ± 4
ASpline-IWilmink	8	16	19	8	9	8	15	18	11	12 ± 5

3.3.2 Models' Principal Component Analysis

Principal component analysis (PCA) was performed on the eighteen lactation curve models within each of the nine resilience indicators (AbsAvg, AC1, AUC, LMS, LnVar, LnVaRel, RMSE, Skew, and ΔAvg) to better visualize the similarities and differences among models in generating the values of resilience indicators.

The first two principal components explained a substantial proportion of the variance across all resilience indicators (Supplementary Table S2): PC1 accounted for 42–61% of the variance (highest for Skew at 61% and LnVaRel at 60%, lowest for ΔAvg at 42% and AUC at 43%), while PC2 accounted for 14–30% of the variance (highest for AUC at 30% and ΔAvg at 27%, lowest for Skew at 14%). Together, PC1 and PC2 explained 69–81% of the total variance across indicators (Supplementary Table S2).

To provide an overview of model behavior across all resilience indicators, Figure 3.2 presents the distribution of PC1 (Figure 3.2a) and PC2 (Figure 3.2b) loadings for each lactation curve model across the nine indicators analyzed. Details of PC1 and PC2 loadings are provided through the Figure 3.1 and Table S2.
The loading distributions of Ali, Poly3, and Poly4, Wood, Wilmink, and Nelder, in conjunction with MM models and ASpline, demonstrate a minimal interquantile range in the negative sign, accompanied by one or two extreme outliers in the positive sign (see Figure 3.2a). The lactation curve model considered before are distinctly different from other models, including WSpline, Q07, Q08, IAli, and IWood. The latter models exhibit a wider interquantile range, spanning from negative to positive values.
PC1 loadings show a reversed pattern for many models, with values ranging from -0.91 to 0.84. Most parametric models like Q07, Q08, and WSpline show strong positive PC1

loadings (0.73 to 0.81), while Wood and Ali show strong negative loadings (-0.91 and -0.74 respectively).

PC2 loadings reveal a more complex organizational structure that differentiates models within their parametric or non-parametric categories. Non-parametric smoothing models consistently cluster in a very wide range: MM5 (0.86 to -0.81), MM11 (0.88 to -0.84), MM21 (0.84 to -0.84), and ASpline (0.84 to -0.48), indicating they form a coherent group along the second dimension. Among parametric models, PC2 loadings show greater variability and indicator-specific patterns. Most parametric models span both positive and negative PC2 values depending on the specific resilience indicator, with Ali showing relatively small but consistent PC2 loadings (0.01 to 0.19) for traditional indicators compared to models like Wood (-0.43 to -0.17). Less broad distributions are achieved even by Q07, Q08, Poly3 and WSpline.

The WSpline model demonstrates behavior similar to other parametric models with strong negative PC1 loadings (-0.91 to -0.95) for traditional indicators but shows consistently positive PC2 loadings (0.12 to 0.30) (Figure 3.1). For ΔAvg, however, WSpline shifts to a positive PC1 loading (0.73) with a strong negative PC2 loading (-0.50), grouping it with other quadratic models.
Iterated models (IAli, IWilmink, IWood) show distinct patterns from their non-iterated counterparts. IAli and IWood maintain strong negative PC1 loadings for most indicators but exhibit different PC2 behavior compared to their base models. For ΔAvg, these iterated models show positive PC1 loadings (IAli: 0.84, IWood: 0.76), in direct contrast to traditional indicators. IWilmink stands out with both strong negative PC1 (-0.52 to -0.72) and consistently negative PC2 loadings (-0.59 to -0.51) for traditional indicators, making it one of the most distinctive among iterated models.

The analysis reveals that model relationships vary significantly depending on the resilience indicator examined. AUC and ΔAvg show the most distinctive patterns, with several models exhibiting markedly different loading patterns compared to other indicators. For instance, the IAli model shows a positive PC1 loading for ΔAvg (0.84) compared to strong negative loadings for other indicators (-0.76 to -0.89). Similarly, Q07 and Q08 show strong positive PC1 loadings for ΔAvg (0.78 and 0.81 respectively) compared to strong negative loadings for variance-based indicators (-0.90 to -0.92).
LnVaRel and Skew demonstrate the strongest model separation along PC1 (60% and 61% variance explained respectively), while AUC and ΔAvg show the strongest separation along PC2 (30% and 27% variance explained respectively, see Figure 3.1 and Table S2).

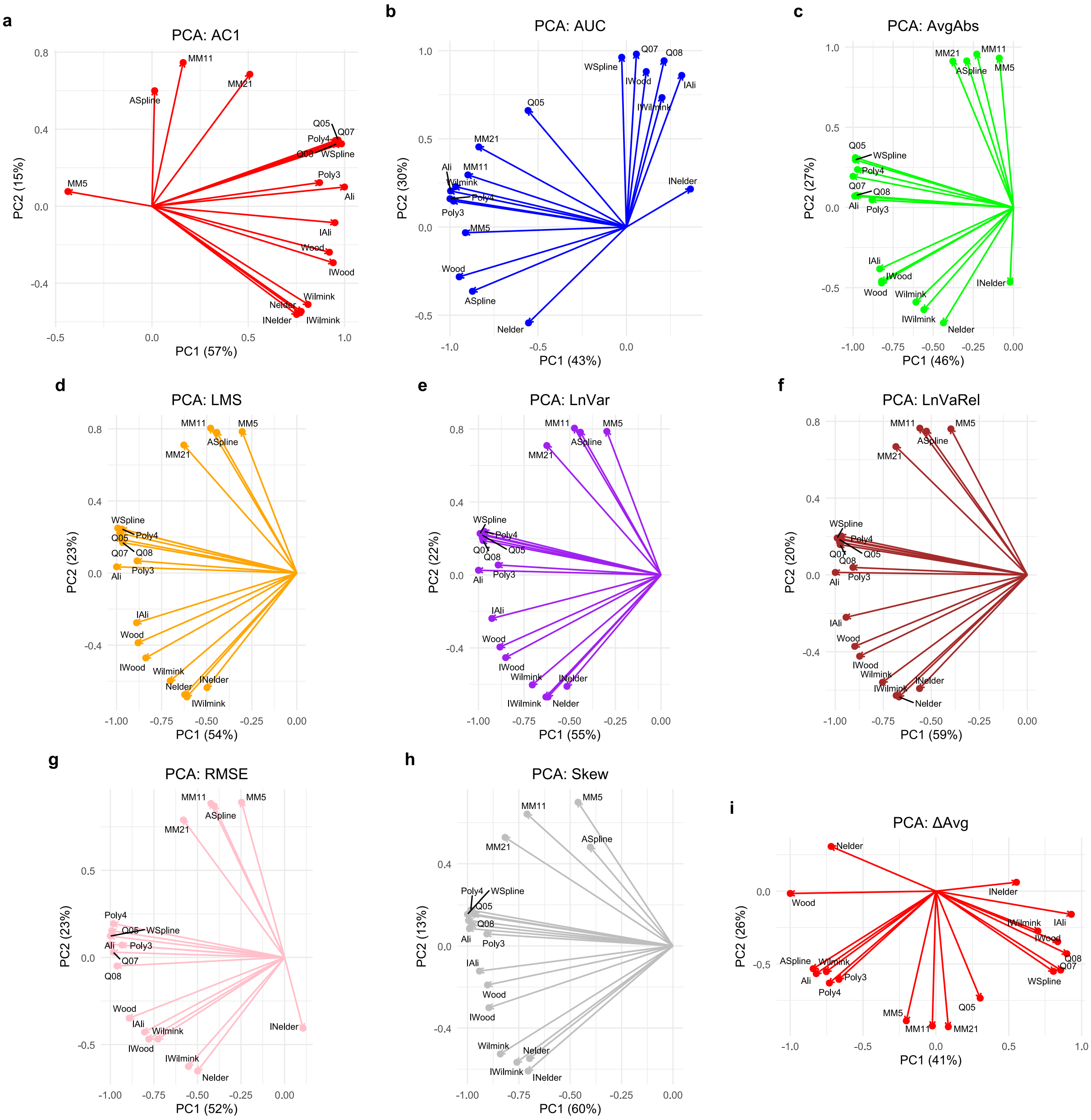


Figure 3.1: PC1 and PC2 loadings for each lactation curve model across all resilience indicators. Each plot represents the loading of the lactation curve models for a given resilience indicator. This figure provides a detailed view of how each model behaves across various resilience metrics, highlighting both consistent patterns and specific variations.

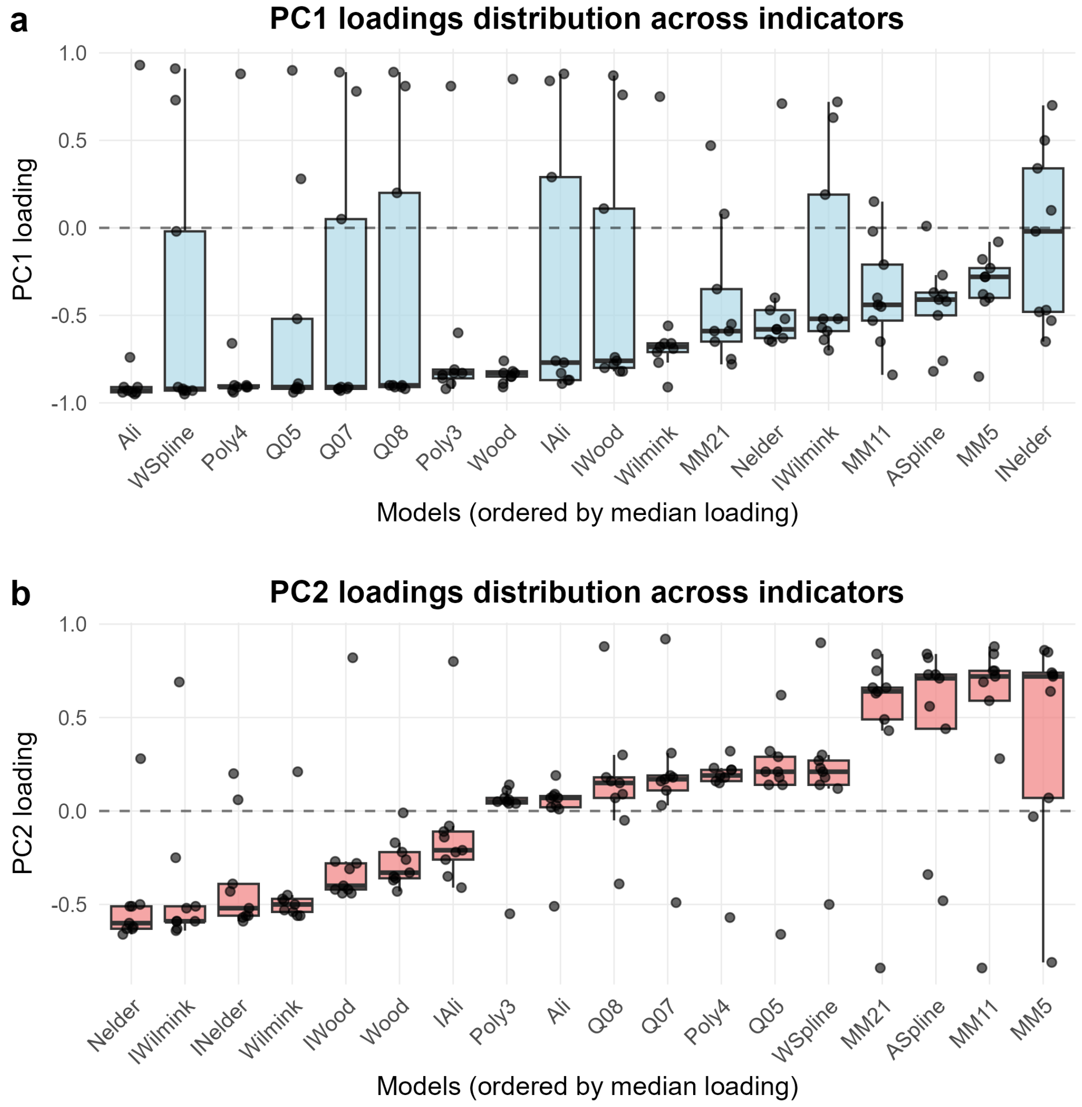


Figure 3.2: Distribution of PCA loadings across all resilience indicators. Panel (a) shows PC1 loadings and panel (b) shows PC2 loadings. Models are ordered by median loading values. Box width indicates consistency across indicators: narrow boxes represent models with consistent behavior across different resilience metrics, while wider boxes indicate models with variable behavior depending on the indicator used.

3.4 Discussion

3.4.1 Model dependency in resilience assessment

The substantial variability in cow rankings across different lactation curve models observed in this study highlights a fundamental challenge in resilience assessment. Our findings demonstrate that the choice of lactation curve model significantly influences which cows are phenotypically identified as resilient, with model pair agreements ranging from 12% to 86%. This model dependency is in line with previous studies interested to estimate genetic parameters of resilience indicators (Elgersma et al. 2018; Poppe et al. 2020). Elgersma

et al. (2018) posed attention to the fact that deviation-based indicators depend on the lactation shape, and this can be even driven by the fact that different deviations quantified with different lactation curve models can lead to different results; Poppe et al. (2020) stated that their different curve-fitting methods led to similar genetic LnVar phenotypes, but small differences were detected in the genetic parameters: Moving average and Moving median resulted in the highest heritability while quantile regression (here Q07) gave the strongest genetic correlations with health, longevity, fertility and metabolic traits.
The superior performance of quantile regression models (Q07-Q08) and their consistent ranking behavior aligns with recent methodological advances in lactation curve modelling. With a degree of similarity score of 86% ± 8%, this pair demonstrates remarkable consistency across multiple resilience indicators. Similarly, the strong performance of combinations involving WSpline (Q07-WSpline: 77% ± 10%, Q08-WSpline: 75% ± 8%) validates the weighted approach developed by Kessler et al. (2024). Our results extend these findings by demonstrating that quantile regression generates more consistent resilience rankings across different indicators.
Models' agreement across resilience indicators reveals a clear hierarchy of consistency in how different model combinations evaluate cow resilience. Analysis of all model pairs demonstrates striking variability in agreement levels, from highly consistent combinations (86% ± 8% for Q07-Q08) to extremely inconsistent ones (12% ± 5% for ASpline-IWilmink). Most notably, quantile regression models maintain the highest overall consistency, with Q07-Q08, Q07-WSpline, and Q08-WSpline achieving mean agreements of 75-86% across all indicators. These high-performing combinations show relatively low standard deviations (8-10%), indicating stable performance across different resilience metrics. In contrast, model pairs involving adaptive splines (ASpline) consistently demonstrate poor agreement (12-29%) with parametric approaches, with particularly low consistency when paired with iterative models (IWilmink, INelder). The moving median approaches (MM5, MM11, MM21) exhibit moderate but highly variable agreement patterns (13-66%), suggesting that their performance is highly dependent on both the specific indicator and the comparison model. This systematic gradient in agreement levels across model types reinforces the fundamental importance of model selection in resilience assessment. The poor performance of the ASpline model, particularly its low agreement with parametric approaches (12-18% overall similarity with models like IWilmink and INelder), reflects the fundamental differences between parametric and non-parametric smoothing methods highlighted by Codrea et al. (2011). The consistently poor performance of ASpline combinations suggests that this approach, while theoretically sound, may introduce methodological inconsistencies that significantly affect resilience rankings.

So far, other studies have applied different lactation curve models to derive resilience indicators, and the most prevalent approach has been the use of quantile polynomial curves, pioniereed by Poppe et al. (2020) and then widely adopted in subsequent studies (even for comparison) (Poppe et al. 2021a; Poppe et al. 2021b; Wang et al. 2022a; Wang et al. 2022b; Zefreh et al. 2024; Punturiero et al. 2025; Guinan et al. 2025a). Another relevant branch of lactation curves adopted as a basis for estimate the unperturbes state of an individual cow is the one of parametric models with iterated regression to remove perturbations, as proposed by Adriaens et al. (2018) and then adopted in subsequent studies (Adriaens et al. 2020; Adriaens et al. 2021; Zefreh et al. 2024). Noteworthy is the fact that the two approaches, quantile regression and iterated regression, lie on the same conceptual basis of estimating the unperturbed lactation curve by the mean of reducing the influence of lactation drops of the acutal curves.

Our results suggest that the choice of lactation curve model can lead to substantial differences in resilience assessment, with quantile regression models showing superior consistency across indicators.

3.4.2 Indicator-specific sensitivity and biological interpretation

The differential performance of resilience indicators observed in this study provides important insights into their biological relevance and practical utility. Variance-based indicators (LnVar, LnVarRel, LMS, RMSE, AvgAbs) demonstrated superior consistency across model combinations, with agreement percentages frequently exceeding 80% for similar model pairs. Even in the Q07-Q08 pair, which shows the highest overall agreement (86% ± 8%), variance-based indicators achieve high consistency (LnVar: 93%, LnVarRel: 96%, LMS: 87%, RMSE: 84%, AvgAbs: 85%). This finding supports the theoretical foundation proposed by Elgersma et al. (2018), who conceptualized resilience as the ability to maintain stable production levels despite environmental perturbations and naturally captured by variance-based metrics.

The consistently poor performance of AUC across model combinations raises important questions about its utility as a resilience indicator. Even the most consistent model pair (Q07-Q08) achieved only 75% agreement for AUC, while maintaining 85-96% agreement for variance-based indicators. More dramatically, combinations like Poly4-Q05 showed extremely low AUC agreement (41%) despite maintaining 88-92% agreement for variance-based metrics. This indicator-specific inconsistency persists across all quantile regression comparisons with Poly4, suggesting that AUC may be overly sensitive to model-specific characteristics. While AUC was found to be a good indicator by Zefreh et al. (2024) for classifying resilient animals and showing robustness to measurement frequency, our results suggest significant limitations in its cross-model reliability.

The ΔAvg indicator emerges as the most problematic and inconsistent metric across nearly all model combinations. With agreement percentages as low as 6-7% even among otherwise similar models (Poly4-Q07, Poly4-WSpline), ΔAvg demonstrates extreme sensitivity to model selection. Particularly striking is how Poly4-Q07 maintains high agreement (82-88%) for variance-based indicators while dropping to just 6% for ΔAvg. Similarly, Poly4-Q05 shows a mere 15% agreement for ΔAvg while maintaining 88-92% for variance-based indicators. This finding is further supported by the PCA results, where ΔAvg displays a unique pattern with reversed PC1 loadings compared to other indicators. While ΔAvg was proposed by Guinan et al. (2025a) to detect perturbations at the pen level and worked very well when perturbations were individuated, our analysis suggests its limited utility for individual animal general resilience assessment due to its extreme model dependency. The intermediate performance of AC1 (autocorrelation) aligns with findings from Zefreh et al. (2024), who validated the statistical properties of autocorrelation-based resilience indicators. Our results show that AC1 maintains good consistency for similar models (94% for Q07-Q08) but deteriorates more rapidly than variance-based indicators as model dissimilarity increases, suggesting its application should be limited to contexts where model selection is carefully considered.

Our analysis revealed fundamental differences between lactation curve models that aim to remove perturbations through different mathematical approaches. Iterated models (IAli, IWilmink, IWood, INelder) and quantile regression approaches (Q05, Q07, Q08) showed consistently poor agreement (24-50%), despite both aiming to estimate underlying lactation potential. Interestingly, the iterative process introduces substantial variation, with iterated models differing from their non-iterated counterparts (IAli-Ali: 56% ± 22%, IWood-Wood: 66% ± 24%, IWilmink-Wilmink: 65% ± 28%) to a greater extent than different non-iterated models differ from each other (e.g., Ali-Wood: 50% ± 7%, Wilmink-Wood: 53% ± 6%).

This finding suggests that methodological decisions regarding the removal of perturbations can have a substantial impact on the assessment of resilience, potentially leading to alterations in the biological interpretation of the resulting indicators.

3.4.3 Methodological insights from PCA analysis

The PCA results provide valuable insights into the underlying structure of model-indicator relationships. The clear separation of models along PC1 for traditional indicators (explaining 42-61% of variance) demonstrates a fundamental divide between parametric and non-parametric approaches. This division is particularly evident in the consistent negative PC1 loadings for parametric models across variance-based indicators.

The striking reversal observed for ΔAvg, where PC1 loadings range from -0.91 to 0.84 with opposite patterns compared to other indicators, highlights the unique nature of this metric. This pattern suggests that ΔAvg captures a different dimension of variation in lactation curves, possibly related to systematic rather than random deviations from expected production. The lower PC1 variance explained by ΔAvg (42%) compared to other

indicators (up to 61% for Skewness) further supports the conclusion that this indicator represents a distinct aspect of resilience that may not align with traditional ones.
The behavior of iterated models (IAli, IWilmink, IWood, INelder) in both the ranking comparisons and PCA analysis raises important questions about the iterative approach to curve fitting. While iteration aims to improve curve fit by removing perturbations, our results suggest that this process may introduce methodological artifacts that significantly affect resilience rankings. The distinct PC2 behavior of iterated models compared to their base versions indicates that iteration fundamentally changes the nature of the curve fit, potentially removing informative biological variation along with noise.

3.4.4 Methodological considerations and future directions

Based on our comprehensive analysis of model and indicator performance, we propose several methodological considerations for future resilience studies:

1. Model selection should be explicit and justified: the strong model dependency observed across all indicators emphasizes the need for researchers to explicitly justify their choice of lactation curve model. Wang et al. (2022a) proposed a framework to select the best model based on statistical criteria.

2. Variance-based indicators demonstrated superior consistency across model combinations, making them more reliable for resilience assessment. However, they have usually large and unfavorable genetic correlations with production level as compared to the other resilience indicators, which can be a practical problem for their implementation in breeding programs.

3. ΔAvg and AUC should be used with caution: the extreme model sensitivity of these indicators suggests they should not be used in isolation and should always be interpreted alongside more robust variance-based metrics.

4. Quantile regression approaches show promise: the consistent performance of Q07-Q08 and Q07-WSpline combinations suggests that quantile regression, particularly at the 0.7-0.8 levels, provides a robust foundation for resilience assessment.

5. Non-parametric smoothing approaches require further validation: the poor performance of ASpline combinations indicates that non-parametric smoothing techniques, while theoretically appealing, may introduce methodological inconsistencies that affect resilience rankings.

6. This study can not determine which model and indicator combination is the most biologically meaningful, as we just compared the consistency of the rankings across models and indicators, but we did not evaluate the biological relevance of the resulting

rankings. Do it would require a true value of resilience to be known, which is currently not possible. Future studies should aim to evaluate the biological relevance of different model and indicator combinations, potentially through longitudinal health and welfare data or experimental perturbation studies.

7. These results and discussion refer to the phenotypic assessment, but for a clearer understanding of the genetic of the traits inside the herd, it would be beneficial to compare these models based on their covariance components and genetic parameters, such as genetic correlations. However, the interpretability of genetic correlations in cows raised in a single environment is limited as shared environmental covariance acts as a confounding effect on genetic covariance.

8. The dataset used in this study may not represent conditions in other dairy farms, which could limit the generalizability of the findings. At least partially, the farm used in this study is equipped with recently built barns that facilitate heat stress mitigation. Furthermore, the farmer is attentive to the space allocated to the cows, thus providing the intrinsic best possible conditions for the cows. However, high temperature and humidity conditions are still present during the summer months, surely leading to some degree of heat stress. Future studies should aim to validate these findings across a wider range of farm conditions and management practices to ensure their applicability in diverse contexts.

3.5 Conclusions

This comprehensive study systematically evaluated the impact of lactation curve model selection on resilience indicator calculation and cow ranking in dairy cattle. Using daily milk yield data from 1,423 lactations collected via AMS, we implemented eighteen different lactation curve models combined with nine resilience indicators to characterize their consistency and reliability. Our findings demonstrate that the choice of lactation curve model profoundly influences resilience assessment, with model pair agreements ranging from 12% to 86%, highlighting a fundamental challenge in the standardization of resilience phenotyping. Quantile regression models, particularly Q07-Q08 (86% ± 8% agreement), emerged as the most consistent approaches, followed closely by combinations involving weighted penalized splines (WSpline). These high-performing models demonstrated remarkable stability across variance-based indicators (LnVar, LnVarRel, LMS, RMSE, AvgAbs), frequently achieving agreement levels exceeding 80-90%. In stark contrast, model pairs involving adaptive spline interpolation (ASpline) consistently showed poor agreement with parametric approaches (12-18%), while moving median models exhibited highly variable performance depending on the specific indicator and comparison model. Critically, our results expose significant differences in indicator behavior across model combinations.

Variance-based indicators demonstrated superior consistency and cross-model reliability, supporting their theoretical foundation as robust measures of production stability. However, the area under the curve (AUC) and particularly ΔAvg showed extreme model sensitivity, with agreement percentages as low as 6-7% even among otherwise similar model pairs. This finding raises important questions about the universal applicability of these indicators and suggests they should be interpreted with caution, particularly when comparing studies using different lactation curve modelling approaches.

They also have important implications for both research and practical implementation of resilience assessment in dairy cattle breeding. The strong model dependency we observed suggests that standardization of lactation curve modelling approaches is crucial for comparing resilience estimates across studies and breeding programs. Our results support the preferential use of quantile regression models (particularly at the 0.7-0.8 quantile levels) combined with variance-based resilience indicators for robust and reproducible resilience assessment.

3.6 Supplementary materials

Table S1. Each cell represents the percentage of cows ranked within the same tails of the distribution when comparing two rankings of the same indicator, each calculated using two different model. The table is sorted based on the degree of similarity, expressed as the mean ± standard deviation of the values in each row.
Acronym expansions are provided in Table 3.1

Model	AC1	AUC	AvgAbs	LMS	LnVar	LnVa	RelRMSE	Skew	ΔAvg	Degree of similarity
Q07-Q08	94	75	85	87	93	96	84	89	71	86 ± 8
Q07-WSpline	81	67	78	83	81	83	86	75	56	77 ± 10
Q08-WSpline	80	68	77	79	82	85	77	72	59	75 ± 8
Q05-Q07	89	35	92	86	91	92	84	74	28	75 ± 25
Poly4-Q05	84	41	88	88	92	91	90	75	15	74 ± 27
Q05-Q08	85	27	81	78	87	92	77	68	18	68 ± 27
Q05-WSpline	81	32	76	82	81	84	81	70	24	68 ± 23
Poly4-Q07	81	26	85	82	87	88	83	72	6	68 ± 30
Poly4-WSpline	81	28	77	81	85	84	83	77	7	67 ± 29
MM11-MM21	50	72	74	64	63	69	70	67	69	66 ± 7
IWood-Wood	81	21	76	81	82	80	80	61	28	66 ± 24
Poly4-Q08	81	24	73	79	85	89	74	67	10	65 ± 28
Ali-Poly4	66	80	56	65	68	66	69	60	51	65 ± 8
IWilmink-Wilmink	89	21	77	76	82	80	75	69	12	65 ± 28
INelder-IWilmink	68	63	60	63	61	61	67	68	59	63 ± 4
Ali-Poly3	55	69	61	61	61	69	68	66	35	61 ± 11
Ali-WSpline	71	32	57	72	72	72	78	67	9	59 ± 23
INelder-Nelder	76	16	66	69	70	73	74	51	20	57 ± 23
Poly3-Poly4	51	79	49	48	49	59	60	61	56	57 ± 10
Ali-Q05	68	42	57	67	67	69	69	59	12	57 ± 19
Ali-IAli	71	19	62	63	68	74	59	73	17	56 ± 22
MM11-MM5	15	61	74	54	54	60	70	53	64	56 ± 17
Ali-Q07	69	25	60	64	65	67	70	62	10	55 ± 22
IWilmink-IWood	62	61	48	46	49	56	53	51	56	54 ± 6
Ali-Q08	66	24	57	63	66	69	65	59	11	53 ± 21
Wilmink-Wood	57	55	51	52	52	56	60	53	39	53 ± 6
IAli-IWood	61	44	46	50	53	57	52	55	41	51 ± 6
IAli-WSpline	53	35	52	51	55	59	54	63	34	51 ± 10
ASpline-MM11	25	32	70	66	65	69	73	36	19	51 ± 22
IAli-Q08	51	37	51	47	54	57	54	56	41	50 ± 7
IWilmink-Nelder	59	22	57	57	62	65	60	43	23	50 ± 17
Nelder-Wilmink	56	25	54	59	59	65	60	57	13	50 ± 18
INelder-Wilmink	66	22	53	57	60	56	60	62	12	50 ± 19
Ali-Wood	53	49	51	51	51	50	56	54	32	50 ± 7
IWood-Wilmink	63	27	51	55	55	64	66	50	15	50 ± 17
Poly3-Wood	54	57	49	47	46	57	53	52	28	49 ± 9
Poly3-WSpline	53	26	47	55	59	64	65	65	8	49 ± 20
IAli-Q07	53	38	49	47	53	56	51	55	37	49 ± 7
INelder-IWood	53	59	43	46	45	48	53	45	44	48 ± 5

(continued)

Model	AC1	AUC	AvgAbs	LMS	LnVar	LnVa	RelRMSE	Skew	ΔAvg	Degree of similarity
Poly3-Q05	50	41	49	49	48	59	60	54	14	47 ± 14
IAli-Wood	56	19	46	49	53	56	52	44	31	45 ± 12
Poly3-Q07	50	23	49	49	50	56	63	53	9	45 ± 17
ASpline-MM5	6	48	61	47	48	53	65	44	28	44 ± 18
Poly3-Q08	50	22	47	49	50	59	60	51	11	44 ± 17
Ali-Wilmink	41	78	35	36	35	40	42	48	39	44 ± 13
IAli-Q05	52	20	48	49	53	55	49	48	18	44 ± 14
Poly4-Wood	43	54	38	43	42	45	46	52	27	43 ± 8
Ali-IWood	54	23	54	46	49	52	49	53	9	43 ± 16
Nelder-Wood	46	37	45	45	46	44	50	42	33	43 ± 5
IAli-Poly4	50	18	47	46	52	55	50	52	13	43 ± 16
IAli-Poly3	49	16	42	43	46	59	52	56	17	42 ± 16
IWood-Nelder	47	20	45	51	50	51	57	32	25	42 ± 13
ASpline-MM21	17	31	64	50	51	59	61	25	15	41 ± 19
IWood-WSpline	47	28	40	41	45	46	44	52	28	41 ± 8
IAli-INelder	45	38	48	40	35	38	53	37	36	41 ± 6
Wood-WSpline	47	23	38	45	46	49	47	47	25	41 ± 10
MM21-MM5	6	49	57	35	36	43	51	36	54	41 ± 15
Poly4-Wilmink	28	75	32	33	34	36	37	39	46	40 ± 14
Q08-Wood	43	28	36	42	41	45	50	41	32	40 ± 7
IAli-IWilmink	47	35	40	35	38	42	40	44	35	40 ± 4
Q07-Wood	43	26	36	41	40	44	49	45	32	40 ± 7
IWood-Q07	44	28	40	37	37	45	44	47	30	39 ± 7
Q05-Wood	44	31	37	41	42	43	49	46	15	39 ± 10
IWood-Q08	43	25	40	37	37	45	44	47	28	38 ± 8
IWood-Poly3	46	22	42	39	40	50	46	50	9	38 ± 14
IWilmink-Wood	55	15	44	40	43	45	45	38	17	38 ± 13
MM21-Q05	30	50	21	40	41	42	36	40	39	38 ± 8
IAli-Wilmink	46	21	36	41	41	47	45	47	15	38 ± 12
IWood-Q05	44	26	41	33	37	44	46	40	20	37 ± 9
IAli-Nelder	38	20	42	43	40	45	48	32	23	37 ± 10
MM21-Poly4	32	47	22	41	42	46	37	44	19	37 ± 10
Poly3-Wilmink	28	70	26	25	24	34	34	39	39	35 ± 14
IWood-Poly4	41	23	40	36	39	45	43	42	9	35 ± 12
MM21-WSpline	33	33	19	39	43	40	32	49	23	35 ± 9
INelder-Wood	45	15	38	39	41	38	45	35	12	34 ± 12
Ali-MM21	30	48	20	35	37	32	33	46	16	33 ± 10
Q05-Wilmink	28	41	32	36	34	35	39	36	10	32 ± 9
Ali-Nelder	35	26	31	33	33	34	40	37	19	32 ± 6
MM21-Q07	30	26	20	36	38	42	31	45	18	32 ± 9
Q07-Wilmink	28	26	33	35	31	36	38	40	12	31 ± 8
Wilmink-WSpline	35	28	28	36	32	37	38	38	7	31 ± 10
Q08-Wilmink	27	23	30	33	31	36	38	41	15	30 ± 8
MM11-Poly4	22	49	22	31	30	39	26	32	20	30 ± 9
MM11-Q05	19	41	21	31	31	36	27	28	33	30 ± 7
Ali-INelder	39	21	31	28	28	30	37	31	18	29 ± 7
Ali-IWilmink	40	17	33	28	31	31	33	37	10	29 ± 10

(continued)

Model	AC1	AUC	AvgAbs	LMS	LnVar	LnVa	RelRMSE	Skew	ΔAvg	Degree of similarity
MM21-Q08	30	22	16	33	37	43	27	42	10	29 ± 11
Ali-ASpline	6	47	21	22	22	27	22	23	69	29 ± 18
Nelder-WSpline	32	20	30	27	28	32	36	26	24	28 ± 5
IWilmink-Q07	28	25	28	26	28	30	27	37	25	28 ± 4
Ali-MM11	19	52	21	26	26	28	25	37	20	28 ± 10
Nelder-Q05	26	23	28	30	30	32	35	27	22	28 ± 4
Nelder-Q08	26	24	27	28	28	33	34	27	23	28 ± 4
Nelder-Q07	25	22	28	27	28	33	34	27	25	28 ± 4
IWilmink-WSpline	34	20	27	27	28	27	28	35	23	28 ± 5
Nelder-Poly4	28	25	31	27	28	32	35	28	14	28 ± 6
ASpline-Poly4	5	47	23	25	25	31	24	24	43	27 ± 12
IWilmink-Q08	27	23	26	25	28	28	28	38	23	27 ± 4
MM11-WSpline	21	30	19	28	28	36	23	34	20	27 ± 6
INelder-WSpline	33	16	28	27	28	28	33	28	17	26 ± 6
INelder-Q07	27	20	27	27	26	28	32	26	21	26 ± 4
IWilmink-Q05	27	19	28	26	31	28	27	30	18	26 ± 5
INelder-Poly4	28	25	28	25	28	27	32	26	14	26 ± 5
Nelder-Poly3	25	27	23	24	23	32	33	30	15	26 ± 6
IWilmink-Poly4	27	24	28	25	31	28	26	32	9	26 ± 7
MM21-Poly3	20	43	15	25	24	25	27	34	15	25 ± 9
INelder-Q08	26	15	24	26	27	28	33	28	20	25 ± 5
MM11-Q07	19	23	20	26	28	36	21	31	17	25 ± 6
INelder-Q05	26	20	27	25	28	27	31	23	13	24 ± 5
IAli-MM21	23	17	20	27	27	24	28	37	13	24 ± 7
MM5-Poly4	20	43	18	20	19	25	24	18	23	23 ± 8
MM11-Poly3	16	46	16	21	20	26	20	27	15	23 ± 10
Ali-MM5	22	46	18	16	15	17	22	23	27	23 ± 9
ASpline-Poly3	9	42	16	20	20	27	18	23	30	23 ± 9
INelder-Poly3	27	19	23	21	21	25	31	28	9	23 ± 6
MM5-Q05	20	28	18	19	21	21	24	18	31	22 ± 5
MM11-Q08	18	22	15	23	26	36	20	30	9	22 ± 8
IWilmink-Poly3	28	20	23	15	18	25	24	34	8	22 ± 8
MM21-Wood	15	32	17	20	20	18	25	34	13	22 ± 7
ASpline-Wood	7	47	18	16	16	17	18	24	31	22 ± 12
ASpline-Q05	6	25	20	23	26	30	26	21	10	21 ± 8
IAli-MM11	21	19	20	21	18	24	21	28	13	21 ± 4
IWood-MM21	15	22	16	19	22	16	25	31	14	20 ± 6
MM11-Wood	14	35	18	17	17	19	17	30	13	20 ± 7
MM21-Wilmink	14	46	13	14	15	12	24	28	8	19 ± 12
MM5-Poly3	18	39	15	15	15	16	18	17	19	19 ± 8
MM21-Nelder	13	22	18	18	18	13	24	24	21	19 ± 4
MM5-WSpline	20	17	16	18	17	23	22	20	15	19 ± 3
MM11-Nelder	13	28	19	15	14	13	20	27	19	19 ± 6
MM11-Wilmink	13	46	13	10	10	12	19	31	14	19 ± 12
ASpline-Q07	5	17	19	21	24	28	21	20	11	18 ± 7
ASpline-WSpline	6	16	18	21	24	30	22	21	8	18 ± 8
MM5-Wood	19	34	17	15	14	14	18	20	14	18 ± 6

(continued)

Model	AC1	AUC	AvgAbs	LMS	LnVar	LnVaRel	RMSE	Skew	ΔAvg	Degree of similarity
ASpline-Q08	6	18	15	19	23	28	19	21	14	18 ± 6
ASpline-Wilmink	6	40	15	8	8	8	17	26	35	18 ± 13
IWood-MM11	18	21	18	13	17	17	19	25	14	18 ± 4
ASpline-IAli	7	14	20	22	19	22	21	17	15	17 ± 5
MM5-Q07	18	11	17	18	19	21	19	19	14	17 ± 3
INelder-MM21	13	23	20	16	16	12	22	21	10	17 ± 5
MM5-Wilmink	18	37	12	7	7	9	19	24	18	17 ± 10
MM5-Nelder	17	24	16	11	11	12	19	22	18	17 ± 5
IAli-MM5	22	15	20	13	14	17	21	14	13	17 ± 4
INelder-MM11	17	24	18	12	11	13	17	21	14	16 ± 4
IWilmink-MM21	14	18	19	14	16	13	20	20	9	16 ± 4
ASpline-Nelder	9	25	18	13	12	10	16	20	20	16 ± 5
MM5-Q08	16	12	13	16	18	21	18	17	9	16 ± 4
ASpline-IWood	7	17	20	14	16	17	18	18	9	15 ± 4
IWood-MM5	18	18	17	9	12	13	20	17	11	15 ± 4
IWilmink-MM11	13	19	20	11	11	13	16	22	8	15 ± 5
INelder-MM5	20	19	16	8	7	11	16	17	14	14 ± 5
IWilmink-MM5	17	15	16	9	9	10	16	20	9	13 ± 4
ASpline-INelder	9	17	19	10	9	8	16	16	17	13 ± 4
ASpline-IWilmink	8	16	19	8	9	8	15	18	11	12 ± 5

Table S2. Relative (expl) and cumulative relative (cum) variance explained by the Principal Components (PC) of lactation curve models across all the resilience indicators.

PC	AC1		AUC		AvgAbs		LMS		LnVar		LnVaRel		RMSE		Skew		ΔAvg	
	Expl	Cum	Expl	Cum	Expl	Cum	Expl	Cum	Expl	Cum	Expl	Cum	Expl	Cum	Expl	Cum	Expl	Cum
1	0.57	0.57	0.43	0.43	0.46	0.46	0.55	0.55	0.56	0.56	0.60	0.60	0.53	0.53	0.61	0.61	0.42	0.42
2	0.16	0.73	0.30	0.73	0.27	0.73	0.23	0.78	0.23	0.79	0.21	0.81	0.24	0.77	0.14	0.75	0.27	0.69
3	0.10	0.83	0.07	0.80	0.10	0.83	0.10	0.88	0.10	0.89	0.09	0.90	0.09	0.86	0.08	0.83	0.10	0.79
4	0.05	0.88	0.05	0.85	0.05	0.88	0.04	0.92	0.04	0.93	0.03	0.93	0.05	0.91	0.04	0.87	0.04	0.83
5	0.04	0.92	0.04	0.89	0.03	0.91	0.02	0.94	0.02	0.95	0.02	0.95	0.02	0.93	0.03	0.90	0.04	0.87
6	0.02	0.94	0.03	0.92	0.02	0.93	0.02	0.96	0.01	0.96	0.01	0.96	0.02	0.95	0.02	0.92	0.03	0.90
7	0.02	0.96	0.02	0.94	0.02	0.95	0.01	0.97	0.01	0.97	0.01	0.97	0.01	0.96	0.02	0.94	0.02	0.92
8	0.02	0.98	0.01	0.95	0.01	0.96	0.01	0.98	0.01	0.98	0.01	0.98	0.01	0.97	0.01	0.95	0.02	0.94
9	0.01	0.99	0.01	0.96	0.01	0.97	0.01	0.99	0.01	0.99	0.01	0.99	0.01	0.98	0.01	0.96	0.02	0.96
10	0.01	1.00	0.01	0.97	0.01	0.98	0.01	1.00	0.01	1.00	0.01	1.00	0.00	0.98	0.01	0.97	0.01	0.97
11	0.00	1.00	0.01	0.98	0.00	0.98	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.98	0.01	0.98	0.01	0.98
12	0.00	1.00	0.00	0.98	0.00	0.98	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.98	0.01	0.99	0.01	0.99
13	0.00	1.00	0.00	0.98	0.00	0.98	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.98	0.00	0.99	0.00	0.99
14	0.00	1.00	0.00	0.98	0.00	0.98	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.98	0.00	0.99	0.00	0.99
15	0.00	1.00	0.00	0.98	0.00	0.98	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.98	0.00	0.99	0.00	0.99
16	0.00	1.00	0.00	0.98	0.00	0.98	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.98	0.00	0.99	0.00	0.99
17	0.00	1.00	0.00	0.98	0.00	0.98	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.98	0.00	0.99	0.00	0.99
18	0.00	1.00	0.00	0.98	0.00	0.98	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.98	0.00	0.99	0.00	0.99

Table S3. Loadings of the first two principal components (PC1 and PC2) for each lactation curve model within each resilience indicator. These loadings indicate the contribution of each model to the principal components derived from the PCA analysis across all resilience indicators, providing insights into how different models relate to the underlying dimensions of resilience in dairy cattle.

	AC1		AUC		AvgAbs		LMS		LnVar		LnVaRel		RMSE		Skew		ΔAvg	
Model	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Ali	0.93	0.09	-0.93	0.19	-0.91	0.07	-0.94	0.03	-0.94	0.02	-0.94	0.01	-0.95	0.07	-0.91	0.08	-0.74	-0.51
ASpline	0.01	0.56	-0.82	-0.34	-0.27	0.84	-0.42	0.73	-0.41	0.73	-0.50	0.71	-0.38	0.82	-0.37	0.44	-0.76	-0.48
IAli	0.88	-0.08	0.29	0.80	-0.77	-0.35	-0.83	-0.26	-0.87	-0.22	-0.89	-0.21	-0.76	-0.41	-0.87	-0.11	0.84	-0.14
INelder	0.70	-0.52	0.34	0.20	-0.02	-0.43	-0.47	-0.59	-0.48	-0.57	-0.53	-0.56	0.10	-0.39	-0.65	-0.56	0.50	0.06
IWilmink	0.72	-0.51	0.19	0.69	-0.52	-0.59	-0.57	-0.64	-0.59	-0.63	-0.64	-0.59	-0.52	-0.59	-0.70	-0.52	0.63	-0.25
IWood	0.87	-0.27	0.11	0.82	-0.76	-0.42	-0.78	-0.44	-0.80	-0.42	-0.82	-0.40	-0.74	-0.44	-0.82	-0.28	0.76	-0.31
MM11	0.15	0.69	-0.84	0.28	-0.21	0.88	-0.45	0.75	-0.44	0.75	-0.53	0.72	-0.40	0.84	-0.65	0.59	-0.02	-0.84
MM21	0.47	0.64	-0.78	0.43	-0.35	0.84	-0.59	0.66	-0.59	0.66	-0.65	0.63	-0.55	0.75	-0.75	0.49	0.08	-0.84
MM5	-0.40	0.07	-0.85	-0.03	-0.08	0.86	-0.28	0.73	-0.28	0.74	-0.38	0.72	-0.23	0.85	-0.42	0.64	-0.18	-0.81
Nelder	0.71	-0.51	-0.52	-0.51	-0.40	-0.66	-0.58	-0.63	-0.58	-0.63	-0.63	-0.60	-0.47	-0.62	-0.64	-0.50	-0.65	0.28
Poly3	0.81	0.11	-0.92	0.14	-0.81	0.04	-0.83	0.06	-0.84	0.05	-0.86	0.04	-0.89	0.07	-0.83	0.05	-0.60	-0.55
Poly4	0.88	0.32	-0.94	0.15	-0.90	0.22	-0.91	0.23	-0.91	0.22	-0.91	0.19	-0.93	0.18	-0.90	0.16	-0.66	-0.57
Q05	0.90	0.32	-0.52	0.62	-0.91	0.29	-0.92	0.21	-0.92	0.21	-0.92	0.18	-0.94	0.14	-0.89	0.14	0.28	-0.66
Q07	0.89	0.31	0.05	0.92	-0.92	0.18	-0.91	0.17	-0.92	0.19	-0.92	0.16	-0.93	0.03	-0.91	0.11	0.78	-0.49
Q08	0.89	0.30	0.20	0.88	-0.90	0.07	-0.90	0.16	-0.91	0.18	-0.92	0.15	-0.91	-0.05	-0.90	0.09	0.81	-0.39
Wilmink	0.75	-0.47	-0.91	0.21	-0.56	-0.54	-0.66	-0.56	-0.66	-0.56	-0.71	-0.53	-0.69	-0.45	-0.77	-0.48	-0.68	-0.50
Wood	0.85	-0.22	-0.89	-0.26	-0.76	-0.43	-0.82	-0.36	-0.83	-0.37	-0.85	-0.35	-0.85	-0.33	-0.83	-0.17	-0.91	-0.01
WSpline	0.91	0.30	-0.02	0.90	-0.91	0.27	-0.93	0.23	-0.93	0.21	-0.93	0.18	-0.95	0.12	-0.92	0.14	0.73	-0.50

Chapter 4

Simulation study: detecting true resilience with proxies derived from different lactation curve models

Andrea Delledonne[ab], Christian Maltecca[b], Francesco Tiezzi[c], Chiara Punturiero[a], Maria Giuseppina Strillacci[a], Alessandro Bagnato[a]

[a]*Department of Veterinary Medicine and Animal Sciences, University of Milan LA STATALE, Lodi, Italy;* [b]*Department of Animal Science, North Carolina State University, Raleigh, NC, USA;* [c]*Department of Science and Agricultural, Food, Environmental and Forest Technologies (DAGRI), University of Firenze, Florence, Italy*

This manuscript is the outcome of my research period as visiting Ph.D. student at North Carolina State University and it will be submitted soon, in conjunction with the characterization study (Chapter 3).

4.1 Introduction

This study is strictly connected to the previous one (Chapter 3), where different lactation curve models and resilience indicators were characterized. The current chapter aims to evaluate how effectively these models and indicators capture true resilience in dairy cows through a simulation study.

Proxies of resilience (also known as resilience indicators) are based on a simple concept: measuring and aggregating the deviation between an estimated unperturbed state and the actual state. A key underlying assumption is that the actual state can be observed precisely, whereas the unperturbed state cannot. In the previous study (Chapter 3), various lactation curve models were implemented and compared for estimating the unperturbed state, including parametric and nonparametric models. Subsequently, different resilience indicators were calculated based on the deviations from the estimated unperturbed state. Results from Chapter 3 highlighted that from at least the phenotypic standpoint, different combinations of lactation curve models and resilience indicators can lead to varying assessments of resilience.

While the heritability of variance-based resilience indicators has been established (Poppe et al. 2020; Chen et al. 2023; Kessler et al. 2024; Guinan et al. 2025a), it is unclear to what extent these indicators truly capture the resilience of cows. This is due to the fact that the unperturbed state is not directly observable, and because resilience indicators are influenced by various factors. This study aims to address this issue by answering two key research questions: i) To what extent can resilience indicators capture the resilience of cows? ii) Which model-indicator combination is most effective for capturing resilience?

4.2 Materials and methods

A small group of resilient cows, validated through health records and rumination patterns, served as the basis for simulating resilience-modulated events. The simulation generated breeding values for production and breeding values for three resilience components (intensity, duration, recovery) with a predefined genetic correlation structure. Four types of resilience-modulated events (mastitis, subclinical mastitis, heat stress, lameness) were simulated, each with specific probabilities and associated phenotypic ranges for intensity, duration, and recovery. These represented the biological responses modulated by the resilience BV components. Additionally, two resilience-unmodulated environmental events (feed shortage and facility breakdown) were included to introduce environmental variation, along with changes in autocorrelation patterns. These elements were incorporated to mimic real-world complexities, acknowledging that certain factors can obscure the genetic signal of resilience when measured by the studied resilience indicators.

4.2.1 Data collection and preprocessing

The data used in this study were exactly the same as in the previous Chapter 3. Data were preprocessed to have a dataset with lactations equally represented and composed of 514,131 observations, for a total of 819 cows and 1,423 lactations.

4.2.2 Simulation study design

We began by establishing a “ground truth” set of resilient cows. To detect and define these resilient individuals, we relied on multiple health-related indicators, including medical treatments received, somatic cell count (SCC) levels, and rumination patterns. Medical treatments were recorded by farm staff, SCC were recorded by AMS, and cows were provided with neck collars equipped with sensors that recorded data on rumination time. Specifically, the 200 most resilient cows as indicated by the LnVar indicator obtained from the Ali model were selected for inspection. Cows exhibiting gaps greater than one day in the continuous series of DIM were excluded to avoid including those potentially affected by disorders that could lead to milk withdrawal. Subsequently, SCC and treatment records were observed to identify cases of mastitis, subclinical mastitis and other medical issues, resulting in the removal of additional cows. Daily rumination time (minutes) was inspected to exclude cows exhibiting significant decreases indicative of poor health; cows exhibiting decreases exceeding 20% from their rolling 7-day baseline for more than 2 consecutive days were excluded. Changes in rumination time are reliable indicators of various health conditions, including mastitis, lameness, ketosis, and abomasal displacement (Llonch et al. 2018), making rumination patterns valuable proxies for identifying cows in compromised health states. These factors served as proxies for resilience, enabling us to identify cows exhibiting phenotypes consistent with resilience. Establishing this reliable subset of 137 resilient cows was relevant, as they represent the ground truth and provide the foundation for our subsequent simulation study.

Consequently, upon a cohort of 137 resilient individuals, we systematically approach the simulation by decomposing observed daily milk yield into genetic and environmental components, generating resilience breeding values with controlled genetic architecture, and simulating resilience-modulated events for 150 repetitions. Next sections detail each part of the simulation framework.

Lactation curve decomposition

Individual lactation curves were decomposed into genetic and environmental components using a linear mixed-effects model:

$$y_{ijk} = \mu + \delta_j + \lambda_k + u_{0i} + u_{1i} + \varepsilon_{ijk} \tag{4.1}$$

where y_{ijk} is the milk production for cow i on day j in lactation k, μ is the population mean, δ_j and λ_k are the fixed effects for DIM and lactation, u_{0i} is the individual random intercept for cow i, u_{1i} is the individual random slope for cow i and ε_{ijk} is the residual error.
For each cow, the genetic component was extracted:

$$G_{ij} = \hat{u_{0i}} + \hat{u_{1i}}. \tag{4.2}$$

Production breeding values ($BV_{production,i}$) were then obtained as the average genetic component across the entire lactation for each cow (i) (n equals to the DIM at the end of the lactation):

$$BV_{production,i} = \frac{1}{n}\sum_{j=1}^{n} G_{ij} \tag{4.3}$$

These breeding values were subsequently standardized to have mean = 0 and standard deviation = 1.

Resilience Breeding Value generation

The simulation implements a multivariate approach (described later) that generates four correlated breeding values for each cow:

1. Production breeding value ($BV_{production}$)
2. Intensity breeding value ($BV_{intensity}$); resistance to production drop
3. Duration breeding value ($BV_{duration}$); ability to shorten event duration
4. Recovery breeding value ($BV_{recovery}$); capacity for quick recovery

The genetic correlation matrix was defined as following:

$$\mathbf{R}_G = \left[\begin{array}{c|ccc} 1.00 & -0.30 & -0.30 & -0.30 \\ \hline -0.30 & 1.00 & 0.90 & 0.90 \\ -0.30 & 0.90 & 1.00 & 0.90 \\ -0.30 & 0.90 & 0.90 & 1.00 \end{array}\right] \tag{4.4}$$

$$\mathbf{R}_G = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{bmatrix} \tag{4.5}$$

where $\mathbf{R}_{11}$ represents the production variance, $\mathbf{R}_{12} = \mathbf{R}'_{21}$ represent the covariance between production and resilience components, and $\mathbf{R}_{22}$ represents the variance-covariance matrix of the resilience components.

This structure reflects a moderate negative correlation between production and resilience components ($r_G = -0.30$, derived from genetic correlations between milk production and resilience indicators (Maskal et al. 2024)), and strong positive correlations among resilience components, due to the assumption of shared underlying biological mechanisms that make a cow more or less resilient for all the components ($r_G = 0.90$).

Given that production breeding values were retrieved from actual data, resilience breeding values were generated using the conditional multivariate normal distribution:

$$BV_{resilience_i} | BV_{production_i} \sim \mathcal{N}(\mu_{conditional_i}, \Sigma_{conditional_i}) \tag{4.6}$$

where:

$$\boldsymbol{\mu}_{conditional_i} = \mathbf{R}_{21} \times BVstd_{production_i} \tag{4.7}$$

$$\boldsymbol{\Sigma}_{conditional_i} = \mathbf{R}_{22} - \mathbf{R}_{\mathbf{21}} \times \mathbf{R}_{11}^{-1} \times \mathbf{R}_{12} = \mathbf{R}_{22} - \mathbf{R}_{\mathbf{21}} \times \mathbf{R}_{12} \tag{4.8}$$

This approach ensures that resilience breeding values are properly correlated with production breeding values and the genetic correlation structure is exactly preserved.

Event-specific phenotype generation

For each component (biological response) at each round of simulation, the phenotypic value was calculated as:

$$\text{Phenotype}_{ij} = \mu_{ij} + \sqrt{h^2} \times \frac{\Delta_{ij}}{4} \times BV^{std}_{component,i} + \sqrt{1-h^2} \times \frac{\Delta_{i,j}}{4} \times \varepsilon \tag{4.9}$$

where i is the biological response – intensity, duration, recovery speed –, j is the randomly simulated event type (see following section), μ_{ij} is the midpoint of the phenotypic range for component i and event j, Δ_{ij} is the phenotypic range (max - min) for component i and event j, $BV^{std}_{component,i}$ is the standardized breeding value for component i and $BV^{std}_{component,i}$ is the standardized breeding value for component i and ε is a random variable $\mathcal{N}$ (0, 1). This formulation ensures constant and proportional genetic contribution across all event types ($\sqrt{h^2} \times BV^{std}$) and event-specific scaling reflecting different baseline severities ($\frac{\Delta_{ij}}{4}$).

For each resilience component, genetic variance was derived from phenotypic ranges reported in the literature (Gröhn et al. 2004; Rajala-Schultz et al. 1999; Lescourret et al. 1994) using the approximation

$$\sigma_a^2 = h^2 \times \left(\frac{\Delta_{phenotype}}{4}\right)^2 \tag{4.10}$$

where h^2 is equal to 0.15 for all resilience components (intensity, duration, and recovery), and $\Delta_{phenotype}$ is the range (max - min) for each component-event. This approach was selected because the genetic variance of resilience could not be obtained directly from the genetic variances of published resilience indicator parameters. Those indicators are not direct measures of resilience, but rather proxies derived from deviations in milk yield. Therefore, the phenotypic ranges of biological responses to specific events were used to derive genetic variances, even if this is a big approximation.

Resilience event simulation

Four resilience-modulated event types were simulated, each with specific base probabilities and severity ranges (Gröhn et al. 2004; Rajala-Schultz et al. 1999; Lescourret et al. 1994):

1. Mastitis (base probability = 0.15)
    - Intensity range: 18-55% production loss
    - Duration range: 7-35 days
    - Recovery range: 7-28 days
2. Subclinical Mastitis (base probability = 0.25)
    - Intensity range: 1-10% production loss
    - Duration range: 14-42 days
    - Recovery range: 7-21 days
3. Heat Stress (base probability = 0.20)
    - Intensity range: 10-25% production loss
    - Duration range: 3-14 days
    - Recovery range: 3-10 days
4. Lameness (base probability = 0.10)
    - Intensity range: 5-20% production loss
    - Duration range: 7-28 days
    - Recovery range: 7-14 days

Additionally, two environmental (non-genetic) events were included, events that are not modulated by resilience breeding values and serve to introduce additional environmental variation:

1. Feed shortage (probability = 0.90)
    - Intensity range: 5-15% production loss
    - Duration range: 7-21 days

- Recovery range: 7-14 days

2. Facility breakdown (probability = 0.90)
    - Intensity range: 10-30% production loss
    - Duration range: 3-14 days
    - Recovery range: 3-10 days

Events were applied multiplicatively to the reconstructed lactation curves during the randomly selected event period:

$$P_{event,t} = P_{reconstructed,t} \times (1 + \text{Intensity}_i) \tag{4.11}$$

for $t \in [\text{start}, \text{start} + \text{duration}]$, where Intensity_i is the individual-specific intensity (expressed as negative fraction, e.g., -0.30 for 30% loss).

During the recovery period, a linear recovery multiplier was applied, ensuring a gradual return to the baseline curve over the recovery period:

$$\text{Recovery\_multiplier}_t = \text{Intensity}_i \times \left(1 - \frac{t - \text{end\_event}}{\text{Recovery\_days}_i}\right), \tag{4.12}$$

$$P_{event,t} = P_{reconstructed,t} \times (1 + \text{Recovery\_multiplier}_t). \tag{4.13}$$

4.2.3 Lactation curve models and resilience indicators

Seventeen lactation curve models were implemented to determine the unperturbed state and calculate the deviations from the simulated state. These models are the same as those used in Chapter 3 and are listed in Table 4.1, along with the nine resilience indicators subsequently calculated, which are also the same. This means that for each simulated lactation (a total of 150 repetitions for each of the 137 cows), a total of 162 (18 models × 9 indicators) resilience indicators values were computed.

Table 4.1: Lactation curve models and resilience indicators included in this study.

Resilience indicator	**Parametric model**	**Non-parametric model**
AC1	Wilmink	MM5
AbsAvg	IWilmink	MM11
AUC	Wood	MM21
ΔAvg	IWood	ASpline
LnVar	Ali	WSpline
LnVaRel	IAli	
LMS	Nelder	
RMSE	INelder	
Skew	Poly3	
	Poly4	
	Q05	
	Q07	
	Q08	

4.2.4 Performance evaluation

The performance of each resilience indicator was evaluated by calculating Spearman correlations with the true breeding values for the production component and each resilience component (intensity, duration, recovery):

$$\rho_{Spearman}(\text{Indicator}, BV_{component_i}). \tag{4.14}$$

Spearman correlation was chosen because it is a non-parametric measure of rank correlation, making it robust to non-normal distributions. This approach is particularly suitable given that resilience indicators often exhibit skewed distributions and may show threshold effects rather than strictly linear associations with underlying resilience capacity (Schober et al. 2018).

Given the genetic architecture with $h^2 = 0.15$ and since that (Falconer 1989; Wright 1921):

$$\begin{aligned} h^2 &= b_{AP} \\ r_{A,P} &= b_{AP} \times \frac{\sigma_P}{\sigma_A} \\ &= h^2 \times \frac{1}{h} \\ &= \sqrt{h^2} \end{aligned}$$

the expected correlation between resilience indicators and resilience breeding values should be approximately equal to the square root of heritability, approximately 0.38, assumed that the resilience indicators are optimal phenotypes for capturing resilience. Under the same assumption, the expected correlation between production BV and resilience indicators is approximately $-0.3 \times \sqrt{h^2_{production}} \times \sqrt{h^2_{resilience}}$, or approximately -0.10. The value of the genetic correlation between production and resilience components was chosen to be -0.30 but it is important to note that was arbitrarily selected and there are studies that reported higher unfavorable values, such as around 0.70 for LnVar with milk yield (Poppe et al. 2020) and 0.80 for the logarithm of the standard deviation with milk production (Wang et al. 2022a). LnVar shown the highest unfavorable genetic correlation with milk production.

To statistically assess the performance of resilience indicators, a linear model was fitted to the Spearman correlation values, with fixed effects for the resilience indicator, the lactation curve model, and their interaction:

$$Y_{i,j} = \mu + \alpha_i + \beta_j + \iota_{i,j} + \varepsilon, \tag{4.15}$$

where $Y_{i,j}$ is the Spearman correlation between the resilience phenotype and the breeding value under analysis, α_i is the effect of the i^{th} resilience indicator, β_j is the effect of the j^{th} lactation curve model and $\iota_{i,j}$ is the effect of their interaction. This model was run four times, once for each resilience BV component (intensity, duration, recovery speed) and once for the production BV.

4.3 Results

A small group of resilient cows, validated through health records and rumination patterns, served as the basis for simulating resilience-modulated events. The aim was to better understand how lactation curve models and resilience indicators capture the genetic component of resilience.

Spearman correlations between resilience indicators and breeding values for production and resilience components were analyzed using a linear model. In this model, the Spearman correlation value served as the response variable, modeled as a linear combination of the fixed effects of the resilience indicator, the lactation curve model, and their interaction (Equation 4.15). Table 4.2 presents the ANOVA table. As expected, the interaction between resilience indicators and lactation curve models was highly significant, indicating that the combined choice of these factors is relevant.

Table 4.2: ANOVA tables for the linear model of Equation 4.15, applied to the Spearman correlations between resilience indicators and breeding values for resilience components (intensity, duration, recovery) and production.

(a) Spearman correlations between resilience indicators and breeding values for resilience duration.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Model	17	3.81	0.22	96.91	<0.001
Indicator	8	47.75	5.96	2577.60	<0.001
Model x Indicator	136	17.34	0.12	55.05	<0.001
Residuals	24138	55.90	0.00		

(b) Spearman correlations between resilience indicators and breeding values for resilience intensity.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Model	17	6.23	0.366	160.16	<0.001
Indicator	8	62.85	7.85	3433.31	<0.001
Model x Indicator	136	15.11	0.11	48.56	<0.001
Residuals	24138	55.23	0.00		

(c) Spearman correlations between resilience indicators and breeding values for resilience recovery.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Model	17	5.25	0.30	130.47	<0.001
Indicator	8	41.29	5.16	2176.49	<0.001
Model x Indicator	136	10.72	0.08	33.26	<0.001
Residuals	24138	57.24	0.00		

(d) Spearman correlations between resilience indicators and breeding values for production.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Model	17	32.89	1.93	850.46	<0.001
Indicator	8	136.84	17.10	7518.68	<0.001
Model x Indicator	136	68.63	0.50	221.84	<0.001
Residuals	24138	54.91	0.00		

Based on the significant interaction, the ability of resilience indicators to capture genetic resilience was further explored by visualizing the least-squares means (LSmeans) of the Spearman correlations for the resilience indicators (Figure 4.1), lactation curve models (Figure 4.2), and their interaction (Figure 4.3), all of theme based on the linear model of Equation 4.15.

LSmeans for the resilience indicators revealed that AvgAbs, AC1, RMSE, LMS and LnVar were the most effective in capturing the genetic component of resilience duration, with correlations around 0.16 (Figure 4.1). These values are slightly higher for the intensity

component (around 0.17), and slightly lower for the recovery component (around 0.15). The other indicators (AUC, ΔAvg, LnVaRel and Skew) showed lower correlations (lower than 0.10), with ΔAvg and LnVaRel performing the worst (around 0.05). This scheme was valid for all three resilience components.
Regarding the correlation with milk production BV, AvgAbs RMSE LMS LnVar and LnVaRel showed the highest absolute correlations (between 0.25 and 0.30), while AC1 showed the least correlation (lower than 0.10).

LSmeans for the lactation curve models (Figure 4.2) indicated that the best-performing models achieved correlations of approximately 0.15 with the resilience components (intensity, duration, recovery), while the lowest-performing models reached around 0.10. WSpline was the top model for all three components; however, for Duration, its leading position was contested by the IWood model, with other models following closely but showing a gradual decline in performance. For Intensity and Recovery, WSpline was followed by two groups of models that consistently performed slightly lower than WSpline.

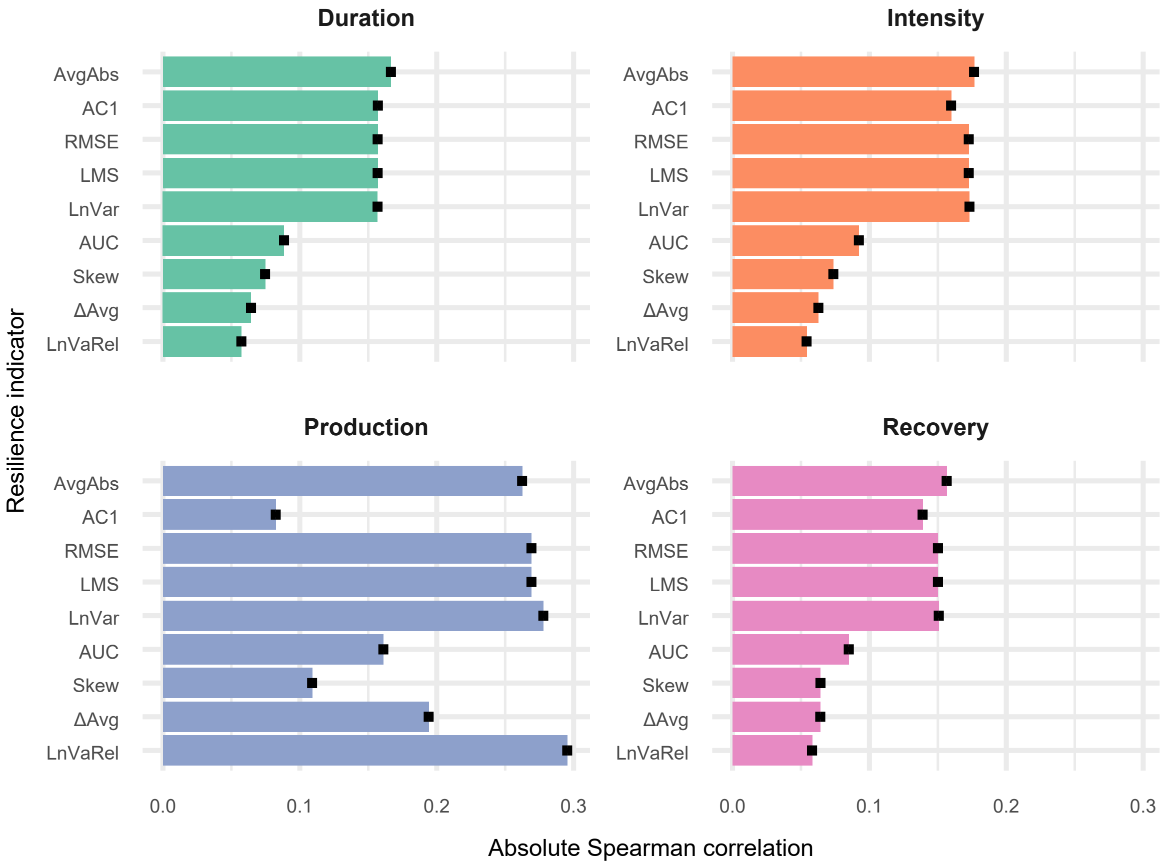


Figure 4.1: Histograms of LSmeans for the effects of the indicators upon the Spearman correlations between resilience proxies and breeding values for production and resilience components (intensity, duration, recovery).

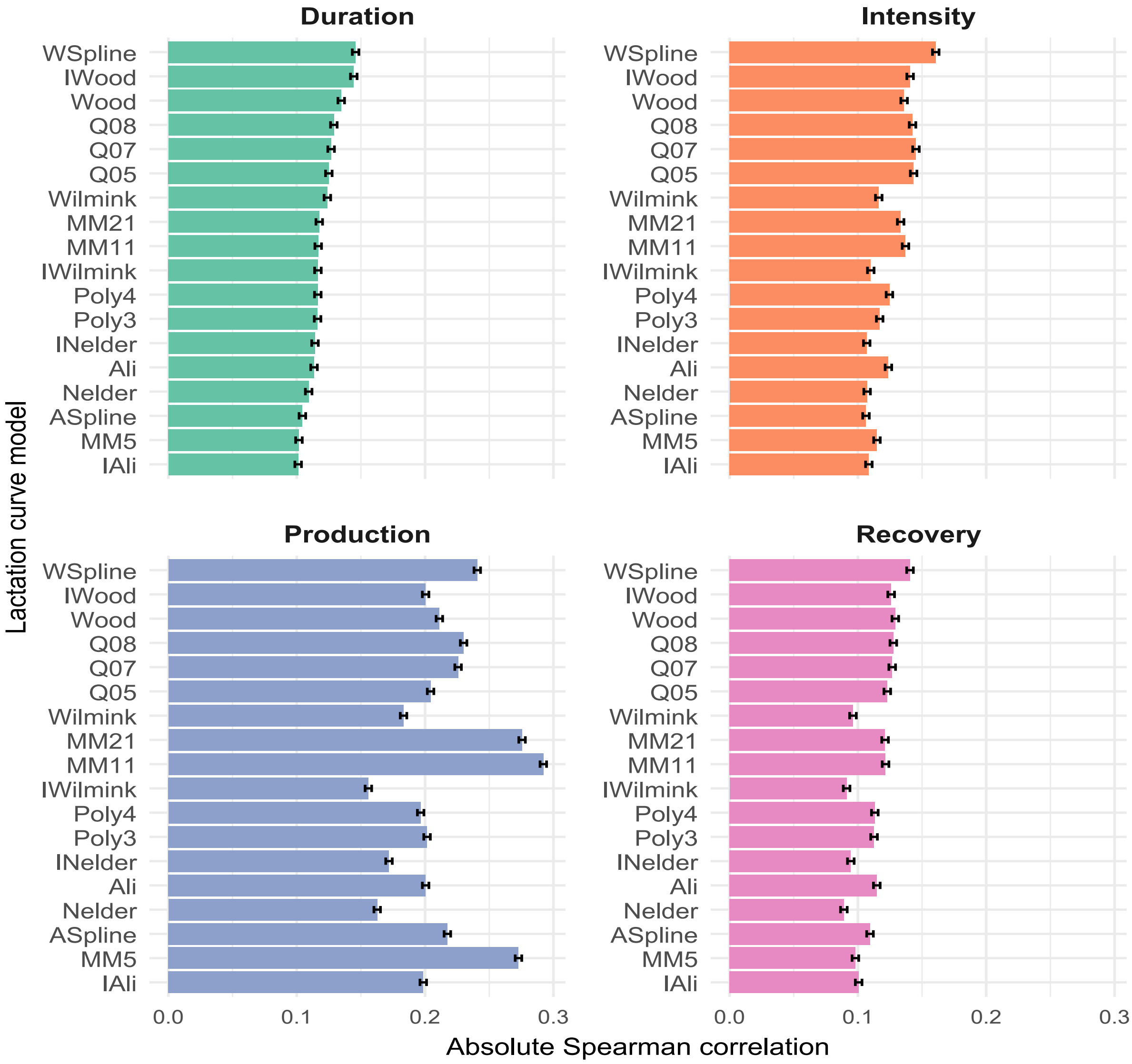
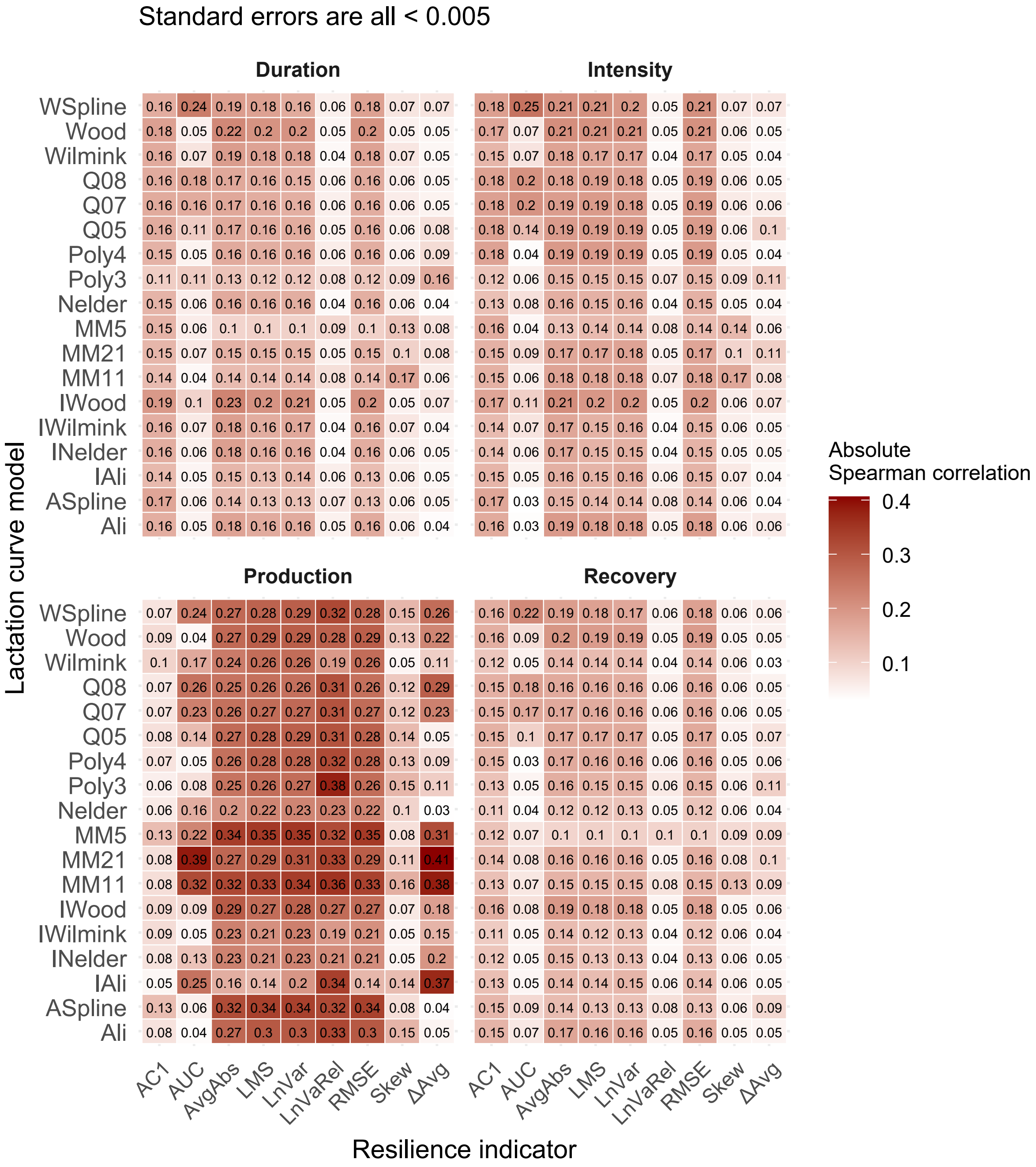


Figure 4.2: Histograms of LSmeans for the effects of the lactation curve models upon the Spearman correlations between resilience proxies and breeding values for production and resilience components (intensity, duration, recovery).

The LSmeans of the Spearman correlations for the interaction between resilience indicators and lactation curve models (Figure 4.3) revealed that the best-performing combinations achieved correlation values between 0.20 and 0.25 for all three resilience components (intensity, duration, recovery). AvgAbs, LMS and LnVar estimated upon the Wood and IWood models raised the highest correlation values, followed by AUC in combination with WSpline. Looking at the pattern of colours in the heatmap, it is evident that AvgAbs, LMS, LnVaRel and RMSE consistently perform well across various models, as emerged in the LSmeans for resilience indicators, while for AUC the only good combination was with WSpline (4.1). However, values were generally higher for the correlation among resilience proxies and the production BV, where the best-performing combinations are those that reach the smallest correlations with production BV.



Duration

Lactation curve model	AC1	AUC	AvgAbs	LMS	LnVar	LnVaRel	RMSE	Skew	ΔAvg
WSpline	0.16	0.24	0.19	0.18	0.16	0.06	0.18	0.07	0.07
Wood	0.18	0.05	0.22	0.2	0.2	0.05	0.2	0.05	0.05
Wilmink	0.16	0.07	0.19	0.18	0.18	0.04	0.18	0.07	0.05
Q08	0.16	0.18	0.17	0.16	0.15	0.06	0.16	0.06	0.05
Q07	0.16	0.16	0.17	0.16	0.16	0.06	0.16	0.06	0.05
Q05	0.16	0.11	0.17	0.16	0.16	0.05	0.16	0.06	0.08
Poly4	0.15	0.05	0.16	0.16	0.16	0.06	0.16	0.06	0.09
Poly3	0.11	0.11	0.13	0.12	0.12	0.08	0.12	0.09	0.16
Nelder	0.15	0.06	0.16	0.16	0.16	0.04	0.16	0.06	0.04
MM5	0.15	0.06	0.1	0.1	0.1	0.09	0.1	0.13	0.08
MM21	0.15	0.07	0.15	0.15	0.15	0.05	0.15	0.1	0.08
MM11	0.14	0.04	0.14	0.14	0.14	0.08	0.14	0.17	0.06
IWood	0.19	0.1	0.23	0.2	0.21	0.05	0.2	0.05	0.07
IWilmink	0.16	0.07	0.18	0.16	0.17	0.04	0.16	0.07	0.04
INelder	0.16	0.06	0.18	0.16	0.16	0.04	0.16	0.06	0.05
IAli	0.14	0.05	0.15	0.13	0.14	0.06	0.13	0.06	0.05
ASpline	0.17	0.06	0.14	0.13	0.13	0.07	0.13	0.06	0.05
Ali	0.16	0.05	0.18	0.16	0.16	0.05	0.16	0.06	0.04

Intensity

Lactation curve model	AC1	AUC	AvgAbs	LMS	LnVar	LnVaRel	RMSE	Skew	ΔAvg
WSpline	0.18	0.25	0.21	0.21	0.2	0.05	0.21	0.07	0.07
Wood	0.17	0.07	0.21	0.21	0.21	0.05	0.21	0.06	0.05
Wilmink	0.15	0.07	0.18	0.17	0.17	0.04	0.17	0.05	0.04
Q08	0.18	0.2	0.18	0.19	0.18	0.05	0.19	0.06	0.05
Q07	0.18	0.2	0.19	0.19	0.18	0.05	0.19	0.06	0.06
Q05	0.18	0.14	0.19	0.19	0.19	0.05	0.19	0.06	0.1
Poly4	0.18	0.04	0.19	0.19	0.19	0.05	0.19	0.05	0.04
Poly3	0.12	0.06	0.15	0.15	0.15	0.07	0.15	0.09	0.11
Nelder	0.13	0.08	0.16	0.15	0.16	0.04	0.15	0.05	0.04
MM5	0.16	0.04	0.13	0.14	0.14	0.08	0.14	0.14	0.06
MM21	0.15	0.09	0.17	0.17	0.18	0.05	0.17	0.1	0.11
MM11	0.15	0.06	0.18	0.18	0.18	0.07	0.18	0.17	0.08
IWood	0.17	0.11	0.21	0.2	0.2	0.05	0.2	0.06	0.07
IWilmink	0.14	0.07	0.17	0.15	0.16	0.04	0.15	0.06	0.05
INelder	0.14	0.06	0.17	0.15	0.15	0.04	0.15	0.05	0.05
IAli	0.15	0.05	0.16	0.15	0.16	0.06	0.15	0.07	0.04
ASpline	0.17	0.03	0.15	0.14	0.14	0.08	0.14	0.06	0.04
Ali	0.16	0.03	0.19	0.18	0.18	0.05	0.18	0.06	0.06

Production

Lactation curve model	AC1	AUC	AvgAbs	LMS	LnVar	LnVaRel	RMSE	Skew	ΔAvg
WSpline	0.07	0.24	0.27	0.28	0.29	0.32	0.28	0.15	0.26
Wood	0.09	0.04	0.27	0.29	0.29	0.28	0.29	0.13	0.22
Wilmink	0.1	0.17	0.24	0.26	0.26	0.19	0.26	0.05	0.11
Q08	0.07	0.26	0.25	0.26	0.26	0.31	0.26	0.12	0.29
Q07	0.07	0.23	0.26	0.27	0.27	0.31	0.27	0.12	0.23
Q05	0.08	0.14	0.27	0.28	0.29	0.31	0.28	0.14	0.05
Poly4	0.07	0.05	0.26	0.28	0.28	0.32	0.28	0.13	0.09
Poly3	0.06	0.08	0.25	0.26	0.27	0.38	0.26	0.15	0.11
Nelder	0.06	0.16	0.2	0.22	0.23	0.23	0.22	0.1	0.03
MM5	0.13	0.22	0.34	0.35	0.35	0.32	0.35	0.08	0.31
MM21	0.08	0.39	0.27	0.29	0.31	0.33	0.29	0.11	0.41
MM11	0.08	0.32	0.32	0.33	0.34	0.36	0.33	0.16	0.38
IWood	0.09	0.09	0.29	0.27	0.28	0.27	0.27	0.07	0.18
IWilmink	0.09	0.05	0.23	0.21	0.23	0.19	0.21	0.05	0.15
INelder	0.08	0.13	0.23	0.21	0.23	0.21	0.21	0.05	0.2
IAli	0.05	0.25	0.16	0.14	0.2	0.34	0.14	0.14	0.37
ASpline	0.13	0.06	0.32	0.34	0.34	0.32	0.34	0.08	0.04
Ali	0.08	0.04	0.27	0.3	0.3	0.33	0.3	0.15	0.05

Recovery

Lactation curve model	AC1	AUC	AvgAbs	LMS	LnVar	LnVaRel	RMSE	Skew	ΔAvg
WSpline	0.16	0.22	0.19	0.18	0.17	0.06	0.18	0.06	0.06
Wood	0.16	0.09	0.2	0.19	0.19	0.05	0.19	0.05	0.05
Wilmink	0.12	0.05	0.14	0.14	0.14	0.04	0.14	0.06	0.03
Q08	0.15	0.18	0.16	0.16	0.16	0.06	0.16	0.06	0.05
Q07	0.15	0.17	0.17	0.16	0.16	0.06	0.16	0.06	0.05
Q05	0.15	0.1	0.17	0.17	0.17	0.05	0.17	0.05	0.07
Poly4	0.15	0.03	0.17	0.16	0.16	0.06	0.16	0.05	0.06
Poly3	0.13	0.05	0.16	0.15	0.15	0.06	0.15	0.06	0.11
Nelder	0.11	0.04	0.12	0.12	0.13	0.05	0.12	0.06	0.04
MM5	0.12	0.07	0.1	0.1	0.1	0.1	0.1	0.09	0.09
MM21	0.14	0.08	0.16	0.16	0.16	0.05	0.16	0.08	0.1
MM11	0.13	0.07	0.15	0.15	0.15	0.08	0.15	0.13	0.09
IWood	0.16	0.08	0.19	0.18	0.18	0.05	0.18	0.05	0.06
IWilmink	0.11	0.05	0.14	0.12	0.13	0.04	0.12	0.06	0.04
INelder	0.12	0.05	0.15	0.13	0.13	0.04	0.13	0.06	0.05
IAli	0.13	0.05	0.14	0.14	0.15	0.06	0.14	0.06	0.05
ASpline	0.15	0.09	0.14	0.13	0.13	0.08	0.13	0.06	0.09
Ali	0.15	0.07	0.17	0.16	0.16	0.05	0.16	0.05	0.05

Figure 4.3: Heatmaps of LSmeans for the interaction between resilience indicators and lactation curve models upon the Spearman correlations between resilience proxies and breeding values for production and resilience components (intensity, duration, recovery).

4.4 Discussion

The present simulation study offers a comprehensive evaluation of the ability of various lactation curve models and resilience indicators to capture the genetic resilience of dairy cows. By establishing a ground truth set of resilient individuals based on health records and rumination patterns, and simulating breeding values with a controlled genetic architecture, the study addresses two central research questions: whether resilience indicators can effectively capture genetic resilience, and which combinations of lactation curve models and indicators are most informative.

Some studies have previously calculated the genetic parameters of resilience indicators derived from longitudinal observations of milk yield, yielding promising results: heritability estimates ranging from 0.13 to 0.25 (Poppe et al. 2020; Chen et al. 2023; Kessler et al. 2024; Guinan et al. 2025a). These findings suggest that resilience indicators have a genetic basis and can potentially be improved through selective breeding. However, because resilience indicators are proxies rather than direct measures of resilience, it remains uncertain how accurately they reflect the genetic resilience of cows. This highlights the importance of the definition of resilience, as this simulation study is based on defining resilience as a cow's capacity to be minimally affected by disturbances (through modulation of the intensity and duration components) and to recover rapidly from a disturbance (through the recovery component) (Colditz et al. 2016).

Based on this definition of resilience and its modulation in the lactation curve of each cow, the simulation study aimed to understand how well resilience proxies capture the cows' resilient capacity when affected by simulated perturbations. These perturbations were based on values reported in studies of mastitis and heat stress, and phenotypic ranges were used as the basis for environmental variances in the simulation. Genetic variance was determined by fixing heritability of 0.15, derived from previous genetic parameter estimates of resilience indicators. Our goal was not to perfectly simulate real-world scenarios, where responses to perturbations depend on various factors, but to ensure a credible simulation framework by setting wide phenotypic ranges and allocating plausible variance components to the quantitative models implemented.

Certain resilience indicators demonstrated a mediocre ability to capture genetic resilience, with Spearman correlations reaching up to 0.15-0.17 with true resilience breeding values for intensity, duration, and recovery components. These indicators are AvgAbs, AC1, RMSE, LMS and LnVar. Among these, AvgAbs consistently exhibited the highest correlations across all three resilience components. However, it is evident that no single resilience indicator achieved the expected correlation of approximately 0.38, which would be anticipated if they were optimal phenotypes for capturing resilience. This discrepancy suggests that while some indicators are effective, there remains room for improvement in developing more accurate measures of genetic resilience, if the aim is to capture the dynamics of resilience as defined in this study. Regarding the correlation with production breeding values, among the top-performing indicators for resilience, only AC1 showed low correlation (<0.1) with production BV, while the others exhibited moderate correlations (slightly less than 0.3). 0.1 was the expected values based on the genetic correlation structure defined in the simulation.

Among the lactation curve models, WSpline, IWood, Wood, Q08, Q07 and Q05 were the most effective in capturing genetic resilience, achieving correlations of approximately 0.15 with the resilience components. These models, except for Wood (Wood 1967) and Q05 (Guinan et al. 2025a), aim to fit a lactation curve less affected by perturbations, thus better reflecting the unperturbed state. The superior performance of these models suggests that accurately estimating the unperturbed lactation curve is crucial for deriving resilience indicators that effectively capture genetic resilience. Furthermore, as WSpline (Kessler et al. 2024) slightly outperformed the other models, it may indicate that approach may combine the right flexibility and robustness to perturbations in estimating the unperturbed lactation curve.

The interaction between resilience indicators and lactation curve models was significant, indicating that the combined choice of these factors is relevant for capturing genetic resilience. The best-performing combinations achieved correlation values between 0.20 and 0.25 for all three resilience components, suggesting that certain pairings of indicators and models can enhance the ability to capture genetic resilience. The combination of AUC estimated upon the WSpline model emerged as the top-performing pair, followed closely by AvgAbs, LMS and LnVar estimated upon the Wood and IWood models. For AUC, WSpline was the unique model that allowed it to perform well, while for AvgAbs, LMS, RMSE and LnVar, Wood, IWood and quantile polynomials also allowed good performances, suggesting the generability of these indicators across different lactation curve models, and viceversa.

Unfortunately, recently proposed resilience indicators such as LnVaRel and ΔAvg (Kessler et al. 2024; Guinan et al. 2025a) did not perform well in capturing genetic resilience. This is compounded by their high sensitivity to lactation curve models, which further limits their effectiveness (Chapter 3). Yet, noteworthy is that ΔAvg was specifically designed to capture resilience not from the entire set of lactation deviations, but only from negative deviations once perturbations at the pen level were individuated.

The goodness of AUC, LnVar and LMS was confirmed by Zefreh et al. (2024), who evaluated their capacity to predict the resilience of dairy cows under a classification analysis. These indicators were the most reliable because of their robustness to measure frequency and observation period variations. All together, these findings underscore the potential of these indicators as valuable tools for assessing genetic resilience in dairy cattle, even if their capacity to capture true resilience is still moderate, as this simplicist simulation study revealed.

Inherent limitations that should be acknowledged include: i) production BV estimation: the framework assumes estimated breeding values from the mixed model represent true genetic values, which may introduce noise in relatively small datasets; ii) genetic and environmental variances derivation: variances are derived from phenotypic ranges using the range divided by four approximation, rather than directly from genetic parameter estimates; iii) event independence: events are simulated independently, though in biological reality they may interact (e.g., mastitis potentially increasing susceptibility to heat stress); iv) component-specific vs. event-specific BVs: each resilience component (intensity, duration, recovery) has its own breeding value, but individual event types do not have event-specific breeding values; v) event occurrence: the current approach does not guarantee a minimum number of events per lactation, which could result in some cows experiencing few or no resilience challenges; this can mask the genetic penetrance of true resilience.

4.5 Conclusions

Thanks to the simulation study, in which a group of lactation curves was assumed to be the explicit phenotypic representation of resilient cows and perturbation events were introduced to allow modulation of three biological response functions corresponding to three resilience components – intensity, duration, and recovery speed – we were able to address our research questions posed at the beginning of the study: 1) Are resilience indicators capable of capturing the genetic resilience of cows? 2) If so, which model-indicator combination is the most effective?

First, resilience indicators can capture genetic resilience, albeit to a relatively lower degree, as the expected correlation is higher. These results may reflect the fact that resilience indicators serve as proxies designed to capture the underlying mechanisms of disturbance coping. Second, the AUC indicator based on deviations obtained with the WSpline model was the best-performing combination, with performance slightly higher than variance-based indicators derived from other lactation curve models. However, this should be interpreted with caution, as the performance of AUC was notable only in combination with WSpline, while variance-based indicators showed good performance across multiple models; to this should be added the consideration that AUC's sensitivity to the lactation curve model further limits its generalizability as seen in Chapter 3.

The results of this study specifically identify which combinations are more effective in capturing genetic resilience through resilience indicators. However, the limitations of this study should be acknowledged when interpreting the results, in order to understand the extent to which these findings can be generalized.
Future improvements to this simulation study may include more detailed refinement of the simulated perturbations and their corresponding variances.

Chapter 5

Mapping genomic regions affecting resilience traits in a large dairy farm of Holstein cows

Chiara Punturiero[a*], Andrea Delledonne[a*], Carlotta Ferrari[a], Raffaella Milanesi[a], Maria Giuseppina Strillacci[a], Alessandro Bagnato[a]

[a] *Department of Veterinary Medicine and Animal Sciences, University of Milan, Italy*

[*] *These authors contributed equally to this work.*

Published on Front. Anim. Sci., August 11, 2025
Sec. Animal Breeding and Genetics
Volume 6 - 2025 | `https://doi.org/10.3389/fanim.2025.1627086`

5.1 Abstract

This study evaluated the genetic architecture of resilience indicators in Holstein cows managed in a herd equipped with automatic milking systems (AMS) from 2017 to 2024. Four resilience indicators were calculated based on deviations in daily milk yield: log-transformed variance (LnVar), autocorrelation of residuals (rauto), weighted frequency of perturbations (wfPert), and accumulated milk losses due to perturbations (dPert). Polynomial quantile regression models were applied to 594,481 daily records from 966 cows, with data filtered for completeness and lactation duration. Descriptive statistics revealed that LnVar increased with parity, indicating greater production variability in older cows, while rauto remained stable, suggesting a consistent ability of cows to recover from production perturbations. Both dPert and wfPert increased across lactations, reflecting greater cumulative losses and perturbation frequencies. GWAS were performed using the selective genotyping approach coupled to the statistics of DNA pooling. Genes related to immune response, energy metabolism and tissue integrity were identified. These findings suggest a multifactorial complex genetic nature of resilience and disclose the involvement of several genes that can explain both the physiology related to production and response

to stressors.

5.2 Introduction

Over the past few years, the introduction of advanced monitoring systems in dairy farms such as automatic milking systems (AMS), have significantly enhanced data collection capabilities. The adoption of AMS provides continuous and high-resolution data. This system enables an unprecedented ability to monitor fluctuations in milk yield and to detect health-related disturbances. Such systems offer valuable insight into resilience, particularly by facilitating the longitudinal monitoring required to capture production responses to environmental stressors and management interventions (Niloofar et al. 2021).

Resilience in dairy cows is a critical aspect, especially in the face of environmental disruptions such as calving, infection by various pathogens, extreme weather (e.g. heat waves), or fluctuations in feed availability and management practice (Hansen et al. 2012; Aradotlu Parameshwarappa et al. 2019). Maintaining high milk yield is essential for the profitability of dairy farms, but health issues can lead to substantial losses in milk production and its quality (Liang et al. 2017). These losses often appear as disruptions in the lactation curve (Hertl et al. 2014; Ben Abdelkrim et al. 2021b), due to reduced feed intake, immune system activation (Daniel et al. 2016) and intramammary infections affecting the udder's functionality (Heikkilä et al. 2018). Resilient cows are minimally affected by these disturbances and are able to recover quickly, resulting in reduced labor requirements, lower treatment costs, and decreased milk yield losses compared to their less resilient counterparts (Colditz et al. 2016; Berghof et al. 2019b). In order to quantify general resilience, two primary indicators have been recently proposed: the natural log-transformed variance of deviations from an expected lactation curve (LnVar) and the lag-1 autocorrelation of yield deviations (rauto). These indicators are based on the hypothesis that cows with stable production (low LnVar) and quick recovery from disturbances (low rauto) exhibit greater resilience (Scheffer et al. 2018; Poppe et al. 2020). Lower LnVar values indicate more stable production and higher resilience to disturbances (Scheffer et al. 2018). Differently, the parameter rauto assesses recovery times from disturbances by measuring the autocorrelation of yield deviations over time. A lower rauto value signifies a faster return to baseline performance following disturbances, indicating quicker recovery (Poppe et al. 2020; Wang et al. 2022a). In addition to these established indicators, more recent published research has introduced novel metrics for resilience assessment, daily perturbations (dPert) and weighted frequency of perturbations (wfPert) (Chen et al. 2023).

These metrics propose a more dynamic and real-time evaluation of resilience by capturing the frequency and impact of deviations in daily milk yield. Similar approaches have been applied to dairy cattle using daily step counts and to other livestock species, including pigs, chickens, and lambs, using longitudinal records of feed intake and body weight (BW) (Berghof et al. 2019a; Nguyen-Ba et al. 2020; Ben Abdelkrim et al. 2021b; Garcia-Baccino et al. 2021; Poppe et al. 2022a). Despite these advances, the integration of genomic information to uncover the genetic basis of resilience is still in its infancy and remains an open topic that requires further investigation. Studies have shown that LnVar has the highest heritability among resilience indicators, ranging from 0.13 to 0.21 depending on the lactation stage (Poppe et al. 2021a; Chen et al. 2023). It is also significantly genetically correlated with health, longevity, fertility, and metabolic traits, highlighting its potential value for breeding more resilient cows. In contrast, rauto exhibits low heritability, ranging from 0.02 to 0.08, suggesting that while it provides some insight into recovery times, it may be less effective for selection purposes. Similarly, the heritability values of wfPert and dPert are low, spanning from 0.01 to 0.06 across different parities (Chen et al. 2023). Genome-wide association studies (GWAS) have revealed differences in the genetic architecture of resilience indicators derived from milk yield variability in North American Holstein cattle (Chen et al. 2024). Relevant genomic regions and biological pathways, particularly those related to intestinal homeostasis, were identified and Mendelian Randomization (MR) analyses indicated an unfavorable causal association between daily milk yield (DMY) and LnVar, suggesting caution in its use for breeding resilient cattle (Chen et al. 2024). Additionally, herd management significantly affects resilience indicators, as variations across herd-years highlight the impact of practices like feed management on environmental disturbances (Chen et al. 2024).

Based on this concept, the aims of the present study were: i) to calculate the resilience indicators described above for cows farmed in a single dairy herd with a large amount of longitudinal data available from an AMS; and ii) to perform a GWAS based on a selective genotyping (Darvasi et al. 1992) using the selective DNA pooling statistics methodology to identify linkage between QTLs and SNP markers (Darvasi et al. 1994).

5.3 Material and methods

5.3.1 Data editing

In this study, we used data collected from nine AMS units between January 2017 and August 2024 on a farm in the Lombardy region (Italy). The dataset consisted of 594,481 daily observations from 966 individual cows, 113 of which had three completed lactations. To ensure the use of high-quality data, we conducted rigorous preprocessing. The first step involved thorough data editing, focusing on filtering out records with missing values or instances where daily milk production was recorded as zero. Next, we removed from the analysis the production of the first 10 Days in Milk (DIM). Then the data analysis was stratified by parity while only those from 1 to 3 were retained. Lactations were filtered by selecting only those that started before the 30th DIM and had at least 250 daily recordings (i.e. DIM); we considered for the study the conventional lactation at 305 DIM. Table 5.1 provides a detailed summary of the number of observations at the start of the analysis and the remaining counts after each data filtering step, alongside the corresponding relative distribution of cows analyzed. After the application of all filtering criteria, the final dataset included 627 cows in first parity, 320 in second parity, and 136 in third parity. This progression ensures a robust dataset with representative samples across the different parity orders for subsequent analyses.

Table 5.1: Data-editing adopted before computing the resilience indicators.

Editing step	N. of observations	N. of cows
Initial data set	594481	906
Removed daily records with missing information	520509	865
Removed first 10 days	452891	846
Lactation 1	193113	681
Max DIM $<$ 305 & Min DIM $>$ 10	178069	627
Lactation 2	139690	582
Max DIM $<$ 305 & Min DIM $>$ 10	92213	320
Lactation 3	62819	277
Max DIM $<$ 305 & Min DIM $>$ 10	38535	136

5.3.2 Expected lactation curve

For each cow and each parity order, we applied a fourth-degree polynomial quantile regression using the ‘quantreg’ package in R and the poly function to generate orthogonal polynomial terms (Koenker et al. 2018; Poppe et al. 2020). This approach allowed us to capture nuanced variations in milk production. For each cow and parity, a polynomial quantile regression curve with the 70th percentile ($\tau = 0.7$) was fitted using daily milk production data and days in milk (DIM), capturing the expected milk yield under normal conditions:

$$\beta_0 + \beta_1 \cdot DIM_i + \beta_2 \cdot DIM_i^2 + \beta_3 \cdot DIM_i^3 + \beta_4 \cdot DIM_i^4 + e_i, \tag{5.1}$$

where: y_i is the predicted daily milk production for day DIM_i; β_0, β_1, β_2, β_3, β_4 are the intercept and the first to fourth order coefficients of the polynomial regression describing the relationship between DIM_n and y_i; e_i is the error term.

The choice of the 70th percentile in the quantile regression was guided by its ability to provide a balanced estimate of the expected milk yield under normal physiological conditions. This quantile reflects production levels that are less influenced by short-term fluctuations or extreme negative deviations, making it particularly suitable for evaluating how well cows sustain higher yields in the face of potential disturbances. This approach has been supported in previous studies (Poppe et al. 2020) which employed the 70^{th} percentile for similar purposes in resilience assessment. Then we calculated the daily yield deviations (residuals) as follows:

$$Res_i = ObservedDailyMilkProduction_i - y_i, \tag{5.2}$$

where Res_i represents the deviation between the observed and expected milk yield on day DIM_i, as derived from model 5.1.

5.3.3 Resilience indicators calculation

Four resilience indicators were then calculated starting from the daily yield deviations residuals, Res_i, all of which are based on the cited studies (Poppe et al. 2020; Adriaens et al. 2021; Chen et al. 2023):

i) Natural log-transformed variance (LnVar): obtained as the natural logarithm of the variance of all Res_i:

$$\text{LnVar} = \ln\left(\frac{1}{n-1}\sum_{i=1}^{n}(Res_i - \bar{Res}_i)^2\right), \tag{5.3}$$

where n represents the total number of observations, x denotes the i^{th} observation, spanning from the first to the last available DIM, and $\bar{x}$ represents the average of the i observation. LnVar provides a comprehensive measure of milk production stability within each lactation.

ii) Autocorrelation of Residuals (rauto): is calculated as the autocorrelation of residuals over consecutive days:

$$r_{\text{auto}} = \frac{n}{s^2(n-1)}\sum_{i=1}^{n_i}(x_i - \bar{x})(x_{i+1} - \bar{x}), \tag{5.4}$$

where n represents the total number of observations, x denotes the i^{th} observation, spanning from the first to the last available DIM, $\bar{x}$ represents the average of the i observations, and $^{\text{s}}$ represents the standard deviation of the l observations.
This indicator helps identify patterns of persistence or volatility in milk production dynamics over time.

iii) Weighted occurrence frequency of yield perturbations (wfPert): this indicator quantifies the occurrence of perturbations. Only sequences of at least four consecutive days with negative deviations from expected values in milk production were considered. Each sequence was required to include at least one day in which total daily milk production fell below 80% of the estimated unperturbed lactation curve. The duration of each sequence was weighted (1/4 per day) to assess its overall impact on milk production efficiency (Adriaens et al. 2021; Chen et al. 2023):

$$wfPert = \sum_{p=1}^{n} \frac{l_p}{4}, \tag{5.5}$$

where l_p is the length of the perturbation and n is the total number of perturbation events. At the end the indicator was normalized to a ratio per 100 DIM, dividing it by the number of DIM and multiplied by 100.

iv) Accumulated milk losses of yield perturbations (dPert): it stands for the total milk losses associated with perturbation and was calculated as the differences between expected productions and actual ones, in percentage (Adriaens et al. 2021; Chen et al. 2023).

$$dPert = \sum_{p=1}^{n} \sum_{d=1}^{l} p \frac{yexp_d - yobs_d}{yexp_d} \times 100 \tag{5.6}$$

where n is the total number of perturbation events, lp is the length of each perturbation (DIM); $yexp$ and $yobs$ are the estimated and observed daily milk yield, respectively. Finally, the indicator was normalized to a ratio per 100 DIM, dividing it by the number of DIM and multiplied by 100.

The Pearson correlations between pairs of indicators were assessed both within and across lactations. Within each lactation, correlations were calculated between the four resilience indicators (LnVar, rauto, dPert, wfPert), grouped by parity order. Additionally, to evaluate consistency across lactations, correlations were computed between the same indicators for cows with complete records across all three lactations.

5.3.4 Heritability of resilience indicators

Genetic parameters were estimated using the following animal model:

$$y = X\beta + Zu + e, \tag{5.7}$$

where y is the vector of phenotypic observations for each resilience indicator, β is the vector of fixed effects including age at calving (in months), month of calving (January to

December), and year of calving (2017–2024), u is the vector of random additive genetic effects assumed to follow a distribution u∼N(0, $A\sigma^2 u$) where A is the pedigree-based relationship matrix, and e is the vector of residual errors e∼N(0,$I\sigma^2 e$).

Heritability was estimated using AIREMLF90 software (Misztal et al. 2018) based on the animal model described above. Estimates were obtained separately for parity order 1 and 2, not for parity order 3 because of the small sample size. The fixed effects include: age at calving expressed in months (from month 22 to 36 for parity 1 and 33 to 49 for parity 2), the month of calving (January to December), the year of calving (from 2017 to 2024). The relationship matrix was calculated using the available genealogical information. All cows included in the study were genotyped using the GGP 100K SNP chip, and the pedigrees were all previously verified by ANAFIBJ (Associazione Nazionale Allevatori della Razza Frisona, Bruna e Jersey Italiana; National Association of Breeders of the Friesian, Brown and Jersey Italian Breed). This validation step based on genomic information ensured the accuracy and consistency of the genealogical data provided for all the animals included in the study, reducing possible bias in estimating variance components due to incorrect genealogical recording. The pedigree information for each individual, included all available generations of ancestors was employed to build the relationship matrix. The overall pedigree file included a total of 3,158 animals, including individuals with phenotypes and individuals non-phenotyped. The pedigree file was used to build the additive relationship matrix A implemented in the animal model.

5.3.5 Genome-Wide Association Studies

To investigate the genetic basis of resilience in dairy cows, we conducted a genome-wide association study (GWAS) using a selective genotyping approach (Darvasi et al. 1992), an efficient strategy that involves genotyping only individuals with extreme phenotypes to enhance the detection of SNPs associated with the trait (Darvasi et al. 1992). This methodology is particularly efficient in detecting QTL, as a considerable amount of genetic information resides in individuals with extreme phenotypes (Xing et al. 2009; Lipkin et al. 2016).

Cows were selected based on the 10% most extreme values (highest and lowest) for each indicator: the most resilient (RE) and least resilient (NRE) cows. Since each indicator captures different aspects of lactation curve perturbations, animals ranked in the top or bottom 10% for one indicator are not necessarily the same as those in another (Medugorac et al. 2001). All genotypes (mapped according to the ARS-UCD1.2 bovine genome assembly) were already available for both selected RE and NRE cows. To simulate the selective DNA pooling strategy (Darvasi et al. 1994) using individual genotype data, each RE and NRE group was randomly divided into two biological replicates (RE1/RE2 and

NRE1/NRE2) with comparable sample sizes. For each replicate, allele frequencies at each SNP marker were calculated using the 'genotype statistics by marker' function of Golden Helix's SNP & Variation Suite (SVS v8.9, Golden Helix Inc., Bozeman, MT, USA). The GWAS was conducted only for resilience indicators with heritability greater than 0.05, both in our dataset and consistently reported in the literature (Poppe et al. 2020; Chen et al. 2023). This dual criterion ensured a focus on traits with a strong biological and genetic basis, making them more reliable for downstream genomic analysis. The analysis was further restricted to the first and second lactations, which included the largest number of animals (627 and 320 cows, respectively), to ensure robust estimation of both phenotypic indicators and genetic parameters. The GWAS analyses were conducted by comparing the allele frequency of each SNP (for one of the possible alleles) between resilient (RE1 and RE2) and non-resilient (NRE1 and NRE2) cows using an in-house R script (R version 4.0.5). Monomorphic markers, which do not provide useful information, were excluded from the analysis. Additionally, SNPs showing a high degree of variability between replicates (i.e., the top 5% of the absolute value of allele differences between RE1 vs RE2 and NRE1 vs NRE2) were also removed to reduce potential false positives calls. After the filtering process, the following numbers of SNPs were retained, out of the original 89,764 autosomal markers, for the analysis: i) LnVar1: 63,862 SNPs; ii) LnVar2: 60,513 SNPs; iii) rauto1: 69,347 SNPs; iv) rauto2: 65,066 SNPs; v) dPert2: 67,124 SNPs. This filtered dataset ensures high-quality genetic markers are available for robust downstream analyses. According to Darvasi et al. (1994), a single-marker association test was applied, calculating a Z-test for each marker to assess the association between the A allele frequency difference and resilience status. The Z-test was defined as follows:

$$Ztest = \frac{Dtest}{SD(Dnull)} \tag{5.8}$$

where *Dtest* is the difference in the A allele frequencies between the resilient and non-resilient groups, and *Dnull* is the difference within groups (i.e., between the two biological replicates within RE or NRE groups).

The association results were visualized using a Manhattan plot, generated with the qqman R package (Turner 2018). After the analysis, the False Discovery Rate (FDR) and Bonferroni correction thresholds were used for determining the statistically significant SNPs.

5.3.6 Gene annotation

All SNPs over the 5% FDR threshold were annotated, and the SNP's rsID code (Reference SNP cluster ID) of each of the Illumina SNP markers has been obtained. The Variant Effect Predictor (VEP) tool of the Ensembl database allowed to annotate the significant

SNPs through the rsID codes according to the Bos taurus genome assembly ARS-UCD1.2 (Annotation Release: 106). Candidate genes were identified as: i) if a significant SNP was annotated within a gene, this latter was considered as a candidate gene; ii) for intergenic SNPs, the candidate gene was the one mapping closest, either upstream or downstream, within a maximum distance of 500 kb. QTL associated with each indicator were identified using the QTLdb's 'Search by associated gene' option available within the Cattle Quantitative Trait Locus (QTL) Database of Animal QTLdb (`https://www.animalgenome.org/cgi-bin/QTLdb/BT/index`).

5.4 Results and discussion

5.4.1 Modelling lactation curves

The modelling of lactation curves using polynomial quantile regression allowed for the estimation of the expected daily milk production. This approach, as detailed in Poppe et al. (2020), captures the variability in milk production across different days in milk (DIM) and provides a robust basis for calculating deviations from expected milk yield and consequently the resilience indicators. The graphs in Figure 5.1 illustrate the lactation performance of two cows: one classified as more resilient (Figure 5.1a) and the other as less resilient (Figure 5.1b).

The first visible difference between the curves is the fluctuations of the daily milk yield. In Figure 5.1a, with LnVar value 1.06, the cow shows a relatively stable milk production over time: the milk yield remains consistent, with minor day-to-day variability, assuming the cow's strong ability to cope with stressors. In contrast, the cow in Figure 5.1b has a value of LnVar equal to 2.98 and reveals more significant fluctuations in daily milk yield and a sharper decline in production. The pattern of milk yield by DIM respect to the expected lactation curve suggests that the resilient cow is more capable to maintain high levels of productivity along the lactation. The less resilient cow, on the other hand, exhibits greater variability in daily milk production, which is hypothesized to reflect a reduced capacity to recover from stressors. Periods of stress are marked by sharp declines in production, followed by slower recoveries, which can be interpreted as a diminished ability to maintain production efficiency over time.

5.4.2 Descriptive statistics of the derived resilience indicators

Table 5.2 shows the descriptive statistics for the calculated resilience indicators across different lactations. When compared to recent studies, LnVar across parity orders are similar to the results of Chen et al. (2023) (average from 1.39 to 1.83 and SD from 0.57 to 0.62), while they differ from Poppe et al. (2021a) (average from 4.40 to 4.99 and SD

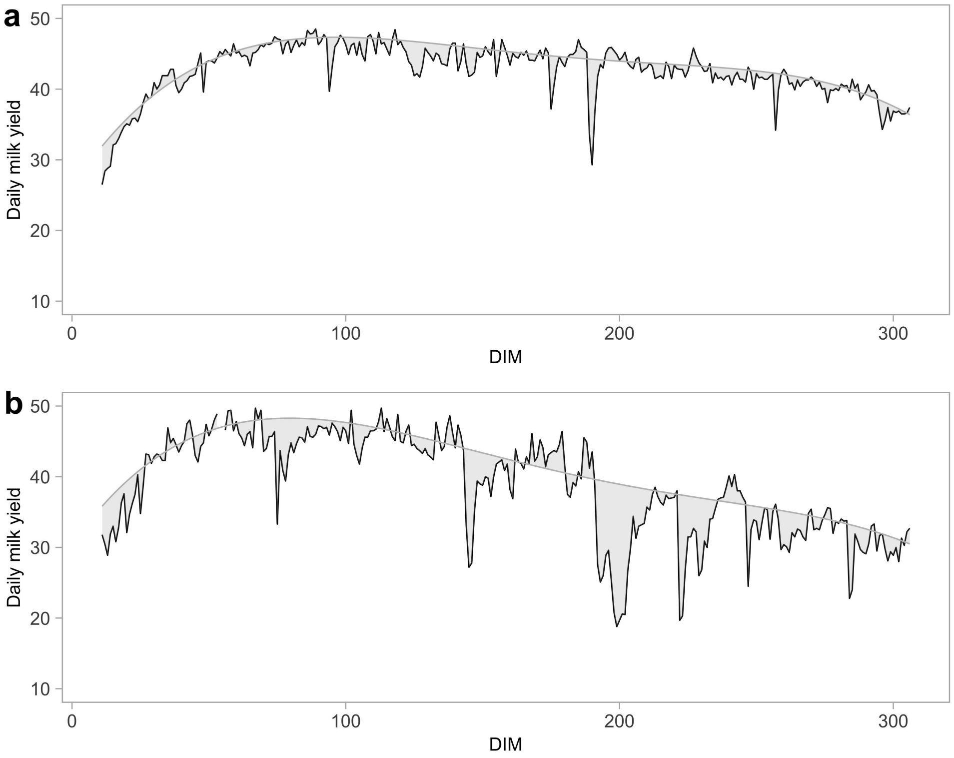


Figure 5.1: In this figure two lactation curves are reported as examples of how resilience is captured and how two different cows are considered resilient (a) and non-resilient (b). The represented lactations belong to cows in their first order of parity.

from 0.66 to 0.79). The observed increase in LnVar through the parity orders is consistent with findings from the mentioned studies, showing that milk yield variability tends to rise proportionally with aging. Possibly, this is due to physiological changes that occur with aging, such as changes in mammary gland function or metabolic regulation. Additionally, prolonged exposure to environmental stressors, including variation in management or environmental conditions over time, may contribute to the increase of the variability (Friggens et al. 2007; Wathes et al. 2007).

The rauto values indicate a consistent pattern of autocorrelation through parities. In our study, the mean rauto values were 0.62, 0.63, and 0.62 for the first, second, and third parity, respectively. In comparison, Chen et al. (2023) reported mean rauto values of 0.37, 0.43, and 0.44 for these parities, while Poppe et al. (2021a) observed values of 0.55, 0.56, and 0.55. This stability suggests that the persistence in milk production patterns stays relatively constant over different parity orders. This contrasts with the increasing variability observed in LnVar, highlighting that while overall variability in milk yield increases with parity, the persistence of production patterns (as measured by autocorrelation) may remain stable. This difference underscores the complexity of lactation dynamics, where variability and persistence can be influenced by different factors and may not always align. In this context, higher values of these indicators correspond to lower resilience and higher milk production (Poppe et al. 2021b). This is an explicit trade-off

between resilience and production. However, accurate measures of cow life events that may be available to help calculate fertility and reproductive phenotypes, together with treatment records kept by farmers, could help to assess a better definition of resilience. In addition, AMS can provide more than just daily milk yield, but longitudinal phenotypes such as feeding behavior and milk contents, which can be used to improve understanding of the physiological responses of lactating cows throughout their lactation. This was not the aim of this study.
The genetic correlations between milk yield in the first and second lactations have been estimated at approximately $r_g = 0.82$, indicating a strong but not perfect relationship between the two (Dong et al. 1989). This suggests that while a common genetic component influences both lactations, specific factors unique to each parity also contribute to milk production dynamics. The potential roles of some candidate genes in resilience are then discussed by resilience indicator below.

The upward trend in dPert across lactations suggests that milk production losses due to perturbations accumulate more significantly as cows' parity increases. In our study, the average dPert ranged from 19.1 to 20.2, with a standard deviation (SD) ranging from 2.44 to 2.74 across lactations. These values are higher than those reported by Chen et al. (2023), who found averages between 15.62 and 17.43 for the same parities. This discrepancy may be explained by the generally higher milk production in multiparous cows, where any deviation from expected yield results in proportionally greater losses. Consequently, higher yields in older cows are likely to contribute to more substantial accumulated milk losses during perturbation periods.

The wfPert values show a relatively consistent frequency of perturbations across parities, with only slight variations seen in the third one. In our study, there were 24 cows in the first lactation without any perturbations (3.8%), 10 in the second lactation (2.8%), and 4 in the third lactation (2.9%). The wfPert showed an average of 4.68 ± 2.67 SD in the first lactation, compared to 5.03 ± 2.81 in the second lactation and 5.57 ± 2.45 in the third lactation. Although the average for the third lactation is a bit higher compared to the first and second lactations, the differences are not large. In Chen et al. (2023), the results are slightly higher: from the first to the third parity order, averages and standard deviations are 5.43 ± 1.75, 5.85 ± 1.91, and 6.11 ± 1.94, respectively.

Table 5.2: Summary statistics of resilience indicators.

Indicator	Avg			SD			Min			Max		
Parity	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
LnVar	1.86	2.31	2.49	0.48	0.57	0.53	0.44	0.64	1.38	3.40	4.25	4.07
rauto	0.62	0.63	0.62	0.11	0.12	0.23	0.23	0.15	0.29	0.90	0.93	0.87
dPert	19.07	19.39	20.02	2.44	2.74	2.65	13.87	13.17	15.20	28.47	31.00	27.39
wfPert	4.68	5.03	5.57	2.67	2.81	2.45	0.00	0.00	0.00	13.01	12.46	13.18

The correlation analysis of resilience indicators within and across lactations provides valuable insights into the relationships between different measures of milk production stability and perturbations. The results of the correlation matrix are visualized in Figure 5.2.

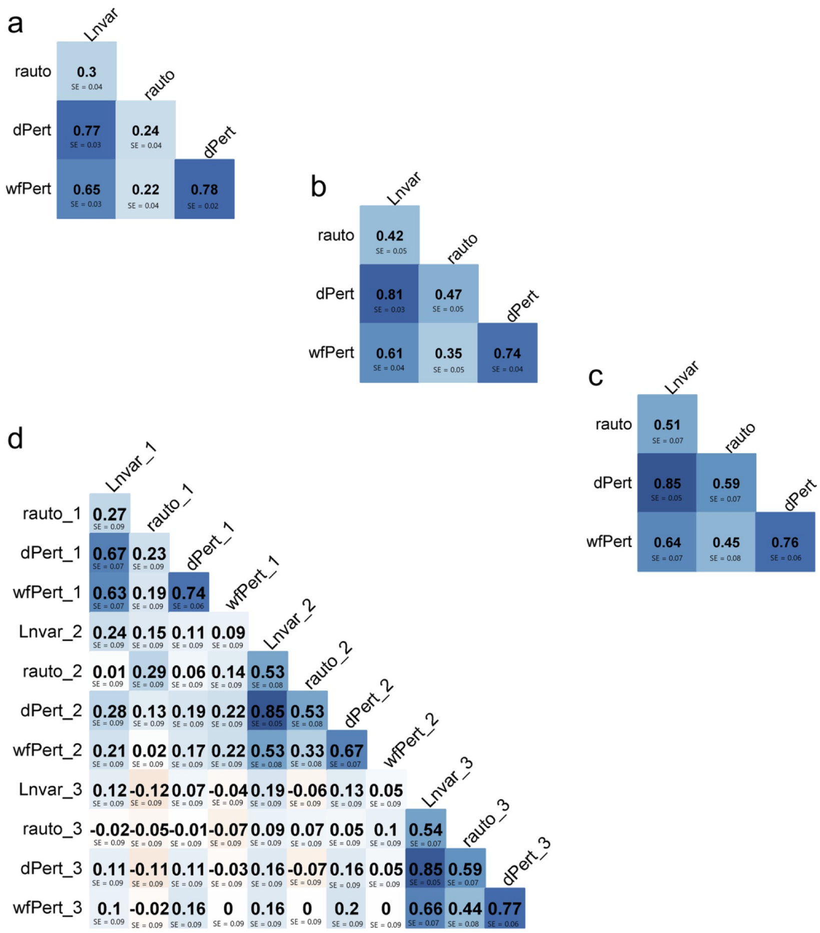
In the first lactation (Figure 5.2a), pairwise correlations between indicators are all positive and range from moderate to high, varying from 0.22 (between rauto and wfPert) to 0.78 (between wfPert and dPert). This pattern is expected, as all indicators are derived from the same residuals obtained through the quantile regression model and are designed to capture different aspects of resilience. The strongest correlation, observed between dPert and wfPert, reflects the similarity in their definitions – both quantify production losses due to perturbations, albeit from different perspectives.

Biologically, it is reasonable that cows with more frequent production losses exhibit greater total loss. Strong correlations are also observed between LnVar–dPert and LnVar–wfPert, suggesting that cows with higher variability in milk production (LnVar) also experience more frequent and severe production losses. In contrast, the lowest correlations involve rauto, which measures the persistence of deviations. Interestingly, in lactations 2 and 3 (Figures 5.2b and 5.2c), correlations between rauto and the other indicators become stronger, while the remaining pairwise correlations remain relatively stable. As shown in Figures 5.2a–c, correlations within lactations tend to strengthen across parities.

However, this pattern is not evident in Figure 5.2d, where correlations between indicators across lactations are generally low and tend to weaken. As cows age, they do not exhibit consistent resilience patterns across parities, suggesting that their responses to perturbations evolve over time. Furthermore, Figure 5.2d confirms that strong correlations between indicators are primarily observed within the same lactation.

Calculating these resilience indicators within a single farm can help farmers better understand the productive life of cows at the end of their lactation. The indicators can also be used to monitor ongoing production and behavior, allowing attention to be paid to animals that show different behavioral patterns. Productive animals are known to be less resilient, depending on how resilience indicators are measured. Since some resilience indicators are constructed such that higher values reflect greater instability or poorer recovery, they are often positively genetically correlated with milk yield (Poppe et al. 2021b). In this context, higher values of these indicators correspond to lower resilience and higher milk production, highlighting an explicit trade-off between resilience and production.

Accurate measures of cow life events, fertility and reproductive traits, together with treatment records, could help to refine the definition of resilience. Additionally, AMS can provide more than just daily milk yield, including longitudinal traits such as feeding behavior and milk composition, which could improve understanding of physiological responses throughout lactation. This was not the aim of this study.



a

	Lnvar	rauto	dPert
rauto	0.3 SE = 0.04		
dPert	0.77 SE = 0.03	0.24 SE = 0.04	
wfPert	0.65 SE = 0.03	0.22 SE = 0.04	0.78 SE = 0.02

b

	Lnvar	rauto	dPert
rauto	0.42 SE = 0.05		
dPert	0.81 SE = 0.03	0.47 SE = 0.05	
wfPert	0.61 SE = 0.04	0.35 SE = 0.05	0.74 SE = 0.04

c

	Lnvar	rauto	dPert
rauto	0.51 SE = 0.07		
dPert	0.85 SE = 0.05	0.59 SE = 0.07	
wfPert	0.64 SE = 0.07	0.45 SE = 0.08	0.76 SE = 0.06

d

	Lnvar_1	rauto_1	dPert_1	wfPert_1	Lnvar_2	rauto_2	dPert_2	wfPert_2	Lnvar_3	rauto_3	dPert_3
rauto_1	0.27 SE = 0.09										
dPert_1	0.67 SE = 0.07	0.23 SE = 0.09									
wfPert_1	0.63 SE = 0.07	0.19 SE = 0.09	0.74 SE = 0.06								
Lnvar_2	0.24 SE = 0.09	0.15 SE = 0.09	0.11 SE = 0.09	0.09 SE = 0.09							
rauto_2	0.01 SE = 0.09	0.29 SE = 0.09	0.06 SE = 0.09	0.14 SE = 0.09	0.53 SE = 0.08						
dPert_2	0.28 SE = 0.09	0.13 SE = 0.09	0.19 SE = 0.09	0.22 SE = 0.09	0.85 SE = 0.05	0.53 SE = 0.08					
wfPert_2	0.21 SE = 0.09	0.02 SE = 0.09	0.17 SE = 0.09	0.22 SE = 0.09	0.53 SE = 0.08	0.33 SE = 0.08	0.67 SE = 0.07				
Lnvar_3	0.12 SE = 0.09	-0.12 SE = 0.09	0.07 SE = 0.09	-0.04 SE = 0.09	0.19 SE = 0.09	-0.06 SE = 0.09	0.13 SE = 0.09	0.05 SE = 0.09			
rauto_3	-0.02 SE = 0.09	-0.05 SE = 0.09	-0.01 SE = 0.09	-0.07 SE = 0.09	0.09 SE = 0.09	0.07 SE = 0.09	0.05 SE = 0.09	0.1 SE = 0.09	0.54 SE = 0.07		
dPert_3	0.11 SE = 0.09	-0.11 SE = 0.09	0.11 SE = 0.09	-0.03 SE = 0.09	0.16 SE = 0.09	-0.07 SE = 0.09	0.16 SE = 0.09	0.05 SE = 0.09	0.85 SE = 0.05	0.59 SE = 0.07	
wfPert_3	0.1 SE = 0.09	-0.02 SE = 0.09	0.16 SE = 0.09	0 SE = 0.09	0.16 SE = 0.09	0 SE = 0.09	0.2 SE = 0.09	0 SE = 0.09	0.66 SE = 0.07	0.44 SE = 0.08	0.77 SE = 0.06

Figure 5.2: Correlation between resilience indicators within lactation one, two and three [respectively, (a–c)] and across lactation (d).

5.4.3 Variance components and heritability

The results of the heritability calculations for the traits are shown in Table 5.3 and briefly discussed below. However, an important premise must be made before the discussion. As the heritability estimates shown here were obtained in only one herd, they may not fully capture the existing population additive genetic variability, even if the sires used in this farm are selected by the owner himself from all available bulls on the market, i.e. from different AI companies, countries and genetic programs, and as such they are a representative sample of the population. Furthermore, since our goal was not to estimate heritability per se or to introduce it into breeding programs, heritability values were retained for comments and to select the resilience indicator to be the subject of the GWAS, also based on the fact that they confirmed already published values, as discussed below. The heritabilities of LnVar in the first and second lactation are both 0.086, which is considerably lower than previous results where 0.13 and 0.15 were estimated for the first lactation and 0.18 and 0.20 for the second lactation (Poppe et al. 2021a; Chen et al. 2023). As shown here and in a previous study, heritability exhibits small variations across parities but does not follow a specific trend, suggesting that the heritability of LnVar may be considered stable throughout a cow's life (Poppe et al. 2021b). For rauto, the estimated values are 0.14 and 0.09 for lactations 1 and 2. These results are slightly higher than those reported previously, ranging from 0.08 for the first lactation to 0.07 for the second one (Poppe et al. 2021a). Our results are also higher than those in another study, where estimates range from 0.04 for the first lactation to 0.02 for the second lactation (Chen et al. 2023). Only one study has estimated the heritability of the resilience indicators dPert and wfPert (Chen et al. 2023). The results we obtained are compared by placing in curly brackets the corresponding values from that study. The heritability of dPert ranges from 0.01 (0.03) in the first lactation to 0.13 (0.06) in the second lactation. Regarding wfPert, the estimated heritabilities obtained here were zero, while another study reported values of 0.04 and 0.02 for the first and second lactation, respectively (Chen et al. 2023).

Table 5.3: Genetic (Gen.) and residual (Res.) variance components and heritability (h^2) for each indicator within lactations 1 and 2.

Indicator	Lactation	Gen. variance (SE)	Res. variance (SE)	Heritability (SE)
LnVar	1	0.002 (0.017)	0.204 (0.019)	0.086 (0.076)
rauto	1	0.0016 (0.0011)	0.0098 (0.0011)	0.143 (0.092)
dPert	1	0.033 (0.261)	5.85 (0.42)	0.01 (0.044)
wfPert	1	0 (0)	7.08 (0.42)	0 (0)
LnVar	2	0.026 (0.038)	0.275 (0.042)	0.086 (0.128)
rauto	2	0.0014 (0.002)	0.014 (0.0022)	0.09 (0.14)
dPert	2	0.895 (1.00)	6.14 (1.05)	0.127 (0.14)
wfPert	2	0 (0)	7.34 (0.63)	0 (0)

5.4.4 Genome-Wide Association Study

As hereinbefore described, GWAS were performed using the selective genotyping approach coupled to the statistics of DNA pooling to investigate the genetic basis of resilience in the population only for those indicators that have shown heritability. The descriptive statistics shown in Table S1 refer to the full RE and NRE groups prior to replicate creation.

For the animals selected for the GWAS (top 10% most resilient – RES and bottom 10% least resilient – NRES), the average gEBVs for milk yield and SCC of the cows were plotted in Figure S1. The figure includes only the resilience indicators selected for GWAS analysis (i.e., LnVar, rauto, and dPert). Interestingly, the analysis revealed that RES animals consistently exhibited lower gEBV values for milk yield and higher values for SCC across the three resilience indicators (Figure S1). This trend was particularly evident for LnVar-based indicators. The results obtained here are in agreement with the findings of Chen et al. (2024), which showed an unfavorable causal association between daily milk yield (DMY) and LnVar.

Results of each GWAS are graphically shown in the Manhattan plots of Figure 5.3 (panels a–e) and reported in Table 5.4. A total of 66, 22, 60, 33, and 8 significant SNPs were identified above the 5% FDR threshold (blue line) for LnVar1, LnVar2, rauto1, rauto2, and dPert2, respectively (Table S2). These SNPs were located in intronic (n = 79), intergenic (n = 99), 5' UTR (n = 2), 3' UTR (n = 4), and intragenic regions (n = 2; one missense and one synonymous). Based on the location of these SNPs, a total of 124 candidate genes were identified for functional interpretation. Out of these, the Animal Genome Cattle Database linked 40 genes to six main "QTL Terms – Trait_class" categories, comprising 77 unique "QTL Trait_Name" entries (Table S2). As shown in Figure S2, milk-related traits were more frequently associated with LnVar1 than with LnVar2. QTL terms associated with LnVar2 were predominantly linked to milk-related traits (87.0%), whereas for LnVar1, only 30.8% were related to milk production. The remaining QTLs associated with LnVar1 were distributed among other categories, including exterior traits (26.2%), reproduction (16.9%), production traits (16.9%), meat and carcass traits (6.2%), and others. This discrepancy may reflect physiological and genetic differences in milk production dynamics between first and second parity. The genetic correlations between milk yield in the first and second lactations have been estimated at approximately rg = 0.82, indicating a strong but not perfect relationship between the two (Dong et al. 1989). This suggests that while a common genetic component influences both lactations, specific factors unique to each parity also contributes to milk production dynamics. The potential roles of some of candidate genes in resilience are then herein below discussed by resilience indicator.

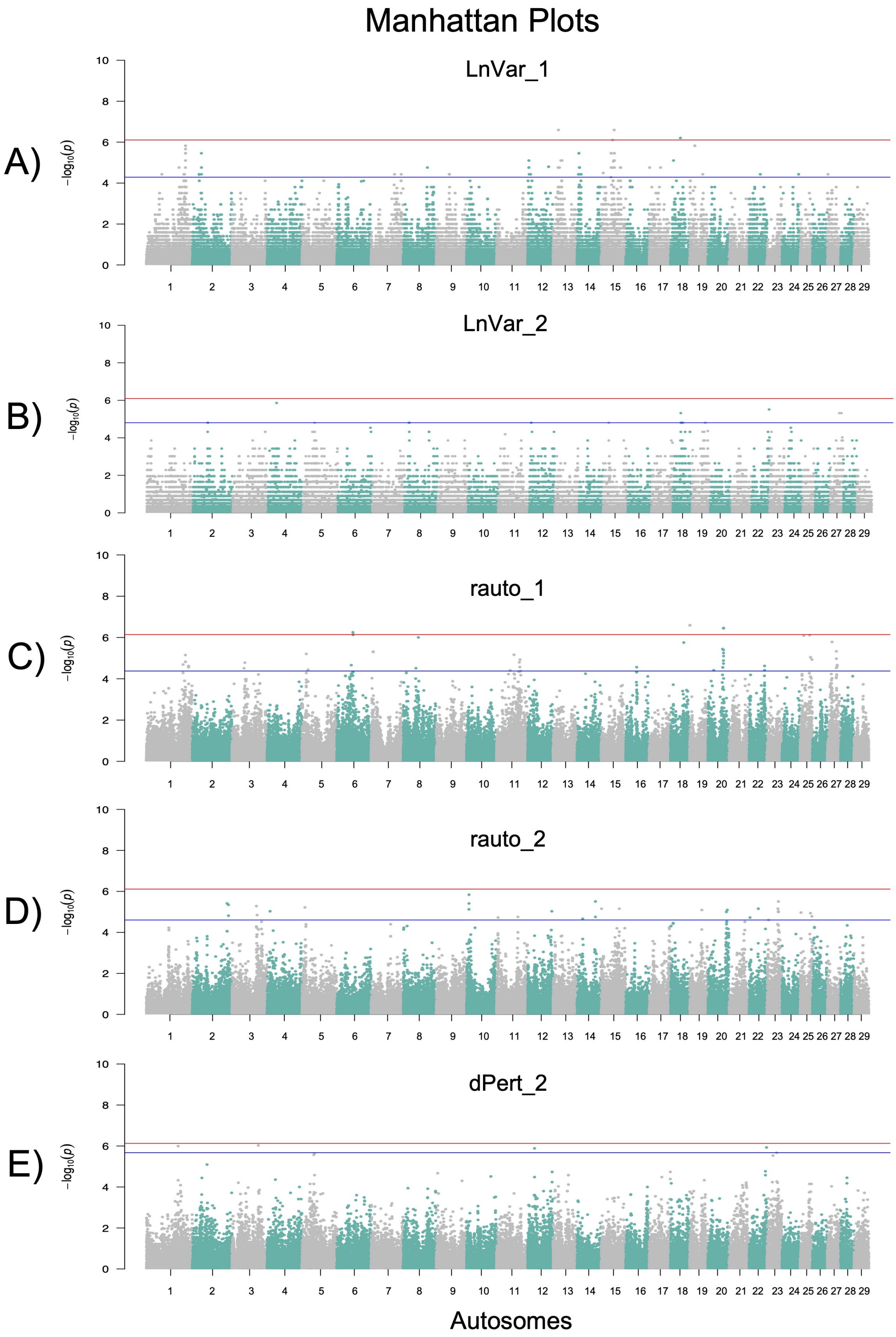


Figure 5.3: Manhattan plots of the GWAS results for the resilience indicators showing heritability: LnVar - first (A) and second lactation (B); rauto - first (C) and second lactation (D); dPert - second lactation. The lines represent the False Discovery Rate (blue) and Bonferroni (red) thresholds, respectively, set at 5% genome-wide.

Table 5.4: Results of the GWAS analysis. The Table lists only SNPs located within genes, grouped by chromosome. When multiple SNPs map to the same gene, the reported P-value (*) corresponds to the average of the individual pvalues. Full results, including all significant SNPs (both genic and intergenic), the corresponding flanking genes, and the associated QTLs for each gene, are provided in Supplementary Table S2.

SNP	Chr	bp_Start	bp_End	P value	Gene	Variant
LnVar1						
Hapmap60237-rs29019540	1	112581930	112581930	1.74E-05	MME	intron
BovineHD0100038618	1	134531830	134531830	1.74E-05	EPHB1	intron
ARS-BFGL-BAC-29669, BovineHD0200009212, Hapmap27145-BTA-116154	2	30980589	31096836	1.94E-05	SCN2A	intron
BovineHD0900012694, BovineHD0900012701	9	45190557	45228937	3.73E-05	HACE1	intron
BovineHD1300005194	13	17948346	17948346	1.74E-05	PDSS1	intron
ARS-BFGL-NGS-17819	13	31353706	31353706	7.91E-06	CUBN	intron
ARS-BFGL-BAC-24839	14	4586387	4586387	3.73E-05	FAM135B	intron
DB-838-seq-rs109494080	14	7881268	7881268	3.73E-05	ST3GAL1	intron
Hapmap53910-rs29021775	14	8266712	8266712	1.74E-05	TG	intron
BovineHD1500001927	15	7384011	7384011	3.21E-05	TRPC6	intron
BovineHD1500009549, ARS-BFGL-NGS-118826, BovineHD1500009559	15	34468497	34497030	1.74E-05	TPH1, SERGEF	intron
BovineHD1500011481	15	41240619	41240619	1.74E-05	GALNT18	intron
BovineHD1500013117	15	45782481	45782481	7.91E-06	ZNF215	intron
BovineHD1500019157	15	65873261	65873261	1.74E-05	SLC1A2	intron
Hapmap53277-rs29025667	18	10305903	10305903	7.91E-06	NECAB2	synonyomus
chr19_18330921	19	17991086	17991086	1.50E-06	RNF135	3' UTR
BovineHD2200010872	22	38049578	38049578	3.73E-05	SYNPR	intron
BovineHD2700000363	27	2286136	2286136	3.73E-05	CSMD1	intron
LnVar2						
BovineHD0400009324, BovineHD0400009326, BovineHD0400009327, BovineHD0400009328, BovineHD0400009334	4	32682347	32692288	1.65E-05	RUNDC3B	intron
BovineHD0500011732, BovineHD0500011741	5	40817911	40874992	9.33E-06	SLC2A13	intron
BovineHD0500012967	5	44906731	44906731	6.09E-06	CPM	intron
BovineHD0500013030	5	45080266	45080266	2.15E-06	NUP107	intron
BovineHD0500013061	5	45169905	45169905	7.27E-07	RAP1B	intron
ARS-BFGL-NGS-19337	19	41942159	41942159	6.09E-06	HAP1	intron
BovineHD1900012186	19	42107233	42107233	1.65E-05	ODAD4	5'UTR
BovineHD2400006407	24	23149184	23149184	1.65E-05	NOL4	intron
rauto1						
BTB-01935567	1	126208399	126208399	2.05E-05	PCOLCE2	intron
ARS-BFGL-NGS-106296, ARS-BFGL-NGS-101030, BovineHD0100038576	1	134365317	134439762	1.24E-05	EPHB1	intron
BovineHD0300013636	3	44311093	44311093	1.66E-05	PLPPR5	intron
BovineHD0500006098	5	21026129	21026129	3.64E-05	DCN	intron
Hapmap25928-BTA-18390	6	55450495	55450495	5.64E-07	ARAP2	intron
BovineHD0800015665	8	51978193	51978193	9.89E-07	PCSK5	intron
Hapmap43807-BTA-21732	11	59822134	59822134	6.86E-06	C11H2orf74	3' UTR
BovineHD1100022461	11	78167522	78167522	1.60E-05	LDAH	intron
Hapmap50260-BTA-121056	16	37673618	37673618	2.75E-05	KIFAP3	intron
ARS-BFGL-NGS-18897	18	45789853	45789853	1.76E-06	HPN	intron
BovineHD1900000212, ARS-BFGL-NGS-92544, BovineHD1900002018	19	1120842	1131844	2.58E-07	CA10	intron
BovineHD2000005959	20	19858203	19858203	3.87E-05	PDE4D	intronic
BovineHD2000014687, BTB-00788976, BovineHD2000014688, BovineHD2000014691, BovineHD2000014724	20	53442839	53557191	1.87E-06	CDH18	intron
ARS-BFGL-NGS-117960	22	52783681	52783681	3.87E-05	TMIE	intron
BovineHD2500003400	25	12057939	12057939	8.16E-07	SHISA9	intron
ARS-BFGL-NGS-59828	25	32764383	32764383	7.77E-07	NCF1	intron
BovineHD2500011569	25	40434604	40434604	1.18E-05	CARD11	intron
BovineHD2700004421	27	16219404	16219404	1.65E-06	FAM149A	5' UTR
BovineHD2700008510, BovineHD2700008511, BovineHD2700008539	27	31017430	31056395	1.51E-05	UNC5D	intron
rauto_2						
BTB-01466122	3	83872828	83872828	5.22E-06	PATJ	intron
BTA-68603-no-rs	3	86312172	86312172	1.43E-05	FGGY	intron
BTA-00702-no-rs	4	10147900	10147900	9.33E-06	CDK6	intron
ARS-BFGL-NGS-69509	5	10190011	10190011	6.04E-06	PTPRQ	intron
ARS-BFGL-NGS-108825, chr10_7895525	10	7960731	7961933	5.70E-06	F2R	synonymous, 3' UTR
BovineHD1100020894	11	73073273	73073273	1.76E-05	OTOF	missense
BovineHD1200025294	12	82887622	82887622	9.33E-06	FAM155A	intron
BovineHD1400018052, BovineHD1400018067	14	62668956	62692420	1.04E-05	NCALD	intron
BovineHD1500018110	15	62392291	62392291	6.99E-06	ELP4	intron
BovineHD2200000724	22	2715544	2715544	1.89E-05	CMC1	intron
Hapmap51784-BTA-97575	23	5530326	5530326	2.49E-05	FAM83B	intron
BovineHD2300011375, BovineHD2300011378, BovineHD2300011379, BovineHD2300011381, BovineHD2300011380	23	39585474	39610535	8.49E-06	KIF13A	intron
ARS-BFGL-NGS-74596	25	35220858	35220858	1.16E-05	COL26A1	3' UTR
Hapmap51005-BTA-60474	25	40111236	40111236	1.64E-05	SDK1	intron
dPert_2						
BovineHD0100031225	1	109410007	109410007	1.02E-06	RSRC1	intron
BovineHD0500011732	5	40817911	40817911	2.70E-06	SLC2A13	intron
BovineHD0500012310	5	42830228	42830228	2.30E-06	PTPRR	intron
BovineHD2200017497	22	59585131	59585131	1.18E-06	KBTBD12	intron

LnVar

For LnVar1, on BTA 1 a total of 9 significant SNPs were above the FDR threshold (Figure 5.3a). Among them, rs110792885 is located in intron position of the *EPHB1* gene (EPH Receptor B1), one of the Eph (erythropoietin-producing hepatocellular carcinoma) receptors representing one of the largest known family of receptor tyrosine kinases in mammals: there is evidence the Eph receptors and ephrin ligands may mediate immune cell activation and the immune cell trafficking required for optimal functioning of immune system (Darling et al. 2019). Given that Eph receptors are involved in various normal cellular processes during development and play a crucial role in maintaining adult tissue homeostasis, their role in both non-infectious and infectious diseases is well established. As a result, the pathways in which they operate may indirectly impact stress responses and resilience traits. The *EPHB1* gene was also associated with udder traits (udder suspension and teat score) in Angus cattle (Devani et al. 2020). On BTA 2, three above FDR threshold SNP are in intron sequences of *SCN2A* and *SCN3A* genes, but none of their functions that may be related to resilience has been reported to date. The SNP rs110164494, is located in the intron of the PDSS1 gene on BTA 13. *PDSS1* encodes the enzyme decaprenyl diphosphate synthase subunit 1, which is involved in the synthesis of coenzyme Q (ubiquinone), a key component of mitochondrial electron transport and cellular energy production. The association of PDSS1 with female fertility traits has been highlighted in previous studies, such as Mohammadi et al. (2020), who identified it in a GWAS focused on Iranian Holstein cattle. This study found that PDSS1 could play a role in female fertility through pathways linked to cellular energy metabolism, which is crucial for reproductive performance and resilience during energy-intensive periods like lactation. On *BTA14* at about 6 Mbp, two intergenic markers (rs110769987 and rs42211697) were found to be significantly associated with LnVar1 and were located closely to *KHDRBS3*. This gene has previously been associated with milk production traits in both Chinese Holstein and crossbred populations (Jiang et al. 2010; Cruz et al. 2020).

The *ST3GAL1* gene was also found to be associated with milk traits, including milk yield, milk fat yield, and milk fat percentage (Wickramasinghe et al. 2011). Thyroglobulin gene (TG), identified as candidate gene on BTA14 (rs29021775), is a key glycoprotein involved in the synthesis of thyroid hormones. TG plays an important role in different physiological processes including regulating metabolism, adipocyte growth, differentiation and homeostasis of fat depots (Dubey et al. 2014). Given its function TG can be easily linked to resilience. In fact, studies in rat and livestock have demonstrated that thyroid function, influenced by TG, is critical for coping with temperature extremes, supporting its role in resilience to environmental stressors (Sejian et al. 2018; Rial-Pensado et al. 2022). At about 82 Kbp from rs41629530 on BTA 15 is located the *GRAMD1B* gene, part of the *GRAM* (glucosyltransferases, Rab-like GTPase activators and myotubularin) domaincontaining gene group, involved in maintaining cholesterol homeostasis, apoptosis and cancer (Yang et al. 2011). In livestock, *GRAMD1B* was already associated with feed efficiency, production and reproduction traits in cattle (Kunej et al. 2024). *SERGEF* at positions about 34 Mbp (identified as candidate gene by rs110010916 and rs109595542) has been linked to pig adaptation to high-altitude conditions, suggesting potential resilience to environmental stressors that may indirectly influence lactation stability (Dong et al. 2014). The significant rs110249272 maps in the 3'UTR region of the *RNF135* gene (Ring Finger Protein 135) involved in the activation of the NF-kB signaling pathway, which is essential for immune responses, inflammation regulation, and cell survival (Kiser et al. 2018). In cattle, activation of this pathway helps increase resilience to infections like Mycobacterium avium paratuberculosis (Map), the pathogen causing Johne's disease. The activation of NF-kB by these proteins boosts the cattle's resilience by ensuring the survival of immune cells, especially monocytes, which are key to fighting infection. This ability to withstand and adapt to immune challenges contributes to overall health and disease resistance (Calderón-Chagoya et al. 2023).

For LnVar2, the major part of candidate genes seems to be involved both in functional and productive traits (as reported in Table S3, QTL). In addition, on BTA 4, all the significant SNPs mapped in the intronic position of *RUNDC3B* gene, that is considered a backfat gene (Yang et al. 2011). As of now, the *RUNDC3B* gene has not been directly associated with backfat thickness in Holstein cattle. However, it is well known that backfat genes have role in body composition, energy reserves, and overall productivity (Schmidtmann et al. 2024). We may speculate that genes influencing fat deposition (backfat genes) might relate to resilience indirectly, as fat reserves can buffer energy deficits during stress, thereby supporting resilience. The gene *Rap1b*, identified as a candidate by the SNP rs133340933, belongs to the Ras superfamily, a group of proteins involved in regulating B-cell development, homing, and T-cell-dependent humoral immunity (Chu et al. 2008). It is well known that adaptive immune response, characterized by its specificity and memory, plays a critical role in maintaining long-term resilience against pathogens (Alotiby 2024). Two significant SNPs were annotated in intronic positions of *SLC2A13*, one of the glucoses transporters, resulted associated with milk, protein and fat yields in Buffaloes species (Du et al. 2019). Even the gene *NOL4* seems to be involved in milk traits according with was reported in different cattle breeds (Bekele et al. 2023).

rauto

Regarding rauto1, the QTL region at about 134 Mbp and defined by three significant SNPs (in intronic position of the EPHB1 gene) overlaps the one identified for LnVa1. Based on its functions (as described in the LnVar1 GWAS results), we hypothesize that for rauto (reflecting stability and recovery after perturbation) *EPHB1* could regulate pathways that maintain homeostasis under stress by guaranteeing robust intercellular signaling. Instead, for LnVar1 (which measures variability, where lower variability suggests more robust responses to environmental or physiological stress) *EPHB1* might reduce variability by ensuring consistent signaling pathways, supporting immune responses across diverse environmental challenges, and thereby contributing to more consistent physiological outcomes.

On BTA 1, a second significant region is defined by three SNPs, all mapping close to the *POFUT2* gene (max distance 59 Kbp). This gene has been associated both with female fertility in Nordic dairy cattle (Mesbah-Uddin et al. 2022) and with body conformation traits in Holstein (Wang et al. 2022b). On BTA 3, the genetic variant rs133042560 is located in the intronic position of *DCN* gene, that encodes for decorin, a small leucine-rich proteoglycan involved in connective tissue structure. Decorin binds collagen fibrils and regulates collagen assembly, influencing fibril uniformity (Khatib 2005). This function is essential for tissue integrity and may impact resilience, particularly in response to metabolic and physical changes during early lactation. *DCN* is maternally expressed in placental tissue in mice, highlighting its potential role in tissue development and adaptation, which could be relevant for bovine resilience during the first lactation (Mizuno et al. 2002).

On BTA 11, two significant SNPs are located near the *IL1B* gene. The IL1B gene, as a member of the interleukin-1 (*IL-1*) family, plays a crucial role in inflammation and immune responses (Moghaddam et al. 2019). Its expression is upregulated in milk somatic cells as part of the immune response during udder infections (Lee et al. 2006). Additionally, increased expression of IL1B has been reported in response to infectious agents, such as Mycoplasma bovis and Klebsiella pneumoniae (Bannerman et al. 2004; Kauf et al. 2007). The *IL1B* gene is also associated with bovine respiratory disease susceptibility (Tizioto et al. 2015), a key trait impacting animal health and productivity (Neupane et al. 2018). The rs135712530 is an intronic variant of the *LDAH* gene. *LDAH* has been associated with hoof and leg disorders (Wu et al. 2016), suggesting a potential role in health traits. Hoof and leg issues are critical in livestock, affecting both productivity and welfare, indicating the relevance of this locus for improving resilience and health in cattle. It may influence rauto1 by modulating the stability of physiological responses under stress.

On BTA 20, the SNP rs109908751, located in the intergenic region near the *PDE4D* may be associated with resilience. The expression of *PDE4D* has been detected in mammary glands, indicating its potential involvement in milk production. In fact, previous studies have suggested a possible role of *PDE4* in the regulation of mammary gland function and lactation (Dostaler-Touchette et al. 2009). This function may be related to rauto1, as *PDE4D* could influence the consistency and stability of physiological responses during early lactation. In fact, the *PDE4D* gene is part of the *PDE4* family, involved in regulating cAMP signaling pathways, which are critical for cell desensitization, signal compartmentalization, and cross-talk between cellular signals. By maintaining cAMP homeostasis, *PDE4D* plays a key role in regulating various physiological processes (Dostaler-Touchette et al. 2009).

The regulation of cAMP signaling by *PDE4D* could also contribute to the stability of immune responses and metabolic adaptations, both critical for resilience in dairy cattle during the early lactation period.

Another SNP, rs42070678, is located in an intronic region of *NCF1* on BTA 25. This gene encodes a cytosolic subunit of neutrophil NADPH oxidase, which plays a critical role in the production of reactive oxygen species (ROS). ROS are key mediators in host defense and the regulation of inflammation (Kennedy et al. 2009). Although this gene has not been directly associated with a known QTL, the role of *NCF1* in modulating ROS production suggests it may be involved in the immune responses and inflammatory control, both of which are crucial for maintaining resilience during the early lactation period. Also, the *CARD11* gene in which rs109938921 maps, is involved in immune signaling, being crucial in the activation of T-cells and the differentiation of peripheral B-cells (Stepensky et al. 2013). In dairy cattle, *CARD11* has been linked to feed efficiency (FE), with studies showing its involvement in residual feed intake (RFI) in Danish Holstein cattle. *CARD11* was downregulated in animals with high RFI compared to those with low RFI, suggesting its role in regulating energy balance and metabolic efficiency (Salleh et al. 2017). *CARD11* may then influence resilience to metabolic and immune stress during early lactation.

On BTA 20, a genomic region at approximately 53.4 Kbp is defined by 9 significant SNPs, including 5 intronic variants and 4 intergenic ones (see Table 5.3). This region harbors the *CDH18* gene, which belongs to the canonical cadherin (*CDH*) gene family. The cadherin family is composed of a series of cell adhesion molecules that play a dominant role in tissue morphogenesis and regulate adhesion interactions. Some studies have shown that the *CDH18* gene locus resulted strongly associated with milk and fat yields in dairy cattle (Laodim et al. 2017).

Finally, three significant SNPs (rs133164649, rs42130478, and rs109905892) are all located in the intron position of the *UNC5D* (BTA 27). *UNC5D* is implicated in the development and maintenance of udder structure and conformation (Cole et al. 2011), which are crucial for efficient milk production and the animal's ability to cope with physiological stress during lactation. Rauto1 may reflect how well an animal's udder structure and other related traits maintain stability during early lactation.

The genes associated with rauto2 are involved in various traits, as described by the examples reported below. On BTA 3, two significant SNPs were located in the intronic regions of the *PATJ* and *FGGY* genes. These genes have been reported to be associated with fertility and reproduction traits, as well as body structure and finishing precocity, respectively. *CDK6* gene can be considered a candidate gene involved in body traits (Liu et al. 2011; Silva et al. 2019). *PTPRQ* on BTA 5, as identified by Robakowska-Hyżorek et al. (2016), may influence meat production traits in beef cattle, possibly through the regulation of *MRF* (myogenic regulatory factors) gene expression. Two closed SNPs on BTA 10 lie in the coding sequence (synonymous and 3'UTR) of the *F2R* gene, proposed as novel and promising candidates for regulation of hypoxic adaptation in the heart by Wang et al. (2021), a study that compared the hypoxic adaptation of the yak (*Bos grunniens*) against different cattle species.

On BTA 14, rs110970186 and rs43430961 are annotated in intronic position of the *NCALD* (Neurocalcin Delta) gene that was associated with the Bovine Respiratory Disease (Kiser et al. 2017). On BTA 25, rs42073064, located in the 3'UTR region of *COL26A1*, which encodes collagen type XXVI, was identified as one of the differentially expressed genes potentially involved in host resistance against ticks (Mantilla Valdivieso et al. 2022). *KRT14*, located 193 bp from the rs43727762 SNP, plays a role in mammary epithelial cell lineage changes, which are essential for the proper development of the mammary epithelium during the cow's life and, consequently, for milk production (Finot et al. 2019). The same authors described also the role of *KRT14* (together with other cell line) in the development of the bovine mammary gland at puberty (Finot et al. 2018).

Finally, we found five significant SNPs mapping in the intronic position of *KIF13A*. This gene belongs to the kinesin superfamily, a large group of motor proteins involved in intracellular transport and recycling endosome dynamics. These functions are crucial for maintaining cellular homeostasis and responding to environmental stressors, particularly when *KIF13A* interacts with other proteins such as Rab GTPases (Thankachan et al. 2022). The interaction between these two classes of proteins could be a key mechanism in stress adaptation and recovery. Therefore, we may speculate that this represents an indirect link with rauto, which measures autoregulation – the ability of a system to autonomously regulate itself without external intervention.

dPert

Eight above FDR threshold SNPs sparse along the chromosomes were associated with dPert2. One of these SNPs was already associated with LnVar2 (rs133894374 annotated in intronic position of *SLC2A13*). As reported here and by Chen et al. (2023), LnVar2 and dPert2 were highly correlated (0.85, see Figure 5.2d). This strong association is expected, as both metrics reflect an animal's ability to maintain stable performance despite challenges.

The PTPRR gene harbors the intronic rs43440584. The protein encoded by *PTPRR* is a member of the protein tyrosine phosphatase (PTP) family and appears to be involved in mammary gland involution, possibly contributing to the remodelling of udder tissue for subsequent parturitions. Researchers reported that weaning (in mice) increased PTP activity in the mammary gland (Tolleson et al. 2017). *PTPRR* gene has also been associated with various mammary traits in different cattle breeds (Tolleson et al. 2017; Sinha et al. 2023).

The rs43219764 and rs132905517 are intronic SNPs of *RSRC1* and *KBTBD12* genes, that were associated with reproduction traits and with milking temperament in Holstein cattle, respectively (Chen et al. 2020; Grigoletto et al. 2020). Opposite genetic correlation between immune response traits (the most ones related to resilience) and fertility traits are reported by König and May (König et al. 2019). In their review, the gestation length resulted positively and negatively correlated with antibody- and cell-mediated immune response, respectively (+0.17; -0.17). The same opposite correlation values were found also for other fertility traits including calving ease, maternal calving ease, and daughter fertility (König et al. 2019). The potential link between milk temperament and resilience in milk production suggests that also behavioral traits might indirectly influence the physiological stability of lactating cows. However, some evidences are contrasting: i) Stepancheva et al. (2024), investigating how milking temperament affects milk productivity, found that Buffalos with higher milking behavior scores (4 or 5, more reactive cows) had the quite higher LS means for TDMY; ii) Marçal-Pedroza et al. (2023) reported that calm and intermediate cows produced more milk, had shorter milking times and a greater average milk flow; iii) Antanaitis et al. (2021) observed a negative genetic correlation between the temperament of cows and milk yield; instead, temperament was positively correlated with SCS.

5.5 Conclusions

This study offers a high-resolution analysis of resilience indicators in Holstein cows, leveraging daily milk yield data from automatic milking systems collected under standardized management and environmental conditions. The four indicators assessed (LnVar, rauto, dPert, and wfPert) captured different resilience dimensions. LnVar showed the most robust biological and genetic signals, particularly in the second parity. Rauto showed moderate heritability and improved biological coherence with age, while dPert and wfPert, despite lower heritabilities, provided valuable insights into the dynamics of short-term production perturbations.

Genome-wide association studies identified a complex genetic basis for resilience, involving immune function, metabolic regulation, and tissue integrity. Candidate genes such as *EPHB1*, *IL1B*, *PDSS1*, *GRAMD1B*, and *DCN* were associated with processes including inflammation, energy homeostasis, and extracellular matrix remodelling. The *EPHB1* and *SLC2A13* genes were linked to multiple indicators or parities, suggesting shared regulatory mechanisms, while others appeared only in later parities, pointing to age-related physiological adaptations. Several genes have also been previously associated with production traits, supporting potential pleiotropy and the importance of considering resilience in breeding decisions.

Working within a single, large and well-monitored herd minimized environmental variability, allowing clearer detection of individual differences and genetic signals. This approach delivers practical value to farmers by supporting herd management decisions, especially for low-resilience cows, and informing breeding strategies that prioritize resilient phenotypes. Future studies across diverse herds and environments, enriched with health records and external stressor data (e.g., disease, heat), may improve and validate these findings and further disentangle intrinsic resilience from environmental effects.

Data availability statement All the procedures were approved by the Animal Welfare Body of the Università degli Studi di Milano (OPBA) and by the Italian Minister of Health (protocol number OPBA_68_2023). The study was conducted in accordance with the local legislation and institutional requirements.

Ethics statement No animal care committee approval was necessary for the purposes of this study, as all genotypes and data were available from the pre-existing database of the GENORIP project, funded by the Lombardy Region.

5.6 Supplementary material

Table S1. Summary statistics of resilience indicators belonging to the resilient cows (RE) and not resilient cows (NRE) selected for the GWAS.

	RE					NRE				
Indicator	Obs	Mean	SD	Min	Max	Obs	Mean	SD	Min	Max
LnVar_1	52	1.04	0.22	0.44	1.30	52	2.79	0.21	2.52	3.40
rauto_1	60	0.42	0.05	0.24	0.47	55	0.80	0.04	0.76	0.91
LnVar_2	30	1.33	0.27	0.64	1.67	30	3.40	0.29	3.05	4.25
rauto_2	32	0.40	0.06	0.15	0.46	31	0.85	0.03	0.81	0.94
dPert_2	29	15.37	0.84	13.17	16.29	30	24.85	1.53	22.97	28.89

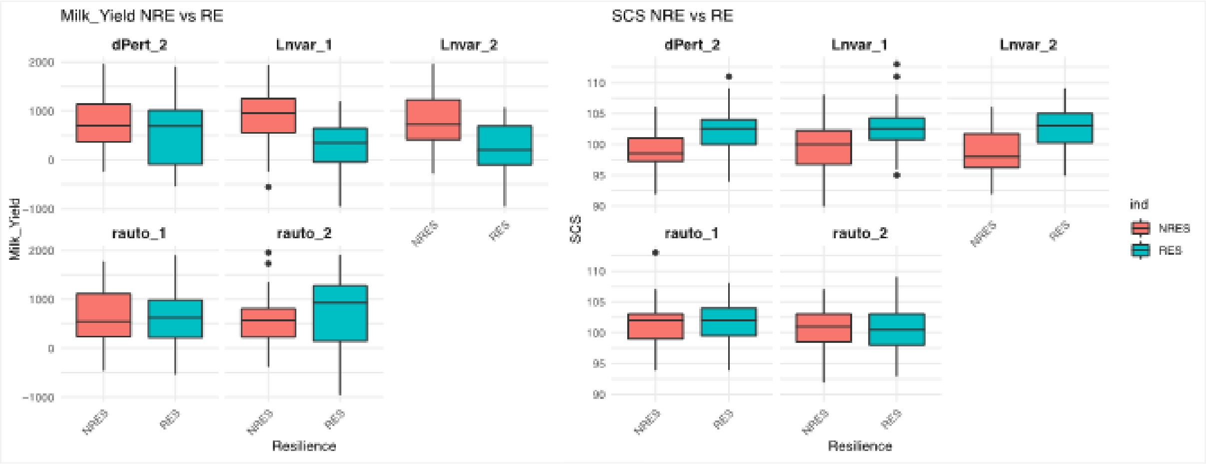


Figure S1.Boxplots of genomic selection indexes for milk yield and Somatic Cell Score (SCC), reported for the most resilient (RE) and least resilient (NRE) cows according to each resilience indicator used in the GWAS.

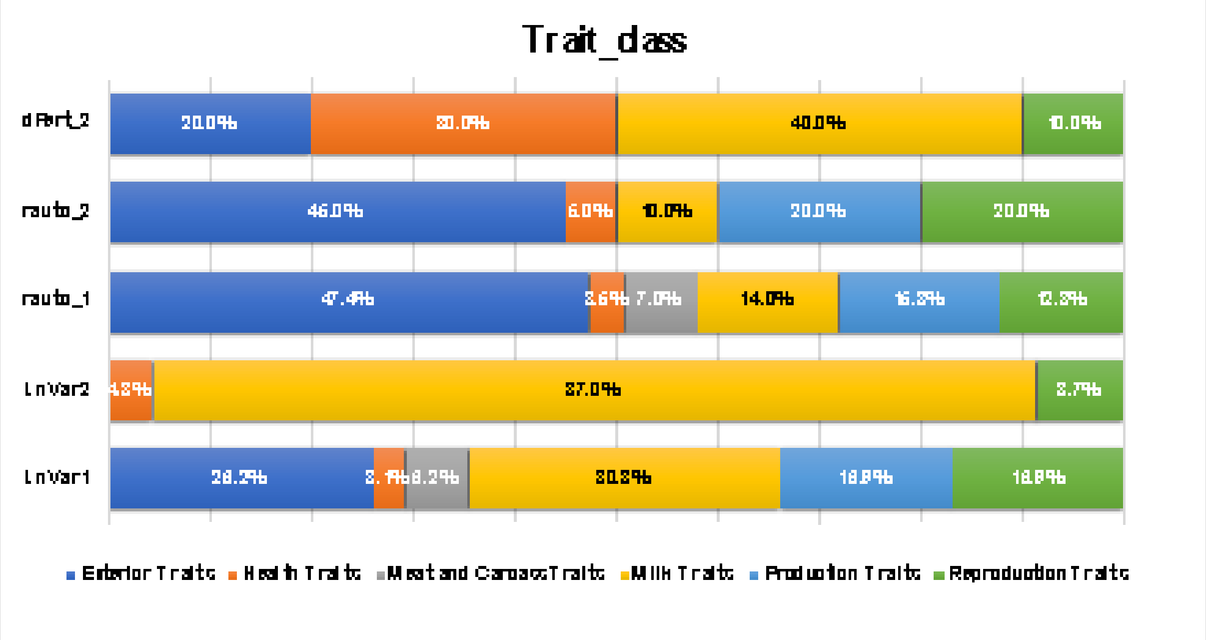


Figure S2. Proportion of QTL_Terms (trait class) for each of indicators of resilience.

Due to the large size of the table, Table S2 is not included in the thesis but is available in the published paper at `https://www.frontiersin.org/articles/10.3389/fanim.2025.1627086/full`.

Chapter 6

Copy number variant scan in more than four thousand Holstein cows bred in Lombardy, Italy

Andrea Delledonne[a], Chiara Punturiero[a], Carlotta Ferrari[a], Francesca Bernini[a], Raffaella Milanesi[a], Alessandro Bagnato[a], Maria G. Strillacci[a]

[a]*Department of Veterinary Medicine and Animal Science, Università degli Studi di Milano*

Published on PLOS ONE, May 21, 2024
`https://doi.org/10.1371/journal.pone.0303044`

6.1 Abstract

Copy Number Variants (CNV) are modifications affecting the genome sequence of DNA, for instance, they can be duplications or deletions of a considerable number of base pairs (i.e., greater than 1,000 bp and up to millions of bp). Their impact on the variation of the phenotypic traits has been widely demonstrated. In addition, CNV are a class of markers useful to identify the genetic biodiversity among populations related to adaptation to the environment. The aim of this study was to detect CNV in more than four thousand Holstein cows, using information derived by a genotyping done with the GGP (GeneSeek Genomic Profiler) bovine 100K SNP chip. To detect CNV the SVS 8.9 software was used, then CNV regions (CNVR) were detected. A total of 123,814 CNV (4,150 non redundant) were called and aggregated into 1,397 CNVR. The PCA results obtained using the CNV information showed that there is some variability among animals. For many genes annotated within the CNVR, the role in immune response is well known, as well as their association with important and economic traits object of selection in Holstein, such as milk production and quality, udder conformation and body morphology. Comparison with reference revealed unique CNVR of the Holstein breed, and others in common with Jersey and Brown. The information regarding CNV represents a valuable resource to understand how this class of markers may improve the accuracy in prediction of genomic

value, nowadays solely based on SNPs markers.

6.2 Introduction

For millennia, humans have established a profound relationship with cattle, domesticating them to exploit their resources, obtain food as milk and meat, and meet various needs (Frantz et al. 2020). Since the 20th century, the selection to improve production traits in animal species, such as the Holstein cattle breed, represents a fundamental step in the development of modern animal husbandry. The Holstein breed, nowadays worldwide recognized for its milk production, has undergone a strong selection effort aimed at improving milk yield, quality, and in the last two decades in enhancing overall functionality and health (Egger-Danner et al. 2015). In recent years, the evolution of nanotechnology made available the SNP genotyping platforms that made possible the genomic selection revolution in cattle breeding theorized by Meuwissen et al. (Meuwissen et al. 2001). The utilization of SNP chips in genotyping has proven to be a potent tool in animal selection, empowering breeders to make well-informed decisions based on the collective genetic information (Wiggans et al. 2017). SNP genotyping data also enable the detection of Copy Number Variants (CNV) through the computation of the Log R Ratio (LRR) and B Allele Frequency (BAF). LRR represents a normalized measure of the total signal intensity for two alleles of a SNP, BAF, the one measuring the allelic intensity ratio at marker level (Wang et al. 2008). The LRR and BAF facilitate the assessment of CNV status (loss vs gain, LRR; homozygote vs heterozygote, BAF). CNV represent a category of genomic structural variants recognized to influence phenotypic diversity through the deletion (loss status) or duplication (gain status) of DNA segments, potentially affecting gene structure and regulating expression (Stranger et al. 2007; Margareto et al. 2009). These variations typically range in size from 1 kilobase (kb) to 5 megabase (Mb) (Mills et al. 2011).

The functional impact of CNV has been studied across various animal species, highlighting their role in influencing a range of traits (Wright et al. 2009; Olsson et al. 2011; Kropatsch et al. 2013; Venhoranta et al. 2013; Awasthi Mishra et al. 2017). The fact that CNV affect a multitude of traits across different animal species underlines their role also in adaptive responses to various environmental conditions (Pierce et al. 2018; Xu et al. 2016; Arendt et al. 2016; Perry et al. 2007). In several studies on Holstein cattle, CNVR have been identified to impact economically important traits as milk production, residual feed intake, fertility and somatic cell score (Hou et al. 2012; Glick et al. 2011; Xu et al. 2014; Durán Aguilar et al. 2017b).

Although CNVR cover a small part of bovine genome length (about 2–10%), as reported by Hay et al. (Hay et al. 2018), these structural variants can be integrated with SNP information in genomic prediction, offering new insights to explain complex traits and understand the proportion of missing heritability not explained by SNP.

Thus, taking into account all information related to Copy Number Variations (CNV), the objectives of this study were to examine a substantial population comprising 4,282 Holstein cows from seven distinct farms in Italy, with the purpose of mapping CNV across the autosomal genome. Additionally, within the more frequent CNVR, the goal encompassed the annotation of genes and of quantitative trait loci (QTL) associated with relevant traits in this breed. To validate our findings, we conducted a comparative analysis both within and across different cattle breeds, drawing on insights from prior research studies.

6.3 Materials and methods

6.3.1 Animal sampling, genotyping and ethics statement

All cows of 7 herds of the Lombardy region were genotyped with the Illumina GGP Bovine 100K (GeneSeek) from 2019 to 2023 for a total of 4,282 individuals. These seven herds are representative of the possible farming systems and selection objectives of Holstein farmers: they in fact span from a small family run farm (110 cows in lactation) with historically low selection, to a large farm with Automatic Milking Systems and with more than three decades of directional selection to improve production and functionality (about 550 lactating cows) and a medium size farm producing Parmigiano Reggiano cheese and thus, requiring specific nutritional practices (no silage) and selection for milk quality. Log R Ratio (LRR) available from the SNP chip processing were used to map CNV. The quality assessment of LRR and the mapping of CNV was performed with the Golden Helix Inc. SVS 8.9 software (SVS). The sampling of individuals was approved by the OPBA (i.e., Animal Welfare Organisation) of the University of Milan (Protocol number 160_2019), by Directive 2010/63/EU of the European Parliament and the Council of 22 September 2010, updating Directive 86/609/EEC on the protection of animals used for scientific purposes.

6.3.2 Quality control of genotyping data

The quality assessment of LRR values was performed considering the Derivative Log Ratio Spread (DLRS) as described by Pinto et al. (2011) and the GC Wave Factor (GCWF) (Diskin et al. 2008), both affecting signal intensity and possible cause of bias in CNV mapping. A total of 47 samples were excluded due to their high DLRS values, while other 135 samples were excluded because of the elevated GCWF values. The detection of CNV was then conducted on a dataset of 4,100 samples.

6.3.3 CNV and CNVR detection

CNV detection was obtained on autosomes with SNPs mapped on the ARS UCD1.2 assembly reference genome. The detection was performed using the Copy Number Analysis

Module (CNAM) of SVS by means of the univariate analysis based on LRR values. Default parameters for CNV calling in CNAM were set as follows: i) a maximum of 100 segments per 10,000 markers; ii) a minimum of 3 markers per segment; iii) 2,000 permutations per pair with a p-value cut-off of 0.005.

To identify animals with outlier CNV frequencies and length, their distributions were analysed using QQ plots (R routine in ggplot2 library (Wickham 2016)). Outliers were identified as samples having CNV length greater than 7.5 Mbp. After the identification and exclusion of the individuals considered outliers (3,809 subjects were left), the individual frequency of gain and loss in relation to each sample mean CNV length was plotted with the ggplot2 library of R.

Using the Bedtools -mergeBed command (Quinlan et al. 2010), CNV that overlapped by at least one bp and were shared by a minimum of two animals were combined to generate CNV regions (CNVR). Then, CNVR were classified as gain, loss, or complex if comprising both deletions (loss) and duplications (gain). A CNV found in a single individual was classified as a singleton CNVR. To be representative, only CNVR shared by at least 2% of the population were selected for descriptive statistics as well as for downstream analyses. The R package HandyCNV (Zhou et al. 2021) was used to visualize the physical distribution of CNVR on autosomes.

6.3.4 Genes and QTL annotations

The genes list with official "gene name ID" was downloaded from NCBI online Database. Genes were then annotated within the detected CNVR using the Bedtools "-intersectBed" command (Quinlan et al. 2010), while the QTL associated with the genes found in the CNVR were identified thanks to the cattle QTL database (`https://www.animalgenome.org/cgi-bin/QTLdb/BT/search`) by gene name, using the "Search by associated gene" option of QTLdb.

The Cytoscape plugin ClueGo was used to identify potential biological connections among candidate genes identified in the CNVR (Shannon et al. 2003; Bindea et al. 2009). The network construction relied on information from GO and KEGG database. This analysis utilized the bovine databases integrated into the ClueGO app. Only connections with a p-value lower than 0.05 were considered.

6.3.5 Diversity at the population level

To study the diversity within the breed we recoded CNV defining a CNVR for each cow as follows: i) '1' for loss state; ii) '0' for normal state; iii) '2' for gain state. We used the Past 4.03 software to perform a principal component analysis (PCA).

6.3.6 Comparison with results from the literature

Our identified CNVR were compared with the results reported in recent literature studies using the HandyCNV library of R-Studio software (compare_cnvr() function). As reported in Table 6.3, two distinct comparisons were performed in order to validate Holstein specific CNVR (comparison within breed), and to identify genomic regions shared by different breeds (comparison among breeds), i.e. Jersey (JER) and Brown Swiss (BSW). For studies with CNVR using a different genome assembly from ARS-UCD1.2, the positions were remapped using the UCSC Lift Genome Annotations tool (`https://genome.ucsc.edu/cgi-bin/hgLiftOver`). A graphical visualization of overlapped CNVR was realized through a Venn diagram built using an online tool (`http://bioinformatics.psb.ugent.be/webtools/Venn/`).

6.4 Results

6.4.1 CNV and CNVR detections

According to the number of CNV per cow and their total length (sum of each CNV length), 291 samples were identified as outliers and subsequently removed to avoid the introduction of possible false positive CNV; the final dataset comprising 123,814 CNV was obtained in 3,809 cows (see Supplementary Table S1); with a total of 4,150 non-redundant CNV.

As reported in Table 6.1, CNV have a maximum, minimum, and average length of 1,860,579, 1,005 and 86,166 bp, respectively. The frequency of loss CNV doubles the frequency of gain CNV and the mean length of losses (90,439) is longer than the mean length of gains (77,785).

Figure 6.1A shows the different distribution of gain and loss CNV according to the relationship between the CNV mean length and their number per samples. Furthermore, as shown in Figure 6.1B, the majority of CNV falls into the first three classes of length. Over 30,000 loss state CNV exhibited a length below 0.05, falling in the first length class. Conversely, the majority of gain CNV had a length ranging between 0.05 and 1 Mb. The longest CNV were low represented for both of CNV states.

The 123,814 CNV were aggregated into 1,397 CNVR (Table 6.2 and Supplementary Table S2), covering 9.18% (228 Mbp) of the total autosomal length (2,489 Mbp). After removing singletons and CNVR shared by less than 2% of the population, 267 CNVR remained (CNVR_2% in Table 6.2 and Supplementary Table S2): 76 in gain state, 129 in loss and 62 categorized as complex. CNV in CNVR_2% are listed in the Supplementary Table S2. These latter CNVR cover 2.92% of the autosomal genome length and their physical distribution on autosomes is shown according to their states in Figure 6.2. Values (%) on this graph represent the genomic proportion covered by CNVR with respect to each chromosome length. CNVR on chromosomes 12, 18 and 23 covered more than 5% of chromosomal length, 9.5%, 7.4% and 5.1% respectively, while all other chromosomes were impacted by a lower proportion of CNVR. The CNVR shared by the largest number of cows were on BTA 10 at 22,676,353 bp (n. 3,528 cows, loss) and on BTA 2 at 93,926,090 (n. 3,107 cows, loss). Instead, CNVR shared by the lowest number of cows, i.e. 76 animals, were found in gain state within chromosome 20 (at 66,818,777 bp).

Table 6.1: Descriptive statistics of identified CNV.

Statistic	Value
N. CNV	123,814
N. Loss	82,258
Min-Max CNV per ID (mean)	13–51 (32.5)
Min-Max (mean) coverage per ID*	947–7,483 (2,792)

*Values expressed in Mbp.

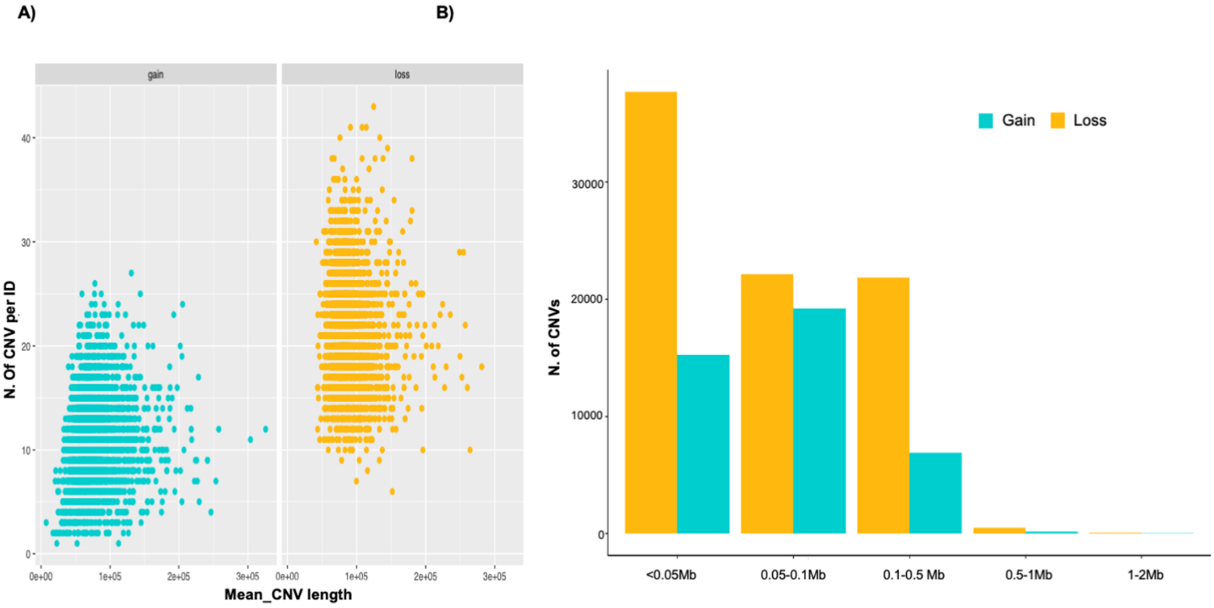


Figure 6.1: Summary statistics for detected CNV. A) Relationship between number and mean total length (bp) of CNV identified in each sample by state (gain vs loss); B) Number of CNV for five classes of length.

Table 6.2: Descriptive statistics of identified CNVR.

			CNVR State			CNVR length		
CNVR	Tot n. CNVR	Tot n. Singleton	Loss	Gain	Complex	Min	Max	Mean
CNVR	1,397	329	714	513	170	1,005	2,286,232	163,678
CNVR_2%	267	-	129	76	62	1,716	2,286,232	272,307

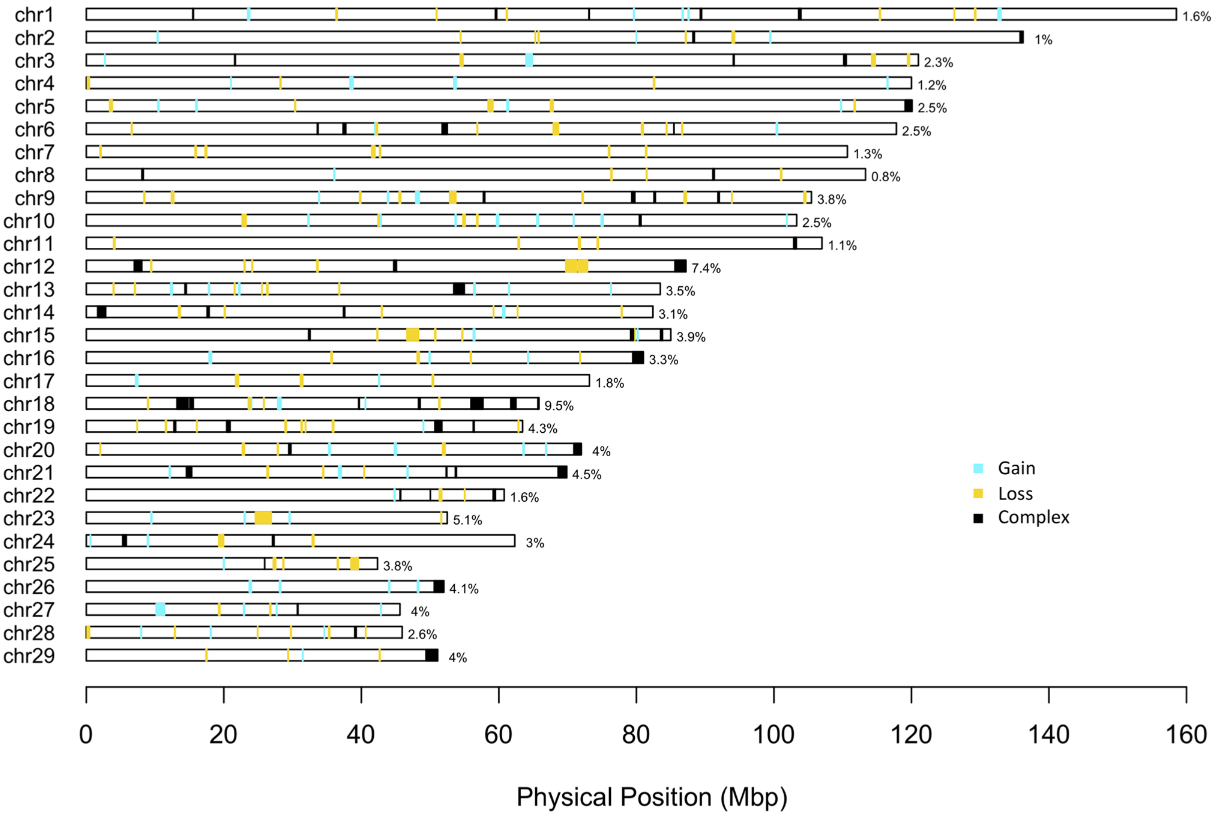


Figure 6.2: Physical distribution of the Copy Number Variants Regions (CNVR) according to states (complex, gain and loss) on the Bos taurus ARS-UCD 1.2 assembly. Plotted CNVR are those shared by at least 2% of individuals. Percentage values refer to the genomic proportion covered by CNVR respect to the BTA length.

S1 Figure shows the genome-wide distribution of the 267 CNVR across the chromosomes together with the mean CNVR coverage length. The maximum number of CNVR are on BTA 1 and BTA 9. The mean CNVR length is not uniform along all chromosomes, and the maximum mean CNVR length was on BTA 12 (717,015 bp).

Principal component analysis results (Figure 6.3A and 6.3B) depict the genetic variability in the 3,809 cows analyzed, according to the presence or absence of CNV in the identified CNVR, considering their state. Each point in the scatter plot represents an individual animal, coloured as unique population (Figure 6.3A) or taking into account the herd from which it was sampled (Figure 6.3B).

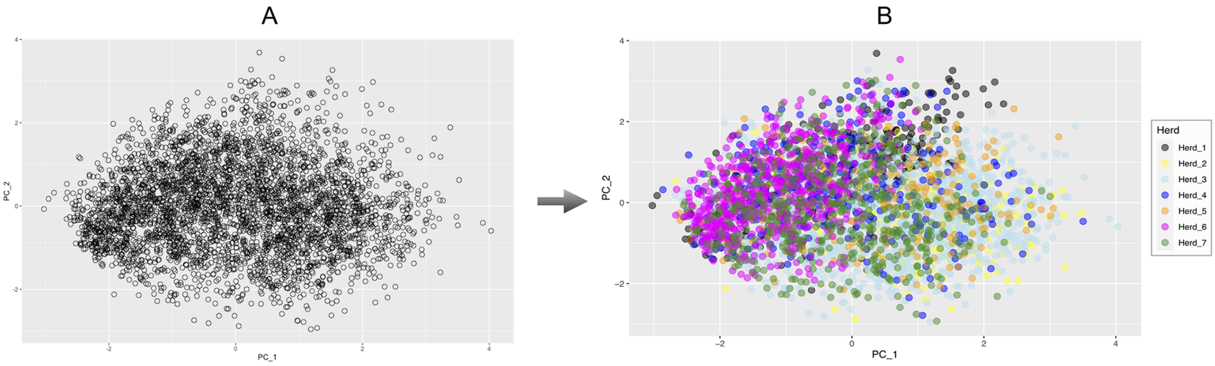


Figure 6.3: Principal Component Analysis results. A) Samples are coloured in black as unique Holstein breed; B) Samples are coloured according to the herds in which the cows were sampled.

6.4.2 Gene content and annotation

A total of 996 genes were annotated within 194 Holstein CNVR (72.6% of the CNVR_2%). Their functional classification, according to the David database, is reported in the S3 Table (recognized gene IDs = 942).

In S2 Figure (ClueGo network) it's possible to observe the presence of five macro-groups of genes associated with the following categories: troponin complex, sensory perception of smell, nervous system process, tuberculosis, and MHC class II protein complex. The KEGG pathway comprising the majority of genes is the one connected to tuberculosis, the same result has been obtained with David analysis.

After consulting the Cattle QTLdb, 142 genes were associated with a total of 122 different "Trait Name", grouped into 24 "Trait Types" corresponding to 6 "Trait Classes" (Exterior, Healthy, Meat and Carcass, Milk, Production, and Reproduction Traits), in concordance with the database nomenclature (Figure 6.4). As Figure 6.4 shows, the most of traits associated with the genes annotated in the CNVR are related to the phenotypes for which the Holstein population has been selected for years.

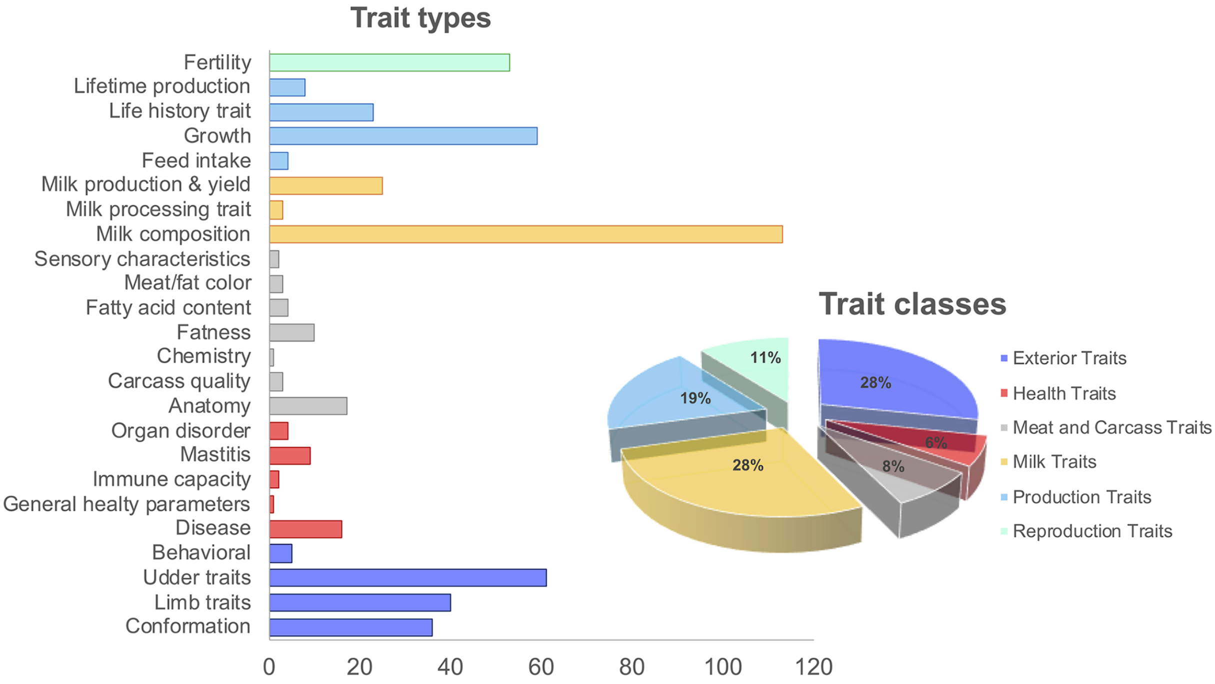


Figure 6.4: Graphical representation of QTL terms (Trait types and Trait classes) associated with genes mapped in CNVR. Colours of Trait types correspond to the ones in Trait classes.

6.4.3 Comparison with references

CNVR identified here were compared with those found in three other Holstein populations (comparison within breed) and in two different breeds (comparison among breeds; one dairy cattle – Jersey; one dual-purpose cattle – Brown Swiss) (Table 6.3, Figure 6.5, Supplementary Figure S2, and Supplementary Table S4). As reported in Table 6.3, the minimum and maximum number of overlapping regions were 7 and 27, respectively. The 48 overlapping CNVR (Supplementary Table S4) included 32 regions identified in other Holstein samples, i.e., CNVR mapped in at least two studies (shared_HOL), as shown in Figure 6.5A. When the comparison was performed with the JER and BSW cattle, the 32 shared_HOL regions in Figure 6.5B resulted in 11 Holstein-specific CNVR and 4 regions found in all breeds. As shown in Figure 6.5B, the BSW breed shared the largest number of overlapping CNVR. The total overlapping CNVR length was similar for those studies in which CNV were identified with the same software (2 Mb – PennCNV and 10 Mb – SVS, Table 6.3).

Table 6.3: Comparison with literature.

Platform	Software	Breed (N. of IDs)	Reference Genome	N. CNVR[a]	N. overlapped CNVR[b]	Overlapping length (bp)	Ref
			Comparison within breed				
Illumina HD; 50K; GGP150K	PennCNV	Holstein (96)	ARS-UCD1.2	36	7 (2.6%; 19.4%)	1,239,370	Butty et al. (2020)
Illumina HD	PennCNV	Holstein (315)	ARS-UCD1.2	135	14 (5.2%; 10.4%)	1,374,082	Lee et al. (2020)
Illumina HD	CNAM (SVS)	Holstein (242)	UMD3.1 remapped	112	23 (8.6%; 20.5%)	12,017,083	Durán Aguilar et al. (2017a)
			Comparison among breeds with different aptitude				
Illumina HD	PennCNV	Jersey (107)	ARS-UCD1.2	142	15 (5.6%; 10.5%)	1,915,749	Lee et al. (2020)
Illumina HD	CNAM (SVS)	Brown Swiss (1,116)	UMD3.1 remapped	233	27 (8.6%; 11.6%)	10,651,728	Prinsen et al. (2016)

[a] When remapped, this number refers to the CNVR resulting after the positions remapping.
[b] Proportion of overlapping: calculated as n. overlapped CNVR/n. CNVR in this study; n. overlapped CNVR/n. CNVR in other study.

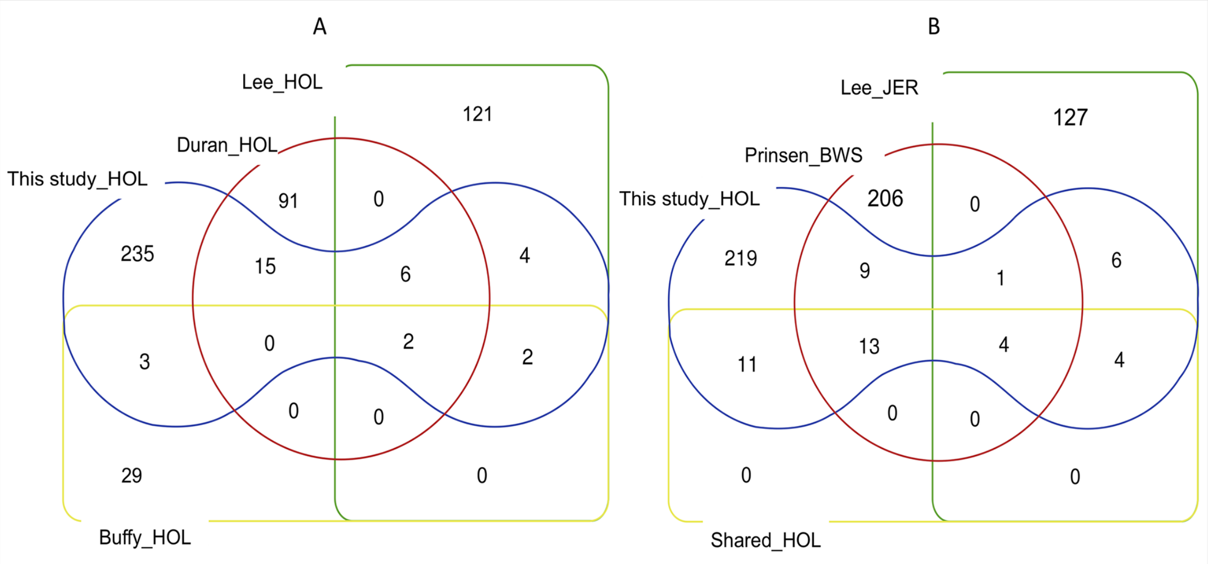


Figure 6.5: Comparison of CNVR identified in different Holstein populations (A) and in others two breeds (B). Shared_HOL are those CNVR (n.32) identified in at least two studies (part A of this Figure).

6.5 Discussion

In the literature there are several studies investigating genetic variability of Holstein' population using SNPs, and to increase knowledge on this breed, a large set of Italian Holstein cows has been inhere analyzed through CNV detection. CNV, a class of structural variation, can inform about population variability and are known to occur in the genome in response to environmental stressors, including positive selection, as a consequence of farming strategies (Strillacci et al. 2018).

This study, based on a medium density SNP chip, i.e. the Illumina GGP Bovine 100K, allowed the identification of a high number of CNV in a substantial number of Holstein cows. The number of CNV per sample (32, on average), is relatively higher compared to studies that rely on non-dense SNP chips, but lower compared to studies that rely on dense SNP chips or use sequences to call CNV (Ahmad et al. 2022; Sassi et al. 2016; Hou et al. 2011). As reported in the majority of CNV mapping studies performed with Illumina SNP chips, the number of deletions calls was approximately 1.98 more recurrent

than duplications (Durán Aguilar et al. 2017b; Lee et al. 2020; Prinsen et al. 2016). The mean length of deletion calls inhere (90,439 bp) is bigger than the mean length of found gains (77,785 bp). Interestingly Lee et al. (2020), using the Illumina BovineHD BeadChip, found that duplications are longer then deletions.
Overlapping CNV resulted in 1,397 CNVR covering 9.18% of the cattle genome. This value is much higher than the ones reported in the literature for Holsteins, which range from 0.5% to 2.8% (Lee et al. 2020; Jiang et al. 2013), but in line with the coverage found by Butty et al. (Butty et al. 2020), depending on the density of the SNP chip and the detection algorithm used (Butty et al. 2020; Xu et al. 2013). When CNV regions shared by at least 2% of the population were selected, the percentage of genome covered by CNVR decreased (2.9% of the autosomal genome length, Figure 6.2), a value similar to those reported by other authors (Lee et al. 2020; Jiang et al. 2013).
As shown in Supplementary Figure S1, CNVR are not uniformly distributed on the autosomes, and the distribution of CNVR according to their length class (Figure 6.1B) shows that the majority are short to medium in length and only a few are observed in the long classes, consistently with previous findings (Lee et al. 2020).

To visualize the genomic variability related to CNV detected in our study population, we performed a Principal Component Analysis and the results in Figure 6.3A, at first glance, show that all animals are spread in the graph without any clustering tendency.
The homogeneous grouping in this study appears to be related to the fact that all the cows, although bred on different farms, undergone similar intensive farming system. Nevertheless, the genetic selection performed by the farmers seems to produce an effect: when the grouping animals by herd (Figure 6.3B) a slight clustering can be observed, mainly for animals in Herd_6 (magenta colour). In Herd_6, mating plans have been based on bulls from a unique AI center for years, while all other herds use sires from different semen providers (Punturiero et al. 2023). When the gain/loss ratio was calculated in each herd to explain our findings, it was equal to 0.40 in Herd_6 (this value correspond to a loss/gain ratio = 2.40) and up to 0.49 in all the others herds (maximum value was 0.70 in Herd_5; loss/gain ratio = 1.41). The lower proportion of gain CNV found in Herd_6 may be linked to the highest number of daughters for sire in Herd_6, with a reduction of variability in specific genomic regions. The lower number of common bulls across all herds (as reported by Punturiero et al. (Punturiero et al. 2023)) can explain the cows' distribution of Herd_6 respect to the ones belong to all other farms. In Herd_5, the number of daughters per sire is one of the lower.

6.5.1 Gene content and annotation

According to the David database (Supplementary Table S3), the genes annotated within the CNVR were classified in 91 Go-Terms. The KEGG pathway analysis revealed that among the genes under analysis 56 are mainly represented in the pathway of immune system, namely, in the classes "Tuberculosis" and "*Staphylococcus aureus* infection", and in the pathway of thermogenesis. Disease resistance (or susceptibility) is a complex trait and interestingly it could be affected by genomic variations, as found by different authors reporting a substantial immune gene enhancement within CNV regions (Durán Aguilar et al. 2017b; Suchocki et al. 2015; Szyda et al. 2019; Lee et al. 2021). The network constructed with ClueGO (Supplementary Figure S2) aligns with the results found with the David analysis. It's possible to see genes connected to different GO categories linked to nervous system, troponin complex, sensory perception of smell, nervous system process, together with the KEGG category of susceptibility to tuberculosis. Some genes are connected with more than one category, for example BOLA genes.

Variation in gene copy number leads to phenotypic variation among animals. After consulting the AnimalQTLdb for cattle we grouped the QTL in 24 trait types. As listed in the Supplementary Table S3 and showed in Figure 6.4, the most common trait type is milk composition, for which 102 QTL were found. This result is in line with the expectations, being the animals part of commercial farms that sell milk for the dairy industry. Milk composition, together with udder conformation, fertility, and growth (more representative trait types in Figure 6.4) are all object traits of selection for high-productive breeds, such as the Holstein.

6.5.2 Noteworthy CNVR and comparison with references

Nine CNVR resulted over-represented due to a high number (> 2,000) of CNV defining these regions: 4 CNVR do not harbor genes, and most of them are in loss state. The only duplication region is the cnvr_234 identified on BTA 25 (in 2,400 cows) (see Supplementary Table S2). In this CNVR, the *EEF2K* and *POLR3E* genes are mapped, which are involved in the cellular response to oxidative stress (Sanchez et al. 2019) and the host innate immune defense against viruses (Ramanathan et al. 2020), respectively. Similarly, genes mapped in cnvr_024 (3,107 cows) on BTA 2 (*PARD3B*, *NRP2*) have reported roles in immune response (Schramek et al. 2009; Raphaka et al. 2017). Finally, cnvr_069 located on BTA 7 (2,156 cows) overlaps the *CNVR20* (complex state) identified by Butty et al. (Butty et al. 2020). This region harbors five genes belonging to the family 2 of olfactory receptor genes (*OR*). CNV are frequently found within *OR* genes, and this variability may contribute to individual or breed-specific differences in olfactory capacity (Hasin et al. 2008), which is also associated with feed intake and efficiency (Connor et al. 2018). This aligns with the findings in our research; indeed, conducting gene ontology analysis with ClueGO (see Supplementary Figure S2) yielded results for 35 genes in a copy number variation state linked with the following functional categories: sensory perception of smell, detection of stimulus involved in sensory perception, detection of chemical stimulus involved in sensory perception, olfactory perception activity. Nonetheless, these results only contribute to a small portion of our understanding given the size and complexity of this gene family comprising more than 1,000 known OR genes.

Regarding the comparison with references, as reported in Table 6.4, among the 267 CNVR, 11 overlapped with those identified only in Holstein populations and 4 in all the considered breeds (Holstein, Jersey, and Brown). It is important to note that the size of the CNVR identified in this study decreases after comparison (only regions with perfect overlap are reported). This is particularly evident for cnvr_225, which is split into two small regions as listed in Table 6.4. The entire cnvr_225 harbors genes belonging to the *BOLA* family, a well-known gene implicated in host immune response. In cnvr_133, located on BTA 13 (both in loss and complex states, according to breeds, see Supplementary Table S4), lies the *SIRPB1* gene, also involved in the immune response (Van Beek et al. 2005). Across the identified CNVR specific to Holstein cows, a wider variability in the region state can be observed; more than 70% are in fact in complex state. Only 4 CNVR harbor genes. Among them, cnvr_137 contains genes such as *LY6D*, *LYNX1*, *LYPD2*, *SLURP1*, *THEM6*, *PSCA*, *TSNARE1*, and *ARC*, which have been associated with clinical mastitis in US Holstein dairy cows (Tiezzi et al. 2015). The cnvr_245 includes the *BNIP3* gene, which plays a critical role in inducing autophagy during heat stress and was associated with the immune response phenotype (Livernois et al. 2021). The same region partially overlaps CNVR_1549_P (the region comprising *JAKMIP3*, *DPYSL4*, *STK32C*, *LRRC27*, *PWWP2B*), which was associated with clinical mastitis in Mexican Holstein cattle (Durán Aguilar et al. 2017a).

Table 6.4: CNVR in common between our study and the ones found in Holstein and in different cattle breeds.

CNVR_ID	Chr	Start	End	State	Genes	QTL
Common CNVR (HOL, JER, BSW)						
cnvr_075	8	76336567	76348332	loss		
cnvr_121	12	71701903	71765886	complex		
cnvr_133	13	53463194	53511604	complex	SIRPB1	SIRPB1: Milk protein percentage (QTL: 174904)
cnvr_225	23	25953514	26064642	complex		
cnvr_225	23	26113327	26350925	complex		
Only HOL CNVR						
cnvr_034	3	119662571	119718948	loss	COPS9, OTOS	
cnvr_035	4	182210	217902	loss		
cnvr_055	6	37695352	37736960	complex		
cnvr_058	6	51884459	52200066	complex		
cnvr_068	7	17374656	17409367	complex		
cnvr_072	7	81385397	81392696	loss		
cnvr_137	14	1645654	2064157	complex	LY6D, LYNX1, LYPD2, SLURP1,THEM6, PSCA, TSNARE1, ARC, ADGRB1, JRK	LY6D: Milk fat percentage (QTL:33308; 166962; 161706), Milk protein percentage (QTL:161824)
cnvr_176	18	27856333	28303561	complex		
cnvr_181	18	57234258	57254890	complex		
cnvr_216	21	69040055	69788216	complex	C21H14orf180, TMEM179, INF2, ADSSL1, SIVA1, AKT1, ZBTB42, CEP170B, PLD4, AHNAK2, CLBA1, CDCA4, GPR132, JAG2, NUDT14, BRF1, BTBD6, PACS2, TEX22, MTA1, CRIP2, CRIP1, TEDC1, TMEM121	AKT1: Bovine respiratory disease susceptibility (QTL: 160320; 160321); BRF1: Conception rate (QTL: 123998)
cnvr_245	26	50598736	51990348	complex	KNDC1, ADGRA1, CFAP46, NKX6-2, INPP5A, BNIP3, JAKMIP3, DPYSL4, STK32C, LRRC27, PWWP2B	

6.6 Conclusions

The study provides novel insights into CNV mapped within the Italian Holstein cows. To date, this is the only study that conducted a CNV analysis on such a large number of animals within this breed. Based on CNV, the Principal Component Analysis (PCA) revealed a homogeneous distribution of cows, indicating a shared effect of the intensive farming system on these animals. The slight clustering observed among cows from the same farm implies that genetic selection may influence CNV distribution, underscoring the potential impact of selective breeding practices. The functional analysis of genes annotated in the more common CNVR revealed biological mechanism related to immune resistance to infection and adaptability. QTL linked with the main traits object of directional selection overlapped with many CNVR here identified. Genes involved in immune response and defense against oxidative stress were identified within CNVR, suggesting that genetic variability could affect the animals' ability to respond to environmental stressors. The analysis of CNV not only provides an additional dimension of genetic information, but also represents a valuable resource to optimise (new perspective) genomic selection in a more complete and accurate way.

6.7 Supporting information

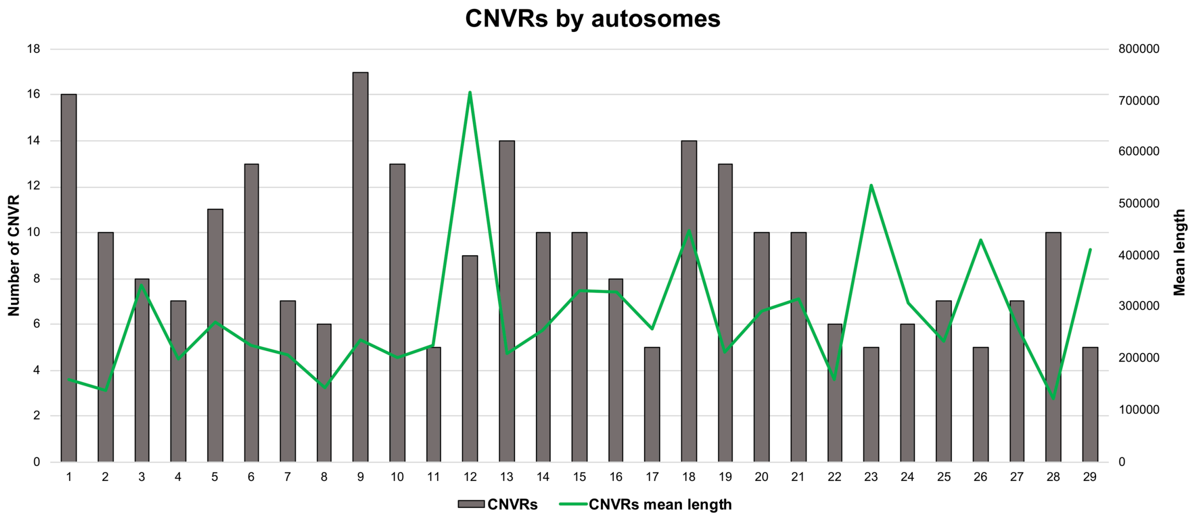


Figure S1. Graphical representation of CNVRs number and mean CNVR coverage length on autosomes.

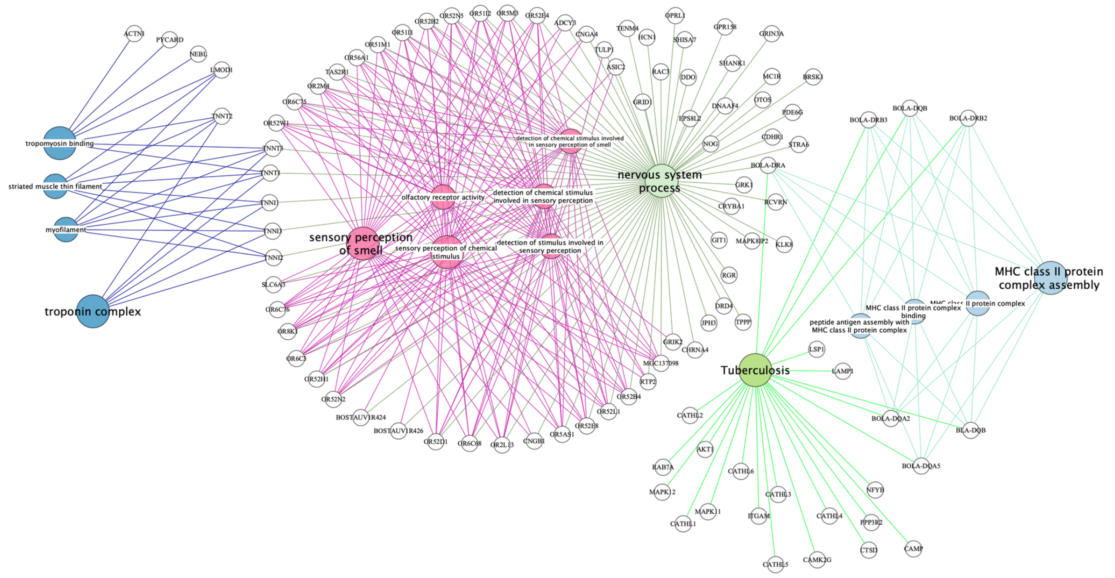


Figure S2. ClueGo network of genes annotated in CNVRs identified in at least 2% of cows.

For the tables, the following files are available but not included in the thesis due to their size:

- S1 Table. CNV identified in Holstein breed. `https://doi.org/10.1371/journal.pone.0303044.s001` (XLSX).

- S2 Table. List of the total CNVR (sheet_1); CNVR identified in at least 2% of cows (sheet_2); list of CNV defining CNVR identified in at least 2% of cows. `https://doi.org/10.1371/journal.pone.0303044.s002` (XLSX).
- S3 Table. Gene functional annotation from David database. `https://doi.org/10.1371/journal.pone.0303044.s003` (XLSX).
- S4 Table. CNVR comparison with references. `https://doi.org/10.1371/journal.pone.0303044.s004` (XLSX).

Chapter 7

Discussion

This thesis investigates resilience of dairy cattle through a multifaceted approach that combined phenotypic characterization, simulation analysis and genomic analyses, leveraging longitudinal data from automatic milking systems (AMS) and comprehensive genomic information from a commercial Holstein herd in Northern Italy. The overarching goal was to advance our understanding of resilience as a complex trait and provide practical tools for its assessment and genetic improvement.

7.1 The complexity of resilience assessment

The concept of resilience, borrowed here from ecological systems (Holling 1973), has gained increasing relevance in the context of livestock production as the sector faces mounting environmental and socioeconomic pressures. As reported in the literature review (Chapter 1), planetary boundaries are being exceeded (Richardson et al. 2023; The Lancet Planetary Health 2025) while climate extremes are becoming more frequent (Sauer et al. 2024; Monteleone et al. 2024). In this context, breeding resilient animals capable of maintaining productivity under suboptimal conditions is important for sustainable agriculture.

However, defining and measuring resilience in dairy cattle is challenging. Due to the nature of the concept, translating it into measurable phenotypes is difficult.

7.1.1 Characterization of resilience indicators and lactation curve models

The first research objective (Chapter 3) addressed a fundamental question: do different lactation curve models rank cows differently when evaluating resilience indicators? This question was motivated by the lack of consensus in the literature regarding the best method for estimating the unperturbed lactation trajectory. Our phenotypic characterization of eighteen lactation curve models and nine resilience indicators revealed substantial heterogeneity in how models rank individual cows for the same indicator. The ranking agreement analysis showed that while some model pairs produced highly concordant

rankings, others diverged considerably. This variation has important implications for practical breeding: if different models produce different rankings, then the choice of model becomes a critical methodological decision that directly influences breeding outcomes. Thus, the choice of lactation curve model and resilience indicator is not neutral and demands careful consideration.
A critical limitation emerged from this characterization: we could rank models relative to each other, but we could not determine which model-indicator combinations best capture "true" resilience. This was because we did not know which cows were genuinely resilient. This limitation motivated the subsequent simulation study, which provided an opportunity to validate resilience indicators against simulated breeding values for resilience.

7.1.2 Simulation study: validating resilience indicators

The characterization study revealed substantial discrepancies among models and indicators, but a critical question remained unsolved: to what extent do these indicators actually capture resilience? Chapter 4 addressed this question through a simulation study that incorporated simulated perturbation events applied to lactation curves, designed to represent plausible biological challenges to lactation.

The simulation study generated resilience breeding values with a controlled genetic architecture, applied both resilience-responsive and resilience-independent perturbations to simulated lactation trajectories, and then evaluated the correlation between true resilience breeding values and resilience indicator values computed from deviations estimated by various lactation curve models. This approach balanced biological realism with precise control over genetic parameters, enabling a test of whether resilience indicators successfully detect resilience and its genetic variation. The results were sobering yet informative. The best-performing resilience indicators (AvgAbs, AC1, RMSE, LMS, LnVar) achieved Spearman correlations with true resilience breeding values of approximately 0.15-0.17. These correlations fall substantially below what would be expected if resilience indicators were optimal phenotypes for capturing resilience under the assumed genetic architecture. This gap reveals a fundamental limitation: current resilience indicators capture only a fraction of the genetic component in this resilience setting.

Several aspects of the simulation design likely contribute to this limitation. First, the inclusion of environmental perturbations unrelated to resilience (e.g., feed shortage, facility breakdown) introduced noise that obscured the genetic signal – a condition that mirrors real-world production environments where multiple independent stressors simultaneously affect milk yield. Second, autocorrelation patterns in environmental components could introduce temporal structure difficult to distinguish from resilience-related deviations, though this aspect was not explicitly investigated. Third, the resilience indicators themselves may not be optimal transformations of the underlying resilience.

Despite measurable heritability (Poppe et al. 2020; Chen et al. 2023; Kessler et al. 2024), the modest correlations between resilience indicators and resilience BV (0.15–0.20) raise an important practical question: are these indicators sufficiently accurate for effective genetic selection? Published heritability estimates for variance-based indicators suggest that genetic progress is achievable, however the low correlation between indicators and resilience imply that selection will be substantially less efficient than heritability estimates alone would suggest.

7.1.3 Genomic background of resilience

Chapter 5 complemented the phenotypic works done in Chapter 3 and Chapter 4 by investigating the genomic regions associated with resilience indicators in the same Holstein population. Resilience indicators under analysis were the logarithm of the variance (LnVar), autocorrelation (rauto), weighted frequency perturbations (WfPert) and daily perturbations (dPert), all of them computed from deviations derived from estimating the unperturbed lactation curve with a fourth degree quantile polynomial of order 0.70. Using a selective genotyping approach with DNA pooling, the GWAS identified genes related to immune response, energy metabolism, and tissue integrity. This finding aligns with the biological understanding that resilience is multifactorial, involving the animal's ability to mount appropriate immune responses to pathogens while maintaining energy balance and tissue homeostasis. This genetic background favors genomic selection approaches that can simultaneously account for many small-effect loci distributed across the genome, thus making it difficult to detect the true phenotypic variation underlying resilience.

7.1.4 Copy Number Variants: an additional genomic resource

Chapter 6 provided a map of copy number variants (CNV) in Italian Holstein cattle, representing the first mid-large scale CNV characterization of a sample of the Italian Holstein, considered representative as the sire genetic component comes from a wide number of populations and AI centers. The detection of 123,814 CNV aggregated into 1,397 CNV regions (CNVR) covering approximately 2-10% of the bovine genome aligns with previous reports (Hay et al. 2018). The annotation of genes within CNVR revealed many involved in immune response, milk production, udder conformation, and body morphology – traits relevant to both productivity and resilience.

From a breeding perspective, CNV represent an additional source of genetic variation that could explain portions of the missing heritability not captured by SNP. CNV can affect gene dosage (through deletions or duplications), disrupt gene structure, and influence regulatory elements. The integration of CNV information with SNP-based genomic predictions has shown promise in some studies (Hou et al. 2012; Xu et al. 2014). For resilience traits, where our simulation study revealed that current indicators capture only a fraction of genetic variation, CNV might provide additional explanatory power, particularly if CNVR overlap with genes involved in stress response, immune function, or metabolic regulation. The CNV map produced in this thesis serves as a valuable resource for future research. As sequencing costs decline and whole-genome sequence data become more accessible, refined CNV detection and improved understanding of their functional effects will enable more sophisticated analyses of CNV contributions to complex traits.

Chapter 8

Conclusions

This work has advanced the idea that resilience can be measured, modeled, and genetically improved, marking a crucial step toward sustainable livestock breeding.

8.1 Methodological innovations and limitations

This thesis introduced methodological innovations. A comprehensive characterization of lactation curve models and resilience indicators was conducted using ranking correlations and PCA, thereby providing a systematic framework for evaluating model consistency. The simulation study's approach to generating breeding values with controlled genetic architecture while maintaining biological realism offers a simple yet compelling framework for assessing the effectiveness of different modelling choices and resilience indicators.

However, it is imperative to acknowledge the inherent limitations of these studies. The utilization of a solitary dataset from a single farm imposes limitations on the generalizability of the findings. The performance of lactation curve models and the ranking of cows for resilience indicators may be influenced by various farming systems, management practices, and environmental conditions. Multi-farm studies, encompassing a range of production environments, would serve to bolster conclusions regarding the robustness of the models. The simulation study's use of the range divided by four approximation to derive genetic variances from phenotypic ranges constitutes a simplification. This approximation assumes normally distributed phenotypes and may not accurately reflect true genetic variances. The simulation also assumed independence between events; however, in reality, mastitis has been shown to increase susceptibility to other health challenges, creating cascading perturbations. In the future, simulation frameworks may have the capacity to model event interactions and utilize empirically derived genetic parameters as they become available. The GWAS has limitations in resolution and power due to the small sample size. The results should be considered exploratory, requiring validation in independent populations and functional studies to confirm causal relationships between identified genomic regions and resilience phenotypes.

8.2 Integration across chapters: a coherent picture

Viewed collectively, the chapters of this thesis paint a coherent picture of resilience in dairy cattle. The characterization study revealed that methodological choices in estimating the unperturbed state significantly affect resilience indicator values and cow rankings, highlighting the need for careful model selection. The simulation study demonstrated that current resilience indicators, while heritable and useful for breeding, capture only a fraction of true genetic resilience. This finding is sobering but valuable because it sets realistic expectations for genetic progress and motivates continued research into improved phenotypes. The GWAS results confirmed the complex, polygenic nature of resilience and identified specific genomic regions for further investigation. Finally, the CNV map provides an additional genomic resource that may contribute to explaining resilience variation.

Together, these studies advance the field from descriptive characterizations of resilience indicators toward a mechanistic understanding of what these indicators capture and the genetic architecture underlying them. This progression from phenotype to simulation to genomics represents a comprehensive approach to studying complex traits.

8.3 Practical implications for breeding programs

The findings of this thesis have practical implications for dairy cattle breeding programs. First, both lactation curve models and resilience indicators should be carefully selected when implementing genetic evaluations for resilience. Based on our simulation results, variance-based indicators (LnVar, LMS, RMSE, AvgAbs) and the less reliable AUC derived from the Weighted Spline (Kessler et al. 2024) or Iterated Wood (Adriaens et al. 2020) models will maximize genetic gain. Second, realistic expectations should be set regarding the rate of genetic progress.

8.4 Future research directions

This thesis opens avenues for future research. First, the simulation framework could be extended to model event interactions (e.g., mastitis increasing heat stress susceptibility), incorporate temporal patterns in resilience across lactation stages, and use empirically derived genetic parameters rather than approximations, while a bigger sample size would enhance the robustness and generalizability of the findings. Second, alternative functional forms for resilience indicators could be explored to optimize the transformation of raw deviations into resilience phenotypes, a field that recently gained improvement (Kessler et al. 2024; Guinan et al. 2025a; Guinan et al. 2025b). Here, the collection of accurate data on veterinary treatments and associated diseases, together with additional sensor data, could provide more efficient resilience indicators: adding new dimensions to the existing indicator classes could pave the way for new opportunities to analyse the genetic basis of resilience and apply accurate selection. Third, genetic correlations among resilience indicators and other functional and production traits need to be evaluated in order to understand the potential for correlated responses to selection and therefore being properly included within selection indices. Fourth, functional genomics studies of candidate genes identified in the GWAS will elucidate biological mechanisms underlying resilience variation. Fifth, investigation of CNV functional effects and integration of CNV information with SNP-based genomic predictions may capture additional genetic variation.

Bibliography

Abera, Michael (2024). “Response of dairy cattle to thermal stress: Implications for designing animal breeding strategies for sustainable productivity under changing climate”. In: *Applied Veterinary Research* 3.4, pp. 2024011–2024011. DOI: `10.31893/avr.2024011`.

Adams, D. (1979). *The Hitchhiker’s Guide to the Galaxy.* New York: Harmony Books.

Adriaens, I., I. van den Brulle, L. D’Anvers, J. M. E. Statham, K. Geerinckx, S. De Vliegher, S. Piepers, and B. Aernouts (Jan. 2021). “Milk losses and dynamics during perturbations in dairy cows differ with parity and lactation stage”. In: *Journal of Dairy Science* 104.1, pp. 405–418. ISSN: 0022-0302. DOI: `10.3168/jds.2020-19195`. (Visited on 08/13/2025).

Adriaens, I., N. C. Friggens, W. Ouweltjes, H. Scott, B. Aernouts, and J. Statham (2020). “Productive life span and resilience rank can be predicted from on-farm first-parity sensor time series but not using a common equation across farms”. In: *Journal of dairy science* 103.8, pp. 7155–7171. DOI: `10.3168/jds.2019-17826`.

Adriaens, Ines, Tjebbe Huybrechts, Ben Aernouts, Katleen Geerinckx, Sofie Piepers, Bart De Ketelaere, and Wouter Saeys (Nov. 2018). “Method for short-term prediction of milk yield at the quarter level to improve udder health monitoring”. In: *Journal of Dairy Science* 101.11, pp. 10327–10336. DOI: `10.3168/jds.2018-14696`. URL: `https://doi.org/10.3168/jds.2018-14696`.

Ahmad, S. F., A. Singh, S. Panda, W. A. Malla, A. Kumar, and T. Dutt (2022). “Genome-wide elucidation of CNV regions and their association with production and reproduction traits in composite Vrindavani cattle”. In: *Gene* 830, p. 146510. DOI: `10.1016/j.gene.2022.146510`.

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* 67.3, pp. 637–644.

Alotiby, A. (2024). “Immunology of Stress: A Review Article”. In: *Journal of Clinical Medicine* 13. DOI: `10.3390/jcm13216394`.

Antanaitis, R., V. Juozaitienė, V. Jonike, V. Čukauskas, D. e Urbšienė, A. Urbšys, et al. (2021). “Relationship between Temperament and Stage of Lactation, Productivity and Milk Composition of Dairy Cows”. In: *Animals* 11. DOI: `10.3390/ani11071840`.

Aradotlu Parameshwarappa, Madhusoodan, Sejian Veerasamy, Rashamol Veettiparambil Pandarathil, Savitha Sira Tulasiramu, Bagath Madiajagan, Krishnan Govindan, and Bhatta Raghavendra (2019). “Resilient capacity of cattle to environmental challenges –

An updated review”. In: *Journal of Animal Behaviour and Biometeorology* 7.3, pp. 104–118. ISSN: 23181265. DOI: `10.31893/2318-1265jabb.v7n3p104-118`.

Arendt, M., K. M. Cairns, J. W. O. Ballard, P. Savolainen, and E. Axelsson (2016). “Diet adaptation in dog reflects spread of prehistoric agriculture”. In: *Heredity* 117, pp. 301–306. DOI: `10.1038/hdy.2016.48`.

Awasthi Mishra, N., C. Drögemüller, V. Jagannathan, I. Keller, D. Wüthrich, R. Bruggmann, et al. (2017). “A structural variant in the 5’-flanking region of the TWIST2 gene affects melanocyte development in belted cattle”. In: *PLoS One* 12, e0180170. DOI: `10.1371/journal.pone.0180170`.

Bannerman, D. D., M. J. Paape, W. R. Hare, and J. C. Hope (2004). “Characterization of the bovine innate immune response to intramammary infection with Klebsiella pneumoniae”. In: *Journal of Dairy Science* 87, pp. 2420–2432. DOI: `10.3168/jds.S0022-0302(04)73365-2`.

Beckett, Emma L., Tim Cassettari, Carlene Starck, and Flávia Fayet-Moore (2024). “Dairy milk: There are alternatives but no equivalents”. In: *Food Science & Nutrition* 12.10, pp. 8470–8482. DOI: `10.1002/fsn3.4301`.

Bekele, R., M. Taye, G. Abebe, and S. Meseret (2023). “Genomic regions and candidate genes associated with milk production traits in Holstein and its crossbred cattle: A review”. In: *International Journal of Genomics*.

Ben Abdelkrim, A., L. Puillet, P. Gomes, and O. Martin (2021a). “Lactation curve model with explicit representation of perturbations as a phenotyping tool for dairy livestock precision farming”. In: *Animal* 15.1, p. 100074. ISSN: 1751-7311. DOI: `https://doi.org/10.1016/j.animal.2020.100074`.

— (2021b). “Lactation curve model with explicit representation of perturbations as a phenotyping tool for dairy livestock precision farming”. In: *Animal* 15, p. 100074. DOI: `10.1016/j.animal.2020.100074`.

Berghof, T. V. L., H. Bovenhuis, and H. A. Mulder (2019a). “Body Weight Deviations as Indicator for Resilience in Layer Chickens”. In: *Frontiers in Genetics* 10. DOI: `10.3389/fgene.2019.01216`.

Berghof, T. V. L., M. Poppe, and H. A. Mulder (2019b). “Opportunities to Improve Resilience in Animal Breeding Programs”. In: *Frontiers in genetics* 10.Jen. ISSN: 1664-8021. DOI: `10.3389/fgene.2018.00692`.

Bindea, G., B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, et al. (2009). “ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks”. In: *Bioinformatics* 25, pp. 1091–1093. DOI: `10.1093/bioinformatics/btp101`.

Bowles, Nicholas, Samuel Alexander, and Michalis Hadjikakou (2019). “The livestock sector and planetary boundaries: A ’limits to growth’ perspective with dietary implications”.

In: *Ecological Economics* 160, pp. 128–136. ISSN: 0921-8009. DOI: `10.1016/j.ecolecon.2019.01.033`.

Brito, L.F., N. Bedere, F. Douhard, H.R. Oliveira, M. Arnal, F. Peñagaricano, A.P. Schinckel, C.F. Baes, and F. Miglior (2021). “Review: Genetic selection of high-yielding dairy cattle toward sustainable farming systems in a rapidly changing world”. In: *Animal* 15. Sustainable livestock systems for high-producing animals, p. 100292. ISSN: 1751-7311. DOI: `10.1016/j.animal.2021.100292`. URL: `https://www.sciencedirect.com/science/article/pii/S175173112100135X`.

Brito, Luiz F., Heringstad Bjørg, Klaas Ilka C., Schodl Katharina, Cabrera Victor E., Stygar Anna, Iwersen Michael, Haskell Marie J., Stock Kathrin F., Gengler Nicolas, Bewley Jeffrey, Hostens Miel, Vasseur Elsa, and Egger-Danner Christa (2025). “Using data from sensors and other precision farming technologies to enhance the sustainability of dairy cattle breeding programs”. In: *Journal of Dairy Science* 108.10, pp. 10447–10480. DOI: `10.3168/jds.2024-26554`.

Butty, A. M., T. C. S. Chud, F. Miglior, F. S. Schenkel, A. Kommadath, K. Krivushin, et al. (2020). “High confidence copy number variants identified in Holstein dairy cattle from whole genome sequence and genotype array data”. In: *Scientific Reports* 10, pp. 1–13. DOI: `10.1038/s41598-020-64680-3`.

Calderón-Chagoya, R., V. E. Vega-Murillo, A. García-Ruiz, Á. Ríos-Utrera, G. Martínez-Velázquez, and M. Montaño-Bermúdez (2023). “Genome and chromosome wide association studies for growth traits in Simmental and Simbrah cattle”. In: *Animal Bioscience* 36, pp. 19–28. DOI: `10.5713/ab.21.0517`.

Call, D. R., M. A. Davis, and A. A. Sawant (2008). “Antimicrobial resistance in beef and dairy cattle production”. In: *Animal Health Research Reviews* 9.2, pp. 159–167. DOI: `10.1017/S1466252308001515`.

Cartwright, S. L., J. Schmied, N. Karrow, and B. A. Mallard (2023). “Impact of heat stress on dairy cattle and selection strategies for thermotolerance: a review”. In: *Frontiers in veterinary science* 10, p. 1198697. ISSN: 2297-1769. DOI: `10.3389/fvets.2023.1198697`.

Chen, S.-Y., J. P. Boerman, L. S. Gloria, V. B. Pedrosa, J. Doucette, and L. F. Brito (2023). “Genomic-based genetic parameters for resilience across lactations in North American Holstein cattle based on variability in daily milk yield records”. In: *Journal of dairy science* 106.6, pp. 4133–4146. DOI: `10.3168/jds.2022-22754`.

Chen, S.-Y., L. S. Gloria, V. B. Pedrosa, J. Doucette, J. P. Boerman, and L. F. Brito (2024). “Unraveling the genomic background of resilience based on variability in milk yield and milk production levels in North American Holstein cattle through genome-wide association study and Mendelian randomization analyses”. In: *Journal of Dairy Science* 107, pp. 1035–1053. DOI: `10.3168/jds.2023-23650`.

Chen, S.-Y., H. R. Oliveira, F. S. Schenkel, V. B. Pedrosa, M. G. Melka, and L. F. Brito (2020). “Using imputed whole-genome sequence variants to uncover candidate mutations and genes affecting milking speed and temperament in Holstein cattle”. In: *Journal of Dairy Science* 103, pp. 10383–10398. DOI: `10.3168/jds.2020-18897`.

Chu, H., A. Awasthi, G. C. White, M. Chrzanowska-Wodnicka, and S. Malarkannan (2008). “Rap1b regulates B cell development, homing, and T cell-dependent humoral immunity”. In: *Journal of Immunology* 181, pp. 3373–3383. DOI: `10.4049/jimmunol.181.5.3373`.

Codrea, M. C., S. Højsgaard, and N. C. Friggens (Oct. 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: *Journal of Animal Science* 89.10, pp. 3089–3098. ISSN: 0021-8812. DOI: `10.2527/jas.2010-3753`. eprint: `https://academic.oup.com/jas/article-pdf/89/10/3089/23613109/3089.pdf`.

Colditz, Ian G. and Brad C. Hine (2016). “Resilience in farm animals: biology, management, breeding and implications for animal welfare”. In: *Animal Production Science* 56.12, pp. 1961–1983. ISSN: 1836-0939. DOI: `10.1071/AN15297`.

Cole, J. B., G. R. Wiggans, L. Ma, T. S. Sonstegard, T. J. Lawlor, B. A. Crooker, et al. (2011). “Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows”. In: *BMC Genomics* 12, pp. 1–17.

Cole, J.B. and P.M. VanRaden (2018). “Symposium review: Possibilities in an age of genomics: The future of selection indices”. In: *Journal of Dairy Science* 101.4, pp. 3686–3701. ISSN: 0022-0302. DOI: `10.3168/jds.2017-13335`. URL: `https://www.sciencedirect.com/science/article/pii/S0022030217309694`.

Connor, E. E., Y. Zhou, and G. E. Liu (2018). “The essence of appetite: does olfactory receptor variation play a role?” In: *Journal of Animal Science* 96, pp. 1551–1558. DOI: `10.1093/jas/sky068`.

Cruz, A. S. da, D. C. Silva, L. B. Minasi, L. K. de Farias Teixeira, F. M. Rodrigues, C. C. da Silva, et al. (2020). “Single-Nucleotide Polymorphism Variations Associated With Specific Genes Putatively Identified Enhanced Genetic Predisposition for 305-Day Milk Yield in the Girolando Crossbreed”. In: *Frontiers in Genetics* 11, p. 573344. DOI: `10.3389/fgene.2020.573344`.

Daniel, J. B., N. C. Friggens, P. Chapoutot, H. van Laar, and D. Sauvant (2016). “Milk yield and milk composition responses to change in predicted net energy and metabolizable protein: a meta-analysis”. In: *Animal* 10, pp. 1975–1985. DOI: `10.1017/S1751731116001245`.

Darling, T. K. and T. J. Lamb (2019). “Emerging Roles for Eph Receptors and Ephrin Ligands in Immunity”. In: *Frontiers in Immunology* 10, p. 1473. DOI: `10.3389/fimmu.2019.01473`.

Darvasi, A. and M. Soller (1992). “Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus”. In: *Theoretical and Applied Genetics* 85, pp. 353–359.

— (1994). “Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus”. In: *Genetics* 138, pp. 1365–1373.

Devani, K., G. Plastow, K. Orsel, and T. S. Valente (2020). “Genome-wide association study for mammary structure in Canadian Angus cows”. In: *PLoS One* 15, e0237818. DOI: `10.1371/journal.pone.0237818`.

Diskin, S. J., M. Li, C. Hou, S. Yang, J. Glessner, H. Hakonarson, et al. (2008). “Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms”. In: *Nucleic Acids Research* 36, e126–e126. DOI: `10.1093/nar/gkn556`.

Dong, K., N. Yao, Y. Pu, X. He, Q. Zhao, Y. Luan, et al. (2014). “Genomic Scan Reveals Loci under Altitude Adaptation in Tibetan and Dahe Pigs”. In: *PLoS One* 9, e110520. DOI: `10.1371/journal.pone.0110520`.

Dong, M. C. and L. D. Van Vleck (1989). “Correlations among first and second lactation milk yield and calving interval”. In: *Journal of Dairy Science* 72, pp. 1933–1936.

Dostaler-Touchette, V., F. Bédard, C. Guillemette, F. Pothier, P. Y. Chouinard, and F. J. Richard (2009). “Cyclic adenosine monophosphate (cAMP)-specific phosphodiesterase is functional in bovine mammary gland”. In: *Journal of Dairy Science* 92, pp. 3757–3765. DOI: `10.3168/jds.2009-2065`.

Du, C., T. Deng, Y. Zhou, T. Ye, Z. Zhou, S. Zhang, et al. (2019). “Systematic analyses for candidate genes of milk production traits in water buffalo (Bubalus Bubalis)”. In: *Animal Genetics* 50, pp. 207–216. DOI: `10.1111/age.12739`.

Dubey, P. K., S. Goyal, A. K. Yadav, B. R. Sahoo, N. Kumari, S. K. Mishra, et al. (2014). “Genetic diversity analysis of the thyroglobulin gene promoter in buffalo and other bovines”. In: *Livestock Science* 167, pp. 65–72. DOI: `10.1016/j.livsci.2014.06.007`.

Durán Aguilar, M., S. I. Román Ponce, F. J. Ruiz López, E. González Padilla, C. G. Vásquez Peláez, A. Bagnato, et al. (2017a). “Genome-wide association study for milk somatic cell score in Holstein cattle using copy number variation as markers”. In: *Journal of Animal Breeding and Genetics* 134.1, pp. 49–59. DOI: `10.1111/jbg.12238`.

— (2017b). “Genome-wide association study for milk somatic cell score in holstein cattle using copy number variation as markers”. In: *Journal of Animal Breeding and Genetics* 134, pp. 49–59. DOI: `10.1111/jbg.12238`.

Egger-Danner, C., J. B. Cole, J. E. Pryce, N. Gengler, B. Heringstad, A. Bradley, et al. (2015). “Invited review: overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits”. In: *Animal* 9, pp. 191–207. DOI: `10.1017/S1751731114002614`.

Elgersma, G. G., G. de Jong, R. van der Linde, and H. A Mulder (2018). “Fluctuations in milk yield are heritable and can be used as a resilience indicator to breed healthy cows”. In: *Journal of Dairy Science* 101.2, pp. 1240–1250.

Falconer, Douglas S. (1989). *Introduction to Quantitative Genetics*. 3rd. Harlow, UK: Longman Scientific and Technical.

Finot, L., E. Chanat, and F. Dessauge (2018). “Molecular signature of the putative stem/progenitor cells committed to the development of the bovine mammary gland at puberty”. In: *Scientific Reports* 8, p. 16194. DOI: `10.1038/s41598-018-34691-2`.

— (2019). “Mammary epithelial cell lineage changes during cow’s life”. In: *Journal of Mammary Gland Biology and Neoplasia* 24, pp. 185–197.

Frantz, Laurent A. F., Daniel G. Bradley, Greger Larson, and Ludovic Orlando (2020). “Animal domestication in the era of ancient genomics”. In: *Nature Reviews Genetics* 21, pp. 449–460. DOI: `10.1038/s41576-020-0225-0`.

Friggens, N. C. and J. H. Badsberg (2007). “The effect of breed and parity on curves of body condition during lactation estimated using a non-linear function”. In: *Animal* 1, pp. 565–574. DOI: `10.1017/S1751731107691861`.

Garcia-Baccino, C. A., C. Marie-Etancelin, F. Tortereau, D. Marcon, J.-L. Weisbecker, and A. Legarra (2021). “Detection of unrecorded environmental challenges in high-frequency recorded traits, and genetic determinism of resilience to challenge, with an application on feed intake in lambs”. In: *Genetics Selection Evolution* 53, p. 4. DOI: `10.1186/s12711-020-00595-x`.

Glick, G., A. Shirak, E. Seroussi, Y. Zeron, E. Ezra, J. I. Weller, et al. (2011). “Fine Mapping of a QTL for Fertility on BTA7 and Its Association With a CNV in the Israeli Holsteins”. In: *G3: Genes| Genomes| Genetics* 1, pp. 65–74. DOI: `10.1534/g3.111.000299`.

Goepp, V., O. Bouaziz, and G. Nuel (2018). “Spline Regression with Automatic Knot Selection”. In: URL: `http://arxiv.org/pdf/1808.01770`.

Goepp, V. and G. Nuel (2022). *aspline: Spline Regression with Adaptive Knot Selection*. R package version 0.2.0. URL: `https://CRAN.R-project.org/package=aspline`.

Graham, Jason R., Masoomeh Taghipoor, Leonardo S. Gloria, Jacquelyn P. Boerman, Jarrod Doucette, Artur O. Rocha, and Luiz F. Brito (2024). “Trait development and genetic parameters of resilience indicators based on variability in milk consumption recorded by automated milk feeders in North American Holstein calves”. In: *Journal of Dairy Science* 107.12, pp. 11180–11194. DOI: `10.3168/jds.2024-25192`.

Graulet, B. (Apr. 2014). “Ruminant milk: A source of vitamins in human nutrition”. In: *Animal Frontiers* 4.2, pp. 24–30. ISSN: 2160-6056. DOI: `10.2527/af.2014-0011`. eprint: `https://academic.oup.com/af/article-pdf/4/2/24/32410289/24.pdf`.

Grigoletto, L., M. H. A. Santana, F. F. Bressan, J. P. Eler, M. F. G. Nogueira, H. N. Kadarmideen, et al. (2020). “Genetic Parameters and Genome-Wide Association Studies for Anti-Müllerian Hormone Levels and Antral Follicle Populations Measured After Estrus Synchronization in Nellore Cattle". In: *Animals* 10. DOI: `10.3390/ani10071185`.

Gröhn, Y. T., David J. Wilson, R. N. González, J. A. Hertl, H. Schulte, G. Bennett, and Y. H. Schukken (2004). "Effect of pathogen-specific clinical mastitis on milk yield in dairy cows". In: *Journal of dairy science* 87.10, pp. 3358–3374.

Guinan, F. L., R. H. Fourdraine, F. Peñagaricano, and K. A. Weigel (2025a). "Dairy cow resilience: Data-driven detection and quantification of perturbations using daily milk weights". In: *Journal of Dairy Science.* ISSN: 0022-0302. DOI: `10.3168/jds.2025-26485`.

— (2025b). "Genetic parameters for alternative resilience phenotypes in Holstein cows". In: *JDS Communications.* ISSN: 2666-9102. DOI: `10.3168/jdsc.2025-0823`.

Hansen, J., M. Sato, and R. Ruedy (2012). "Perception of climate change". In: *Proceedings of the National Academy of Sciences* 109, E2415–E2423. DOI: `10.1073/pnas.1205276109`.

Hasin, Y., T. Olender, M. Khen, C. Gonzaga-Jauregui, P. M. Kim, A. E. Urban, et al. (2008). "High-resolution copy-number variation map reflects human olfactory receptor diversity and evolution". In: *PLoS Genetics* 4, e1000249. DOI: `10.1371/journal.pgen.1000249`.

Hay, E. H. A., Y. T. Utsunomiya, L. Xu, Y. Zhou, H. H. R. Neves, R. Carvalheiro, et al. (2018). "Genomic predictions combining SNP markers and copy number variations in Nellore cattle". In: *BMC Genomics* 19, pp. 1–8. DOI: `10.1186/s12864-018-4787-6`.

Heikkilä, A.-M., E. Liski, S. Pyörälä, and S. Taponen (2018). "Pathogen-specific production losses in bovine mastitis". In: *Journal of Dairy Science* 101, pp. 9493–9504. DOI: `10.3168/jds.2018-14824`.

Hertl, J. A., Y. H. Schukken, F. L. Welcome, L. W. Tauer, and Y. T. Gröhn (2014). "Pathogen-specific effects on milk yield in repeated clinical mastitis episodes in Holstein dairy cows". In: *Journal of Dairy Science* 97, pp. 1465–1480. DOI: `10.3168/jds.2013-7266`.

Hogeveen, Henk, Wilma Steeneveld, and Christopher A. Wolf (2019). "Production Diseases Reduce the Efficiency of Dairy Production: A Review of the Results, Methods, and Approaches Regarding the Economics of Mastitis". In: *Annual Review of Resource Economics* 11.Volume 11, 2019, pp. 289–312. ISSN: 1941-1359. DOI: `https://doi.org/10.1146/annurev-resource-100518-093954`.

Holling, C. S. (1973). "Resilience and Stability of Ecological Systems". In: *Annual Review of Ecology, Evolution, and Systematics* 4, pp. 1–23. DOI: `10.1146/annurev.es.04.110173.000245`.

Hou, Y., D. M. Bickhart, H. Chung, J. L. Hutchison, H. D. Norman, E. E. Connor, et al. (2012). "Analysis of copy number variations in Holstein cows identify potential mechanisms contributing to differences in residual feed intake". In: *Functional & Integrative Genomics* 12, pp. 717–723. DOI: `10.1007/s10142-012-0295-y`.

Hou, Y., G. E. Liu, D. M. Bickhart, M. F. Cardone, K. Wang, E. Kim, et al. (2011). “Genomic characteristics of cattle copy number variations”. In: *BMC Genomics* 12, pp. 1–11. DOI: `10.1186/1471-2164-12-127`.

Huxley, J.N. (2013). “Impact of lameness and claw lesions in cows on health and production”. In: *Livestock Science* 156.1. Lameness and claw lesions in sows, cows and small ruminants, pp. 64–70. ISSN: 1871-1413. DOI: `https://doi.org/10.1016/j.livsci.2013.06.012`.

Jiang, L., J. Jiang, J. Yang, X. Liu, J. Wang, H. Wang, et al. (2013). “Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins”. In: *BMC Genomics* 14, pp. 1–10. DOI: `10.1186/1471-2164-14-131`.

Jiang, L., J. Liu, D. Sun, P. Ma, X. Ding, Y. Yu, et al. (2010). “Genome wide association studies for milk production traits in Chinese Holstein population”. In: *PLoS One* 5, e13661. DOI: `10.1371/journal.pone.0013661`.

Kašná, E., L. Zavadilová, J. Vařeka, and J. Kyselová (2022). “General resilience in dairy cows: A review”. In: *Czech Journal of Animal Science* 67.12.

Kauf, A. C. W., R. F. Rosenbusch, M. J. Paape, and D. D. Bannerman (2007). “Innate immune response to intramammary Mycoplasma bovis infection”. In: *Journal of Dairy Science* 90, pp. 3336–3348. DOI: `10.3168/jds.2007-0058`.

Kennedy, A. D. and F. R. DeLeo (2009). “Neutrophil apoptosis and the resolution of infection”. In: *Immunologic Research* 43, pp. 25–61. DOI: `10.1007/s12026-008-8049-6`.

Keßler, F., R. Wellmann, M. G. G. Chagunda, and J. Bennewitz (2025a). “Toward a resilience selection index with indicator traits in German Holstein dairy cattle”. In: *Journal of dairy science* 108.1, pp. 726–734. DOI: `10.3168/jds.2024-25323`.

Keßler, Franziska, Maximilian Zölch, Robin Wellmann, and Jörn Bennewitz (Oct. 2025b). “Mapping genes for resilient dairy cows by means of across-breed genome-wide association analysis”. In: *BMC Genomics* 26.1, p. 879. DOI: `10.1186/s12864-025-11940-z`. URL: `https://doi.org/10.1186/s12864-025-11940-z`.

Kessler, F., R. Wellmann, M. G. G. Chagunda, and J. Bennewitz (2024). “Resilience indicator traits in 3 dairy cattle breeds in Baden-Württemberg”. In: *JOURNAL OF DAIRY SCIENCE* 107.6, pp. 3780–3793. ISSN: 0022-0302. DOI: `10.3168/jds.2023-24305`.

Khatib, H. (2005). “The COPG2, DCN, and SDHD genes are biallelically expressed in cattle”. In: *Mammalian Genome* 16, pp. 545–552. DOI: `10.1007/s00335-005-0029-8`.

Kiser, J. N., T. E. Lawrence, M. Neupane, C. M. Seabury, J. F. Taylor, J. E. Womack, et al. (2017). “Rapid Communication: Subclinical bovine respiratory disease - loci and pathogens associated with lung lesions in feedlot cattle”. In: *Journal of Animal Science* 95, pp. 2726–2731. DOI: `10.2527/jas.2017.1548`.

Kiser, J. N., M. Neupane, S. N. White, and H. L. Neibergs (2018). “Identification of genes associated with susceptibility to Mycobacterium avium ssp. paratuberculosis (Map) tissue infection in Holstein cattle using gene set enrichment analysis-SNP”. In: *Mammalian Genome* 29, pp. 539–549. DOI: `10.1007/s00335-017-9725-4`.

Koenker, R., S. Portnoy, P. T. Ng, A. Zeileis, P. Grosjean, and B. D. Ripley (2018). *Package ’quantreg’*. Reference manual available at R-CRAN: `https://cran.rproject.org/web/packages/quantreg/quantreg.pdf`.

Koenker, Roger (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press.

— (2023). *quantreg: Quantile Regression*. R package version 5.97. URL: `https://CRAN.R-project.org/package=quantreg`.

Kołoszycz, Ewa, Artur Wilczyński, Alberto Menghi, Chiara Serena Soffiantini, and Marija Klopčič (2024). “Searching for Dairy Farm Resilience with the R4D Project: Innovation Needs to Be Sustainable”. In: *Sustainability* 16.9, p. 3520. DOI: `10.3390/su16093520`.

König, S. and K. May (2019). “Invited review: Phenotyping strategies and quantitative-genetic background of resistance, tolerance and resilience associated traits in dairy cattle”. In: *Animal* 13.5, pp. 897–908. ISSN: 1751-7311. DOI: `https://doi.org/10.1017/S1751731118003208`.

Kropatsch, R., G. Dekomien, D. A. Akkad, W. M. Gerding, E. Petrasch-Parwez, N. D. Young, et al. (2013). “SOX9 duplication linked to intersex in deer”. In: *PLoS One* 8, e73734. DOI: `10.1371/journal.pone.0073734`.

Kunej, T., M. Šimon, B. Luštrek, S. Horvat, and K. Potočnik (2024). “Examining genotype-phenotype associations of GRAM domain proteins using GWAS, PheWAS and literature review in cattle, human, pig, mouse and chicken”. In: *Scientific Reports* 14, p. 28889. DOI: `10.1038/s41598-024-80117-7`.

Laodim, T., M. A. Elzo, S. Koonawootrittriron, T. Suwanasopee, and D. Jattawa (2017). “Identification of SNP markers associated with milk and fat yields in multibreed dairy cattle using two genetic group structures”. In: *Livestock Science* 206, pp. 95–104. DOI: `10.1016/j.livsci.2017.10.015`.

Lathuillière, M.J., R. Flach, L. Wang-Erlandsson, et al. (2025). “International reliance on Brazil’s water through soy and beef supply chains”. In: *Communications Earth & Environment* 6, p. 688. DOI: `10.1038/s43247-025-02658-7`.

Lee, J.-W., D. D. Bannerman, M. J. Paape, M.-K. Huang, and X. Zhao (2006). “Characterization of cytokine expression in milk somatic cells during intramammary infections with Escherichia coli or Staphylococcus aureus by real-time PCR”. In: *Veterinary Research* 37, pp. 219–229. DOI: `10.1051/vetres:2005051`.

Lee, Y.-L., M. Bosse, E. Mullaart, M. A. M. Groenen, R. F. Veerkamp, and A. C. Bouwman (2020). “Functional and population genetic features of copy number variations in two dairy cattle populations". In: *BMC Genomics* 21, p. 89. DOI: `10.1186/s12864-020-6496-1`.

Lee, Y.-L., H. Takeda, G. Costa Monteiro Moreira, L. Karim, E. Mullaart, W. Coppieters, et al. (2021). "A 12 kb multi-allelic copy number variation encompassing a GC gene enhancer is associated with mastitis resistance in dairy cattle". In: *PLoS Genetics* 17, e1009331. DOI: `10.1371/journal.pgen.1009331`.

Leip, Adrian, Gilles Billen, Josette Garnier, Bruna Grizzetti, Luis Lassaletta, Stefan Reis, David Simpson, Mark A Sutton, Wim de Vries, Franz Weiss, and Henk Westhoek (Nov. 2015). "Impacts of European livestock production: nitrogen, sulphur, phosphorus and greenhouse gas emissions, land-use, water eutrophication and biodiversity". In: *Environmental Research Letters* 10.11, p. 115004. DOI: `10.1088/1748-9326/10/11/115004`.

Leng, Guoyong and Jim W. Hall (2021). "Where is the Planetary Boundary for freshwater being exceeded because of livestock farming?" In: *Science of The Total Environment* 760, p. 144035. ISSN: 0048-9697. DOI: `https://doi.org/10.1016/j.scitotenv.2020.144035`.

Lescourret, Francoise and Jean Baptiste Coulon (1994). "Modeling the impact of mastitis on milk production by dairy cows". In: *Journal of Dairy Science* 77.8, pp. 2289–2301.

Li, Chaohui, Prajal Pradhan, Xudong Wu, Zhi Li, Jingyu Liu, Klaus Hubacek, and Guoqian Chen (2024). "Livestock sector can threaten planetary boundaries without regionally differentiated strategies". In: *Journal of Environmental Management* 370, p. 122444. ISSN: 0301-4797. DOI: `https://doi.org/10.1016/j.jenvman.2024.122444`.

Liang, D., L. M. Arnold, C. J. Stowe, R. J. Harmon, and J. M. Bewley (2017). "Estimating US dairy clinical disease costs with a stochastic simulation model". In: *Journal of Dairy Science* 100, pp. 1472–1486. DOI: `10.3168/jds.2016-11565`.

Lipkin, E., M. G. Strillacci, H. Eitam, M. Yishay, F. Schiavini, M. Soller, et al. (2016). "The use of Kosher phenotyping for mapping QTL affecting susceptibility to Bovine respiratory disease". In: *PLoS One* 11. DOI: `10.1371/journal.pone.0153423`.

Liu, Y. F., L. S. Zan, W. T. Cui, Y. P. Xin, Y. Jiao, and K. Li (2011). "Molecular cloning, characterization and association analysis of the promoter region of the bovine CDK6 gene". In: *Genetics and Molecular Research* 10, pp. 1777–1786. DOI: `10.4238/vol10-3gmr1466`.

Livernois, A. M., B. A. Mallard, S. L. Cartwright, and A. Cánovas (2021). "Heat stress and immune response phenotype affect DNA methylation in blood mononuclear cells from Holstein dairy cows". In: *Scientific Reports* 11, p. 11371. DOI: `10.1038/s41598-021-89951-5`.

Llonch, Pol, Eva Mainau, Ignacio R. Ipharraguerre, Fernando Bargo, Gemma Tedó, Marta Blanch, and Xavier Manteca (2018). "Chicken or the Egg: The Reciprocal Association Between Feeding Behavior and Animal Welfare and Their Impact on Productivity in

Dairy Cows”. In: *Frontiers in Veterinary Science* Volume 5 - 2018. ISSN: 2297-1769. DOI: `10.3389/fvets.2018.00305`.

Mancin, E., C. Sartori, E. Rulli, A. Oian, G. Gomez Proto, F. Tiezzi, and R. Mantovani (2025). “Indicators of resilience in cattle based on test-day milk yield: comparison of within- and between-lactation and their relationship with milk production”. In: *animal* 19.9, p. 101596. ISSN: 1751-7311. DOI: `doi.org/10.1016/j.animal.2025.101596`. URL: `https://www.sciencedirect.com/science/article/pii/S175173112500179X`.

Mantilla Valdivieso, E. F., E. M. Ross, A. Raza, M. N. Naseem, M. Kamran, B. J. Hayes, et al. (2022). “Transcriptional changes in the peripheral blood leukocytes from Brangus cattle before and after tick challenge with Rhipicephalus australis”. In: *BMC Genomics* 23, p. 454. DOI: `10.1186/s12864-022-08686-3`.

Marçal-Pedroza, M. G., M. M. Campos, M. F. Martins, M. V. B. Silva, M. J. R. Da Paranhos Costa, J. A. Negrão, et al. (2023). “Is the temperament of crossbred dairy cows related to milk cortisol and oxytocin concentrations, milk yield, and quality?” In: *PLoS One* 18, e0286466. DOI: `10.1371/journal.pone.0286466`.

Margareto, J., O. Leis, E. Larrarte, I. C. Pomposo, J. M. Garibi, and J. V. Lafuente (2009). “DNA copy number variation and gene expression analyses reveal the implication of specific oncogenes and genes in GBM”. In: *Cancer Investigation* 27, pp. 541–548. DOI: `10.1080/07357900802563044`.

Maskal, Jacob M., Victor B. Pedrosa, Hinayah Rojas de Oliveira, and Luiz F. Brito (May 2024). “A comprehensive meta-analysis of genetic parameters for resilience and productivity indicator traits in Holstein cattle”. In: *Journal of Dairy Science* 107.5, pp. 3062–3079. DOI: `10.3168/jds.2023-23668`.

Medugorac, I. and M. Soller (2001). “Selective genotyping with a main trait and a correlated trait”. In: *Journal of Animal Breeding and Genetics* 118.5, pp. 285–295. DOI: `10.1046/j.1439-0388.2001.00308.x`.

Mesbah-Uddin, M., B. Guldbrandtsen, A. Capitan, M. S. Lund, D. Boichard, and G. Sahana (2022). “Genome-wide association study with imputed whole-genome sequence variants including large deletions for female fertility in 3 Nordic dairy cattle breeds”. In: *Journal of Dairy Science* 105, pp. 1298–1313. DOI: `10.3168/jds.2021-20655`.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard (2001). “Prediction of total genetic value using genome-wide dense marker maps”. In: *Genetics* 157, pp. 1819–1829. DOI: `10.1093/genetics/157.4.1819`.

Meyer, David, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel, and Friedrich Leisch (2023). *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien.* R package version 1.7-14. URL: `https://CRAN.R-project.org/package=e1071`.

Miglior, Filippo, Allison Fleming, Francesca Malchiodi, Luiz F. Brito, Pauline Martin, and Christine F. Baes (Dec. 2017). “A 100-year review: Identification and genetic selection of economically important traits in dairy cattle". In: *Journal of Dairy Science* 100.12, pp. 10251–10271. DOI: `10.3168/jds.2017-12968`. URL: `https://doi.org/10.3168/jds.2017-12968`.

Mills, R. E., K. Walter, C. Stewart, R. E. Handsaker, K. Chen, C. Alkan, et al. (2011). "Mapping copy number variation by population-scale genome sequencing". In: *Nature* 470, pp. 59–65. DOI: `10.1038/nature09708`.

Min, Li, Dagang Li, Xiong Tong, Xuemei Nan, Diyun Ding, Bin Xu, and Gang Wang (Sept. 2019). "Nutritional strategies for alleviating the detrimental effects of heat stress in dairy cows: a review". In: *International Journal of Biometeorology* 63.9. Epub 2019 Jun 19, pp. 1283–1302. DOI: `10.1007/s00484-019-01744-8`.

Min, Seung Hyun and Jiawei Zhou (2021). "smplot: an R package for easy and elegant data visualization". In: *Frontiers in Genetics* 12, p. 802894. DOI: `10.3389/fgene.2021.802894`.

Misztal, I., S. Tsuruta, D. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica (2018). *Manual for BLUPF90 family of programs*. University of Georgia. Athens, USA.

Mizuno, Y., Y. Sotomaru, Y. Katsuzawa, T. Kono, M. Meguro, M. Oshimura, et al. (2002). "Asb4, Ata3, and Dcn are novel imprinted genes identified by high-throughput screening using RIKEN cDNA microarray". In: *Biochemical and Biophysical Research Communications* 290, pp. 1499–1505. DOI: `10.1006/bbrc.2002.6370`.

Moghaddam, S. H. H., M. Moridi, and S. Z. Mirhoseini (2019). "IL1B gene expression and association with SCS in Guilan native and Holstein cow populations and their reciprocal crosses". In: *Journal of Applied Biological Sciences* 13, pp. 150–153.

Mohammadi, A., S. Alijani, S. A. Rafat, and R. Abdollahi-Arpanahi (2020). "Genome-Wide Association Study and Pathway Analysis for Female Fertility Traits in Iranian Holstein Cattle". In: *Annals of Animal Science* 20, pp. 825–851. DOI: `10.2478/aoas-2020-0031`.

Monteleone, Beatrice and Iolanda Borzí (2024). "Drought in the Po Valley: Identification, Impacts and Strategies to Manage the Events". In: *Water* 16.8. ISSN: 2073-4441. DOI: `10.3390/w16081187`.

Morris, R. F. (1963). "The Dynamics of Epidemic Spruce Budworm Populations". In: *Memoirs of the Entomological Society of Canada* 95.S31, pp. 1–12. DOI: `10.4039/entm9531fv`.

Nelder, J. A. (1966). "Inverse Polynomials, a Useful Group of Multi-Factor Response Functions". In: *Biometrics* 22.1, pp. 128–141. ISSN: 0006341X, 15410420. URL: `http://www.jstor.org/stable/2528220` (visited on 08/07/2025).

Neupane, M., J. N. Kiser, and H. L. Neibergs (2018). "Gene set enrichment analysis of SNP data in dairy and beef cattle with bovine respiratory disease". In: *Animal Genetics* 49, pp. 527–538. DOI: `10.1111/age.12718`.

Nguyen-Ba, H., M. Taghipoor, and J. van Milgen (2020). “Modelling the feed intake response of growing pigs to diets contaminated with mycotoxins”. In: *Animal* 14, s303–s312. DOI: `10.1017/S175173112000083X`.

Niloofar, P., D. P. Francis, S. Lazarova-Molnar, A. Vulpe, M.-C. Vochin, G. Suciu, et al. (2021). “Data-driven decision support in livestock farming for improved animal health, welfare and greenhouse gas emissions: Overview and challenges”. In: *Computers and Electronics in Agriculture* 190, p. 106406. DOI: `10.1016/j.compag.2021.106406`.

Nugrahaeningtyas, Eska, Jong-Sik Lee, and Kyu-Hyun Park (Nov. 2024). “Greenhouse gas emissions from livestock: sources, estimation, and mitigation”. In: *Journal of Animal Science and Technology* 66.6. Epub 2024 Nov 30. PMID: 39691613; PMCID: PMC11647415, pp. 1083–1098. ISSN: 2055-0391. DOI: `10.5187/jast.2024.e86`.

OECD and FAO (2025). *OECD-FAO Agricultural Outlook 2025-2034*. Paris/Rome: OECD Publishing/FAO. DOI: `10.1787/601276cd-en`. URL: `https://doi.org/10.1787/601276cd-en`.

Oliveira, Charles Paranhos, Fernanda Campos de Sousa, Alex Lopes da Silva, Érica Beatriz Schultz, Roger Iván Valderrama Londoño, and Pedro Antônio Reinoso de Souza (2025). “Heat Stress in Dairy Cows: Impacts, Identification, and Mitigation Strategies—A Review”. In: *Animals* 15.2. ISSN: 2076-2615. DOI: `10.3390/ani15020249`.

Olsson, M., J. R. S. Meadows, K. Truve, G. R. Pielberg, F. Puppo, E. Mauceli, et al. (2011). “A novel unstable duplication upstream of HAS2 predisposes to a breed-defining skin phenotype and a periodic fever syndrome in Chinese Shar-Pei dogs”. In: *PLoS Genetics* 7, e1001332. DOI: `10.1371/journal.pgen.1001332`.

Online Etymology Dictionary (2025). Accessed: October 21, 2025. Online Etymology Dictionary. URL: `https://www.etymonline.com/search?q=resilien` (visited on 2025).

Parlasca, Martin C. and Matin Qaim (2022). “Meat Consumption and Sustainability”. In: *Annual Review of Resource Economics* 14.Volume 14, 2022, pp. 17–41. ISSN: 1941-1359. DOI: `10.1146/annurev-resource-111820-032340`.

Perrin, Augustine, Rebecka Milestad, and Guillaume Martin (2020). “Resilience applied to farming: organic farmers’ perspectives”. In: *Ecology and Society* 25.4, p. 5. DOI: `10.5751/ES-11897-250405`.

Perry, G. H., N. J. Dominy, K. G. Claw, A. S. Lee, H. Fiegler, R. Redon, et al. (2007). “Diet and the evolution of human amylase gene copy number variation”. In: *Nature Genetics* 39, pp. 1256–1260. DOI: `10.1038/ng2123`.

Pierce, M. D., K. Dzama, and F. C. Muchadeyi (2018). “Genetic diversity of seven cattle breeds inferred using copy number variations”. In: *Frontiers in Genetics* 9, p. 163. DOI: `10.3389/fgene.2018.00163`.

Pinto, D., K. Darvishi, X. Shi, D. Rajan, D. Rigler, T. Fitzgerald, et al. (2011). "Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants". In: *Nature Biotechnology* 29, pp. 512–520. DOI: `10.1038/nbt.1852`.

Poppe, M., G. Bonekamp, M. L. van Pelt, and H. A. Mulder (2021a). "Genetic analysis of resilience indicators based on milk yield records in different lactations and at different lactation stages". In: *Journal of Dairy Science* 104, pp. 1967–1981. DOI: `10.3168/jds.2020-19245`.

Poppe, M., H. A. Mulder, M. L. van Pelt, E. Mullaart, H. Hogeveen, and R. F. Veerkamp (2022a). "Development of resilience indicator traits based on daily step count data for dairy cattle breeding". In: *Genetics Selection Evolution* 54, p. 21.

Poppe, M., H. A. Mulder, and R. F. Veerkamp (2021b). "Validation of resilience indicators by estimating genetic correlations among daughter groups and with yield responses to a heat wave and disturbances at herd level". In: *Journal of Dairy Science* 104, pp. 8094–8106. DOI: `10.3168/jds.2020-19817`.

Poppe, M., R. F. Veerkamp, M. L. Van Pelt, and H. A. Mulder (2020). "Exploration of variance, autocorrelation, and skewness of deviations from lactation curves as resilience indicators for breeding". In: *Journal of Dairy Science* 103.2, pp. 1667–1684.

Poppe, Marieke, Han A. Mulder, Mathijs L. van Pelt, Erik Mullaart, Henk Hogeveen, and Roel F. Veerkamp (2022b). "Development of resilience indicator traits based on daily step count data for dairy cattle breeding". In: *Genetics Selection Evolution* 54.1, p. 21. ISSN: 1297-9686. DOI: `10.1186/s12711-022-00713-x`.

Prinsen, R. T. M. M., M. G. Strillacci, F. Schiavini, E. Santus, A. Rossoni, V. Maurer, et al. (2016). "A genome-wide scan of copy number variants using high-density SNPs in Brown Swiss dairy cattle". In: *Livestock Science* 191. DOI: `10.1016/j.livsci.2016.08.006`.

Punturiero, C., A. Delledonne, C. Ferrari, A. Bagnato, and M. G. Strillacci (2025). "Mapping genomic regions affecting resilience traits in a large dairy farm of Holstein cows". In: *Frontiers in Animal Science* Volume 6 - 2025. ISSN: 2673-6225. DOI: `10.3389/fanim.2025.1627086`.

Punturiero, C., R. Milanesi, F. Bernini, A. Delledonne, A. Bagnato, and M. G. Strillacci (2023). "Genomic approach to manage genetic variability in dairy farms". In: *Italian Journal of Animal Science* 22, pp. 769–783. DOI: `10.1080/1828051X.2023.2243977`.

Quinlan, A. R. and I. M. Hall (2010). "BEDTools: a flexible suite of utilities for comparing genomic features". In: *Bioinformatics* 26, pp. 841–842. DOI: `10.1093/bioinformatics/btq033`.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: `https://www.R-project.org/`.

Rajala-Schultz, PJ, YT Gröhn, CE McCulloch, and CL Guard (1999). "Effects of clinical mastitis on milk yield in dairy cows". In: *Journal of dairy science* 82.6, pp. 1213–1220.

Ramanathan, A., M. Weintraub, N. Orlovetskie, R. Serruya, D. Mani, O. Marcu, et al. (2020). “A mutation in POLR3E impairs antiviral immune response and RNA polymerase III”. In: *Proceedings of the National Academy of Sciences* 117, pp. 22113–22121. DOI: `10.1073/pnas.2009947117`.

Ramsay, J. O., N. Heckman, and B. W. Silverman (Mar. 1997). “Spline smoothing with model-based penalties”. In: *Behavior Research Methods, Instruments, & Computers* 29.1, pp. 99–106. ISSN: 1532-5970. DOI: `10.3758/BF03200573`.

Ramsey, J. and B. Ripley (2024). *pspline: Penalized Smoothing Splines.* R package version 1.0-21. URL: `https://CRAN.R-project.org/package=pspline` (visited on 08/12/2025).

Ranzato, G., B. Aernouts, I. Lora, I. Adriaens, A. Ben Abdelkrim, M.J. Gote, and G. Cozzi (2024). “Comparison of 3 mathematical models to estimate lactation performance in dairy cows”. In: *Journal of Dairy Science* 107.9, pp. 6888–6901. ISSN: 0022-0302. DOI: `https://doi.org/10.3168/jds.2023-24224`.

Raphaka, K., O. Matika, E. Sánchez-Molano, R. Mrode, M. P. Coffey, V. Riggio, et al. (2017). “Genomic regions underlying susceptibility to bovine tuberculosis in Holstein-Friesian cattle”. In: *BMC Genetics* 18, pp. 1–10. DOI: `10.1186/s12863-017-0493-7`.

Rauw, W.M, E Kanis, E.N Noordhuizen-Stassen, and F.J Grommers (1998). “Undesirable side effects of selection for high production efficiency in farm animals: a review”. In: *Livestock Production Science* 56.1, pp. 15–33. ISSN: 0301-6226. DOI: `https://doi.org/10.1016/S0301-6226(98)00147-X`. URL: `https://www.sciencedirect.com/science/article/pii/S030162269800147X`.

Resilience dictionary (2025). Accessed: 2025-08-29. URL: `https://www.stockholmresilience.org/research/resilience-dictionary.html` (visited on 08/29/2025).

Rial-Pensado, E., V. Rivas-Limeres, C. Grijota-Martínez, A. Rodríguez-Díaz, V. Capelli, O. Barca-Mayo, et al. (2022). “Temperature modulates systemic and central actions of thyroid hormones on BAT thermogenesis”. In: *Frontiers in Physiology* 13, p. 1017381. DOI: `10.3389/fphys.2022.1017381`.

Richardson, K., W. Steffen, W. Lucht, J. Bendtsen, S. E. Cornell, J. F. Donges, M. Drüke, I. Fetzer, G. Bala, W. von Bloh, G. Feulner, S. Fiedler, D. Gerten, T. Gleeson, M. Hofmann, W. Huiskamp, M. Kummu, C. Mohan, D. Nogués-Bravo, S. Petri, M. Porkka, S. Rahmstorf, S. Schaphoff, K. Thonicke, A. Tobian, V. Virkki, L. Wang-Erlandsson, L. Weber, and J. Rockström (2023). “Earth beyond six of nine planetary boundaries”. In: *Science Advances* 9.37, eadh2458. DOI: `10.1126/sciadv.adh2458`.

Robakowska-Hyżorek, D., R. R. Starzyński, B. Żelazowska, J. Oprządek, T. Sadkowski, and L. Zwierzchowski (2016). “Effects of polymorphism in the bovine PTPRQ gene on the expression of MYF6 and MYF5 genes in skeletal muscle and on meat production traits in beef bulls”. In: *Czech Journal of Animal Science* 61, pp. 232–242. DOI: `10.17221/118/2014-CJAS`.

Rockström, J., W. Steffen, K. Noone, Å Persson, F. S. Chapin III, E. Lambin, T. M. Lenton, M. Scheffer, C. Folke, H. J. Schellnhuber, et al. (2009a). "A safe operating space for humanity". In: *nature* 461.7263, pp. 472–475.

— (2009b). "Planetary boundaries: exploring the safe operating space for humanity". In: *Ecology and society* 14.2.

Salleh, M. S., G. Mazzoni, J. K. Höglund, D. W. Olijhoek, P. Lund, P. Løvendahl, et al. (2017). "RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high- and low-residual feed intake in Nordic dairy cattle". In: *BMC Genomics* 18, p. 258. DOI: `10.1186/s12864-017-3622-9`.

Sanchez, M., Y. Lin, C.-C. Yang, P. McQuary, A. R. Campos, P. A. Blanc, et al. (2019). "Cross talk between eIF2alfa and eEF2 phosphorylation pathways optimizes translational arrest in response to oxidative stress". In: *iScience* 20, pp. 466–480. DOI: `10.1016/j.isci.2019.09.031`.

Sassi, N. Ben, Ó. González-Recio, R. de Paz-Del Río, S. T. Rodríguez-Ramilo, and A. I. Fernández (2016). "Associated effects of copy number variants on economically important traits in Spanish Holstein dairy cattle". In: *Journal of Dairy Science* 99, pp. 6371–6380. DOI: `10.3168/jds.2015-10487`.

Sauer, J. et al. (2024). "Characterising farming resilience capacities in the face of drought and animal diseases: Findings from Australia, Estonia and France". In: *OECD Food, Agriculture and Fisheries Papers* 211. DOI: `10.1787/d3d592c9-en`.

Scheffer, M., J. E. Bolhuis, D. Borsboom, T. G. Buchman, S. M. W. Gijzel, D. Goulson, J. E. Kammenga, B. Kemp, I. A. van de Leemput, S. Levin, C. M. Martin, R. J. F. Melis, E. H. van Nes, L. M. Romero, and M. G. M. Olde Rikkert (2018). "Quantifying resilience of humans and other animals". In: *Proceedings of the National Academy of Sciences* 115.47, pp. 11883–11890. DOI: `10.1073/pnas.1810630115`.

Schmidtmann, C., J. Mugambe, I. Blaj, C. Harms, and G. Thaller (2024). "Genetic investigations on backfat thickness and body condition score in German Holstein cattle". In: *Journal of Animal Breeding and Genetics* 141, pp. 602–613. DOI: `10.1111/jbg.12867`.

Schober, Patrick, Christa Boer, and Lothar A. Schwarte (May 2018). "Correlation Coefficients: Appropriate Use and Interpretation". In: *Anesthesia & Analgesia* 126.5, pp. 1763–1768. DOI: `10.1213/ANE.0000000000002864`.

Schramek, H., R. Sarközi, C. Lauterberg, A. Kronbichler, M. Pirklbauer, R. Albrecht, et al. (2009). "Neuropilin-1 and neuropilin-2 are differentially expressed in human proteinuric nephropathies and cytokine-stimulated proximal tubular cells". In: *Laboratory Investigation* 89, pp. 1304–1316. DOI: `10.1038/labinvest.2009.96`.

Sejian, V., R. Bhatta, J. B. Gaughan, F. R. Dunshea, and N. Lacetera (2018). "Review: Adaptation of animals to heat stress". In: *Animal* 12, s431–s444. DOI: `10.1017/S1751731118001945`.

Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, et al. (2003). "Cytoscape: a software environment for integrated models of biomolecular interaction networks". In: *Genome Research* 13, pp. 2498–2504. DOI: `10.1101/gr.1239303`.

Silpa, M. V., S. König, V. Sejian, P. K. Malik, M. R. R. Nair, V. F. C. Fonseca, A. S. C. Maia, and R. Bhatta (2021). "Climate-Resilient Dairy Cattle Production: Applications of Genomic Tools and Statistical Models". In: *Frontiers in Veterinary Science* Volume 8 - 2021. ISSN: 2297-1769. DOI: `10.3389/fvets.2021.625189`.

Silva, R. P., M. P. Berton, L. Grigoletto, F. E. Carvalho, R. M. O. Silva, E. Peripolli, et al. (2019). "Genomic regions and enrichment analyses associated with carcass composition indicator traits in Nellore cattle". In: *Journal of Animal Breeding and Genetics* 136, pp. 118–133. DOI: `10.1111/jbg.12373`.

Sinha, R., B. Sinha, R. Kumari, V. M R, N. Sharma, A. Verma, et al. (2023). "Association of udder type traits with single nucleotide polymorphisms in Sahiwal (Bos indicus) and Karan Fries (Bos taurus x Bos indicus) cattle". In: *Animal Biotechnology* 34, pp. 2745–2756. DOI: `10.1080/10495398.2022.2114083`.

Stepancheva, T., I. Marinov, and Z. Gergovska (2024). "Milking Temperament and Its Association with Test-Day Milk Yield in Bulgarian Murrah Buffaloes". In: *Animals* 14. DOI: `10.3390/ani14070987`.

Stepensky, P., B. Keller, M. Buchta, A.-K. Kienzler, O. Elpeleg, R. Somech, et al. (2013). "Deficiency of caspase recruitment domain family, member 11 (CARD11), causes profound combined immunodeficiency in human subjects". In: *Journal of Allergy and Clinical Immunology* 131, 477–85.e1. DOI: `10.1016/j.jaci.2012.11.050`.

Stranger, B. E., M. S. Forrest, M. Dunning, C. E. Ingle, C. Beazlsy, N. Thorne, et al. (2007). "Relative impact of nucleotide and copy number variation on gene expression phenotypes". In: *Science* 315, pp. 848–853. DOI: `10.1126/science.1136678`.

Strillacci, M. G., E. Gorla, M. C. Cozzi, M. Vevey, F. Genova, K. Scienski, et al. (2018). "A copy number variant scan in the autochthonous Valdostana Red Pied cattle breed and comparison with specialized dairy populations". In: *PLoS One* 13, pp. 1–18. DOI: `10.1371/journal.pone.0204669`.

Suchocki, T. and J. Szyda (2015). "Genome-wide association study for semen production traits in Holstein-Friesian bulls". In: *Journal of Dairy Science* 98, pp. 5774–5780. DOI: `10.3168/jds.2014-8951`.

Szyda, J., M. Mielczarek, M. Frąszczak, G. Minozzi, J. L. Williams, and K. Wojdak-Maksymiec (2019). "The genetic background of clinical mastitis in Holstein-Friesian cattle". In: *Animal* 13, pp. 2156–2163. DOI: `10.1017/S1751731119000338`.

Taghipoor, M., M. Pastell, O. Martin, H. Nguyen Ba, J. van Milgen, A. Doeschl-Wilson, C. Loncke, N. C. Friggens, L. Puillet, and R. Muñoz-Tamayo (2023). "Animal board invited review: Quantification of resilience in farm animals". In: *Animal : an international journal of animal bioscience* 17.9. DOI: `10.1016/j.animal.2023.100925`.

Thankachan, J. M. and S. R. G. Setty (2022). “KIF13A-A Key Regulator of Recycling Endosome Dynamics”. In: *Frontiers in Cell and Developmental Biology* 10, p. 877532. DOI: `10.3389/fcell.2022.877532`.

The Lancet Planetary Health (2025). “Evidence with a quiet voice”. In: *The Lancet Planetary Health* 9.9. ISSN: 2542-5196. DOI: `10.1016/j.lanplh.2025.101371`.

Thornton, P., G. Nelson, D. Mayberry, and M. Herrero (Mar. 2022). “Impacts of heat stress on global cattle production during the 21st century: a modelling study”. In: *The Lancet Planetary Health* 6.3, e192–e201. ISSN: 2542-5196. DOI: `10.1016/S2542-5196(22)00002-X`.

Tiezzi, F., K. L. Parker-Gaddis, J. B. Cole, J. S. Clay, and C. Maltecca (2015). “A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure”. In: *PLoS One* 10, e0114919. DOI: `10.1371/journal.pone.0114919`.

Tilman, David, Christian Balzer, Jason Hill, and Belinda L. Befort (2011). “Global food demand and the sustainable intensification of agriculture”. In: *Proceedings of the National Academy of Sciences of the United States of America* 108.50, pp. 20260–20264. ISSN: 00278424. URL: `http://www.jstor.org/stable/23060109` (visited on 08/29/2025).

Tizioto, P. C., J. Kim, C. M. Seabury, R. D. Schnabel, L. J. Gershwin, A. L. Van Eenennaam, et al. (2015). “Immunological response to single pathogen challenge with agents of the bovine respiratory disease complex: an RNA-sequence analysis of the bronchial lymph node transcriptome”. In: *PLoS One* 10, e0131459.

Tolleson, M. W., C. A. Gill, A. D. Herring, P. K. Riggs, J. E. Sawyer, J. O. Sanders, et al. (2017). “Association of udder traits with single nucleotide polymorphisms in crossbred Bos indicus-Bos taurus cows”. In: *Journal of Animal Science* 95, pp. 2399–2407. DOI: `10.2527/jas.2017.1475`.

Turner, S. (2018). “qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots”. In: *The Journal of Open Source Software*. DOI: `10.21105/joss.00731`.

Van Beek, E. M., F. Cochrane, A. N. Barclay, and T. K. van den Berg (2005). “Signal regulatory proteins in the immune system”. In: *The Journal of Immunology* 175, pp. 7781–7787. DOI: `10.4049/jimmunol.175.12.7781`.

Venhoranta, H., H. Pausch, M. Wysocki, I. Szczerbal, R. Hänninen, J. Taponen, et al. (2013). “Ectopic KIT copy number variation underlies impaired migration of primordial germ cells associated with gonadal hypoplasia in cattle (Bos taurus)”. In: *PLoS One* 8, e75659. DOI: `10.1371/journal.pone.0075659`.

Wang, A., L. F. Brito, H. Zhang, R. Shi, L. Zhu, D. Liu, et al. (2022a). “Exploring milk loss and variability during environmental perturbations across lactation stages as resilience indicators in Holstein cattle". In: *Frontiers in Genetics* 13, p. 1031557. DOI: `10.3389/fgene.2022.1031557`.

Wang, H., J. Zhong, J. Wang, Z. Chai, C. Zhang, J. Xin, et al. (2021). "Whole-Transcriptome Analysis of Yak and Cattle Heart Tissues Reveals Regulatory Pathways Associated With High-Altitude Adaptation". In: *Frontiers in Genetics* 12, p. 579800. DOI: `10.3389/fgene.2021.579800`.

Wang, K. and M. Bucan (2008). "Copy number variation detection via high-density SNP genotyping". In: *Cold Spring Harbor Protocols*. DOI: `10.1101/pdb.top46`.

Wang, P., X. Li, Y. Zhu, J. Wei, C. Zhang, Q. Kong, et al. (2022b). "Genome-wide association analysis of milk production, somatic cell score, and body conformation traits in Holstein cows". In: *Frontiers in Veterinary Science* 9, p. 932034.

Wathes, D. C., Z. Cheng, N. Bourne, V. J. Taylor, M. P. Coffey, and S. Brotherstone (2007). "Differences between primiparous and multiparous dairy cows in the inter-relationships between metabolic traits, milk yield and body condition score in the periparturient period". In: *Domestic Animal Endocrinology* 33, pp. 203–225. DOI: `10.1016/j.domaniend.2006.05.004`.

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

Wickramasinghe, S., S. Hua, G. Rincon, A. Islas-Trejo, J. B. German, C. B. Lebrilla, et al. (2011). "Transcriptome profiling of bovine milk oligosaccharide metabolism genes using RNA-sequencing". In: *PLoS One* 6, e18895. DOI: `10.1371/journal.pone.0018895`.

Wiggans, G. R., J. B. Cole, S. M. Hubbard, and T. S. Sonstegard (2017). "Genomic selection in dairy cattle: the USDA experience". In: *Annual Review of Animal Bioscience* 5, pp. 309–327. DOI: `10.1146/annurev-animal-021815-111422`.

Willett, W., J. Rockström, B. Loken, M. Springmann, T. Lang, S. Vermeulen, T. Garnett, D. Tilman, F. DeClerck, A. Wood, M. Jonell, M. Clark, L. J. Gordon, J. Fanzo, C. Hawkes, R. Zurayk, J. A. Rivera, W. De Vries, L. Majele Sibanda, A. Afshin, A. Chaudhary, M. Herrero, R. Agustina, F. Branca, A. Lartey, S. Fan, B. Crona, E. Fox, V. Bignet, M. Troell, T. Lindahl, S. Singh, S. E. Cornell, K. Srinath Reddy, S. Narain, S. Nishtar, and C. J. L. Murray (2019). "Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems". In: *The Lancet* 393.10170, pp. 447–492. ISSN: 0140-6736. DOI: `10.1016/S0140-6736(18)31788-4`.

Wilmink, J.B.M. (1987). "Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation". In: *Livestock Production Science* 16.4, pp. 335–348. ISSN: 0301-6226. DOI: `https://doi.org/10.1016/0301-6226(87)90003-0`.

Wood, P. D. P. (1967). "Algebraic Model of the Lactation Curve in Cattle". In: *Nature* 216.5111, pp. 164–165. ISSN: 1476-4687. DOI: `10.1038/216164a0`.

Wright, D., H. Boije, J. R. S. Meadows, B. Bed'Hom, D. Gourichon, A. Vieaud, et al. (2009). "Copy number variation in intron 1 of SOX5 causes the Pea-comb phenotype in chickens". In: *PLoS Genetics* 5, e1000512. DOI: `10.1371/journal.pgen.1000512`.

Wright, Sewall (1921). "Systems of mating". In: *Genetics* 6, pp. 111–178.

Wu, X., B. Guldbrandtsen, M. S. Lund, and G. Sahana (2016). "Association analysis for feet and legs disorders with whole-genome sequence variants in 3 dairy cattle breeds". In: *Journal of Dairy Science* 99, pp. 7221–7231. DOI: `10.3168/jds.2015-10705`.

Xing, C. and G. Xing (2009). "Power of selective genotyping in genome-wide association studies of quantitative traits". In: *BMC Proceedings*, S23.

Xu, L., J. B. Cole, D. M. Bickhart, Y. Hou, J. Song, P. M. VanRaden, et al. (2014). "Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins". In: *BMC Genomics* 15, pp. 1–10. DOI: `10.1186/1471-2164-15-683`.

Xu, L., Y. Hou, D. M. Bickhart, J. Song, and G. E. Liu (2013). "Comparative analysis of CNV calling algorithms: literature survey and a case study using bovine high-density SNP data". In: *Microarrays* 2, pp. 171–185. DOI: `10.3390/microarrays2030171`.

Xu, L., Y. Hou, D. M. Bickhart, Y. Zhou, E. H. A. Hay, J. Song, et al. (2016). "Population-genetic properties of differentiated copy number variations in cattle". In: *Scientific Reports* 6, p. 23161. DOI: `10.1038/srep23161`.

Yang, B., A. Bassols, Y. Saco, and M. Pérez-Enciso (2011). "Association between plasma metabolites and gene expression profiles in five porcine endocrine tissues". In: *Genetics Selection Evolution* 43, p. 28. DOI: `10.1186/1297-9686-43-28`.

Zefreh, M. Ghaderi, R. Pong-Wong, and A. Doeschl-Wilson (2024). "Validating statistical properties of resilience indicators derived from simulated longitudinal performance measures of farmed animals". English. In: *ANIMAL* 18.8. ISSN: 1751-7311. DOI: `10.1016/j.animal.2024.101248`.

Zeileis, Achim and Gabor Grothendieck (2005). "zoo: S3 Infrastructure for Regular and Irregular Time Series". In: *Journal of Statistical Software* 14.6, pp. 1–27. DOI: `10.18637/jss.v014.i06`.

Zhou, J., L. Liu, T. J. Lopdell, D. J. Garrick, and Y. Shi (2021). "HandyCNV: Standardized Summary, Annotation, Comparison, and Visualization of Copy Number Variant, Copy Number Variation Region, and Runs of Homozygosity". In: *Frontiers in Genetics* 12, p. 731355. DOI: `10.3389/fgene.2021.731355`.