Alma Mater Studiorum · Università di Bologna

DOTTORATO DI RICERCA IN SCIENZE E TECNOLOGIE AGRARIE, AMBIENTALI E ALIMENTARI
Curriculum in Ingegneria Agraria
Settore Concorsuale: 07/C1
Settore Scientifico Disciplinare: AGR/10

Numerical modeling for PLF in dairy cattle farms: methods for the analysis of environmental, productive and behavioral parameters

Supervisore:
Ing.
Stefano Benni

Co-Supervisore:
Prof.
Daniele Torreggiani

Coordinatore Dottorato:
Chiar.mo Prof.
Giovanni Dinelli

Presentata da:
Filippo Bonora

Ciclo XXX
Esame Finale Anno 2018

Abstract

English version

In the last years, marked by considerable social, environmental and economic changes, even the agricultural world has not been exempt from important and continuous transformations, like all aspects of everyday life. In particular, the steady increase of the world population, new eating and life habits, the climate changes and the related and growing attention to the environmental impact are transforming the production and the management of animal husbandry.

As already happened in other sectors, the most selected answer at these continuous challenges is the "technological innovation", in this field it is called Precision Livestock Farming (PLF). It is so becoming increasingly common to see in cattle barns, swine and poultry farms cameras, weather stations, collars, pedometers and other unusual sensors and instruments until a few decades ago in these places.

The study carried out in this thesis is part of this scientific field, particularly with regard to dairy cowshed. The objective of the work is to identify standard methods of analysis aimed at optimizing the management of the herd in terms of welfare and production of individual animals, but also to facilitate the daily work of the farmer.

Two selected case studies were examined in three macro-areas: 1) herd management optimization; 2) relationship between milk production and environmental conditions; 3) behavioral characteristics of the individual animal.

The obtained results underline the importance and usefulness of the devices installed today in the barns and the studies connected to them: if one can manage the enormous amount of data collected through ICT systems and transform it into clear information (for examples graphs), it is possible to create a system capable of optimizing the management of the farm in terms of better quality and greater quantity of product, animal welfare and simplified and more profitable working conditions.

Keywords: Animal Behavior, Dairy cattle, Environmental Parameters, Numerical Modeling, Precision Livestock Farming (PLF).

Italian version

Come tutti gli aspetti del vivere quotidiano, anche il mondo agricolo non è esente da importanti e continue trasformazioni in questi anni segnati da notevoli cambiamenti a livello sociale, ambientale ed economico. In particolare, il costante aumento della popolazione mondiale, le nuove abitudini alimentari, e più in generale di vita, da una parte e i cambiamenti climatici e le connesse e crescenti attenzioni verso l'impatto ambientale del proprio lavoro quotidiano dall'altra stanno trasformando anche la produzione e la gestione delle attività zootecniche di tutto il mondo.

Come già successo in altri settori, tra le varie e differenti soluzioni che le persone del mestiere stanno scegliendo come risposta a queste continue sfide occupa sicuramente uno dei primi posti il concetto di "innovazione tecnologica", che in questo campo prende il nome di zootecnia di precisione. Ecco quindi che diventa sempre più comune vedere telecamere, centraline meteo, collari, pedometri e altri sensori e strumenti non abituali fino a qualche decennio fa in stalle, porcilaie e allevamenti avicoli.

Lo studio effettuato in questa tesi rientra in questo ambito di ricerca, in particolare è incentrato sugli allevamenti di bovine da latte. L'obiettivo del lavoro è stato quello di individuare replicabili metodologie di analisi atte a ottimizzare la gestione della mandria in termini di benessere e di produzione del singolo animale, ma anche al fine di facilitare l'operato quotidiano dell'allevatore.

A questo scopo sono stati presi in esame due casi studio appositamente selezionati e sono state progettate e testate sperimentazioni che possono essere racchiuse in tre macro-aree: a) gestione ottimizzata della mandria, b) rapporto produzione - condizioni ambientali e, infine, c) abitudini comportamentali del singolo capo.

I risultati ottenuti sottolineano l'importanza e l'utilità dei dispositivi tecnologici installati oggigiorno negli allevamenti e degli studi ad essi connessi.

Se infatti si è in grado di gestire l'ingente ammontare di dati acquisiti mediante i dispositivi, e trasformarlo in pure e semplici informazioni, è possibile creare un sistema capace di ottimizzare la gestione dell'intera struttura lavorativa in termini di migliore qualità e maggiore quantità di prodotto, elevato benessere animale e condizioni lavorative semplificate e rese più proficue.

Contents

Abstract **i**

1 Introduction and aims of the study **1**

1.1 Introduction 1

1.2 Goals 2

1.3 Thesis structure 3

2 State of the art **5**

3 Materials and methods **17**

3.1 Study cases 17

3.1.1 Farm A 19

3.1.2 Farm B 22

3.2 Data monitoring 24

3.2.1 Devices 24

3.2.2 Data processing 26

3.3 Models 28

3.3.1 Cluster-Graph model 28

3.3.2 Correlation analysis 32

3.3.3 Stepwise Multilinear Regression 33

3.3.4 GAM (Generalized Additive Model) 34

3.3.5 Viola-Jones algorithm 36

3.4 Software 37

3.4.1 MATLAB 37

3.4.2 Gephi 37
3.4.3 R 38

4 Results 41
4.1 Cow Clustering 41
4.1.1 Seasonal Cow Clustering 41
4.1.2 Monthly Cow Clustering 48
4.2 Environmental conditions and milk production 54
4.2.1 THI and Milk Production 54
4.2.2 Production forecasting 56
4.2.3 Heat Stress variability on a single cow basis 60
4.3 Cow Detection 69

5 Conclusions 75

Bibliography 81

List of Figures

2.1 Number of lactating cows (pre/post AMS) 8
2.2 Average daily milk production per cow (pre/post AMS) 9
2.3 THI in Temperature and Relative Humidity axes 12
2.4 Combined parameters 13

3.1 Plan layout (farm A) 19
3.2 Resting area (farm A) 20
3.3 Feeding area (farm A) 20
3.4 Robotic milking system (farm A) 20
3.5 Milking area (farm A) 20
3.6 Ventilation (farm A) 21
3.7 Internal roof (farm A) 21
3.8 The herd and the cubicles (farm A) 21
3.9 External view (farm A) 21
3.10 Plan layout (farm B) 22
3.11 External view (farm B) 23
3.12 Feeding area (farm B) 23
3.13 External resting area (farm B) 23
3.14 Internal resting area (farm B) 23
3.15 Internal THI (farm A) 25
3.16 External environmental trend (farm A) 26
3.17 PCA of a single cow (farm A) 28

4.1 Gephi Cluster-Graph (farm A) 44

4.2 Cluster-Graph statistics (farm A) 45
4.3 Monthly Gephi plot: January 2016 (farm A) 49
4.4 Monthly Gephi plot: February 2016 (farm A) 49
4.5 Monthly Gephi plot: March 2016 (farm A) 50
4.6 Monthly Gephi plot: April 2016 (farm A) 50
4.7 Monthly Gephi plot: May 2016 (farm A) 50
4.8 Monthly Gephi plot: June 2016 (farm A) 50
4.9 Monthly histogram: January 2016 (farm A) 51
4.10 Monthly histogram: February 2016 (farm A) 51
4.11 Monthly histogram: March 2016 (farm A) 52
4.12 Monthly histogram: April 2016 (farm A) 52
4.13 Monthly histogram: May 2016 (farm A) 52
4.14 Monthly histogram: June 2016 (farm A) 52
4.15 Yearly trend of the main variable: Cw (farm A) 52
4.16 Yearly trend of the variable: My (farm A) 52
4.17 Yearly trend of the main variable: Ma (farm A) 53
4.18 Yearly trend of the main variable: Mr (farm A) 53
4.19 Yearly trend of the variable: $\#M$ (farm A) 53
4.20 Yearly trend of the variable: $d\alpha$ (farm A) 53
4.21 Yearly trend of the main variable: Ld (farm A) 53
4.22 Yearly trend of the variable: Pa (farm A) 53
4.23 Correlation between $ITHI$ and SMy (farm A) 55
4.24 Trend of z-scored $ITHI$ and SMy (farm A) 55
4.25 Training phase (farm A) 58
4.26 Validation phase (farm A) 58
4.27 Test phase (farm B) 59
4.28 Lactation spline (farm A) 61
4.29 Feed spline (farm A) 62
4.30 THI' Coefficients vs. Intercept (farm A) 63
4.31 Daily My and THI' cow 101 ($b < 0$) (farm A) 65
4.32 Daily My and THI' cow 148 ($b < 0$) (farm A) 65

4.33 Daily My and THI' cow 207 ($b < 0$) (farm A) 65
4.34 Daily My and THI' cow 249 ($b > 0$) (farm A) 65
4.35 Daily My and THI' cow 189 ($b > 0$) (farm A) 66
4.36 Daily My and THI' cow 236 ($b > 0$) (farm A) 66
4.37 $d\alpha$ of the three clusters (farm A) 66
4.38 Box of the three clusters (farm A) 66
4.39 $\#M$ of the three clusters (farm A) 66
4.40 $Cond$ of the three clusters (farm A) 66
4.41 Cw of the three clusters (farm A) 67
4.42 My of the three clusters (farm A) 67
4.43 Ld of the three clusters (farm A) 67
4.44 Sf of the three clusters (farm A) 67
4.45 $My/\#M$ of the three clusters (farm A) 67
4.46 Pa of the three clusters (farm A) 67
4.47 Mr of the three clusters (farm A) 68
4.48 The size of the three clusters (farm A) 68
4.49 $Temp$ values of the three clusters (farm A) 68
4.50 Vel of the three clusters (farm A) 68
4.51 Example a): application of the Viola-Jones (farm A) 71
4.52 Example b): application of the Viola-Jones (farm A) 71
4.53 Example c): application of the Viola-Jones (farm A) 72
4.54 Example d): application of the Viola-Jones (farm A) 72

List of Tables

2.1 Comparison pre/post AMS (pre/post AMS) 10

4.1 $\#M$ in k-means clusters (farm A) 42
4.2 Pa in k-means clusters (farm A) 42
4.3 $d\alpha$ in k-means clusters (farm A) 43
4.4 Mr in k-means clusters (farm A) 43
4.5 Cw in k-means clusters (farm A) 44
4.6 Cluster-Graph statistics (part 1) (farm A) 46
4.7 Cluster-Graph statistics (part 2) (farm A) 47
4.8 Cluster-Graph statistics (part 3) (farm A) 47
4.9 Cardinality of the monthly clusters (farm A) 49
4.10 Forecasting statistics (test phase) (farm A) 57
4.11 Forecasting results (farm A) 60
4.12 GAM groups (farm A) 64
4.13 Correlation between My and THI' (farm A) 64
4.14 Cow detection results (farm A) 73

Chapter 1

Introduction and aims of the study

1.1 Introduction

Livestock farming is facing up to big challenges and profound changes nowadays. The increase of the global demand, food alarms and new dietary trends are straining the farmers: on the one hand, they have to guarantee the welfare and adequate conditions of life for the animals and to reduce the environmental footprint; on the other hand, they have also to develop new strategies in order to improve the management of the farm reducing the costs. One of the most selected way to obtain these two goals is the technological development.

As the single aspects of our daily lives, the farms are becoming more and more modern and scientific. Many device systems are implemented both on the animals and on the buildings to control every single aspect of the management. Therefore, farms could be considered data generators with long rows of numbers coming from collars and accelerometers, video to control the animal positions, pedometers and weather stations etc.

In particular, the work of this thesis is inserted in the context of the Precision Livestock Farming, a macro topic including animal and computer science, engineering, maths and biostatistics. PLF utilizes technologies to monitoring many different aspects of animals (behavioral, productive, reproductive) in order to improve the management and the performance of the herd and to reduce the environmental footprint. These techniques are more and more applied in the farms and they are becoming very reliable and precise: their benefits are obviously more evident in the big farms where every worker has to monitor many animals. As reported by Koeleman (2017), the need of PLF in the farms can be considered driven by 5 trends in the industry and society:

1. "Farms have become larger and more complex in terms of management, with more animals per employee";
2. "There is a need to reduce costs and create higher value to improve competitiveness";
3. "Productivity needs to be improved to meet rising animal protein demand";
4. "The environmental footprint of farming needs to be reduced";
5. "Farmers need to address societal concerns about animal welfare, improving transparency in the sector".

1.2 Goals

The main aim of this PhD research is so the development, validation and calibration of numerical models useful for the livestock building design and management, focused on production efficiency, quality and sustainability of livestock production, in particular for what concerns dairy cattle farming. Therefore, the specific goals of the study are:

- The development and the test of innovative procedures for the comprehensive analysis of AMS-generated multi-variable and accelerometers time-series in order to support dairy livestock farm management, in particular a model to provide an automatic grouping of the cows based on production and behavioral features (subsection 3.3.1);

- The monitoring and the analysis of the relationships between milk production and climatic conditions. In particular:

  1. The investigation of the negative effects of THI (Temperature-Humidity Index) on the milk production (subsection 3.3.2);
  2. The design of a forecasting model based on the integration of milking data and temperature, humidity and wind speed levels (subsection 3.3.3);
  3. The study of the different characteristics of the animals in relation to their reactions to the heat stress (subsection 3.3.4).

- The test of an object detection algorithm analysis in order to verify the applicability in the livestock management to identify specific patterns in the behavior of the herd (subsection 3.3.5).

1.3 Thesis structure

In order to achieve all these objectives, different mathematical-numerical models have been applied in this work and they are explained in the next chapters. In particular, the thesis is structured as follow:

- Chapter 2 includes the state of art of Precision Livestock Farming, with a particular attention to the dairy cattle farms and the most recent and interesting developments and studies concerning the subjects of this research;

- Chapter 3 illustrates the materials and methods of the work. The description of the case study farms with all the technical characteristics about milking and management is the main aspects of the first part of the chapter. The second part is focused on data, models and software used in this research;

- Chapter 4 explains the obtained results with the help of plots and graphs highlighting the important consequences that could arise from a correct use of devices and technologies in this field;

- Chapter 5 summarizes the work and underlines possible future developments.

Chapter 2

State of the art

The rise of purchase of animal proteins (in particular the insertion of new markets as Asia and South America), with the increase of demands for animal products of 70% by 2050, the increment of the number of livestock but the decrease of the number of farmers, the need of productivity reducing the costs and the environmental and "animal social" impacts and other changes in the dairy industry (Barkema et al. (2015)) are the main factors causing the involvement of not only animal scientists and veterinaries in the livestock sector but also engineers, mathematicians and computer technicians (Berckmans (2017)).

Moreover, due to the reduction of the use of antibiotics from one side and, from the other, the high number of zoonosis diseases (Quammen (2012)) with the consequent attention to the safety and the quality of the farm products and the urgency of monitoring animal welfare and health, the multidisciplinary discipline called Precision Livestock Farming (PLF) is constantly growing.

PLF could be defined as "the use of information and communication technologies for improved control of fine-scale animal and physical resource variability to optimize economic, social, and environmental dairy farm performance"(Eastwood et al. (2012)).

Similarly, another possible definition is given by (Bartzanas et al. (2013)): "intelligent management and care of (individual) animals in livestock farming by continuous automated monitoring and controlling of the production and reproduction, health and welfare of (individual) animals, thereby allowing quick corrections when deviations from normal are monitored".

Therefore, the main expected benefits from PLF deal with real time monitoring of animal welfare and health, early disease alerting, increment of production, reduction in production costs, and improvement of farmers' work conditions and life quality (Berckmans (2014)): differently from many other fields, this supervision is very complex because, following the definition coming from the first European PLF Conference in 2003 (Berlin), the main subjects of this action are CITD systems (Complex, Individually different, Time-varying and Dynamic) (Berckmans (2017)).

PLF technologies are widespread in all the global livestock sector: so for example, Shao and Xin (2008) have realized a real-time image processing system to detect movement and classify thermal comfort state of group-housed pigs and Romanini et al. (2013) have defined a system for real-time monitoring of the egg shell temperature (OvoScanTM, Petersime) during incubation.

ICT (Information and Communications Technologies) and IoT (Internet of Things) have so become ever more popular also in agriculture and it is clear that their application in Precision Livestock Farming has increased very quickly in the last decades in the various livestock farming sectors, with peculiar applications for dairy cows barn.

In the past, it was not so difficult to monitor every single animal in a little-medium size tie-stall barn, but the introduction of the free-stall cowshed with the aim of enlarging the number of reared cows and increasing the productivity per employee has changed the management of the farm and the farmer has transferred his attention toward a more "herd-orientated" approach.

Nowadays, the installation of new devices and systems has the goal to go back to a specif analysis per single cow in order to spend time only for the problematic ones without using new employees.

Automatic Milking Systems are "one of the earliest precision livestock farming developments" (John et al. (2016)) and their introduction in late Nineties has deeply changed the barn layout and the herd management in dairy farms, as reported by de Koning (2010) and Rodenburg (2017).

The first studies about the automatic systems in dairy farms date back to the last years of the 1970s: the beginning tests have been conducted by the researchers of IMAG Institute of Wageningen (The Netherlands) in order to reduce the production costs and to improve the quality of the work conditions of the farmers. The first milking robots have been introduced into the market only in 1992 and their use have started to increase from 1998: in 2004 there were over than 2200 farms in the world with an AMS system (de Koning and Rodenburg (2004)) and more than 2400 in 2008 (Reinemann (2008)). One robot unit can manage 60-70 cows, with an increase of milk production per cow by 6-35% (due to the increment of the milking frequency) and a reduction of 20-30 % in farm labor (de Koning and Rodenburg (2004), Heikkilä et al. (2010)).

AMSs measure and record specific data about milk production and cow behavior, providing farmers with useful real-time information for each animal ("specific zoom"), but at the same time the remarkable amount of info stored in their database has also a great potential for herd characterization and management optimization ("general zoom").

All these data have also represented a very important source for the academic research in the livestock farming world. A significant example of the outcomes of the introduction of AMS technologies in a dairy farm can be provided by the analysis of the differences between "pre" and "post" AMS introduction in the quality and the quantity of the milk production.

In this regard, an example is hereafter provided concerning a study case analyzed in this research. The data from a farm (Azienda Aletti) located in Grontardo (Cremona, WGS84 coordinates 45°12'N 10°09'E, 46 m a.s.l) have been collected. The milking parameters about the herd before and after the use of AMS robot have been analyzed: in particular the years 2010 and 2011 represent the No-AMS period and the data from January 2012 to June 2016 are the ones with the milking systems. Now, the farm is equipped with two milking robots in order to rear around 150 dairy cows, with average values of 2.5 milking events/day and 36 liters/day per cow. It is important to notice that during the study period the herd has maintained regular characteristics (number of cows, building conditions ...), so it is possible to make a comparison (see for example Figure 2.1).

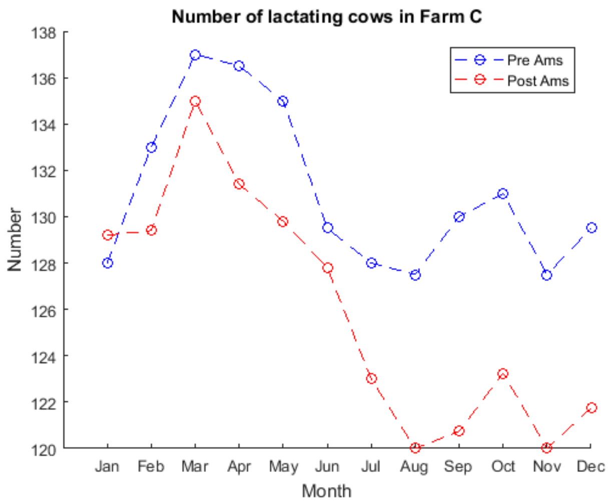

Figure 2.1: Number of lactating cows (pre/post AMS). The size of the herd does not change in a significantly way.

In Figure 2.2 and in Table 2.1, the mean values of the fundamental indices to analyze the herd pre and post the Automatic Milking System are shown.

From these results, it is easy to confirm two typical consequences of the introduction of the AMS already known in the scientific literature.

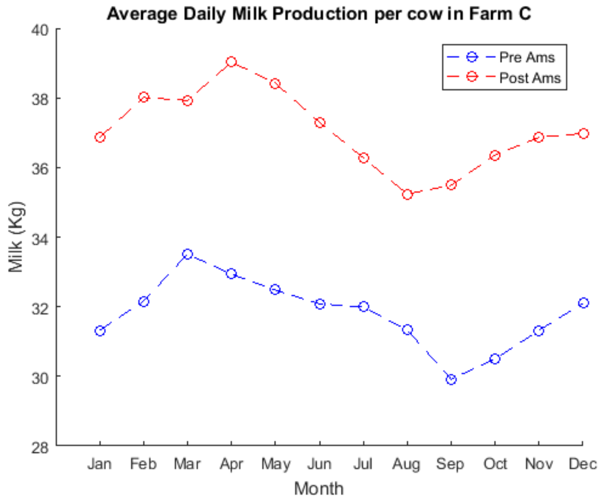

Figure 2.2: Average daily milk production per cow (pre/post AMS).

First, the most important trend is the increase of daily milk yield per cow (Hayashi and Kawamura (2002), Sitkowska et al. (2015)): in particular, after the introduction of AMS, in this farm the milk production has increased by 19%. Moreover, in a more general view, the generic trend of the milk production curve could be observed, i.e. the decrease in the warmest months (Figure 2.2).

In Table 2.1, on the other side, the principal change could be observed in somatic cells: in the first months after the installation of the milking system robot, the farmer has noticed a light increment of the somatic cells in herd (Castro et al. (2017), Poelarends et al. (2004)).

As described before, the introduction of automatic milking systems has changed the layout of the barn and, in particular, the cow traffic in the farm: the milking permission frequency, i.e. the frequency of the milking for each cow, is now up to the farmer and it can be also set individually for each animal. For example, Gaworski et al. (2016) has studied three different barn cow traffics:

- Free Cow (FT). In FT the animals have free access to the milking robot;

Table 2.1: Evaluation of the most important indices in the herd (pre/post AMS).

Index	Pre AMS	Post AMS
Milk/Cow	32 Kg	37 Kg
% Fat	3.8	3.7
% Protein	3.4	3.4
Dry Matter	21.7 Kg	23.2 Kg
Somatic Cells (x 1000)	222.7	268.8

- Milk First (MF). In MF there is a selection gate which gives access to the cows into the milking area only if they have the permission from the robot, otherwise the gate sends them into the feeding and the resting area;
- Feed First (FF). In FF there is a one-way gate from the resting area to the feeding area and then another "checkpoint" sends them into the milking area (if they have the permission) otherwise they have to come back into the resting area again.

Linked to these different types of cow traffic, another meaningful variable is the capacity, i.e. the calculation and the optimization of the milk yield per AMS unit per day (Gygax et al. (2007)).

An additional automation used in the farms to reduce labor time requirements is the Automatic Feeding System (AFS): the act of feeding requires the 16% of the total labor in a farm. The connection between these two systems (AMS and AFS) has been studied, for example, by Oberschätzl-Kopp et al. (2016): in this work, the influence of different feed delivery (six times per day vs. two times per day) in a herd milked by AMS has been investigated. The study permits the conclusion that the distribution of fresh food rations more often (six times per day) could be linked to a more efficiency for the AMS as well as to the animal welfare.

Besides at the automated systems, indoor and outdoor climatic conditions, especially in terms of temperature, humidity and wind speed, represent another important point of interest within a dairy livestock barn for the PLF. They are a well-known crucial issue in farm building design and management, since these parameters can remarkably influence cows behavior, milk yield and animal welfare.

The perfect habitat for a dairy cow is characterized by:

- Temperature: the cow is a warm blooded animal, so it can maintain a bodily temperature around 38.5°C independently of the external climatic conditions (if they are not so adverse). A dairy cow has the best productive performance in a thermo-neutral environment, i.e. when the animal does not need heat and so it could use all the available energy for the milk production. This climatic zone is defined between -5°C and 22-25 °C (Armstrong (1994), Kadzere et al. (2002));

- Relative Humidity: between 50% and 80% (Dragovich (1979)). A different climatic condition forces the animals to use energy to regulate the bodily temperature and not to produce milk;

- Wind: the animals do not require an intense flow, but just a sufficient fresh air in order to remove HN_3, CO_2 and methane from the barn, and the heat coming from the cow bodies. The movement of air layers in the barn can increase the body convective heat transfer and evaporation rate, independently of air temperature, but the real effect is difficult to estimate because it depends on many factors (body mass, air temperature, ...) (CIGR Section II Working Group N 14 (Cattle Housing) (2014)).

It is well known that many scholars have focused their attention on the relationships between cows and environmental conditions and, in particular, an index has been defined in order to have a real estimation of the authentic climatic impact perceived by the cows.

It is the THI (Temperature-Humidity Index), a very common parameter used to monitor the risk of heat stress for cows and production yield.

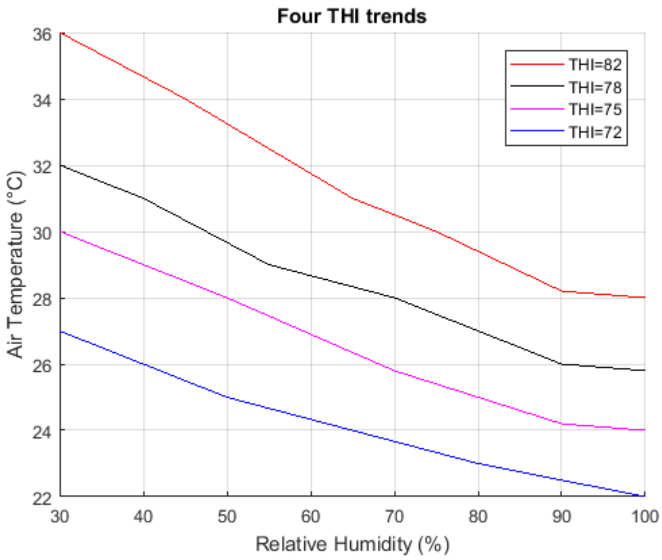

Figure 2.3: Reminding about the connection between relative humidity and dew point, the plot shows the trend of four THI values in a Relative Humidity-Temperature plot (CoolCows (2017)).

Following Asabe (1986), THI is defined as

$$THI = T + 0.36 * DP + 41.2 \tag{2.1}$$

where T is the temperature (° C) and DP is the dew point (° C).

In particular, Samal (2013) associated THI thresholds to significant levels of heat stress in cows:

- $72 < THI < 80$ causes a mild stress level, with increase in respiration rate and blood vessels dilatation;
- $80 \leq THI < 90$ causes a moderate stress level, with water consumption increase, body temperature growth and milk production decrease (from 1% to 20%);

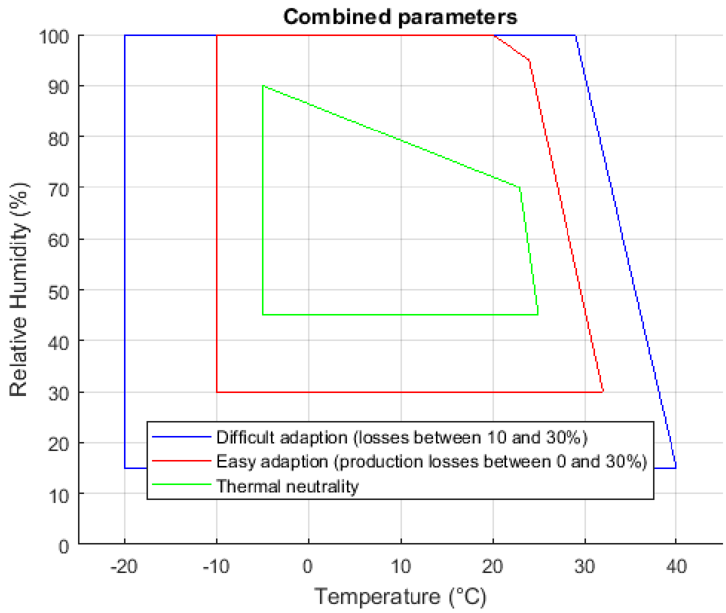

Figure 2.4: The plot represents the relationship between production and (relative humidity + temperature) at an air speed of 0.5 m/s. (CIGR Section II Working Group N 14 (Cattle Housing) (2014)).

- $90 \leq THI < 98$ causes a severe stress level, with high body temperature, panting, excessive saliva production and a marked decrease in reproduction and milk production;

- $THI \geq 98$ causes a very dangerous stress level with potential death.

Several studies have so provided evidence of a relation between heat stress and many different problems regarding the dairy cows:

- Behavior. The increase of temperature causes a change of cow behavior and an increment of the standing position with the following growth of the body temperature (Hillman et al. (2005), Overton et al. (2002));

- Genetics. West (2003) focused the attention on southeastern United States and showed interests in the link between heat stress and genetic advancements, with the selection for heat tolerance or the identification of genetic traits connected with it;

- Milk production and quality. Bouraoui et al. (2002) measured the heat stress effects on lactating Frisian-Holstein cows in the Mediterranean area and they found that summer heat stress reduces milk yield and also it alters milk composition. Mote et al. (2016) evaluated the effect of different macro climatic variables on lactation milk yield and lactation length. It was observed that maximum temperature has a significant correlation with lactation milk yield, whereas maximum temperature, minimum temperature, sunshine hours and wind speed have significant correlation with lactation length;

- Cow wellness and reproduction. Hagiya et al. (2017) examined the effects of heat stress (HS) on production traits, somatic cell score (SCS) and conception rate (CR) at first insemination in Holsteins in Japan. Summer calving caused the greatest increase in SCS and in the first and second lactations this increment became greater as THI rose. In cows, CR was affected by the interaction between HS group and insemination month: in the summer and early autumn insemination, there was a reduction in CR, and it was much larger in the mild- and moderate-HS groups than in the no-HS group.

There are two different kinds of techniques used to reduce the negative effects of the climatic conditions: the passive methods and the active ones. The passive methods are appropriate architectural materials or advantageous shapes and structures of the building: for example, it is possible to use coverage materials reflecting and converting the solar radiation in a new type of energy (Berdahl and Bretz (1997)). The natural ventilation (chimney and wind effect), the plant and the orientation of the farm building have also an important role for the shading. In particular, Collier et al. (1981) and Roman-Ponce et al. (1977) have shown that if shade treatments have been applied to lactating cows in the summer, the animals show a lower rectal temperature and respiration/min, and a higher conception rate. It is interesting to notice that there is also the possibility to install movable shading elements in a farm (Bucklin et al. (1991)).

On the other side, the active methods include artificial ventilation and refreshment systems (cooling by air and cooling by surface). In both of passive and active solutions, the ventilation plays a fundamental role to assure the wellness of the cows.

Finally, other studies have tested the use of specific devices and algorithms into barns to respond to the decrease of the direct human-cattle contact caused by the increase of the farm and herd size: pedometers (Mattachini et al. (2013)), video analysis (Kaihilahti et al. (2007), Mattachini et al. (2011), Noldus et al. (2001)), ear-attached movement sensors (Bikker et al. (2014)), instrumentation for cow positioning and tracking indoors (Huhtala et al. (2007)) are ever more adopted in farms, nowadays.

Anyway, despite this great interest, as reported by Steeneveld et al. (2015), the economic benefits from investments in technological systems come still more from the reduction in labor costs than from a more precision in the measure of health and wellness of the cows: there is still much to be done.

Chapter 3

Materials and methods

In the following sections the materials and methods of the thesis are presented. In the section 3.1 the technical features of the two case study farms are described and in the section 3.2 the devices installed and the data processing are illustrated. The different models applied in this work are shown in the section 3.3 and a summary of the mathematical softwares used in the research is reported in the section 3.4.

3.1 Study cases

Two farms (A and B) have been selected as testers in this thesis. The decision is in according to the following points:

- Milking system. The aim of the study was to develop methods that can be applied independently of the adopted milking system. Therefore, it has been selected a farm equipped with the automatic milking system (farm A) and another one with a conventional milking parlor (farm B) in order to describe both of the milking management;

- Size/dimension of the herd. The two selected farms have different dimensions: farm B is the biggest ones and it has the capacity to host 270 lactating cows; in the farm A around 65 cows are reared;

- Location in Po valley. First, this region is really affected by the climate changes and it is becoming one of the hot spot in the world (Segnalini et al. (2011)): it could be thus considered a significant area to study the heat stress in dairy cows.

  Moreover, as reported by the Italian National Institute of Statistics (Istat) in 2015, Emilia-Romagna is the second most productive Italian region of cow milk (the 14.7% of the Italian milk is yielded here).

  Finally, Po valley is the Italian leader of the emerging and increasing diffusion of the new technological devices in the farms. It has been so considered a proper area to conduct the study;

- Innovation. Both of the selected farmers are really open-mind and willing to experiment new techniques and to collaborate with the research world. In particular, farmer B is also building a new barn section in collaboration with the University of Bologna in order to renew the farm.

3.1.1 Farm A

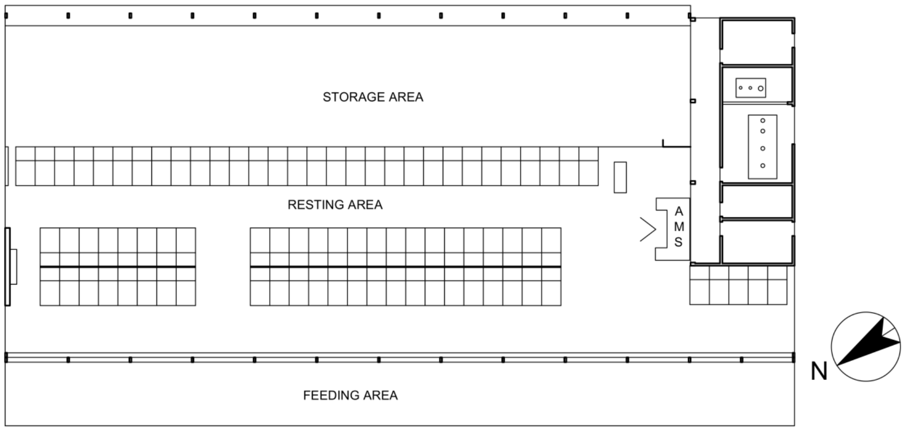

Figure 3.1: Plan layout.

Farm A (Azienda Piazzi) is placed in the municipality of Budrio, about 15 km north-east of Bologna (WGS84 coordinates $44°33'32.7''N$ $11°31'09.7''E$, 25 m a.s.l.). The barn (whose layout is shown in Figure 3.1) is a 51 m long and 23 m wide rectangular building, with SW-NE-oriented longitudinal axis, consisting of a hay storage area on the SE side, a resting area in the central part of the building (Figure 3.2), and feeding area and a feed delivery lane on the NW side (Figure 3.3).

The resting area, whose floor is partially slatted, hosts 78 cubicles with straw bedding where about 65 Frisian cows are housed. Two blocks of head-to-head rows are located in its central part, and another row runs along the entire length of the resting area.

The milk-room is located on the SW side of the building, next to the offices and technical plant rooms (Figure 3.4 and 3.5). Cow milking is performed by means of a robotic milking system "Astronaut A3 Next" by Lely (Maassluis, The Netherlands), marked as "AMS" in the right part of Figure 3.1.

Figure 3.2: Resting area.

Figure 3.3: Feed delivery lane.

Figure 3.4: AMS robot.

Figure 3.5: Milking area.

Figure 3.6: Ventilation system.

Figure 3.7: Internal roof.

Figure 3.8: The herd and the cubicles.

Figure 3.9: External view of the farm.

The robot is programmed to assure a number of daily visits for each cow depending on the cow productivity and its expected optimal milk yield per visit, with a minimum and a maximum number of daily visits as constraints.

The ventilation is controlled by three high volume low speed (HVLS) fans with five horizontal blades which are activated by a temperature-humidity sensor situated in the middle of the barn (Figure 3.6).

Data collected from farm A have been analyzed for the clustering procedures, for the training and the validation phases of the correlation analyses, for the development of GAM models and for the cow detection algorithm, as described in the following subsections.

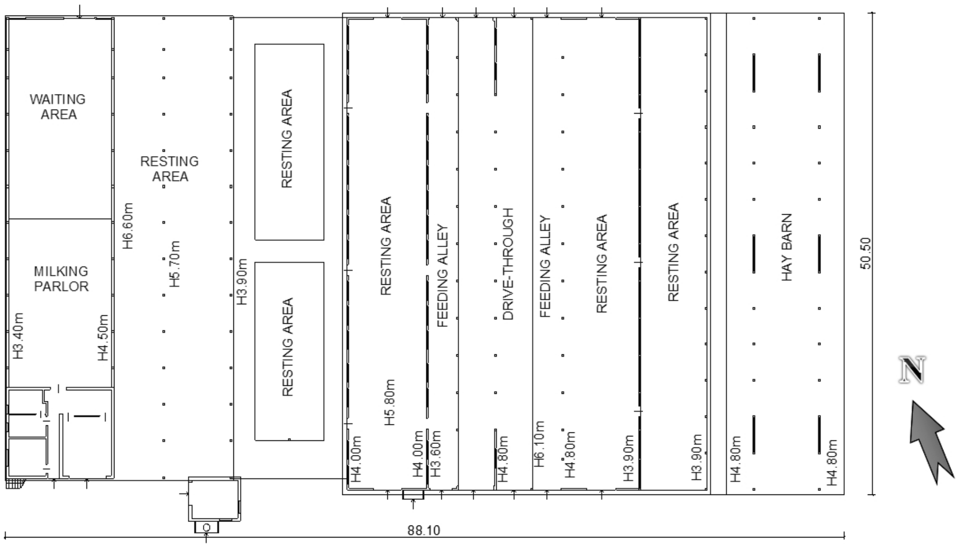

Figure 3.10: Plan layout of farm B.

3.1.2 Farm B

Farm B (Azienda Montagnini) is located in the municipality of San Pietro in Casale, about 25 km north of Bologna (WGS84 coordinates $44°42'59.2''N$ $11°27'04.9''E$, 17 m.a.s.l.), see Figure 3.10 and Figure 3.11.

This farm has the capacity to host 270 lactating Frisian cows and the barn is arranged into three sectors located into three adjacent barns built in different periods:

- A barn, built in the 2000s, with effective systems of natural and forced ventilation hosts 90 cows in the first three months of lactation;
- A masonry barn built in 1986 and an adjacent shelter with steel structure and double metal sheet roof where 110 cows in the intermediate lactating period (4 months) are reared;
- The third and last sector on the SE side is for 70 cows in the last three lactating months and it is in a barn built in the 1990s.

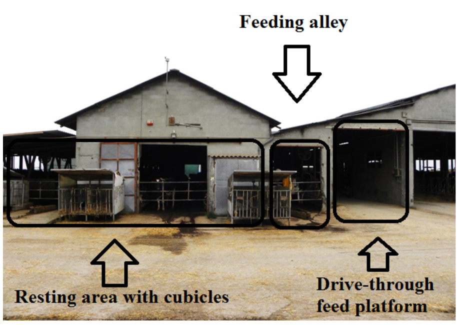

Figure 3.11: External view.

Figure 3.12: Feeding area.

Figure 3.13: Resting area (1).

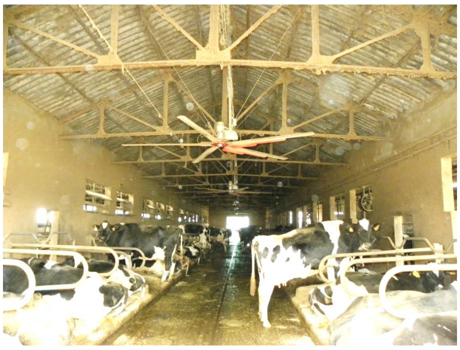

Figure 3.14: Resting area (2).

The second sector has shown some problems of ventilation and cooling, and it has been therefore monitored for this study and for a broader research in collaboration with the farm. It has concrete floor with scrapers, two rows of cubicles and the feeding alley are inside the masonry building and all cubicles have straw bedding.

The milking parlor, a 2x15 herringbone, and the milk room have been built in the 2000s and they are in the NW side of the building.

Data collected from farm B have been adopted for the test phase of the correlation analysis model, as described in the following sections.

3.2 Data monitoring

3.2.1 Devices

The following five devices have been used for this study:

1. AMS robot (farm A). After every milking events, it records data about milk quality and quantity and it also manages the supplement feeding. The accesses or the refusals of the cow depend on the average milk production, the milk yield prediction and the day of lactation;

2. Temperature and humidity sensors (farm A and farm B). Internal temperature, humidity and dew point are measured and recorded every 30 minutes in both farms by two PCE-HT71 stand-alone data loggers, which have resolution of 0.1° C and accuracy of ±0.5° C (see an example of the trend in the Figure 3.15).

   In the farm A two devices are located in the central cubicle rows, at a height of 1 meter from the ground.

   In the farm B, three sensors have been positioned respectively at the center of the indoor resting area, at the center of the outdoor cubicles row and in the middle of the feeding alley (at 2.2 m from the ground in all cases);

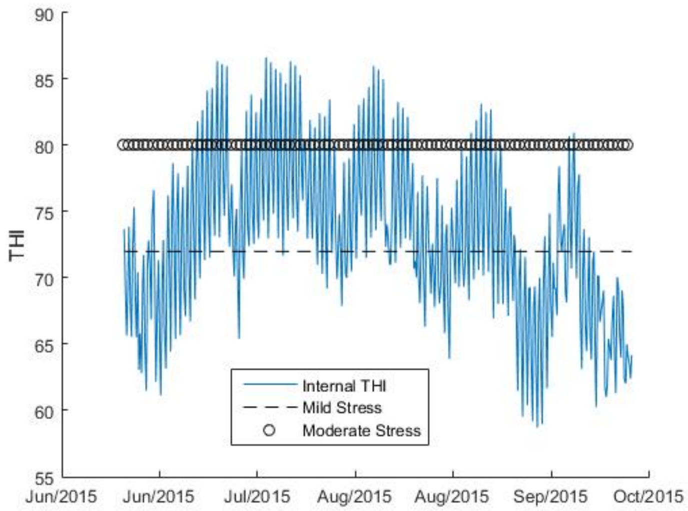

Figure 3.15: Internal THI (farm A) with the thresholds of moderate and mild stress.

3. Weather station (farm A). External climate data have been recorded by a PCE-FWS20 weather station: it has been installed besides the barn and it measured temperature, humidity, rain amount and rate, solar radiation, wind speed and direction every 30 minutes (see an example of the trend of each variable in the Figure 3.16).

   Climate data concerning farm B have been acquired from a weather station located 7 km from the barn (Environmental Service ARPAE of the Emilia-Romagna Region);

4. Activity collar (farm A). Cows behavior data have been measured in 2 hours blocks by means of a collar by SCR (Netanya, Israel) mounted for cow identification and activity sensor. It monitors activity levels (α) of each animal by means of an acceleration sensor measuring movement duration and intensity.

   This variable is widely used in livestock management for automated heat detection (Shahriar et al. (2016));

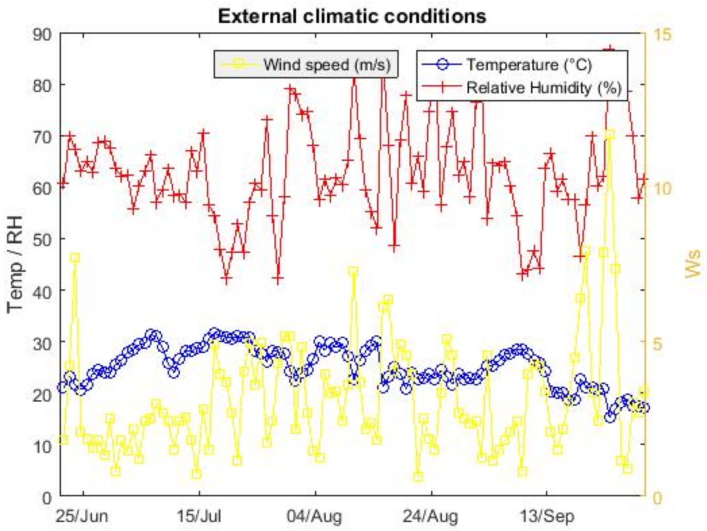

Figure 3.16: Temperature, relative humidity and wind speed trends in the summertime in the farm A.

5. Camera (farm A). It has been positioned above the cubicles in order to start a preliminary study of the animal behavior in the farm through the application of an object-detection algorithm.

3.2.2 Data processing

As described in the subsection 3.2.1, cow-related and milk production data of farm A have been recorded by the AMS at each cow passage. Based on the database downloaded from the AMS management software, a matrix called 'Visit' has been created where each row corresponds to a cow passage and the columns contain the following parameters:

- Cow identification number (Cin);
- Date and time of the cow passage (tcp);
- Milk Yield (My) in liter;
- Day of lactation (Ld);

- Supplement Feeding (*Sf*), supplementation with additional concentrates in kilogram. Total daily amount is calculated by the AMS based on the daily milk yield and the day of lactation;
- Cow body weight (*Cw*) in kilogram;
- Parity (*Pa*);
- Mastitis (*Ma*): in every single milking event, it is equal to 1 if the AMS identifies a mastitis, otherwise it is equal to 0. The final output is the average of the variable, i.e. it represents the probability of mastitis for a cow in every milking;
- Average of daily conductivity (*Cond*). The conductivity values are usually estimated to prevent mastitis;
- Average of daily milk-robot velocity (*Vel*) in liter/minute;
- Average of milk temperature (*Temp*) in °C;
- Total time spent in milking box (*Box*) in second;
- Milking regularity (*Mr*) in hour. It represents the standard deviations between two milking events.

Data are downloaded also from the collars at each passage through the AMS robot and they have been collected for farm A in a matrix called 'Activity', where each row contains the two-hour activity of each cow. 'Visit' and 'Activity' matrices and the internal and external climatic data have been finally jointly processed.

Before the application of the models, the three datasets have been analyzed and examined accurately in order to confirm the correct download of the files: for example, in the Figure 3.17 a PCA (Principal Component Analysis) applied on data of a single cow is shown. Here,the results show the already known high connection between climatic data and activity. Then, it is possible to notice also that the milking production and supplementary feeding are correlated (it depends obviously on the definition of Sf).

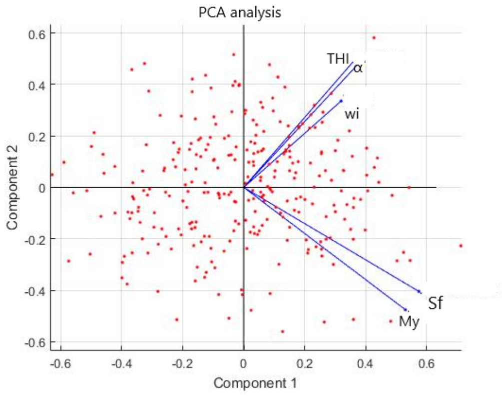

Figure 3.17: PCA (Principal component analysis) for a single cow in the period June 2014 - August 2015. THI, wind speed (wi), α and My of a single cow have been studied.

3.3 Models

In the following subsections the principal models applied in this work are explained. Summer has been selected as the head time period in the majority of exemplifying case studies: as described above, heat stress is the bitter enemy for farmers (Garcia et al. (2015), Lessire et al. (2015), Rhoads et al. (2009), Smith et al. (2013)), not only for milk production but more in general for the welfare of the animal (for example, it has negative effects also on reproduction (Pavani et al. (2015))).

3.3.1 Cluster-Graph model

As reported by Adamczyk et al. (2017), the great diversity and individuality of the dairy cows and their related management makes difficult to compare and analyze their behavior and characteristics. Despite the fact that it has been applied only in few researches, a cluster analysis could be therefore an important tool to identify objectively their welfare: it allows to compare the characteristics of a single cow in a littler group composed only by similar animals.

In particular, differently from Adamczyk et al. (2017) where the Ward's method and the Kohonen's networks have been selected as algorithms to study the herd, in this thesis a cluster-graph model has been applied. It converts the huge amount of data available in the modern farms in an unique visual output in order to give farmers useful and immediate information.

In the following subsections two linked clustering approaches are presented: the first one takes in account an entire season, the second is a more dynamic version and it "takes a photo" of the herd in every single months.

Seasonal Clustering

Cow clustering based on production and behavioral features has been performed on data surveyed in a whole season. Here summer 2015 (from June 21st to September 30th) of farm A has been selected as time period for the study. In this case the clustering has been performed on an overall number of 88 cows: it is worth noticing that this is number of cows accounts for all the animals reared in the barn in the considered study period.

The k-means algorithm (MacQueen (1967)) has been used to study the average of the following parameters on a daily basis in the study period:

1. Number of daily milking ($\#M$);
2. Pa;
3. Average daily activity ($d\alpha$);
4. Mr;
5. Cw.

A cow is so represented by a single value (the mean) for each parameter in the above k-means analysis. The study of five representative values for the entire season allows to describe the cow characteristics of various months in a stationary way.

For each variable, different k values have been selected a posteriori to highlight some particular trends. Applying this procedure for all the five variables above, different subgroups of the starting set have been found: in particular, $C_i^j(\cdot)$ represents the value of the cluster j for the parameter i and c_i^j its centroid.

Then, the clusters obtained through the k-means algorithm have been joined in a network graph (Barabási (2015)) with Gephi (subsection 3.4.2): the network has been designed assigning each cow a node and linking two nodes if the cows belong to the same cluster (at least one). The weight of the link, W, between two general nodes A and B is defined by the summation of the 5 "similarity index" S_i as follow,

$$W(A,B) = \sum_{i=1}^{5} S_i(A,B) \tag{3.1}$$

where i identifies one of the previous five variables, A and B represent two analyzed cows and S_i is calculated for each parameter i as in Equation 3.2.

$$S_i(A,B) = \begin{cases} 1 - \frac{|C_i^j(A) - C_i^{j'}(B)|}{c_i^j}, & \text{if } j = j' \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

The graph is then analyzed to find subnetworks basing on modularity, i.e. a measure that minimizes the number of edges between two different clusters (Newman (2006)).

In summary, the procedure is composed by three steps:

1. Five k-means analysis, one for each parameter described above;
2. Creation of a network with nodes and edges and definition of a measure of connectivity;
3. Analysis of the network in order to find subnetworks.

Differently from a classic k-means cluster on five indicators, this methodology allows:

- To monitor every single parameter choosing the most appropriate k for each of them;
- To design a network and so to convert numbers and matrices in nodes and colors, i.e. a more immediate and easy way to monitor the characteristics of the herd. Moreover, the clustering approach based on modularity does not require to define "a priori" how many clusters the dataset must have.

Monthly Clustering

The same procedure described in the previous subsection could be applied in a shorter time period after appropriate modifications. The seasonal clustering has allowed to characterize the herd from a static and general viewpoint; on the other side the monthly one examines the animals in a more detailed way and it allows to find some interesting trends.

This second cow clustering based on production and behavioral features has been focused on data surveyed from January 2016 to June 2016 in farm A and applied to the overall number of lactating cows which were reared in the barn in the considered study period (obviously the number of animals is different in every single month).

Here, the k-means algorithm has been adopted to study the following parameters for each cow:

1. $\#M$;
2. Pa;
3. $d\alpha$;
4. Mr;
5. Ld.

In order to maintain the same structure of the previous procedure, five variables have been selected also in the monthly clustering. In particular, in this case, the parameter Ld has been preferred to Cw: the study period is composed by only one month, so the days in milk have been considered more distinctive and useful to describe the herd than cow weight. It was not possible to insert Ld in the seasonal clustering because it loses all its descriptive potentiality if it is averaged on an entire season.

For each variable, different k values have been selected “a posteriori” to highlight specific trends. In particular, the same ks decided in the seasonal clustering have been used also in this evaluation for the number of milking events per day, milking regularity and activity. For Pa, $k = 3$ has been selected on the basis of the results of the seasonal study (subsection 4.1.1), which indicate that three main groups can clearly be recognized. It is also known that three fundamental phases can be identified in the typical lactation curve with transitions corresponding to about 90 and 210 days, therefore k has been selected equal to 3 also for the variable days in milk in the monthly clustering.

The clusters obtained through the k-means algorithm have been joined in a network graph through Gephi as described in the previous subsection with the same definitions, weights and procedures.

3.3.2 Correlation analysis

As mentioned before, several studies have investigated the influence of climatic conditions on milk production. In particular, Bohmanova et al. (2008) analyzed the duration of the time lag between heat stress and production loss. This research found that in USA the climatic conditions 3 days before the considered milking show stronger correlation with the variability of milk production than those of 1 or 2 days before.

In this thesis, a similar test has been conducted in the summertime 2015 (July 1st - September 14th) through the analysis of the Bravais-Pearson correlation coefficient ρ of the following two variables in the farm A:

- Exponential Moving average of daily internal THI values ($ITHI$).

  Bohmanova et al. (2008) used the "Heat stress degree" (Hsd), defined for each time t as

  $$Hsd(t) = \begin{cases} 0, & \text{if } THI(t) \leq 72 \\ THI(t) - 72, & \text{if } THI > 72 \end{cases}$$

  Here, the $ITHI$ value is computed as the exponential moving average of the daily mean of THI values recorded inside the barn every 30 min between 12 pm and 6 pm. This time interval has been chosen as it proved to be the period of the day when THI most significantly exceeds the heat stress threshold;

- Simple moving average of the difference of daily values of milk production of the herd (SMy) with a certain forward time lag from the days considered for the computation of the variable $ITHI$. Such time difference (in days) has been defined through correlation analysis.

3.3.3 Stepwise Multilinear Regression

After the analysis described in the subsection 3.3.2, a model for forecasting the milk yield of the herd has been implemented by means of a stepwise multilinear regression: here data about outdoor climatic conditions - i.e. wind speed wi (m/s), relative humidity rh (%), temperature T (° C) and time D (positive integers as days, i.e. 1= 1st January, 32 = 1st February) along the study period - are assumed as the predictive terms, while the milk yield of the herd (My) represents the dependent variable. These variables are selected to understand how much the environmental conditions affect the milk production.

The data recorded in farm A in summer-autumn 2014 (from June to December) have been used as the training set (196 daily data, derived from a dataset of 1860 milking records and 9408 temperature records), while the data of farm A recorded in summer 2015 (from June to August) have been adopted as the validation set (74 daily data from a dataset of 13270 milking records and 3552 temperature records). Finally, the data of summer 2015 (from June to August) of farm B have been used as the test set (74 daily data from a dataset of 148 milking records (2 milkings per day) and 1776 temperature records).

The starting expression of the model is the multi-linear relationship of the Equation 3.3 for the generic time t for a generic cow with coefficients a_0 (intercept), a_1 (wind speed), a_2 (relative humidity), a_3 (temperature) and a_4 (day of the year):

$$My(t) = a_0 + a_1 * wi(t) + a_2 * rh(t) + a_3 * T(t) + a_4 * D(t) \tag{3.3}$$

Then this equation evolves in a new form depending on the analyzed data, as shown in subsection 4.2.2.

3.3.4 GAM (Generalized Additive Model)

As described above, heat stress is one of the hardest enemy for a farmer in the summer and heat waves have repercussions on wellness, milk production and reproduction of the cows. In the following lines a GAM (Generalized Additive Model) clustering procedure has been applied on data surveyed in the summer 2016 in the farm A in order to characterize three different groups of animals in according to their response to the heat stress (severe/moderate/mild stress). The results could be useful to indicate the most suffering animals in the barn in order to help them with particular treatments (shading, cooling, ...).

In according to Yano et al. (2014), the Equation 3.4 has been applied for each cow in order to analyze the relationship between the milk production and THI.

In this case, the dataset contains the values surveyed in summer 2016 (from June to September) in farm A: in particular, 44 cows have selected for the study, i.e. the total number of cows without blanks or strange quantities of the previous parameters, and enough monitoring days (70% of the total time period).

$$ln(My_i(t)) = a'_i + b'_i * THI'(t) + s1_i(Ld_i(t)) + s2_i(Sf_i(t)) \qquad (3.4)$$

where

- i represents a generic cow;
- t is the analyzed day of the year;
- $My_i(t)$ is the milk yield of the cow i in the day t;
- $THI(t)'$ is the average of the THI calculated in the 12 pm - 6pm of the day t. This time period has been selected because it has proved to be characterized by persistent thermo-hygrometric conditions causing heat stress for cows;
- $Ld_i(t)$ is the lactation day of the cow i in the day t;
- $Sf_i(t)$ is the robotic feed of the cow i in the day t;
- $s1_i$ and $s2_i$ are two smoothing spline functions for the cow i.

$s1_i$ and $s2_i$ have been estimated for each cow from the data using the Generalized Addictive Modeling: GAM is a generalized linear model in which the relationships between the independent variables and the dependent one are not linear, but they are simply described by regulated and non parametric functions (Hastie et al. (2009), James et al. (2013)). Therefore it could be considered as a combination of a linear model and the "black box" of the machine learning.

Then, spline1 and spline2 are calculated as the average of the each s_i in order to have two generic functions for the herd. Therefore, the Equation 3.4 has been updated with the aim of estimating the coefficients a_i and b_i with a curve fitting algorithm:

$$ln(My(t)_i) = a_i + b_i * THI'(t) + s1(Ld(t)_i) + s2(Fi(t)_i). \qquad (3.5)$$

In the results section, the value of a_i and b_i are analyzed and three different groups have been found and characterized.

3.3.5 Viola-Jones algorithm

The Viola-Jones algorithm (Viola and Jones (2001)) is the model selected for the preliminary study for the cow detection. It has been developed by Paul Viola and Michael Jones for the identification of the faces in 2001, then it has been improved by Rainer Lienhart and Jochen Maydt to recognize also rotated objects. The algorithm allows to analyze an image in a very fast way and it is a widespread solution also for what concerns the dairy cattle farms (Porto et al. (2013), Porto et al. (2015)). It is essentially based on the following four points:

- A new representation of the image, "integral image", allows a quick elaboration and computation of the features;
- The AdaBoost algorithm and the Haar transform improve the learning and survey ability for a more efficient classification;
- The classifiers are trained with "positive samples" (where the selected object is represented) and "negative samples" (what it is not searched) to recognize if in a section of the analyzed image there is what the algorithm is trying to find;
- The cascade architecture allows to remove and to not analyze the areas of the image where is highly unlikely to find the searched object.

The work showed in this thesis could be considered a very first step of a research that can be useful for many different goals. For example, it could help the farmer for an automatic and real time monitoring of locomotion and posture behavior of specif animals (i.e. pregnant cows (Cangar et al. (2008)) or also the control of the cow comfort status (Cook et al. (2005), Haley et al. (2000)).

This kind of studies are also used to analyze different indices (hourly or daily) measuring the wellness of the cows, as the following ones for example:

- Cow Comfort Index (CCI) $= \frac{\#\text{ cows in the cubicles}}{\#\text{ herd}}$;
- Cow Stress Index (CSI) $= \frac{\#\text{ standing cows}}{\#\text{ herd}}$;
- Eligible Cow Comfort Index (ECCI) $= \frac{\#\text{ cows in the cubicles}}{\#\text{ cows not in the feeding area}}$.

3.4 Software

3.4.1 MATLAB

MATLAB (MATrix LABoratory) is a software for the numerical methods that allows to use matrices, functions and plots in an easy and fast way developed by MathWorks: it is one of the most used scientific software for research purposes. The first official version has been released in 1984 and it has written in C, but it has an own language (Otto and Denier (2005)). In this thesis, it has been used especially to manage the big amount of CSV files coming from the robot and different MATLAB toolboxes have been also very useful. For example, the regressions and smoothing procedures have been applied with the Curve Fitting app and the Computer Vision app has been utilized for the Viola-Jones algorithm.

3.4.2 Gephi

Gephi is an open source software for networks and complex systems for Windows, MacOS and Linux (Bastian et al. (2009)). It has written in Java on Netbeans platform and it has been selected by the Google Summer of Code for five consecutive years (from 2009 to 2013). The penultimate stable release (0.9.1) has been presented in February 2016 and it has been used in this research.

Gephi is one of the most widely-known software to explore and manipulate networks and, for all intents and purposes, it is considered and used as a complementary tool to the traditional statistics in many different fields (Barabási and Bonabeau (2003), Leetaru (2011)).

Its interface is very intuitive and constituted by three work tabs:

1. Overview, where the graph could be manipulated and the principal network measurements could be calculated;
2. Data Laboratory, where all the nodes and edges info are listed;
3. Preview, where it is possible to edit the graph and export it.

Gephi can import many graph file formats: in this thesis GEFX (Graph Exchange XML Format) files have been used. A GEFX file is usually so structured:

- The first part is composed by general information about the graph (software version, date, author, ...);
- The central section is the list of the nodes with their attributes (Cin, Pa, ...);
- The final segment is the list of the edges with their weights.

3.4.3 R

R is an open source software (GNU GPL License) for statistical computing and analysis. It allows to manage easily datasets, to produce graphic outputs and to use different statistical tools, but also suites for vectors and matrices.

R has been developed by Robert Gentleman and Ross Ihaka (Department of Statistics) in the University of Auckland (New Zealand) with the aim of creating a free high quality system for statistical purpose: it has been written in C, Fortran and R (Dalgaard (2008)) and the first official version has been released in 1993. R is available for Unix, Linux, Windows and MacOS and it is widespread in the academic world, but its importance is also increasing in the companies one.

In this study, the software has been used useful to apply the GAM model (Wood (2006)), as described in the subsection 3.3.4.

Chapter 4

Results

In the following sections the results coming from the application of the models described in the section 3.3 are shown.

4.1 Cow Clustering

Both seasonal and monthly methodologies have demonstrated to be effective tools to control the herd. They allow to monitor the single animal as the whole herd and to make comparisons in a very fast and precise way.

4.1.1 Seasonal Cow Clustering

As seen in the subsection 3.3.1, the first step of the Cluster-Graph procedure is to identify the subgroups for the five selected parameters.

The results of herd clustering of farm A according to the single descriptive variables have been reported in the following tables:

- The cluster analysis has provided four groups of cows with different milking habits in terms of mean daily events (Table 4.1);

Table 4.1: Cardinality and minimum / maximum / median values and centroids for each cluster about the mean number of daily milking events ($\#M$).

Cluster	Cardinality	Min	Max	Median	Centroid
$C1_{\#M}$	10	1.2	1.7	1.5	1.5
$C2_{\#M}$	38	1.8	2.3	2.0	2.0
$C3_{\#M}$	30	2.3	2.8	2.6	2.6
$C4_{\#M}$	10	2.9	3.7	3.0	3.1

Table 4.2: Cardinality of each cluster about the number of parity (Pa) i.e. number of different times a female has had offspring.

Cluster	Cardinality	Number of parity
$C1_{Pa}$	35	1
$C2_{Pa}$	33	2
$C3_{Pa}$	10	3
$C4_{Pa}$	7	4
$C5_{Pa}$	2	5
$C6_{Pa}$	1	6

- The herd has been subdivided into six groups depending on parity (Table 4.2). The group composed by first-calf cows ($C1_{Pa}$) and that of cows with parity equal to 2 ($C2_{Pa}$) have similar cardinality and together represent the largest part of the herd (more than 77 %);

- In the Table 4.3 the herd has been split into four groups of animals with similar cardinality and significantly different levels of activity;

- Table 4.4 points out the different habits of herd for the act of milking, in terms of standard deviations of time interval between two milking events in succession;

Table 4.3: Cardinality and minimum / maximum / median values and centroids for each cluster about the mean daily activity ($d\alpha$).

Cluster	Cardinality	Min	Max	Median	Centroid
$C1_{d\alpha}$	21	28.4	44.5	40.7	39.6
$C2_{d\alpha}$	24	45.8	54.9	49.7	50.2
$C3_{d\alpha}$	19	55.3	65.9	59.3	60.0
$C4_{d\alpha}$	14	66.8	83.8	72.1	72.2

Table 4.4: Cardinality and minimum / maximum / median values and centroids for each cluster about the standard deviations of time interval (in hour) between milking events in succession (Mr).

Cluster	Cardinality	Min	Max	Median	Centroid
$C1_{Mr}$	4	0	1.3	0.7	0.7
$C2_{Mr}$	39	1.7	3.0	2.5	2.5
$C3_{Mr}$	26	3.2	4.4	3.7	3.7
$C4_{Mr}$	14	4.5	6.3	5.2	5.2
$C5_{Mr}$	5	7.0	9.6	7.3	7.7

- Finally Table 4.5 spotlights the differences within the herd in terms of body weight.

Then the obtained clusters have been joined in a graph and the subnetworks have been found through the modularity procedure. The clusters resulting from the final modularity process are shown in Figure 4.1 where the graph is drawn based on the 'Force Atlas' layout (Repulsion strength = 10.000 and Gravity = 400 and adjusted by sizes), based on Noack's edge-directed force layout (Gephi Consortium (2011)). The size of each node is proportional to daily milk yield: the higher is the milk yield, the longer is the node radius. Furthermore, parity is indicated as a number inside each circle.

Table 4.5: Cardinality and minimum / maximum / median values and centroids for each cluster about the mean cow weight(Cw).

Cluster	Cardinality	Min	Max	Median	Centroid
$C1_{Cw}$	28	541.8	589.4	570.1	568.6
$C2_{Cw}$	24	591.6	641.4	609.6	612.9
$C3_{Cw}$	18	649.3	698.7	676.7	676.5
$C4_{Cw}$	15	710.0	743.9	725.3	726.4
$C5_{Cw}$	3	761.5	792.9	768.8	774.4

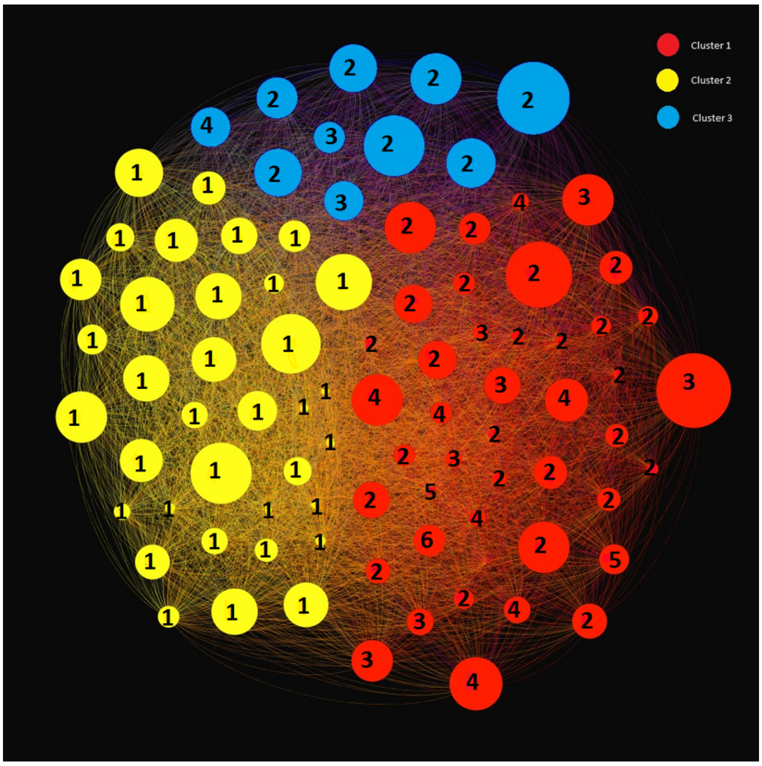

Figure 4.1: Cluster of farm A with Gephi. The graph has 5920 edges and 88 nodes, one for each cow. The number inside each circle represents the parity of each cow.

The mean values of every parameter for the three clusters, together with the respective standard deviations, are shown in Figure 4.2.

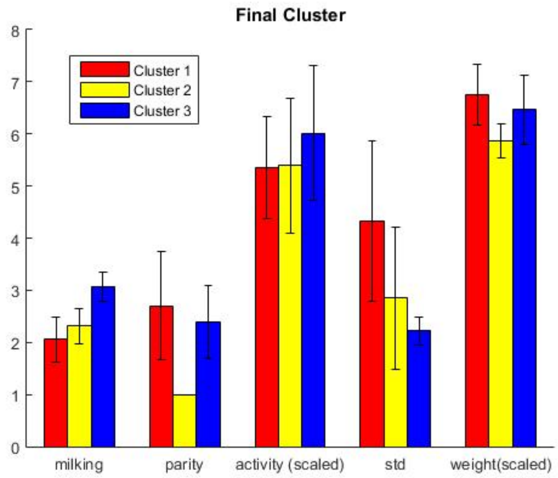

Figure 4.2: Mean and standard deviation of each final cluster: $\#M$, Pa, $d\alpha/10$, Mr and $Cw/100$.

The resulting subdivision of the herd shows a clear differentiation of each cluster from one another for what concerns all the considered descriptive parameters. The main diversification concerns parity as far as cluster 2 shows a mean of one, which indicates that the group includes only first-calf cows. This separation confirms the results of Hart et al. (2013) where the different behaviors of primiparous and multiparous cows are highlighted and the feasible benefits coming from the grouping and the management of the animals based on parity are underlined.

In particular, in the farm A the first-calf cows of the cluster 2 have intermediate mean values of number of milking events and milking regularity, while their mean body weight is the lowest, according to expectations.

The remaining two groups have similar mean parity, between two and three, and mostly differ as for number of milking events and milking regularity. More specifically, cluster 1 has a small number of effective AMS visits (daily average of 2) with poor regularity (the standard deviation of time interval between visits is almost 5 h): this cluster includes nearly half of the animal population.

On the contrary, cluster 3 has a significantly smaller cardinality than the other two groups, but it is strongly characterized by good milking performances. These are expressed by the average number of daily AMS visits which exceeds three and by their good regularity, given by a standard deviation of the time intervals which is about one half of that of cluster 1. It is interesting to observe that the mean activity score of this cluster is significantly higher than that of the other ones, while the average body weight is lower than cluster 1.

Important confirmations of the diversification of the three clusters in terms of productive characteristics are provided in Tables 4.6, 4.7 and 4.8, which contain also the averages and standard deviations of daily milk yields of the clusters, besides their cardinality. The milk yield has not been selected as a parameter for clustering because it has been considered the dependent variable according to which the effectiveness of the clustering procedure and the productive feature of each resulting group of cows have been assessed. It easy to notice that not only the clustering identified groups of animals with different behavior and physical conditions, but it also provided clusters with clearly different average productivity.

Table 4.6: Statistics for the final cluster (part 1): mean (m) and standard deviation (std) of M, Pa and $d\alpha$.

Cluster	$m(\#M)$	$std(\#M)$	$m(Pa)$	$std(Pa)$	$m(d\alpha)$	$std(d\alpha)$
1	2.0	0.4	2.7	1.0	53.5	9.8
2	2.3	0.3	1.0	0.0	53.9	12.9
3	3.0	0.3	2.4	0.7	60.2	13.0

The average milk yield of the whole herd is 27 liters per day and it can be considered as a reference value to assess the results of each cluster. In fact, cluster 3 performs much higher milk yield than clusters 1 and 2.

Table 4.7: Statistics for the final clusters (part 2): mean (m) and standard deviation (std) of My in liter and Mr in hour.

Cluster	*m(My)*	*std(My)*	*m(Mr)*	*std(Mr)*
1	24.7	10.0	4.3	1.5
2	27.0	10.4	2.9	1.3
3	36.7	7.0	2.2	0.3

Table 4.8: Statistics for the final clusters (part 3): mean (m) and standard deviation (std) of Cw in kilogram and Cardinality.

Cluster	*m(Cw)*	*std(Cw)*	Cardinality
1	674.5	58.0	43
2	587.4	33.0	35
3	646.4	66.8	10

The performance of cluster 2 is almost 10 liter/day below cluster 3 and represents the intermediate result. The negative difference between this value and the total mean is likely to be due to the condition of first delivery of the cows. The worst average performance is given by cluster 1, which mostly includes animals with scarce productivity.

The results highlight that lactating cows have very diversified conditions and the cluster analysis is an effective tools to identify the most significant groups, provided that, as it is in case of AMS use, data about animal behavior and conditions are available.

The identification of clusters can contribute to define feeding strategies based not only on the milk yield and lactation period, but also on the other descriptive variables, which prove to be effective in characterizing herd groups with different production potentials.

The results suggest also that particular attention should be paid to single animals whose milk yield is poorly consistent with the average value of the cluster they belong to. In particular, cows in cluster 1 showing high milking values should be monitored in order to prevent that their production decrease to the common values of that cluster. In these cases, proper measures should be defined to increase the number of daily visits to AMS and their regularity in time.

Another aspect deserving attention is represented by cows of cluster 3 showing milking production significantly below the cluster mean. Proper investigations should be performed on these animals to identify the causes of low milking and to identify possible corrective strategies.

Finally, cows of cluster 2 with the lowest production rates should be checked to verify if it is due to their normal lactating curve pattern or there are other factors hampering the expected productivity.

The resulting clusters are farm-dependent, but the methodology developed - consisting in the computational procedure and the approach to critical analysis of results - has general validity and represents a supplemental tool for a real time highly informative knowledge of herd's conditions.

4.1.2 Monthly Cow Clustering

Differently from the seasonal one, the monthly cow clustering model allows to organize the herd in groups based on data from the 30-days period of the study time span. It only requires the definition of a "main variable" in order to link the clusters in the successive months.

Here, a monthly clustering procedure has been applied on data surveyed in the first six months of 2016 in the farm A. The main variable adopted in the following figures is the cow identification number, i.e. every month, after the identification of the clusters, the color and the name of each found group has been selected in order to maximize the number of cows remaining in the same subdivision.

In the Figures 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, the light-colored lines are the edges linking two different nodes, and every cow is a node and its color represents a specific cluster: Blue = cluster a, Green = cluster b and, finally, Magenta = cluster c.

Table 4.9: Cardinality of the three clusters (a, b and c) in the first six months of 2016.

Month / Cluster	a	b	c
January	18	19	21
February	17	21	21
March	25	23	13
April	9	27	23
May	23	20	6
June	22	11	21

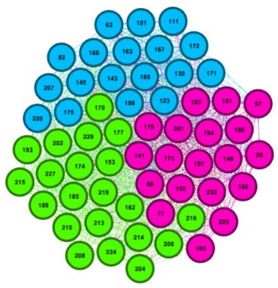

Figure 4.3: Monthly Gephi plot: January 2016.

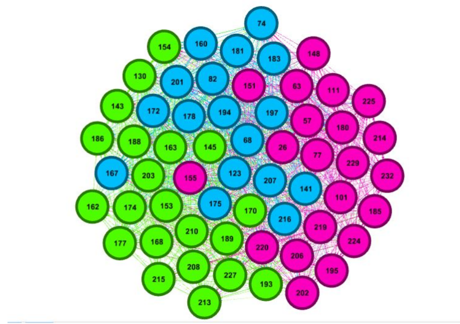

Figure 4.4: Monthly Gephi plot: February 2016.

It is interesting to notice that despite this cluster procedure does not require the definition "a priori" of the number of searched clusters (i.e. the k in the k-means algorithm), three groups have been exactly identified in each month (see Table 4.9).

Then, in the Figures 4.9, 4.10, 4.11, 4.12, 4.13 and 4.14 the monthly values of the main characteristics of herd are analyzed with a histogram representation. These plots confirm the capacity of the procedure to identify the most significant groups in according to the data from AMS and collars.

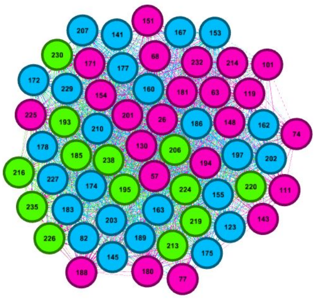

Figure 4.5: Monthly Gephi plot: March 2016.

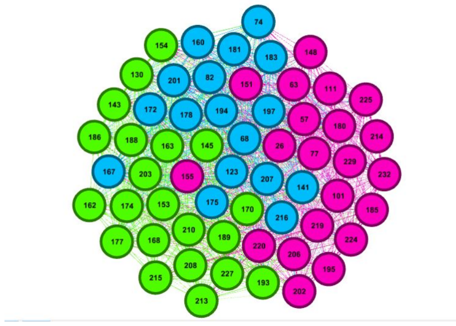

Figure 4.6: Monthly Gephi plot: April 2016.

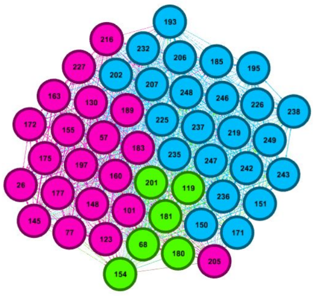

Figure 4.7: Monthly Gephi plot: May 2016.

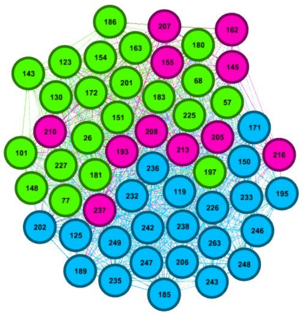

Figure 4.8: Monthly Gephi plot: June 2016.

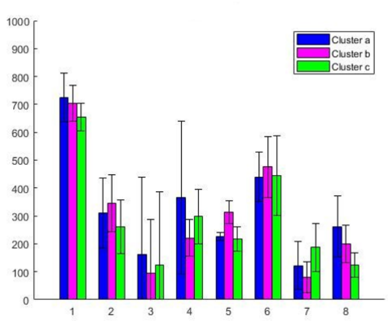

Figure 4.9: Monthly histogram: January 2016.

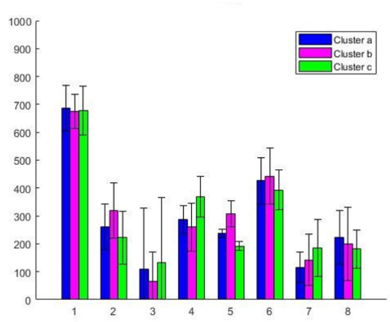

Figure 4.10: Monthly histogram: February 2016.

In particular, the averages and the standard deviations of the following variables are shown:

1. Cw;
2. Mean Daily of My in deciliter;
3. mMa, i.e. Ma x 10^{-3};
4. hMr, i.e. Mr x 10^2;
5. h$\#M$, i.e. $\#M$ x 10^2;
6. da$d\alpha$, i.e. $d\alpha$ x 10;
7. Ld;
8. hPa, i.e. Pa x 10^2.

Finally, Figures 4.15, 4.16, 4.17, 4.18, 4.19, 4.20, 4.21, 4.22 illustrate the summary of each parameter for one year in terms of monthly averaged values, thus they represent the herd trends of each cluster. The data from the last six months of 2015 have been added in order to have a more complete curve for all the selected parameters.

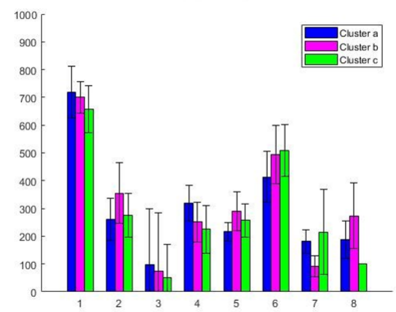

Figure 4.11: Monthly histogram: March 2016.

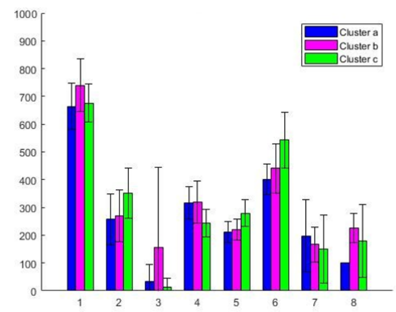

Figure 4.12: Monthly histogram: April 2016.

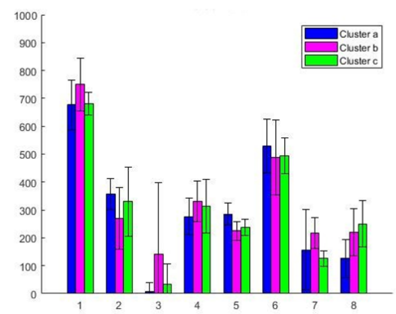

Figure 4.13: Monthly histogram: May 2016.

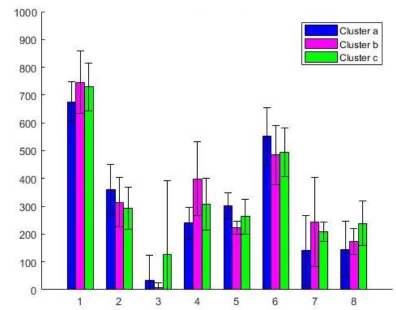

Figure 4.14: Monthly histogram: June 2016.

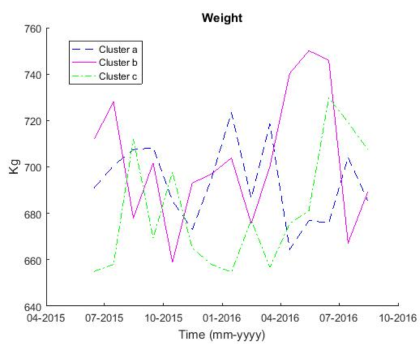

Figure 4.15: Yearly trend of the main variable: Cw.

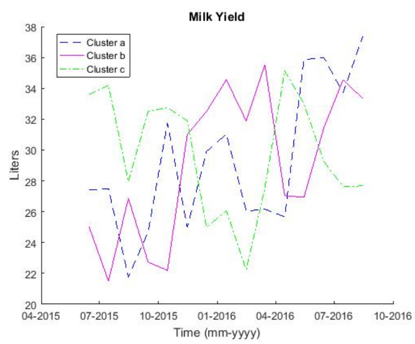

Figure 4.16: Yearly trend of the main variable: My.

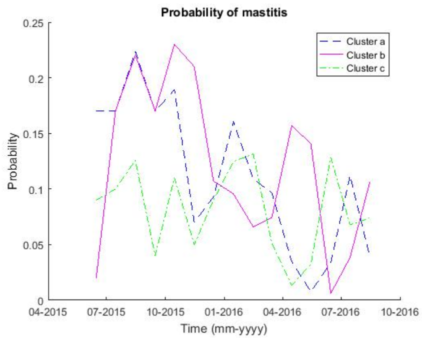

Figure 4.17: Yearly trend of the main variable: Ma.

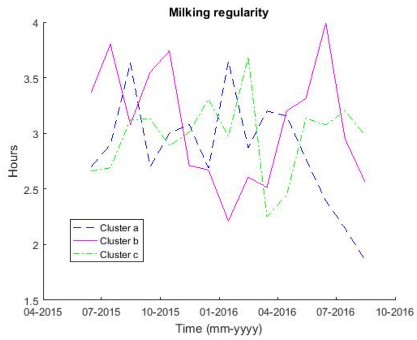

Figure 4.18: Yearly trend of the main variable: Mr.

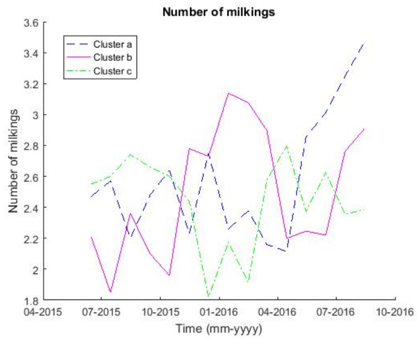

Figure 4.19: Yearly trend of the main variable: $\#M$.

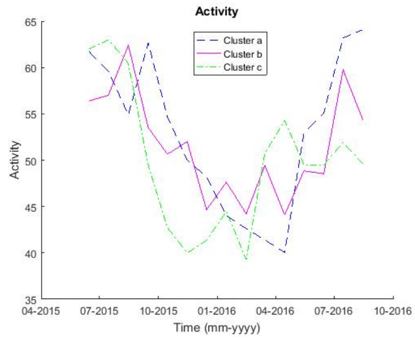

Figure 4.20: Yearly trend of the main variable: $d\alpha$.

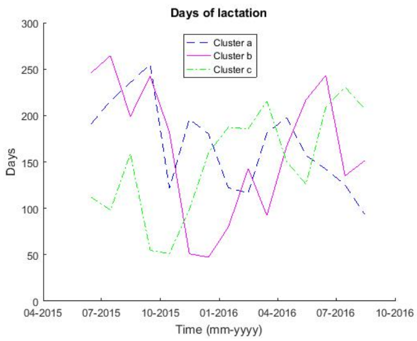

Figure 4.21: Yearly trend of the main variable: Ld.

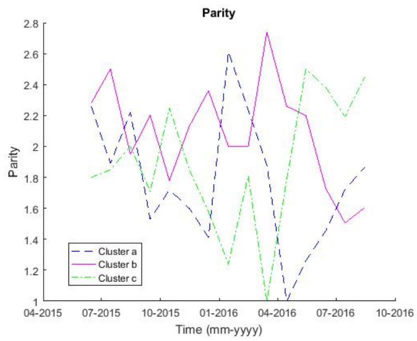

Figure 4.22: Yearly trend of the main variable: Pa.

The study of the dynamic distributions resulting from the analysis allows to analyze the animals in specific time intervals and to compare their conditions with those of previous months, thus achieving a detailed control of the herd. Moreover the trends of the parameters defining the clusters provide a dynamic characterization of the herd.

4.2 Environmental conditions and milk production

4.2.1 THI and Milk Production

The effects of climatic conditions on milk production have been analyzed by testing the correlation of the two vectors $ITHI$ and SMy with different time lags (from 1 to 7 days of difference) and recorded in farm A in the summer 2015. As shown by Figure 4.23, the maximum correlation (in absolute value) is for lag of 5 days, where a Bravais-Pearson coefficient $\rho =$ -0.7 have been obtained, with significance level of correlation of 0.05. It means that the negative effects of the heat stress do not occur within the same day when the causing environmental conditions take place, but especially in the summer they could affect the life of the cows for many days. In particular, the five lags show the maximum value of the negative correlation, but it is possible to notice high values also for many days after (for lag = 12, $\rho < -0.5$).

The negative correlated trend between $ITHI$ and SMy could be also noticed from the Figure 4.24. Here the two z-scored curves are plotted (i.e. for each day t, $X'(t) = \frac{X(t) - \text{mean}(X)}{\sigma(X)}$, where σ is the standard deviation and X' is the z-scored X) in order to facilitate the comparison between them.

Then, the same relationship between the two variables in the previous summer (2014) has been checked and a correlation with $\rho = -0.5$ for the fifth day has been obtained.

A further test has been carried out also on farm B, where THI and production data analyzed are referred to the same time period.

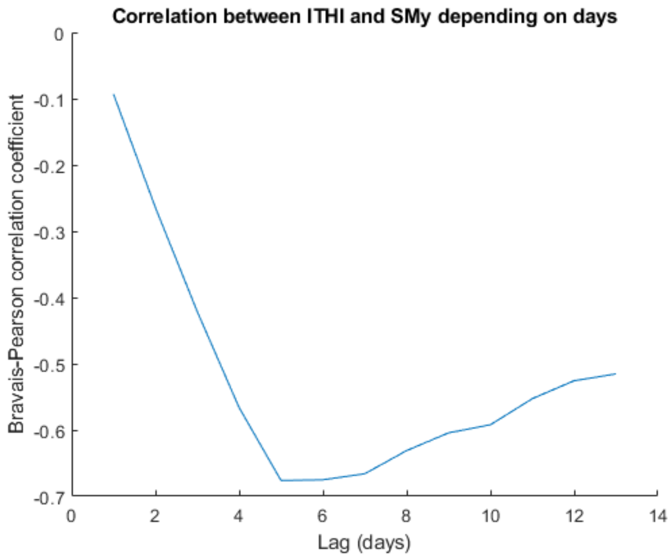

Figure 4.23: Correlation between $ITHI$ and SMy depending on days. The lag-days are reported in the x-axis and the Bravais-Pearson coefficient ρ in the y-axis.

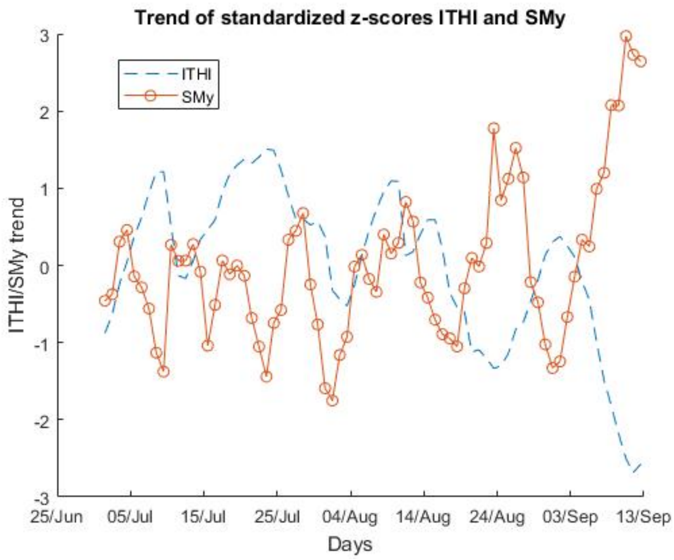

Figure 4.24: Trend of standardized z-scores $ITHI$ and SMy.

The results show also that in this case, when about 230 Frisian cows were housed in the barn under investigation, a significant correlation exists for time lag of 5 days, as ρ is -0.6.

Therefore, it is possible to affirm that the milk production is not affected significantly only by the daily THI, but also the reduction in milk production appears dependent on the high levels recorded some days before. This result differs slightly from that of the study by Bohmanova et al. (2008), which was carried out considering the average values of manifold barns in the USA. The difference may be due to the different geographic and environmental conditions and to the specific features of the Italian study case considered. In Bernabucci et al. (2014), the most significant effects were found for 3-4 days before the test day and the estimated duration of the THI effect is 8 days. The light differences could be caused by the difference of the dataset (time period and location of the farms).

In conclusion, the results confirm that THI is not a simple daily problem, but it could have negative shock waves for more than one week on the milk production.

4.2.2 Production forecasting

The correlation highlighted in the subsection 4.2.1 has confirmed the opportunity to develop a predictive model of milk yield for a generic cow depending on the environmental conditions.

In particular, a multilinear stepwise regression applied to temperature, relative humidity and wind speed has been selected as algorithm in order also to identify which variables have a higher impact on the milk production in the study period.

After the training phase, the model expressed by Equation 3.3 has provided the p-values displayed in Table 4.10. The p-value is the probability of finding the observed, or more extreme, results under the assumption of null hypothesis. In a stepwise regression, the null hypothesis is that the parameter would have a coefficient equal to zero if it is computed in the model.

Table 4.10: Statistics for forecasting (test phase). Status "In" means that the variable has been considered in the final model; "Out" means not considered in the final model.

Variable	P-Value	Status
Temperature	8.01e-13	In
Relative humidity	1.48e-01	Out
Wind Speed	1.93e-01	Out
Day	1.20e-25	In

Based on such values, T and D have been selected as variables to forecast milk yield in farm A and so only the regression coefficients a_0, a_3 and a_4 have been estimated. The results of the algorithm appear reasonable: the wind speed effects have been probably mitigated by the building, as the animals are always indoor; the exclusion of relative humidity can be explained by the its high correlation with temperature (Bravais-Pearson correlation equal to 0.7).

The relationship expressed by Equation 3.3 has thus become:

$$My(t) = 41.2 - 0.37 * T(t) - 0.007 * D(t) \tag{4.1}$$

It describes the relation between daily milk production of a generic cow (My) and temperature (T) in a specific time period (D) in the summertime and it could be used for prevision purpose.

The goal of this analysis is not to predict the milk production of every single day but it is to know and understand the general trend of the milk yield for a cow in the summer: the estimated and the real curves have been so smoothed in order to make the two curves flatter.

Different moving averages have been tested and, following the above correlation results, a time lag of 5 days has been adopted in the Figure 4.25 and in the Figure 4.26, as it resulted in the maximum correlation in farm A.

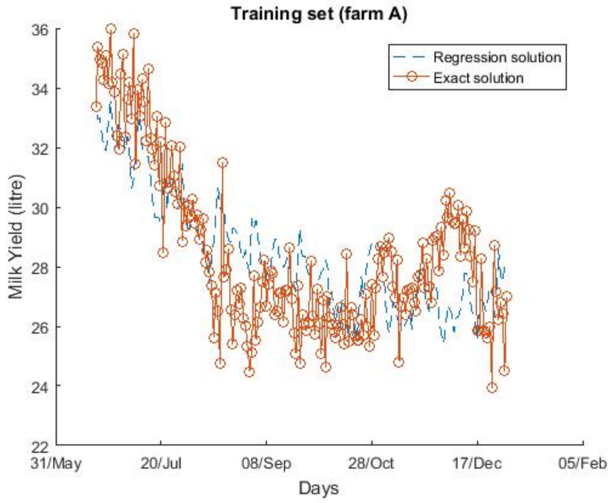

Figure 4.25: Training phase in farm A. This plot shows the training phase (June 19th - December 31st 2014) with regression and exact curve of daily milk yield for a generic cow in the analyzed barn.

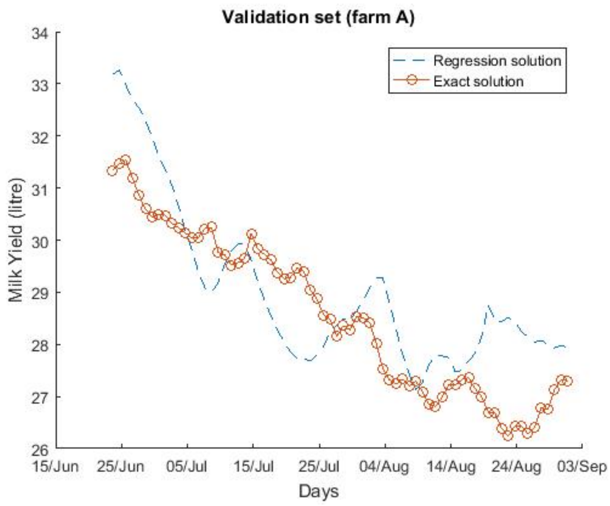

Figure 4.26: Validation phase in farm A. The plot displays the trend of regression and real data of the daily mean of the milk yield for a generic cow in the barn in summer 2015.

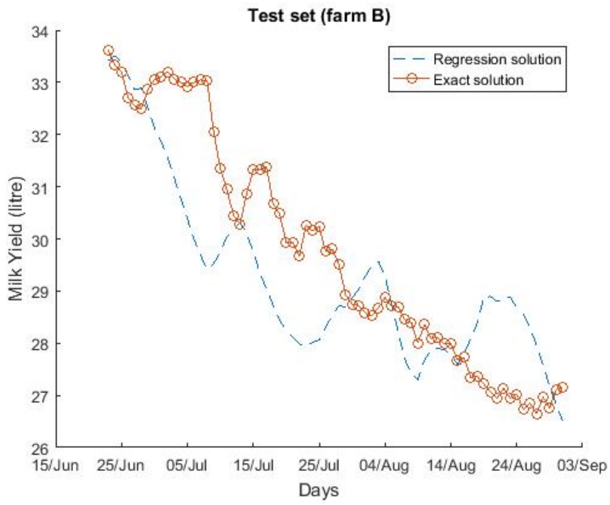

Figure 4.27: Test phase in farm B. This plot shows the test phase: after training and validation, the model has been tested with summer 2015 data in a different farm (farm B).

The mean errors of milk yield foreseen by the model and their standard deviations have been calculated and they are respectively equal to 1.7% and 3.7%. Therefore, in case an uncertainty range of 2% is considered acceptable, this analysis represents a successful validation test.

Furthermore, the farm B has been used to test the model (Test set): Equation 4.1 has been applied to the new climatic data of farm B in 2015 in order to forecast the milk yield of the whole summer. Also in this case a moving average procedure has been applied on the two curves to smooth the trends: in particular the same time lag has been selected because the idea is to apply only one model for both of the farms.

The results are illustrated in Figure 4.27, which shows a substantial agreement of theoretic results with measured data. Mean error is 1.7% with a standard deviation of 4.7%, therefore the test has confirmed the reliability of the model. The result is satisfying as far as the goal of this study was not to define a daily prediction, but rather a general trend without significant errors on a weekly basis.

Finally, Table 4.11 summarizes the statistics of the analysis performed in the three cases (Training, Validation and Testing):

- R^2 (Coefficient of Determination);
- RMSEP (Root Mean Square Error of Prediction);
- MAE (Mean Absolute Error);

Table 4.11: Statistics about forecasting in the three cases (Training, Validation and Testing).

Phase	R^2	RMSEP	MAE
Training (farm A, 2014)	0.51	6.83%	5.54%
Validation (farm A, 2015)	0.47	3.98%	3.35%
Testing (farm B)	0.58	4.95%	3.95%

The results can lend support to cow monitoring, through the comparison between the actual and the expected cow behavior. The forecasting model can be applied to assess the potential milk loss due to heat stress in a given climatic context, and therefore it represents a tool suitable for a direct assessment of the expected benefit that can be achieved through proper investments for the control of the indoor climate of a dairy barn. It also could be implemented to study milk yield trends depending on the expected environmental conditions.

4.2.3 Heat Stress variability on a single cow basis

A Generalized Additive Model (GAM) has been also selected to study the relationship between the THI and the herd, in particular the milk yield production. Differently from the previous results, this analysis is more "single animal" oriented and it does not focus only on the milk production, but it controls all the cow variables.

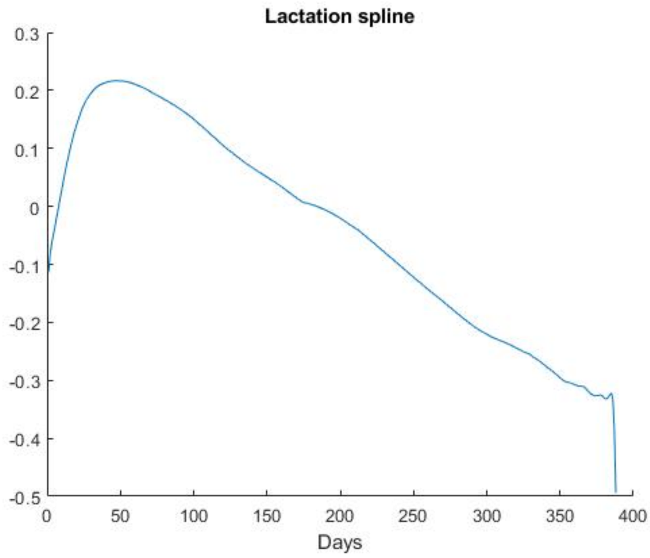

Figure 4.28: The plot shows the spline1, i.e. the smooth spline function linking the day of lactation and the daily milk production.

As illustrated in the subsection 3.3.4, in the first step of this methodology a general spline for the lactation day and a general spline for the supplement feeding are calculated in according to the milk production.

Figures 4.28 and 4.29 show $s1$ and $s2$, i.e. the trend of the milk production in according to the days of lactation and supplementary feeding respectively.

In the lactation spline it is possible to recognize easily the three different sections, already known in the literature:

1. First period. Here from one side there are the highest values of the milk production, on the other one there is a scarce capability for what concerns the ingestion and the digestion, so the farmer has to pay many attentions (even more than usual) to the feeding in order to avoid extreme weight losses.

   It usually corresponds to the first 90 days of the lactation, in this farm it is lightly shorter;

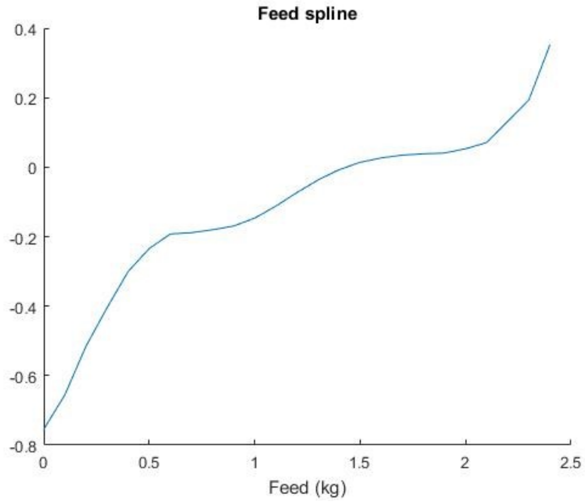

Figure 4.29: The plot shows the spline2, i.e. the smooth spline function linking the supplementary feeding and the daily milk production.

2. Second period. In these days (91-210 from the literature, here is shorter) the milk production is lower than in the previous phase and the cow re-balances her weight;

3. Third period. It is the last phase before the dry period, it is important to avoid an excessive increase in weight.

The feed spline is clearly connected with the decisions of the farmer and the veterinary about the AMS feeding. In the analyzed farm it depends on the day of lactation and the daily milk production.

After that, the coefficients a_i and b_i have been calculated for every cows through the Equation 3.5 and, in the Figure 4.30, the values of the two parameters for each cow are plotted.

In according to them, the herd has been split in two categories:

- $b < 0$ (red stars), i.e. animals suffering significantly due to heat stress;

- $b > 0$ (green stars), i.e. animals apparently not affected by high THI.

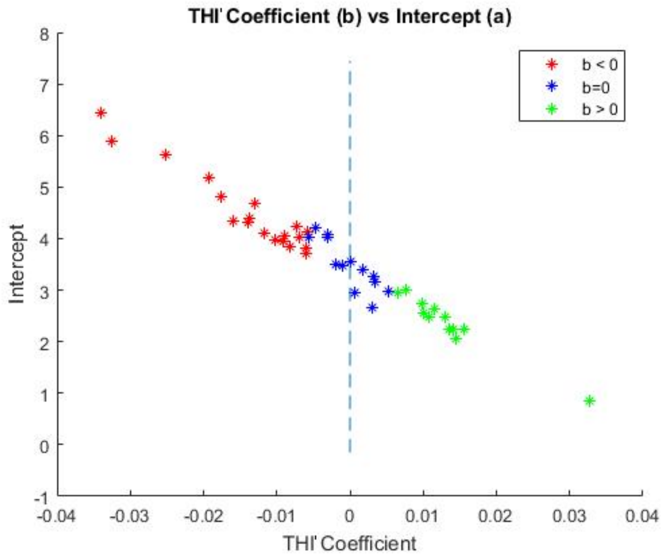

Figure 4.30: The plot shows the coefficients for each analyzed cow with THI' coefficient (b) in the x-axis and the intercept coefficient (a) in the y-axis. Every "*" is a cow, and it is green if $b > 0$, red if $b < 0$ and blue if it is close to 0 (30^{th} percentile).

Moreover, another group has been added: it represents cows with b very close to 0, selected in according to the 30^{th} percentile (blue group).

From the results in Table 4.12, it is clear that the correlation coefficients of "$b = 0$" and "$b > 0$" are very low, so these two groups are considered respectively as moderate and mild stressed by high values of THI.

It is easy to notice a negative correlation between the slope (b) and the intercept (a): it confirms the results of Yano et al. (2014), where it has been proved that the relatively highly productive cows are more sensitive to heat, therefore their production may decrease in the summer. Moreover, the highly productive cows have high body temperature and they could suffer more for the heat stress.

In particular, the trend of the milk production of some specific cows of the two main groups (three animals with $b < 0$ and three with $b > 0$) in relation to the THI' have been analyzed (Table 4.13).

Table 4.12: Statistics about the correlation (Bravais-Pearson) ρ between My and THI' in the three groups.

Group	$median(\rho)$	Stress
$b < 0$	-0.27	Severe
$b = 0$	-0.09	Moderate
$b > 0$	-0.05	Mild

Table 4.13: Correlation values between My and THI' of six selected cows: three from the group "$b < 0$" and three from "$b > 0$".

Cow	Correlation (Bravais-Pearson)
101 ($b < 0$)	-0.36
148 ($b < 0$)	-0.41
207 ($b < 0$)	-0.35
249 ($b > 0$)	0.07
236 ($b < 0$)	0.06
189 ($b < 0$)	0.09

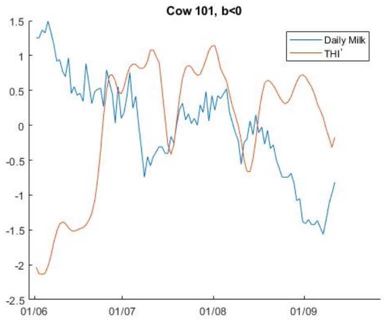

Figure 4.31: Z-scored My and THI' trends of cow 101 ($b < 0$) in the study period.

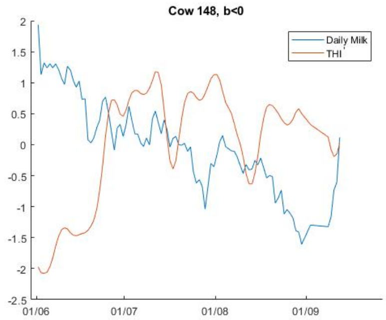

Figure 4.32: Z-scored My and THI' trends of cow 148 ($b < 0$) in the study period.

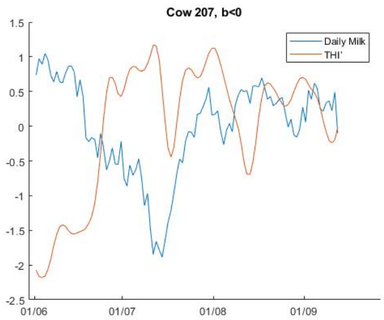

Figure 4.33: Z-scored My and THI' trends of cow 207 ($b < 0$) in the study period.

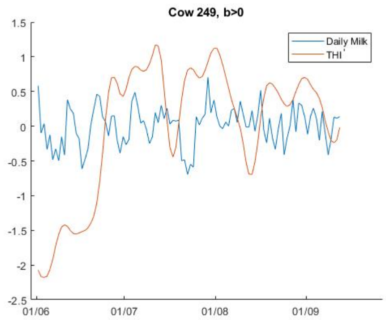

Figure 4.34: Z-scored My and THI' trends of cow 249 ($b > 0$) in the study period.

Moreover, in order to compare graphically these two variables, a z-score procedure has been applied on them before plotting in Figures from 4.31 to 4.36.

Finally, in the Figures from 4.37 to 4.50 the characteristics of the three groups are reported in terms of $d\alpha$, Box, $\#M$, $Cond$, Cw, My, Ld, Sf, My, Pa, Mr, $Temp$ and Vel.

The analyses highlight that the level of dependence of milk production upon THI in the herd under study is affected by parity.

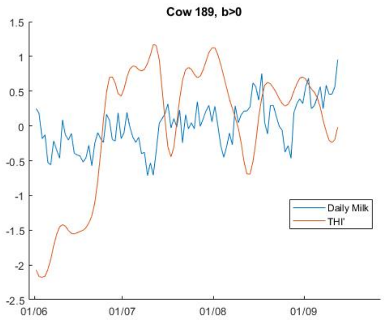

Figure 4.35: Z-scored My and THI' trends of cow 189 ($b > 0$) in the study period.

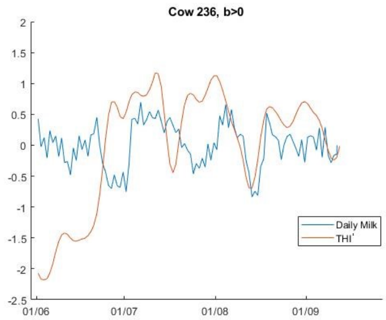

Figure 4.36: Z-scored My and THI' trends of cow 236 ($b > 0$) in the study period.

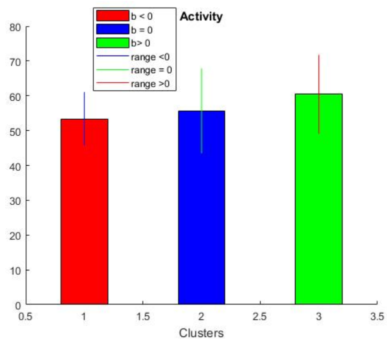

Figure 4.37: The activity values ($d\alpha$) of the three clusters.

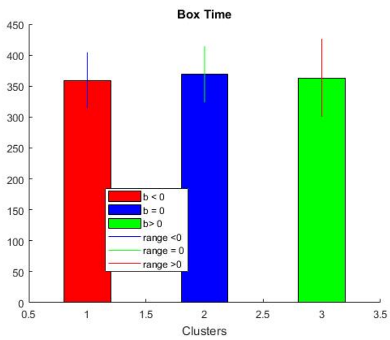

Figure 4.38: The Box Time (Box) of the three clusters.

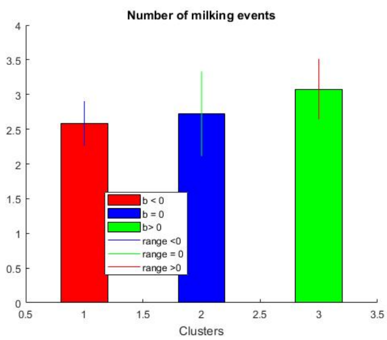

Figure 4.39: The number of daily milking events ($\#M$) of the three clusters.

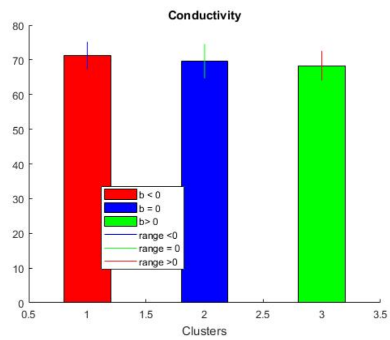

Figure 4.40: The Conductivity values ($Cond$) of the three clusters.

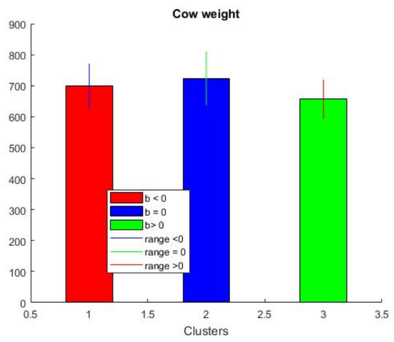

Figure 4.41: The cow weights (Cw) of the three clusters.

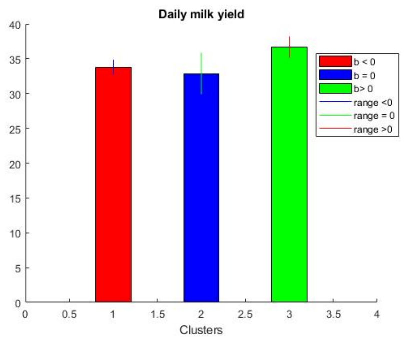

Figure 4.42: The Daily Milk Yield values (My) of the three clusters.

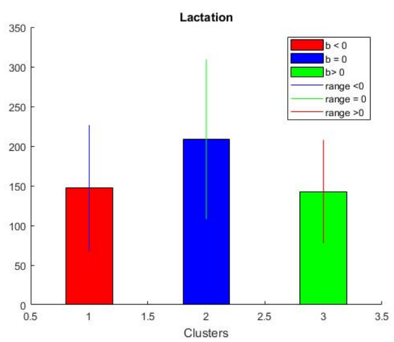

Figure 4.43: The days in lactation (Ld) of the three clusters.

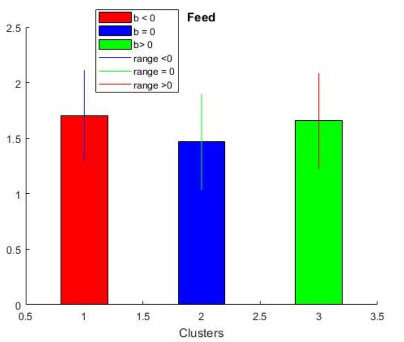

Figure 4.44: The supplementary feeding (Sf) of the three clusters.

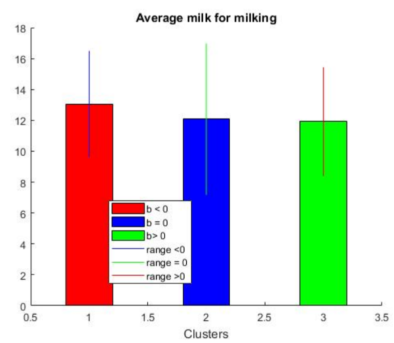

Figure 4.45: The average My for milking of the three clusters.

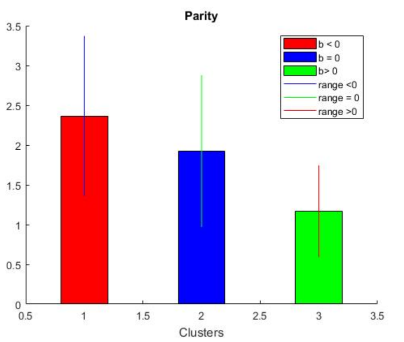

Figure 4.46: The parity values (Pa) of the three clusters.

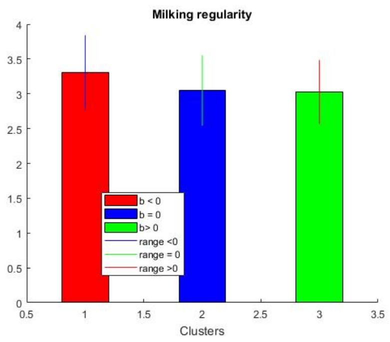

Figure 4.47: The milking regularity values (Mr) of the three clusters.

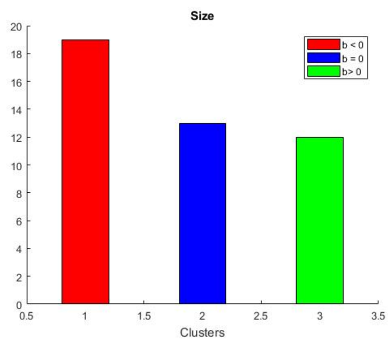

Figure 4.48: The size of the three clusters (severe / moderate / mild stress).

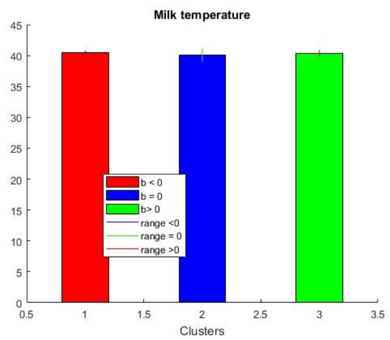

Figure 4.49: The milk ($Temp$) values of the three clusters.

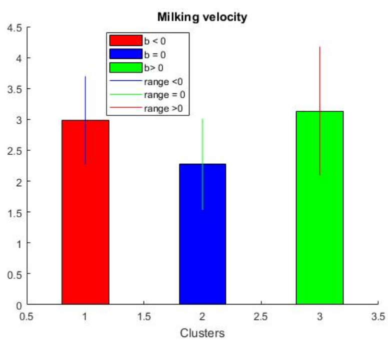

Figure 4.50: The milking velocity values (Vel) of the three clusters.

Furthermore the cows of the green group have in fact a smaller average weight and moreover they appear more active and have also more frequent milking events, with higher regularity than cows of the other groups.

Then, the production quantities of cows which are in the final part of the lactation curve appear to be independent on THI.

The most suffering cows (the red group) have the minimum activity score of the herd and their number of milking events per day is the lowest in the farm, and it also contains the oldest animals. On the other side, the green group shows high values for the activity and the number of daily milking, and it includes all the youngest cows. The heat stress has more effects on the animals which can be considered the weakest section of the herd, causing a reduction of the movements inside the barn and therefore also of milking events and production. This analysis represents an effective tool to optimize the management of the herd, identifying the most suffering cows in order to implement some specific actions in terms of cooling treatments, enhancement of the feeding properties, etc. on them.

4.3 Cow Detection

As described in the subsection 3.2.1, in the farm A a camera has been installed on a pillar in the middle of the barn in order to shoot six cubicles. This device has been studied with the aim of starting a preliminary research about the positions and the routines of the herd: the future plan is to collect information about the number of cows in resting area, in feeding area and in the milking, how many animals are standing and lying animals and other variables and indices connected to the welfare of the herd.

The idea is that if all these data are conveniently integrated with the information collected from collars, pedometers and milking system robot, they could represent a useful tool to manage the herd in a more optimized way.

In this work, 106 images are captured and selected from the videos recorded by the camera in different days:

- 97 (85 with and 12 without cows) for the training of the Viola-Jones model. In each pictures, the cows have been counted if they occupied entirely one of the six filmed cubicles;

- 9 for the testing phase.

Then, the precision and the sensibility of the Viola-Jones algorithm have been calculated evaluating how many cows in the cubicles have been identified by the model in the nine images.

Figures 4.51, 4.52, 4.53 and 4.54 show some results obtained testing the procedure, for a total of 31 cows to detect (test dataset).

In the pictures the positions of the cubicles have been schematized as the following matrix:

$$\begin{pmatrix} (1,1) & (1,2) \\ (2,1) & (2,2) \\ (3,1) & (3,2) \end{pmatrix}$$

The result in the Table 4.14 highlights only two cases of false positives (i.e. the algorithm indicates the presence of a cow, when it does not) and one false negative (i.e. a cow is not identified by the method), so it could be considered a promising field of research for future developments.

Figure 4.51: Example a): application of the Viola-Jones.

Figure 4.52: Example b): application of the Viola-Jones.

Figure 4.53: Example a): application of the Viola-Jones. In (2,2) position the cow does not occupy entirely the cubicle, so properly it is not hooped.

Figure 4.54: Example d): application of the Viola-Jones. In (1,1) position the cow occupies entirely the cubicles, so it is an error.

Table 4.14: Results of the object detection algorithm in the nine test-images.

Photo	Cows	True Positives	False Positives	False Negatives
1	3	3	1	0
2	5	5	1	0
3	0	0	0	0
4	3	3	0	0
5	1	1	0	0
6	5	4	0	1
7	4	4	0	0
8	4	4	0	0
9	6	6	0	0

It is important to underline that all the false positives are cows staying not entirely in a cubicle, so in order to develop the accuracy of the model, it will be probably sufficient to define more accurately which cows the algorithm has to count and which ones no adding new photos in the starting dataset.

Chapter 5

Conclusions

The principal phases of the PhD research could be summarized in the following three steps:

1. Planning of the data monitoring and acquisition in a dairy cattle farm. The data collected have concerned the milk production, the cow welfare and the thermal conditions in the three year of work;

2. Management and data processing of the files collected from the various devices;

3. Development, validation and test of numerical models, by applying different methods in order to analyze and characterize some particular cow behaviors with the aim of obtaining methods for an optimized management of the farm.

The results confirm the important contribution that could derive from a correct use of numerical models in a farm. Moreover, the procedures defined could be a suitable tool to fully use and take advantage of ICT devices and the consequent enormous amount of data and files available in modern barns. In particular:

- In the cluster-graph analysis, a dataset derived from AMS source has been considered. It contains the time series of the main parameters recorded for each cow in the summertime in an Italian barn (Seasonal Cow Clustering) and the first six months of 2016 (Monthly Cow Clustering).

In both of them, the combination of cluster analysis and graph theory has allowed to obtain objective results on the basis of statistical and mathematical methodologies. The herd has been characterized according to three main clusters, which proved diversified from one another in terms of productivity and animal behavior. The result depends obviously on the herd characteristics, but the methodology could be considered as general and thus suitable to be implemented in farms with different characteristics.

This study can lend support to cow monitoring, through the comparison between the actual and the expected cow behavior. In fact, herd partitioning could help to regulate the number of milking events or the supplementary feeding of animals belonging to specific groups and the identification of the clusters can contribute to define proper feeding strategies;

- An important segment of the thesis has been dedicated to the study of the herd productivity in relation to the environmental conditions, in particular in the summertime.

  In the first part of the study, the cow milking performances have been confirmed to be negatively affected by the heat stress, as indicated in literature, with a delay of five days after the occurrence of critical THI.

  Therefore, according to the results of the analysis of the influence of high THI levels on milk yield reduction, in the second part a model capable of forecasting milk yield in a farm in the warm season has been developed and tested on two farms with different herd sizes and milking systems, and it proved reliable within a predefined acceptable interval of variability. The forecasting model can be applied to assess the potential milk loss due to heat stress in a given climatic context and therefore it represents a suitable tool for a direct assessment of the expected benefit that can be achieved through proper investments for the control of the indoor climate of a dairy barn. It also could be implemented to study milk yield trends depending on the expected environmental conditions, thus it can be used also to predict milk production variation due to global climate change.

Finally, in the third and last subsection, the productive characteristics of the cows have been studied with a GAM analysis. The results could be useful to identify the most problematic animals for what concerns high values of THI and so to implement specific cooling strategies in order to mitigate their critical responses;

- The cow detection study has been implemented in order to analyze captured images of the barn with an object detection model (Viola-Jones algorithm). It is a pilot study, in which the results have confirmed the enormous potentiality of the automatic detection of the cow. If a highly quality system is implemented, i.e. low percentages of false positives and false negatives, this solution could allow to monitor the position of the animals with an approach that can be alternative or complementary to RFID schemes and have the potential to detect also behavioral patterns.

This work has also opened interesting points to additional and future tests. Further developments of the research are ongoing and firstly consist in the application of the models to other farms in different geographic contexts and under different climatic and farming conditions for a finer calibration.

Moreover:

- In the graph analysis, new variables can be added in the dataset, also through the use of further devices, with the aim of improving the consistency and effectiveness of the graph-clustering.

  Furthermore, another clustering approach may be tested (for example, multi-dimensional) in order to compare also the similarity in the trend of the analyzed variables.

Finally, the graph theory can be further developed by exploiting a broader potential of the network system (e.g. including direction of the edges, measures and metrics, ... and all the other features already used successfully in other fields, as described in the subsection 3.4.2);

- In the analysis of the relationship between cows and environmental conditions, in particular for what concerns the GAM model, the introduction of new devices and variables connected to the milk production and cow welfare could allow to characterize further the critical cows.

  Further insights include the monitoring of the trends and the values of the animals in the cluster "heat waves" (the red group in the subsection 4.2.3) and the "pre" and "post" analysis of the cooling actions.

  Finally, the developments of the research could focus also on the implications of the climatic conditions for innovations in the definition of spatial layouts, with expected benefits for the design of dairy barns;

- The cow detection method tested in this thesis is one of the most interesting and open question created by the work. The Viola Jones algorithm has shown the potentiality of the camera analysis and so it is clear that the implementation of an accurate and precise technique could allow to monitor the herd behavior in the barn (for example, percentage of animals in the feeding area, percentage of animals in resting area and percentage of animals in the milking area calculated in real time).

  Future developments include the positioning of cameras in the specific points in the barn and the systematic monitoring of the herd, but also the analysis of selected cows (drawing a number or another characteristic symbols on her body) in order to identify peculiar behavioral patterns in relationship with productivity and welfare scores.

  Moreover, new object detection models can be applied, e.g. with reference to Gao and Kasabov (2016) who applied spiking neural networks to detect the cow movement.

Finally, the integration of the results of the previous sections with this video position analysis can provide a better comprehensive behavioral description of dairy cattle in a farm.

Bibliography

K. Adamczyk, D. Cywicka, P. Herbut, and E. Trześniowska. The application of cluster analysis methods in assessment of daily physical activity of dairy cows milked in the Voluntary Milking System. *Computers and Electronics in Agriculture*, 141:65–72, 2017.

D.V. Armstrong. Heat Stress Interaction with Shade and Cooling. *Journal of Dairy Science*, 77(7):2044–2050, 1994.

Asabe. Design of Ventilation Systems for Poultry and Livestock Shelters. *ASAE*, 270.5:1–33, 1986.

A.L. Barabási. *Network Science*. Cambridge University Press, 2015.

A.L. Barabási and E. Bonabeau. Scale Free Networks. *Scientific American*, (May):50–59, 2003.

H.W. Barkema, M.A.G. von Keyserlingk, J.P. Kastelic, T.J.G.M. Lam, C. Luby, J.-P. Roy, S.J. LeBlanc, G.P. Keefe, and D.F. Kelton. Invited review: Changes in the dairy industry affecting dairy cattle health and welfare. *Journal of Dairy Science*, 98(11):7426–7445, 2015.

T. Bartzanas, B. Amon, S. Calvet, M. Mele, D. Morgavi, T. Norton, D. Yanez-Ruiz, and C. Vandongen. Reducing livestock emissions from Cattle farming Mini-paper â Precision Livestock Farming. Technical report, 2013.

M. Bastian, S. Heymann, and M. Jacomy. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Third International AAAI Conference on Weblogs and Social Media*, pages 361–362, 2009.

D. Berckmans. Precision livestock farming technologies for welfare management in intensive livestock systems. *Rev. sci. tech. Off. int. Epiz*, 33(1): 189–196, 2014.

D. Berckmans. General introduction to precision livestock farming. *Animal Frontiers*, 7(1):6, 2017.

P. Berdahl and S.E. Bretz. Preliminary survey of the solar reflectance of cool roofing materials. *Energy and Buildings*, 25(2):149–158, 1997.

U. Bernabucci, S. Biffani, L. Buggiotti, A. Vitali, N. Lacetera, and A. Nardone. The effects of heat stress in Italian Holstein dairy cattle. *Journal of Dairy Science*, 97(1):471–486, 2014.

J.P. Bikker, H.V. Laar, P. Rump, J. Doorenbos, K.V. Meurs, G.M. Griffioen, and J. Dijkstra. Technical note: Evaluation of an ear-attached movement sensor to record cow feeding behavior and activity. *Journal of Dairy Science*, 97(5):2974–2979, 2014.

J. Bohmanova, I. Misztal, S. Tsuruta, H.D. Norman, and T.J. Lawlor. Short communication: genotype by environment interaction due to heat stress. *Journal of Dairy Science*, 91:840–846, 2008.

R. Bouraoui, M. Lahmar, A. Majdoub, M. Djemali, and R. Belyea. The relationship of temperature-humidity index with milk production of dairy cows in a Mediterranean climate. *Animal Research*, 51(6):479–491, 2002.

R.A. Bucklin, L.W. Turner, D.K. Beede, D.R. Bray, and R.W. Hemken. Methods to relieve heat stress for dairy cows in hot, humid climates. *Applied Engineering in Agriculture*, 7(2):241–247, 1991.

Ö. Cangar, T. Leroy, M. Guarino, E. Vranken, R. Fallon, J. Lenehan, J. Mee, and D. Berckmans. Automatic real-time monitoring of locomotion and posture behaviour of pregnant cows prior to calving using online image analysis. *Computers and Electronics in Agriculture*, 64(1):53–60, 2008.

A. Castro, J.M. Pereira, C. Amiama, and M. Barrasa. Long-term variability of bulk milk somatic cell and bacterial counts associated with dairy farms moving from conventional to automatic milking systems. *Italian Journal of Animal Science*, 0(0):1–8, 2017.

CIGR Section II Working Group N 14 (Cattle Housing). Recommendations of Dairy Cow and Replacement Heifer Housing. Technical report, 2014.

R.J. Collier, R.M. Eley, A.K. Sharma, R.M. Pereira, and D.E. Buffington. Shade Management in Subtropical Environment for Milk Yield and Composition in Holstein and Jersey Cows. *Journal of Dairy Science*, 64(5): 844–849, 1981.

N.B. Cook, T.B. Bennett, and K.V. Nordlund. Monitoring Indices of Cow Comfort in Free-Stall-Housed Dairy Herds. *Journal of Dairy Science*, 88 (11):3876–3885, 2005.

CoolCows. Temperature Humidity Index (THI). *Web*, 2017. URL `http://www.coolcows.com.au/`.

P. Dalgaard. *Introductory Statistics with R.* Springer, New York, second edition, 2008.

K. de Koning. Automatic Milking - Common Practice on Dairy Farms. *The First North American Conference on Precision Dairy Management*, page 16, 2010.

K. de Koning and J. Rodenburg. Automatic milking: State of the art in Europe and North America. *Automatic milking: a better understanding. Wageningen Academic Publishers*, pages 27–37, 2004.

D. Dragovich. Effect of high temperature-humidity conditions on milk production of dairy herds grazed on farms in a pasture-based feed system. *International Journal of Biometeorology*, 23(1):15–20, 1979.

C.R. Eastwood, D.F. Chapman, and M.S. Paine. Networks of practice for co-construction of agricultural decision support systems: Case studies of precision dairy farms in Australia. *Agricultural Systems*, 108:10–18, 2012.

T. Gao and N. Kasabov. Adaptive cow movement detection using evolving spiking neural network models. *Evolving Systems*, 7(4):277–285, 2016.

A.B. Garcia, N. Angeli, L. Machado, F.C. de Cardoso, and F. Gonzalez. Relationships between heat stress and metabolic and milk parameters in dairy cows in southern Brazil. *Tropical Animal Health and Production*, 47 (5):889–894, 2015.

M. Gaworski, A. Leola, O. Sada, P. Kic, and J. Priekulis. Effect of cow traffic system and herd size on cow performance and automatic milking systems capacity. *Agronomy Research*, 14(1), 2016.

Gephi Consortium. Gephi Tutorial Layouts. 2011.

L. Gygax, I. Neuffer, C. Kaufmann, R. Hauser, and B. Wechsler. Comparison of functional aspects in two automatic milking systems and auto-tandem milking parlors. *Journal of Dairy Science*, 90(9):4265–4274, 2007.

K. Hagiya, K. Hayasaka, T. Yamazaki, T. Shirai, T. Osawa, Y. Terawaki, Y. Nagamine, Y. Masuda, and M. Suzuki. Effects of heat stress on production, somatic cell score and conception rate in Holsteins. *Animal Science Journal*, 88(1), 2017.

D.B. Haley, J. Rushen, and A.M. de Passillé. Behavioural indicators of cow comfort: activity and resting behaviour of dairy cows in two types of housing. *Canadian Journal of Animal Science*, 80(2):257–263, 2000.

K.D. Hart, B.W. McBride, T.F. Duffield, and T.J. DeVries. Effect of milking frequency on the behavior and productivity of lactating dairy cows. *Journal of Dairy Science*, 96(11):6973–6985, 2013.

T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning. *Bayesian Forecasting and Dynamic Models*, 1:1–694, 2009.

T. Hayashi and T. Kawamura. Effect of an automatic milking system on milking traits. *Animal Science Journal*, 73(5):403–408, 2002.

A.M. Heikkilä, L. Vanninen, and E. Manninen. Economics of Small-Scale Dairy Farms Having Robotic Milking. In *The First North American Conference on Precision Dairy Management*, 2010.

P.E. Hillman, C.N. Lee, and S.T. Willard. Thermoregulatory responses associated with lying and standing in heat-stressed dairy cows. *Transcations of the ASABE*, 48(2):795–801, 2005.

A. Huhtala, K. Suhonen, P. Mäkelä, M. Hakojärvi, and J. Ahokas. Evaluation of Instrumentation for Cow Positioning and Tracking Indoors. *Biosystems Engineering*, 96(3):399–405, 2007.

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning with Applications in R*. Springer, 2013.

A.J. John, C.E.F. Clark, M.J. Freeman, K.L. Kerrisk, S.C. Garcia, and I. Halachmi. Review: Milking robot utilization, a successful precision livestock farming evolution. *Animal*, 10(9), 2016.

C.T. Kadzere, M.R. Murphy, N. Silanikove, and E. Maltz. Heat stress in lactating dairy cows: A review. *Livestock Production Science*, 77(1):59–91, 2002.

J. Kaihilahti, A. Suokannas, and S. Raussi. Observation of Cow Behaviour in an Automatic Milking System using Web-based Video Recording Technology. *Biosystems Engineering*, 96(1):91–97, 2007.

E. Koeleman. 5 trends driving precision livestock farming. *Web*, 2017. URL http://www.allaboutfeed.net.

K.H. Leetaru. Culturomics 2.0: Forecasting Large-scale human behavior using global news media tone in time and space. *First Monday*, 16(9), 2011.

F. Lessire, J.L. Hornick, J. Minet, and I. Dufrasne. Rumination time, milk yield, milking frequency of grazing dairy cows milked by a mobile automatic system during mild heat stress. *Advances in Animal Biosciences*, 6 (01):12–14, 2015.

J.B. MacQueen. Kmeans Some Methods for classification and Analysis of Multivariate Observations. *5th Berkeley Symposium on Mathematical Statistics and Probability*, 1(233):281–297, 1967.

G. Mattachini, E. Riva, and G. Provolo. The lying and standing activity indices of dairy cows in free-stall housing. *Applied Animal Behaviour Science*, 129(1):18–27, 2011.

G. Mattachini, A. Antler, E. Riva, A. Arbel, and G. Provolo. Automated measurement of lying behavior for monitoring the comfort and welfare of lactating dairy cows. *Livestock Science*, 158(1-3):145–150, 2013.

S.S. Mote, D.S. Chauhan, and N. Ghosh. Effect of environment factors on milk production and lactation length under different seasons in crossbred cattle. *Indian Journal of Animal Research*, 50(2), 2016.

M. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–82, 2006.

L.P. Noldus, A.J. Spink, and R.A. Tegelenbosch. EthoVision: a versatile video tracking system for automation of behavioral experiments. *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc*, 33(3):398–414, 2001.

R. Oberschätzl-Kopp, B. Haidn, R. Peis, K. Reiter, and H. Bernhardt. Studies on dairy cow behaviour with automatic feeding in a herd milked by an AMS. *Landtechnik*, 71(2), 2016.

S.R. Otto and J.P. Denier. *An introduction to programming and numerical methods in MATLAB*. 2005.

M.W. Overton, W.M. Sischo, G.D. Temple, and D.A. Moore. Using Time-Lapse Video Photography to Assess Dairy Cattle Lying Behavior in a Free-Stall Barn. *Journal of Dairy Science*, 85(9):2407–2413, 2002.

K. Pavani, I. Carvalhais, M. Faheem, A. Chaveiro, F.V. Reis, and F.M. da Silva. Reproductive Performance of Holstein Dairy Cows Grazing in Dry-summer Subtropical Climatic Conditions: Effect of Heat Stress and Heat Shock on Meiotic Competence and In vitro Fertilization. *Asian-Australasian journal of animal sciences*, 28(3):334–42, 2015.

J.J. Poelarends, O.C. Sampimon, F. Neijenhuis, J.D.H.M. Miltenburg, J.E. Hillerton, J. Dearing, and C. Fossing. Cow factors related to the increase of somatic cell count after introduction of automatic milking. Technical report, 2004.

S.M.C. Porto, C. Arcidiacono, U. Anguzza, and G. Cascone. A computer vision-based system for the automatic detection of lying behaviour of dairy cows in free-stall barns. *Biosystems Engineering*, 115(2):184–194, 2013.

S.M.C. Porto, C. Arcidiacono, U. Anguzza, and G. Cascone. The automatic detection of dairy cow feeding and standing behaviours in free-stall barns by a computer vision-based system. *Biosystems Engineering*, 133:46–55, 2015.

D. Quammen. *Spillover: Animal Infections and the Next Human Pandemic*. W W Norton & Co Inc, 2012.

D.J. Reinemann. Robotic Milking : Current Situation. In *National Mastitis Council Annual Proceedings*, pages 75–80, 2008.

M.L. Rhoads, R.P. Rhoads, M.J. VanBaale, R.J. Collier, S.R. Sanders, W.J. Weber, B.A. Crooker, and L.H. Baumgard. Effects of heat stress and plane of nutrition on lactating Holstein cows: I. Production, metabolism, and aspects of circulating somatotropin. *Journal of Dairy Science*, 92(5): 1986–1997, 2009.

J. Rodenburg. Robotic milking: Technology, farm design, and effects on work flow. *Journal of Dairy Science*, 100:11715, 2017.

H. Roman-Ponce, W.W. Thatcher, D.E. Buffington, C.J. Wilcox, and H.H. Van Horn. Physiological and Production Responses of Dairy Cattle to a Shade Structure in a Subtropical Environment1. *Journal of Dairy Science*, 60(3):424–430, 1977.

C.E.B. Romanini, V. Exadaktylos, Q. Tong, I. Mcgonnel, T.G.M. Demmers, H. Bergoug, N. Eterradossi, N. Roulston, P. Garain, C. Bahr, and D. Berckmans. Monitoring the hatch time of individual chicken embryos. *Poultry Science*, 92(2):303–309, 2013.

L. Samal. Heat Stress in Dairy Cows - Reproductive Problems and Control Measures. *International Journal of Livestock Research*, 3(3), 2013.

M. Segnalini, A. Nardone, U. Bernabucci, A. Vitali, B. Ronchi, and N. Lacetera. Dynamics of the temperature-humidity index in the Mediterranean basin. *International Journal of Biometeorology*, 55(2):253–263, 2011.

M.S. Shahriar, D. Smith, A. Rahman, M. Freeman, J. Hills, R. Rawnsley, D. Henry, and G. Bishop-Hurley. Detecting heat events in dairy cows using accelerometers and unsupervised learning. *Computers and Electronics in Agriculture*, 128:20–26, 2016.

B. Shao and H. Xin. A real-time computer vision assessment and control of thermal comfort for group-housed pigs. *Computers and Electronics in Agriculture*, 62:15–21, 2008.

B. Sitkowska, D. Piwczynski, J. Aerts, and M. Waskowicz. Changes in milking parameters with robotic milking. *Archiv Tierzucht*, 58:137–143, 2015.

D.L. Smith, T. Smith, B.J. Rude, and S.H. Ward. Short communication: comparison of the effects of heat stress on milk and component yields and somatic cell score in Holstein and Jersey cows. *Journal of Dairy Science*, 96(5):3028–33, 2013.

W. Steeneveld, J.C.M. Vernooij, and H. Hogeveen. Effect of sensor systems for cow management on milk production, somatic cell count, and reproduction. *Journal of Dairy Science*, 98(6):3896–905, 2015.

P. Viola and M. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:511–518, 2001.

J.W. West. Effects of Heat-Stress on Production in Dairy Cattle. *Journal of Dairy Science*, 86(6):2131–2144, 2003.

S.N. Wood. *Generalized Additive Models: An Introduction with R.* Chapman and Hall/CRC, 2006.

M. Yano, H. Shimadzu, and T. Endo. Modelling temperature effects on milk production: a study on Holstein cows at a Japanese farm. *SpringerPlus*, 3 (1):129, 2014.

Acknowledgments

The study has been developed within research agreements between the Department of Agricultural Sciences of the University of Bologna and the dairy farms Piazzi (Budrio, BO, Italy), Montagnini (San Pietro in Casale, BO, Italy) and Aletti (Grontardo, CR, Italy).

The author wishes to thank the land reclamation consortium "Consorzio della Bonifica Renana" (in particular the Agro-Environmental and Irrigation Sector (Agricultural and Environmental Area)) for the support in providing climatic data.

Finally, the writer would like to express his sincere gratitude to Prof. Mirko Degli Esposti (University of Bologna) for his collaboration and support and Matti Pastell (senior research scientist at LUKE - Natural Resources Institute Finland, Helsinki) for his insightful comments, the supervision during the visiting Ph.D. period in Finland and feedback.