REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

INFLUÊNCIA DA PARIDADE E DO ESTRESSE TÉRMICO (THI) NA PRODUÇÃO MÉDIA DE LEITE DE VACAS BRAUNVIEH CRIADAS EM PASTAGENS EM AMBIENTES MONTANHOSOS: UMA ANÁLISE EM NÍVEL POPULACIONAL UTILIZANDO MODELOS DE APRENDIZADO DE MÁQUINA

INFLUENCE OF PARITY AND THERMAL STRESS (THI) ON THE AVERAGE MILK YIELD OF PASTURE-BASED BRAUNVIEH COWS RAISED IN MOUNTAIN ENVIRONMENTS: A POPULATION-LEVEL ANALYSIS USING MACHINE LEARNING MODELS

INFLUENCIA DE LA PARIDAD Y EL ESTRÉS TÉRMICO (EIT) EN LA PRODUCCIÓN MEDIA DE LECHE DE VACAS BRAUNVIEH CRIADAS EN PASTO EN AMBIENTES DE MONTAÑA: UN ANÁLISIS A NIVEL POBLACIONAL MEDIANTE MODELOS DE APRENDIZAJE AUTOMÁTICO

Jean Carlo Silva de Macedo
Student in Systems Analysis, Federal Institute of São Paulo (Campinas), Brazil
E-mail: m.jean@aluno.ifsp.edu.br

Leandro Ataide Barbosa de Oliveira
Doctor in Applied Linguistics, Federal Institute of São Paulo (Campinas), Brazil
E-mail: leandro.ataide@ifsp.edu.br

Diego Fernandes Goncalves Martins
Master's degree in Electrical Engineering, Federal Institute of São Paulo (Campinas), Brazil
E-mail: diego.martins@ifsp.edu.br

Jackson Gomes Soares Souza
Doctor in School Education, Federal Institute of São Paulo (Campinas), Brazil
E-mail: jackson@ifsp.edu.br

Danilo Douradinho Fernandes
Doctor of Business Administration, Federal Institute of São Paulo (Campinas), Brazil
E-mail: danilo.douradinho@ifsp.edu.br

Glauber da Rocha Balthazar
Doctor of Agricultural Systems Engineering, Federal Institute of São Paulo (Campinas), Brazil
E-mail: glauber.balthazar@ifsp.edu.br

Resumo

A produção leiteira em sistemas de pastagem é fortemente influenciada pelo estresse ambiental, particularmente o estresse térmico medido pelo Índice de Temperatura e Umidade (ITU). Este estudo

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

investigou como o ITU interage com a paridade (número de partos) para afetar a produção média diária de leite em vacas da raça Braunvieh, utilizando um conjunto de dados aberto com 17.396 registros independentes. A análise estatística inferencial revelou uma associação negativa significativa entre o ITU e a produção de leite em todas as classes de paridade ($p < 0,001$). Para cada aumento de uma unidade no ITU, a produção diminuiu em 0,18 kg/dia em vacas primíparas, 0,26 kg/dia em vacas de segunda parição e 0,23 kg/dia em vacas multíparas. Isso se traduz em reduções de 0,9 a 1,3 kg/dia para um aumento de cinco unidades no ITU. Os modelos apresentaram poder explicativo moderado (R^2 entre 0,076 e 0,119), o que é típico para dados observacionais em nível populacional. A modelagem preditiva utilizando regressão polinomial e Random Forest identificou padrões não lineares: o conforto térmico atinge um platô em valores de THI mais baixos, com perdas de produção acelerando sob maior estresse térmico. A paridade foi a variável mais fortemente associada às diferenças de produção basal, enquanto o THI esteve associado às variações na produção de leite sob diferentes condições ambientais. O estudo conclui que a produção média de leite em vacas Braunvieh é melhor descrita por uma abordagem não linear em nível populacional. Os resultados sugerem que os padrões de produção de leite em nível populacional estão associados tanto à estrutura de paridade quanto às condições de THI. Essas descobertas apoiam o uso de modelos de aprendizado de máquina juntamente com inferência estatística clássica para melhor compreender as complexas interações ambientais e demográficas em sistemas de produção leiteira.
Palavras-chave: produção de leite em nível populacional; estresse térmico (THI); efeitos do número de partos; modelos de aprendizado de máquina; sistemas leiteiros baseados em pasto.

Abstract

Milk production in pasture-based systems is heavily influenced by environmental stress, particularly thermal stress measured by the Temperature–Humidity Index (THI). This study investigated how THI interacts with parity to affect average daily milk yield in Braunvieh cows, using an open dataset of 17,396 independent records. Inferential statistics revealed a significant negative association between THI and milk yield across all parity classes ($p < 0.001$). For each one-unit increase in THI, production decreased by 0.18 kg/day in primiparous cows, 0.26 kg/day in second-parity cows, and 0.23 kg/day in multiparous cows. This translates to reductions of 0.9 to 1.3 kg/day for a five-unit rise in THI. The models showed moderate explanatory power (R^2 between 0.076 and 0.119), which is typical for observational population-level data. Predictive modeling using polynomial regression and Random Forest identified non-linear patterns: thermal comfort plateaus at lower THI values, with production losses accelerating under higher thermal stress. Parity was the variable most strongly associated with baseline production differences, while THI was associated with variations in milk yield under different environmental conditions. The study concludes that average milk yield in Braunvieh cows is best described by a non-linear population-level approach. The results suggest that population-level

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

milk yield patterns are associated with both parity structure and THI conditions. These findings support using machine learning models alongside classical statistical inference to better understand complex environmental-demographic interactions in dairy systems.

Keywords: Population-level milk production; Thermal stress (THI); Parity effects; Machine learning models; Pasture-based dairy systems.

Resumen

La producción de leche en sistemas basados en pastoreo está fuertemente influenciada por el estrés ambiental, particularmente el estrés térmico medido por el Índice de Temperatura-Humedad (ITH). Este estudio investigó cómo el ITH interactúa con la paridad (número de partos) para afectar el rendimiento diario promedio de leche en vacas Braunvieh, utilizando un conjunto de datos abierto de 17,396 registros independientes. Las estadísticas inferenciales revelaron una asociación negativa significativa entre el ITH y el rendimiento de leche en todas las clases de paridad ($p < 0.001$). Por cada aumento de una unidad en el ITH, la producción disminuyó en 0.18 kg/día en vacas primíparas, 0.26 kg/día en vacas de segunda paridad y 0.23 kg/día en vacas multíparas. Esto se traduce en reducciones de 0.9 a 1.3 kg/día por un aumento de cinco unidades en el ITH. Los modelos mostraron un poder explicativo moderado (R^2 entre 0.076 y 0.119), lo cual es típico para datos observacionales a nivel poblacional. El modelado predictivo mediante regresión polinómica y Random Forest identificó patrones no lineales: el confort térmico se estabiliza a valores bajos de THI, mientras que las pérdidas de producción se aceleran bajo un mayor estrés térmico. La paridad fue la variable más fuertemente asociada con las diferencias de producción iniciales, mientras que el THI se asoció con variaciones en la producción de leche bajo diferentes condiciones ambientales. El estudio concluye que la producción media de leche en vacas Braunvieh se describe mejor mediante un enfoque no lineal a nivel poblacional. Los resultados sugieren que los patrones de producción de leche a nivel poblacional están asociados tanto con la estructura de paridad como con las condiciones de THI. Estos hallazgos respaldan el uso de modelos de aprendizaje automático junto con la inferencia estadística clásica para comprender mejor las complejas interacciones ambientales-demográficas en los sistemas lecheros.

Palabras clave: producción de leche a nivel poblacional; estrés térmico (THI); efectos del número de partos; modelos de aprendizaje automático; sistemas lecheros basados en pastoreo.

1. Introduction

Milk production from dairy cows directly influences the economies of many

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

countries. In low-income countries, milk production improves human nutrition, health, and economic opportunities for subsistence farmers (Tricarico et al., 2020); in high-income countries, such as those of the European Union, production increased from 151 million tons in 1998 to 165 million tons in 2017, with milk being a major contributor to gross domestic product, since household final consumption of milk and dairy products is one of the main factors driving this increase (Bórawski et al., 2020). This influence arises from the fact that milk is a primary product for the manufacture of several derivatives, such as butter, cheese, and yogurt. As a result, the dairy industry contributes to the economies of many regions and countries, directly influencing their gross domestic product (Buleca et al., 2018; Domingos et al. 2026).

To achieve expressive milk production results, it is necessary to meet nutritional, productive, zootechnical, physiological, health, environmental, and management requirements, as well as to apply appropriate technical management and genetic improvement strategies that favor the production of high-quality milk with elevated fat and protein contents (Hu et al., 2023; Kamaldinov et al., 2022; Terry et al., 2020; Bach et al., 2019). Healthy animals produce more milk of better quality and, consequently, exhibit a well-defined estrous cycle, being limited only by natural factors such as seasonal variation and genetics (Silva, 2023).

Thus, understanding the reproductive efficiency of dairy cows can significantly improve economic and environmental performance, with milk price having the greatest impact on overall net productive returns at the farm level (Krpalkova et al., 2020; Castro Júnior et al., 2019). Several authors (Luna, 2023; Micic, 2022; Knegsel et al., 2022; Krpalkova et al., 2020; Nemeckova et al., 2015) report that dairy production efficiency is related to pregnancy rate and farm environmental conditions. In this sense, milk production is related to: (a) animal fertility (LeBlanc, 2010), as longer calving intervals indicate a reduction in average milk yield per cow, since animals that spend more time at the end of lactation generate lower financial returns due to decreased productivity and feed costs (Nemeckova et al., 2015); and (b) environmental conditions (Micic, 2022), as daily

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

milk yield in dairy cows is significantly influenced by thermal stress conditions according to breed, region, year of calving, season of calving, parity, and stage of lactation.

Thermal stress is a combination of internal and external stimuli that affect the cow by disrupting body temperature regulation, inducing physiological responses such as increased energy and heat expenditure to maintain thermoregulation and, consequently, reducing milk production efficiency (Asmarasari et al., 2023; Utami & Widiarso, 2022). Animal thermal stress can be evaluated through several equations that determine comfort levels, with the Temperature–Humidity Index (THI) being one of the most widely used indices in dairy production, as it combines the effects of air temperature and humidity associated with thermal stress levels (Bohmanova, 2007). In addition, population-level and cross-sectional approaches are widely employed in animal science and animal production to identify average patterns, structural associations, and environmental responses in production systems, being methodologically appropriate when the objective is not to infer individual trajectories over time, but rather to understand variability and average behavior in heterogeneous populations (Dohoo et al., 2009; Field, 2018; James et al., 2021; Kraetzer, Balthazar, 2021).

This study presents the hypothesis that it is possible to identify and predict population-level patterns of milk production in Braunvieh cows based on demographic and environmental variables (parity and thermal index [THI]), by analyzing the joint influence of these variables on average daily milk yield, without attempting to reconstruct individual longitudinal trajectories, focusing instead on population-level associations among productive and environmental variables.

The general objective of this study is to analyze population-level patterns of milk production in Braunvieh cows raised in mountain pasture systems by relating average daily milk yield to demographic indicator variables and environmental conditions, with the purpose of identifying associations and modeling population trends under different environmental scenarios. The specific objectives are to: a) obtain and preprocess the Braunvieh cow dataset published by Duruz et al. (2020),

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

ensuring data integrity, standardization, and consistency of the variables of interest (average milk yield and THI); b) perform descriptive and exploratory analyses of continuous and categorical variables, identifying patterns of distribution and population-level variability; c) examine statistical associations between the number of parities (lact_num) and continuous variables (avg_milk, avg_thi3), considering lact_num as a demographic factor representative of the herd population structure; d) apply and compare supervised machine learning models to predict average milk yield based on environmental and demographic variables, evaluating predictive performance using R^2, RMSE, and MAE; and e) interpret predictive results from a population-level and ecological perspective, discussing the role of environmental and structural variables (THI and parity) in the average productive variability of the herd.

2. Material and Methods

Dataset

This study used the dataset derived from the research of Duruz et al. (2020), which compiled productive and environmental information from Braunvieh cows raised in mountain pasture systems. The dataset contains independent records from different individuals and does not necessarily correspond to the same cows over time. Therefore, it is characterized as a generalist population sample, suitable for cross-sectional and associative analyses. The dataset is publicly available in the Dryad[1], GitHub[2], and Zenodo[3] repositories under the General Public License (GNU), as published by the Free Software Foundation. The following variables were considered:

- avg_milk — average milk yield (kg/day);

[1] Available at: https://datadryad.org/stash/dataset/doi:10.5061/dryad.z612jm68g#usage

[2] Available at: https://github.com/SolangeD/lactModel/tree/master

[3] Available at: https://zenodo.org/records/3889931

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

- avg_thi3 — three-day average thermal index (dimensionless, indicator of thermal comfort); and
- lact_num — number of lactations (parities), used as a demographic indicator rather than as an individual temporal sequence (throughout this manuscript, the term "parity" is used to refer to the biological concept represented by the variable lact_num).

The three-day average THI (avg_thi3) was selected because milk yield responses to heat stress are not necessarily immediate and may reflect short-term accumulated thermal exposure. West (2019) and West et al. (2003) have shown that lagged thermal conditions, including THI values from previous days, can influence milk yield more strongly than same-day weather conditions. Therefore, the use of a three-day average THI provides a biologically plausible indicator of recent thermal load, smoothing short-term daily fluctuations while preserving the short-term environmental signal relevant to milk production.

Data processing was performed in the Python programming language using the Pandas and NumPy libraries. The analyzed dataset comprised a total of 17,396 records. Data treatment followed the recommendations of the study conducted by Zhu et al. (2024) as a multi-stage process: during the preprocessing stage, dataset integrity was assessed, and no missing values or inconsistencies were identified in the variable of interest (milk production); in the outlier verification stage, the Interquartile Range (IQR) method was applied, establishing an upper cutoff limit of 25.95 kg/day; finally, the last stage consisted of exploratory analysis and graphical visualization (histograms, boxplots, and scatter plots), conducted using the Matplotlib and Seaborn libraries. Lastly, outliers in average milk production were identified using the Interquartile Range (IQR) criterion, adopting a globally defined upper cutoff limit for the entire sample, and for the machine learning stage, continuous variables were retained on their original scale, without prior standardization or normalization, as these procedures are not assumptions of the algorithms employed.

From the original dataset containing 20,000 records, 2,604 observations

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

(13.02%) were removed during the data-cleaning process, resulting in a final analytical dataset of 17,396 records. Among these removals, 180 observations (0.90% of the original dataset) were identified as high-end outliers according to the Interquartile Range (IQR) criterion applied to average milk yield (avg_milk), using a cutoff value of 26.645 kg/day (Q3 + 1.5 × IQR). The remaining exclusions were associated with additional preprocessing and data-quality procedures described in the dataset preparation workflow. Although a formal sensitivity analysis comparing models with and without outlier exclusion was not performed, the IQR procedure identified only 180 observations (0.90% of the original dataset) as outliers. Therefore, the exclusion affected a very small fraction of the available observations and is unlikely to substantially alter the overall population-level patterns reported in this study. The adopted procedure was intended to reduce the influence of extreme observations while preserving the underlying structure of the dataset.

As a consequence of the structure of the publicly available analytical dataset, the present study focuses on population-level associations and predictive patterns rather than on herd-level, farm-level, or individual-animal inference.

Observational unit and analytical assumptions

The original study by Duruz et al. was based on more than 5 million milk-record observations collected from Braunvieh cows between 2000 and 2015. However, the public dataset made available through Dryad corresponds to a reduced and preprocessed subset of the original database and does not provide access to the complete longitudinal structure of the records. In the present study, the observational unit corresponds to a productive-environmental record containing average milk yield (avg_milk), thermal index (avg_thi3), and parity number (lact_num). The analytical dataset consisted of 17,396 records obtained after preprocessing and filtering procedures applied to the public data source.

Because the public dataset does not allow complete reconstruction of individual longitudinal trajectories across all animals and parities, the analyses were conducted from a population-level perspective. Therefore, the objective was not to

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

infer individual responses over time but rather to identify average associative patterns between environmental conditions, parity structure, and milk production. Consequently, record-level independence should be interpreted as an analytical assumption associated with the structure of the available dataset rather than as proof of complete biological independence among all original observations.

The analytical dataset used in this study contained only four variables:parity (lact_num), average milk yield (avg_milk) and three-day average thermal index (avg_thi3). No identifiers related to individual animals, farms, herds, geographic locations, altitude, production units, years, or seasons were available in the final analytical dataset. Consequently, grouped-validation procedures, hierarchical models, and mixed-effects approaches based on these factors could not be implemented.

Exploratory data analysis

This approach involved examining and visualizing the dataset to understand its characteristics, patterns, and relationships prior to the application of statistical models or machine learning algorithms. The variable lact_num was used as a demographic factor rather than as a temporal variable, allowing the assessment of population-level associations between parity structure and average environmental conditions. The analysis was conducted through: (a) descriptive analyses aimed at describing and summarizing the main characteristics of the data using measures such as mean, median, mode, standard deviation, minimum, and maximum values; (b) graphical visualizations to understand the distribution of continuous variables; and (c) calculation of correlations between demographic structure and environmental conditions in order to examine how the variables are related using appropriate statistical coefficients.

Inferential analysis

With the objective of investigating whether average milk yield exhibits statistically significant differences among parity classes, an inferential analysis was

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

conducted while controlling for the effect of environmental variables, in order to isolate the demographic component of productive variation. Thus, to quantify the population-level thermal sensitivity of milk production, the following question was formulated: “Within each parity class (primiparous, second-parity, and multiparous), how much does average daily milk yield decrease as the environment becomes hotter and more humid (i.e., as THI increases)?” This approach aimed to identify population-level responses to thermal stress according to lactation order, using THI as an indicator of thermal discomfort in cattle (Bouraoui et al., 2023; Polsky & von Keyserlingk, 2017). Considering that the dataset is observational and cross-sectional (composed of independent records from different individuals), a stratified associative analysis was adopted, in which the dataset was subdivided into three parity groups (lact_num = 1; lact_num = 2; lact_num ≥ 3), allowing the thermal impact to be estimated separately within each group. For each stratum, a simple linear regression model was fitted, as shown in Equation 1.

$$avg_{milk} = \alpha k + \beta(THI, k)\, avg_thi3 + \varepsilon$$

Equation 1 - Linear regression model

In which:

- avg_{milk}: dependent variable (kg day⁻¹): observed average daily milk yield for each record in the sample;
- αk: regression intercept (kg day⁻¹): theoretical interpretation of the expected production value when the THI index equals zero, serving only as a reference for fitting the regression line;
- $\beta_{(THI,k)}$: regression coefficient $(kg\ day^{-1} \cdot THI^{-1})$: measures the average change in milk yield for each one-unit increase in the THI index within parity class k.

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

Negative values indicate an association between higher THI values and lower average milk yield;

- avg_thi3: three-day average Temperature–Humidity Index (dimensionless): composite environmental indicator integrating air temperature and relative humidity; and
- ε: random error term: represents the variation not explained by the model, corresponding to residual factors.

Inference was based on HC3 robust standard errors, which are suitable for heteroskedastic data and widely recommended in observational animal science studies (White, 1980; Marumo et al., 2022). A 95% confidence interval was adopted. The following hypotheses were tested:

H0: $\beta(THI, k) = 0$: there is no linear relationship between the thermal index (THI) and average daily milk yield; that is, the thermal environment does not significantly affect productive performance;

H1: $\beta(THI, k) < 0$: given a negative regression coefficient, this indicates a systematic reduction in milk yield as THI increases; that is, higher THI values were associated with lower average milk yield within each parity group.

Finally, the practical magnitude of the effect was expressed as the expected change in milk production for an increase of +5 THI units , corresponding to the typical range between thermal comfort and heat stress conditions reported in the literature (West, 2019; Carvalho et al., 2020). This formulation allows a comparative assessment of thermal sensitivity across parity classes, providing a direct measure of the population-level impact of the thermal environment on milk production.

The Temperature–Humidity Index (THI) was adopted because it is a widely used and well-established indicator of heat stress in dairy cattle, combining air temperature and relative humidity into a single measure of thermal discomfort (Bohmanova et al, 2007). Its use allows comparison with previous studies on heat stress and milk production and is consistent with the environmental variables made

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

available in the public dataset. Although other bioclimatic indicators may capture additional dimensions of thermal load, THI was selected due to its interpretability, availability, and broad use in dairy science.

Machine learning

The relationship between the linear inferential model (Equation 1) and supervised learning arises from the fact that both aim to estimate a mapping function between explanatory variables and the response variable based on labeled data. The parity-centered predictor developed in the inferential stage (Equation 1) constitutes a specific form of supervised learning in the form of linear regression, in which parameters are adjusted to minimize the mean squared error between observed and estimated values (Hastie, Tibshirani, & Friedman, 2009). This model therefore serves as an interpretable statistical baseline, in which the coefficient $\beta(THI, k)$ quantifies the average thermal impact on milk production within each parity class.

From this point onward, supervised learning techniques for continuous regression were employed to enhance predictive capacity by exploring potential nonlinearities and interactions between the thermal index (THI) and demographic variables. Models such as regularized linear regression (Ridge and Lasso), regression trees, and Random Forests expanded the structure of the linear predictor by incorporating regularization, hierarchical partitioning of the feature space, and combinations of multiple estimators, respectively (Breiman, 2001; James et al., 2021). These approaches made it possible to quantify the relative importance of environmental variables and to evaluate productive responses under diverse thermal conditions, while maintaining coherence with the inferential statistical framework and the zootechnical assumptions established in the literature (Bouraoui et al., 2023; Marumo et al., 2022; West, 2019).

In other words, supervised learning models allowed investigation of whether the relationship between the thermal index (THI) and milk production followed a constant linear pattern—as observed in the inferential regression—or whether it exhibited curvature and thresholds indicative of nonlinear responses, as well as

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

potential interactions between environment and parity. This analysis combined the statistical rigor of linear modeling with the flexibility of learning algorithms, enabling the identification of complex thermal patterns and demographic heterogeneity in the productive performance of the studied population.

Results evaluation

The analytical dataset was randomly divided into three mutually exclusive subsets: training set (70%), validation set (15%), and test set (15%). The training set was used to fit the models, the validation set was used for model comparison and hyperparameter adjustment, and the test set was reserved exclusively for final performance evaluation. The split was performed using a fixed random seed to ensure reproducibility. No simulated observations were included in the training, validation, or test sets. Model performance was evaluated using regression metrics appropriate for continuous outcomes, namely the coefficient of determination (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). R^2 was used to quantify the proportion of variance explained by each model, whereas RMSE and MAE were used to measure prediction errors in the original milk-yield scale (kg/day). All metrics were calculated exclusively from observed records in the validation set.

All subsets were composed exclusively of observed records derived from the public dataset. No simulated observations were used during model training, validation, or testing.

Computational environment and model configuration

The machine learning analyses were implemented in Python (3.12.3) using Pandas, NumPy, scikit-learn, Matplotlib and Seaborn. The dataset was split into training, validation and test subsets using train_test_split with a fixed random seed (random_state = 42) to ensure reproducibility. The predictive target was average milk yield (avg_milk), and the explanatory variables were the three-day average THI (avg_thi3) and parity class derived from lact_num. Parity class was encoded using one-hot encoding with handle_unknown = "ignore" and drop = "first".

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

The evaluated models included: linear regression as the baseline model; Ridge regression with default regularization (alpha = 1.0); Random Forest regression with n_estimators = 100, max_depth = 8, random_state = 42, and n_jobs = -1; and polynomial regression with a second-degree polynomial term for THI (degree = 2, include_bias = False) combined with parity class. An additional Random Forest model using only THI was evaluated with n_estimators = 200, max_depth = 15, min_samples_leaf = 5, random_state = 42, and n_jobs = -1.

3. Results and Discussions

The dataset used in this research consisted of the informative sample of dairy production provided by Duruz et al. (2020), from which the following variables were utilized: number of parities, average milk yield (kg/day), and the three-day average thermal index (THI, dimensionless). Results from the exploratory analysis and data preprocessing stages indicated that application of the Interquartile Range (IQR) method led to the exclusion of only a small fraction of the total records, thereby preserving the population structure of the dataset and preventing the influence of extreme values on subsequent modeling. The upper cutoff limit for average milk production was defined globally for the entire sample, rather than being stratified by parity class, in order to maintain a uniform data treatment criterion. Furthermore, for the machine learning models, continuous variables were used on their original scale, without prior standardization or normalization, as the employed algorithms do not require these procedures as underlying assumptions.

Descriptive statistical analysis

Considering the continuous variables of the study, the following descriptive statistics were obtained: average milk yield (avg_milk): mean = 15.51, sample mode = 14.60, standard deviation = 3.74 (n = 17,396); and three-day average thermal index (avg_thi3): mean = 60.10, mode = 59.58 (multimodal distribution), standard deviation = 4.96 (n = 17,396).

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

Milk production exhibited well-defined central tendencies with moderate dispersion (relative SD of approximately 24%), suggesting variability compatible with heterogeneous production systems. That is, the observed variability in the dataset (CV ≈ 24% for milk production; ≈ 8% for THI) is consistent with heterogeneous production systems, in which differences among herds, years, seasons, management practices, and cow biology accumulate and increase dispersion—especially in mountain pasture-based systems such as those of the Braunvieh breed, where environmental conditions (THI) and altitude exert a measurable effect on milk production (Duruz et al., 2020; Micic, 2022; Bohmanova, Misztal, & Cole, 2007). In addition, several studies demonstrate that thermal stress reduces milk yield and that sensitivity to heat and humidity varies according to climate and the index adopted, reinforcing environmental sources of intra- and inter-herd variation (Bohmanova, Misztal, & Cole, 2007; Micic, 2022). Furthermore, parity and stage of lactation modulate daily yield, with primiparous cows producing less and intermediate lactations tending to peak, contributing to the observed heterogeneity (Kertz et al., 1997; Dürr et al., 2008; Capuco et al., 2003; LeBlanc, 2010; Nemeckova, Stadnik, & Citek, 2015).

The multimodal distribution observed for THI is consistent with the presence of seasonal and regional structures in the thermo-hygrometric regime. In temperate and mountainous climates, THI tends to be organized into seasonal ranges (e.g., lower values in winter and peaks in summer), creating natural groupings throughout the year. Additionally, regional differences (altitude and continentality) modulate the relative contributions of temperature and humidity, generating distinct profiles of thermal stress risk among regions and, consequently, multiple modes in the annual distribution (Bohmanova, Misztal, & Cole, 2007; Vitali et al., 2009; Nam et al., 2024). From a mechanistic perspective, Bohmanova (2007) states that thresholds and slopes of production loss vary across indices and regions (indices with greater weight on humidity perform better in humid climates, whereas those more sensitive to temperature respond better in semi-arid climates), reinforcing the concept of climatic clusters with different responses. There is also evidence of behavioral and productive

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

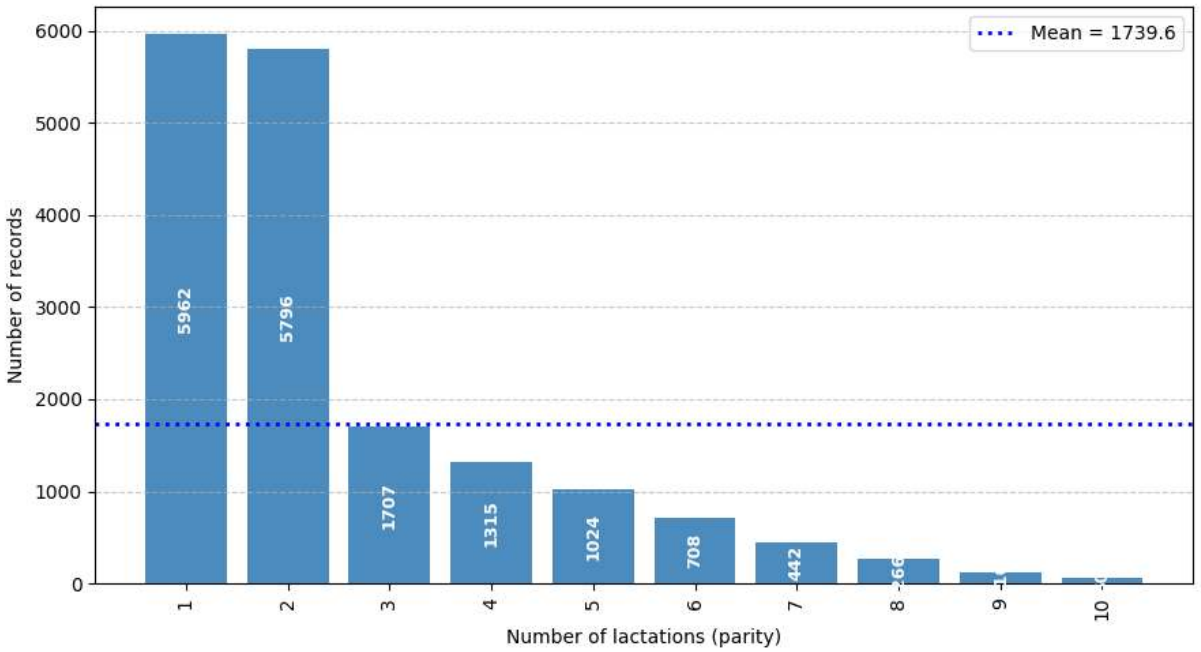
seasonality associated with THI (changes in standing/lying time, rumination, intake, and production decline), with impairments concentrated in the warmest months, which is reflected in seasonal peaks of risk and contributes to the multimodal distribution (Vitali et al., 2009; Toledo et al., 2024; Nam et al., 2024). Finally, Ravagnolo (2000) and Bohmanova (2007) demonstrate that heat responses follow functions with breakpoints (thresholds) and subsequent declines in production, with thresholds differing by region and index, consistent with the formation of distinct climatic clusters in THI.

Regarding parity, there is a strong predominance of cows in the first two lactations (Figure 1), with a progressive reduction in advanced lactations, corroborating the expected pattern in dairy herds, in which high replacement rates and technical culling occur as age advances. This trend reflects the typical demographic distribution of herds, characterized by a predominance of primiparous and second-parity cows, resulting from the balance between annual replacement and the culling of older cows (Kruszyński et al., 2021; Tassinari et al., 2022). In other words, there are many more cows in the first and second lactations and fewer cows in advanced lactations due to the continuous replacement process by newly calved heifers, combined with the natural reduction in fertility, increased incidence of diseases (such as mastitis), and decline in lactation persistency in older cows (Narváez-Solarte et al., 2020; Knegsel et al., 2022).

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

Figure 1. Distribution of cow parities in the dataset.



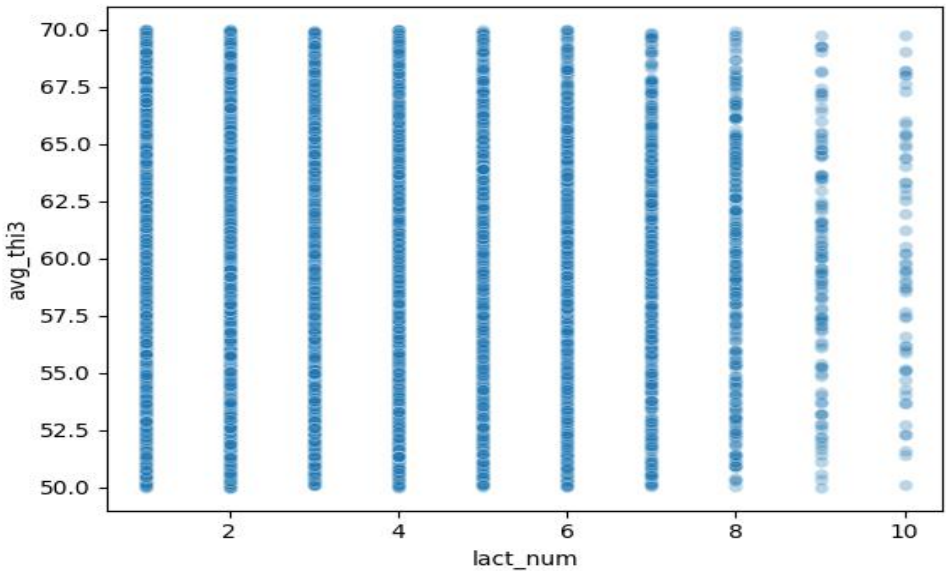
Source: Duruz et al., 2020

Spearman's correlation coefficient was adopted to evaluate associations between the demographic variable (lact_num) and the environmental variable (avg_thi3), as it is suitable for the statistical characteristics of the dataset. This approach was intended to interpret the relationship between the variables as a demographic–environmental association rather than as a linear physiological causal relationship. To this end, the nature of the variables was examined to meet the assumptions of Spearman's coefficient: lact_num is a discrete ordinal variable (1 to 10 lactations), whereas avg_thi3 (three-day average thermal index) is continuous, satisfying the requirement of compatible scales for rank-based analysis; Shapiro–Wilk normality tests indicated a non-normal distribution for the environmental variable ($p < 0.001$ for avg_thi3), and scatter plots (Figure 2) revealed non-linear patterns, with no evident linear relationship between lact_num and the environmental factor (conditions under which Pearson's coefficient would be inappropriate).

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

Figure 2. Assessment of linearity and homoscedasticity through data dispersion: demographic (lact_num) versus environmental (avg_thi3).



Based on these considerations, Spearman's coefficient was deemed the most appropriate, as it evaluates monotonic relationships without assuming bivariate normality or linearity. The analysis yielded p = 0.081 for the correlation between lact_num and avg_thi3, a value close to zero and not statistically significant at the 5% level. This result indicates the absence of a significant monotonic trend between demographic structure and average environmental conditions, demonstrating that cows observed across different lactation classes were measured under similar climatic conditions, with no sampling bias related to parity. Similar results were reported by Duruz et al. (2020) and Micic (2022), who also observed no differential microclimatic effects among parity classes in mountain dairy herds, confirming that environmental variations represent general productive conditions rather than demographic factors.

Inferential statistical analysis

Considering that the dataset adopted in this study was population-based, cross-sectional, and generalist (with no correspondence between the same cows), the inferential analysis was conducted with the aim of avoiding longitudinal

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

conclusions and focusing on associations between demographic groups (parity) and environmental/productive variables.

To statistically evaluate the relationship between THI and average daily milk yield within each parity class, a script was developed in the Python programming language using the Statsmodels library, which fitted simple linear regression models according to Equation 1, stratifying the dataset into primiparous (lact_num = 1), second-parity (lact_num = 2), and multiparous cows (lact_num ≥ 3). Each model was estimated using ordinary least squares (OLS) with HC3 robust standard errors, in order to ensure inferential consistency under potential heteroscedasticity, a frequent condition in zootechnical studies with productive data derived from different individuals and environments (White, 1980; Marumo et al., 2022; Field, 2018). For each stratum, the analysis returned the regression coefficient ($\beta_{\mathrm{THI},k}$), 95% confidence intervals, t-statistics, two-sided and one-sided (left-tailed) p-values, and the coefficient of determination (). Thus, the formulated hypotheses — $H_0: \beta_{\mathrm{THI},k} = 0 \text{and } H_1: \beta_{\mathrm{THI},k} < 0$ — were tested for each lactation class, allowing identification of whether increases in THI were associated with statistically significant reductions in average milk yield within each population group. The results are presented in Table 1.

The $\beta_{\mathrm{THI},k}$ coefficients were negative and highly significant ($p < 0.001$) across all parity classes, indicating that increases in the average thermal index lead to reductions in average milk yield. In primiparous cows, the estimated effect was −0.18 kg/day per unit increase in THI (95% CI [−0.196; −0.165]), corresponding to a decline of approximately 0.9 kg/day for a 5-unit increase in THI. In second-parity cows, the reduction was more pronounced (−0.26 kg/day; 95% CI [−0.278; −0.242]; Δ +5 THI = −1.30 kg/day), while in multiparous cows a similar decline was observed (−0.23 kg/day; 95% CI [−0.254; −0.216]; Δ +5 THI = −1.18 kg/day). All models exhibited moderate explanatory power, with R^2 values ranging from 0.076 to 0.119, which is consistent with population-level observational data in which zootechnical and environmental factors jointly explain productive variability.

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

Table 1. Regression coefficients (β_{THI}) and significance of the effect of THI on milk production by parity.

Parity	n	β_THI (kg/day·THI⁻¹)	IC95%	p (one-sided)	R²	Δ (+5 THI) (kg/day)	Decision
1st (primiparous)	5962	−0.180	[−0.196; −0.165]	<0.001	0.076	−0.90	Reject H₀
2nd (second-parity)	5796	−0.260	[−0.278; −0.242]	<0.001	0.119	−1.30	Reject H₀
3rd+ (multiparous)	5638	−0.235	[−0.254; −0.216]	<0.001	0.091	−1.18	Reject H₀

The results presented in Table 1 confirm the rejection of the null hypothesis $H_0: \beta_{\mathrm{THI},k} = 0$ in favor of the alternative $H_1: \beta_{\mathrm{THI},k} < 0$, demonstrating that the thermal-hydric environment exerts a significant and adverse influence on the milk production of Braunvieh cattle. The observed behavior aligns with related research (Bouraoui et al., 2023; West, 2019; Polsky & von Keyserlingk, 2017), which identifies productivity declines under thermal stress, especially when the THI exceeds comfort ranges between 60 and 72. The higher sensitivity observed in the second-parity and multiparous cows can be explained by the increased energy metabolism associated with higher milk production and lower thermoregulatory efficiency as age progresses (Marumo et al., 2022; Micic, 2022). Furthermore, the obtained population pattern reinforces the cumulative effect of heat and humidity, as described by Bohmanova, Misztal, and Cole (2007), who demonstrated progressive production declines in cows subjected to high thermal load conditions.

Machine learning model performance

The predictive performance of the regression models was evaluated using R², RMSE, and MAE. These metrics were selected because milk yield is a continuous variable and therefore requires regression-oriented evaluation criteria.

Table 2 and Figure 3 summarize the results obtained for the validation dataset. Among the evaluated models, Random Forest achieved the best overall performance, presenting the highest coefficient of determination (R² = 0.156) and the lowest prediction errors (RMSE = 3.39 kg/day; MAE = 2.69 kg/day). Polynomial regression

REMUNOM
ISSN 2178-6925

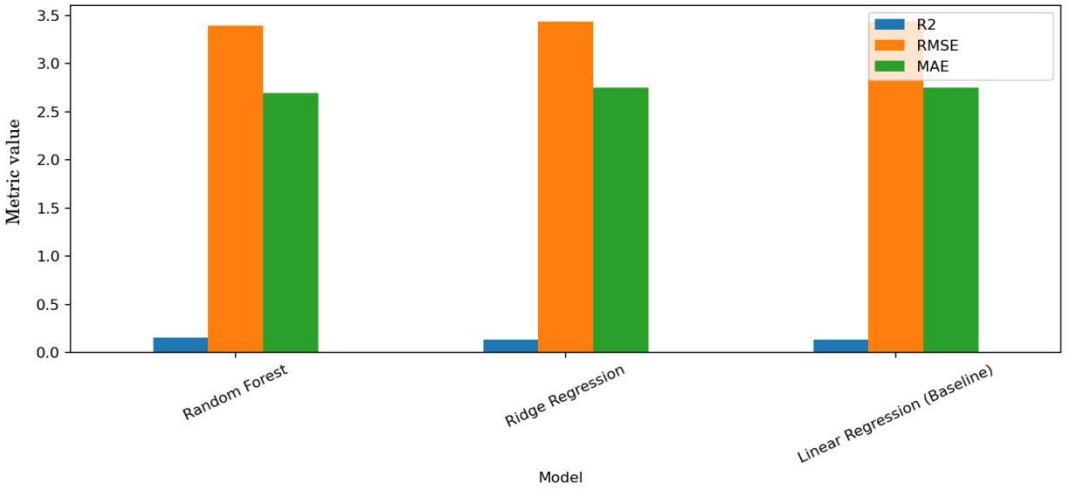
Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

showed slightly lower performance (R^2 = 0.146), while Ridge Regression and the baseline linear model exhibited similar results (R^2 = 0.136).

Table 2. Comparative performance of regression models evaluated using coefficient of determination (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE).

Model	R2	RMSE	MAE
Random Forest	0.15	3.39	2.69
Polinomial_THI2	0.15	3.40	2.71
Ridge Regression	0.13	3.43	2.75
Linear Regression (Baseline)	0.13	3.43	2.75
Random Forest_THI_Puro	0.11	3.52	2.80

Figure 3. Comparison of regression model performance based on R^2, RMSE, and MAE calculated on the validation dataset.



Although the explanatory power of the models was modest, the observed performance is compatible with the characteristics of observational population-level datasets, in which milk production is influenced by numerous biological, environmental, nutritional, and management-related factors not available in the analytical dataset. The superior performance of Random Forest suggests the

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

presence of nonlinear relationships between parity, thermal conditions, and milk production.

Predictive statistical analysis

Inferential linear regression demonstrated that THI negatively affects milk production; however, it imposes a strong assumption: that the effect is constant (the same decline across the entire THI range) and independent of parity. Nevertheless, physiological and empirical studies suggest the existence of thermal comfort and stress thresholds (THI ≈ 68–72) beyond which the decline accelerates (West, 2019; Bouraoui et al., 2023); cows at different parity levels respond differently to these thresholds (Marumo et al., 2022); and environmental and demographic factors (THI and parity) may interact (for example, the same increase of 5 THI units does not result in the same production loss in primiparous and multiparous cows). Therefore, investigating nonlinearity and interaction effects is essential to determine whether the impact of the thermal environment is proportional and continuous (linear) or saturating, threshold-based, or exponential (nonlinear), and whether this effect is consistent across groups (no interaction) or varies among them (with interaction).

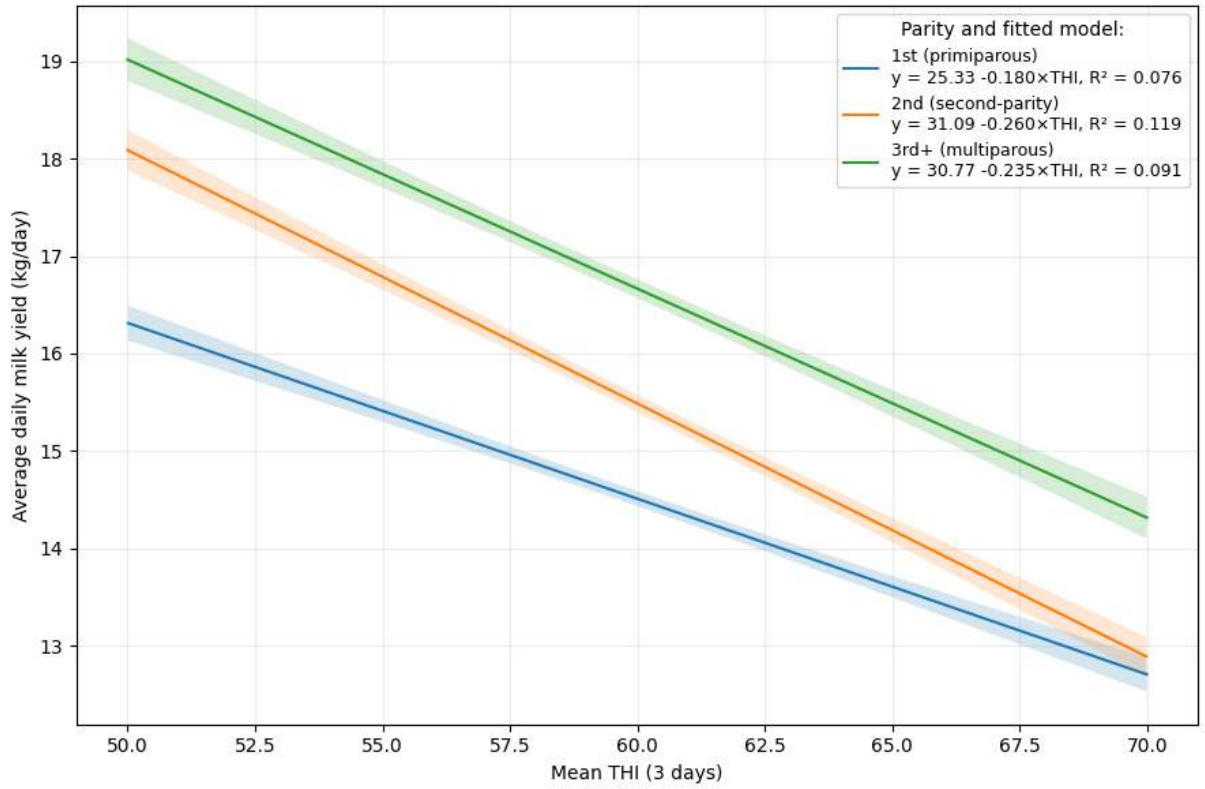
Thus, unlike parametric models—where the functional form of the relationship is imposed—supervised learning algorithms automatically “discover” complex patterns among variables. In this context, the forest plot (Figure 4) shows that the thermal effect (β_{THI}) is negative across all parity classes but differs in magnitude (2nd > 3rd+ > 1st), indicating increasing thermal sensitivity with advancing parity. However, the predicted curves (Figure 5) reflect the linearity assumed by the inferential models and the downward trend in production as THI increases, while also suggesting potential curvature and slope changes at the extremes of the thermal range. Taken together, these results—along with the moderate R^2 values—indicate that although the effect of THI on milk production is statistically consistent, it may not be strictly linear nor uniform across parity groups. This finding underpins the subsequent analytical stage using supervised learning models, which are capable of investigating nonlinearities and complex interactions between the thermal

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

environment and demographic variables without imposing a predefined functional form on the THI × production × parity relationship.

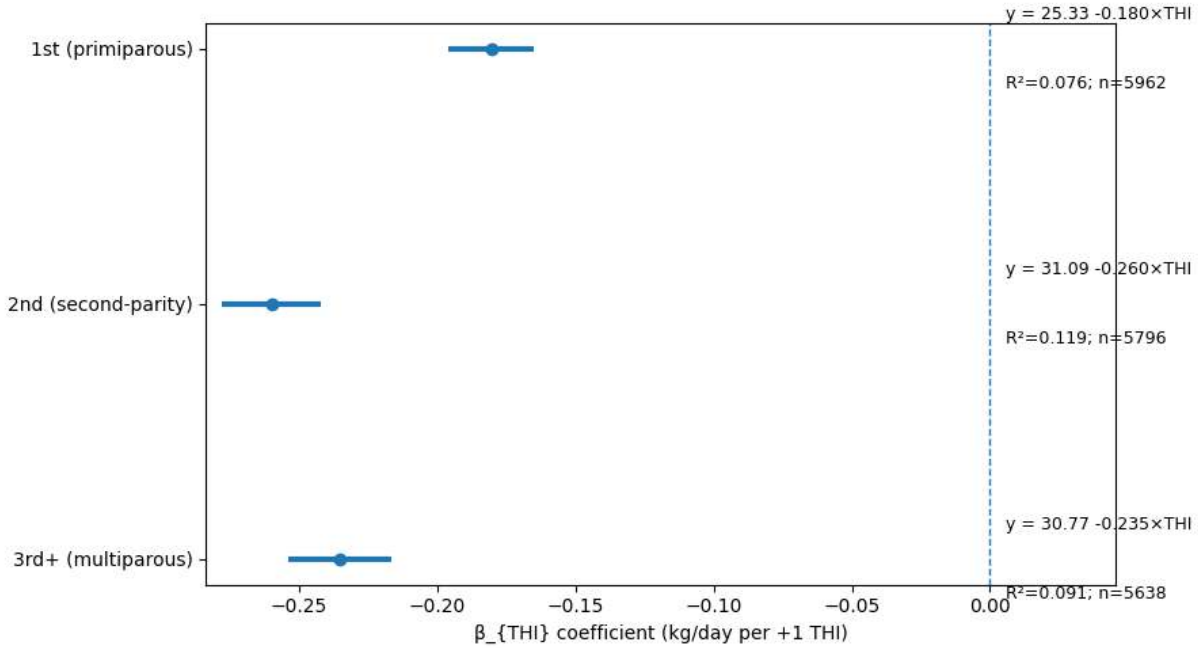
Figure 4. Predicted curves of average daily milk yield ($\hat{y}_k$) as a function of the mean thermal index (THI) by parity class, with 95% Confidence.



REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

Figure 5. Estimates of the β_{THI} coefficients (points) and 95% confidence intervals (horizontal lines) by parity class, obtained using OLS regression with HC3 robust standard errors.



The superiority of the polynomial model does not represent merely a mathematical improvement, but rather faithfully reflects animal biology. The resulting curves visually illustrate thermal homeostasis, in which milk production remains stable within comfort ranges (THI 50–60) and declines sharply only when thermoregulatory mechanisms are overloaded (THI > 65). This provides population-level validation of the physiological stress thresholds described in experimental literature.

The nonlinearity observed in the predictive curves is consistent with studies reporting that milk production losses under thermal stress do not occur in a strictly linear manner, but instead exhibit thresholds and changes in slope associated with increasing THI, varying according to region, the index adopted, and herd characteristics (Bohmanova et al., 2007; Ravagnolo & Misztal, 2000; West, 2019; Bouraoui et al., 2023). Research based on large observational datasets also indicates that productive responses to heat intensify after the transition from thermal comfort ranges, reinforcing that flexible models are more suitable to represent this population-level dynamic (Vitali et al., 2009; Micic, 2022; Nam et al., 2024). From a methodological perspective, machine learning approaches such as Random Forests

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

and polynomial models have been widely employed in animal science precisely because of their ability to capture nonlinearities and interactions between environmental variables and animal characteristics, resulting in superior predictive performance compared with strictly linear models (Breiman, 2001; Hastie et al., 2009; James et al., 2021). Thus, the observed superiority of non-linear models in this study should be interpreted as being consistent with existing biometeorological knowledge, suggesting that the effect of THI on average milk yield is heterogeneous and context-dependent in terms of production system and demographic structure, rather than as evidence of an individual causal physiological mechanism.

Limitations and future research directions

The results of this study should be interpreted considering the limited set of variables available in the analytical dataset. Milk production is a multifactorial trait influenced by physiological, nutritional, environmental, and management-related factors. Variables such as stage of lactation, days in milk, diet composition, feeding strategy, herd management, seasonality, health status, and genetic background were not available in the analytical dataset and therefore could not be incorporated into the models.

Consequently, the associations identified between THI, parity, and milk yield should not be interpreted as exhaustive explanations of productive variability. The relatively modest R^2 values observed in the models suggest that a substantial proportion of milk yield variation is associated with factors not represented in the present dataset. Future studies integrating productive, environmental, nutritional, and management variables may provide a more comprehensive understanding of milk production dynamics under thermal stress conditions. Therefore, the findings should be interpreted as population-level associations derived from observational data and should not be considered direct evidence of causal relationships or individual physiological responses.

Although a formal THI × parity interaction model could provide an additional statistical assessment of effect modification, this was not the primary objective of the present study. The analytical strategy adopted here was based on stratified population-level analyses, allowing thermal sensitivity to be estimated and interpreted

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

separately within each parity group. This approach provides direct biological interpretability and facilitates comparison of thermal-response patterns among demographic strata. Future studies may explore interaction-based models to further investigate potential effect modification between parity and thermal conditions.

4. Conclusions

This study achieved its goal of identifying and modeling the population-level milk production patterns in Braunvieh cattle, integrating both inferential and predictive approaches to understand the joint influence of demographic factors (parity) and environmental factors (Temperature-Humidity Index - THI).

The inferential analysis supported the hypothesis that the thermal environment has a significant adverse impact on milk production, with negative regression coefficients (β_THI) in all parity classes. However, the application of machine learning models allowed for a refinement of this understanding, showing that the productive response to thermal stress is not strictly linear. The Polynomial Regression model showed a superior fit (R^2 = 0.152) compared to the linear model (R^2 = 0.136), highlighting the existence of a "comfort plateau" followed by a negative acceleration in production as THI increases.

The demographic structure, represented by the number of lactations, proved to be the determining factor of the herd's productive potential, as validated by the feature importance analysis of the Random Forest algorithm. The prediction curves revealed that, while multiparous cows have a higher absolute production capacity, the form of the response to thermal stress follows a physiological pattern common to the species, modulated by the intensity of the environmental challenge.

It is concluded that the variability in milk production within the studied population is best described by a curvilinear function, where parity defines the productive level (intercept) and THI acts as a non-linear limiting factor. This finding reinforces the importance of considering non-linearity and demographic heterogeneity in management strategies and in predictive models for animal

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

husbandry.

Data availability

Dataset: https://github.com/JeancMacedo/Data.csv

Source code: https://github.com/JeancMacedo/IC

References

ASMARASARI, S.; AZIZAH, N.; SUTIKNO, S.; PUASTUTI, W.; AMIR, A.; PRAHARANI, L.; RUSDIANA, S.; HIDAYAT, C.; HAFID, A.; KUSUMANINGRUM, D.; SAPUTRA, F.; TALIB, C.; HERLIATIKA, A.; SHIDDIEQY, M. e HAYANTI, S. **A review of dairy cattle heat stress mitigation in Indonesia**. Veterinary World, v. 16, p. 1098-1108, 2023. DOI: https://doi.org/10.14202/vetworld.2023.1098-1108.

BACH, A.; TERRÉ, M. e VIDAL, M. **Symposium review: Decomposing efficiency of milk production and maximizing profit**. Journal of Dairy Science, 2019. DOI: https://doi.org/10.3168/jds.2019-17304.

BOHMANOVA, J.; MISZTAL, I. e COLE, J. B. **Temperature-Humidity Indices as Indicators of Milk Production Losses due to Heat Stress**. Journal of Dairy Science, v. 90, n. 4, p. 1947-1956, 2007. DOI: 10.3168/jds.2006-513.

BÓRAWSKI, P.; PAWLEWICZ, A.; PARZONKO, A.; HARPER, J. e HOLDEN, L. **Factors Shaping Cow's Milk Production in the EU. Sustainability**, v. 12, p. 420, 2020. DOI: https://doi.org/10.3390/su12010420.

BOURAOUI, R.; MARTINS, M.; SILVA, J.; FERREIRA, L.; COSTA, A. e OLIVEIRA, P. **Heat stress effects on dairy cow performance and physiological responses: a global perspective and implications for management**. Animals, v. 13, n. 5, p. 856-870, 2023.

BREIMAN, L. **Random forests. Machine Learning**, v. 45, p. 5-32, 2001.

BRUCE, Peter. **Estatística prática para cientistas de dados: 50 conceitos essenciais**. Rio de Janeiro: Alta Books, 2019. 299 p.

BULECA, J.; KOVÁČ, V. e ŠUBOVÁ, N. **Milk production related to price of raw cow's milk in selected european countries**. Potravinarstvo, 2018. DOI: https://doi.org/10.5219/1002.

CAPUCO, A. V.; SANTOS, J. E. P.; STAFFORD, D. M. e TUCKER, H. A. **Mammary cell turnover in lactating dairy cows**. Domestic Animal Endocrinology, v. 25, n. 2, p. 161-174, 2003.

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

CASTRO JUNIOR, Sérgio Luís de; BALTHAZAR, Glauber da Rocha e SILVA, Iran José Oliveira da. **Diagnóstico preditivo em tempo real do conforto térmico de animais de produção em Sistema Operacional Android**. In: Anais... Lavras: DEA-UFLA, 2019. p. 1091 il. Disponível em: http://www.infobibos.com/anais/siapas2/EBOOK_SIAPAS_SIMCRA.pdf. Acesso em: 15 dez. 2025.

DOHOO, I.; MARTIN, W. e STRYHN, H. **Veterinary Epidemiologic Research.** 2. ed. Charlottetown, Canada: VER Inc., 2009.

DOMINGOS, Murilo Alves; OLIVEIRA, Leandro Ataide Barbosa de; MARTINS, Diego Fernandes Goncalves; SOUZA, Jackson Gomes Soares; FERNANDES, Danilo Douradinho e BALTHAZAR, Glauber da Rocha. **STATISTICAL AND PREDICTIVE ANALYSIS OF BROILER CHICKEN PRODUCTION IN THE WORLD'S MAJOR PRODUCING COUNTRIES: a time-series approach**. Revista de Geopolítica, v. 17, n. 3, p. 1-22, 24 mar. 2026. DOI: http://dx.doi.org/10.56238/revgeov17n3-162.

DÜRR, J. W.; COSTA, C. N.; MARTINS, E. N. e GUIMARÃES, M. P. **Lactation curves in dairy cows**. Brazilian Journal of Animal Science (R. Bras. Zootec.), v. 37, n. 4, p. 692-699, 2008.

DURUZ, Solange; VAJANA, Elia; BURREN, Alexander; FLURY, Christine e JOOST, Stéphane. **Big dairy data to unravel effects of environmental, physiological and morphological factors on milk production of mountain-pastured Braunvieh cows**. Royal Society Open Science, v. 7, n. 7, p. 200638, jul. 2020. DOI: http://dx.doi.org/10.1098/rsos.200638.

ESCOVEDO, Tatiana. **Introdução a data science: algoritmos de machine learning e métodos de análise**. São Paulo: Casa do Código, 2022. 272 p.

FACELI, Katti; LORENA, Ana Carolina; GAMA, João; ALMEIDA, Tiago Agostinho de e CARVALHO, André Ponce de Leon F. de. **Inteligência artificial: uma abordagem de aprendizado de máquina**. 2. ed. Rio de Janeiro: LTC, 2021. 400 p.

FIELD, A. **Discovering statistics using IBM SPSS statistics**. 5. ed. London: Sage, 2018.

HARRISON, Matt. **Machine learning: guia de referência: trabalhando com dados estruturados em Python**. São Paulo: Novatec, 2020. 272 p.

HASTIE, T.; TIBSHIRANI, R. e FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2. ed. New York: Springer,

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

2009.

HU, T.; ZHANG, J.; ZHANG, X.; CHEN, Y.; ZHANG, R. e GUO, K. **The Development of Smart Dairy Farm System and Its Application in Nutritional Grouping and Mastitis Prediction**. Animals: an Open Access Journal from MDPI, v. 13, 2023. DOI: https://doi.org/10.3390/ani13050804.

JAMES, G.; WITTEN, D.; HASTIE, T. e TIBSHIRANI, R. **An Introduction to Statistical Learning: with Applications in Python**. New York: Springer, 2021.

KAMALDINOV, E.; PETROV, A.; SHATOKHIN, K.; NAROZHNYKH, K.; MARENKOV, V.; ZHIGULIN, T.; BOGDANOVA, O.; PALCHIKOV, P. e PLAKHOVA, A. **Reliability of primary zootechnical records in dairy farming.** Bulletin of NSAU (Novosibirsk State Agrarian University), v. 63, n. 2, p. 76-83, 2022. DOI: https://doi.org/10.31677/2072-6724-2022-63-2-76-83.

KERTZ, A. F.; REUTZEL, L. F. e COLE, W. J. **First-lactation milk yield of Holstein cows**. Journal of Dairy Science, v. 80, n. 8, p. 1858-1863, 1997.

KNEGSEL, A.; BURGERS, E.; MA, J.; GOSELINK, R. e KOK, A. **Extending lactation length: consequences for cow, calf, and farmer**. Journal of Animal Science, v. 100, 2022. DOI: https://doi.org/10.1093/jas/skac220.
KRAETZER, Thiago Lorente e BALTHAZAR, Glauber da Rocha. **FISHBOARD: an electronic device for analysis of productive data in pisciculture (fish-farming) / fishboard**. Brazilian Journal of Development, v. 7, n. 3, p. 28513-28533, 2021. DOI: http://dx.doi.org/10.34117/bjdv7n3-526.

KRPALKOVA, L.; MAHONY, N.; CARVALHO, A.; CAMPBELL, S. e WALSH, J. **Evaluating the economic profit of reproductive performance through the integration of a dynamic programming model on a specific dairy farm**. Czech Journal of Animal Science, v. 65, p. 124-134, 2020. DOI: https://doi.org/10.17221/38/2020-cjas.

KRUSZYŃSKI, W.; KARWACKI, M.; MICHALSKI, A. e SOBCZUK, M. **Analysis of culling reasons and longevity of Holstein-Friesian cows in different herd sizes**. Animals, v. 11, n. 9, 2541, 2021. DOI: 10.3390/ani11092541.

LAERD STATISTICS. **One-way Welch ANOVA and Games-Howell post-hoc test using SPSS Statistics**. 2023. Disponível em: https://statistics.laerd.com.

LEBLANC, S. **Assessing the association of the level of milk production with reproductive performance in dairy cattle**. Journal of Reproduction and Development, v. 56, supl. S, p. S1-S7, 2010.

LUNA, M. **Effect of reproductive indices and bovine somatotropin on milk**

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

yield in Holstein cows experiencing climatic sub-fertility. Mljekarstvo, 2023. DOI: https://doi.org/10.15567/mljekarstvo.2023.0302.

MARUMO, J. L. M.; YOKOO, M. J. e CUNHA, D. N. F. V. **Environmental and physiological factors affecting milk yield and heat-stress tolerance in Holstein cows under tropical conditions**. Journal of Dairy Science, v. 105, n. 3, p. 2448-2461, 2022.

MICIC, N. **The effect of animal-related and some environmental effects on daily milk production of dairy cows under the heat stress conditions**. Mljekarstvo, 2022. DOI: https://doi.org/10.15567/mljekarstvo.2022.0406.

MOORE, David S. **A estatística básica e sua prática**. 9. ed. Rio de Janeiro: LTC, 2023. 626 p.

NAM, K. T.; LEE, S. H.; KIM, J. W.; PARK, S. B. e CHOI, H. J. **Effect of the Temperature–Humidity Index on Productivity and Recovery in Dairy Cows**. Animals, v. 14, n. 19, 2848, 2024. DOI: 10.3390/ani14192848.

NEMECKOVA, D.; STADNIK, L. e CITEK, J. **Associations between milk production level, calving interval length, lactation curve parameters and economic results in Holstein cows**. Mljekarstvo, v. 65, n. 4, p. 243-250, 2015. DOI: 10.15567/mljekarstvo.2015.0404.

RAVAGNOLO, O. e MISZTAL, I. **Genetic Component of Heat Stress in Dairy Cattle, Parameter Estimation**. Journal of Dairy Science, v. 83, n. 9, p. 2120-2125, 2000.

SILVA, Emanuel I. C. da. **Fisiologia do Ciclo Estral dos Animais Domésticos**. Instituto Agronômico de Pernambuco. Disponível em: https://hal.science/hal-04177588/document. Acesso em: 23 out. 2024.

SILVA, Adryan Felipe Marques da; BALTHAZAR, Glauber da Rocha; FERREIRA, Eduardo Augusto; SANTOS, Everton Souza dos. **Comparative analysis of blocking and non-blocking models in rest APIS**. Cuadernos de Educación y Desarrollo, v. 17, n. 8, e9049, 2025. DOI: https://doi.org/10.55905/cuadv17n8-013.

TERRY, S.; BASARAB, J.; GUAN, L. e MCALLISTER, T. **Strategies to improve the efficiency of beef cattle production**. Canadian Journal of Animal Science, v. 101, p. 1-19, 2020. DOI: https://doi.org/10.1139/cjas-2020-0022.

TOLEDO, I. M.; SOARES, R. F.; ALMEIDA, G. L. e PEREIRA, M. C. **Seasonal pattern of multiparous dairy cow behavior in a free-stall farm in a humid subtropical climate**. JDS Communications, v. 5, n. 4, p. 243-250, 2024. DOI: 10.3168/jdsc.2023-0032.

REMUNOM
ISSN 2178-6925

Received: 04/06/2026 - **Accepted:** 04/07/2026
Vol: 13.15
DOI: 10.66104/cg3m1879
Pages: 1-31

TRICARICO, J.; KEBREAB, E. e WATTIAUX, M. MILK **Symposium review: Sustainability of dairy production and consumption in low-income countries with emphasis on productivity and environmental impact**. Journal of Dairy Science, v. 103, n. 11, p. 9791-9802, 2020. DOI: https://doi.org/10.3168/jds.2020-18269.

UTAMI, K. e WIDIARSO, B. **Assessment of heat stress in dairy cows related to physiological responses**. Jurnal Ilmu-Ilmu Peternakan, v. 32, n. 2, 2022. DOI: https://doi.org/10.21776/ub.jiip.2022.032.02.14.

VITALI, A.; CONTI, G.; BELLO, R. e BRAMBILLA, L. **Seasonal pattern of mortality and relationships between mortality and temperature-humidity index in dairy cows**. Journal of Dairy Science, v. 92, n. 8, p. 3781-3790, 2009. DOI: 10.3168/jds.2009-2127.

WEST, J. W. **Effects of heat stress on production in dairy cattle**. Journal of Dairy Science, v. 102, n. 5, p. 376-388, 2019.

WEST, J.W.; MULLINIX, B.G.; BERNARD, J.K.. **Effects of Hot, Humid Weather on Milk Temperature, Dry Matter Intake, and Milk Yield of Lactating Dairy Cows**. Journal Of Dairy Science, [S.L.], v. 86, n. 1, p. 232-242, jan. 2003. American Dairy Science Association. http://dx.doi.org/10.3168/jds.s0022-0302(03)73602-9.

ZHU, Jingyu; ZHAO, Xintong; SUN, Yu; SONG, Shaoxu e YUAN, Xiaojie. **Relational Data Cleaning Meets Artificial Intelligence: a survey**. Data Science And Engineering, v. 10, n. 2, p. 147-174, 20 dez. 2024. DOI: http://dx.doi.org/10.1007/s41019-024-00266-7.