ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

SCIENZE E TECNOLOGIE AGRARIE, AMBIENTALI E ALIMENTARI

Ciclo 37

Settore Concorsuale: 07/C1 - INGEGNERIA AGRARIA, FORESTALE E DEI BIOSISTEMI

Settore Scientifico Disciplinare: AGR/10 - COSTRUZIONI RURALI E TERRITORIO AGROFORESTALE

SMART MONITORING AND DATA-DRIVEN ADAPTIVE CONTROL OF LIVESTOCK ENVIRONMENT IN DAIRY CATTLE BARNS. DEVELOPMENT OF A PREDICTIVE MODELLING AND FORECASTING FRAMEWORK TO ENHANCE ANIMAL WELFARE AND ENERGY EFFICIENCY

Presentata da: Carlos Alejandro Perez Garcia

Coordinatore Dottorato

Diana Di Gioia

Supervisore

Stefano Benni

Co-supervisore

Daniele Torreggiani

Esame finale anno 2026

PHRASE

"Prediction is very difficult, especially if it's about the future"

Niels Bohr

DECLARATION

I hereby declare that this thesis is the outcome of the research conducted during my Ph.D. program. All contributions from other authors have been duly acknowledged and cited in the bibliography. This work is original and has not been submitted, in whole or in part, to any other institution or university for the award of a degree or qualification.

Carlos Alejandro Perez Garcia

Bologna, October 2025

ABSTRACT

The growing demand for sustainable and welfare-oriented livestock production has intensified the need for more intelligent and adaptive environmental control systems in dairy farming. This research proposes a data-driven framework for forecasting key environmental indices in dairy cattle barns by integrating Internet of Things (IoT) technologies with neural-network-based predictive models, with the aim of enhancing both thermal comfort and energy efficiency. Specifically, two of the most widely used thermal indices the Temperature-Humidity Index (THI) and the Equivalent Temperature Index (ETI); were selected. These indices were modeled using high-resolution environmental data collected by a Smart Monitoring System deployed in two commercial dairy farms.

The IoT system enabled the continuous acquisition of environmental parameters at several representative points within the barns, allowing the generation of multizonal representations of THI through spatial interpolation. This approach provided new insights into the spatial dynamics of thermal variability, including the estimation of conditions in unmeasured areas. Moreover, the integration of an anemometer within the monitoring system represents one of the principal innovations of this research, as it allowed for the inclusion of wind speed, a critical factor influencing animal thermal comfort, in the computation of ETI.

A complementary case study enabled the simulation of energy consumption in the barn's ventilation system by comparing traditional control methods based on THI measurements with an alternative approach using ETI as the control variable. The results demonstrated that the proposed ETI-based strategy, combined with multi-point monitoring, enables a more localized and adaptive control of ventilation, resulting in significant energy savings while maintaining optimal animal welfare. These findings highlight the potential of predictive modeling and IoT integration to foster sustainable, data-driven management of environmental conditions in modern dairy farming, addressing one of the major challenges in the current environmental and geopolitical context.

SUMMARY

NOMENCLATURE TABLE

Symbol/Term	Description	unit (SI)
THI	Temperature-Humidity Index	–
ETI	Equivalent Temperature Index	℃
IoT	Internet of Things	–
MAE	Mean Absolute Error	–
RMSE	Root Mean Square Error	–
IDW	Inverse Distance Weighting	–
SMS	Smart Monitoring System	–
AMS	Automatic Milking System	–

CHAPTER 1. INTRODUCTION AND LITERATURE REVIEW

1.1 Research Context

Environmental conditions play a pivotal role in livestock farming, parameters such as temperature, relative humidity, solar radiation, and wind speed directly and indirectly affecting both animal welfare and production. High ambient temperatures can lead to an increase in body temperature affecting the animal's thermoregulatory capacity and evoke a physiological response that results in voluntary feed intake, animal weight, fertility, and production (Adjassin et al., 2020).

The sum of all external factors that invoke an adverse physiological response in animals is known as heat stress (Ji et al., 2020). This phenomenon has received significant attention from stakeholders, largely due to the observed increase in global temperatures and the frequency of heat waves in recent years. In the case of dairy cattle, the effects of heat stress can be observed in physiological and behavioral changes. Documented physiological coping strategies include an increased respiration rate, panting, and sweating, as well as a reduction in milk yield and reproductive performance. Behavioral responses to heat stress include modified drinking and feed intake (e.g., increased water intake and shifting feeding times to cooler periods during the day), increased standing time and shade seeking, and decreased activity and movement (Polsky & von Keyserlingk, 2017).

Commonly used methods to evaluate cattle heat stress include a variety of physiological and environmental indicators. While physiological indicators provide direct insight into the animal's condition, their observation is often time-consuming and labor-intensive. Additionally, the equipment required for such measurements is not feasible for large-scale commercial operations, limiting its practicality. Consequently, thermal comfort index models are more frequently employed to assess heat stress. These models integrate various environmental factors, offering a more efficient and scalable approach to monitoring and managing heat stress in cattle.

Thermal Indices (TIs) are formulated by integrating various thermal environmental parameters that represent the influence of latent and sensible heat transfer between the animals and its environment. Among the TIs relevant to dairy cattle, the Temperature-Humidity Index (THI) is the most widely used. Originally derived from the human discomfort, THI has undergone numerous modifications to improve its applicability and accuracy for cattle (Herbut et al., 2018; Thom, 1959). Over the years, other TIs have been developed to address specific dairy needs and conditions.

Other commonly used thermal indices in dairy cattle applications are the Respiratory Rate Predictor (RRP) and the Heat Load Index (HLI). These indices were developed for feedlot cattle and focus on predicting and managing heat stress in these environments (Hoffmann et al., 2020). The Equivalent Temperature Index for Cattle (ETIC) was developed with high-yielding dairy cows in mind and provides a tailored approach to assessing heat stress in these highly productive animals (Wang et al., 2018). In addition, the Index of Thermal Stress for Cows (ITSC) was created to assess the thermal stress experienced by cows in tropical environments where the challenges of heat and humidity are particularly pronounced (Da Silva et al., 2014).

A dynamic data-based approach to TIs is particularly important in environments where conditions are constantly changing. This approach recognizes that due to the dynamic nature of inputs, such as animal heat production and ventilation rates, as well as external factors like outdoor climate, the calculated indoor conditions fluctuate rapidly over time (Banhazi et al., 2009). Therefore, dynamic data-based models operate on the premise that animal responses to their microenvironment must be continuously measured, and that this information should be feedback into the system's controller. Additionally, these models rely on mathematical representations that accurately predict process responses under both steady-state and dynamic conditions. By using advanced mathematical identification techniques, dynamic data-based models can estimate unknown parameters in real-time, based on ongoing measurements of process inputs and outputs. This results in an adaptive model capable of managing the time-variant nature of bioprocesses, such as those in livestock farming (Aerts et al., 2000).

1.1.1 IoT and Data Analytics in Agriculture

The dairy industry's shift towards larger farms and increased automation has given rise to significant challenges in the monitoring of cow welfare, health, and productivity. Farmers are less frequently on-site and have limited time to assess each cow individually daily (Norton et al., 2019). As a result, new strategies are needed to effectively oversee and manage dairy cows. This requirement has prompted the adoption of advanced Internet of Things (IoT) technologies in livestock management (Navarro et al., 2020). These technologies enhance data management by linking sensors, controllers to communication networks, creating an information-based, automated, and intelligent system. Such a system allows for the remote handling and integration of data from multiple sensors through data fusion, facilitating continuous and automatic monitoring of animals (Zhang et al., 2021). Consequently, it provides farmers with valuable support in the control and management of their herds (Berckmans, 2017).

In Alonso et al., (2020), a platform was developed with the objective of monitoring the state of dairy cattle and feed grain in real time. This platform integrates a number of emerging technologies, including the IoT, Edge Computing (EC), Artificial Intelligence (AI), and Blockchain, within the novel Global Edge Computing Architecture (GECA). This architectural approach optimizes the utilization of computational, storage, and network resources by processing data at the network edge, thereby enhancing service response times, quality of service, and security. The platform was evaluated on a real dairy farm, demonstrating notable enhancements in data traffic reduction and communication reliability between the IoT-Edge layers and the Cloud. In conclusion, this platform offers a comprehensive solution for the dairy industry to meet the demands of a globalized and competitive market by leveraging advanced technologies to optimize processes, ensure product quality, and promote sustainability.

On the other hand, Amiri-Zarandi et al., (2022) present a platform approach of smart farming information processing, with the objective of addressing the growing challenges in agricultural productivity resulting from the rapid population increase and heightened food demand. This platform employs technologies, including IoT and AI, to optimize agricultural efficiency and productivity. The system collects data from a variety of sources, including wireless sensor networks, weather stations, monitoring cameras, and smartphones, and then integrates these data sets into a unified system. A noteworthy aspect of this work is its emphasis on addressing the challenges posed by the diversity of agricultural data formats and meanings, as well as the limited interoperability among different smart farming services and technologies. The platform was designed to ensure interoperability, reliability, scalability, real-time data processing, end-to-end security and privacy, and adherence to standardized regulations and policies. The practical deployment and assessment of this platform have revealed its capacity to markedly enhance the productivity, profitability, and overall performance of smart farms. By providing a unified solution that supports data sharing and cooperation among various stakeholders, this approach highlights the necessity of integrating modern technologies to meet the demands of a globalized agricultural market.

Moreover, in the study Leliveld et al., (2024), an automatic system for integrating multiple sensors relating to barn environment and cow behavior was developed. This system collects sensor data that measures various parameters, including indoor conditions, CO_2 concentration, water use, and cow behavior, such as lying, eating, and ruminating. The system also uses a weather system from the outside and video recording devices. The architecture of the system has four main components, which are the sensors, nodes, gateways, and backend. The data recorded by the sensors are locally processed and then transmitted through radio channels to the local gateways, which collect and relay the data from many nodes and forward it across a 3G/4G cellular network to distributed digital storage in "the cloud" where the information is process and stored in a database. The users can visualize the date recorded in real time online an internet-based dashboard. Overall, while the system presents a sophisticated solution for integrating environmental and behavioral monitoring in dairy farming, its practical implementation may be constrained by cost, connectivity, and technical requirements. Our research aims to address these challenges by exploring more cost-effective sensors, alternative data transmission methods, and user-friendly interfaces to enhance accessibility and reliability for all scales of dairy farming operations.

1.1.2 Machine Learning in Agriculture

To effectively integrate the advancements in the IoT and data analytics with the transformative capabilities of Machine Learning (ML) in agriculture, it is essential to understand how these technologies interact to foster innovation within the sector. The IoT facilitates the continuous acquisition of real-time environmental and animal health data on dairy farms, while ML serves as a powerful analytical tool that processes this information to derive meaningful, actionable insights

(Vranken & Berckmans, 2017). ML models can process large volumes of complex, multidimensional data, uncovering patterns and trends that would be difficult to detect manually (Banhazi et al., 2012; Banhazi & Black, 2009; Cockburn, 2020). This capability not only enhances predictive accuracy for environmental indices but also facilitates the development of adaptive, data-driven strategies for improving animal welfare and optimizing production. By integrating IoT data streams with advanced machine learning algorithms, farmers can make informed decisions, leading to more sustainable and efficient agricultural practices (Neethirajan, 2020).

The development of ML models can be categorized into three main approaches, which are applicable to a variety of applications, including the prediction of environmental indicators and the assessment of animal welfare. The initial approach entails the development of models based on fundamental physical principles, which are often referred to as White-box models. These models require substantial modeling effort and ongoing maintenance, particularly as the system or its components evolve. In contrast, purely data-driven (Black-box) models take advantage of the data availability generated by IoT systems, and reproduce highly complex nonlinear phenomena, including those for which theories have not been proposed yet, and they are defined as models where no understanding of the model structure and its parameters is required. Lastly, Grey-box models integrate a partial theoretical structure with empirical data to yield a more comprehensive representation of the phenomenon under study. This approach can address the limitations of Black-box models, which lack a clear theoretical foundation and explanatory power. Nevertheless, the incorporation of parameters that are grounded in underlying biological or physical processes represents a significant challenge. Consequently, despite its limitations, the use of Black-box models persists as one of the most prevalent approaches among researchers in this field.

There is a considerable amount of research that has been conducted to investigate the potential of ML to address various challenges in agriculture. (Milan et al., 2018) put forth a comprehensive discourse on the implementation of ML algorithms for the monitoring of diverse physiological and behavioral parameters in dairy cows. The authors demonstrated the potential of ML techniques for accurately predicting a range of physiological and behavioral parameters, including internal and surface temperatures, respiration rates, sweating rates, gait patterns, behaviors, physical dimensions, weight, and body condition scores. The study underscored the importance of individualized monitoring to optimize animal health and productivity, particularly in precision dairy farming. It also addressed the challenges associated with these technologies, such as the time-consuming and computationally intensive nature of ML algorithms and the difficulty of generalizing models across different farm environments. Despite these challenges, the research demonstrated the promising role of ML in advancing the precision and effectiveness of dairy farming practices, ultimately contributing to better management and welfare of dairy cattle.

Gorczyca & Gebremedhin, (2020) examine the impact of environmental heat stressors on the physiological responses of dairy cows, a critical factor affecting both their health and productivity. In contrast to conventional models that frequently depend on researcher-defined assumptions regarding the correlation between environmental factors and physiological responses, this study employs ML algorithms to assess these relationships in a more data-driven and impartial manner. By employing nonlinear machine learning techniques such as random forests and neural networks, the study accurately predicted key physiological responses (respiration rate, skin temperature, and vaginal temperature) with high accuracy, as evidenced by low root mean squared error (RMSE) values. Among the environmental variables considered, air temperature had the most significant effect on all physiological responses, while wind speed had a comparatively smaller impact. The innovative aspect of this study is its use of machine learning to rank the effects of these environmental stressors, thereby offering a more nuanced understanding of their interactions without sacrificing model performance. The study’s conclusions indicate that future research could expand these models to include larger datasets and multiple livestock species, which may enhance the precision and applicability of heat stress management strategies across diverse farming environments.

In their research (Ji et al., 2022) examines the increasing utilization of Automatic Milking System (AMS) in contemporary livestock farming, emphasizing their capacity to diminish labor costs and augment animal welfare and productivity through the implementation of enhanced monitoring. Specifically, the study utilizes a dataset collected over a five-year period from 80 cows on a robotic dairy farm, with the objective of developing a machine learning framework. The objective of this framework is to facilitate the automatic training of models with up-to-date farm data, enabling the prediction of key variables such as daily milk yield, composition (fat and protein content), and milking frequency of individual cows over a 28-day period. A noteworthy advancement highlighted in this research is the high degree of accuracy achieved by the machine learning models, with a coefficient of determination (R^2) greater than 0.90 and an overall accuracy exceeding 80%. The practical implications of this research indicate that the developed framework has the potential to significantly enhance management efficiency and animal welfare in robotic dairy farms. By predicting important metrics related to milk production and cow behavior, farmers can make more informed decisions, which ultimately leads to improved farm productivity and animal health.

1.1.3 Advanced Environmental Control

The automatic control of the microenvironment in livestock buildings has traditionally been based on feedback from internal air temperature measurements taken at a single location within the building. Thermostatically controlled climate systems traditionally use set points for environmental variables assumed to be ideal for an average animal. These set points are typically derived from a combination of small-scale laboratory experiments. Despite the application of modern techniques, this approach frequently fails to achieve the desired outcomes, as it oversimplifies the complex interactions between an animal and its environment. This simplification often results in suboptimal environmental conditions, as it fails to account for the variability in individual animals' needs and the dynamic nature of the microenvironment (Fournel et al., 2017).

Consequently, livestock farmers may struggle to meet consumer expectations for animal products or adhere to the increasingly stringent regulations aimed at reducing the environmental impact of production systems and enhancing animal welfare (Aerts et al., 2003). The growing demand for more responsive livestock management practices has resulted in the adoption of precision livestock farming systems. These systems include: (i) methods for electronically measuring critical components of production systems related to resource use efficiency; (ii) software tools designed to interpret the captured information; and (iii) control processes that optimize animal productivity. By integrating real-time monitoring and control systems, these technologies promise to radically enhance the production efficiency of livestock enterprises (Kuczynski et al., 2011).

The importance of precise environmental control is underscored by case studies, such as those reported by Black & Banhazi, (2013). These cases demonstrate the efficacy of environmental control in the prevention of significant economic losses. In both cases, the failure to maintain optimal conditions for animal rearing resulted in adverse effects on growth rates and feed intake, leading to substantial financial losses.

The implementation of advanced control systems based on predictive information remains relatively limited in the context of livestock buildings. The integration of machine learning tools with IoT technologies has the potential to transform existing control paradigms, offering more efficacious and responsive strategies. While the present study is concerned with the development of multizonal predictive models, future research will be directed towards the application of this knowledge for the enhancement of control systems in livestock management.

1.2 Research Problem Statement

The application of predictive models provides valuable insights into future scenarios of indoor environmental parameters, thereby facilitating proactive management of conditions that impact dairy cattle welfare. By integrating multizonal forecasts of the Temperature-Humidity Index (THI) with spatial interpolation, this research enhances understanding of thermal index behavior across various barn zones, enabling a more detailed assessment of animal welfare throughout the indoor space. Furthermore, the advent of the Internet of Things (IoT) has enabled the measurement of additional parameters beyond temperature and humidity, including wind speed. The aforementioned capability enables the calculation of indices such as the Equivalent Temperature Index (ETI), which provides a more comprehensive approach to thermal comfort management. The incorporation of ETI into a data-driven control system is intended to optimize the operational regimen of ventilation systems, thereby balancing the need for animal welfare with the objective of energy efficiency. This research addresses critical challenges in livestock environmental control by integrating predictive modeling, IoT, and advanced thermal indices in order to provide actionable insights for the development of more effective and sustainable management practices in dairy cattle barns.

1.3 Research Aim

The objective of this research is to enhance the environmental control systems in dairy cattle barns through the utilization of a neural-network predictive modelling framework, multizonal forecasting, and the incorporation of advanced thermal indices. By integrating data-driven approaches with Internet of Things (IoT) technologies, the study seeks to enhance both animal welfare and energy efficiency in managing indoor environmental conditions.

1.4 Specific Research Objectives

- Develop a predictive model using NeuralProphet to forecast the Temperature-Humidity Index (THI) across various zones within the dairy barn, providing insights into future environmental conditions that impact animal welfare.
- Implement a spatial interpolation technique to generate a multizonal THI distribution within the barn, enabling a comprehensive understanding of thermal conditions across different areas and heights.
- Utilize wind speed measurements from integrated anemometers to compute the Equivalent Temperature Index (ETI) and develop a predictive model for ETI, assessing its potential as a control variable for optimized ventilation systems.
- Evaluate the energy efficiency of an ETI-based control system compared to traditional ventilation control methods, quantifying the potential energy savings and identifying opportunities for sustainable barn management.

1.5 Outline of the Thesis

This doctoral thesis stems from the research activities carried out during the Ph.D. programme and is based on the scientific works developed and published within this period, including "*3D Numerical Modelling of Temperature and Humidity Index Distribution in Livestock Structures: A Cattle-Barn Case Study*" (Perez Garcia et al., 2023), "*Indoor Temperature Forecasting in Livestock Buildings: A Data-Driven Approach*" (Perez Garcia et al., 2024) and "*Predictive Modeling of Energy Consumption for Cooling Ventilation in Livestock Buildings: A Machine Learning Approach*" (Perez Garcia et al., 2025). The thesis manuscript is structured in the following manner:

- CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

This chapter presents an overview of the research context and explores the importance of environmental control in livestock production, particularly for dairy cattle. It reviews key concepts and recent advances in IoT, data analytics, and machine learning in agriculture, focusing on thermal indices and their applications to improve animal welfare and productivity.

- CHAPTER 2: MATERIALS AND METHODS

The second chapter details the methods used for data collection, system architecture, and computational tools, with a focus on Python-based data analysis for time series prediction. It explains the multi-layered structure of the Smart Monitoring System, its components, and the integration of the NeuralProphet framework for predictive modeling.

- CHAPTER 3: MODEL DEVELOPMENT AND VALIDATION

This chapter describes the development of predictive models for temperature-humidity index (THI) and equivalent temperature index (ETI) forecasting. It covers model selection, the training process, and the validation metrics used to assess accuracy. The chapter also discusses the importance of incorporating exogenous variables and environmental indicators.

- CHAPTER 4: RESULTS AND DISCUSSION

The results of the prediction models are presented and analyzed. The chapter includes a comparative evaluation of different scenarios based on THI and ETI predictions and discusses their implications for energy consumption, ventilation control, and animal welfare. The results of spatial interpolation methods and energy consumption simulations are also detailed.

- CONCLUSIONS

In this final chapter, the research findings are synthesized to highlight the contributions of a data-driven, neural network-based predictive modeling approach for optimizing environmental control in dairy cattle barns.

CHAPTER 2. MATERIALS AND METHODS

This chapter presents a thorough examination of the materials and methods employed in the development of a framework for time series forecasting, with the objective of monitoring key environmental indices within livestock facilities. This study employs a variety of Python libraries to enhance the monitoring of environmental conditions, which are critical to ensuring animal welfare and optimizing productivity, with a particular focus on the extensive capabilities of Python in data analysis and machine learning. The chapter initiates with an examination of the rationale behind the selection of Python as the principal tool for data processing, model development, and visualization. Subsequently, the architecture and components of the Smart Monitoring System (SMS) are described, which integrates Internet of Things (IoT) devices for the acquisition of real-time data within the barn. Furthermore, essential Python libraries and the NeuralProphet framework, which is employed for predictive modeling, are discussed alongside the methodology for data processing and model validation. Additionally, the use of interpolation techniques, specifically Inverse Distance Weighting (IDW), to estimate climatic variables in unmonitored areas is highlighted, as this demonstrates the system's adaptability in managing microclimatic variations across the barn. This comprehensive approach provides the foundation for the predictive models and environmental control strategies explored in subsequent chapters.

2.1 Python-Powered Data Analysis for Time Series Forecasting

Monitoring and controlling environmental parameters within livestock buildings are fundamental to safeguarding animal welfare and optimizing production efficiency, particularly in relation to milk yield. In this study, data acquired through the Smart Monitoring System (SMS) are processed and analyzed using advanced machine learning techniques implemented in Python. This approach enables the extraction of meaningful patterns from environmental data, facilitating the prediction of key thermal indices and supporting informed decision-making for improved environmental management in dairy production systems.

2.1.1 Application of Python in the Research Framework

Python was selected as the primary programming language for this research due to its versatility, ease of use, and extensive support for scientific computing and data analysis. In the context of this study, Python serves as the foundation for the development of time series forecasting models that predict critical environmental indices within farm buildings. The extensive library ecosystem provided by Python makes it an optimal choice for handling complex time series data, facilitating detailed analysis and model development.

The versatility and flexibility of the Python programming language represent its most significant strengths. The language supports multiple programming paradigms, including procedural, object-oriented, and functional programming. This allows developers to approach problems from various angles, thereby enhancing the flexibility and adaptability of the solutions developed. This is of particular significance in the dynamic and multifaceted environment of the IoT and SMS, where the capacity to rapidly adapt to new data or changing conditions is crucial.

Furthermore, the ease of use and readability of the Python programming language contribute to its suitability for this research project. The language's uncomplicated syntax minimizes the intricacy of coding, enabling accelerated development cycles and facilitating expeditious prototyping and iterative testing. This is especially advantageous in research contexts where the capacity to rapidly test and refine models is crucial. Furthermore, the readability of Python facilitates collaboration among team members, thereby ensuring that the codebase remains accessible and maintainable over time. Python is readily capable of interfacing with a multitude of hardware components, establishing connections to databases, and interacting with cloud services. This versatility allows for the deployment of time series forecasting models in real-world environments with minimal effort. This capability ensures that the models developed are not only theoretically robust but also practically implementable in dynamic and complex systems.

Another significant benefit of Python is its extensive and dynamic community, which drives continuous development and rapid evolution of its libraries. This guarantees that researchers have access to the most recent developments in machine learning and data science, which can be readily applied to their work. Furthermore, the extensive documentation and resources available within the Python community facilitate the troubleshooting of issues and ensure researchers remain apprised of new methodologies.

The capacity of Python to process voluminous datasets and its compatibility with high-performance computing frameworks render it an optimal choice for the expansion of time series forecasting models, a necessity in IoT environments where data is generated at a continuous and high rate. The scalability of Python ensures that the forecasting models remain responsive and accurate even as the system undergoes expansion.

2.1.2 Key Python Libraries for Data Analysis

This research relies on a variety of Python libraries, each essential to different phases of data processing, analysis, and model development:

- *Pandas* is an indispensable tool for data manipulation and preprocessing. It offers a range of tools for handling time series data, enabling efficient data cleaning, filtering, and transformation. The Dataframe structure provided by Pandas ensures that datasets are optimally prepared for the training of forecasting models.
- *Scikit-learn* is a widely recognized tool for machine learning, particularly useful in the preprocessing and evaluation stages of model development. It encompasses utilities for data scaling, feature selection, and the partitioning of datasets into training and testing sets. Furthermore, Scikit-learn offers a comprehensive set of metrics for evaluating the accuracy and reliability of forecasting models.
- *Statsmodels* is a tool for statistical modeling, which is employed to construct conventional time series models such as ARIMA and SARIMA. These models serve as benchmarks within the research, offering a comparative basis for evaluating the effectiveness of more advanced models. Furthermore, Statsmodels facilitates comprehensive statistical analysis, elucidating intricate patterns and structures within the time series data.
- *Plotly* facilitates enhanced visualization of research data through the creation of interactive graphs and visualizations. It supports a wide variety of chart types, including three-dimensional plots, which are particularly useful for visualizing complex datasets and relationships between variables. Plotly’s interactive features permit users to explore data in

a dynamic manner, thereby facilitating a more profound comprehension of trends and patterns within time series data and ensuring that visual representations are both informative and engaging.

- The *InfluxDB* Python client library is a key component in the process of interfacing with the InfluxDB database, which is utilized in the context of this research. The library enables efficient access to time series data stored in the database, facilitating seamless data retrieval and integration with other analysis tools. The library provides the functionality to query, write data, and manage the database, thereby ensuring that the forecasting models have real-time access to the most up-to-date environmental data.

While several Python libraries are utilized for data processing and analysis in this research, the development of the forecasting models specifically relies on the *NeuralProphet* library. NeuralProphet is a time series forecasting tool that extends the functionality of the Prophet library by incorporating neural network components, thereby enabling it to capture more complex patterns in data. The introduction of NeuralProphet in this section is intended merely to provide an overview of its role in the forecasting models, which will be explored in greater detail in a subsequent section.

2.2 Smart Monitoring System

The Smart Monitoring System (SMS) has been deployed in the frame of the project “Smart Dairy Farming: Innovative Solutions to Improve Herd Productivity” (Italian PRIN, Research Projects of Relevant National Interest - Call for Proposals 2017) following a multi-layered architecture (Figure 2.1).

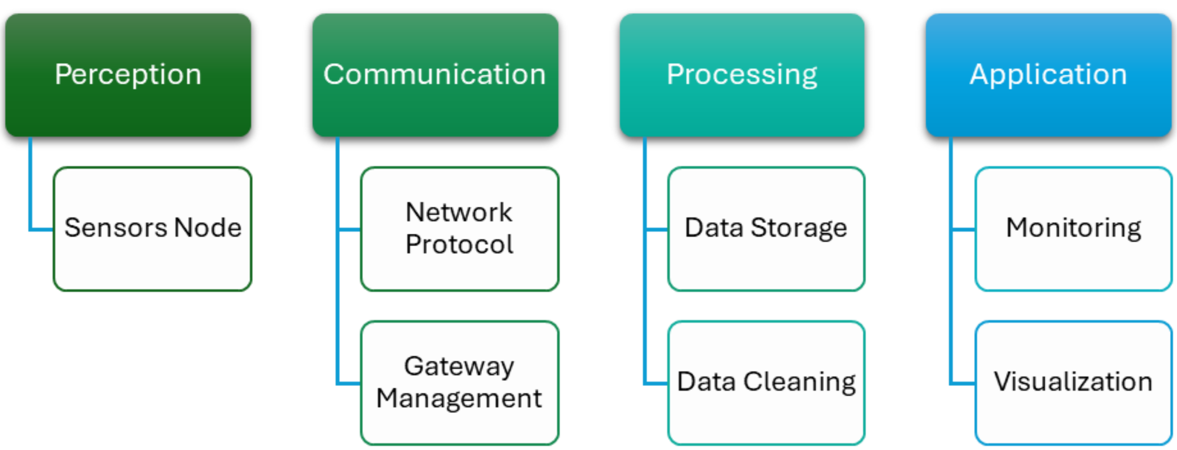

Figure 2.1 Smart Monitoring System architecture.

This architecture ensures the efficient collection, transmission, processing, and visualization of environmental data that are crucial for monitoring barn conditions. Each layer of the system plays an essential role in its overall functionality, enabling real-time data analysis and decision-making that are crucial for optimizing herd productivity and welfare in diverse farming environments. The following sections provide a detailed overview of the components and functions of each layer within the SMS, highlighting their roles in achieving an effective and responsive monitoring system.

2.2.1 Perception Layer

The Perception layer serves as the foundation of the SMS, responsible for directly interacting with the physical environment through a network of strategically placed sensor nodes across the farm. The system is designed to measure critical environmental parameters that affect animal welfare, such as ambient temperature, relative humidity and air velocity.

The sensor nodes (Figure 2.2) are based on electronic boards with a flexible architecture, enabling the integration of different transducers according to the specific application requirements. At the core of these devices an ARM Cortex-M0+ microcontroller processes the electrical signals received from the sensors and performs preliminary data processing tasks. These tasks include data type conversion (e.g., from float to integer) and information compaction, among others. Once the data has been preprocessed, it is packaged into a data frame and transmitted to the communication module, that operates wirelessly using a radio module on the ISM band at 868 MHz, employing LoRa modulation. Given the need to install these nodes in strategic locations, including areas with limited access to electrical power, the sensor nodes are designed to be battery-operated. To minimize long-term maintenance costs, particularly regarding battery replacement, the nodes are equipped with energy harvesting capabilities.

Figure 2.2 Sensor node prototype

The deployment at the Montagnini family farm (later described as the second case study) utilized a standard sensor node complemented by an anemometer for measuring wind speed and direction (Figure *2.3*). This supplementary measurement is crucial for understanding the airflow within the barn and its impact on the thermal environment, as air velocity significantly influences heat dissipation and animal comfort. The inclusion of wind speed data is particularly important for the calculation of the Equivalent Temperature Index (ETI), a more representative measure of animal welfare compared to indices that only consider temperature and humidity

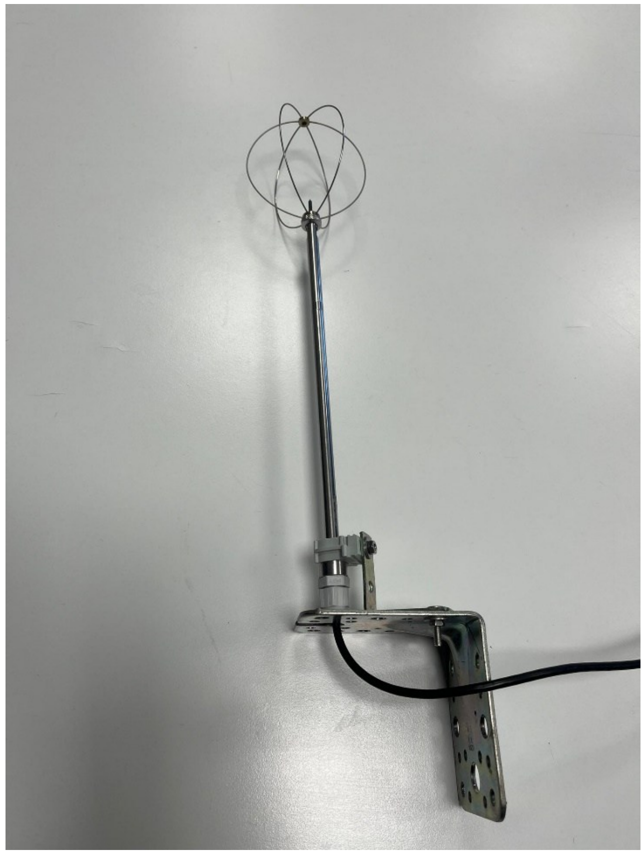

Figure 2.3 Anemometer probe

An important consideration in the development of the SMS has been the environmental protection of the hardware components. Both the sensor nodes and the gateway are equipped with physical protection to withstand the high levels of contamination often present within dairy farms. This includes protection against dust, humidity, airborne particles, and potentially harmful gases. This robust design ensures the durability and reliability of the system in challenging agricultural environments, enabling continuous and accurate monitoring of critical environmental conditions.

2.2.2 Communication Layer

In order to comprehend the capabilities of the LoRaWAN protocol as implemented in the SMS in their totality, it is imperative to understand its fundamental features. The following Table *2-1* offers a concise summary of these features. The combination of these features positions LoRaWAN as an optimal solution for SMS communication requirements. By enabling reliable, long-range, and energy-efficient communication, LoRaWAN ensures that the data collected from sensors across the farm is transmitted securely and effectively to the central processing systems.

Table 2-1 LoRaWAN specifications

Features	LoRaWAN
Modulation	Chirp based spectrum
Data rate	$0.3 - 50.0$ kbps
Interference immunity	Very high
Transmission range	up to 15 km in rural areas
Bandwidth	$433, 868, 915$ MHz

To support a modular architecture that allows for the seamless addition of new sensors, an extended *Star Topology* was implemented within the network. In this configuration, gateways function as transparent bridges, forwarding messages between the end devices (sensors node) and a central network server (LoRaWAN Network Server, LNS). This setup enhances the system's scalability and flexibility, enabling the network to grow and adapt as the monitoring needs of the farm evolve.

LoRaWAN operates at the physical layer of the OSI model (PHY), utilizing the air as the medium for transmitting LoRa radio waves from an RF transmitter in an IoT device to an RF receiver in a gateway. Once these Radio Frequency (RF) messages are received, the gateway forwards them to the LNS via an IP backbone, which can be supported by various communication technologies such as Wi-Fi, Ethernet, or cellular connections (Figure 2.4). A key strength of this protocol is the absence of a fixed association between end devices and specific gateways. This feature significantly enhances the network's resilience, allowing multiple gateways to serve the same sensor. As a result, data transmission remains uninterrupted even if one gateway encounters issues, thereby ensuring the continuous operation of the monitoring system.

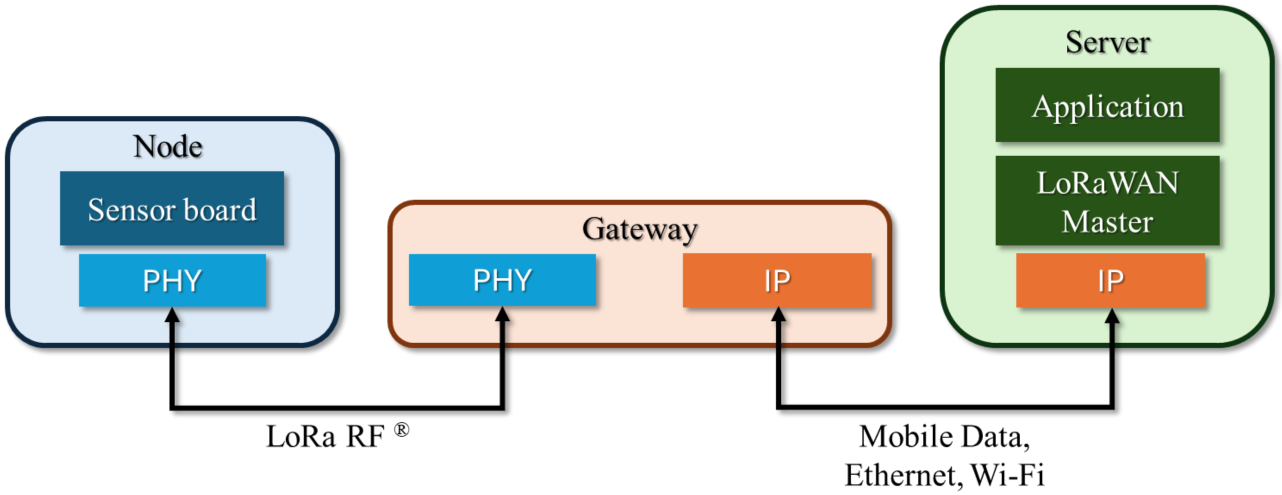

Figure 2.4 LoRaWAN communication diagram

The gateway module employed in the SMS is built around a Raspberry Pi III, a versatile and widely used single-board computer known for its robust features and flexibility. The Raspberry Pi III includes a 1.2 GHz 64-bit quad-core ARM Cortex-A53 processor, 1 GB of RAM, integrated Wi-Fi and Bluetooth connectivity, and multiple USB ports. The gateway operates on RAKPiOS, a Linux-based operating system developed by RAKwireless. RAKPiOS comes preconfigured with essential tools and software for the development and deployment of IoT applications, such as routing encrypted measurements from the sensor nodes to the cloud via the internet.

The gateway is connected to the internet through a network access point, this direct connection ensures that communications arriving at the gateway from the local wireless network are instantly forwarded to the cloud, where they are securely stored. Consequently, the communications from the gateway to the server maintain the same level of reliability as any other system directly connected to an internet access point. The monitoring system, therefore, does not introduce any additional level of uncertainty.

For Ethernet configurations, the gateway is equipped with a Power over Ethernet (PoE) adapter, which simplifies wiring by allowing a single cable to provide both power and data connection to the gateway. The PoE adapter must be connected to the electrical network, and the local LAN Ethernet cable must be connected for internet access. A single Ethernet cable is then routed from the PoE adapter to the gateway. The default configuration enables the gateway to acquire network settings via DHCP, allowing it to automatically connect to the cloud and be ready to forward information from the local LoRa network to the remote server.

The gateway is equipped with either an integrated antenna or an external high-gain antenna (with higher dBi) to enhance radio reception, ensuring reliable communication with sensor nodes even at extended distances. This setup is crucial for maintaining the integrity and efficiency of the SMS, enabling continuous data flow and effective monitoring of environmental parameters across the farm.

A critical function of the LoRaWAN network server is to manage the overall network, especially adapting to changing environmental conditions and maintaining communication security. The LNS dynamically controls the network parameters and establishes secure 128-bit AES connections for data transport. This ranges from end-to-end data that is going from the LoRaWAN end device to the cloud-based application to control traffic, which goes between the end device and the LNS. The server is responsible for authenticating each sensor on the network and verifying the integrity of each message so that the smart monitoring system it hosts is reliable and safe.

2.2.3 Processing Layer

The Processing Layer plays a pivotal role in the SMS, facilitating the seamless transfer and management of data as it moves from the communication layer to the final application. The tools selected for this layer were chosen to align with the LoRaWAN communication protocol and to ensure efficient data handling and processing.

ChirpStack

The first major component in this layer is ChirpStack, an open-source LoRaWAN Network Server and serves as the central hub for managing the LoRaWAN network, handling the registration and configuration of all network components, including sensor nodes and gateways. Through its web interface, users can easily manage devices, assigning them to specific gateways and monitoring their status in real-time. ChirpStack is responsible for processing data packets received from the sensor nodes. After being collected by the gateways, these packets are forwarded to the ChirpStack server, where they are decrypted and prepared for further processing. By ensuring that the data is correctly formatted, ChirpStack makes it ready for subsequent stages in the Processing Layer. Tis server is also responsible for handling encryption and decryption processes, which are essential to ensuring the security and integrity of data as it is transmitted across the network.

MQTT

Once the data has been processed by ChirpStack, it is transmitted using MQTT (Message Queuing Telemetry Transport). MQTT is a lightweight messaging protocol specifically designed for scenarios where bandwidth and power consumption are limited, making it ideal for IoT applications (MQTT, n.d.). MQTT operates on a publish-subscribe model, where data from the sensor nodes (publishers) is sent to a central broker, which then distributes the messages to interested parties (subscribers). This method of communication reduces network load and ensures that data is transmitted efficiently and reliably, even in environments with limited resources.

Node-RED

Following the transmission via MQTT, Node-RED takes over as the next key component of the Processing Layer. Node-RED is a programming tool that allows for the integration of various hardware devices, APIs, and online services into cohesive workflows. With its intuitive, browser-based editor, users can create and manage data flows by connecting different nodes representing data inputs, outputs, and processing steps (Node-RED, n.d.). In this system, Node-RED is tasked with unpacking the MQTT messages received from ChirpStack. It then routes the relevant data to the database for storage and further analysis.

By coordinating these components (ChirpStack for network management, MQTT for data transmission, and Node-RED for data routing and processing) the Processing Layer ensures that the SMS operates efficiently, securely, and is capable of evolving with the farm's changing needs.

2.2.4 Application Layer

The Application Layer is the final and most versatile component of the SMS, focusing on the customization and deployment of applications that utilize the data collected and processed by the underlying layers. This layer is designed to provide flexibility in how the data is utilized, visualized, and analyzed, adhering to the development paradigms commonly associated with IoT applications.

InfluxDB for Time Series Data Storage

At the core of the Application Layer is the InfluxDB database, a highly efficient solution for storing time series data, which is a critical aspect of IoT applications. InfluxDB is an open-source time series database designed to handle high write and query loads, making it particularly suitable for environments where data is continuously generated and needs to be analyzed in near real-time. Its architecture is optimized for rapid storage and retrieval of time-stamped data, which aligns perfectly with the requirements of the SMS, where environmental parameters are continuously monitored and recorded.

In IoT applications, InfluxDB is widely adopted due to its ability to efficiently manage and store large volumes of time series data. This database supports high availability and horizontal scaling, which ensures that it can handle the large data sets typical of modern IoT deployments. Furthermore, InfluxDB provides a powerful query language (InfluxQL) that allows users to perform complex queries on the time series data, facilitating advanced analytics and data-driven decision-making.

In the Smart Monitoring System, measurements of different types such as temperature, humidity, and wind speed; are systematically organized, with each type stored in its own dedicated table within the InfluxDB database. Each data entry is time-stamped, ensuring that the sequence and timing of measurements are accurately preserved. Unique identifiers, referred to as “sign ids”, are used as keys to differentiate between various data sources and ensure that the data is correctly associated with the corresponding sensor nodes. This structuring allows for efficient data retrieval and analysis, as queries can be tailored to specific time frames, sensor nodes, or types of measurements. The use of InfluxDB in the Application Layer not only enhances the system’s ability to store large volumes of data efficiently but also supports the flexibility needed to customize the application according to specific requirements.

Grafana for Data Visualization

To complement InfluxDB, Grafana was employed as the primary visualization tool within the Application Layer. Grafana is an open-source platform known for its powerful and flexible capabilities in data visualization, making it an ideal choice for monitoring systems that require real-time insights and historical data analysis.

Grafana connects seamlessly with InfluxDB, allowing users to create dynamic, interactive dashboards that visually represent the data collected by the SMS. These dashboards can be customized to display various environmental parameters such as temperature, humidity, and wind speed, providing a comprehensive overview of the farm’s conditions at any given time. Key features of Grafana include:

- Customizable Dashboards: Grafana allows users to build and customize dashboards to fit specific monitoring needs. Users can choose from a wide range of visualization options, including graphs, heatmaps, and gauges, to represent the data in the most meaningful way.
- Real-time Monitoring: Grafana supports real-time data updates, ensuring that the information displayed on the dashboards is current. This feature is critical for making timely decisions based on the latest environmental conditions.
- Alerting System: Grafana includes an alerting mechanism that can be configured to notify users when certain thresholds are reached, such as when temperature or humidity levels exceed predefined limits. This proactive approach helps in maintaining optimal conditions within the farm.
- Data Querying: Through its intuitive interface, Grafana allows users to perform complex queries on the data stored in InfluxDB, enabling detailed analysis and the extraction of actionable insights.

The combination of ChirpStack, Node-RED, InfluxDB, and Grafana creates a powerful workflow that support the SMS, as illustrated in Figure 2.5. ChirpStack processes and routes data via MQTT to Node-RED, which then directs it to the InfluxDB database for storage. Finally, Grafana visualizes the stored data, providing users with interactive and customizable dashboards for real-time monitoring and analysis. This integrated approach ensures that the system is not only efficient in data storage but also effective in presenting the data in an accessible and actionable format, supporting informed decision-making and optimized farm management.

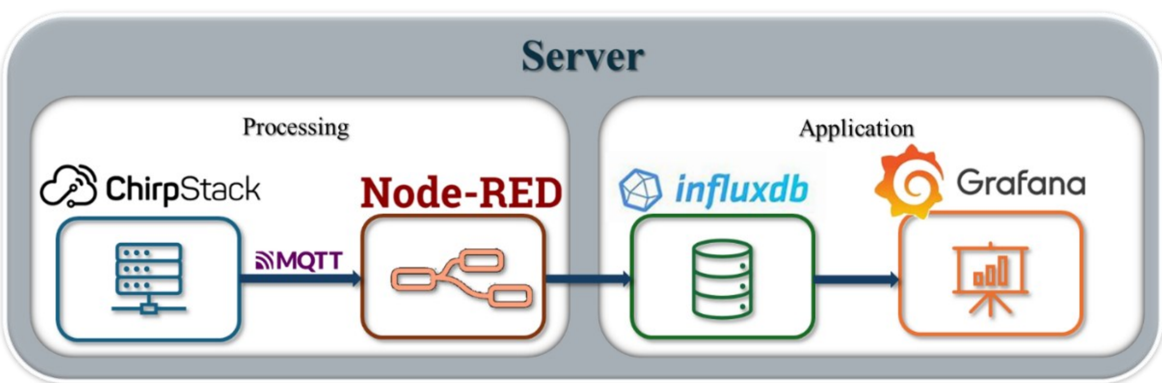

Figure 2.5 Overview of the Processing and Application Layers in the Smart Monitoring System

2.3 NeuralProphet Framework

NeuralProphet represents an advanced evolution of the Prophet algorithm, originally developed by Facebook, and is specifically tailored for time series forecasting (Triebe et al., 2021). This model integrates the strengths of statistical algorithms, such as ARIMA and GARCH, which are known for their mathematical explainability, with the powerful capabilities of neural network models that have proven highly effective but are inherently more complex. This integration facilitates a balance between efficiency and the retention of interpretability, essential for practical applications where understanding the model's predictions is crucial.

A pivotal aspect of the NeuralProphet model is its modular composability. This design allows the model to be assembled from distinct modules, each contributing an additive component to the overall forecast. These components can often be configured to scale with the trend, introducing a multiplicative effect that enriches the model's flexibility and responsiveness to varying data dynamics. NeuralProphet is designed to forecast multiple future time steps simultaneously, defined by the horizon h, which specifies the number of future steps to predict at once. This capability allows the model to efficiently generate predictions over the intended forecast horizon, adapting dynamically to changes in input data through the relationship specified in Eq. *2.1*:

$$\begin{aligned}\hat{y}_{t+k-1} = {} & T(t+k-1) + S(t+k-1) + E(t+k-1) \\ & + F(t+k-1) + A(t+k-1) + L(t+k-1)\end{aligned} \qquad \forall\, k \in [1, h] \qquad \text{Eq. 2.1}$$

where:

- $\hat{y}_{t+k-1}$: is the predicted value
- T: Trend at time t
- S: Seasonal effects at time t
- E: Event Holiday at time t
- F: Regression effects at time t for future-known exogenous variables
- A: Auto-Regression effects at time t based on past observations
- L: Regression effects at time t for lagged observations of exogenous variables

In the context of time series analysis, the "trend" represents the long-term progression of the data, showing an overall directional movement over time. It is a crucial component in understanding underlying patterns and making future predictions. In NeuralProphet, this term is referred to as "change points" and delineate periods where the trend exhibits noticeable shifts, essential for accurate modeling and forecasting. The model utilizes a piecewise linear approach to capture the trend component, assuming that the trend between identified change points follows a linear pattern. This methodology is analogous to performing segmented linear regressions between pairs of change points, allowing for flexibility and adaptation to sudden changes in the trend's direction. The trend at any given time t is mathematically represented in Eq. 2.2:

$$T(t) = (\delta_0 + \delta_1 + \cdots + \delta_k) \cdot t + (\rho_0 + \rho_1 + \cdots + \rho_k) \quad \text{Eq. 2.2}$$

Where:

- δ_0 denotes the slope or growth rate change at each change point
- ρ_i represents the intercept adjustments
- k is indexed to the last change point occurring before time t.

The framework begins model training by initially placing change points uniformly across the time series. These points, however, are adjustable during the training phase to optimize model fit and enhance performance. The algorithm employs L1 regularization to streamline the model by minimizing the number of change points, thus averting overfitting and improving interpretability.

The seasonality component of a time series is characterized by recurrent fluctuations in the signal, typically linked to the natural periodicity of the underlying data. To accurately model this aspect, the NeuralProphet framework posits that the target series behaves as a periodic and continuous function, which can be effectively represented using a Fourier series approach (Harvey & Shephard, 1993). This mathematical formulation involves defining a number of Fourier terms for each type of seasonality, capturing the essence of both fixed and non-integer periodic patterns. Specifically, the seasonality for a given periodicity p is expressed through the Eq. 2.3.

$$S_p(t) = \sum_{j=1}^{k} \left[a_j \cos\left(\frac{2\pi j t}{p}\right) + b_j \sin\left(\frac{2\pi j t}{p}\right) \right] \quad \text{Eq. 2.3}$$

Here, k represents the number of Fourier terms designated for the seasonality, and the terms are structured as pairs of sine and cosine functions. This setup not only accommodates multiple seasonalities within a dataset but also manages seasonalities with non-integer periodicities such as a yearly seasonality represented in daily data ($p = 365.25$) or weekly data ($p = 52.18$). In scenarios with multiple seasonal patterns, different values for k can be assigned to each distinct periodicity.

Fourier terms are particularly useful for modeling seasonality because they yield smooth functions that are straightforward to interpret and robust when fitted to data. However, they primarily capture deterministic seasonal shapes, which are presumed to be constant over time. While employing a greater number of Fourier terms enables the model to accommodate more complex seasonal patterns, excessive flexibility might lead to overfitting or induce spurious patterns between observations. Each Fourier term corresponds to a frequency proportional to j/p, and is modeled through a weighted combination of sine and cosine transformations, with each seasonality associated with $2k$ coefficients. If the trend component is specified as "multiplicative," the seasonality can interact with the trend component, enhancing the model's ability to adapt to situations where seasonal effects are proportionally related to the trend level (Eq. 2.4).

$$S_p^*(t) = \begin{cases} T(t) \cdot S_p(t) & \text{Seasonality mode is "multiplicative"} \\ S_p(t) & \text{otherwise} \end{cases} \qquad \text{Eq. 2.4}$$

Autoregressive (AR) models are fundamental in time series analysis, where the future value of a variable is presumed to be a function of its past values. In NeuralProphet, this concept is extended to accommodate multi-step forecasting directly, leveraging the structure of AR-Net, a variant that enhances traditional AR models to produce forecasts for multiple future steps simultaneously (Triebe et al., 2019). In traditional AR models, the relationship for a one-step-ahead forecast is expressed in Eq. 2.5 where c represents the intercept, θ_i are the coefficients affecting the influence of past values, and e_t is the error term. However, to generate forecasts for multiple future steps ($h > 1$), traditional AR models would require fitting separate models for each step, which is computationally inefficient and complex.

$$y_t = c + \sum_{i=1}^{i=p} \theta_i \cdot y_{t-i} + e_t \qquad \text{Eq. 2.5}$$

In contrast, the AR module in NeuralProphet, based on AR-Net, streamlines this process by using a single model to predict all h future values concurrently. This is achieved by taking the last p

observations as inputs (referred to as lags) and producing h outputs that correspond to the AR effects for each forecast step up to h. The formulation is given by Eq. 2.6 where each $A^t(t+k)$ represents the predicted AR effect for the forecast at time $t+k$, predicted at the initial time t using data observed up to $t-1$.

$$A^t(t), A^t(t+1), \cdots, A^t(t+h-1) = AR-NET\left(y_{t-1}, y_{t-2}, \cdots, y_{t-p}\right) \qquad \text{Eq. 2.6}$$

A key advantage of this approach is that it generates multiple forecasts for a given point in time, each based on forecasts made at different points in the past, using different sets of available data. The result is a comprehensive understanding of possible future scenarios that reflect changes in data over time. Choosing an appropriate value for p is critical, as it should reflect the length of the relevant historical context. By integrating AR module, NeuralProphet provides a flexible, efficient approach to time series forecasting, capable of handling complex patterns and dependencies with enhanced accuracy. This feature is particularly beneficial in scenarios where understanding and forecasting based on historical trends is critical, such as agricultural yield forecasting, financial market analysis, or energy demand estimation.

Lagged regressors are used to correlate other observed variables to our target time series. They differ from future regressors in that their future values are not known at the time of forecasting. At any given forecasting time t, the model can only access past values of these regressors up to and including $t-1$. This constraint is mathematically represented in Eq. 2.7 where X is the set of covariates with dimensions $\mathbb{R}^{T\times n_l}$, and p is the number of past observations considered (lags).

$$L(t) = \sum_{x\in\mathbb{X}} L_x\left(x_{t-1}, x_{t-2}, \cdots, x_{t-p}\right) \qquad \text{Eq. 2.7}$$

For each covariate x in the dataset, a separate lagged regressor module is constructed, allowing the model to assess and attribute the individual impacts of each covariate on the forecast outcomes. This modular approach ensures that the effects of each covariate are independently evaluated, providing a clear picture of how external factors influence the target variable. Functionally, each lagged regressor module mirrors the structure of the autoregressive (AR) module, but with a key difference in the inputs used. Instead of the target series itself, the inputs to each lagged regressor module consist of the last p observations of the covariate x. The outputs are of identical form, each module producing h additive components to the overall forecasts Eq. 2.8.

$$L_x^t(t), L_x^t(t+1), \cdots, L_x^t(t+h-1) = AR - NET\left(x_{t-1}, x_{t-2}, \cdots, x_{t-p}\right) \quad \text{Eq. 2.8}$$

Future Regressors are variables that are known in the future. Given the set of future regressors as $\mathbb{F} \in \mathbb{R}^{T \times n_l}$, where nf is the number of regressors, the effect from all future regressors at time step t can be denoted by $F(t)$ as in Eq. 2.9, where d_f stands for the coefficient of the model for future regressor $f \in \mathbb{V}$. By default, future regressors have an additive effect, which can be configured to be multiplicative instead.

$$F(t) = \sum_{f \in \mathbb{F}} F_f^{\star}(t) \quad \text{Eq. 2.9}$$

Where:

$$F_f(t) = d_f f(t)$$

$$F_f^{\star}(t) = \begin{cases} T(t) \cdot F_f(t) & \text{If } f \text{ is “multiplicative”} \\ F_f(t) & \text{otherwise} \end{cases}$$

In the context of modeling special events and holidays, NeuralProphet provides a dedicated module designed to account for the impact of such occurrences on time series data. This module permits the incorporation of known holidays or special events that may result in deviations from typical patterns, thereby enhancing forecast accuracy during these periods. Nevertheless, despite the availability of this methodology within the NeuralProphet framework, this research does not consider its application. Consequently, the holiday module is not pertinent to the context of this study and will not be analyzed further in this documentation.

2.3.1 Processing

In time series analysis, dealing with missing data is a critical aspect of preprocessing, particularly when using models like NeuralProphet. The framework offers specific tools to address gaps in the data, especially when working with modules like auto-regression or lagged regressors, where missing values can have a significant impact on the analysis. When non-lagged input variables are used, handling missing data is relatively straightforward corresponding timesteps with missing values can simply be dropped, resulting in the loss of one data sample per missing entry. However, the challenge intensifies with auto-regression or lagged regressor modules, as a single missing data point can lead to the loss of multiple data samples due to the missing forecast targets and lag values.

To mitigate this issue, NeuralProphet approximates missing values using a bi-directional linear interpolation technique. This method utilizes the last known value before the gap and the first known value after the gap as anchor points for the interpolation. The interpolation is applied for up to five missing values in each direction. If the gap exceeds 10 consecutive missing values, they remain as NaNs after this step. The number of missing values to be interpolated is adjustable according to the user's needs.

For remaining missing values that are not addressed by linear interpolation, NeuralProphet applies a centered rolling average to fill the gaps. This rolling average is computed over a window size of 30 data points and is used to fill up to 20 consecutive missing values. Again, the number of missing values that can be filled using this method is user-configurable. However, if more than 30 consecutive data points are missing, the imputation algorithm aborts the interpolation and instead drops all the missing data points.

Equally important is the normalization of time series modeling, particularly when developing predictive models that incorporate diverse input variables with varying scales and magnitudes. In the context of environmental monitoring, for instance, variables such as temperature and relative humidity not only differ in magnitude but also exhibit inverse proportional relationships. These disparities can significantly affect the performance and accuracy of predictive models if not appropriately addressed. NeuralProphet provides a flexible tool for normalizing time series data, allowing the user to specify the type of normalization to be applied. This capability is essential for ensuring that the model processes each input variable on a comparable scale, thereby enhancing the model's ability to learn from the data effectively. Table 2-2 outlines in detail the available normalization options in NeuralProphet.

Table 2-2 Available data normalization options available in NeuralProphet

Option name	Description
'auto'	'minmax' if binary, else 'soft'

'off'	bypasses data normalization
'minmax'	scales the minimum value to 0.0 and the maximum value to 1.0
'standardize'	zero-centers and divides by the standard deviation
'soft'	scales the minimum value to 0.0 and the 95th quantile to 1.0
'soft1'	scales the minimum value to 0.1 and the 90th quantile to 0.9

2.3.2 Training

The training process in NeuralProphet is designed to be both flexible and robust, leveraging modern techniques to optimize model performance. At the core of this process is the use of mini-batch Stochastic Gradient Descent (SGD), a widely adopted fitting procedure in the field of Deep Learning. Mini-batch SGD is particularly effective because it can efficiently handle large datasets and accommodate the complexity of advanced model components. One of the key advantages of using SGD in NeuralProphet is its compatibility with a variety of model components. Any component that can be trained using SGD can be seamlessly integrated as a module within the framework.

The default loss function employed during training is the Huber loss, also known as smooth L1-loss (Eq. 2.10). The Huber loss is a hybrid between mean squared error (MSE) and mean absolute error (MAE), which offers several advantages in training neural networks. Specifically, for prediction errors below a certain threshold β, the Huber loss behaves like MSE, providing the benefits of MSE's smooth gradient for small errors (Girshick, 2015). However, for larger errors, the loss function transitions to MAE, which is less sensitive to outliers. This characteristic helps to mitigate the problem of exploding gradients, which can occur when the model encounters extreme data points, thereby enhancing the stability and reliability of the training process.

$$L_{huber}(y, \hat{y}) = \begin{cases} \frac{1}{2\beta}(y_i - \hat{y}_i)^2 & \text{For } |y_i - \hat{y}_i| < \beta \\ |y_i - \hat{y}_i| - \frac{\beta}{2} & \text{otherwise} \end{cases} \qquad \text{Eq. 2.10}$$

The effectiveness of the learning process undertaken when training the NeuralProphet model is contingent upon the adjustment of several key parameters. These include the learning rate, batch size and training epochs. The specific characteristics of the dataset and the requirements of the model inform the dynamic adjustment of these parameters.

The learning rate η is a critical hyperparameter that controls how much the model's parameters are adjusted with respect to the loss gradient during each update. NeuralProphet employs a learning rate range test to determine the optimal learning rate. This test is performed over a series of iterations, where the learning rate begins at a very low value (e.g., $\eta = 1e - 7$) and increases exponentially up to a high value (e.g., $\eta = 1e + 2$). The number of iterations for this test is calculated as shown in Eq. 2.11, where T is the length of the dataset.

$$100 + \log_{10}(10 + T) * 50 \qquad \text{Eq. 2.11}$$

Throughout the test, the steepest learning rate is identified by selecting the learning rate that maximizes the negative gradient of the losses, indicating the most rapid improvement in model performance. To enhance reliability, this test is conducted three times, and the final selected learning rate is the geometric mean of these three runs, calculated as:

$$\eta^\star = \frac{1}{3}\left(\log_{10}(\eta_1) + \log_{10}(\eta_2) + \log_{10}(\eta_3)\right) \qquad \text{Eq. 2.12}$$

$$\eta = 10^{\eta^\star}$$

The batch size B is another important parameter that determines the number of data samples used in each iteration of the model training process. If the user does not specify a batch size, NeuralProphet uses a heuristic to automatically determine an appropriate value based on the length T of the dataset. The batch size is calculated using Eq. 2.13.

$$B^\star = 2^{2+[\log_{10}(T)]} \qquad \text{Eq. 2.13}$$

The final batch size is then constrained within a range, ensuring that it is neither too small (which could make training inefficient) nor too large (which could lead to memory issues), as shown in Eq. 2.14

$$B = min\left(T, max\left(16, min(256, B^{\star})\right)\right) \quad \text{Eq. 2.14}$$

The number of training epochs N_{epoch} refers to the total number of times the entire dataset is passed through the model during training. Like the batch size, the number of epochs can be automatically determined if not specified by the user. NeuralProphet calculates a suggested number of epochs using the heuristic in Eq. 2.15:

$$N_{epoch}^{\star} = \frac{1000 \cdot 2^{\frac{5}{2}\log_{10}(T)}}{T} \quad \text{Eq. 2.15}$$

This value is then adjusted to ensure it falls within a reasonable range, as shown in Eq. *2.16*

$$N_{epoch} = min\left(500, max\left(50, [N_{epoch}^{\star}]\right)\right) \quad \text{Eq. 2.16}$$

2.3.3 Validation

In the NeuralProphet model, evaluating the performance of the predictive model involves splitting the available data into training and validation sets, calculating key metrics that indicate the model's performance, and using these metrics to fine-tune the model. The $split\_df$ function in NeuralProphet is designed to facilitate the creation of a validation dataset ($validation\_df$) from the original dataset. This validation dataset is used during the training process to evaluate the model's performance on unseen data.

The $split\_df$ function divides the dataset into two parts: a training set that the model uses to learn the patterns in the data, and a validation set that is used to test the model's accuracy. By using the $validation\_df$ during the fitting process, NeuralProphet can continuously monitor the model's performance on the validation data. This approach allows for the early detection of issues such as overfitting and underfitting, and it provides valuable insights into how the model is likely to perform in real-world scenarios.

After the model fitting process, NeuralProphet automatically calculates and stores several key performance metrics in the metrics parameter, which is an output of the fitting process, when $validation\_df$ is used. In this study, forecasting performance was evaluated using Mean Absolute

Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), which together provide complementary information on absolute and relative prediction errors.

- **Mean Absolute Error (MAE):** measures the average magnitude of the errors between the predicted values and the actual values. It is calculated as:

$$MAE = \frac{1}{T}\sum_{i=1}^{T}|y_i - \hat{y}_i| \qquad \text{Eq. 2.17}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and T is the number of observations.

- **Root Mean Squared Error (RMSE):** provides a measure of the standard deviation of the residuals (prediction errors). It is calculated as:

$$RMSE = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(y_i - \hat{y}_i)^2} \qquad \text{Eq. 2.18}$$

RMSE penalizes larger errors more strongly than MAE, making it particularly useful for identifying models that may exhibit occasional but substantial deviations from observed values.

- **Mean Absolute Percentage Error (MAPE):** expresses prediction errors as a percentage of the observed values, providing a scale-independent measure of forecasting accuracy. It is defined as:

$$MAPE = \frac{100}{T}\sum_{i}^{T}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \qquad \text{Eq. 2.19}$$

MAPE is particularly useful for comparing forecasting performance across different thermal indices and sensor locations, as it highlights relative deviations that may not be fully captured by absolute error metrics alone. In this study, the inclusion of MAPE supports a more comprehensive evaluation of model behavior under varying thermal conditions, including periods characterized by elevated index values.

2.4 Weather Forecast Data as Exogenous Variables

In time series forecasting, exogenous variables represent external drivers that influence the target variable but are not contained within its historical observations. Their inclusion is particularly relevant in environmental applications, where outdoor weather conditions exert a strong influence on indoor microclimatic dynamics. Incorporating such variables enables forecasting models to account explicitly for external forcing mechanisms that cannot be inferred from sensor data alone.

In this study, short-term weather forecasts were obtained from the Open-Meteo service (Open-Meteo, n.d.), which provides high-resolution meteorological data (1–11 km) derived from national weather services. These forecasts were integrated as exogenous inputs within the NeuralProphet framework to represent anticipated outdoor conditions influencing the barn microclimate. The inclusion of weather-based exogenous variables enables the model to combine observed indoor measurements with forecasted external conditions, supporting short-term predictions under dynamically evolving climatic scenarios, particularly during periods of rapid ambient change that may not be fully captured by autoregressive components alone. The use of short-term weather forecasts as exogenous predictors in agricultural and livestock applications has been previously reported in the literature, including earlier work by the author (Perez Garcia et al., 2024).

Weather APIs such as Open-Meteo estimate site-specific meteorological variables through spatial interpolation based on data from surrounding meteorological stations. The accuracy of these interpolated values depends on the distance between the location of interest and the contributing stations. To address potential uncertainties associated with this process, external field sampling campaigns were conducted for both application scenarios. On-site measurements of environmental variables were collected and subsequently compared with the corresponding Open-Meteo data to verify their suitability as exogenous inputs for the forecasting models.

2.5 Environmental Indicators

The thermal environment within farm buildings is a complex system, encompassing key elements such as air temperature, relative humidity, and wind speed. The interaction between these environmental factors and the animals housed, is a complex and intricate process. These interactions have a significant impact on the physiological responses of the animals. Thermal stress can impact production metrics, including milk yield, body weight gain, and the conception success rate. Furthermore, it can also affect the immune system and alter animal behavior, thereby complicating the assessment of overall animal welfare (Fournel et al., 2017).

In light of the challenges associated with directly measuring physiological indicators, immune function, and behavior, the development of thermal indices that link environmental conditions to observable animal responses has emerged as a pivotal area of research. These indices offer a practical method for assessing animal comfort by correlating readily measurable environmental variables with the more complex physiological responses of the animals.

One of the foundational indices in this field is the Temperature-Humidity Index (THI), originally developed by Thom, (1959) to quantify human discomfort during summer. This index was later adapted by Bianca, (1962) to estimate heat stress in cattle. THI is calculated using Eq. *2.20* that considers both temperature and relative humidity, providing a quantitative measure of the environmental impact on animal welfare.

$$\mathrm{THI} = (1.8 * T_a + 32) - [(0.55 - 0.0055 * RH) * (1.8 * T_a - 26.8)] \qquad \text{Eq. 2.20}$$

Where:

- T_a is the indoor temperature in °C
- RH is the indoor relative humidity in %

However, traditional THI models do not account for the effects of wind speed or ventilated air movement, which are crucial factors in mitigating heat stress. Higher air movement speeds can enhance convection cooling, thereby reducing the impact of THI values and improving the thermal comfort and productivity of dairy cows.

In addition to the THI, several other thermal indices have been developed to provide a more comprehensive assessment of heat stress in livestock. Notable among these are the Black Globe Humidity Index (BGHI), the Environmental Stress Index (ESI), and the Heat Load Index (HLI). Each of these indices incorporates different combinations of environmental parameters, providing different perspectives on the impact of the thermal environment on cattle. However, studies have shown that although indices such as the THI and BGHI are commonly used, they may have poor correlation with certain physiological responses such as rectal temperature and respiration rate (Silva et al., 2007).

To overcome these limitations, the Equivalent Temperature Index (ETI) was developed by da Costa Baêta et al., (1987). The ETI (Eq. *2.21*) incorporates wind speed, making it a more robust indicator of heat stress in dairy cattle, particularly in indoor housing systems where natural ventilation is limited. The inclusion of wind speed in the ETI is particularly important because it directly affects the animal's ability to dissipate heat, thereby affecting both animal welfare and productivity.

$$ETI = 27.88 - 0.456t_a + 0.010754{t_a}^2 - 0.4905RH + 0.00088RH^2 + 1.1507U - 0.126447U^2 + 0.019876t_aRH - 0.046313t_aU \quad \text{Eq. 2.21}$$

Where:

- T_a is the indoor temperature in °C
- RH is the indoor relative humidity in %
- U is the indoor wind speed in m/s

In this study, different thermal indices were applied and adapted to the specific conditions of each farm. For the Piazzi family farm, the Temperature-Humidity Index (THI) was employed due to the absence of wind speed measurements. On the other hand, at the Montagnini family farm, where wind speed measurements are available, the ETI) is used. By incorporating wind speed, the ETI provides a more comprehensive assessment of heat stress, taking into account the animals' ability to dissipate heat by convection. This approach ensures that the chosen thermal index is appropriately matched to the environmental data available on each farm, optimizing the assessment of animal welfare and productivity.

2.5.1 Multi-Zonal Modeling Approach

The analysis of the spatial distribution of climatic variables within buildings is a key application of Computational Fluid Dynamics (CFD). CFD is a method that utilizes computer simulations to control dimensions, areas, and volumes within a defined space. The process involves dividing the area into smaller segments, known as cells, through a technique called meshing (Sugiono et al., 2016). The accuracy of the simulation results is closely tied to the size of these cells; larger cells create a more homogeneous mesh that may not accurately reflect real conditions. In contrast, smaller cells can produce more accurate results but consume significantly more computing power and extend computation times, making CFD impractical for near-real-time applications.

To address these challenges, a common practice in environmental monitoring and control systems, particularly in farm buildings, has been to rely on averages of environmental parameters over the entire indoor area. This approach has been widely adopted due to two primary issues: the inherent difficulties in obtaining accurate environmental measurements within buildings, and traditional environmental control systems that typically rely on a single setpoint.

Farm buildings, in particular, present additional challenges for indoor environmental sensing. These challenges include the difficulty of installing sensors in areas occupied by animals, the polluted farm environment, and the complexity of deploying wired infrastructure due to the construction characteristics of buildings. The development of IoT devices has significantly mitigated these problems by enabling the design of Smart Monitoring systems based on sensor nodes, as investigated in this research. These systems allow the monitoring and control of key environmental variables at different representative points within the buildings, facilitating the creation of zones of interest based on the clustering of measurements. Given the nature of this technology, it is possible to obtain measurements over short time intervals, approaching real-time monitoring.

The development of a multi-zone modeling approach could represent a paradigm shift in environmental control strategies by enabling the implementation of more advanced control techniques. In the specific application contexts addressed in this research, it becomes feasible to develop predictive models that estimate the behavior of thermal indices in various predefined zones of interest. This approach requires the use of models capable of producing multiple outputs or the development of separate models for each zone.

2.5.2 Interpolating Climatic Variables in Livestock Buildings

In livestock buildings, two main scenarios frequently arise where interpolation techniques can be invaluable for estimating climatic variables in unmeasured areas. The initial scenario pertains to areas of interest for the animals, particularly those where they are prone to congregate during their daily routines. In the context of dairy farms, these zones may correspond to the proximity of milking robots, feeding and drinking areas, and other locations where social interactions among animals may be concentrated. However, it should be noted that this pattern is not fixed, as animals may gather in other areas based on various factors, such as social relationships or environmental conditions. This variability presents a challenge for ventilation control systems, which typically rely on fixed-point measurements of internal conditions. Therefore, there may be occasions when the measured point indicates favorable conditions, while specific zones of interest may still experience discomfort due to microclimatic variations.

The second scenario relates to the vertical stratification of temperatures within these frequently tall livestock buildings. While the height at which animals are present is of primary importance, discomfort or extreme values at higher elevations can directly impact the overall indoor environment, influencing airflow and thermal distribution. To effectively address these challenges, a strategic approach to IoT sensor placement is required, along with methods to estimate conditions in unmeasured zones.

The deployment of IoT devices for multi-point measurement within the barn provides a practical solution to this problem. However, the locations of these measurement points must be selected with care to ensure that they do not impede the animals' free movement throughout the facility. Consequently, while a multi-point measurement strategy can provide valuable data on conditions at key areas, the utility of this approach can be further enhanced by the application of interpolation algorithms that estimate the environmental variables across the entire interior area. This approach not only provides a more comprehensive understanding of the barn's microclimate but also facilitates proactive decision-making with regard to environmental control.

A range of interpolation techniques is available; however, considering the characteristics of the environmental variables and the specific context of this study, the application of the Inverse Distance Weighting (IDW) algorithm was identified as the most suitable approach. IDW algorithm provides a means of estimating values at unmeasured locations. This is achieved by weighting the influence of each measured point according to its proximity to the estimation location. The degree of influence exerted by a given point on the estimated value is directly proportional to its proximity. The simplicity of implementation, minimal computational resources required, and effective handling of climatic data variability within confined spaces, such as livestock barns, make IDW an advantageous choice in this context. Additionally, it allows for flexibility in managing irregular sensor placement and can accommodate any shape or distribution of points, which makes it particularly useful for estimating conditions in unevenly monitored spaces.

CHAPTER 3. MODEL DEVELOPMENT AND VALIDATION

This chapter delineates the processes and methodologies associated with the development and validation of predictive models for environmental monitoring within dairy barns. By utilizing the case studies of two commercial dairy farms, this chapter offers a comprehensive description of the procedures involved in the collection, preparation, training, and evaluation of models. The use of sensor networks and advanced machine learning techniques enabled the development of models capable of forecasting the Temperature-Humidity Index (THI) and Equivalent Temperature Index (ETI) based on time series data of environmental parameters. These indices are of critical importance for the assessment and management of heat stress, which has a direct impact on animal welfare and productivity.

3.1 Case Study I: Piazzi Farm

The first on-farm data set was based on long-term measurements carried out at a commercial dairy building, located in Budrio, a municipality in the metropolitan city of Bologna, Emilia-Romagna, Italy (coordinates: 44.559075 N, 11.519375 E, 25 m above sea level). The dairy building is 50.0 m long and 27.00 m wide. The height of the sheet metal roof varies from 6.0 m at the sides to 8.5 m at the gable peak (Figure 3.1).

Figure 3.1 Three-dimensional view of the farm building

The internal area of the barn host about 70 Holstein-Friesian lactating and 15 dried cows and is divided into three main areas: resting, feeding, and milking (Figure 3.2). The milking process is carried out through an AMS. The resting area has a partially slatted floor and comprises 78 cubicles with sawdust bedding. These cubicles are arranged in two blocks of head-to-head rows in the central part of the resting area, with an-other row running along its entire length. On the opposite side, the feeding area spans the entire width of the building and is on the north-east side. Lastly, the AMS station houses a robotic milking system “Astronaut A3 Next” by Lely. This system ensures that each cow receives a specific number of daily visits based on its productivity and the expected optimal milk yield per visit. The number of daily visits per cow ranges between a minimum of 2 and a maximum of 4, depending on production metrics.

Mechanical ventilation in the barn is provided by three high-volume, low-speed fans with five horizontal blades each. These fans are activated based on the THI, calculated using a temperature and humidity sensor installed centrally within the barn. This system ensures that the barn’s environmental conditions are controlled to mitigate heat stress and improve animal welfare (Perez Garcia et al., 2024).

3.1.1 Data Collection

The temperature and relative humidity within the barn were monitored continuously using RH/T sensors (TE Connectivity Inc., 2014), which offer high precision and reliability for environmental data collection. The sensors exhibit an accuracy of ±0.3°C for temperature measurements within the range of 5°C to 60°C and ±3% for relative humidity measurements across the full range of 0%

to 100%. The sensors were programmed to log data at 10-minute intervals, thereby providing a high-resolution dataset that captured the subtle indoor environmental conditions within the facility.

The resulting dataset cover one year period, including the four distinct seasons, thereby ensuring that the full range of environmental conditions experienced within the barn were captured. This approach permitted the identification of both high-temperature, low-humidity periods characteristic of summer and low-temperature, high-humidity periods typical of winter. The input of this seasonal variation is vital for ensuring that the resulting predictive models can accommodate both the extremes and the subtler fluctuations in temperature and humidity. This renders the models robust and capable of providing effective environmental management strategies year-round.

The placement of the sensor nodes was designed with the objective of monitoring the main environmental parameters at different heights within the barn. The nodes were organized into four distinct clusters based on their height, at 1.0, 2.0, 2.5, and 3.5 meters, respectively. The vertical distribution of sensors permitted a comprehensive evaluation of the microclimatic conditions across disparate zones of the barn, which could fluctuate due to air circulation patterns and structural influences. For example, the lower zones are more susceptible to humidity accumulation, while the higher zones are prone to elevated temperatures caused by rising heat. By monitoring these varying heights, the collected data provided a comprehensive understanding of the barn's internal environment. This information is critical not only for analyzing how temperature and humidity impact animal welfare but also for optimizing operational efficiency within the facility (Figure 3.2).

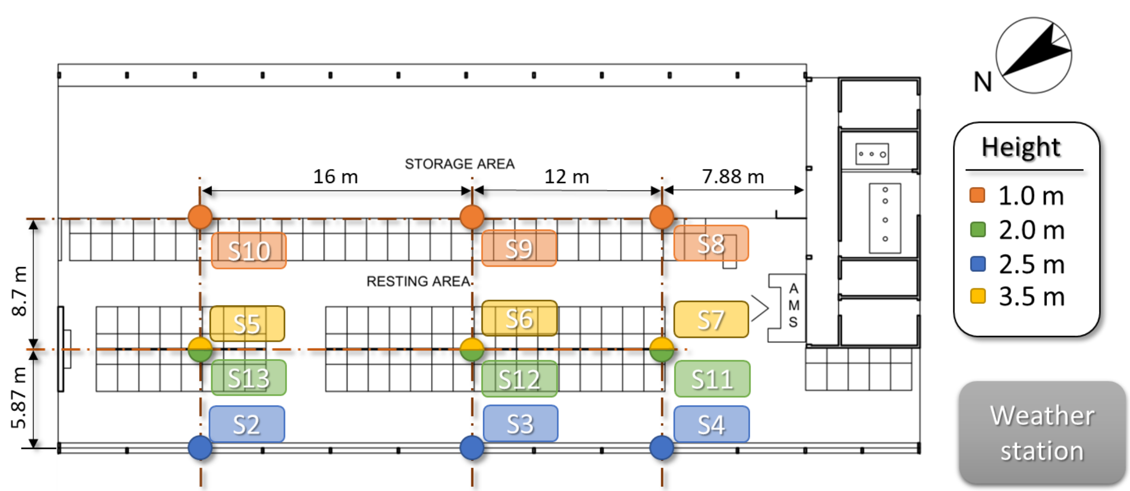

Figure 3.2 Farm building top view and sensor location

3.1.2 Data Preprocessing

Data preprocessing is a crucial step in preparing the dataset for time series forecasting, ensuring that the data is clean, consistent and suitable for predictive modelling. In this case study, the raw data was resampled to a 30-minute interval to improve computational efficiency, make the predictive models easier to use and to fill in missing values over short time periods. This resampling approach allowed the retention of key trends and patterns within the dataset while reducing the overall volume of data, making the subsequent analysis more manageable and ensuring a smooth modelling process.

Following the resampling process, an outlier detection procedure was implemented using a two-stage cascading strategy. First, the values of the time series were clipped based on the measuring range of the instruments. This step ensured that data points outside the physical limits of the sensor's capabilities, which could be due to hardware faults or erroneous recordings, were removed. After this initial cleaning, the Mean Absolute Deviation (MAD) algorithm was applied to identify further outliers. MAD was chosen for its robustness and ability to detect deviations from the central tendency of the data. Its main advantage over standard deviation-based methods is that MAD is less sensitive to extreme values. These anomalies are often caused by the highly polluted environment within the farm, where dust particles and other contaminants can interfere with sensor operation.

Figure 3.3 shows a schematic representation of the outlier detection process, focusing specifically on temperature measurements. This illustration highlights the various steps involved in identifying and handling anomalous values to ensure that the data set is sufficiently clean for accurate predictive modelling.

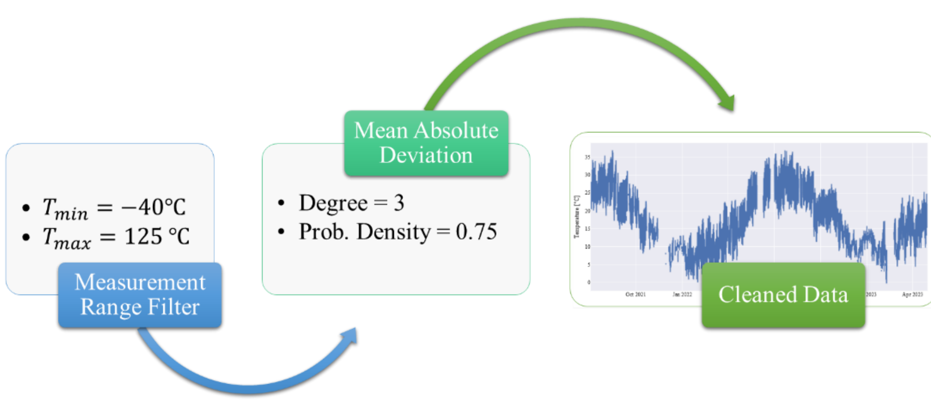

Figure 3.3 Data cleaning workflow for indoor temperature

It is important to note that the procedures for imputing missing values were primarily applied to large data gaps (more than 24 consecutive missing points), while smaller gaps were handled directly by the regression models. For substantial gaps, a prediction model was fitted to the available data, supplemented by external weather measurements from a nearby meteorological station. This allowed a more reliable estimation of missing data and ensured that the continuity of the time series was maintained without introducing artificial trends that could distort the prediction models.

3.1.3 Model Training

In this context, the target variable is the THI, which is calculated to assess heat stress within the barn environment. After dealing with missing values in accordance with the data pre-processing steps previously outlined, which were necessary to maintain the continuity and integrity of the time series data, the THI was calculated using Eq. *2.20* based on temperature and relative humidity readings from the sensor nodes. Given the spatial variability within the barn, captured by twelve different sensor nodes.

The observed similarity between the twelve THI measurements, shown in Figure *3.4*, supports the decision to develop twelve prediction models using a unified set of parameters. The following parameters were selected based on a hyperparameter tuning process to ensure robust and accurate predictions across all sensor nodes. The parameter $n\_forecasts = 48$ allows the model to generate forecasts for up to 24 hours with 30 minute intervals, providing actionable insights for daily operations. A parameter $n\_lags = 6$ was chosen to capture the influence of past measurements over a three-hour period, which is the typical time frame over which environmental conditions can significantly affect livestock.

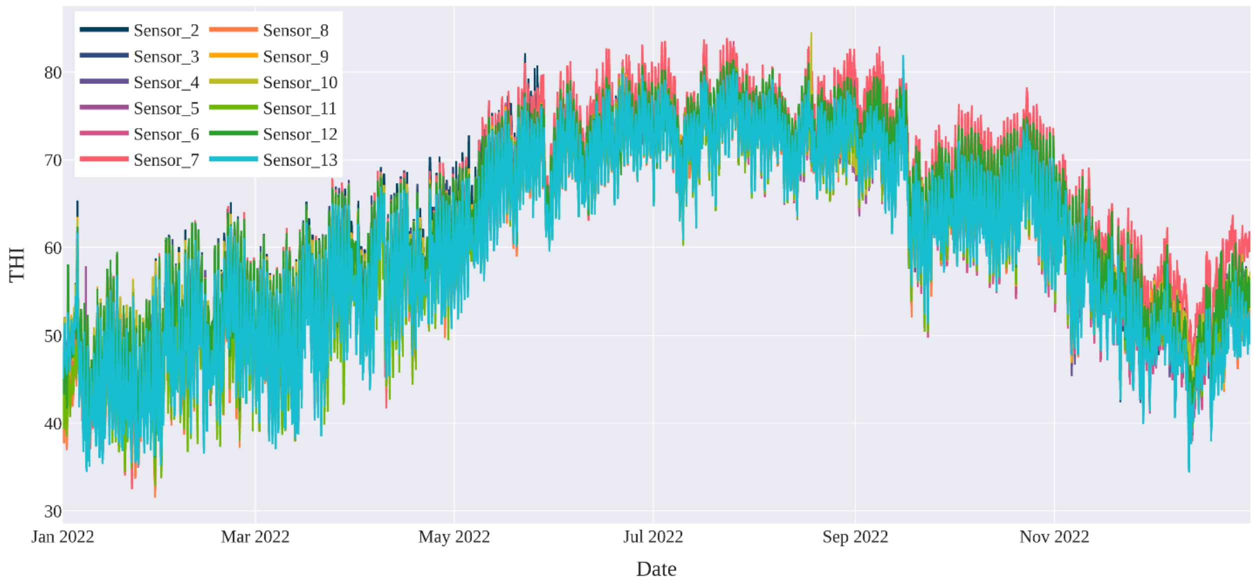

Figure 3.4 THI Measurements by Sensor: Training Dataset for Developing Predictive Models

The model captures recurring patterns with $yearly\_seasonality = True$ and $daily\_seasonality = True$, addressing both annual and daily variations that are critical to barn conditions. $weekly\_seasonality = False$ was used as weekly cycles were deemed less relevant based on the data analysis. With $seasonality\_mode = "additive"$, the model assumes cumulative effects in the environmental data, which is appropriate for the nature of THI variations. $normalize = "auto"$ facilitates consistent input scaling by automatically adapting to the needs of each sensor dataset.

For training, a batch size of 64 strikes was set as balance between computational efficiency and training stability, while setting $epochs = 100$ ensures sufficient training iterations for model convergence without overfitting. A $learning\_rate = 0.001$ was chosen for stable convergence, avoiding overly aggressive updates. The parameter $n\_changepoints = 10$ was included to account for potential abrupt shifts in THI trends due to environmental changes. To ensure robustness to outliers, $loss\_func = \text{"Huber"}$ balances sensitivity and efficiency. $Quantiles = [0.1, 0.5, 0.9]$ allow probabilistic forecasting and provide insight into THI variability under different conditions. Enabling $impute\_missing = True$ supports the model's handling of data gaps, maintaining the continuity of the time series. In addition, external regressors, temperature_2m and relative_humidity_2m, provide predicted environmental inputs, enhancing the model's ability to anticipate indoor THI variations.

3.1.4 Model Validation

In the model validation stage, the dataset was divided into two distinct sets: training and validation set. The ratio of the two sets was 80-20%, with 80% of the data allocated for training and 20% for validation. This approach guarantees that the model is trained on a significant portion of the data while a portion is reserved for evaluating its performance on previously unseen data. The NeuralProphet framework's built-in function for dataset splitting and model fitting was employed, whereby the training and validation data were automatically assigned and passed into the fit process for the model.

To assess the predictive accuracy of each model, several performance metrics were obtained, including the mean absolute error (MAE), root mean squared error (RMSE), and the mean absolute percentage error (MAPE), along with the loss functions available within the NeuralProphet framework. The formulations of MAE, RMSE, and MAPE, discussed in detail in Section 2.3.3, provide complementary measures for evaluating model performance in terms of both absolute and relative errors. The validation results for the twelve predictive models are summarized in Table 3-1.

Table 3-1 Validation Metrics for Predictive Models Across Sensors

Sensor Location	MAE	MAPE	RMSE	Loss	Sensor Location	MAE	MAPE	RMSE	Loss
Sensor_2	4.2327	0.0805	4.8082	0.0982	Sensor_8	4.1593	0.0458	4.6862	0.092
Sensor_3	6.9665	0.1187	7.7317	0.1521	Sensor_9	6.0102	0.16528	6.9359	0.1422
Sensor_4	7.777	0.1767	8.7547	0.1773	Sensor_10	3.9395	0.04822	4.6376	0.0956
Sensor_5	4.8042	0.1597	5.5027	0.1099	Sensor_11	3.4363	0.0540	4.0112	0.0867
Sensor_6	6.7153	0.1854	7.7791	0.1479	Sensor_12	6.5294	0.1869	7.1972	0.1385
Sensor_7	4.8912	0.0987	5.565	0.1075	Sensor_13	8.7301	0.1962	9.8113	0.2001

The MAE values range between 3.43 and 8.73 THI units, indicating that the models are generally capable of capturing the dominant temporal patterns of THI across the monitored sensor locations with limited average error. Similarly, RMSE values range from 4.01 to 9.81 THI units, with higher values associated with sensors exposed to greater environmental variability. This behavior reflects the sensitivity of RMSE to larger deviations and suggests that localized microclimatic fluctuations influence prediction accuracy at specific sensor positions.

The inclusion of the MAPE provides additional insight into the relative performance of the predictive models. MAPE values range from approximately 4.6% to 19.6%, indicating that while absolute errors remain within acceptable bounds, relative deviations increase at sensor locations characterized by higher THI variability. Sensors exhibiting lower MAE and RMSE values generally correspond to lower MAPE values, whereas sensors located in more dynamic microclimatic zones show elevated relative errors. This result highlights the complementary role of MAPE in identifying sensor-specific differences in relative forecasting performance that may not be fully captured by absolute error metrics alone.

The loss on test data, a direct output from NeuralProphet's loss function, further indicates model stability across the board, with values between 0.09 and 0.20. Lower loss values underscore the model's robustness in capturing the underlying patterns of THI, validating its effectiveness for real-time applications.

The summarized metrics indicate that the models are effectively suited for practical applications in monitoring and managing THI within the barn. The models offer a high degree of accuracy that consistently supports the generation of reliable insights across a range of sensor locations. As a result, they have the potential to become invaluable resources for maintaining optimal environmental conditions and enhancing animal welfare within the facility.

3.2 Case Study II: Montagnini Farm

The second on-farm data set was collected from the Montagnini Family Farm, located in San Pietro in Casale, Bologna, Emilia-Romagna, Italy (coordinates: 44.716038 N, 11.451233 E, 10 m above sea level). The farm building, measuring 80.0 meters in length and 42.32 meters in width, is designed with a variable-height roof that ranges from 4.0 meters at the sides to 12.15 meters at the gable peak (Figure 3.5).

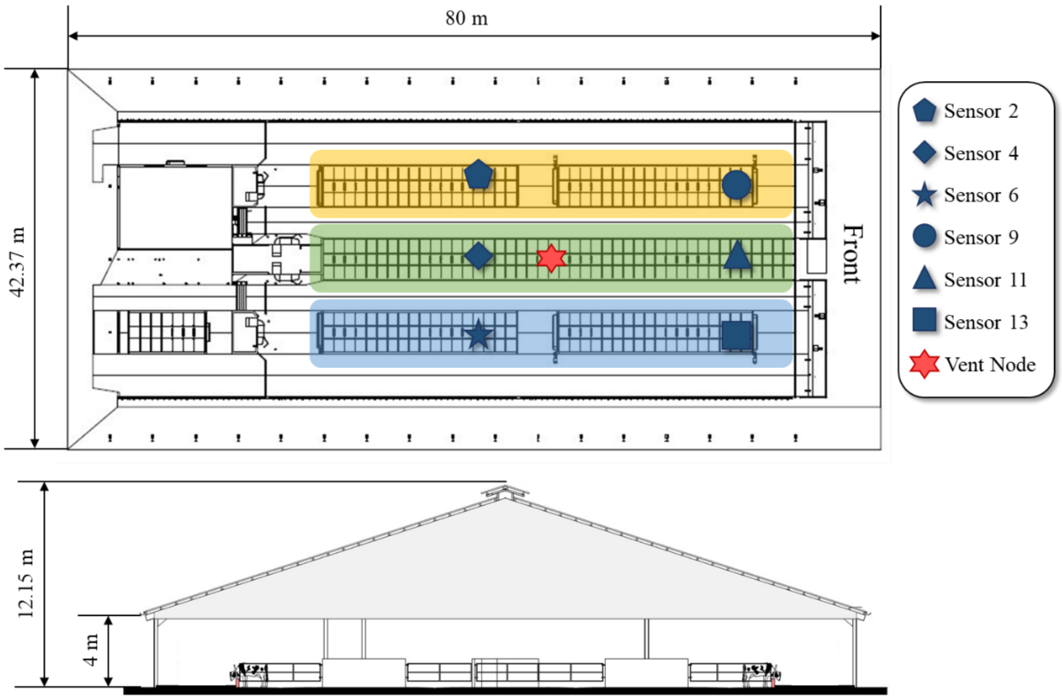

Figure 3.5 Layout and Sensor Placement of Dairy Building: Overhead View with Sensor Locations at Top and Front View with Dimensions at Bottom

The internal layout hosts approximately 260 Holstein-Friesian dairy cows and is organized into functional areas, including resting, feeding, milking, and an infirmary zone. The cows are milked up to four times daily using two Lely Astronaut A4 robotic milking systems. These systems provide individualized milking schedules that cater to each cow's productivity needs. This setup not only supports optimal milk production but also enhances animal welfare by allowing cows to follow their natural milking patterns.

The ventilation system is currently composed of 14 fans with horizontal axes operating along the two feeding lanes (i.e., 7 per lane) and 10 fans with vertical axis distributed on two lines and placed over the resting area. The operation of these fans is governed by the THI, which is continuously monitored by a centrally located sensor, denoted by the red six-pointed star in Figure 3.5. This setup ensures that the ventilation system responds dynamically to changes in the barn's microclimate, thereby providing a comfortable environment for the cows and minimizing the risk of heat stress.

3.2.1 Data Collection

The monitoring system architecture for the case study of Montagnini Family Farm is structured in a manner consistent with that of the system presented in the preceding case study, comprising sensor nodes and a LoRaWAN gateway. The primary differences between the two configuration sets relate to the location and configuration of the sensors. In this configuration, all sensor nodes are positioned at a uniform height of 2.5 meters and are arranged into three distinct clusters. This arrangement allows for the internal barn area to be divided into three zones: northwest, central, and southeast, as represented by the yellow (Group 1), green (Group 2), and blue (Group 3) shaded areas in Figure 3.5, respectively. The strategic clustering of sensors enables more precise monitoring of microclimatic conditions within each zone, providing detailed insights into environmental variations across the barn.

A second distinguishing feature of the SMS is the incorporation of an anemometer into one of the sensor nodes (Figure 2.3). The anemometer provides crucial data on internal airflow within the barn, which is of significant value for the effective management of the ventilation system. By measuring wind speed, the system is able to make informed adjustments to optimize air circulation and thermal regulation. The anemometer, which is part of the Weather Sensor Assembly p/n 80422, features a cup design that measures wind speed through a magnetic mechanism. In particular, for every 1.492 MPH (2.4 km/h) of wind speed, the magnet closes a switch once per second.

3.2.2 Data Preprocessing

The data preprocessing methods applied in this case study were consistent with those utilized in the previous case study, ensuring uniformity in data handling and preparation. As was the case in the preceding study, the raw data were resampled to a 30-minute interval. This resampling process aimed to strike a balance between computational efficiency and data granularity, retaining key trends while optimizing the dataset for model development. Furthermore, the resampling process facilitated the handling of missing values, which were imputed for short gaps to maintain continuity and enhance model reliability.

An outlier detection procedure was implemented to ensure the accuracy and validity of the data. Using a cascading approach, initial outliers were identified by clipping values that were outside the expected measurement range of the sensors. This was followed by the application of the MAD algorithm, which is effective in detecting anomalies due to its robustness to extreme values.

The incorporation of wind speed data in this case study required the implementation of supplementary preprocessing steps. Wind speed data, which is crucial for calculating the ETI, was processed in conjunction with temperature and humidity readings to ensure consistency across variables. In cases where data was absent for an extended period, defined as intervals with more than 24 consecutive missing points, external meteorological data and linear interpolation were employed, as previously described.

3.2.3 Model Training

In this study, the target variable is the ETI, which incorporates wind speed along with temperature and humidity measurements. A time series of these environmental parameters was constructed to provide an overview of the conditions within the barn. Figure *3.6* shows the full time series for temperature, humidity, wind speed, and ETI, illustrating the trends and seasonality for each variable. This broad perspective offers valuable insights into the barn's overall environmental conditions, serving as the basis for the development of accurate predictive models.

Figure 3.6 Time Series of Environmental Parameters at Montagnini Farm

A notable variability in wind speed measurements can be observed across the three sensor clusters between January 28 and February 3, 2023. Figure 3.7 provides a detailed illustration of this period, with ETI in the top subplot and wind speed in the bottom subplot. A more detailed analysis reveals the discrepancies in wind speed across different zones and illustrates the subsequent influence on ETI values. Given these observed variations, tailored models were developed for each cluster, rather than employing a single uniform model across all sensors, as was the case in the Piazzi Farm study.

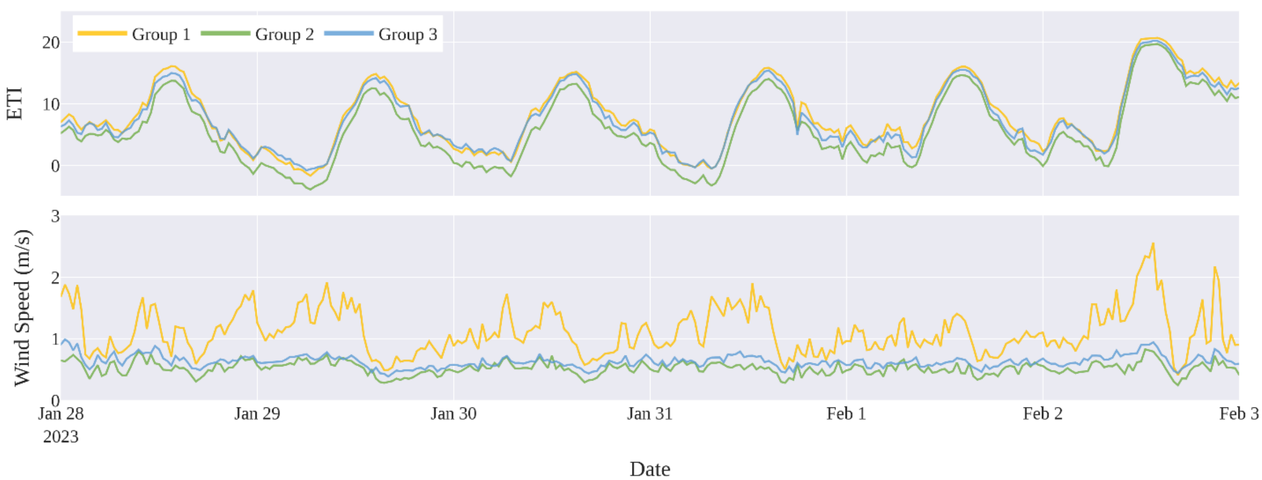

Figure 3.7 Detailed View of Wind Speed and ETI Fluctuations at Montagnini Farm (January 28 - February 3, 2023)

Despite the model customization, several parameters were kept consistent across the three models to reflect the shared trends and seasonal components observed in the environmental data. These key parameters closely align with those utilized in the initial case study. These include $n\_forecasts = 48$ for forecasting 24 hours ahead at 30-minute intervals and $n\_lags = 6$ to capture short-term dependencies up to three hours prior. In consideration of the seasonal and diurnal patterns, the parameters $yearly\_seasonality$ and $daily\_seasonality$ were set to True, while $weekly\_seasonality$ was maintained at False due to the absence of weekly cycles. An additive seasonality mode was employed, which is suitable for cumulative environmental effects. Furthermore, the $normalize = "auto"$ setting was retained to ensure that each dataset was appropriately scaled. Furthermore, the use of $batch\_size = 64$ and $epochs = 100$ enabled efficient and stable training.

In this instance, the learning rate was set at 0.001 for all models, a value that strikes a balance between ensuring stability and facilitating convergence. However, the models differed in their settings for the number of change points, in order to reflect anticipated shifts in trend. Model 1 and Model 3 employed 10 changepoints, whereas Model 2 necessitated 30 to capture more intricate wind patterns. The $lagged\_reg\_layers$ and ar_layers parameters were also tailored, with Model 2 incorporating a more complex layer structure (e.g., [30, 20, 10] for $lagged\_reg\_layers$ and [40, 30, 20] for ar_layers]) to accommodate its higher wind variability. In contrast, Model 1 and Model 3 employed more straightforward configurations, with Model 1 utilising layers of [20, 10] and Model 3 layers of [10].

3.2.4 Model Validation

In order to validate the predictive models developed for the Montagnini Farm case study, the same strategies were employed as were used in the Piazzi Farm case study. The dataset was divided into two distinct sets: a training set, comprising 80% of the data, and a validation set, comprising the remaining 20%. This division was accomplished through the utilization of the NeuralProphet framework's $model.split\_df$ function. During the process of model fitting, the validation dataset (df_val) was incorporated into the $model.fit$ function to evaluate the model's performance on data that had not been seen during training, thereby providing an essential assessment of the model's capacity for generalization.

The validation metrics MAE, RMSE, MAPE, and Loss were derived for each of the three predictive models. Table 3-2 summarizes these metrics across the different barn zones. The MAE values range from 2.16 to 2.96, while the RMSE values span from 2.92 to 3.65, indicating satisfactory absolute accuracy in forecasting ETI across the barn zones. MAPE further provides insight into the relative forecasting performance of the models. MAPE values range from approximately 5.5% to 6.6%, indicating moderate relative deviations between predicted and measured ETI values. Compared to absolute error metrics, MAPE highlights a conservative model behavior during periods of elevated ETI, where relative errors become more pronounced. This behavior is consistent with the increased sensitivity of ETI to high-frequency variations in air velocity, which are more difficult to capture using time-series forecasting approaches.

Table 3-2 Validation Metrics for ETI Predictive Models Across Barn Zones

Model ID	MAE	MAPE	RMSE	Loss
Group 1	2.1699	0.0554	2.9257	0.0556
Group 2	2.9578	0.0655	3.6540	0.1692
Group 3	2.7968	0.0593	3.6311	0.1119

The observed Loss values, ranging from 0.06 to 0.17, further confirm the robustness of the models. These metrics are indicative of the models' ability to generalize well across varying environmental conditions within the barn zones, making them suitable for real-time application. Notably, Zone 2 exhibited slightly higher error margins, which may be attributed to the complex wind speed dynamics observed in that area.

A review of the metrics presented in Table 3-2 suggests that the ETI predictive models are well-suited for practical applications within the Montagnini Farm context. The models' low error rates across sensor locations confirm their accuracy and reliability, which is essential for maintaining optimal environmental conditions and enhancing animal welfare within the barn.

CHAPTER 4. RESULTS AND DISCUSSION

This chapter presents the outcomes and analysis of the predictive modeling approaches applied to monitor and manage indoor environmental conditions in dairy cattle barns. This study focuses on two critical thermal indices, the Temperature-Humidity Index (THI) and the Equivalent Temperature Index (ETI), and explores how these indices can be leveraged to optimize animal welfare and enhance energy efficiency. The initial section of the study assesses the efficacy of THI projections across a range of sensors, utilizing spatial interpolation to gain a comprehensive understanding of the barn's thermal environment. The second section assesses the efficacy of ETI-based ventilation control as a means of reducing energy consumption while maintaining optimal conditions for livestock. By investigating both forecasting and real-time control applications, this chapter demonstrates the potential of data-driven strategies to enhance precision livestock farming through predictive insights and adaptive management.

4.1 THI Prediction Analysis

The summer season is a crucial period for assessing indoor conditions in dairy cattle barns, given the rise in ambient temperatures and the associated increase in animal discomfort indices. For this analysis, June 20, 2023, was selected as a representative date during the summer, encompassing a 24-hour period with 30-minute intervals. This timeframe allows for an in-depth analysis of the model's performance under conditions that are likely to challenge the welfare of the animals.

Figure 4.1 presents a comparison between the predicted THI values (orange line) and the actual measurements obtained by the SMS (light blue line) across all 12 sensor nodes within the barn. The results highlight variation in predictive accuracy among the different sensors. Notably, sensors 5, 9, and 10 exhibited the lowest MAE values of 0.87, 1.29, and 1.17, respectively, demonstrating strong alignment between predicted and observed data in these zones. In contrast, sensor 7 exhibited the highest MAE of 7.06, indicating a diminished predictive accuracy in that zone (in proximity to the AMS).

The RMSE values exhibited variation across sensors, with sensors 5, 9, and 10 demonstrating the lowest RMSE at 1.12, 1.52, and 1.41, respectively. Conversely, sensor 7 exhibited the highest RMSE at 7.32. This variability highlights discrepancies in model performance across the barn, indicating potential areas for improvement through additional refinement to enhance prediction accuracy.

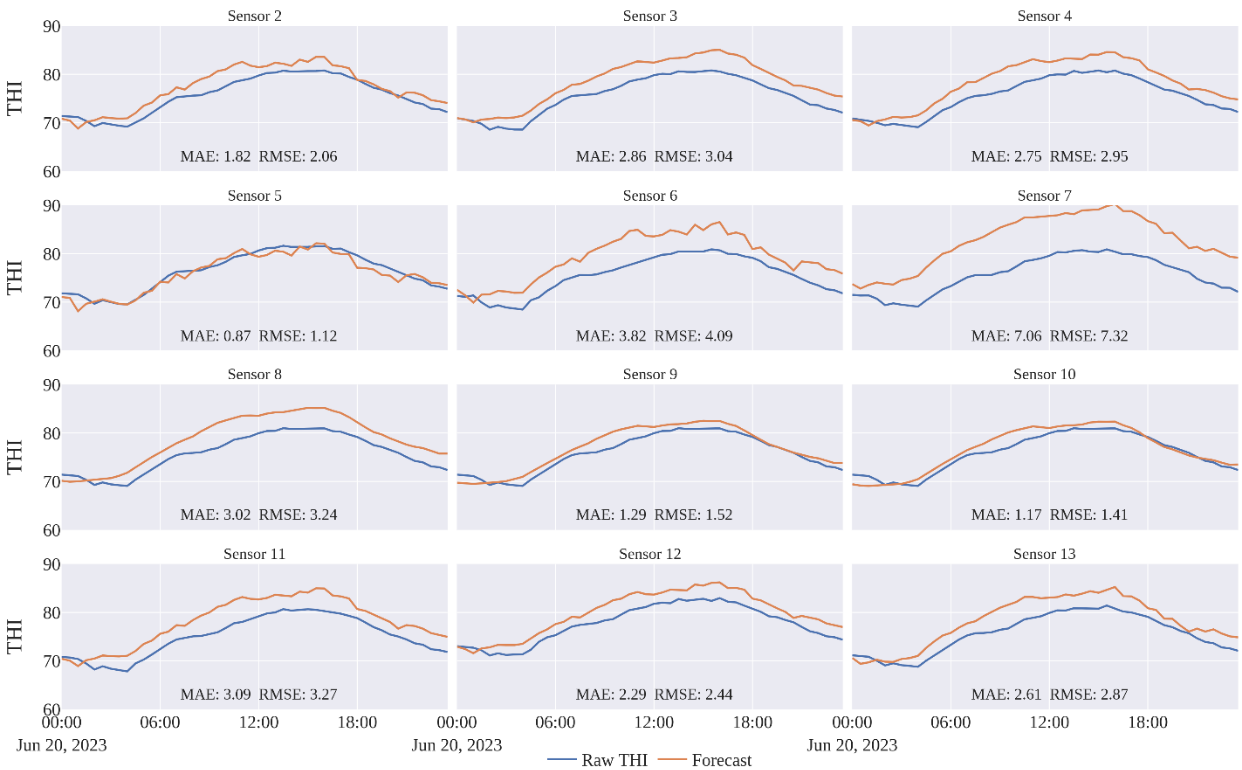

Figure 4.1 Indoor Temporal Forecasting of Temperature-Humidity Index (THI) Across Multisensor Measurement Points

The findings of this analysis demonstrate that the THI models exhibit satisfactory predictive accuracy when applied to all sensor nodes, suggesting that they provide a reliable initial approximation for forecasting THI in this context. Although there is some variation in accuracy among individual sensors, the MAE and RMSE values indicate that the models perform sufficiently well to offer valuable insights into thermal conditions within the barn. It is notable that the tendency of the models to overestimate THI can be regarded as a protective measure, as it ensures that potentially adverse conditions are identified in a timely manner. This approach enhances animal welfare by enabling the implementation of proactive measures to mitigate heat stress before it reaches critical levels.

4.1.1 IDW Interpolated THI Heatmap - Hot Conditions

To further understand the spatial distribution of THI across different zones within the barn, Inverse Distance Weighting (IDW) interpolation was used to estimate THI values in unmeasured areas. The following section will present two different scenarios, each based on distinct environmental conditions observed at different times of day.

Figure 4.2 illustrates the interpolated THI distribution at four different heights (1.0 m, 2.0 m, 2.5, and 3.5 m) within the barn, recorded at 16:00 when THI values were at their peak. The diamonds in magenta represent the locations of the sensor nodes that provided the measured THI values used as input for the interpolation. The heatmap shows significant variability in THI values along both horizontal and vertical dimensions within the barn. In particular, higher THI values are observed near sensor node S7 at a height of 3.5 m. While this height is above the level where animals are typically present, it reflects the dynamics of physical systems where hot air tends to rise and accumulate near the ceiling in livestock buildings. Although animals are not directly exposed to these elevated THI levels, it is important to monitor this zone as it affects the overall indoor environment. Elevated temperatures near the ceiling can affect air circulation patterns and contribute to a warmer overall microclimate in the barn.

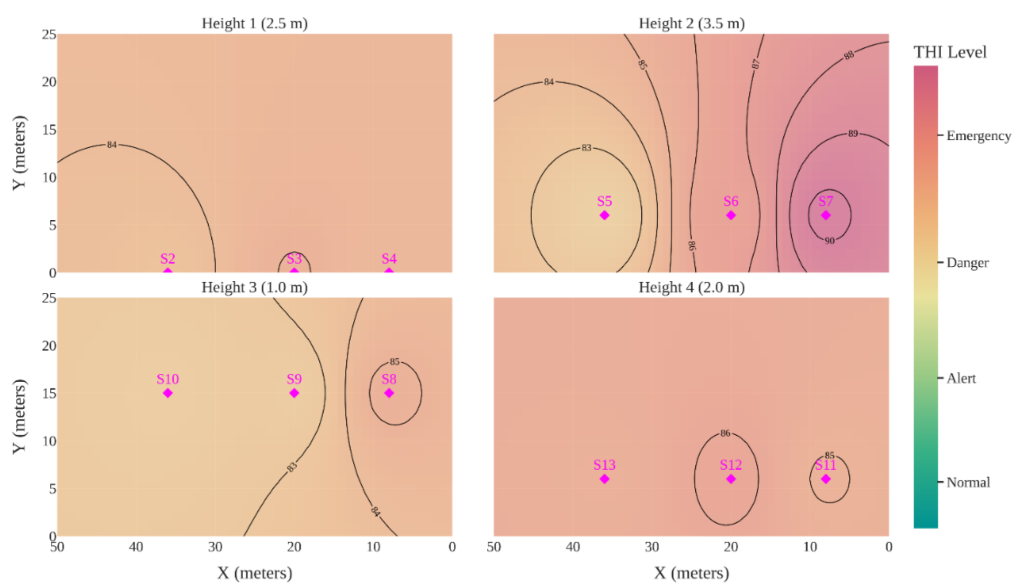

Figure 4.2 IDW Interpolated THI Heatmap - Hot Conditions (2023-06-20 16:00)

These observations suggest that the ventilation strategy may need to be adjusted to target these specific zones, especially during peak heat conditions. The use of IDW interpolation provides a more comprehensive understanding of the THI distribution beyond individual sensor measurements, allowing for more targeted and effective environmental management within the barn.

4.1.2 IDW Interpolated THI Heatmap - Cooler Conditions

To evaluate how THI varies under cooler conditions, Figure 4.3 shows the interpolated THI heatmap at four different heights (1.0 m, 2.0 m, 2.5 m, and 3.5 m) within the barn at 19:00, a time when THI values were significantly lower compared to peak conditions. Similar to the previous heatmap, the magenta diamonds indicate the locations of the sensor nodes that provided the data for interpolation.

The interpolated heatmap indicates relatively moderate THI values across all zones and elevations within the barn, with most readings falling within the Normal or Alert ranges. This indicates that, overall, the barn environment is conducive to animal comfort during this time, with THI levels well below the thresholds for “Danger” or “Emergency” levels. However, higher THI values are still observed in certain areas, particularly near sensor node S7, which again records the highest THI at the 3.5 meter level.

The stratification observed in this map, although less pronounced than during the peak heat conditions, illustrates that even under cooler conditions, there can be areas within the barn where THI remains elevated. These localized pockets of higher THI at upper levels, such as around S7, underscore the importance of monitoring and managing airflow throughout the barn. In this scenario, IDW interpolation provides valuable insight into how environmental conditions may differ at different levels, allowing for the possibility of more refined and energy efficient ventilation strategies.

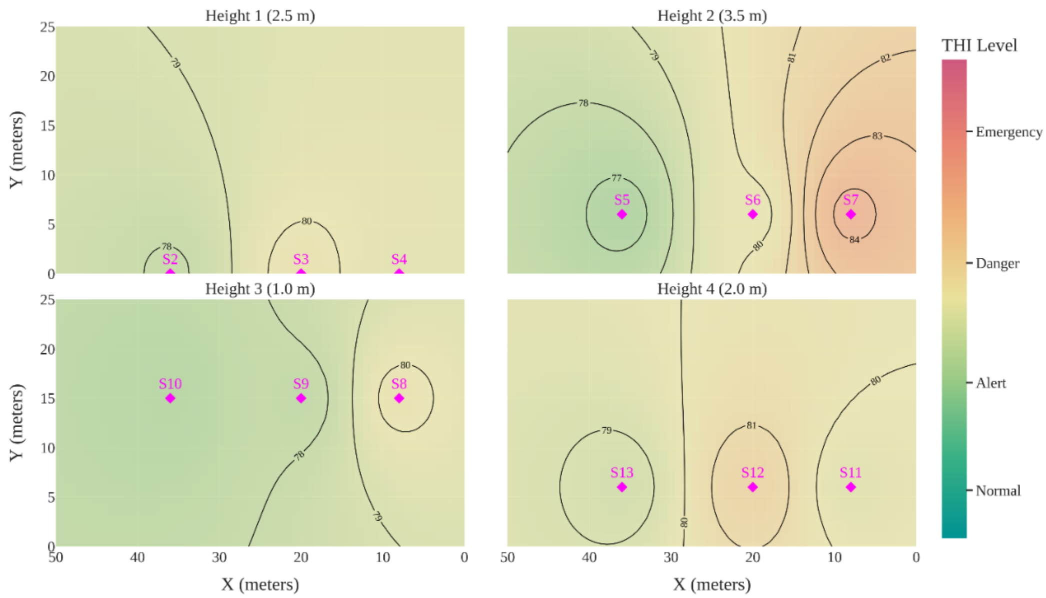

Figure 4.3 IDW Interpolated THI Heatmap - Cooler Conditions (2023-06-20 19:00)

In addition, the presence of moderate THI values near sensor nodes that influence the barn ventilation system suggests a potential area for optimization. Even though cooler conditions prevail, the recorded THI values from the ventilation control sensor may be triggering the fans unnecessarily, resulting in increased energy consumption. This visualization highlights the value of integrating interpolated spatial data with the predictive models, providing a tool for fine-tuning the operation of environmental control systems. By predicting THI distributions, farm managers can make more informed decisions to balance animal comfort with energy efficiency.

4.2 ETI Prediction Analysis

In order to evaluate the forecasted ETI within the barn, June 15, 2024 was selected as a representative date. This date was selected based on the occurrence of high ambient temperatures, which have been demonstrated to result in elevated ETI values and, consequently, greater heat stress risks for livestock. Forecasted ETI values provide a basis for evaluating the temporal evolution of thermal stress conditions within the barn under critical climatic scenarios.

In the ETI forecasting analysis, external weather forecasts were included as exogenous inputs alongside historical sensor measurements. Their inclusion contributed to greater predictive stability during periods characterized by rapidly changing outdoor conditions, particularly under high thermal stress scenarios. This behavior is consistent with recent studies reporting the benefits of integrating short-term weather forecasts into environmental prediction models for livestock buildings.

Figure 4.4 presents a comparison between the forecasted ETI values (orange line) and the actual ETI values computed from the sensor clusters (light blue line) for the three defined zones within the barn. The figure offers insight into the accuracy of the ETI forecasting models developed for each of the three groups of sensors, which were clustered in a strategic manner based on sensor location and environmental similarity within the barn.

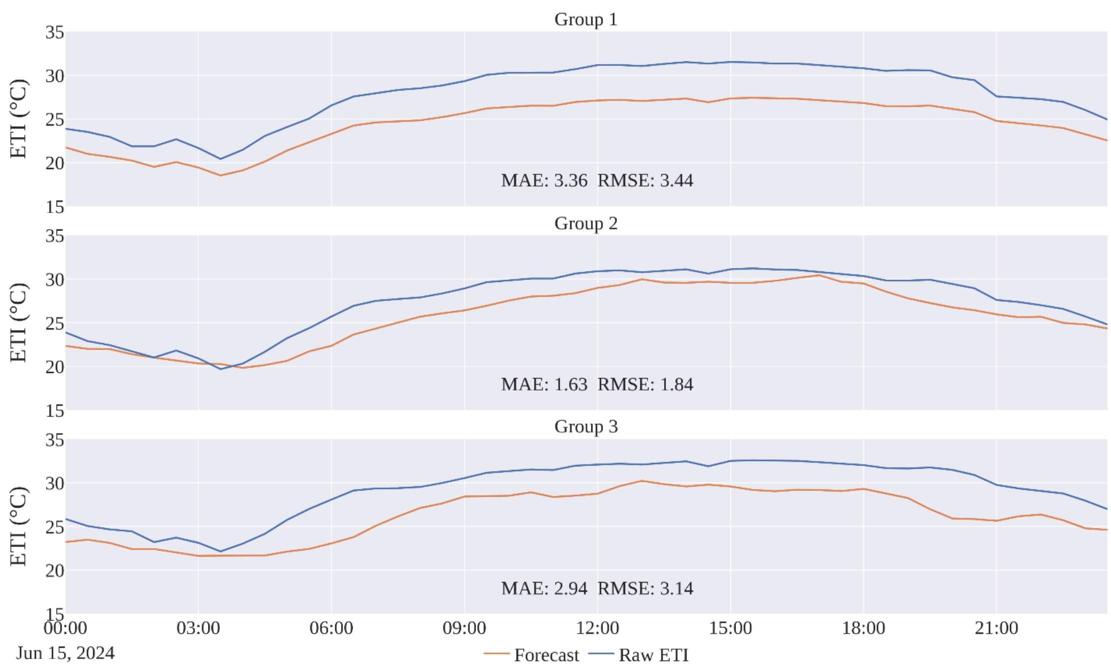

Figure 4.4 Comparison of Predicted vs. Measured ETI Across Three Clusters on June 15, 2024

In terms of performance, Group 1 exhibited a mean absolute error (MAE) of 3.36 and a root mean square error (RMSE) of 3.44, indicating moderate accuracy in forecasting ETI values for this cluster. Group 2, in contrast, exhibited the highest degree of forecasting accuracy, with a MAE of 1.63 and a root mean square error (RMSE) of 1.84. This suggests that the model's predictions were in close alignment with the actual recorded ETI values in this zone. Group 3 exhibited a marginally higher error margin than Group 2, with a mean absolute error (MAE) of 2.94 and a root mean square error (RMSE) of 3.14. This reflects a reasonable level of forecasting precision for this group.

Despite the overall satisfactory performance of the ETI forecasting models, a systematic underestimation of peak ETI values was observed across the analyzed zones. This behavior differs from that observed for THI forecasts and can be attributed to both the formulation of the ETI and the internal structure of the NeuralProphet framework. Unlike THI, ETI explicitly incorporates air velocity, a variable characterized by high temporal variability and localized fluctuations driven by fan operation and animal activity.

Within the NeuralProphet architecture, the trend (T) and seasonal (S) components capture dominant long-term and periodic patterns, while the autoregressive (A) and lagged regressor (L) components model short-term dependencies and external influences. However, abrupt and transient variations in air velocity may not be fully captured by these components, resulting in a smoothing effect that attenuates sharp ETI peaks. Consequently, the model tends to produce conservative estimates under extreme thermal stress conditions, which explains the observed underestimation relative to measured ETI values.

The results obtained from the three groups demonstrate that the forecasting models are effective in estimating future ETI values within the barn, thereby providing valuable insights for the management of thermal conditions. Although there was some variation in forecasting performance between the groups, the overall accuracy was sufficient to provide practical value in guiding ventilation strategies and supporting informed management decisions.

4.2.1 Evaluating ETI as a Control Variable for Ventilation Systems

Ventilation control systems in dairy cow barns commonly rely on THI as the primary control variable. While its widespread adoption is largely due to its straightforward computation and the practical challenges associated with measuring airflow within livestock facilities, THI does not explicitly account for air velocity, a parameter that varies in response to ceiling fan operation.

In this study, the sensor nodes deployed at Montagnini Farm were equipped with anemometers to measure air speed at six internal locations. The integration of these measurements enabled not only the calculation of an alternative thermal index but also the exploration, through simulation, of the potential use of the ETI as a control variable for ventilation systems.

First, an analysis was conducted with the objective of establishing a correlation between ETI and THI on the basis of the discomfort thresholds established for dairy cattle. The thresholds proposed by (Eigenberg et al., 2005) were employed to identify the setpoint for ventilation, with THI values of 74 and above designated as the condition for ventilation activation and values below 68 selected for deactivation. Figure 4.5 presents a scatter plot comparing historical ETI and THI values over the same time period. The strong correlation between ETI and THI, as illustrated in Figure 4.5, facilitated the calibration of a linear model with an R^2 exceeding 0.96 across all three sensor groups. This high correlation enabled the establishment of ETI control thresholds at 27°C for activation and 22°C for deactivation, aligning ETI values with the established discomfort thresholds for dairy cattle.

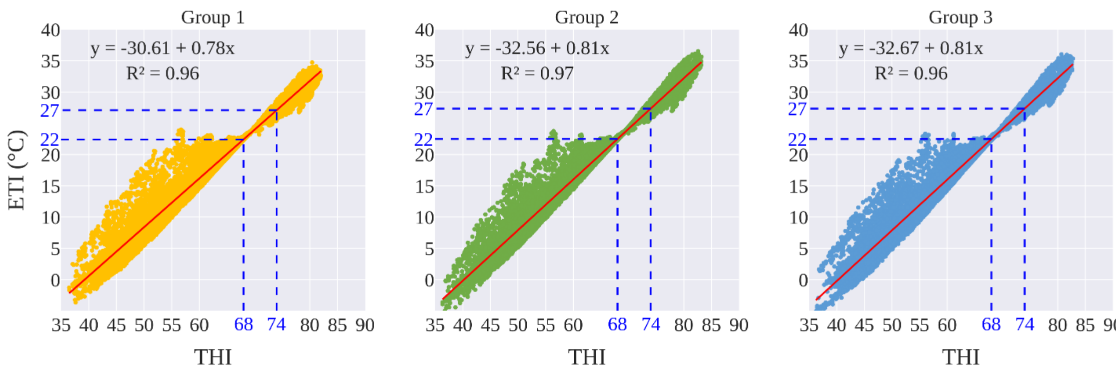

Figure 4.5 Relationship Between ETI and THI for Ventilation Control Threshold Determination

Once the ETI control thresholds had been defined, the forecasted ETI data from the predictive models was employed to simulate the fan control signal. Figure 4.6 illustrates the control signals for the three distinct zones within the barn over the selected time period, demonstrating that these zones have varying thermal requirements.

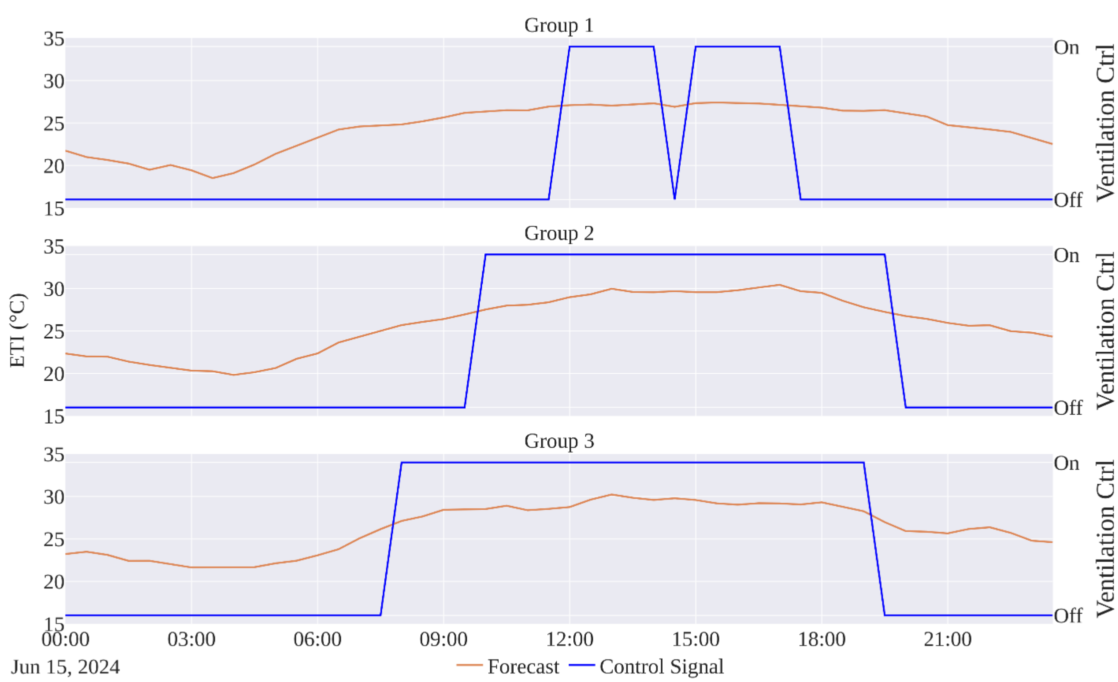

Figure 4.6 Simulated ETI-Based Ventilation Control Signals Across Barn Zones

As illustrated in Figure 4.6, the activation of fans occurs in a staggered pattern, beginning with the southeast zone (Group 3), then the central zone (Group 2), and finally the northwest zone (Group 1). This sequential activation pattern may be attributed to a number of factors, including the barn's open structure and the prevailing wind direction. This pattern demonstrates the value of a customized ETI-based control system, which has the potential to optimize fan operation and reduce unnecessary energy consumption while maintaining animal comfort throughout the barn.

4.2.2 Energy Consumption Comparison: Default Control vs. ETI-Controlled Ventilation

In order to assess the potential energy savings that could be achieved through the implementation of an ETI-based control approach for the barn's ventilation system, two distinct operational scenarios were subjected to analysis. The initial scenario assumes that all fans operate under the same regimen manager by the single point measurement, thereby establishing a maximum consumption baseline. In contrast, the second scenario employs an ETI-based control strategy, whereby fans are activated solely when ETI levels in each barn zone indicate a necessity for cooling in order to maintain optimal animal comfort.

Energy consumption was estimated for both scenarios based on the power rating of each fan (75 W) and the configuration of four fans per zone, totaling twelve fans across the system. The total energy consumption for each scenario was then computed at half-hour intervals throughout the 24-hour evaluation period. In the baseline scenario, all fans operated continuously, resulting in a constant rate of energy use. In contrast, the second scenario exhibited variable energy consumption, governed by the ETI-based control signals simulated for each barn zone according to the predictive model.

The results of the energy analysis are presented in Figure *4.7*. The default operation scenario (orange) represents the energy consumption for all fans under the default control strategy, whereas the ETI-based control scenario (blue) illustrates the consumption when the fans are activated only when necessary. The energy savings were calculated to be approximately 13.65 kWh, which represents a significant reduction in comparison to the always-on scenario.

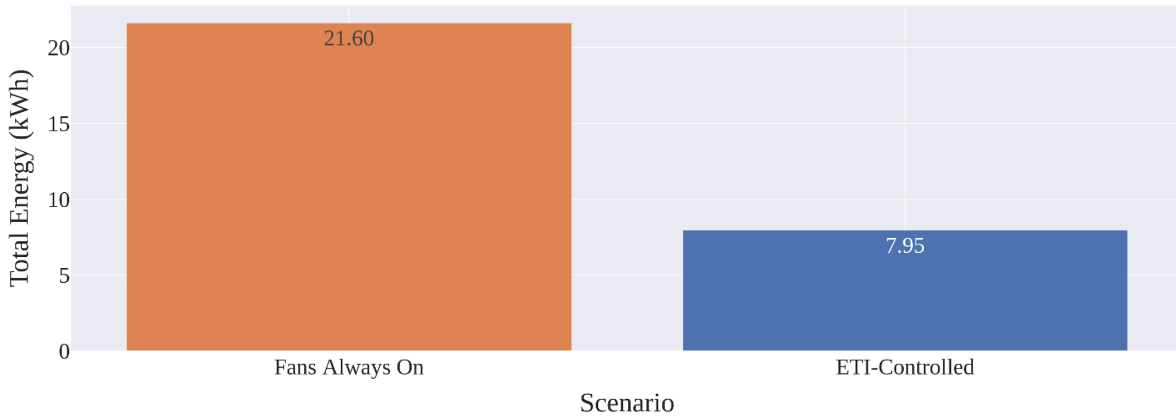

Figure 4.7 Energy Consumption Comparison Between Always-On and ETI-Controlled Ventilation Systems

The ETI-based control scenario was simulated using forecasted ETI values generated by the predictive models. As shown in the ETI validation results (Section 4.2), the forecasts tend to conservatively estimate peak ETI conditions. This bias may lead to an underestimation of fan activation time and, consequently, to a conservative estimate of energy consumption in the ETI-controlled scenario.

A direct recalculation of energy consumption using measured ETI values as the control input would provide a more accurate benchmark for quantifying the impact of forecast bias; however, this comparison was not performed in the present analysis because the energy simulation workflow was implemented exclusively on forecast-based control signals for the selected evaluation period.

4.3 Discussion

The application of predictive models for forecasting thermal indices in livestock barns supports more precise environmental monitoring and control, with direct implications for animal welfare and operational efficiency. This study focused on two main applications: the interpolation-based forecasting of the Temperature–Humidity Index (THI) across the barn and the use of the Equivalent Temperature Index (ETI) to guide energy-efficient ventilation strategies.

4.3.1 Selection of the predictive modeling approach

The selection of the predictive modeling approach was guided by the specific characteristics of the monitored barn environment, which exhibits nonlinear dynamics, multiple seasonal behaviors, and strong interactions with external weather conditions. Such complexity requires a forecasting framework capable of jointly incorporating autoregressive effects and exogenous information while remaining robust to sensor noise and occasional data gaps. In this context, NeuralProphet was adopted as it provides greater flexibility than conventional statistical approaches, such as ARIMA, while avoiding the high data requirements and tuning complexity often associated with purely deep learning models.

4.3.2 Model performance variability across sensor locations

The forecasting accuracy exhibited variability across sensor locations, reflecting the heterogeneous microclimatic conditions within the barn. Higher prediction errors were primarily detected at sensors positioned near the external perimeter of the building. Due to the semi-open structure of the facility, these areas are particularly exposed to rapid and unpredictable fluctuations in temperature, humidity, and air velocity driven by outdoor weather conditions.

Additionally, increased errors were identified at nodes located in zones characterized by intense animal activity, such as the Automatic Milking System (AMS) area. In these locations, frequent animal movement, intermittent ventilation patterns, and localized heat and moisture generation create short-term disturbances that are inherently difficult for time-series forecasting models to capture.

4.3.3 Spatial variability in forecasting performance

The integration of spatial interpolation with predictive forecasting enabled the estimation of thermal indices in unmonitored areas of the barn, providing a more comprehensive representation of the indoor microclimate. The application of the IDW method allowed the reconstruction of continuous THI distributions across different zones and heights, revealing localized thermal gradients that would not be captured through single-point measurements.

Furthermore, the use of ETI as a control variable introduced a more physically representative indicator of thermal comfort by incorporating air velocity alongside temperature and humidity. This resulted in a more responsive ventilation strategy, enabling localized and demand-driven operation of the cooling system. Compared with conventional THI-based control, this approach enabling more precise environmental regulation while reducing unnecessary energy consumption.

4.3.4 Challenges and Limitations

Despite the contributions of this study, several limitations should be acknowledged. In particular, the proposed framework does not incorporate variables related to the number and spatial distribution of animals within the facility. The analysis focuses primarily on environmental parameters, without explicitly accounting for animal presence or movement. However, animals directly influence localized thermal conditions through metabolic heat production and behavioral patterns, which may introduce additional variability in the indoor climate. Incorporating animal-related data, such as population density or real-time location within specific zones, could provide valuable context for the predictive models and improve the accuracy of comfort assessments across the barn.

A further limitation concerns the definition of ETI-based ventilation thresholds. In this study, these thresholds were derived through statistical relationships with THI rather than through direct physiological indicators of heat stress, such as respiration rate or core body temperature, which are commonly regarded as more representative measures of animal discomfort. While this approach enables practical implementation using readily available environmental measurements, it introduces an indirect calibration that may limit the physiological interpretability of the control strategy. Consequently, the ETI-based system should therefore be regarded as an operational refinement of existing THI-based methods rather than a physiologically validated replacement. Future research should incorporate animal-based indicators to establish thresholds more directly linked to welfare outcomes.

Additionally, although the incorporation of wind speed into the ETI-based control strategy represents a methodological advancement, challenges related to the maintenance and calibration of anemometers remain. Environmental factors such as dust and humidity may affect sensor performance over time, requiring periodic maintenance to ensure data reliability. Addressing these operational constraints will be essential for scaling the proposed system to broader commercial applications.

Finally, the estimation of energy savings associated with the ETI-based ventilation strategy was based on a simulated baseline scenario intended to represent conventional THI-based control. In practice, barn management often involves manual adjustments, adaptive setpoints, and operator experience that may not be fully captured by a simplified control model. As a result, the absolute magnitude of the reported savings should be interpreted as indicative rather than definitive.

Although the predictive model operated at a 30-minute temporal resolution, capturing short-term and day-to-day fluctuations in energy demand, the analysis primarily reported aggregated results. It did not explicitly stratify performance across seasons or extreme climatic conditions. Evaluating the robustness of the proposed strategy under different seasonal scenarios would provide a more comprehensive assessment of its generalizability. Additionally, the study did not directly relate energy savings to animal performance indicators such as milk yield, behavior, or physiological responses.

4.3.5 Potential Applications and Future Directions

The results of this study indicate significant potential for the broader application of predictive modeling and interpolation within the domain of Precision Livestock Farming. One promising avenue of inquiry involves integrating these procedures into a standardized pipeline within the SMS framework. By automating the data flow from sensor measurement to prediction and visualization, the SMS can provide farmers with actionable insights on barn conditions in real time, enabling them to respond proactively to forecasted conditions. Such integration could be facilitated by the implementation of a cloud-based infrastructure, which would enable the system to process and disseminate predictive insights directly to farm operators via a centralized platform.

Moreover, the incorporation of animal-specific data, such as individual animal location and heat output, has the potential to facilitate the development of even more sophisticated control strategies in future iterations of this SMS workflow. This would permit the system to optimize ventilation based on environmental conditions and to adjust dynamically based on the real-time distribution of animals within the barn. Such capabilities could markedly enhance the capacity of farm operators to maintain optimal conditions, thereby improving both animal welfare and operational efficiency.

Overall, the integration of THI and ETI forecasting as essential elements of an SMS represents a significant advancement in predictive environmental management for livestock facilities. By enabling a more detailed and responsive assessment of indoor climate conditions, these tools demonstrate strong potential to support energy-efficient and welfare-oriented ventilation strategies, particularly when combined with high-resolution sensing, external weather information, and animal-based data in future implementations.

CONCLUSIONS

This research aimed to improve environmental control systems in dairy cattle barns through a data-driven, neural network-based predictive modeling approach, focusing on multi-zone forecasting and the integration of advanced thermal indices. The study explored the use of NeuralProphet to predict temperature-humidity index (THI) and equivalent temperature index (ETI), leveraging IoT technologies to support animal welfare and energy-efficient barn operation. The main conclusions of the study are summarized as follows:

- The NeuralProphet-based predictive model demonstrated satisfactory accuracy in forecasting THI levels across multiple zones within the barn. The model achieved mean absolute error (MAE) values ranging from 3.43 to 8.73 and root mean square error (RMSE) values ranging from 4.01 to 9.81 across different sensor nodes. These results indicate that the proposed forecasting framework is capable of capturing the temporal dynamics of thermal conditions relevant to animal welfare, particularly during periods of elevated heat stress.
- The application of spatial interpolation enabled the development of a detailed multizonal representation of the Temperature-Humidity Index (THI) across different heights within the barn. This approach offered a thorough characterization of the spatial variability of thermal conditions. This made it easier to identify localized hotspots and zones that could benefit from different ventilation regimens.
- By incorporating wind speed measurements, the study successfully calculated ETI and developed a corresponding predictive model using NeuralProphet. The ETI forecasts exhibited MAE values ranging from 2.16 to 2.96 and RMSE values ranging from 2.92 to 3.65 across the defined barn zones. While the model showed a tendency to conservatively estimate peak ETI values, the results support the relevance of ETI as a more comprehensive

indicator of thermal comfort compared to THI, due to its explicit consideration of air movement. This provides an enhanced basis for evaluating ventilation requirements under both observed and forecasted conditions.

- The ETI-based ventilation control strategy demonstrated the potential to reduce energy consumption compared to a traditional always-on fan operation, with an estimated daily reduction of approximately 13.65 kWh under the evaluated conditions

REFERENCES

Adjassin, J. S., Assani, S. A., Assogba, B. G. C., Idrissou, Y., Worogo, H. S. S., Daramola, J. O., & Alkoiret, I. T. (2020). Heat stress under different climate conditions for dairy livestock in Benin, West Africa. *Livestock Research for Rural Development*, *32*(5).

Aerts, J. M., Wathes, C. M., & Berckmans, D. (2003). Dynamic Data-based Modelling of Heat Production and Growth of Broiler Chickens: Development of an Integrated Management System. *Biosystems Engineering*, *84*(3), 257–266. https://doi.org/10.1016/S1537-5110(02)00285-4

Aerts, J.-M., Berckmans, D., Saevels, P., Decuypere, E., & Buyse, J. (2000). Modelling the static and dynamic responses of total heat production of broiler chickens to step changes in air temperature and light intensity. *British Poultry Science*, *41*(5), 651–659. https://doi.org/10.1080/713654981

Alonso, R. S., Sittón-Candanedo, I., García, Ó., Prieto, J., & Rodríguez-González, S. (2020). An intelligent Edge-IoT platform for monitoring livestock and crops in a dairy farming scenario. *Ad Hoc Networks*, *98*, 102047. https://doi.org/10.1016/j.adhoc.2019.102047

Amiri-Zarandi, M., Hazrati Fard, M., Yousefinaghani, S., Kaviani, M., & Dara, R. (2022). A Platform Approach to Smart Farm Information Processing. *Agriculture*, *12*(6), 838. https://doi.org/10.3390/agriculture12060838

Banhazi, T., Aarnink, A., Thuy, H., Pedersen, S., Hartung, J., Payne, H., Mullan, B., & Berckmans, D. (2009). Review of the Consequences and Control of High Air Temperatures in Intensive Livestock Buildings. *Australian Journal of Multi-Disciplinary Engineering*, *7*(1), 63–78. https://doi.org/10.1080/14488388.2009.11464800

Banhazi, T. M., & Black, J. L. (2009). Precision Livestock Farming: A Suite of Electronic Systems to Ensure the Application of Best Practice Management on Livestock Farms. *Australian Journal of Multi-Disciplinary Engineering*, *7*(1), 1–14. https://doi.org/10.1080/14488388.2009.11464794

Banhazi, T. M., Lehr, H., Black, J. L., Crabtree, H., Schofield, P., Tscharke, M., & Berckmans, D. (2012). Precision Livestock Farming: An international review of scientific and commercial aspects. *International Journal of Agricultural and Biological Engineering*, *5*(3), 1–9. https://doi.org/10.3965/j.ijabe.20120503.00?

Berckmans, D. (2017). General introduction to precision livestock farming. *Animal Frontiers*, *7*(1), 6–11. https://doi.org/10.2527/af.2017.0102

Bianca, W. (1962). Relative Importance of Dry- and Wet-Bulb Temperatures in Causing Heat Stress in Cattle. *Nature*, *195*(4838), 251–252. https://doi.org/10.1038/195251a0

Black, J., & Banhazi, T. (2013). Economic and social advantages from Precision Livestock Farming in the pig industry. *The Proceedings of the 6th European Conference on Precision Livestock Farming*, 199.

Cockburn, M. (2020). Review: Application and Prospective Discussion of Machine Learning for the Management of Dairy Farms. *Animals*, *10*(9), 1690. https://doi.org/10.3390/ani10091690

da Costa Baêta, F., Meador, N. F., Shanklin, & Johnson, H. D. (1987). *Equivalent temperature index at temperatures above the thermoneutral for lactating dairy cows*. https://eurekamag.com/research/001/587/001587256.php

Da Silva, R. G., Maia, A. S. C., & de Macedo Costa, L. L. (2014). Index of thermal stress for cows (ITSC) under high solar radiation in tropical environments. *International Journal of Biometeorology*, *59*(5), 551–559. https://doi.org/10.1007/s00484-014-0868-7

Eigenberg, R. A., Brown-Brandl, T. M., Nienaber, J. A., & Hahn, G. L. (2005). Dynamic Response Indicators of Heat Stress in Shaded and Non-shaded Feedlot Cattle, Part 2: Predictive Relationships. *Biosystems Engineering*, *91*(1), 111–118. https://doi.org/10.1016/j.biosystemseng.2005.02.001

Fournel, S., Rousseau, A. N., & Laberge, B. (2017). Rethinking environment control strategy of confined animal housing systems through precision livestock farming. *Biosystems Engineering*, *155*, 96–123. https://doi.org/10.1016/j.biosystemseng.2016.12.005

Girshick, R. (2015). Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, 1440–1448. http://arxiv.org/abs/1504.08083

Gorczyca, M. T., & Gebremedhin, K. G. (2020). Ranking of environmental heat stressors for dairy cows using machine learning algorithms. *Computers and Electronics in Agriculture*, *168*. https://doi.org/10.1016/j.compag.2019.105124

Harvey, A. C., & Shephard, N. (1993). Structural time series models. In G.S. Maddala, C.R. Rao, & H.D. Vinod (Eds.), *Handbook of Statistics* (Econometrics, Vol. 11, pp. 261–302). North Holland.

Herbut, P., Angrecka, S., & Walczak, J. (2018). Environmental parameters to assessing of heat stress in dairy cattle—a review. *International Journal of Biometeorology*, *62*(12), 2089–2097. https://doi.org/10.1007/s00484-018-1629-9

Hoffmann, G., Herbut, P., Pinto, S., Heinicke, J., Kuhla, B., & Amon, T. (2020). Animal-related, non-invasive indicators for determining heat stress in dairy cows. *Biosystems Engineering*, *199*, 83–96. https://doi.org/10.1016/j.biosystemseng.2019.10.017

Ji, B., Banhazi, T., Perano, K., Ghahramani, A., Bowtell, L., Wang, C., & Li, B. (2020). A review of measuring, assessing and mitigating heat stress in dairy cattle. *Biosystems Engineering*, *199*, 4–26. https://doi.org/10.1016/j.biosystemseng.2020.07.009

Ji, B., Banhazi, T., Phillips, C. J. C., Wang, C., & Li, B. (2022). A machine learning framework to predict the next month's daily milk yield, milk composition and milking frequency for cows in a robotic dairy farm. *Biosystems Engineering*, *216*(March), 186–197. https://doi.org/10.1016/j.biosystemseng.2022.02.013

Kuczynski, T., Blanes-Vidal, V., Li, B., Gates, R. S., De Alencar Nääs, I., Moura, D. J., Berckmans, D., & Banhazi, T. M. (2011). Impact of global climate change on the health, welfare and productivity of intensively housed livestock. *International Journal of Agricultural and Biological Engineering*, *4*(2). https://doi.org/10.3965/j.issn.1934-6344.2011.02.001-022

Leliveld, L. M. C., Brandolese, C., Grotto, M., Marinucci, A., Fossati, N., Lovarelli, D., Riva, E., & Provolo, G. (2024). Real-time automatic integrated monitoring of barn environment and dairy cattle behaviour: Technical implementation and evaluation on three commercial farms. *Computers and Electronics in Agriculture*, *216*, 108499. https://doi.org/10.1016/j.compag.2023.108499

Milan, H. F., Perano, K. M., & Gebremedhin, K. G. (2018). Survey and future prospects in precision dairy farming. *10th International Livestock Environment Symposium (ILES X)*. https://doi.org/10.13031/iles.18-053

MQTT. (n.d.). *MQTT: MQ Telemetry Transport*. Retrieved August 30, 2024, from http://mqtt.org/

Navarro, E., Costa, N., & Pereira, A. (2020). A Systematic Review of IoT Solutions for Smart Farming. *Sensors*, *20*(15), 4231. https://doi.org/10.3390/s20154231

Neethirajan, S. (2020). The role of sensors, big data and machine learning in modern animal farming. *Sensing and Bio-Sensing Research*, *29*, 100367. https://doi.org/10.1016/J.SBSR.2020.100367

Node-RED. (n.d.). *Node-RED: Flow-based development tool for visual programming*. Retrieved August 30, 2024, from https://nodered.org/

Norton, T., Chen, C., Larsen, M. L. V., & Berckmans, D. (2019). Review: Precision livestock farming: building 'digital representations' to bring the animals closer to the farmer. *Animal*, *13*(12), 3009–3017. https://doi.org/10.1017/S175173111900199X

Open-Meteo. (n.d.). *Open-Meteo: Free and open-source weather API*. Retrieved January 9, 2024, from https://open-meteo.com/

Perez Garcia, C. A., Bovo, M., Torreggiani, D., Tassinari, P., & Benni, S. (2023). 3D numerical modeling of THI distribution in livestock structures: a cattle barn case study. *Journal of Agricultural Engineering*. https://doi.org/10.4081/jae.2023.1522

Perez Garcia, C. A., Bovo, M., Torreggiani, D., Tassinari, P., & Benni, S. (2024). Indoor Temperature Forecasting in Livestock Buildings: A Data-Driven Approach. *Agriculture*, *14*(2), 316. https://doi.org/10.3390/agriculture14020316

Perez Garcia, C. A., Tassinari, P., Torreggiani, D., & Bovo, M. (2025). Predictive Modeling of Energy Consumption for Cooling Ventilation in Livestock Buildings: A Machine Learning Approach. *Energies*. https://doi.org/10.3390/en18030633

Polsky, L., & von Keyserlingk, M. A. G. (2017). Invited review: Effects of heat stress on dairy cattle welfare. *Journal of Dairy Science*, *100*(11), 8645–8657. https://doi.org/10.3168/jds.2017-12651

Silva, R. G. da, Morais, D. A. E. F., & Guilhermino, M. M. (2007). Evaluation of thermal stress indexes for dairy cows in tropical regions. *Revista Brasileira de Zootecnia*, *36*(4 suppl), 1192–1198. https://doi.org/10.1590/S1516-35982007000500028

Sugiono, Dewi, H., & Rudy, S. (2016). Measuring Thermal Stress of Dairy Cattle Based on Temperature Humidity Index (THI) in Tropical Climate. *MATEC Web of Conferences*, *68*, 06004. https://doi.org/10.1051/matecconf/20166806004

TE Connectivity Inc. (2014). *HTU21DF RH/T Sensor IC: Digital Relative Humidity sensor with Temperature output*. https://www.te.com/en/product-CAT-HSC0004.datasheet.pdf

Thom, E. C. (1959). The Discomfort Index. *Weatherwise*, *12*(2), 57–61. https://doi.org/10.1080/00431672.1959.9926960

Triebe, O., Hewamalage, H., Pilyugina, P., Laptev, N., Bergmeir, C., & Rajagopal, R. (2021). *NeuralProphet: Explainable Forecasting at Scale*. http://arxiv.org/abs/2111.15397

Triebe, O., Laptev, N., & Rajagopal, R. (2019). *AR-Net: A simple Auto-Regressive Neural Network for time-series*. http://arxiv.org/abs/1911.12436

Vranken, E., & Berckmans, D. (2017). Precision livestock farming for pigs. *Animal Frontiers*, *7*(1), 32–37. https://doi.org/10.2527/af.2017.0106

Wang, X., Gao, H., Gebremedhin, K. G., Bjerg, B. S., Van Os, J., Tucker, C. B., & Zhang, G. (2018). A predictive model of equivalent temperature index for dairy cattle (ETIC). *Journal of Thermal Biology*, *76*(March), 165–170. https://doi.org/10.1016/j.jtherbio.2018.07.013

Zhang, M., Wang, X., Feng, H., Huang, Q., Xiao, X., & Zhang, X. (2021). Wearable Internet of Things enabled precision livestock farming in smart farms: A review of technical solutions for precise perception, biocompatibility, and sustainability monitoring. *Journal of Cleaner Production*, *312*, 127712. https://doi.org/10.1016/j.jclepro.2021.127712